A Study of Translation Rule Classification for Syntax-based Statistical Machine Translation

Hongfei Jiang, Sheng Li, Muyun Yang and Tiejun Zhao
School of Computer Science and Technology
Harbin Institute of Technology
{hfjiang,lisheng,ymy,tjzhao}@mtlab.hit.edu.cn

Abstract

Recently, numerous statistical machine translation models which can utilize various kinds of translation rules are proposed. In these models, not only the conventional syntactic rules but also the non-syntactic rules can be applied. Even the pure phrase rules are includes in some of these models. Although the better performances are reported over the conventional phrase model and syntax model, the mixture of diversified rules still leaves much room for study. In this paper, we present a refined rule classification system. Based on this classification system, the rules are classified according to different standards, such as lexicalization level and generalization. Especially, we refresh the concepts of the structure reordering rules and the discontiguous phrase rules. This novel classification system may supports the SMT research community with some helpful references.

1 Introduction

Phrase-based statistical machine translation models (Marcu and Wong, 2002; Koehn et al., 2003; Och and Ney, 2004; Koehn, 2004; Koehn et al., 2007) have achieved significant improvements in translation accuracy over the original IBM word-based model. However, there are still many limitations in phrase based models. The most frequently pointed limitation is its inefficacy to modeling the structure reordering and the discontiguous corresponding. To overcome these limitations, many syntax-based SMT models have been proposed (Wu, 1997; Chiang, 2007; Ding et al., 2005; Eisner, 2003; Quirk et al., 2005; Liu et al., 2007; Zhang et al., 2007; Zhang et al., 2008a; Zhang et al., 2008b; Gildea, 2003; Galley et al., 2004; Marcu et al., 2006; Bod, 2007). The basic motivation behind syntax-based model is that the syntax information has the potential to model the structure reordering and discontiguous corresponding by the intrinsic structural generalization ability. Although remarkable progresses have been reported, the strict syntactic constraint (the both sides of the rules should strictly be a subtree of the whole syntax parse) greatly hinders the utilization of the non-syntactic translation equivalents. To alleviate this constraint, a few works have attempted to make full use of the non-syntactic rules by extending their syntax-based models to more general frameworks. For example, forest-to-string transformation rules have been integrated into the tree-to-string translation framework by (Liu et al., 2006; Liu et al., 2007). Zhang et al. (2008a) made it possible to utilize the non-syntactic rules and even the phrases which are used in phrase based model by advancing a general tree sequence to tree sequence framework based on the tree-to-tree model presented in (Zhang et al., 2007). In these models, various kinds of rules can be employed. For example, as shown in Figure 1 and Figure 2, Figure 1 shows a Chinese-to-English sentence pair with syntax parses on both sides and the word alignments (dotted lines). Figure 2 lists some of the rules which can be extracted from the sentence pair in Figure 1 by the system used in (Zhang et al., 2008a). These rules includes not only conventional syntax rules but also the tree sequence rules (the multi-headed syntax rules). Even the phrase rules are adopted by

Proceedings of SSST-3, Third Workshop on Syntax and Structure in Statistical Translation, pages 45–50,
Boulder, Colorado, June 2009.

the system. Although the better performances are reported over the conventional phrase-based model and syntax-based model, the mixture of diversified rules still leaves much room for study. Given such a hybrid rule set, we must want to know what kinds of rules can make more important contributions to the overall system performance and what kinds of rules are redundant compared with the others. From engineering point of view, the developers may concern about which kinds of rules should be preferred and which kinds of rules could be discard without too much decline in translation quality. However, one of the precondition for the investigations of these issues is what are the "rule categories"? In other words, some comprehensive rule classifications are necessary to make the rule analyses feasible. The motivation of this paper is to present such a rule classification.

