Evaluating Syntax-Driven Approaches to Phrase Extraction for MT

Ankit Srivastava[†] Sergio Penkale[†] Declan Groves[‡*] John Tinsley[†]

[†]CNGL/NCLT, Dublin City University, Ireland
[‡]Traslán, Co. Wicklow, Ireland
email: {*asrivastava,spenkale,jtinsley*}*@computing.dcu.ie*, *dgroves@traslan.ie*

Abstract

In this paper, we examine a number of different phrase segmentation approaches for Machine Translation and how they perform when used to supplement the translation model of a phrase-based SMT system. This work represents a summary of a number of years of research carried out at Dublin City University in which it has been found that improvements can be made using hybrid translation models. However, the level of improvement achieved is dependent on the amount of training data used. We describe the various approaches to phrase segmentation and combination explored, and outline a series of experiments investigating the relative merits of each method.

1 Introduction

Over the course of the last number of years, there has been a large body of research carried out aimed at investigating different methods of phrase segmentation for the training of data-driven machine translation (MT) models. A significant portion of this research took place in Dublin City University (DCU) during the development of the MaTrEx[1] MT system [23, 13].

Traditionally, phrase-based SMT (PB-SMT) models extract a set of phrase pairs based on a refined word aligned parallel corpus. This refined aligment comprises the intersection of a set of bidirectional word alignments, providing a high precision word alignment. This is then extended with heuristics [18] to obtain a final word alignment. All phrase pairs consistent with this word alignment are then extracted.

In this body of work, various approaches to phrase segmentation, beyond traditional PB-SMT methods, were examined. These include marker-based chunking, as used in example-based MT (EBMT) [4, 6], and more syntax-aware PB-SMT methods, for example, parallel treebank-based phrase extraction [25, 22]. In addition to this, several methods for the combination of the different phrase segmentation approaches were explored to investigate whether improvements could be made over a baseline PB-SMT system.

In this paper, we will describe in detail the various methodologies used for phrase segmentation and how they were exploited to build hybrid data-driven translation systems. At each stage we compare the performance of the hybrid translation systems against that of a baseline PBSMT system. We present a summary of the outcomes of the many experiments carried out, discussing in detail the advantages and shortcomings of each, and offer some insight into the potential of these methods going forward as the MT landscape continues to develop.

2 Marker-Based Chunking

In [28, 4], marker-based chunking is used in our work on EBMT. This approach to chunking is founded on the Marker Hypothesis [5], which suggests that all languages contain a closed set of specific lexemes

[*]This work was carried out during the author's time in DCU.
[1]Machine Translation using Examples

Mikel L. Forcada, Andy Way (eds.)
Proceedings of the 3rd Workshop on Example-Based Machine Translation, p. 19–28
Dublin, Ireland, November 2009

or morphemes that can be used to signal specific grammatical contexts. Chunking sentences whenever we encounter such ‘marker’ words allows us to extract syntactically-motivated chunks which can then be subsequently aligned to provide marker chunk pairs for use in translation.

For the experiments outlined in [6], seven marker categories were defined: determiners, quantifiers, prepositions, conjunctions, WH-adverbs, possessive pronouns and personal pronouns. During a pre-processing stage, the source–target aligned sentences in the parallel corpus are segmented at each occurrence of a marker word with the constraint that each marker chunk must contain at least one non-marker (or content) word, as illustrated in (1).

(1) <DET> that is almost <DET> a personal record <PREP> for <PRON> me <DET> this autumn!

$\Longrightarrow$<DET> c’ est pratiquement <DET> un record personnel <PREP> pour <PRON> moi , <DET> cet automne!

