

Decoding with Syntactic and Non-Syntactic Phrases in a Syntax-Based Machine Translation System

Greg Hanneman and Alon Lavie

Language Technologies Institute

Carnegie Mellon University

Pittsburgh, PA 15213 USA

{ghannema, alavie}@cs.cmu.edu

Abstract

A key concern in building syntax-based machine translation systems is how to improve coverage by incorporating more traditional phrase-based SMT phrase pairs that do not correspond to syntactic constituents. At the same time, it is desirable to include as much syntactic information in the system as possible in order to carry out linguistically motivated reordering, for example. We apply an extended and modified version of the approach of Tinsley et al. (2007), extracting syntax-based phrase pairs from a large parallel parsed corpus, combining them with PBSMT phrases, and performing joint decoding in a syntax-based MT framework without loss of translation quality. This effectively addresses the low coverage of purely syntactic MT without discarding syntactic information. Further, we show the potential for improved translation results with the inclusion of a syntactic grammar. We also introduce a new syntax-prioritized technique for combining syntactic and non-syntactic phrases that reduces overall phrase table size and decoding time by 61%, with only a minimal drop in automatic translation metric scores.

1 Introduction

The dominance of traditional phrase-based statistical machine translation (PBSMT) models (Koehn et al., 2003) has recently been challenged by the development and improvement of a number of new models that explicitly take into account the syntax of the sentences being translated. One simple approach is to limit the phrases learned by a standard

PBSMT translation model to only those contiguous sequences of words that additionally correspond to constituents in a syntactic parse tree. However, a total reliance on such syntax-based phrases has been shown to be detrimental to translation quality, as the space of phrase segmentation of a parallel sentence is heavily constrained by both the source-side and target-side tree structures. Noting that the number of phrase pairs extracted from a corpus is reduced by around 80% when they are required to correspond to syntactic constituents, Koehn et al. (2003) observed that many non-constituent phrase pairs that would not be included in a syntax-only model are in fact extremely important to system performance. Since then, researchers have explored effective ways for combining phrase pairs derived from syntax-aware methods with those extracted from more traditional PBSMT. Briefly stated, the goal is to retain the high level of coverage provided by non-syntactic PBSMT phrases while simultaneously incorporating and exploiting specific syntactic knowledge.

Zollmann and Venugopal (2006) overcome the restrictiveness of the syntax-only model by starting with a complete set of phrases as produced by traditional PBSMT heuristics, then annotating the target side of each phrasal entry with the label of the constituent node in the target-side parse tree that subsumes the span. They then introduce new constituent labels to handle the cases where the phrasal entries do not exactly correspond to the syntactic constituents. Liu et al. (2006) also add non-syntactic PBSMT phrases into their tree-to-string translation system. Working from the other direction, Marton and Resnik (2008) extend a hierarchical PBSMT

system with a number of features to prefer or disprefer certain types of syntactic phrases in different contexts. Restructuring the parse trees to ease their restrictiveness is another recent approach: in particular, Wang et al. (2007) binarize source-side parse trees in order to provide phrase pair coverage for phrases that are partially syntactic.

Tinsley et al. (2007) showed an improvement over a PBSMT baseline on four tasks in bidirectional German–English and Spanish–English translation by incorporating syntactic phrases derived from parallel trees into the PBSMT translation model. They first word align and extract phrases from a parallel corpus using the open-source Moses PBSMT toolkit (Koehn et al., 2007), which provides a baseline SMT system. Then, both sides of the parallel corpus are parsed with independent automatic parsers, subtrees from the resulting parallel treebank are aligned, and an additional set of phrases (with each phrase corresponding to a syntactic constituent in the parse tree) is extracted. The authors report statistically significant improvements in translation quality, as measured by a variety of automatic metrics, when the two types of phrases are combined in the Moses decoder.

