Proceedings of the 16th EAMT Conference, 28-30 May 2012, Trento, Italy

Extending CCG-based Syntactic Constraints in Hierarchical Phrase-Based SMT

Hala Almaghout
CNGL, School of Computing
Dublin City University
Dublin, Ireland
halmaghout@computing.dcu.ie

Jie Jiang
Applied Language Solutions
Delph
UK
jie.jiang@appliedlanguage.com

Andy Way
Applied Language Solutions
Delph
UK
andy.way@appliedlanguage.com

Abstract

In this paper, we describe two approaches to extending syntactic constraints in the Hierarchical Phrase-Based (HPB) Statistical Machine Translation (SMT) model using Combinatory Categorial Grammar (CCG). These extensions target the limitations of previous syntax-augmented HPB SMT systems which limit the coverage of the syntactic constraints applied. We present experiments on Arabic–English and Chinese–English translation. Our experiments show that using extended CCG labels helps to increase nonterminal label coverage and achieve significant improvements over the baseline for Arabic–English translation. In addition, combining extended CCG labels with CCG-augmented glue grammar helps to improve the performance of the Chinese–English translation over the baseline systems.

1 Introduction

Hierarchical Phrase-Based (HPB) Statistical Machine Translation (SMT) (Chiang, 2005) has been demonstrated to be one of the most successful SMT approaches nowadays. Its main idea is to imitate Context-Free Grammar (CFG) production rules in modelling translation rules while maintaining the strength of statistically extracted phrases. However, HPB SMT only models the hierarchical aspect of the language and does not use any linguistic information in rule extraction. A set of approaches (Zollmann and Venugopal, 2006; Almaghout et al., 2010) have tried to incorporate syntactic information extracted according to different grammar theories in the HPB SMT model by annotating phrases and nonterminals with syntactic labels. These systems face many challenges in integrating their syntax-based constraints with the syntax-free statistically extracted HPB SMT translation grammar, which limits the coverage of these syntactic constraints and thus minimizes the benefit obtained from applying them.

In this paper, we try to extend the scope of target-side syntactic constraints in syntax-augmented HPB SMT. More specifically, we try to exploit the flexibility of Combinatory Categorial Grammar (CCG) (Steedman, 2000) to increase the coverage of syntactic labels used to label phrases and nonterminals in hierarchical rules. In addition, we augment HPB glue grammar rules with CCG combinatory operators with the aim of directing the decoding process towards building a full parse tree of the translation output. We apply these constraints in a soft manner through a feature in the log-linear model (Venugopal et al., 2009).

The rest of this paper is organized as follows. Section 2 reviews related work. Section 3 gives an introduction to HPB SMT. Section 4 introduces CCG. Section 5 describes our approach. Section 6 presents our experiments. Finally, Section 7 concludes and provides avenues for future work.

2 Related Work

Syntax Augmented Machine Translation (SAMT) (Zollmann and Venugopal, 2006) tries to improve the grammaticality of the HPB SMT translation output by attaching syntactic labels to target-side phrases and nonterminals. These labels are extracted from context-free phrase structure grammar parse trees of the target-side of the parallel corpus. The function of these syntactic labels is to impose syntactic constraints on phrases replacing nonterminals during decoding, allowing this replacement only when the labels of the nonterminal and the replacing phrase match. CCG-augmented HPB (Almaghout et al., 2010) follows the SAMT approach in labelling nonterminals with syntactic labels. It extracts CCG-based labels from CCG forest trees of the target-side of the parallel corpus. Almaghout et al. (2011) use contextual information presented in CCG categories to extract syntactic labels for nonterminals and phrases in the HPB SMT translation model. Birch et al. (2007) use CCG supertags as a source and target factor in the factored Phrase-Based (PB) SMT translation model (Koehn and Hoang, 2007). Hassan et al. (2009) integrate target-side CCG incremental parsing in the Direct Translation Model (DTM2). They also extract a set of syntactic features based on CCG supertags, combinatory operators and parsing states. This helps to build a fully connected parsing structure during decoding and prune hypotheses which do not constitute a valid parsing state.