The resulting set of bilingual chunks are then aligned making use of marker tag and relative source and target chunk positions. Additionally, cognate and mutual information (MI) scores are used to help further inform the alignment process where poor scoring alignments are disregarded whenever a competing alignment has a higher MI score. The chunk alignments extracted from (1) are given in (2):

(2) a. <DET> that is almost : <DET> c’ est pratiquement

b. <DET> a personal record : <DET> un record personnel

c. <PREP> for me this autumn : <PREP> pour moi cet automne

The chunk pair in (2c) is an example of how the chunking constraint mentioned previously (that every marker chunk must contain at least one content word) comes into play, as segmentation is not performed at the marker words *me* and *this* in English, and *moi* and *cet* in French.

2.1 Combining Marker-Based Chunks & SMT Phrases – Pharaoh

A number of experiments were carried out directly combining marker-based EBMT chunks together with baseline SMT phrase pairs (extracted as described in section 1) in a PB-SMT system [6]. For these experiments, the MT system was built using the Pharaoh[2] decoder together with a trigram language model. The aim of these experiments was to investigate whether making use of the more syntactically-motivated EBMT-style phrases extracted via marker-based methods could improve system performance . During such direct combination approaches, the baseline PB-SMT phrases and the supplementary set of phrase pairs (in this case, the marker-based EBMT chunks) are directly combined in a single phrase table and their counts are combined to produce a new probability distribution. The intended effect here is to complement the baseline phrase table by introducing novel translation pairs while assigning increased probability mass to those hypothetically more reliable pairs found in the intersection of the two sets of phrase pairs.

Making use of French–English data taken from the Europarl corpus [9], MT experiments were carried out with and without the use of the EBMT chunks on training data sets of incremental sizes. It was found that including the EBMT chunks resulted in system improvements for both French–English and English–French tasks [6]. However, it was observed that the gains in system performance decreased as training data size increased, with relative improvements in BLEU score dropping from 24% to 19.6% for French–English and from 17.7% to 12.1% for English–French when comparing the results for the 78K training set against the 322K training set. The full set of results for these experiments are given in Table 1, which compares baseline system performance against a hybrid configuration combining both marker-based and SMT phrases for the various training data sets.

[2]http://www.isi.edu/licensed-sw/pharaoh/

		French–English			English–French		
Data	**System**	**BLEU**	**Prec.**	**Recall**	**BLEU**	**Prec.**	**Recall**
78K	Baseline	0.1943	0.5289	0.5477	0.1771	0.5045	0.4696
	Hybrid	0.2070	0.5368	0.5551	0.1898	0.5152	0.4787
156K	Baseline	0.2040	0.5369	0.5526	0.1855	0.5120	0.4724
	Hybrid	0.2176	0.5437	0.5579	0.1965	0.5208	0.4810
322K	Baseline	0.2102	0.5409	0.5539	0.1933	0.5180	0.4751
	Hybrid	0.2236	0.5483	0.5592	0.2040	0.5284	0.4851

Table 1: Comparing baseline and 'hybrid' PB-SMT systems for French–English and English–French, for training data sets of 78K, 156K and 322K sentence pairs.

Looking at the phrase pairs that were contributed via both approaches for the 322K-sentence training set, EBMT chunking introduced ∼747K new phrase pairs to the system, accounting for ∼13% of the total number of phrase pairs in the combined translation model and only ∼3% of the total number of phrase pairs were common to both the marker-based and SMT set of phrases.

2.2 Combining Marker-Based Chunks & SMT Phrases – Moses

In a more recent series of experiments, we incorporated marker-based phrases into the more sophisticated PB-SMT system implemented by Moses [10]. This system scores translations based on many features, namely: phrase translation probabilities and lexical weighting in both language directions, phrase penalty, lexicalised reordering probabilities and a language model. These features are combined in a log-linear model [17] whose weights are tuned for BLEU using Minimum Error Rate Training (MERT) [16]. For these experiments, as well as the experiments described in later sections of this paper, we made use of MERT-tuned Moses together with a 5-gram language model built using the SRILM toolkit.