Our approach in this paper is structurally similar to that of Tinsley et al. (2007), but we extend or modify it in a number of key ways. First, we extract both non-syntactic PBSMT and syntax-driven phrases from a parallel corpus that is two orders of magnitude larger, making our system competitive in size to state-of-the-art SMT systems elsewhere. Second, we apply a different algorithm for subtree alignment, proposed by Lavie et al. (2008), which proceeds bottom-up from existing statistical word alignments, rather than inducing them top-down from lexical alignment probabilities. Third, in addition to straightforwardly combining syntax-derived phrases with traditional PBSMT phrases, we demonstrate a new combination technique that removes PBSMT phrases whose source-language strings are already covered by a syntax-derived phrase. This new syntax-prioritized technique results in a 61% reduction in the size of the combined phrase table with only a minimal decrease in automatic translation metric scores. Finally, and crucially, we carry out the joint decoding over both syntactic and non-syntactic phrase pairs in a syntax-aware MT sys-

tem, which allows a syntactic grammar to be put in place on top of the phrase pairs to carry out linguistically motivated reordering, hierarchical decoding, and other operations.

After this introduction, we first describe the base MT system we used, its formalism for specifying translation rules, and the method for extracting syntax-derived phrase pairs from a parallel corpus (Section 2). Section 3 gives the two methods for combining PBSMT phrases with our syntactic phrases, and introduces our first steps with including a grammar in the syntax-based translation framework. The results of our experiments are described in Section 4 and are further discussed in Section 5. Finally, Section 6 offers some conclusions and directions for future work.

2 Base Translation System

The base MT system used for our experiments is the statistical transfer (“Stat-XFER”) framework (Lavie, 2008). The core of the framework is a transfer engine using two language-pair-dependent resources: a grammar of weighted synchronous context-free rules, and a probabilistic bilingual lexicon. Once the resources have been provided, the Stat-XFER framework carries out translation in a two-stage process, first applying the lexicon and grammar to synchronously parse an input sentence, then running a monotonic decoder over the resulting lattice of scored translation pieces assembled during parsing to produce a final string output. Reordering is applied only in the first stage, driven by the syntactic grammar; the second-stage monotonic decoder only assembles translation fragments into complete hypotheses.

2.1 Lexicon and Grammar Formalism

Each Stat-XFER bilingual lexicon entry has a synchronous context-free grammar (SCFG) expression of the source- and target-language production rules, shown in abbreviated format below, where c_s and c_t represent source- and target-side syntactic category labels and w_s and w_t represent source- and target-side word or phrase strings.

$$c_s :: c_t \rightarrow [w_s] :: [w_t]$$

Each entry in the lexicon is assigned a pair of rule scores ($r_{t|s}$ and $r_{s|t}$) based on c_s , w_s , c_t , and w_t ¹. The $r_{t|s}$ score is a maximum-likelihood estimate of the distribution of target-language translations and source- and target-language syntactic categories given the source string (Equation 1); this is similar to the usual “target-given-source” phrasal probability in standard SMT systems. The $r_{s|t}$ score is similar, but calculated in the reverse direction to give a source-given-target probability (Equation 2).

$$r_{t|s} = \frac{\#(w_t, c_t, w_s, c_s)}{\#(w_s) + 1} \quad (1)$$

$$r_{s|t} = \frac{\#(w_t, c_t, w_s, c_s)}{\#(w_t) + 1} \quad (2)$$

The add-one smoothing in the denominators counteracts overestimation of the rule scores of lexical entries with very infrequent source or target sides.

Stat-XFER grammar rules have a similar form, shown below via an example.

NP :: NP \rightarrow [DET¹ N² de N³] :: [DET¹ N³ N²]

The SCFG backbone may include lexicalized items, as above, as well as non-terminals and pre-terminals from the grammar. Constituent alignment information, shown here as co-indexes on the non-terminals, specifies one-to-one correspondences between source-language and target-language constituents on the right-hand side of the SCFG rule. Rule scores $r_{t|s}$ and $r_{s|t}$ for grammar rules, if they are learned from data, are calculated in the same way as the scores for lexical entries.