2 Related Works

A few researches have made some exploratory investigations towards the effects of different rules by classifying the translation rules into different subcategories (Liu et al., 2007; Zhang et al., 2008a; DeNeefe et al., 2007). Liu et al. (2007) differentiated the rules in their tree-to-string model which integrated with forest[1]-to-string into fully lexicalized rules, non-lexicalized rules and partial lexicalized rules according to the lexicalization levels. As an extension, Zhang et al. (2008a) proposed two more categories: Structure Reordering Rules (SRR) and Discontiguous Phrase Rules (DPR). The SRR stands for the rules which have at least two non-terminal leaf nodes with inverted order in the source and target side. And DPR refers to the rules having at least one non-terminal leaf node between two terminal leaf nodes. (DeNeefe et al., 2007) made an illuminating breakdown of the different kinds of rules. Firstly, they classify all the GHKM[2] rules (Galley et al., 2004; Galley et al., 2006) into two categories: lexical rules and non-lexical rules. The former are the rules whose source side has no source words. In other words, a non-lexical rule is a purely abstract rule. The latter is the complementary set of the former. And then lexical rules are classified further into phrasal rules and non-phrasal rules. The *phrasal rules* refer to the rules whose source side and the yield of the target side contain exactly one contiguous phrase each. And the one or more non-terminals can be placed on either side of the phrase. In other words, each phrasal rule can be simulated by the conjunction of two more phrase rules. (DeNeefe et al., 2007) classifies non-phrasal rules further into structural rules, re-ordering rules, and non-contiguous phrase rules. However, these categories are not explicitly defined in (DeNeefe et al., 2007) since out of its focus. Our proposed rule classification is inspired by these works.

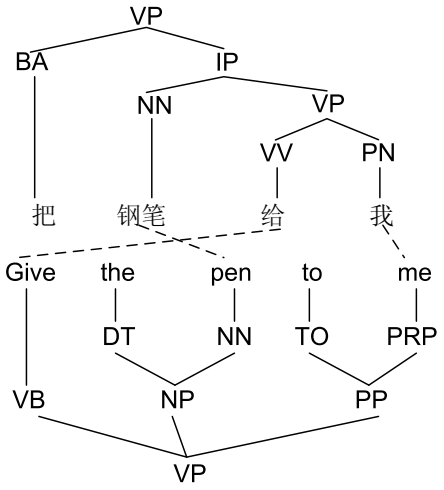

Figure 1: A syntax tree pair example. Dotted lines stands for the word alignments.

3 Rules Classifications

Currently, there have been several classifications in SMT research community. Generally, the rules can be classified into two main groups according to whether syntax information is involved: bilingual phrases (Phrase) and syntax rules (Syntax). Further, the syntax rules can be divided into three categories according to the lexicalization levels (Liu et al., 2007; Zhang et al., 2008a):

1) Fully lexicalized (FLex): all leaf nodes in both the source and target sides are lexicons (terminals)

2) Unlexicalized (ULex): all leaf nodes in both the

[1] A "forest" means a sub-tree sequence derived from a given parse tree

[2] One reviewer asked about the acronym **GHKM**. We guess it is an acronym for the authors of (Galley et al., 2004): Michel **G**alley, Mark **H**opkins, Kevin **K**night and Daniel **M**arcu.

