CCG Contextual Labels in Hierarchical Phrase-Based SMT

Hala Almaghout, Jie Jiang, Andy Way
CNGL, School of Computing
Dublin City University
Dublin, Ireland
{halmaghout, jjiang, away}@computing.dcu.ie

Abstract

In this paper, we present a method to employ target-side syntactic contextual information in a Hierarchical Phrase-Based system. Our method uses Combinatory Categorial Grammar (CCG) to annotate training data with labels that represent the left and right syntactic context of target-side phrases. These labels are then used to assign labels to nonterminals in hierarchical rules. CCG-based contextual labels help to produce more grammatical translations by forcing phrases which replace nonterminals during translations to comply with the contextual constraints imposed by the labels. We present experiments which examine the performance of CCG contextual labels on Chinese–English and Arabic–English translation in the news and speech expressions domains using different data sizes and CCG-labeling settings. Our experiments show that our CCG contextual labels-based system achieved a 2.42% relative BLEU improvement over a Phrase-Based baseline on Arabic–English translation and a 1% relative BLEU improvement over a Hierarchical Phrase-Based system baseline on Chinese–English translation.

1 Introduction

After the successful introduction of Hierarchical Phrase-Based (HPB) MT system (Chiang, 2005), many approaches tried to provide the HPB model with syntactic information extracted from the target side of the training corpus in order to improve translation quality. Methods such as (Zollmann

and Venugopal, 2006; Almaghout et al., 2010) augment nonterminals in hierarchical rules with syntactic labels extracted from target-side parse trees of the training corpus. These syntactic labels act as syntactic constraints on phrases replacing nonterminals during decoding. However, there are some problems which affect the performance of syntax-augmented systems. One problem is the strong syntactic constraints imposed by syntax-augmented rules which restrict the search space of translation and prevent the system in many cases from finding good translations. Another problem is the sparse syntactic labels used in such systems, which cause the generation of low-probability, less reliable rules. This weakens the ability of the system to generalize. As a solution to these problems comes approaches which try to soften syntactic constraints imposed by labeled synchronous rules (Venugopal et al., 2009; Chiang, 2010) have been advanced.

In this paper, we propose a method to label nonterminals in hierarchical rules with target-side syntactic contextual labels extracted using Combinatory Categorial Grammar (CCG) (Steedman, 2000). For each target-side phrase in the training corpus, our method uses CCG supertags assigned to its words to extract the left and right syntactic context, represented in the left and right contextual CCG categories of this phrase. Then we annotate the target side of the training corpus by assigning to each phrase a syntactic label that results from combining its left and right contextual CCG categories. Finally, hierarchical rules are extracted from the annotated training corpus with nonterminals bearing the CCG-based contextual labels, which represent rich syntactic information while imposing softer syntactic constraints compared to the labels which use full CCG supertags (Almagh-

out et al., 2010).

We present experiments examining the performance of our CCG contextual labels-based system compared to HPB and Phrase-Based (PB) baseline systems on data sets from the news and traveling speech expressions domains. Our CCG contextual labels system was the best-performing system on Arabic–English news Chinese–English speech translation. In addition, our experiments demonstrate that the performance of CCG-augmented systems is affected by several factors, including the size and domain of training data and the source language of the translation pairs.

The rest of this paper is organized as follows: in Section 2 we review previous approaches which augment HPB system with syntactic knowledge. Section 3 gives a brief introduction to Combinatory Categorial Grammar (CCG). In Section 4 we introduce our approach. Section 5 presents our experiments. Section 6 concludes, and provides avenues for further work.

2 Related Work

An HPB MT system (Chiang, 2005) extracts a synchronous Context-Free Grammar (Lewis and Stearns, 1966) from a parallel corpus without syntactic annotation in the form of hierarchical rules. Hierarchical rules are phrases which contain gaps called nonterminals that can be replaced by other phrases. These rules capture the hierarchical aspects of language by providing the ability to translate discontinuous phrases and perform lexical reordering. Having no syntactic constraints imposed on phrases replacing nonterminals in the HPB grammar causes the production of ungrammatical translations. This led to approaches trying to provide the basic HPB model with syntactic knowledge in the form of syntactic labels attached to nonterminals in hierarchical rules. These labels restrict nonterminal replacement in hierarchical rules to the phrases which match the syntactic constraints imposed by the labels attached to nonterminals.

