Linguistically-motivated Tree-based Probabilistic Phrase Alignment

Toshiaki Nakazawa **Sadao Kurohashi**

Graduate School of Informatics, Kyoto University
Yoshida-honmachi, Sakyo-ku
Kyoto, 606-8501, Japan
nakazawa@nlp.kuee.kyoto-u.ac.jp kuro@i.kyoto-u.ac.jp

Abstract

In this paper, we propose a probabilistic phrase alignment model based on dependency trees. This model is linguistically-motivated, using syntactic information during alignment process. The main advantage of this model is that the linguistic difference between source and target languages is successfully absorbed. It is composed of two models: Model1 is using content word translation probability and function word translation probability; Model2 uses dependency relation probability which is defined for a pair of positional relations on dependency trees. Relation probability acts as tree-based phrase reordering model. Since this model is directed, we combine two alignment results from bi-directional training by symmetrization heuristics to get definitive alignment. We conduct experiments on a Japanese-English corpus, and achieve reasonably high quality of alignment compared with word-based alignment model.

1 Introduction

Most of statistical machine translation (SMT) systems are based on "word-based" alignment method starting with IBM models (Brown et al., 1993). Based on the word alignment results, some enhanced and successful models which extract phrases have been proposed and established the state-of-the-art Phrase-Based SMT models (Koehn et al., 2003). Another approaches incorporate syntactic information by parsing source or target sentences (Quirk et al., 2005; Galley et al., 2006; Cowan et al., 2006). Chiang (2005) proposed hierarchical phrase-based translation model which was based on a weighted synchronous CFG. This model could handle more complex linguistic phenomena which sequences of words could not do. However, all of these models are on the basis of IBM models, which do not essentially consider syntactic information. Word-based alignment method works well for language pairs whose linguistic structure is not so different from each other (such as English v.s. European languages), but not for language pairs with great difference in linguistic structure (such as Japanese-English). For a linguistically different pairs, deeper natural language processing (NLP) analysis is necessary even during alignment process.

Watanabe et al. (2000) and Menezes and Richardson (2001) proposed a structural alignment methods. These methods use heuristic rules when resolving correspondence ambiguities, not considering the consistency between two dependency structure as a whole. Yamada and Knight (2001) and Gildea (2003) proposed a tree-based probabilistic alignment methods. These methods reorder, insert or delete sub-trees of one side to reproduce the other side. The constraints of using syntactic information is often too rigid. Yamada and Knight flattened the trees by collapsing nodes, Gildea cloned the sub-trees to deal with the problem.

Our method proposed in this paper does not require any operations for controlling tree structures, just align phrase-to-phrase on dependency structure. Though our model is more simple than well-known IBM Model3 or greater, our model can achieve high accuracy of alignment and high quality of translation. We propose a probabilistic tree-based phrase

alignment model. Since it uses dependency structure, our method can overcome the difference of languages even if they are structurally different from each other, which simple statistical word alignment models are not able to. It can be said as a tree-based phrase reordering model.

In section 2, our proposed model is illustrated in a general way, and in the following section we explain the model in detail using simple examples. The symmetrization algorithm is shown in section 4. We performed some experiments to evaluate our proposal, which are reported in section 5. Finally, we give a short conclusion and future work.

2 Tree-based Probabilistic Phrase Alignment Model

We suppose Japanese for source language and English for target language in the description of our model. Note that the model is not specialized for this language pair, it can be applied to any language pairs.

2.1 Dependency Analysis of Sentences

Since our model utilizes dependency tree structures, both source and target sentences are parsed at first. Japanese sentences are converted into dependency structures using the morphological analyzer JUMAN (Kurohashi et al., 1994), and the dependency analyzer KNP (Kurohashi and Nagao, 1994). Japanese dependency structure consists of nodes which correspond to content words. Function words such as post-positions, affixes, and auxiliary verbs are included in the nodes.

For English sentences, Charniak's nlparser is used to convert them into phrase structures (Charniak and Johnson, 2005), and then they are transformed into dependency structures by handmade rules defining head words for phrases. As is the case with Japanese, each node in this dependency tree consists of a content word and related function words[1]. We define function words as the words with tags of "IN"(preposition or subordinating conjunction), "TO", "MD"(modal), "CC"(coordinating conjunction) by nlparser.

[1]There would be some special cases that a phrase has no content word. See a phrase "and" in figure 1. Also, there would be a phrase which has more than one content words.