Recently, applying syntactic constraints in syntax-augmented HPB SMT systems in a soft manner has been demonstrated to improve the performance of these systems (Venugopal et al., 2009; Chiang, 2010). This means that the derivations which violate the syntactic constraints imposed by the model are not prevented per se, but the system learns to favour more grammatical translations. Strong syntactic constraints impose restrictions on the translation search space and consequently have a negative impact on performance. Venugopal et al. (2009) transform the syntactic constraints in the SAMT translation model to a syntactic feature integrated into the log-linear model. They use an unlabelled translation model during decoding. Another SAMT-based syntactic model, which measures the probability of different labellings of each hierarchical rule, is used to calculate the value of the syntactic feature at each nonterminal replacement during decoding.

3 Hierarchical Phrase-Based SMT

HPB SMT (Chiang, 2005) is a tree-based model which extracts a synchronous CFG automatically from the training corpus. HPB SMT extracts hierarchical rules – the fundamental translation units in the HPB model – from phrases extracted according to the PB model (Koehn et al., 2003). Thus, hierarchical rules have the strengths of statistically extracted continuous phrases plus the ability to translate discontinuous phrases and learn phrase-reordering without a separate reordering model. The HPB SMT model has two types of rules: hierarchical rules and glue grammar rules. Hierarchical rules are rewrite rules with aligned pairs of right-hand sides, taking the following form:

$$X \rightarrow < \alpha, \beta, \sim > \qquad (1)$$

where X is a non-terminal, α and β are both strings of terminals and non-terminals, and $\sim$ is a one-to-one correspondence between non-terminal occurrences in α and β. Hierarchical rules are extracted from the training corpus by subtracting continuous phrase-pairs attested in the translation table recursively from longer phrases and replacing them with the nonterminal symbol X. Nonterminals in hierarchical rules act as placeholders that are replaced with other phrases during translation in a bottom-up fashion.

Glue grammar rules perform monotone phrase concatenation, which means that they combine target phrases together without performing any reordering. They consist of the following two rules:

$$S \rightarrow < \ S\,X \ , \ S\,X \ > \qquad (2)$$

$$S \rightarrow < \ X \ , \ X \ > \qquad (3)$$

Their main role is to produce translation when no possible hierarchical rule can be applied. They are also used to reduce the complexity of chart decoding by limiting the application of the hierarchical rules to a certain limit (12 words in our experiments, cf. Section 6) above which only glue grammar rules are applied. Glue grammar rules can also be applied below this limit but their application cannot alternate with hierarchical rules, and they always form a left-balanced binary tree on top of the hierarchical rules in the derivation tree.

4 Combinatory Categorial Grammar

CCG (Steedman, 2000) is a grammar formalism which consists of a lexicon that pairs words with lexical categories (supertags, cf. Bangalore and Joshi (1999)) and a set of combinatory rules which specify how the categories are combined. A supertag is a rich syntactic description that specifies the local syntactic context of the word at the lexical level in the form of a set of arguments. Most of the CCG grammar is contained in the lexicon, which is why CCG has simpler combinatory rules compared to CFG productions.

CCG categories are divided into atomic and complex categories. Examples of atomic categories are S (sentence), N (noun), NP (noun phrase), etc. Complex categories such as S\NP and (S\NP)/NP are functions which specify the type and directionality of their arguments and results. CCG builds a parse tree for a sentence by combining CCG categories using a set of binary combinatory operators. Since most of the CCG grammar resides in the lexicon, CCG has a simple set of combinatory operators. Figure 1 shows a CCG parse tree of the English sentence *Would you like cream and sugar in your coffee ?*

4.1 CCG and SMT

CCG has many unique qualities which make it an attractive grammar formalism to be incorporated into SMT systems. First, CCG allows for flexible structures thanks to its combinatory operators. Thus it is possible to assign a CCG category to phrases which do not represent standard syntactic constituents. This is an important feature for SMT systems as SMT phrases are statistically extracted, and do not necessarily correspond to syntactic constituents. Second, CCG supertags present rich syntactic information at the lexical level about the dependents and local context of each word in the sentence. Therefore, CCG supertags reflect important information about the syntactic structure of the sentence without the need to build a full parse tree. This allows SMT systems to build grammaticality metrics based on examining sequences of CCG supertags of the words of the translation output. Finally, CCG can be efficiently parsed thanks to the process of supertagging (Bangalore and Joshi, 1999), which assigns supertags to the words of the sentence before parsing. This reduces the parsing search space significantly and is especially important for computationally complex SMT systems.