When working with this system, the presence of features other than relative-frequency translation probabilities means the direct-combination approach taken by [6] might not be enough to allow these phrases to provide significant improvements in translation quality, thus, we explored different methods of incorporating this additional knowledge source. The methods we evaluated are:

- **n-count**: We add *n* copies of the extracted EBMT phrases to the baseline phrase tabel. In this way, we artificially increment the count of each EBMT phrase, increasing their phrase translation probability in the log-linear model.

- **feature**: Beginning with normal direct combination, we then add an additional binary feature to the log-linear model which indicates whether a phrase pair was EBMT-based or not. We assign 1 to EBMT phrases and 0 to the others and allow the MERT process to assign the relevant weight for this new feature.

Results for these experiments are given in Table 2. Along with translation quality measures, we report the percentage of phrases used during translation that were extracted by the marker-based method (EBMT%).

From these results, we can see that while the direct combination approach (1-count) increases the amount of marker-based phrases used during decoding, in contrast to [6], it fails to provide a significant improvement in translation quality. Further values for *n* yield similar results. On the other hand, allowing the system to benefit from the knowledge of which source each phrase comes from (Feature) leads to improvements in translation quality.

System	BLEU	NIST	METEOR	EBMT%
Baseline	0.3079	7.5590	0.6025	24.21
1-count	0.3078	7.5775	0.6024	23.47
2-count	0.3076	7.5582	0.6020	23.64
4-count	0.3071	7.5609	0.6015	24.34
8-count	0.3083	7.5969	0.6018	26.64
16-count	0.3042	7.5386	0.5986	29.71
Feature	0.3111	7.6004	0.6055	35.09

Table 2: Experiments carried out using 200k randomly selected sentences from the Spanish-English Europarl corpus, as provided for the Third Workshop on Statistical Machine Translation (WMT08). Testing was performed on the 2000 sentence testset provided for the workshop.

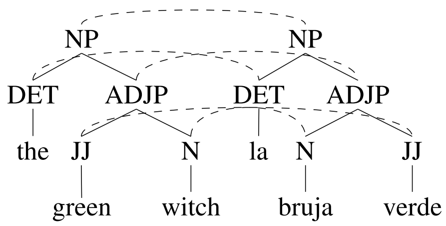
3 Parallel Treebank-Based Phrase Extraction

When working with parallel treebanks, our aligned tree pairs can be analysed using different grammar formalisms. Making use of such linguistic information draws heavily from EBMT research which makes explicit use of syntax during the creation of knowledge bases [21, 27]. Throughout the course of our work, we have exploited both context-free phrase-structure representations and dependency representations. We describe both approaches here.

3.1 Constituency-Based Extraction

The first step in producing a set of constituent-based phrase pairs is to build a parallel treebank. To do this, each side of a sententially aligned parallel corpus is monolingually parsed using available phrase-structure parsers, e.g. the Berkeley parser [19]. Following this, a set of sub-sentential alignments is introduced between the constituent nodes in each tree pairs[26, 30]. These alignments imply translational equivalence between the surface strings dominated by the aligned node pair, thus, given a fully aligned parallel treebank, a set of phrase pairs is extracted based on all node alignments. An example of this process is illustrated in Figure 1.

(a)



(b)

the green witch	↔	la bruja verde
green witch	↔	bruja verde
the	↔	la
green	↔	verde
bruja	↔	witch

Figure 1: Example of (a) a parallel treebank entry and (b) the set of extracted phrases.

A range of experiments were carried out using syntax-based phrase pairs extracted from parallel treebanks. These experiments focused on different techniques for supplementing the PB-SMT phrase tables with the syntax-based data. We describe the various methodologies below, along with results carried out from experiments on the English–Spanish Europarl corpus containing 726,891 training sentences pairs, from which we built a parallel treebank using the Berkeley parser [19] to parse the English side, and the parser of Bikel [1] to parse the Spanish. Again, all baseline components were built using Moses and translation accuracy was evaluated on a 1,000 sentence test set.