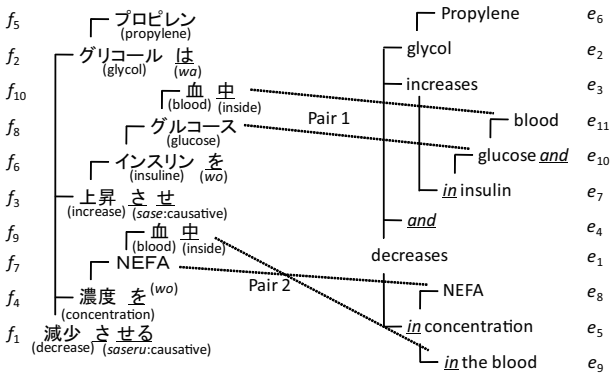

Figure 1: Example of dependency trees.

Figure 1 shows an example of dependency structure. Each node of the tree corresponds to a linguistic phrase. Underlined words are handled as function words, others are content words. Our model uses the linguistic phrase as an unit of alignment rather than a word. The root of a tree is placed at the extreme left and phrases are placed from top to bottom.

2.2 Tree-based Model

In IBM models (Brown et al., 1993), the best alignment $\hat{a}$ between given source (French) sentence f and target (English) sentence e is acquired by the following equation:

$$\hat{a} = \underset{a}{\operatorname{argmax}}\, p(f|e,a) \cdot p(a|e) \qquad (1)$$

where $p(f|e,a)$ is called as "lexicon probability" and $p(a|e)$ is called as "alignment probability".

Since our model is based on dependency tree structure, we can find the best alignment $\hat{a}$ between given source (Japanese) tree T_f and target (English) tree T_e as follows (while focusing on Japanese and English, we use the common notation for indexes as f and e):

$$\hat{a} = \underset{a}{\operatorname{argmax}}\, p(T_f|T_e,a) \cdot p(a|T_e) \qquad (2)$$

Suppose T_f consists of J nodes (equivalent to phrase) $f_1, ..., f_J$, T_e consists of I nodes $e_1, ..., e_I$. In IBM models, f_1 represents the first word of the sentence and f_J represents the last word. On the other hand, since we are handling dependency tree structures, f_1 represents the root node of the tree and f_J represents one of the leaves of the tree in

our model. The parent node is denoted with minus (-) mark on its index, that is, the parent node of f_j is f_{j-}. With these notations, we decompose lexicon probability and alignment probability as follows:

$$p(T_f|T_e, a) \approx \prod_{j=1}^{J} p(f_j|e_{a_j}) \quad (3)$$

$$p(a|T_e) \approx \prod_{j=1}^{J} p(rel(e_{a_j}, e_{a_{j-}})|rel(f_j, f_{j-})) \quad (4)$$

a_j denotes the phrase number which source phrase f_j corresponds to, then e_{a_j} denotes a target side phrase which corresponds to source phrase f_j. $p(f_j|e_{a_j})$ is a phrase translation probability. Words in a phrase are categorized into two groups: content words or function words. We define the phrase translation probability as a product of content word translation probability and function word translation probability.

$$p(f_j|e_{a_j}) = p_{cont.}(f_j|e_{a_j}) \cdot p_{func.}(f_j|e_{a_j}) \quad (5)$$

Here, two or more content or function words in one phrase are considered together. In figure 1, two Japanese function words "さ" and "せ" are combined and considered as one word "させ".

$rel(f_j, f_{j-})$ represents the dependency relation between f_j and f_{j-} on dependency tree. In case no corresponding candidate is found for f_{j-}, we refer the parent node of f_{j-}, and this is repeated until a node with corresponding candidate is found, then the relation between this node and f_j is considered.

Take Pair 1 in figure1 for example. f_{10} is a child of its parent f_8, and corresponding English node e_{11} is also a child of its parent e_{10}. For Pair 2 as an another example, the relation in source side is same to Pair 1 (child), as for target side, e_9 is a child of e_5, which is a parent of e_8 ($e_8 \xrightarrow{\text{parent}} e_5 \xrightarrow{\text{child}} e_9$).

$p(rel(e_{a_j}, e_{a_{j-}})|rel(f_j, f_{j-}))$ represents the dependency relation probability, which is assigned to a pair of source and target side relations between two phrases. This can be said a tree-based reordering model.