5 Our Approach

5.1 Motivation

Although incorporating syntax into HPB SMT has been demonstrated to improve its translation quality (Zollmann and Venugopal, 2006), the coverage of the syntactic constraints in syntax-augmented HPB SMT systems is limited because they include only part of the phrases and the grammar in the model. The mismatch between the notion of the phrase in SMT systems and grammar formalisms leaves many phrases in syntax-augmented HPB SMT systems unlabelled. Almaghout et al. (2010) show that CCG-augmented HPB SMT and SAMT systems fail to label 30% and 50% of the total phrases in the training corpus, respectively. Furthermore, syntax-augmented HPB SMT systems have always focused efforts on augmenting hierarchical rules with syntax, ignoring the other important part of the grammar which is glue grammar rules. Glue grammar rules constitute about 30% to 40% of the total rules used in the derivations, which means that they play an important role in the translation process. Bearing in mind that hierarchical rules have a limited span to reduce the complexity of chart decoding, the application of syntactic constraints is also limited for the same reason. Ignoring these aspects limits the scope of syntactic constraints in syntax-augmented HPB systems which in turn limits their effect on improving the grammaticality of translation output.

In our approach, we try to expand the scope of syntactic constraints in our CCG-augmented HPB system. To achieve this we follow a two-fold approach. First, we try to extend the notion of the syntactic label attached to nonterminal labels and phrases with the aim of increasing label coverage. Secondly, we augment glue grammar rules with CCG combinatory operators. We apply these enhancements in a soft way under the Preference Grammars paradigm for applying soft syntactic constraints in HPB SMT (Venugopal et al., 2009). Thus, we add a syntactic feature to the log-linear model which judges the grammaticality of each nonterminal replacement and glue grammar rule application. We will describe each research strand in detail in the following sections.

5.2 Extended CCG-based Syntactic Labels

In SAMT, a set of CCG-like binary operators are used to increase the coverage of nonterminal labels. This is necessary as SAMT labels are extracted using phrase structure grammar, which has a small set of constituent labels that are insufficient to cover all the different syntactic structures of extracted phrases. Almaghout et al. (2010) use single-category CCG labels as nonterminal labels. Although CCG flexible structures allow a better label coverage than phrase structure grammar-based labels, using single-category CCG labels fails to label about one third of the total phrases. In or-

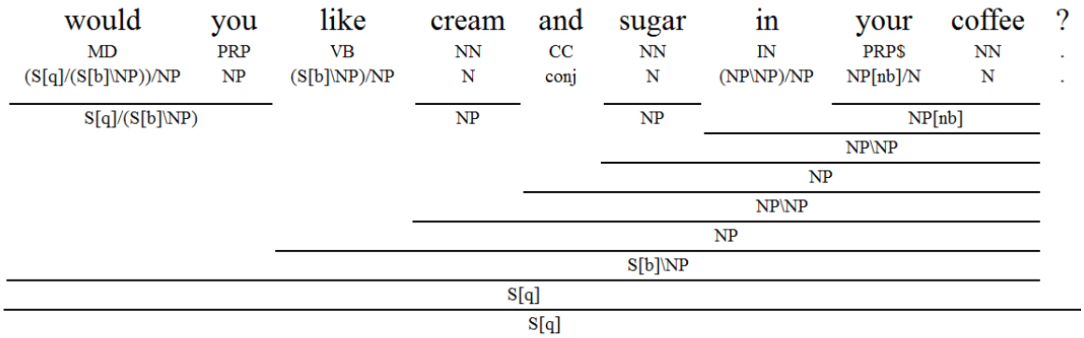

Figure 1: An example of a complete CCG parse tree of an English sentence.

cream and	N+conj
you like cream	S[b]
like cream and	S[b]\NP+conj
sugar in your	N+(NP\NP)/NP+NP[nb]/N

Figure 2: A set of phrases along with their extended CCG labels extracted from the CCG tree in Figure 1.