3.1.1 Direct Combination

We first employed the direct combination approach (as described previously in Section 2.1), the results of which are presented in Table 3.

System	BLEU	NIST	METEOR
Baseline	0.3341	7.0765	0.5739
+Syntax	**0.3397**	**7.0891**	0.5782
Syntax_only	0.3153	6.8187	0.5598
Syntax Prioritised	0.3339	6.9887	0.5723
Baseline Prioritised	0.3381	7.0835	**0.5789**

Table 3: Results of direct and prioritised combination translation experiments with syntax-based phrase pairs.

We see that supplementing the baseline model with syntax-based phrase pairs (Baseline+Syntax) significantly improved translation performance. As with the EBMT experiments of section 2, we attribute this to the introduction of novel phrase pairs to the model, which amounted to 16.79% of the total number of phrase pairs, plus the increased probability mass assigned to those common phrase pairs, which comprised 4.87% of the total number of phrase pairs. We also note that using only syntax-based phrase pairs (Syntax_only) achieves lower scores due to the considerably smaller phrase table size compared to the baseline: there were 24,708,527 and 6,432,771 baseline and syntax-based phrase pairs respectively.

An alternative to direct combination approach is to prioritise one set of phrase pairs during combination. Following this method, originally suggested in [7], given two sets of phrase pairs, for example A and B, we prioritise one set over the other. Assuming we have prioritised set A, when combining the two sets, we only add phrase pairs from set B if their source-side phrases are not already covered by some entries in A. The motivation behind this approach is that we may believe one set of phrase pairs to be more reliable than the other: the prioritised set. Thus, when the prioritised set provides a translation for a particular source phrase, we opt to trust it and not introduce further ambiguity from the other set of phrase pairs. Results using this method for combination can be seen in rows 4 and 5 of Table 3.

No improvements over the direct combination approach are achieved using prioritised combination, which is congruent with the findings of [7]. Given the benefits of direct combination as we discussed previously, we attribute our findings here to the absence of high-probability ”reliable“ phrase pairs found in the overlap between the two sets of phrase pairs.

3.1.2 Weighting Syntax-Based Data

Similar to the n-count experiments of section 2.2, we built a number of translation models in which we weighted the syntax-based based phrase pairs more and less heavily in a direct combination model. The results of this experiment are shown in Table 4.

System	BLEU	NIST	METEOR
Baseline+Syntax	0.3397	7.0891	0.5782
+Syntax x2	0.3386	7.0813	0.5776
+Syntax x3	0.3361	7.0584	0.5756
+Syntax x5	0.3377	7.0829	0.5771
Half-weights	**0.3404**	**7.1050**	**0.5792**

Table 4: Effect of increasing relative frequency of syntax-based phrase pairs in the direct combination model.

Weighting the syntax-based phrase pairs more heavily did not improve upon regular direct combination. We suspect this is due to increased likelihood assigned to lower quality syntax-based phrase pairs, as well as the more reliable ones, such as function word pairs. This issue is discussed in greater detail in [25].

In addition to this, we see a slight improvement when weighting the syntax-based phrase pairs less heavily. Intuitively, this model is similar to the baseline prioritised model in that it will most likely choose a baseline phrase pair where it exists, and default to syntax-based phrase pairs when no baseline pair exists. However, this model has the additional advantage of still further increasing the likelihood of phrase pairs in the intersection as we are not discarding any phrase pairs. It is this combination of factors that ultimately results in the further improved translation accuracy.

3.1.3 Investigating Influence of Training Set Size

Given the findings in all our previous experiments with syntax-based phrase pairs [24, 8, 13] and with marker-based chunks (cf. Sections 2.1 & 2.2), we found that the size of the training set had a significant influence of the effectiveness of direct combination. In order to investigate this further within a single experimental setup, we designed an experiment in which we increased the size of the training corpus incrementally and evaluated translation performance on a common test set. Our findings are presented in Figure 2.