Syntax Augmented Machine Translation (SAMT) (Zollmann and Venugopal, 2006) attaches constituent grammar-based syntactic labels to nonterminals in hierarchical rules. First, each target-side sentence in the training corpus is assigned a parse tree. Next, a syntactic label is assigned to each phrase pair extracted from the sentence pair according to word alignments. This syntactic label corresponds to the constituent in the parse tree cover ing the target-side phrase. After that, hierarchical rules are extracted according to the same basic method presented in (Chiang, 2005), but with syntactic labels attached to nonterminals. Phrases which are not covered by a single constituent in the parse tree will be assigned a composite label that results from combining the constituents spanned by the phrase using a set of combinatory operators. (Almaghout et al., 2010) follow the same approach as SAMT to augment the HPB model with syntactic knowledge by assigning syntactic labels to nonterminals in hierarchical rules using Combinatory Categorial Grammar, which provides richer syntactic labels that accurately reflect the syntactic context and dependents of the phrase in addition to being flexible and efficiently extracted without the need for creating a full parse of the sentence.

(Hassan et al., 2007) integrate CCG supertags into the target language model and the target side of the translation model of the PB model. They also integrate a grammaticality metric into the n-gram language model over supertags. This metric penalizes the number of violations of combinatory operators in a sequence of supertags. (Birch et al., 2007) use CCG supertags as a factor in the factored PB translation model (Koehn and Hoang, 2007) following two approaches. The first approach generates CCG supertags as a target-side factor in the factored translation model, and then apply an n-gram language model over them. The second approach uses supertags as a source-side factor to direct the decoding process.

3 Combinatory Categorial Grammar

CCG (Steedman, 2000) is a grammar formalism in which most of the grammar of the language is stored in the lexicon. The CCG lexicon contains words paired with rich syntactic categories called "supertags". CCG uses a small set of simple combinatory rules to combine supertags.

CCG categories are divided into atomic and complex categories. Examples of atomic categories are: S(sentence), N (noun), NP (noun phrase). Complex categories such as $S \backslash NP$ and $(S \backslash NP)/NP$ are functions which specify the type and directionality of their arguments (primitive or complex categories) and the type of their result (primitive or complex category). Complex categories come in the following formats:

- X\Y is a functor X which takes as an argument the category Y to its left (which might be a primitive or complex category) and the result is the category X (which might also be a primitive or complex category).

- X/Y is a functor which takes as an argument the category Y to its right (which might be a primitive or complex category) and the result is the category X (which might also be a primitive or complex category).

For example, the lexical category of the verb *read* in the sentence *I read* is S\NP, which means that this category needs an NP (which plays the role of the subject in this case) as a left argument and the result of this category when an NP comes to its left is a sentence S. By contrast, in the sentence *I read a book* the lexical category assigned to the verb *read* in this case is (S\NP)/NP, which means that it needs an NP as a left argument (which plays the role of the subject) and another NP as a right argument (the object), and the result of this category when all of its arguments are fulfilled is a whole sentence S. Thus, the complex lexical category (S\NP)/NP represents a transitive verb while the lexical category S\NP represents an intransitive verb.

CCG Supertagging: According to CCG, each word in the lexicon has a number of supertags, each of which corresponds to a specific syntactic context in which the word may appear. Assigning every possible supertag to each word in the sentence before parsing will create a huge search space, thus putting a heavy burden on the parser to disambiguate them. A solution to this problem is CCG supertagging, which tries to disambiguate the set of supertags assigned to the words of the sentence before parsing according to the words context. This reduces the derivations search space and results in a faster parsing. (Bangalore and Joshi 1999) use statistics about supertag co-occurrences collected from a parsed corpus to reduce the number of supertags assigned to the words of the sentence. (Clark and Curran, 2004) build a wide coverage CCG parser which uses a log-linear probabilities to supertag the sentence before parsing, leading to faster, more accurate and robust parsing. After supertagging, the parser has only to use combinatory operators to combine supertags assigned to words in order to parse the sentence. That is why supertagging a sentence is considered as "al-
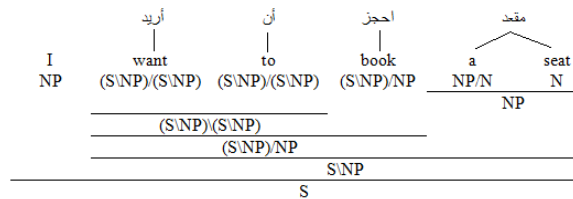