der to increase label coverage, we extend the definition of the nonterminal label to be composed of more than one CCG category. Therefore, if there is no single CCG category at the root of the trees which cover a phrase, the highest-scoring sequence of categories with a minimum number of CCG categories is extracted from CCG trees covering the phrase and used as the phrase label. Figure 2 shows a set of phrases extracted from the sentence illustrated in Figure 1 along with their extended CCG labels. In this example, the phrase *like cream and* has an extended CCG label composed of two categories: S[b]$\backslash$NP+conj. The CCG categories of the words *like* and *cream* are combined into the CCG category S[b]$\backslash$NP. However, this category cannot be combined with the conj category.

We define the degree of the extended label to be the number of CCG categories in the label. In our previous example, the extended CCG label of the phrase *sugar in your* is of degree three while the phrase *cream and* is of degree two. The degree of the system which uses extended CCG labels is defined to be the maximum degree of the labels in the model.

5.3 CCG-augmented Glue Grammar

Instead of concatenating phrases during glue grammar rule application without applying any syntactic constraints, we try to augment glue grammar rules with CCG combinatory operators. CCG combinatory operators are binary operators, which makes them suitable to be applied on glue grammar rules which are also binary rules. First, we change the definition of the glue grammar rule (2) as follows:

$$X \rightarrow < \ X\ X\ ,\ X\ X\ > \tag{4}$$

This removes the left-balance constraints from the construction of glue-grammar rule application. Additionally, this rule allows the application of glue grammar rules and hierarchical rules to alternate, which gives better flexibility. Secondly, we build a metric which judges the grammaticality of concatenating two phrases at each glue grammar rule application based on their extended CCG labels. The calculation of this grammaticality metric is based on an extended CCG label model. This model is extracted using relative frequency counts from the target-side of the training corpus which is annotated with extended CCG labels for each sub-phrase in each sentence.

Whenever two phrases are concatenated under glue grammar rule application, the following steps are applied to calculate the grammaticality features for each extended CCG label pair L1 and L2 from the first and second phrase, respectively:

- Simplify L1+L2 by applying all possible CCG combinatory operators on L1+L2 to derive the extended CCG label L with the minimum number of CCG categories.
- If the resulting label L from the previous step is composed of one CCG category, the two phrases are likely to constitute a grammatical phrase and the grammaticality feature is set to 1.
- Otherwise, the grammaticality feature is set to the probability of L according to the extended CCG labels model.

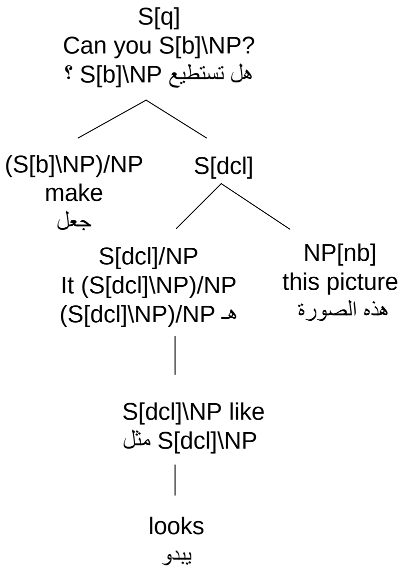

Figure 3: A derivation tree which shows the application of CCG-augmented glue grammar rules.

- Assign L to the phrase resulting from glue grammar rule application.

Augmenting glue grammar rules with CCG combinatory operators enables the building of a full parse tree of the translation output and extends the scope of the syntactic constraints to cover the whole translation output. The grammaticality feature helps to guide the decoding process by awarding the application of hierarchical and glue grammar rules which yield a grammatical output.

Figure 3 shows a derivation tree we obtain when translating a sentence from Arabic into English. Each node in the tree usually has more than one CCG label but the figure shows only the most probable label for the sake of simplicity. The resulting English translation is: *can you make it looks like this picture ?*. Although the translation is not totally grammatical, having the verb *look* in the wrong form *looks*, we can see how hierarchical and glue grammar rules participate in building a full parse tree which covers the whole translation output.