2.2 Syntax-Based Phrase Extraction

In this section, we briefly summarize the automatic resource extraction approach described by Lavie et al. (2008) and recently extended by Ambati and Lavie (2008), which we use here, specifically as applied to the extraction of syntax-based phrase pairs for the bilingual lexicon.

The grammar and lexicon are extracted from a large parallel corpus that has been statistically word-aligned and independently parsed on both sides with

¹If no syntactic category information is available, c_s and c_t can be set to dummy values, but the rule score equations remain unchanged.

automatic parsers. Word-level entries for the bilingual lexicon are directly taken from the word alignments; corresponding syntactic categories for the left-hand side of the SCFG rules are obtained from the preterminal nodes of the parse trees. Phrase-level entries for the lexicon are based on node-to-node alignments in the parallel parse trees. In the straightforward “tree-to-tree” scenario, a given node n_s in one parse tree S will be aligned to a node n_t in the other parse tree T if the words in the yield of n_s are all either aligned to words within the yield of n_t or have no alignment at all. If there are multiple nodes n_t satisfying this constraint, the node in the tree closest to the leaves is selected. Each aligned node pair (n_s, n_t) produces a phrase-level entry in the lexicon, where the left-hand sides of the SCFG rule are the labels of n_s and n_t , and the right-hand sides are the yields of those two nodes in their respective trees. In the expanded “tree-to-tree-string” configuration, if no suitable node n_t exists, a new node n'_s is introduced into T as a projection of n_s , spanning the yield of the words in T aligned to the yield of n_s . At the end of the extraction process in either case, the entry counts are collected and scored in the manner described in Section 2.1.

3 Combination with PBSMT Phrases

Conceptually, we take the opposite approach to that of Tinsley et al. (2007) by adding traditional PBSMT phrases into a syntax-based MT system rather than the other way around. We begin by running steps 3 through 5 of the Moses training script (Koehn et al., 2007)², which results in a list of phrase pair instances for the same word-aligned corpus to which we applied the syntax-based extraction methods in Section 2.2. Given the two sets of phrases, we explore two methods of combining them.

- **Direct Combination.** Following the method of Tinsley et al. (2007), we directly combine the counts of observed syntax-based phrase pairs with the counts of observed PBSMT phrase pairs. This results in a modified probability model in which a higher likelihood is moved onto syntactic phrase pairs that were also extractable using traditional PBSMT heuristics. It

²See also www.statmt.org/moses.

Decoder	Phrase Type	# Phrases	METEOR	BLEU	TER
Stat-XFER	Syntactic only, PHR	917,266	0.5654	0.2734	56.49
Stat-XFER	Syntactic only, frag	1,081,233	0.5653	0.2741	56.54
Stat-XFER	Syntactic only, gra	1,081,233	0.5665	0.2772	56.26
Stat-XFER	PBSMT only	8,069,480	0.5835	0.3018	54.26
Stat-XFER	Direct combination, PHR	8,071,773	0.5835	0.3009	54.21
Stat-XFER	Direct combination, frag	9,150,713	0.5841	0.3026	54.52
Stat-XFER	Direct combination, gra	9,150,713	0.5855	0.3034	54.28
Stat-XFER	Syntax-prioritized, PHR	2,888,154	0.5800	0.2961	54.79
Stat-XFER	Syntax-prioritized, frag	3,052,121	0.5802	0.2979	54.78
Stat-XFER	Syntax-prioritized, gra	3,052,121	0.5813	0.2991	54.73
Moses	PBSMT only, mono	8,145,083	0.5911	0.3139	53.77
Moses	PBSMT only, lex RO	8,145,083	0.5940	0.3190	53.48

Figure 1: Results on the test set for all phrase table configurations. For BLEU, bold type indicates the best Stat-XFER baseline and the configurations statistically equivalent to it (paired bootstrap resampling with $n = 1000$, $p = 0.05$).

also allows either extraction mechanism to introduce new entries into the combined phrase table that were not extracted by the other, thus permitting the system to take full advantage of complementary information provided by PBSMT phrases that do not correspond to syntactic constituents.