Figure 1: An Arabic sentence and its aligned English translation along with CCG derivation tree assigned to the English sentence.

most parsing" (Bangalore and Joshi, 1999). Figure 1 illustrates CCG derivation tree assigned to the English sentence *I want to book a seat*. The supertags assigned to each word in the sentence represent the first level of the CCG derivation tree.

4 CCG-based Contextual Labels in HPB

4.1 Motivation

SMT systems derive their strength from phrase-pairs extracted from the training corpus according to pure statistical methods, which means that phrase-pairs in an SMT system do not necessarily correspond to syntactic constituents. This is one of the main reasons why ungrammatical translations are produced and why it is difficult to incorporate syntax into SMT systems. Restricting SMT systems to use only phrases which represent syntactic constituents caused translation performance to degrade (Koehn et al., 2003). Thus, devising a method which incorporates syntax in SMT systems while maintaining statistically-extracted phrases would be the optimal solution.

Using a constituent-based grammar to annotate statistically extracted phrases with syntactic categories proved to be difficult, because constituent grammar has rigid structures which cause it to fail to annotate many phrases (Almaghout et al., 2010). By contrast, a lexicalized grammar such as CCG allow for flexible structures, which enables a CCG supertag to be assigned for a phrase which does not correspond to a grammatical constituent. In addition, a CCG supertag assigned to a word or phrase reflects complex syntactic constraints imposed on the word or phrase in its local context and at the lexical level, which makes it a rich syntactic description in its own right without the need for full parsing. Moreover, a CCG supertag is designed to contain only those elements on which the lexical item imposes constraints, which makes it a precise

283

syntactic description of all the dependents of the word or phrase.

All the aforementioned features provide many advantages for incorporating CCG supertags in SMT systems. However, rich syntactic information held by CCG supertags might become a disadvantage when using them to label nonterminals in hierarchical rules because of the large number of different labels extracted from the training data, which in turn leads to a very large grammar. As a result, the rule probability will be widely distributed among rules, which causes system performance to deteriorate as low-probability rules weaken the ability of HPB system to generalize. As a solution to this problem, we try to simplify CCG labels by using only part of the syntactic information represented in them. This helps to reduce the sparsity of the labels and soften the syntactic constraints imposed by them while still representing rich syntactic information. In our approach, we use the left and right contextual categories expressed by CCG supertags to label nonterminals in hierarchical rules. In other words, we use the left and right arguments of the CCG supertag assigned to each target-side phrase and drop the functor information from the label representation. For example in Figure 1, the phrase *want to* is assigned NP_NP as a contextual label instead of using the full supertag (S\NP)/NP. The loss of information that results from this step might cause the production of ungrammatical translations. However, we wanted to examine the total effect of the trade-off between the accuracy and the size of the grammar.

4.2 Extraction of CCG Contextual Labels

A CCG category takes the form of C=(R\L1)/L2 where L1 represents the left argument category, L2 represents the right argument category, and R represents the resulting category. For each target-side phrase in the training corpus assigned C as its CCG category, we include the left and right arguments L1 and L2 of C and ignore R in the label presentation. Thus, a syntactic label of the form L1_L2 is assigned for each target-side phrase in the training corpus. If any of the left or right arguments in C do not exist, it is replaced with X symbol. For example, in Figure 1, the CCG category assigned to the phrase *want to book* is (S\NP)/NP. This CCG category has S as a functor, NP as a right argument and NP as a left argument. Therefore, we use NP_NP

as the syntactic label assigned to this phrase.