The unknown parameters θ are determined by maximizing the likelihood on the parallel training corpus which consists of S parallel sentences:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \prod_{s=1}^{S} \sum_{a} p(T_f|T_e, a) \cdot p(a|T_e) \quad (6)$$

EM algorithm is adopted for the parameter estimation, and best alignment $\hat{a}$ will be found. We train the model bi-directionally and acquire two best alignment results. These results are combined into one definitive alignment using symmetrization heuristics.

3 Model Training

Our proposed tree-based phrase alignment model is composed of three probabilities, content word translation probability, function word translation probability, and dependency relation probability as shown in previous section. The content / function word translation probabilities are independent from other phrases than the focusing phrase. Dependency relation probability depends on which phrase the parent phrase is correspond to in the target tree.

In a first step, the model learning on two translation probabilities only (Model1) is conducted. Next, the enhanced model including dependency relation probability (Model2) is learned with the parameter values learned in Model1 as an initial parameters. Model1 can be efficiently learned without approximation like IBM model1 and 2. For Model2, some approximation is necessary as IBM model3 or greater. Here, we adopt beam-search algorithm.

3.1 Model1

Each phrase in source side f_j ($1 \leq j \leq J$) can correspond to an arbitrary phrase in target side e_i ($1 \leq i \leq I$) or a NULL phrase (e_0), if f_j does not correspond to any phrase, independently of other source phrase. A probability of one possible alignment $p(a, T_f|T_e)$ is calculated as follows:

$$p(a, T_f|T_e) = \prod_{j=1}^{J} p_{cont.}(f_j|e_{a_j}) \cdot p_{func.}(f_j|e_{a_j}) \quad (7)$$

Also, $p(T_f|T_e)$ is calculated as:

$$p(T_f|T_e) = \sum_{a} p(a, T_f|T_e) \quad (8)$$

Since there are $(I+1)^J$ possible alignments, we have to do $2J * (I+1)^J$ arithmetic operations to evaluate this expression. However, the expression

above can be transformed like:

$$\sum_{a}\prod_{j=1}^{J} p(f_j|e_{a_j}) = \prod_{j=1}^{J}\sum_{i=0}^{l} p(f_j|e_{a_j}) \qquad (9)$$

This last expression only requires a quadratic number of arithmetic operations to evaluate, therefore no approximation is needed.

As initial parameters, we use uniform probabilities.

3.2 Model2

In the first iteration of Model2, dependency relation probabilities are set to be uniform, and learning result of translation probabilities in Model1 is used as the initial parameters. Dependency relation probabilities are calculated according to the equation 4. It is impossible to enumerate all the possible alignment, we consider only a subset of "good-looking" alignments using beam-search algorithm.

Dependency relation probability refers to the relation between a source side phrase f_j and its parent phrase f_{j-} ($rel(f_j, f_{j-})$), and the relation between corresponding target side phrases e_{a_j} and $e_{a_{j-}}$ ($rel(e_{a_j}, e_{a_{j-}})$). f_{j-} represents the nearest parent phrase which is not aligned to NULL. Dependency relation ($rel(P_1, P_2)$) indicates a path to a phrase (P_2) to another one (P_1). We use the notations below.

- "c-" if P_1 is a pre-child of P_2
- "c+" if P_1 is a post-child of P_2
- "p+" if P_2 is a pre-child of P_1
- "p-" if P_2 is a post-child of P_1
- "INCL" if P_1 and P_2 are same phrase

Relations between two phrases which are far more than 1 node from each other are expressed by putting these marks.

A head phrase of a sentence is supposed to depends on the imaginary root node which is a start point of beam-search. It corresponds to the SOS (start-of-sentence) in word-base models. Figure 2 is an abstract example where beam width = 5. f_1 and e_1 which are the head phrases depend on imaginary root node. First, we focus on f_1 which has three correspondence candidates, e_1, e_2 and NULL (e_0). Alignment probabilities are calculated as $p(f_1|e_1) \cdot p(\text{ROOT}|\text{ROOT})$, $p(f_1|e_2) \cdot p(\text{ROOT}|\text{ROOT;c})$, $p(f_1|\text{NULL}) \cdot p(\text{ROOT}|\text{NULL})$ respectively. Here, "ROOT" means that the phrase is a child of imaginary root node.