- **Syntax-Prioritized Combination.** Under this method, we take advantage of the fact that syntax-based phrase pairs are likely to be more precise translational equivalences than traditional PBSMT phrase pairs, since constituent boundaries are taken into account during phrase extraction. PBSMT phrases whose source-side strings are already covered by an entry from the syntactic phrase table are removed; the remaining PBSMT phrases are combined as in the direct combination method above. The effect on the overall system is to trust the syntactic phrase pairs in the cases where they exist, supplementing with PBSMT phrase pairs for non-constituents.

For each type of phrase-pair combination, we test three variants when jointly decoding syntax-based phrases, which come with syntactic information, along with PBSMT phrases, which do not. In the first configuration (“PHR”), all extracted phrase labels for syntactic phrases are mapped to a generic “PHR” tag to simulate standard SMT monotonic de-

coding; this matches the treatment given throughout to our extracted non-syntactic phrases. In the second variant (“frag”), the phrase labels in the large nonterminal sets used by our source- and target-side parsers are mapped down to a smaller set of 19 labels that we use for both sides. The same translation phrase pair may occur with multiple category labels in this case if it was extracted with different syntactic categories from different trees in the corpus. In a third variant (“gra”), a small manually developed grammar is additionally inserted into the system. The Stat-XFER system behaves the same way in each variant. All phrase pairs are applied jointly to the input sentence during the parsing stage, getting added to the translation according to their syntactic category and scores, although phrases tagged as PHR cannot participate in any grammar rules. The second-stage decoder then receives the joint lattice and assembles complete output hypotheses regardless of syntactic category labels.

4 Experiments

We extracted the lexical resources for our MT system from version 3 of the French–English Europarl parallel corpus (Koehn, 2005), using the officially released training set from the 2008 Workshop in Statistical Machine Translation (WMT)³. This gives us a corpus of approximately 1.2 million sentence

³www.statmt.org/wmt08/shared-task.html

Phrase Table	# Entries	# Source Sides	Amb. Factor
Total syntax-prioritized table	3,052,121	113,988	26.8
Syntactic component	1,081,233	39,105	27.7
PBSMT component	1,970,888	74,883	26.3
Total baseline PBSMT table	8,069,480	113,972	70.8
Overlap with syntax-prioritized	6,098,592	39,089	156.0

Figure 2: Statistical characteristics of the syntax-prioritized phrase table (top) compared with the baseline PBSMT phrase table (bottom). The ambiguity factor is the ratio of the number of unique entries to the number of unique source sides, or the average number of target-language alternatives per source phrase.

pairs. Statistical word alignments are learned in both directions with GIZA++ (Och and Ney, 2003), then combined with the “grow-diag-final” heuristic. For the extraction of syntax-based phrase pairs, we obtain English-side constituency parses using the Stanford parser (Klein and Manning, 2003), and French-side constituency parses using the Xerox XIP parser (Aït-Mokhtar et al., 2001). In phrase extraction, we concentrate on the expanded tree-to-tree-string scenario described in Section 2.2, as it results in a nearly 50% increase in the number of extracted phrase pairs over the tree-to-tree method. For decoding, we construct a suffix-array language model (Zhang and Vogel, 2006) from a corpus of 430 million words, including the English side of our training data, the English side of the Hansard corpus, and newswire data. The “gra” variant uses a nine-rule grammar that is meant to address the most common low-level reorderings between French and English, focusing mainly on the reordering between nouns or noun phrases and adjectives or adjective phrases.

Our test set is the 2000-sentence “test2007” data set, also released as part of the WMT workshop series. We report case-insensitive scores on version 0.6 of METEOR (Lavie and Agarwal, 2007) with all modules enabled, version 1.04 of IBM-style BLEU (Papineni et al., 2002), and version 5 of TER (Snover et al., 2006).