The labels that consist of only the left and right arguments of a CCG supertag still express important syntactic information, namely the syntactic context of phrases, but at the same time are simpler and thus less sparse than complete CCG supertags. Furthermore, CCG context labels can be extracted simply from the CCG supertags assigned to the leftmost and rightmost words of the phrase without the need to parse the sentence, because the left argument of a CCG supertag assigned to a phrase is the same as the left argument of the leftmost supertag in the phrase, and the right argument of a CCG supertag assigned to a phrase is the same as the right argument of the rightmost supertag in the phrase. This makes our method more reliable and efficient that the methods that create full parse trees of the sentences. In the previous example, the leftmost supertag in the phrase *want to book* is (S\NP)/NP assigned to the leftmost word *want*. Thus, the left context of this phrase is NP. The rightmost supertag in the same phrase is (S\NP)/NP assigned to the word *book*. As a result, the right context of this phrase is NP. In this way, we can extract CCG context labels for a phrase from only the left argument of the leftmost supertag and the right argument of the rightmost supertag. The phrase *book a seat* in the previous example has NP_X as its contextual label meaning that this phrase expects an NP to its left but does not impose a constraint on the right context. In contrast, the phrase *a seat* has X_X as its contextual label meaning that this phrase does not impose any constraint on its right nor left context.

After annotating phrases in the training corpus with CCG-based contextual labels, hierarchical rules are extracted from the annotated training corpus according to the method of (Chiang, 2010), but with CCG contextual labels attached to nonterminals in the hierarchical rules. Figure 2 illustrates some of the hierarchical rules augmented with CCG contextual labels extracted from the English sentence *I want to book a seat* and its aligned Arabic sentence in Figure 1.

4.3 Simplifying CCG-based Labels

We tried to examine the effect of further simplifications of CCG labels on the overall performance of HPB system by removing the features carried by some CCG atomic categories from the label representation. These features are analo-

NP_S\NP → أُرِيد أن , want to
X_X → احجز مقعد NP_S\NP , I NP_ S\NP book a seat
NP_X → احجز مقعد , book a seat
X_X → مقعد NP_NP أُرِيد أن , I want to NP_NP a seat

Figure 2: A set of hierarchical rules augmented with CCG contextual labels extracted from the aligned Arabic–English sentence in Figure 1.

gous to latent variables attached to nonterminal symbols in Probabilistic Context Free Grammar in what is called Probabilistic Context Free Grammar with latent annotation (PCFG-LA) (Matsuzaki et al., 2005). For example the atomic category S might have a feature attached to it (described between two brackets after the category symbol) which distinguishes types of sentences such as declarative S[dcl] or wh-question S[wq]. Removing these features turns all the atomic categories S[dcl] and S[wq] into S, and the complex category (S[dcl]\NP)/NP into (S\NP)/NP. This led to further reduction in the total number of different labels, and accordingly the size of the extracted grammar. We applied this simplification on the CCG-augmented HPB system which uses CCG supertags as nonterminals labels (Almaghout et al., 2010) and on CCG contextual labels. The next section presents our experiments on the following CCG-based systems in comparison with PB and HPB baselines:

- CCG Context: uses CCG contextual labels as described in Section 4.2 to label nonterminals in hierarchical rules.

- CCG: uses CCG supertags to label nonterminals in hierarchical rules as described in (Almaghout et al., 2010)

- CCG Context (s): uses simplified CCG contextual labels.

- CCG (s): uses simplified CCG supertag labels.

5 Experiments

We conducted a set of experiments which examine the performance of the HPB system augmented with CCG contextual labels compared to HPB, PB and CCG supertags HPB systems under different settings: language pair, corpus type and corpus

Data	AE	CE
Training	48065	51044
Development	500	500
Testing	500	500

Table 1: Data size used for training, tuning and testing in our Arabic–English and Chinese–English news experiments. AE stands for Arabic–English and CE stands for Chinese–English.

Data	AE	CE
Training	20376	54574
Development	389	507
Testing	507	504

Table 2: Data size used for training, tuning and testing in our Arabic–English and Chinese–English speech experiments.

size. We conducted experiments on Arabic-to-English and Chinese-to-English translation in the news and traveling speech expressions domain using different corpus sizes.