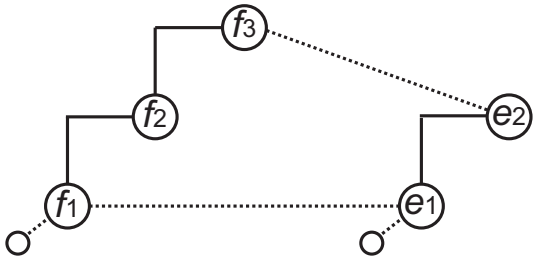

Figure 2: Abstraction of alignment.

As for f_2 in the next place, there are also three correspondence candidates. Consequently, there should be nine alignment candidates in total (table 1 shows each probability). These nine are sorted by the probabilities and only 5-best alignments are preserved. When we consider about f_3, we take the five alignment candidates into account, and this results in generating 15 (5*3) alignment candidates, again we discard except top 5 probable alignments. These steps are repeated until all the source side phrases are aligned to any one target side phrase. Parameters are updated using only the last 5-best alignments.

4 Symmetrization Algorithm

After the model learning is finished in each direction, two alignment results can be acquired. With these results, we generate final alignment by combining two alignment results using heuristic rules like Koehn et al. (2003). The differences between theirs and ours are: 1) alignment unit is phrase rather than word, 2) using tree structure during growing process, 3) n-best alignment results in each direction are considered.

Reduplication N-best alignment results acquired in each model learning, 2n alignments in total are reduplicated, then possible alignment points are scored from 1 to 2n (see figure 3 where n=5). In descending order of the score, possible alignment points are adopted as definitive alignment points if there is no point with higher score than the focusing point in the same row or column. In the example, dark colored six points are adopted. If we choose n=1, this algorithm is same to simple "intersection".

Table 1: Probability calculation example.

f_1	f_2	Probability
e_1	e_1	$p(f_1\|e_1) \cdot p(\text{ROOT}\|\text{ROOT}) \cdot p(f_2\|e_1) \cdot p(\text{INCL}\|\text{c-})$
e_1	e_2	$p(f_1\|e_1) \cdot p(\text{ROOT}\|\text{ROOT}) \cdot p(f_2\|e_2) \cdot p(\text{c-}\|\text{c-})$
e_1	NULL	$p(f_1\|e_1) \cdot p(\text{ROOT}\|\text{ROOT}) \cdot p(f_2\|\text{NULL}) \cdot p(\text{NULL}\|\text{c-})$
e_2	e_1	$p(f_1\|e_2) \cdot p(\text{ROOT;c-}\|\text{ROOT}) \cdot p(f_2\|e_1) \cdot p(\text{p+}\|\text{c-})$
e_2	e_2	$p(f_1\|e_2) \cdot p(\text{ROOT;c-}\|\text{ROOT}) \cdot p(f_2\|e_2) \cdot p(\text{INCL}\|\text{c-})$
e_2	NULL	$p(f_1\|e_2) \cdot p(\text{ROOT;c-}\|\text{ROOT}) \cdot p(f_2\|\text{NULL}) \cdot p(\text{NULL}\|\text{c-})$
NULL	e_1	$p(f_1\|\text{NULL}) \cdot p(\text{NULL}\|\text{ROOT}) \cdot p(f_2\|e_1) \cdot p(\text{ROOT}\|\text{ROOT;c-})$
NULL	e_2	$p(f_1\|\text{NULL}) \cdot p(\text{NULL}\|\text{ROOT}) \cdot p(f_2\|e_2) \cdot p(\text{ROOT;c-}\|\text{ROOT;c-})$
NULL	NULL	$p(f_1\|\text{NULL}) \cdot p(\text{NULL}\|\text{ROOT}) \cdot p(f_2\|\text{NULL}) \cdot p(\text{NULL}\|\text{ROOT;c-})$

Growing Possible points around the definitive points are adopted. Here, "around" does not mean right before or after the phrases, it means the continuous phrases on the dependency trees. Even if a right before or after the definitively aligned phrase is possible point, it would not be adopted when they are not directly connected. In the example, there occurs three growing alignments colored in gray indicated with thick arrows.

Handling Isolation Not adopted possible points which are not aligned to any phrase in both source and target language are adopted. Left bottom isolated part (gray colored points) in the example is adopted by this process.