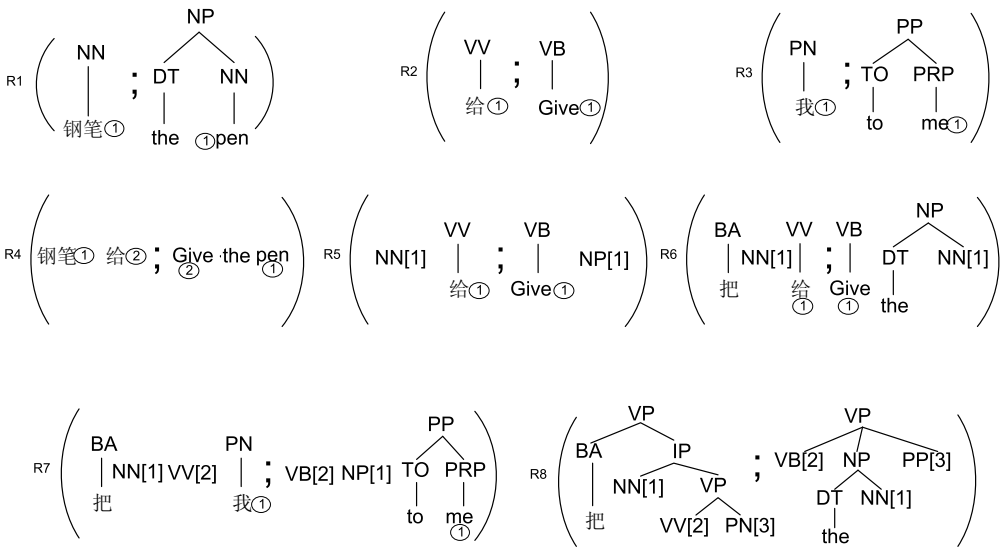

Figure 2: Some rules can be extracted by the system used in (Zhang et al., 2008a) from the sentence pair in Figure 1.

source and target sides are non-lexicons (non-terminals)

3) Partially lexicalized (PLex): otherwise.

In Figure 2, R_1-R_3 are FLex rules, and R_5-R_8 are PLex rules.

Following (Zhang et al., 2008b), a syntax rule r can be formalized into a tuple

$$< \xi_s, \xi_t, A_T, A_{NT} >$$

, where ξ_s and ξ_t are tree sequences of source side and target side respectively, A_T is a many-to-many correspondence set which includes the alignments between the terminal leaf nodes from source and target side, and A_{NT} is a one-to-one correspondence set which includes the synchronizing relations between the non-terminal leaf nodes from source and target side.

Then, the syntax rules can also fall into two categories according to whether equipping with generalization capability (Chiang, 2007; Zhang et al., 2008a):

1) Initial rules (Initial): all leaf nodes of this rule are terminals.

2) Abstract rules (Abstract): otherwise, i.e. at least one leaf node is a non-terminal.

A non-terminal leaf node in a rule is named an **abstract node** since it has the generalization capability. Comparing these two classifications for syntax rules, we can find that a FLex rule is a initial rule when ULex rules and PLex rules belong to abstract rules.

These classifications are clear and easy for understanding. However, we argue that they need further refinement for in-depth study. Specially, more refined differentiations are needed for the abstract rules (ULex rules and PLex rules) since they play important roles for the characteristic capabilities which are deemed to be the advantages over the phrase-based model. For instance, the potentials to model the structure reordering and the discontiguous correspondence. The Structure Reordering Rules (SRR) and Discontiguous Phrase Rules (DPR) mentioned by (Zhang et al., 2008a) can be regarded as more in-depth classification of the syntax rules. In (Zhang et al., 2008a), they are described as follows:

Definition 1: The **Structure Reordering Rule (SRR)** refers to the structure reordering rule that has at least two non-terminal leaf nodes with inverted order in the source and target side.

Definition 2: The **Discontiguous Phrase Rule (DPR)** refers to the rule having at least one non-terminal leaf node between two lexicalized leaf nodes.

Based on these descriptions, R_7, R_8 in Figure 2 belong to the category of SRR and R_6, R_7 fall into the category of DPR. Although these two definitions are easy implemented in practice, we argue that the definition of SRR is not complete. The reordering rules involving the reordering between content word terminals and non-terminal (such as R_5 in Figure 2) also can model the useful structure reorderings. Moreover, it is not uncommon that a rule demonstrates the reorderings between two non-terminals as well as the reorderings between one non-terminal and one content word terminal. The reason for our emphasis of content word terminal is that the reorderings between the non-terminals and function word are less meaningful.