6 Experiments

In our experiments, we try to explore the effect of each method for extending the syntactic constraints in the HPB SMT system presented in Section 5. Sections 6.3.1 and 6.3.2 give the results for each individual method. We then conduct experiments which examine the effect of combining both approaches in a single system.

6.1 Data Used

We used the data provided by the IWSLT 2010 evaluation campaign.[1] The Chinese–English training corpus consists of 55k sentence pairs from the IWSLT 2010 Chinese–English training data for the DIALOG task. The development and test sets are IWSLT evaluation data sets (500 sentence pairs for each) provided for the Chinese–English DIALOG task for 2008 and 2009. The development set has 15 references and the test set has 7 references. The Arabic–English training corpus consists of 20k sentence pairs from the IWSLT 2010 training data provided for the Arabic–English BTEC task. The development and test sets are the IWSLT evaluation data sets provided for the BTEC task for 2007 and 2008 evaluations, with 489 sentence pairs in the development set and 507 sentence pairs in the test set, respectively. The development set has 7 references and the test set has 16 references. All the English data used in our experiments is lower-cased and tokenized. The Arabic data is segmented according to the D3 segmentation scheme using MADA (Morphological Analysis and Disambiguation for Arabic).[2]

6.2 Baseline Systems

We have two baseline systems in our experiments: the HPB SMT baseline system and the CCG-augmented HPB SMT baseline system which uses single-category CCG labels and applies strong syntactic constraints (Almaghout et al., 2010). We built our HPB SMT baseline system using the Moses Chart Decoder.[3] The GIZA++ toolkit[4] is used to perform word and phrase alignment and the "grow-diag-final" refinement method is adopted (Koehn et al., 2003). Maximum phrase length and maximum rule span are both set to 12 words. The maximum span for the chart during decoding is set to 20 words, above which only glue grammar rules are applied. Hierarchical rules extracted contain up to 2 nonterminals. 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

[1] http://iwslt2010.fbk.eu/node/27
[2] http://www1.ccls.columbia.edu/MADA/
[3] http://www.statmt.org/moses/?n=Moses.SyntaxTutorial
[4] http://fjoch.com/GIZA++.html

of the parallel corpus using the SRILM toolkit[5] with modified Kneser-Ney smoothing. Our CCG-augmented HPB system was also built using the Moses Chart Decoder, which has an option to extract syntax-augmented rules from an annotated corpus. We used the same rule extraction and decoding settings as for the HPB baseline system. We use the CCG parser from C&C tools[6] to parse the training data for our CCG-augmented HPB system experiments and to combine CCG categories during glue grammar rule application.

6.3 Experimental Results

6.3.1 Extended CCG Labels Experiments

In this section we examine the effect of our extended CCG labels. We try out extended labels of degrees ranging from one to five under soft syntactic constraints. Tables 1 and 2 show BLEU, TER and METEOR scores of extended CCG labels systems along with the number of different nonterminal labels estimated in thousands and the percentage of unlabelled nonterminals in the rule table of each system for Arabic–English and Chinese–English translation, respectively.

From Table 1, we can see that the 5-category CCG-augmented HPB SMT system is the best-performing system in terms of BLEU and TER scores, outperforming the HPB and CCG-augmented HPB baseline systems by 0.86 and 2.22 absolute BLEU points, which corresponds to a 1.63% and 4.3% relative improvement, respectively. The result of the paired bootstrap resampling test (Koehn, 2004) demonstrates that the improvement achieved over both baseline systems is statistically significant at p-level=0.05. Table 1 also shows that using soft syntactic constraints leads to significant improvements over the CCG-augmented HPB SMT baseline, which uses strong syntactic constraints. Furthermore, using extended CCG labels significantly decreases the percentage of unlabelled nonterminals in the rule table from 28% in the single-category system to 0.05% in the 5-category system.

Table 2 shows that the 3-category CCG-augmented HPB SMT system is the best-performing system in terms of BLEU and TER. The 3-category CCG-augmented HPB SMT system outperformed the HPB and CCG-augmented HPB SMT baseline systems by 1.65 and 3.93 absolute BLEU points, which corresponds to a 3.4% and 8.5% relative improvement, respectively. The paired bootstrap resampling test demonstrates that these improvements are both significant at p-level=0.05. Similar to our Arabic–English experiments, using soft syntactic constraints helps to achieve significant improvements over the strong-constraints CCG-augmented HPB baseline system.