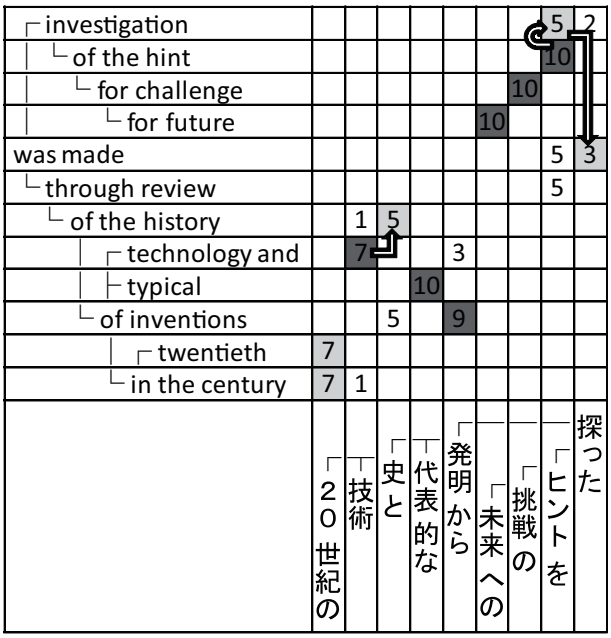

Figure 3: Example of growing the alignment points.

5 Experiments and Results

We conducted both alignment and translation evaluation experiments. JST [2] Japanese-English paper abstract corpus consists of 1M parallel sentences were used for model training. This corpus was constructed from 2M Japanese-English paper abstract corpus belongs to JST by NICT [3] using the method of Uchiyama and Isahara (2007).

We made gold-standard alignment for 100 sentence pairs among the 1M parallel sentences by hand. The annotations are only sure (S) alignments (no possible (P) alignments (Och and Ney, 2003)). The unit of evaluation was morpheme-base for Japanese and word-base for English.

We used precision, recall, and F-measure as evaluation criterion. The evaluation results are shown in table 2. We used "3-best-grow" and "5-best-grow"

[2] http://www.jst.go.jp/

[3] http://www.nict.go.jp/

symmetrization heuristics which uses 3-best and 5-best alignment results respectively from each direction and extend the alignment points using heuristic rules explained in the previous section.

For comparison, we segmented the data using the morphological analyzer JUMAN (Kurohashi et al., 1994) for Japanese sentences and created alignments using freely available word alignment tool GIZA++ (Och and Ney, 2003). We conducted word alignment bidirectionally with its default parameters and merged them using seven types of symmetrization heuristics (Koehn et al., 2003) shown in table 2. Training are run on original forms of words for both proposed model and GIZA++.

For translation evaluation, we use 500 paper abstract sentences which are parts of JST corpus. Note that test sentences are not included in training corpus. As a decoder, we used state-of-the-art phrase-based SMT toolkit Moses (Koehn et al., 2007) with its default options except for phrase table limit (20 → 10) and distortion limit (6 → -1 means infinite). Evaluation was done with all the punctuations being deleted and case-insensitively. The BLEU scores of each alignment methods are shown in table 2, in the last column.

Actually, it is hard to integrate proposed alignment results into Moses decoder because our model is based on "linguistic phrase". If we align all words to all words in a corresponding two phrases, Moses would fail to translate a content word with different function words from the learned phrase pair. To avoid this problem to some extent, we aligned content words to content words, and function words to function words separately, in addition, no article in English sentences was aligned. Of course this is not sufficient at all, Japanese has many kinds of function words which English does not have. Even in translation process, it is necessary to handle the function words carefully.

6 Discussions

Table 2 shows that our proposed model could achieve reasonably high accuracy of alignment, and it is better than word-base models. As an example, word-base alignment result in figure 4 fails to find the correspondence between content words "上昇" and "increase", what is worse, "increase" is incorrectly aligned to a word "は (wa)" which is a function word. This is because, as mentioned in section 1, the statistical methods work well for language pairs that are not so different regarding language structure. Japanese and English have significantly different structure: Japanese sentences consist of SOV word order, but English word order is SVO. For such language pair as Japanese and English, deeper sentence analysis using NLP resources is necessary and useful, like in our method. By using the tree structure in figure 5, these two words (phrases) are correctly aligned.

Table 2: Experimental results of alignment and translation.