One of the theoretical problems with phrase based SMT models is that they can not effectively model the discontiguous translations and numerous attempts have been made on this issue (Simard et al., 2005; Quirk and Menezes, 2006; Wellington et al., 2006; Bod, 2007; Zhang et al., 2007). What seems to be lacking, however, is a explicit definition to the discontiguous translation. The definition of DPR in (Zhang et al., 2008a) is explicit but somewhat rough and not very accurate. For example, in Figure 3(a), non-terminal node pair ([0,'爱'], [0,'love']) is surrounded by lexical terminals. According to Definition 2, it is a DPR. However, obviously it is not a discontiguous phrase actually. This rule can be simulated by conjunctions of three phrases ('我', 'I'; '爱', 'love'; '你','you'). In contrast, the translation rule in Figure 3(b) is an actual discontiguous phrase rule. The English correspondences of the Chinese word '关' is dispersed in the English side in which the correspondence of Chinese word '灯' is inserted. This rule can not be simulated by any conjunctions of the sub phrases. It must be noted that the discontiguous phrase ('关'-"switch ... off") can not be abstracted under the existing synchronous grammar frameworks. The fundamental reason is that the corresponding parts should be abstracted in the same time and lexicalized in the same time. In other words, the discontiguous phrase can not be modeled by the permutation between non-terminals (abstract nodes). Another point to notice is that our focus in this paper is the ability demonstrated by the abstract rules. Thus, we do not pay much attentions to the reorderings and discontiguous phrases involved in the

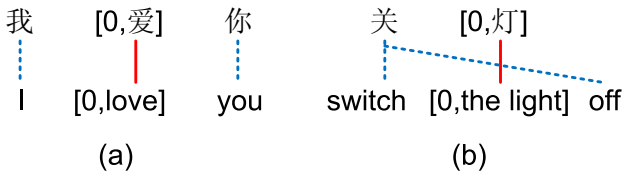

Figure 3: Examples for demonstrating the actual discontiguous phrase. (a) is a negative example for the definition of DPR in (Zhang et al., 2008a), (b) is a actual discontiguous phrase rule.

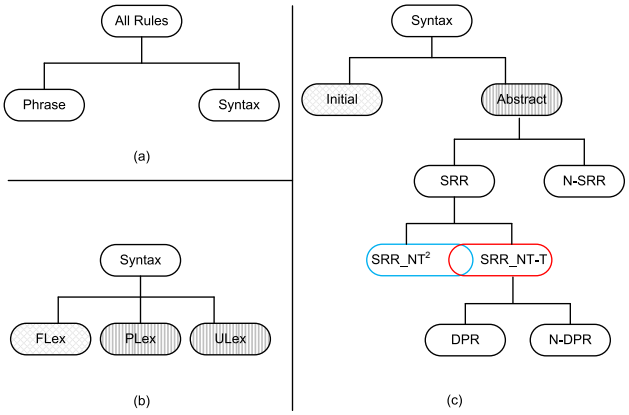

Figure 4: The rule classifications used in this paper. (a) shows that the rules can be divided into phrase rules and syntax rules according to whether a rule includes the syntactic information. (b) illustrates that the syntax rules can be classified into three kinds according to the lexicalization level. (c) shows that the abstract rules can be classified into more refined sub-categories.

phrase rules (e.g. "关 灯"-"switch the light off") since they lack the generalization capability. Therefore, the discontiguous phrase is limited to the relation between non-terminals and terminals.

On the basis of the above analyses, we present a novel classification system for the abstract rules based on the crossings between the leaf node alignment links. Given an abstract rule $r =< \xi_s, \xi_t, A_T, A_{NT} >$, it is

1) a Structure Reordering Rule (SRR), if $\exists$ a link $l \in A_{NT}$ is crossed with a link $l' \in \{A_T \cap A_{NT}\}$
   a) a SRR_NT^2 rule, if the link $l' \in A_{NT}$
   b) a SRR_NT-T rule, if the link $l' \in A_T$

2) not a Structure Reordering Rule (N-SRR), otherwise.