Figure 1 gives an overall summary of our results on the test2007 data. Overall, we train and test 10 different configurations of phrase pairs in the Stat-XFER decoder. We begin by testing each type of phrase separately, producing one set of baseline systems with only phrase pairs that correspond to syntactic constituents (“Syntactic only”) and one baseline system with only phrase pairs that were ex-

tracted from Moses (“PBSMT only”). We then test our two combination techniques, and their variants, as described in Section 3. Statistical significance is tested on the BLEU metric using paired bootstrap resampling (Koehn, 2004) with $n = 1000$ and $p = 0.05$. In the figure, the best baseline system and the configurations statistically equivalent to it are indicated in bold type. In addition to automatic metric scores, we also list the number of unique phrase pairs extracted for each configuration. (Because of the large number of phrase pairs, we pre-filter them to only the set whose source sides appear in the test data; these numbers are the ones reported.)

As an additional point of comparison, we build and tune a Moses MT system on the same data as our Stat-XFER experiments. The Moses system with a 4-gram language model and a distance-6 lexical reordering model (“lex RO”) scores similarly to state-of-the-art systems of this type on the test2007 French–English data (Callison-Burch et al., 2007). Without the reordering model (“mono”), the Moses system is as comparable as possible in design and resources to the Stat-XFER PBSMT-only configuration. We do not propose in this paper a head-to-head performance comparison between the Stat-XFER and Moses decoders; rather, we report results on both to gain a better understanding of the impact of the non-syntactic lexical reordering model in Moses compared with the impact of the syntactic grammar in Stat-XFER.

5 Discussion

5.1 Phrasal Coverage and Precision

One observation apparent in Figure 1 is that we have again confirmed that a total restriction to syntax-

Source:	Il faut que l' opinion publique soit informée pleinement sur les caractéristiques du test dont je parle .
Reference:	Public opinion must be fully informed of the characteristics of the test I am talking about .
Syntax only:	It is that the public be informed fully on the characteristics of the test I am talking about .
PBSMT only:	We must that public opinion gets noticed fully on the characteristics of the test above .
Direct comb.:	We must that public opinion gets noticed fully on the characteristics of the test above .
Syntax-prioritized:	It is important that <i>the public</i> <i>be informed</i> fully on <i>the characteristics</i> <i>of the test</i> <i>I am talking about</i> .

Figure 3: A translation example from the test set showing the output’s division into phrases. In the syntax-prioritized translation, English phrases that derived from syntax-based phrasal entries are shown in italics.

based phrases is detrimental to output quality. A likely reason for this, as Tinsley et al. (2007) suggested, is that the improved precision and informativeness of the syntactic phrases is not enough to overcome their relative scarcity when compared to non-syntactic PBSMT phrases. (The syntactic phrase table is only 11 to 13% of the size of the PBSMT phrase table.) It is important to note that this scarcity occurs at the *phrasal* level: though there are 294 unknown word types in our test set when translating with only syntactic phrase pairs, this number only drops to 277 with the inclusion of PBSMT phrases. The largest phrase table configuration, direct combination, yields statistically equivalent performance to the baseline system created using standard PBSMT extraction heuristics. Its key benefit is that the inclusion of syntactic information in the phrase pairs, where possible, leaves open the door to further improvement in scores with the addition of a larger syntactic grammar. We have thus addressed the syntax-only phrase coverage problem without giving up syntactic information.

An interesting conclusion is revealed in the analysis of the sizes and relative overlaps of the phrase tables in each of our translation conditions. In the absence of significant grammar, the equivalence of scores between the PBSMT-only and direct-combination scenarios is understandable given the minimal change in the size of the phrase table. Out of nearly 8.1 million entries, only 2293 entirely new

entries are provided by adding the syntactic phrase table; further, these phrases are relatively rare long phrases that do not have much effect on the translation of the overall test set. On the other hand, the syntax-prioritized phrase table is extremely different in nature — and only 37.8% of the size of the baseline PBSMT phrase table — yet still attains nearly the same automatic metric scores. There, we can clearly see the effect of the syntactic phrases, since the 3,052,121 phrases used in the fragmented variant of that scenario are more noticeably split between 1,970,888 PBSMT phrases (64.6%) and 1,081,233 syntax-based phrases (35.4%).