5.1 Experimental Setup

Data Used: For Arabic–English news experiments, we used a corpus comprised of sentences selected randomly from the Arabic News corpus from LDC. We also used data provided by the IWSLT 2010 evaluation campaign for the Arabic–English BTEC task which consists of basic traveling speech expressions. For Chinese–English news experiments, we used data which consists of sentences selected randomly from the FBIS corpus. We also used training data from Chinese–English data provided by the IWSLT 2010 evaluation campaign for Chinese–English DIALOG task which consists of spoken dialogues in travel situations. Tables 1 and 2 show the size of training, development and test sets used in Arabic–English and Chinese–English news and speech experiments, respectively.

All the English data used in our experiments is lower-cased and tokenized. We used the CCG supertagger (Clark and Curran, 2004) from C&C tools[1] to supertag the English side of the training data for our CCG augmented hierarchical system experiments.

The Arabic data is segmented according to the D3 segmentation scheme (Sadat and Habash,

---

[1] http://svn.ask.it.usyd.edu.au/trac/candc/

285

2006) using MADA.[2]

Baseline Systems: We built two baseline systems: PB and HPB. We built the PB baseline system using the Moses Phrase-Based Decoder (Koehn et al., 2007) with maximum phrase length=12. The HPB baseline system is built using the Moses Chart-Decoder. For all our hierarchical systems, maximum phrase length is set to 12 and maximum rule span is set to 12. Rules extracted contain up to 2 nonterminals. The GIZA++ toolkit[3] is used to perform word and phrase alignment and the "grow-diag-final" refinement method is adopted (Koehn et al., 2003). Minimum error rate training (Och, 2003) is performed to tune all our SMT systems. The 5-gram language model in all experiments was trained on the target side of the parallel corpus using the SRILM toolkit[4] with modified Kneser-Ney smoothing (Kneser and Ney, 1995).

CCG Systems: We built our CCG-augmented HPB systems using Moses Chart Decoder which has an option to extract syntax-augmented rules from an annotated corpus. [5]

5.2 Experiments Results

Arabic–English: Tables 3 and 4 show the BLEU, METEOR and TER scores for CCG-based systems and baseline systems on Arabic–English news and IWSLT data, respectively. From Table 3 we can see that the best-performing system in terms of BLEU and TER was the simplified CCG contextual labels system, beating the PB baseline by 0.56 absolute BLEU points, which corresponds to a 2.42% relative improvement. We used paired bootstrap resampling to compute statistical significance (Koehn, 2004) of this improvement. The result of the test showed that the simplified CCG contextual labels system is better than the PB system in 98% of the samples at p-level=0.05, which is statistically significance. For Arabic–English experiments on IWSLT data, Table 4 shows that the HPB system has the best BLEU score with just a small improvement (0.06 absolute BLEU points) over the second best performing system, namely the simplified CCG contextual labels systems. We can also see that all CCG-based systems performed better than the PB baseline. In addition, simplifying CCG contextual labels proved to be useful for

System	BLEU	TER	METEOR
CCG Context(s)	**23.69**	**63.78**	55.88
PB	23.13	64.00	**57.11**
CCG	22.97	64.12	56.15
CCG Context	22.68	64.59	56.62
HPB	22.60	64.32	56.21
CCG (s)	20.98	64.84	55.58

Table 3: Experiments results for CCG-augmented HPB systems and baseline systems on Arabic–English news data.

System	BLEU	TER	METEOR
HPB	**53.20**	30.95	71.43
CCG Context(s)	53.14	31.04	**71.50**
CCG Context	52.86	31.13	71.33
CCG (s)	52.70	**30.85**	69.85
CCG	52.32	31.89	70.86
PB	52.31	32.42	70.80

Table 4: Experiments results for CCG-augmented HPB systems and baseline systems on Arabic–English IWSLT data.

both news and IWSLT data. However, simplifying CCG supertags proved to not always be useful.

Chinese–English: Tables 5 and 6 show the BLEU, METEOR and TER scores for CCG-based systems and baseline systems on Chinese–English news and IWSLT data, respectively. For news data, the CCG supertags system had the best BLEU score with just a slight improvement over the PB baseline by 0.03 absolute BLEU points. The CCG contextual labels system achieved the best TER and METEOR scores. For IWSLT data, from Table 6 we can see that the simplified CCG contextual labels system was the best-performing system in terms of BLEU and METEOR scores, beating the HPB baseline by 0.53 absolute BLEU points, which corresponds to 1% relative improvement. However, this improvement was not statistically significant. In addition, results show that simplifying CCG supertags proved to be useful for both news and IWSLT data, while simplifying CCG contextual labels proved to damage the performance on both news and IWSLT data.