System	BLEU	TER	MET	Lab	%X
HPB	52.90	31.06	**71.51**	-	-
CCG	51.54	32.32	70.33	0.5	28
CCG1	52.83	31.13	70.77	0.5	28
CCG2	53.38	30.92	70.60	8.0	6.3
CCG3	53.10	30.76	70.77	18	1.3
CCG4	53.09	30.76	70.62	23	0.3
CCG5	**53.76**	**30.76**	71.05	24	0.05

Table 1: Experimental results of CCG-augmented HPB systems with extended CCG labels from different degrees compared to the baseline systems for Arabic–English translation. Lab indicates the number of different labels used by each system (in thousands). %X indicates the percentage of unlabelled nonterminals in the rule table.

System	BLEU	TER	MET	Lab	%X
HPB	48.29	35.28	65.85	-	-
CCG	46.01	34.86	63.01	0.6	31
CCG1	49.73	34.04	**66.66**	0.6	31
CCG2	48.19	35.46	64.67	12	7.6
CCG3	**49.94**	**34.02**	66.29	30	1.7
CCG4	48.32	34.54	65.07	40	0.4
CCG5	49.44	34.10	65.76	43	0.09

Table 2: Experimental results of CCG-augmented HPB systems with extended CCG labels from different degrees compared to the baseline systems for Chinese–English translation along with the number of different labels and the parentage of unlabelled nonterminals in the model of each system.

6.3.2 CCG-augmented Glue Grammar Experiments

We examined the application of our CCG-augmented glue grammar rules using the single-category CCG labels under soft syntactic constraints. Tables 3 and 4 show the results of using CCG-augmented glue grammar for Arabic–

[5]http://www-speech.sri.com/projects/srilm/
[6]http://svn.ask.it.usyd.edu.au/trac/candc/

English and Chinese–English translation, respectively.

System	BLEU	TER	METEOR
CCG glue1	53.06	31.42	71.00

Table 3: Experimental results of the CCG-augmented HPB system which uses CCG-augmented glue grammar rules with single-category CCG labels for Arabic–English translation.

For both language pairs, CCG-augmentation for glue grammar rules failed to achieve any improvement over the best-performing systems obtained using extended CCG labels. Furthermore, we observe that using CCG-augmented glue grammar rules leads to a significant decrease in BLEU score for Chinese–English translation, even below the baseline performance.

System	BLEU	TER	METEOR
CCG glue1	45.65	36.64	62.91

Table 4: Experimental results of the CCG-augmented HPB system which uses CCG-augmented glue grammar rules with single-category CCG labels for Chinese–English translation.

6.3.3 Extension Approaches in Combination

In this section we try to combine both approaches to extending syntactic constraints described in this paper, namely (i) extended CCG labels and (ii) CCG-augmented glue grammar rules. We try to use CCG-augmented glue grammar rules with the best-performing systems obtained in Section 6.3.1, namely the 5-category and 3-category CCG-augmented HPB SMT systems for Arabic–English and Chinese–English translation, respectively. Tables 5 and 6 show BLEU, TER and METEOR scores when using CCG-augmented glue grammar rules in these systems. Using CCG-augmented glue grammar rules for Arabic–English leads to an improvement of 0.38 absolute TER points, which corresponds to a 1% relative improvement. Using CCG-augmented glue grammar rules for Chinese–English leads to an increase of 0.79 absolute BLEU points over the 3-category CCG-augmented HPB system, which corresponds to a 1.6% relative improvement. This result is corroborated by improvements with respect to TER and METEOR. The paired bootstrap resampling test shows that our CCG-augmented glue grammar system outperforms the 3-category CCG-augmented HPB SMT system in 93 out of 100 samples. However, this improvement is not statistically significant at p-level=0.05.

System	BLEU	TER	METEOR
CCG glue5	53.51	30.38	70.81

Table 5: Experimental results of the CCG-augmented HPB system which uses CCG-augmented glue grammar rules with 5-category extended CCG labels for Arabic–English translation.