Some statistics for the makeup of the syntax-prioritized phrase table, compared to the baseline PBSMT phrase table, are shown in Figure 2. For each, we calculate the “ambiguity factor,” or the average number of target-language alternatives for each source-language phrase in the table. This analysis shows not only that the distribution of traditional PBSMT phrases is rather different from that of the syntactic phrases, it is also different from the non-syntactic PBSMT phrases that are preserved in the syntax-prioritized table. In effect, given a baseline PBSMT phrase table, the syntax prioritization replaces phrase entries for 39,089 source-language phrases, each with an average of 156 different target-language translations, with 39,105 source phrases, each with an average of 27.7 syntactically motivated target translations — a net savings of 5.0 million

Source: Je veux saluer , à mon tour , l' intervention forte et substantielle du président Prodi .
Reference: I too would like to welcome Mr Prodi 's forceful and meaningful intervention .

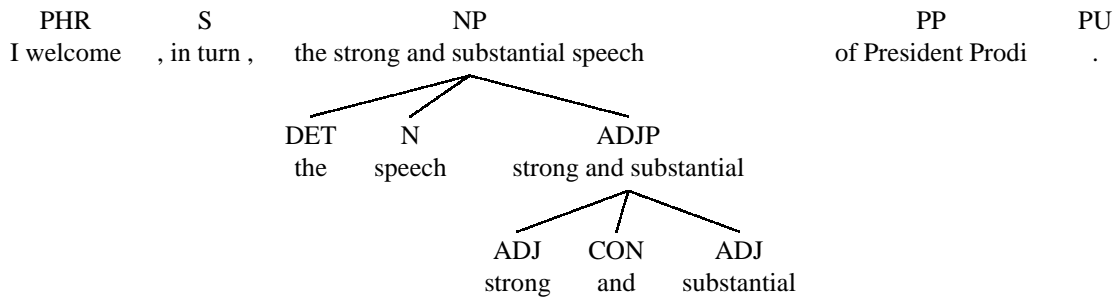


Figure 4: A translation example from the test set showing the result of including the nine-rule grammar in the syntax-prioritized combination. The SMT-only translation of the noun phrase is *the decisive intervention and substantial*.

phrase pairs. This is a strong indication that, because of the more accurate phrase boundary detection, the syntactic phrases are a much more precise representation of translational equivalence. An additional benefit is a significant reduction in decoding time, from an average of 27.3 seconds per sentence with the baseline PBSMT phrase table to 10.7 seconds per sentence with the syntax-prioritized table with the grammar included.

Improved precision due to the inclusion of syntactic phrases can be seen by examining a translation example and the phrasal chunks that produce it (Figure 3). In the syntax-prioritized output, the English phrases deriving from syntax-based phrase pairs are shown in *italic*, while the phrases deriving from PBSMT pairs are in normal type. The example shows an effective combination of on-target translations for syntactic constituents, when they are available, with non-syntactic phrases to handle constituent boundaries or places where parallel constituents are difficult to extract. The translation pieces *be informed* and *I am talking about*, though they exist in the baseline PBSMT phrase table, do not make it into the top-best translation in the PBSMT-only scenario because of its high ambiguity factor.

5.2 Effect of Syntactic Information

Although our current experiments do not show a significant increase in automatic metric scores with the addition of a small grammar, we can see the potential power of grammar in examining further sentences from the output. For example, in Figure 4, standard PBSMT phrase extraction is able to pick up

the adjective–noun reordering when translating from *intervention forte* to *decisive intervention*. However, in this sentence we have an adjective *phrase* following the noun, and there is no pre-extracted phrase pair for the entire constituent, so our system built from only PBSMT phrases produces the incorrect noun phrase translation *the decisive intervention and substantial*. Our nine-rule grammar, specifically targeted for this scenario, is able to correct the structure of the sentence by applying two rules to produce *the strong and substantial speech*.