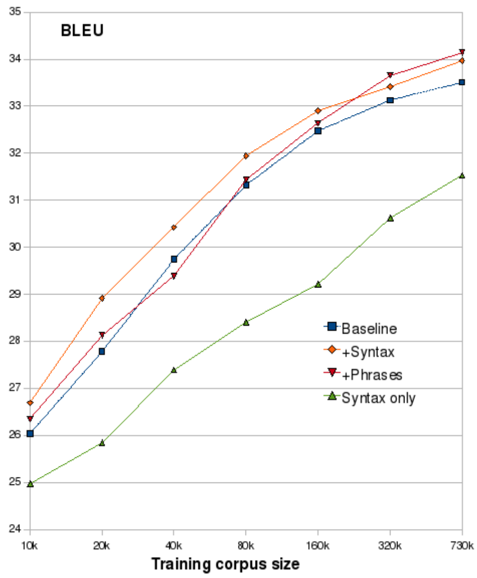


Figure 2: Effect of increasing training corpus size on influence of syntax-based phrase pairs.

We see the gains achieved over the baseline by adding the syntax-based phrase pairs steadily diminish as the training corpus grows. From our analysis in [25], we discovered that the complementary effect of the syntax-based phrase pairs in the combined model was reduced given larger data sets as the baseline phrase extraction learned many of the translation pairs that were originally unique to the syntax-based set.

3.2 Dependency-Based Extraction

Dependency-based phrase pairs, which have been demonstrated to be useful in pure EBMT [11], are extracted in much the same way as constituent-based phrase pairs described in the previous section. The main difference is that the training corpus is parsed with head-dependent relations between words of a sentence instead of constituent phrase structures. In this context, two techniques for dependency parsing have been explored:

- **Direct**: Dependency parse trees obtained by using off-the-shelf dependency parsers such as the Syntex parser [2] for both the source and target languages.
- **Indirect**: Dependency parse trees are obtained, using some heuristics, by exploiting lexical head information in constituency parse trees. A head percolation table [14] is used to select the head node in each constituent structure, which is then percolated, like features, from the head to its parent projection [29]. We label these **percolated dependencies** to distinguish them from the directly obtained dependency parses.

These structures are then converted into a labelled bracketed format (with arbitrary labels) in order to use the tree aligner [30] to obtain sub-sentential alignments between the nodes in each tree pair. This is done by composing a constituent by grouping a head and its dependents arranged as siblings in the order in which they occur in the sentence [8].

Although percolated dependencies are obtained from constituency parse trees and are structurally equivalent to directly obtained (unlabeled) dependencies, it was shown by [22] that the phrase pairs induced from these dependencies were divergent from both constituency- and direct dependency-based phrase pairs.

Experiments were carried out for the French–English translation task on different combinations of the baseline SMT phrase pairs and three syntax-based phrase pairs: CON (constituency-based as in Section 3.1), DEP (direct dependencies), and PERC (percolated dependencies). The BASELINE phrase pair extraction and the MT system setup was performed using Moses as described previously in this paper.

3.2.1 Phrase-table Combination

Implementing the direct combination strategy, 15 possible systems were created from the four types of phrase extractions, as shown in Table 5. The best-performing system combination was BASELINE+CON. The findings of section 3.1 concerning direct combination are also reflected here in the very strong baseline performance of the syntax-supplemented baseline systems (left-hand side of Table 5) against pure syntax-based systems (right-hand-side of Table 5). This is due to the relative differences in the size of these phrase tables; the CON t-table is just 31% of the size of the full BASELINE t-table, with DEP just 27% and PERC even smaller at just 26% of the size. Another similarity is that supplementing the BASELINE phrases with any syntax-based phrases (i.e. systems labelled ”B + ..”) results in improved translation performance across all three evaluation metrics.