5.3 Analysis

From previous experiments, it is obvious that for a specific language pair, the performance of CCG systems differs according to corpus type. IWSLT

---

[2] http://www1.ccls.columbia.edu/ cadim/MADA.html
[3] http://fjoch.com/GIZA++.html
[4] http://www-speech.sri.com/projects/srilm/
[5] http://www.statmt.org/moses/?n=Moses.SyntaxTutorial

System	BLEU	TER	METEOR
CCG (s)	**23.77**	67.08	52.43
PB	23.74	67.78	52.56
HPB	23.65	66.81	52.76
CCG Context	23.45	**66.38**	**53.49**
CCG	23.30	67.50	52.32
CCG Context (s)	22.91	68.29	51.40

Table 5: Experiments results for CCG-augmented HPB systems and baseline systems on Chinese–English news data.

System	BLEU	TER	METEOR
CCG Context	**51.42**	33.00	**67.82**
HPB	50.89	**32.53**	66.81
CCG Context(s)	50.82	32.55	67.59
CCG (s)	48.86	36.12	65.92
CCG	48.26	36.17	64.88
PB	47.69	34.20	65.35

Table 6: Experiments results for CCG-augmented HPB systems and baseline systems on Chinese–English IWSLT data.

System	BLEU	TER	METEOR
HPB	**38.96**	**40.07**	**60.05**
CCG Context	38.40	40.91	59.22
PB	38.37	39.92	59.48
CCG Context (s)	38.36	41.10	59.26
CCG (s)	38.12	40.36	59.38
CCG	36.12	42.17	57.98

Table 7: Experiments results for CCG-augmented HPB systems and baseline systems on 20k of the Chinese–English IWSLT data.

data consists of speech expressions in the travel domain which are mainly short questions with an average sentence length of 9 words. In contrast, news data consists of long sentences in the news domain with an average sentence length of 37 words. However, analyzing the experiments on IWSLT data across Arabic–English and Chinese–English experiments shows that the performance of CCG systems which use contextual labels is better than CCG systems which use supertags. While our CCG contextual labels system outperforms HPB system on Chinese–English IWSLT data, HPB outperforms all CCG-based systems on Arabic–English IWSLT data. However, the size of the Arabic–English IWSLT data is about one third of the size the Chinese–English data, which means that the comparison might be unfair between the systems on these data sets. Thus, we reduced the size of the Chinese–English data to be the same as the size of the Arabic–English data and rerun the experiments. Table 7 presents the results of Chinese–English systems on the reduced data set. We see that the HPB system has the best BLEU score, while other CCG-based systems maintain the same order relative to each other. Reducing data size resulted in an advantage for the HPB system over other CCG-based systems. This shows

that the size of the training data is important for the performance of the CCG-based systems.

While experiments on IWLST data showed similarities in the performance of CCG-based systems, the situation for news data is totally the opposite. The order of Arabic–English CCG-based systems according to their BLEU scores is the exact opposite of that of Chinese–English ones. Furthermore, simplifying CCG contextual labels and supertags had different effects on the performance depending on the source language. Keeping in mind that the average sentence length and the size of the news training data is approximately the same between the two language pairs, this shows that the source langauge also affects the performance of the CCG-based systems.

6 Conclusion and Future Work

In this paper, we presented our approach to augment the Hierarchical Phrase-Based system with contextual labels extracted from CCG supertags assigned to target-side words. Compared to CCG supertags, CCG contextual labels are easier to extract and are less sparse. We presented a set of experiments which examined the performance of CCG contextual labels on Arabic–English and Chinese–English translation compared to Hierarchical and Phrase-Based systems and CCG supertags-based system on data sets from the news and traveling speech expressions domains. Our experiments showed that our CCG contextual labels system was the best-performing system for Arabic–English news translation and for Chinese–English speech translation. Moreover, our experiments demonstrated that the performance of all CCG-based systems is affected by the domain, size and source language of translation.