Analysis of the entire test set further suggests that even our small grammar produces correct and precise output across all phrase table configurations, although the total number of applications of the nine rules remains low. There are 590 rule applications in the one-best output on the test set in the syntax-only configuration, 472 applications in the syntax-prioritized configuration, and 216 applications in the direct combination. In each configuration, we manually inspected all rule applications in the first 200 sentences and classified them as correctly reordering words in the English output (“good”), incorrectly reordering (“bad”), or “null.” This last category denotes applications of monotonic structure-building rules that did not feed into a higher-level reordering rule. The results of this analysis are shown in Figure 5. Overall, we find that the grammar is 97% accurate in its applications, making helpful reordering changes 88% of the time.

Given the preceding analysis — and the fact that our inclusion of a lexicalized reordering model in

Phrase Table	Good	Bad	Null
Syntactic only	47	3	8
Syntax-prioritized	45	1	3
Direct combination	25	0	0

Figure 5: Manual analysis of grammar rule applications in the first 200 sentences of the test set.

Moses resulted in automatic metric gains of only 0.0051 BLEU, 0.0029 METEOR, and 0.29 TER — we believe that further experiments with a much larger syntactic grammar will lead to a more significant improvement in automatic metric scores and translation quality.

6 Conclusions and Future Work

We have extended and applied an algorithm for combining syntax-based phrases from a parallel parsed corpus with non-syntactic phrases from phrase-based SMT within the context of a statistical syntax-based translation framework. Using a much larger corpus than has previously been employed for this approach, we produce jointly decoded output statistically equivalent to a monotonic decoding using standard PBSMT phrase-extraction heuristics, retaining syntactic information and setting the stage for further improvements by incorporating a syntactic grammar into the translation framework. Our preliminary nine-rule grammar, targeted for two specific English–French linguistic phenomena, already shows promise in performing linguistically motivated reordering that cannot be captured formally by a standard PBSMT model.

We present a syntax-prioritized method of combining phrase types into a single phrase table by always selecting a syntax-based phrase pair when one is available for a given source string. This new combination style reduces the size of the resulting phrase table and total decoding time by 61%, with only a minor degradation in MT performance. We suggest that this is because the syntax-derived phrases, when they can be extracted, are a much more precise method of describing correct translational equivalences.

As yet, we have made only minimal use of the Stat-XFER framework’s grammar capabilities. In our experiments, the full tree-to-tree-string rule-

extraction process of Ambati and Lavie (2008) produces more than 2 million unique SCFG rules when applied to a corpus the size of the Europarl. Not only is translating with such a large set computationally intractable, but empirically nearly 90% of the rules were observed only once in the parallel parsed corpus, making it difficult to separate rare but correct rules from those due to noise in the parses and word alignments. With the view of moving beyond our manually written nine-rule grammar, but wanting to get only the most useful rules from the entire automatically extracted set, we are currently investigating methods for automatic scoring or selection of a reasonable number of grammar rules for a particular language pair. Given that the majority of our phrase pairs, even in the syntax-prioritized combination, are non-syntactic, we have also conducted preliminary experiments with “syntactifying” them so that they may also be used by grammar rules to produce larger translation fragments.

The experiments in this paper used the grow-diag-final heuristic for word alignment combination because it has been shown to provide the highest precision on the subtree node alignment method by which we extract syntax-based phrase pairs (Lavie et al., 2008). However, this is a trade-off that sacrifices some amount of recall. Experimenting with different symmetric alignment heuristics may lead to a more optimal configuration for phrase-pair extraction or combination with PBSMT phrases. We also suspect that the choice of source- and target-side parsers plays a significant role in the number and nature of phrase pairs we extract; to address this, we are in the process of re-trying our line of experiments using the Berkeley parser (Petrov and Klein, 2007) for English, French, or both.

Acknowledgments

This research was supported in part by NSF grant IIS-0534217 (LETRAS) and the DARPA GALE program. We thank the members of the Parsing and Semantics group at Xerox Research Center Europe for parsing the French data with their XIP parser.