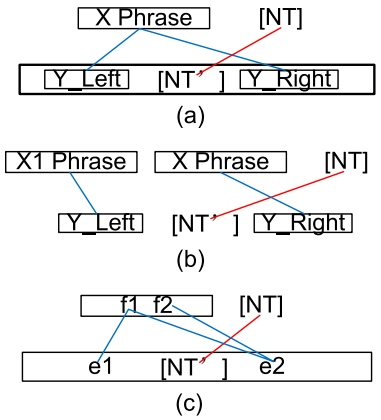

Figure 5: The patterns to show the characteristics of discontiguous phrase rules.

Note that the intersection of SRR_NT^2 and SRR_NT-T is not necessary an empty set, i.e. a rule can be both SRR_NT^2 and SRR_NT-T rule.

The basic characteristic of the discontiguous translation is that the correspondence of one non-terminal NT is inserted among the correspondences of one phrase X. Figure 5 (a) illustrates this situation. However, this characteristic can not support necessary and sufficient condition. For example, if the phrase X can be divided like Figure 5 (b), then the rule in Figure 5 (a) is actually a reordering rule rather than a discontiguous phrase rule. For sufficient condition, we constrain that the phrase $X = w_i \dots w_j$ need to satisfy the requirement: w_i should be connected with w_j through word alignment links (A word is connected with itself). In Figure 5(c), f_1 is connected with f_2 when NT' is inserted between e_1 and e_2. Thus, the rule in Figure 5(c) is a discontiguous phrase rule.

Definition 3: Given an abstract rule $r =< \xi_s, \xi_t, A_T, A_{NT} >$, it is a **Discontiguous Phrase** iff $\exists$ two links l_{t1}, l_{t2} from A_T and a link l_{nt} from A_{NT}, satisfy: l_{t1}, l_{t2} are emitted from the same word and l_{t1} is crossed with l_{nt} when l_{t2} is not crossed with l_{nt}.

Through Definition 3, we know that the DPR is a sub-set of the SRR_NT-T.

4 Conclusions and Future Works

In this paper, we present a refined rule classification system. Based on this classification system, the rules are classified according to different standards, such as lexicalization level and generalization. Especially, we refresh the concepts of the structure reordering rules and the discontiguous phrase rules. This novel classification system may supports the SMT research community with some helpful references.

In the future works, aiming to analyze the rule contributions and the redundances issues using the presented rule classification based on some real translation systems, we plan to implement some synchronous grammar based syntax translation models such as the one presented in (Liu et al., 2007) or in (Zhang et al., 2008a). Taking such a system as the experimental platform, we can perform comprehensive statistics about distributions of different rule categories. What is more important, the contribution of each rule category can be evaluated seriatim. Furthermore, which kinds of rules are preferentially applied in the 1-best decoding can be studied. All these investigations could reveal very useful information for the optimization of rule extraction and the improvement of the computational models for synchronous grammar based machine translation.

Acknowledgments

This work is supported by the Key Program of National Natural Science Foundation of China (60736014), and the Key Project of the National High Technology Research and Development Program of China (2006AA010108).

References

Rens Bod. 2007. Unsupervised syntax-based machine translation: The contribution of discontiguous phrases. In *Proceedings of Machine Translation Summit XI 2007*,Copenhagen, Denmark.

David Chiang. 2007. Hierarchical phrase-based translation. In *computational linguistics*, 33(2).

Ding, Y. and Palmer, M. 2005. Machine translation using probabilistic synchronous dependency insertion grammars In *Proceedings of ACL*.

DeNeefe, S. and Knight, K. and Wang, W. and Marcu, D. 2007. What can syntax-based MT learn from phrase-based MT? In *Proceedings of EMNLP/CONLL*.

Michel Galley, Mark Hopkins, Kevin Knight and Daniel Marcu. 2004. What's in a translation rule? In *Proceedings of NAACL-HLT 2004*, pages 273-280.