While evaluating the number of unique phrase pairs in the four single phrase tables (B/C/D/P), it was interesting that despite the huge relative size of the BASELINE phrase table, there was very little overlap with any of the other methods; the largest overlap with BASELINE was using CON phrases, but this amounted to only 6% of the BASELINE phrase table being also derived via CON, and only 22% of the CON phrase table being also derived via BASELINE. The largest overlap in pure numerical terms was between CON and PERC; 74% of the CON phrase table are common with PERC, whereas 87% of the PERC phrase pairs are common with CON. A detailed study of this can be found in [22].

3.2.2 System Combination

An alternative to combining the phrase tables (either directly or via some prioritised weighting) is to use Minimum Bayes Risk and Confusion Network decoding (MBR-CN framework, [3]) to combine

SMT + Syntax				*Pure Syntax*			
System	**BLEU**	**NIST**	**METEOR**	**System**	**BLEU**	**NIST**	**METEOR**
BASELINE(B)	0.2850	7.00	0.5783				
B + C	**0.2950**	**7.10**	**0.5855**	CON(C)	0.2564	6.55	0.5526
B + D	0.2930	7.08	0.5843	DEP(D)	0.2524	6.59	0.5465
B + P	0.2945	**7.10**	0.5854	PERC(P)	0.2587	6.59	0.5563
B + C + D	0.2929	7.09	0.5848	C + D	0.2632	6.69	0.5556
B + C + P	0.2949	**7.10**	0.5850	C + P	0.2637	6.62	0.5605
B + D + P	0.2939	7.09	0.5849	D + P	0.2657	6.74	0.5583
B + C + D + P	0.2940	7.09	0.5849	C + D + P	0.2690	6.75	0.5614

Table 5: Summary of the results on Europarl French–English trained on 100,000 sentence pairs, tuned on 1,889 sentence pairs, and evaluated on 2,000 sentences

phrase pairs at the system level (after decoding) rather than at the phrase table level (during training). This was evaluated by combining the four base systems – BASELINE, CON, DEP, and PERC – as well as performing a system combination on the entire set of 15 systems. The results of these experiments are shown in Table 6. The results demonstrate that, when combining only the four systems (against the BASELINE, the best performing system in this sub-group), there is 7.16% relative improvement in BLEU score. Furthermore, when all 15 systems are passed through the Confusion Network (row 4 in Table 6, we see a 12.3% relative improvement in BLEU score. The improvements are reflected across all evaluation metrics. We attribute these gains to the fact that the translation output produced by the CON and PERC systems (which had the biggest overlap in their phrase pairs) has a 30% overlap only. Therefore, it seems that despite a huge overlap in the phrase table configurations, the systems are different enough to produce different translations. Consequently, the divergences between the phrase tables produced by the various phrase segmentation strategies can be successfully exploited using a system combination framework.

System	**BLEU**	**NIST**	**METEOR**
MBR (4 systems)	0.2952	6.85	0.5784
CN (4 systems)	0.3070	7.06	0.5852
MBR (15 systems)	**0.3260**	7.32	**0.6050**
CN (15 systems)	0.3251	**7.33**	0.6039

Table 6: Results of MBR-CN system combination on the systems in in Table 5.

4 Conclusions and Future Work

We have demonstrated that supplementing PB-SMT translation models with more syntactically aware EBMT-influenced methods, such as marker-based and treebank-based phrases, can result in improved translation accuracy. In addition to this, we have explored several strategies for combining syntax-based and baseline SMT phrase pairs. Of these strategies, confusion network-based system combination gives rise to the most significant improvements. However, in the case of other combination approaches, the gains achieved decrease as the size of the training set increases. This is due, in part, to the large coverage that SMT methods provide which ultimately diminishes the complementary effect of the syntax-based phrase pairs in the combined models. However, our methods for phrase combination can be particularly beneficial for languages with limited training resources [13], as well as for those MT systems with a smaller footprint requiring smaller phrase tables [20].