We attempted to understand why using CCG-augmented glue grammar rules led to an improvement using 3-category extended labels, but caused a performance degradation when used with single category labels for Chinese–English translation. Accordingly, we measure the percentage of glue grammar rule application in the derivation trees that yield the translation output of each system. We found that glue grammar rules constitute 13.76% of the total rules used by the single-category CCG-augmented HPB SMT system which uses CCG-augmented glue grammar rules, compared to 4.8% used by the 3-category CCG-augmented HPB SMT system which uses CCG-augmented glue grammar rules. We think that this increased usage of glue grammar rules is due to restrictions imposed on the single-category system, which result from the restricted set of the single-category labels. This forces the system to use more glue grammar rules, which perform no reordering, causing the performance of the system to degrade. We think that the reason why using CCG-augmented glue grammar rules did not improve the performance for Arabic–English translation might be because of the small size of the training data (20k only), which increases the sparsity of translation rules extracted.

System	BLEU	TER	METEOR
CCG glue3	50.73	33.50	66.67

Table 6: Experimental results of the CCG-augmented HPB system which uses CCG-augmented glue grammar rules with 3-category extended CCG labels for Chinese–English translation.

7 Conclusion and Future Work

In this paper, we presented two syntactic extensions to HPB SMT system using CCG. The first extension tries to increase the coverage of syntactic labels used to label nonterminals in hierarchical rules by using complex CCG-based labels composed of more than one CCG category. The second extension tries to build a full parse tree which covers the whole translation output by augmenting glue grammar rules with CCG combinatory operators. We presented experiments on Arabic–English and Chinese–English translation. Our experiments showed that using extended CCG labels achieved the best performance for Arabic–English translation, while using a combination of CCG-augmented glue grammar rules and extended CCG labels led to the best performance for Chinese–English translation.

In future work, we will try to integrate the application probability of CCG combinatory operators performed during glue grammar rule application in the grammaticality feature. Furthermore, we will try to integrate more syntactically aware CCG-based evaluation metrics in tuning and evaluation, which enables a higher accuracy in evaluating improvements on the grammaticality of the translation output.

Acknowledgments

This work is supported by Science Foundation Ireland (Grant No. 07/CE/I1142) as part of the Centre for Next Generation Localisation (www.cngl.ie) at Dublin City University.

References

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

Almaghout, H., J. Jiang, and A. Way. 2011. CCG contextual labels in hierarchical phrase-based SMT. In *proceedings of the 15th conference of the European Association for Machine Translation*, pages 281–288, Leuven, Belgium.

Bangalore, S. and A. Joshi. 1999. Supertagging: An approach to almost parsing. *Computational Linguistics*, 25(2):237–265.

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

Chiang, D. 2005. A hierarchical phrase-based model for statistical machine translation. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, ACL '05, pages 263–270.

Chiang, D. 2010. Learning to translate with source and target syntax. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ACL '10, pages 1443–1452, Uppsala, Sweden.

Hassan, H., K. Sima'an, and A. Way. 2009. A syntactified direct translation model with linear-time decoding. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, EMNLP '09, pages 1182–1191, Singapore.

Koehn, P. 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 (EMNLP-CoNLL-2007)*, pages 868–876, Prague, Czech Republic.

Koehn, P., F.J. Och, and D. Marcu. 2003. Statistical phrase-based translation. In *HLT-NAACL 2003: conference combining Human Language Technology conference series and the North American Chapter of the Association for Computational Linguistics conference series*, pages 48–54, Edmonton, Canada.

Koehn, P. 2004. Statistical significance tests for machine translation evaluation. In *Proceedings of the 2004 Conference of Empirical Methods in Natural Language Processing (EMNLP-2004)*, pages 388–395, Barcelona, Spain.

Och, F.J. 2003. Minimum error rate training in statistical machine translation. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1*, pages 160–167, Sapporo Convention Center, Japan.

Steedman, M. 2000. *The syntactic process*. MIT Press, Cambridge, MA, USA.

Venugopal, A., 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*, pages 37–45, Boulder, Colorado.

Zollmann, A. and A. Venugopal. 2006. Syntax augmented machine translation via chart parsing. In *Proceedings of the Workshop on Statistical Machine Translation*, StatMT '06, pages 138–141.