In future work, we intend to experiment on

larger data sets. In addition, we want to examine the effect of word segmentation on the performance of CCG-based systems. Word segmentation affects word alignment, which in turn affects phrase and rule extraction and scoring. We believe that the differences between Arabic and Chinese word segmentation is the key factor behind the varying performance of CCG-based systems on data sets of the same size and domain between Arabic–English and Chinese–English experiments. Finally, we will try to use system combination on CCG-based systems to obtain a better performing system.

References

H. Almaghout, J. Jiang, and A. Way. 2010. *CCG augmented hierarchical phrase-based machine translation.* In Proceedings of the 7th International Workshop on Spoken Language Translation, Paris, France; pp.211-218.

S. Bangalore and A. Joshi. 1999. *Supertagging: An Approach to Almost Parsing.* Computational Linguistics 25(2):237-265.

A. Birch, M. Osborne, and P. Koehn. 2007. *CCG supertags in factored statistical machine translation.* ACL 2007: In Proceedings of the Second Workshop on Statistical Machine Translation, Prague, Czech Republic, pp. 9-16.

D. Chiang. 2005. *A hierarchical phrase-based model for statistical machine translation.* In Proceedings of the 43rd Annual meeting of the Association for Computational Linguistics, Ann Arbor, MI, pp.263-270.

D. Chiang. 2010. *Learning to translate with source and target syntax.* In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Uppsala, Sweden, pp.1443-1452.

S. Clark and J. Curran. 2004. *The Importance of Supertagging for Wide-Coverage CCG Parsing.* In Proceedings of COLING-04, Geneva, Switzerland, pp.282288.

H. Hassan, K. Sima'an, and A. Way. 2007. *Supertagged phrase-based statistical machine translation.* In Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Prague, Czech Republic, pp. 288-295.

R. Kneser and H. Ney. 1995. *Improved Backing-Off for M-Gram Language Modeling.* In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP),Detroit, Michigan, USA, Vol. 1, pp. 181-184.

P. Koehn. 2004. *Statistical significance tests for machine translation evaluation.* In Proceedings of Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, Barcelona, Spain.

P. Koehn and H. Hoang. 2007. *Factored translation models.* In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Prague, Czech Republic, pp. 868-876.

P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin and E. Herbst. 2007. *Moses: open source toolkit for statistical machine translation.* ACL 2007: proceedings of demo and poster sessions, Prague, Czech Republic, pp. 177-180.

P. Koehn, F. Och, and D. Marcu. 2003. *Statistical phrase-based translation.* In Proceedings of the 2003 Human Language Technology Conference and the North American Chapter of the Association for Computational Linguistics, Edmonton, AB, Canada, 2003, pp. 48-54.

P. M. Lewis and R. E. Stearns. 1966. *Syntax directed transduction.* Foundations of Computer Science, Annual IEEE Symposium on, pp. 21-35, 7th Annual Symposium on Switching and Automata Theory (SWAT 1966).

T. Matsuzaki, Y. Miyao and J. Tsujii *Probabilistic CFG with latent annotations .* In Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, ACL '05, Ann Arbor, Michigan, pages 7582.

F. Och. 2003. *Minimum Error Rate Training in Statistical Machine Translation.* In the 41st Annual Meeting of the Association for Computational Linguistics, Sapporo, Japan, pp. 160-167.

F. Sadat and N. Habash. 2006. *Combination of Arabic preprocessing schemes for statistical machine translation.* In Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, Sydney, pp.1-8.

M. Steedman. 2000. *The Syntactic Process.* MIT Press, Cambridge, MA.

A. Venugopal, A. Zollmann, N.A.Smith, and S. Vogel. 2009. Preference grammars: softening syntactic constraints to improve statistical machine translation. In Proceedings of Human Language Technologies: the 2009 annual conference of the North American Chapter of the ACL, Boulder, Colorado, pp.37-45.

A. Zollmann and A. Venugopal. 2006. *Syntax augmented machine translation via chart parsing.* HLT-NAACL 2006: Proceedings of the Workshop on Statistical Machine Translation, New York, NY, USA, pp. 138-141.