In conclusion, we believe it to be quite difficult to improve on the PB-SMT alignment/extraction/decoding pipeline without significant remodelling some components of the system. Unlike baseline phrase pairs, all of the alternative phrase extraction approaches described in this paper inherently encode some degree of syntactic information – for instance, marker-based generalised translation templates [28] and synchronous context-free grammar rules from parallel trees – which can not capitalised upon in baseline PB-SMT models. Going forward, in order to fully exploit the potential of our phrase extraction approaches at DCU, we need to utilise them in syntax-aware MT models, e.g. [31, 15, 12], which have previously demonstrated their capability to make use of such linguistically-enriched resources.

Acknowledgements

This work is supported by Science Foundation Ireland (Grant Nos. 07/CE/I1142 and 05/RF/CMS064). Thanks also to the reviewers for their comments and suggestions.

References

[1] D. Bikel. Design of a Multi-lingual, parallel-processing statistical parsing engine. In *Human Language Technology Conference (HLT)*, San Diego, CA, 2002.

[2] Didier Bourigault, Céile Fabre, Céile Fréot, Marie-Paule Jacques, and Sylwia Ozdowska. Syntex, Analyseur Syntaxique de Corpus. In *Atelier EASy (Évaluation des Analyseurs Syntaxiques), Actes de la Conféence Traitement Automatique des Langues Naturelles*, Dourdan, France, 2005.

[3] Jinhua Du, Yifan He, Sergio Penkale, and Andy Way. MATREX: The DCU MT system for WMT 2009. In *Proceedings of the Fourth Workshop on Statistical Machine Translation*, pages 95–99, Athens, Greece, March 2009. Association for Computational Linguistics.

[4] Nano Gough and Andy Way. Robust Large-Scale EBMT with Marker-Based Segmentation. In *Proceedings of the 10th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-04)*, pages 95–104, Baltimore, MD, 2004.

[5] T Green. The Necessity of Syntax Markers. Two experiments with artificial languages. *Journal of Verbal Learning and Behavior*, 18:481–496, 1979.

[6] Declan Groves and Andy Way. Hybrid Data-Driven Models of Machine Translation. *Machine Translation: Special Issue on Example-Based Machine Translation*, **19**(3-4):301–323, 2005.

[7] Greg Hanneman and Alon Lavie. ”decoding with syntactic and non-syntactic phrases in a syntax-based machine translation system”. In *Proceedings of the Third Workshop on Syntax and Structure in Statistical Translation (SSST-3)*, Boulder, CO., 2009.

[8] Mary Hearne, Sylwia Ozdowska, and John Tinsley. Comparing Constituency and Depedency Representations for SMT Phrase-Extraction. In *15ème Conférence sur le Traitement Automatique des Langues Naturelles (TALN ’08)*, Avignon, France, 2008.

[9] Philipp Koehn. Europarl: A Parallel Corpus for Statistical Machine Translation. In *Machine Translation Summit X*, pages 79–86, Phuket, Thailand, 2005.

[10] Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. Moses: Open Source Toolkit for Statistical Machine Translation. In *45th Annual Meeting of the Association for Computational Linguistics (ACL), demonstration session*, pages 177–180, Prague, Czech Republic, 2007.

[11] Philippe Langlais and Fabrizio Gotti. EBMT by Tree-Phrasing. *Machine Translation*, 20(1):1–23, 2006.

[12] Alon Lavie. Stat-XFER: A General Search-based Syntax-driven Framework for Machine Translation. In *Proceedings of thr 9th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-08) - Invited Paper*, pages 362–375, Haifa, Israel, February 2008.

[13] Yanjun Ma, John Tinsley, Hany Hassan, Jinhua Du, and Andy Way. Exploiting Alignment Techniques in MaTrEx: the DCU Machine Translation System for IWSLT08. In *Proc. of the International Workshop on Spoken Language Translation*, pages 26–33, Hawaii, USA, 2008.