Galley, M. and Graehl, J. and Knight, K. and Marcu, D. and DeNeefe, S. and Wang, W. and Thayer, I. 2006. Scalable inference and training of context-rich syntactic translation models In *Proceedings of ACL-COLING*

Daniel Gildea 2003. Loosely Tree-Based Alignment for Machine Translation. In *Proceedings of ACL 2003*, pages 80-87.

Jason Eisner. 2003. Learning non-isomorphic tree mappings for machine translation. In *Proceedings of ACL 2003*.

Philipp Koehn, Franz Joseph Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *Proceedings of HLT/NAACL 2003*, pages 127-133, Edmonton, Canada, May.

Philipp Koehn. 2004. Pharaoh: a beam search decoder for phrase-based statistical machine translation models. In *Proceedings of the Sixth Conference of the Association for Machine Translation in the Americas*, pages 115-124.

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, Evan Herbst. 2007. Moses: open source toolkit for statistical machine translation. ACL 2007, demonstration session, Prague, Czech Republic, June 2007.

Yang Liu, Qun Liu, Shouxun Lin. 2006. Tree-to-string alignment template for statistical machine translation. In *Proceedings of ACL-COLING*.

Yang Liu, Yun Huang, Qun Liu, and Shouxun Lin. 2007. Forest-to-string statistical translation rules. In *Proceedings of ACL 2007*, pages 704-711.

Daniel Marcu and William Wong. 2002. A phrase based, joint probability model for statistical machine translation. In *Proceedings of EMNLP*.

Daniel Marcu, Wei Wang, Abdessamad Echihabi, and Kevin Knight. 2006. SPMT: Statistical machine translation with syntactified target language Phrases. In *Proceedings of EMNLP*.

Franz Josef Och and Hermann Ney. 2000. Improved statistical alignment models. In *Proceedings of ACL 2000*, pages 440-447.

Franz Josef Och and Herman Ney. 2004. The alignment template approach to statistical machine translation. *Computational Linguistics*, 30(4):417-449.

Chris Quirk, Arul Menezes, and Colin Cherry. 2005. Dependency treelet translation: Syntactically informed phrasal SMT. In *Proceedings of ACL 2005*, pages 271-279, Ann Arbor, Michigan, June.

Chris Quirk and Arul Menezes. 2006. Do we need phrases? Challenging the conventional wisdom in Statistical Machine Translation. In *Proceedings of HLT/NAACL*

Simard, M. and Cancedda, N. and Cavestro, B. and Dymetman, M. and Gaussier, E. and Goutte, C. and Yamada, K. and Langlais, P. and Mauser, A. 2005. Translating with non-contiguous phrases. In *Proceedings of HLT-EMNLP*, volume 2, pages 901-904.

Benjamin Wellington, Sonjia Waxmonsky and I. Dan Melamed. 2006. Empirical Lower Bounds on the Complexity of Translational Equivalence. In *Proceedings of ACL-COLING 2006*, pages 977-984.

Dekai Wu. 1997. Stochastic inversion transduction grammars and bilingual parsing of parallel corpora. In *Proceedings of ACL 1997. Computational Linguistics*, 23(3):377-403.

Min Zhang, Hongfei Jiang, Ai Ti AW, Jun Sun, Sheng Li, and Chew Lim Tan. 2007. A tree-to-tree alignment-based model for statistical machine translation. In *Proceedings of Machine Translation Summit XI 2007*,Copenhagen, Denmark.

Min Zhang, Hongfei Jiang, Ai Ti AW, Haizhou Li, Chew Lim Tan and Sheng Li. 2008a. A tree sequence alignment-based tree-to-tree translation model. In *Proceedings of ACL-HLT*

Min Zhang, Hongfei Jiang, Haizhou Li, Ai Ti AW, and Sheng Li. 2008b. Grammar Comparison Study for Translational Equivalence Modeling and Statistical Machine Translation. In *Proceedings of Coling*