References

Salah Ait-Mokhtar, Jean-Pierre Chanod, and Claude Roux. 2001. A multi-input dependency parser. In

- Proceedings of the Seventh International Workshop on Parsing Technologies*, Beijing, China, October.
- Vamshi Ambati and Alon Lavie. 2008. Improving syntax driven translation models by re-structuring divergent and non-isomorphic parse tree structures. In *Proceedings of the Eighth Conference of the Association for Machine Translation in the Americas*, pages 235–244, Waikiki, HI, October.
- Chris Callison-Burch, Cameron Fordyce, Philipp Koehn, Christof Monz, and Josh Schroeder. 2007. (Meta-) evaluation of machine translation. In *Proceedings of the Second Workshop on Statistical Machine Translation*, pages 136–158, Prague, Czech Republic, June.
- Dan Klein and Christopher D. Manning. 2003. Fast exact inference with a factored model for natural language parsing. In *Advances in Neural Information Processing Systems 15*, pages 3–10. MIT Press, Cambridge, MA.
- Philipp Koehn, Franz Josef Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *Proceedings of HLT-NAACL 2003*, pages 48–54, Edmonton, Alberta, May–June.
- 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. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the ACL 2007 Demo and Poster Sessions*, pages 177–180, Prague, Czech Republic, June.
- Philipp Koehn. 2004. Statistical significance tests for machine translation evaluation. In *Proceedings of EMNLP 2004*, pages 388–395, Barcelona, Spain, July.
- Philipp Koehn. 2005. Europarl: A parallel corpus for statistical machine translation. In *Proceedings of the 10th Machine Translation Summit*, pages 79–86, Phuket, Thailand, September.
- Alon Lavie and Abhaya Agarwal. 2007. METEOR: An automatic metric for MT evaluation with high levels of correlation with human judgments. In *Proceedings of the Second Workshop on Statistical Machine Translation*, pages 228–231, Prague, Czech Republic, June.
- Alon Lavie, Alok Parlikar, and Vamshi Ambati. 2008. Syntax-driven learning of sub-sentential translation equivalents and translation rules from parsed parallel corpora. In *Proceedings of the Second ACL Workshop on Syntax and Structure in Statistical Translation*, pages 87–95, Columbus, OH, June.
- Alon Lavie. 2008. Stat-XFER: A general search-based syntax-driven framework for machine translation. In *Computational Linguistics and Intelligent Text Processing*, Lecture Notes in Computer Science, pages 362–375. Springer.
- Yang Liu, Qun Liu, and Shouxun Lin. 2006. Tree-to-string alignment template for statistical machine translation. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL*, pages 609–616, Sydney, Australia, July.
- Yuval Marton and Philip Resnik. 2008. Soft syntactic constraints for hierarchical phrase-based translation. In *Proceedings of ACL-08: HLT*, pages 1003–1011, Columbus, OH, June.
- Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, PA, July.
- Slav Petrov and Dan Klein. 2007. Improved inference for unlexicalized parsing. In *Proceedings of NAACL HLT 2007*, pages 404–411, Rochester, NY, April.
- Matthew Snover, Bonnie Dorr, Richard Schwartz, Linnea Micciulla, and John Makhoul. 2006. A study of translation edit rate with targeted human annotation. In *Proceedings of the Seventh Conference of the Association for Machine Translation in the Americas*, pages 223–231, Cambridge, MA, August.
- John Tinsley, Mary Hearne, and Andy Way. 2007. Exploiting parallel treebanks to improve phrase-based statistical machine translation. In *Proceedings of the Sixth International Workshop on Treebanks and Linguistic Theories*, pages 175–187, Bergen, Norway, December.
- Wei Wang, Kevin Knight, and Daniel Marcu. 2007. Binarizing syntax trees to improve syntax-based machine translation accuracy. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 746–754, Prague, Czech Republic, June.
- Ying Zhang and Stephan Vogel. 2006. Suffix array and its applications in empirical natural language processing. Technical Report CMU-LTI-06-010, Carnegie Mellon University, Pittsburgh, PA, December.
- Andreas Zollmann and Ashish Venugopal. 2006. Syntax augmented machine translation via chart parsing. In *Proceedings of the Workshop on Statistical Machine Translation*, pages 138–141, New York, NY, June.