[14] David Magerman. Statistical Decision-Tree Models for Parsing. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, pages 276–283, Cambridge, MA., 1995.

[15] Yuvol Marton and Philip Resnik. Soft Syntactic Constraints for Hierarchical Phrased-Based Translation. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (ACL'08)*, pages 1003–1011, Columbus, Ohio, 2008.

[16] Franz Josef Och. Minimum Error Rate Training in Statistical Machine Translation. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 160–167, Sapporo, Japan, 2003.

[17] Franz Josef Och and Hermann Ney. Discriminative Training and Maximum Entropy Models for Statistical Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 295–302, Philadelphia, PA., 2002.

[18] Franz Josef Och, Christoph Tillmann, and Hermann Ney. Improved Alignment Models for Statistical Machine Translation. In *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*, pages 20–28, University of Maryland, College Park, MD, 1999.

[19] Slav Petrov and Dan Klein. Improved Inference for Unlexicalized Parsing. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics*, pages 404–411, Rochester, NY, April 2007.

[20] Felipe Sánchez-Martínez and Andy Way. Marker-based Filtering of Bilingual Phrase Pairs for SMT. In *Proceedings of the 13th Annual Meeting of the European Association for Machine Translation (EAMT-09)*, pages 144–151, Barcelona, Spain, 2009.

[21] Satoshi Sato and Makoto Nagao. Toward Memory-Based Translation. In *Proceedings of the 13th Internation Conference on Computational Linguistics (COLING'90)*, pages 247–252, Helsinki, Finland, 1990.

[22] Ankit Srivastava and Andy Way. Using Percolated Dependencies for Phrase Extraction in SMT. In *Machine Translation Summit XII*, pages 316–323, Ottawa, Canada, 2009.

[23] Nicolas Stroppa and Andy Way. MaTrEx: the DCU machine translation system for IWSLT 2006. In *Proceedings of the International Workshop on Spoken Language Translation*, pages 31–36, Kyoto, Japan, 2006.

[24] John Tinsley, Mary Hearne, and Andy Way. Exploiting Parallel Treebanks to Improve Phrase-Based Statistical Machine Translation. In *Proceedings of the Sixth International Workshop on Treebanks and Linguistic Theories (TLT-07)*, pages 175–187, Bergen, Norway, 2007.

[25] John Tinsley and Andy Way. Automatically-Generated Parallel Treebanks and their Exploitability in Phrase-Based Statistical Machine Translation. *Machine Translation*, 15(4):(in press), 2009.

[26] John Tinsley, Venstislav Zhechev, Mary Hearne, and Andy Way. Robust Language-Pair Independent Sub-Tree Alignment. In *Machine Translation Summit XI*, pages 467–474, Copenhagen, Denmark, 2007.

[27] H. Watanabe, S. Kurohashi, and E. Aramaki. Finding Translation Patterns from Paired Source and Target Dependency Structures. In Michael Carl and Andy Way, editors, *Recent Advances in Example-Based Machine Translation*, pages 397–420. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2003.

[28] Andy Way and Nano Gough. wEBMT: Developing and Validating an Example-Based Machine Translation System using the World Wide Web. *Computational Linguistics: Special Issue on the Web as Corpus*, **29**(3):421–458, 2003.

[29] Fei Xia and Martha Palmer. Converting Dependency Structures to Phrase Structures. In *Proceedings of the Human Language Technology Conference (HLT '01)*, pages 1–5, San Diego, CA., 2001.

[30] Ventsislav Zhechev and Andy Way. Automatic Generation of Parallel Treebanks. In *Proceedings of the 22nd International Conference on Computational Linguistics (CoLing'08)*, pages 1105–1112, Manchester, UK, 2008.

[31] Andreas Zollmann and Ashish Venugopal. Syntax Augmented Machine Translation via Chart Parsing. In *Proceedings of the Workshop on Statistical Sachine Translation: HLT–NAACL 2006*, page 138–141, New York, NY, 2006.