	Pre	Rec	F	BLEU
Proposed				
1-best	**90.92**	41.69	57.17	12.73
1-best-grow	83.30	54.33	65.76	14.97
3-best-grow	81.21	56.52	66.65	15.09
5-best-grow	80.59	**57.33**	**67.00**	15.40
GIZA++				
intersection	88.14	40.18	55.20	16.35
grow	83.50	49.65	62.27	17.05
grow-final	67.19	56.91	61.63	17.85
grow-final-and	78.00	52.93	63.06	17.70
grow-diag	77.34	53.18	63.03	**17.89**
grow-diag-final	67.24	56.63	61.48	17.80
grow-diag-final-and	74.95	54.26	62.95	17.76

Even if the alignment accuracy was improved, this did not lead to improve the translation quality referring to the BLEU score, BLEU score of our proposed model is worse than that of word-base models. One reason of this is that, as mentioned above, the infelicity of integrating our alignment results into Moses decoder. Another reason is that BLEU is essentially insensitive to syntactic structure. The translation result may indeed better from the point of dependency structure. We need to try parsing base line output and the output of the realigned system and see if the parsing results improve.

Some of recent studies suggest that there is less relationship between alignment quality and translation results (Lopez and Resnik, 2006; Ayan and Dorr, 2006). Even if the contribution to translation quality is small, there is no doubt that better alignment quality leads to better translation, which

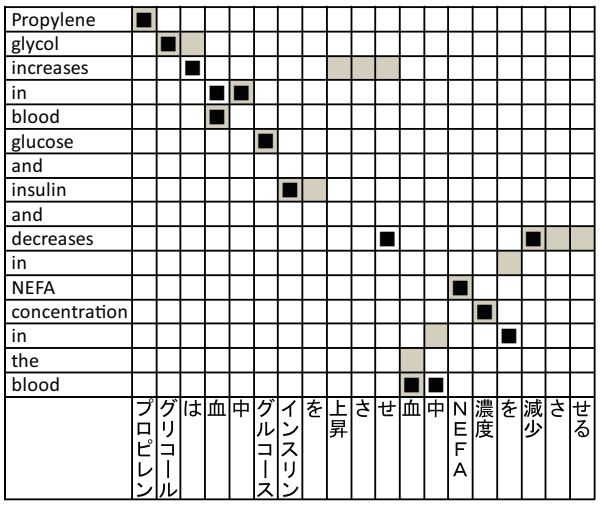

Figure 4: Result of word-base alignment (grow-diag-final-and).

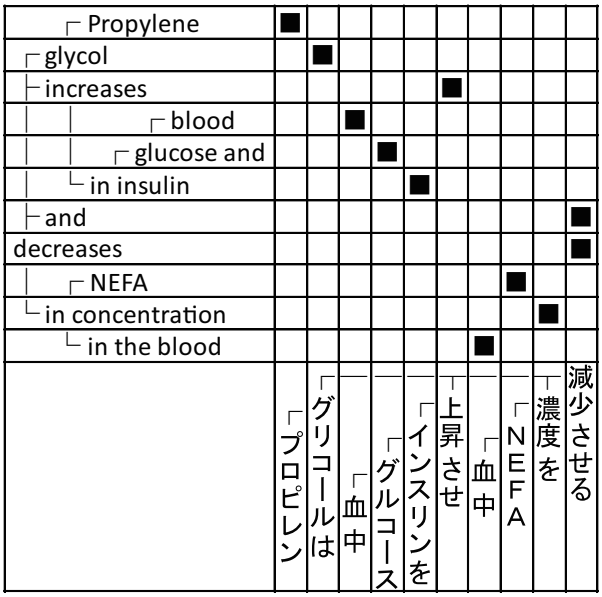

Figure 5: Result of phrase-base alignment.

is more prominence in case only small amount of parallel corpus can be used. What is more, most of these works are done on similar language pairs, such as English v.s. Chinese, French. We are now investigating the contribution of alignment quality improvement to translation quality in Japanese-English, language pair whose structure is very different.

According to the study about alignment quality (Fraser and Marcu, 2007), if the gold standard data only contains sure alignments, precision is much important than recall. Modified F-measure calculated by the equation 10 shows higher correlation to translation quality if α is set to be more than 0.5 (means emphasizing on precision).

$$\text{Modified F-measure} = \frac{1}{\frac{\alpha}{\text{Precision}} + \frac{(1-\alpha)}{\text{Recall}}} \qquad (10)$$

However, our results disagree with this argument. There seems no relation between precision and translation quality, indeed, F-measure has good correlation with BLEU score.

7 Related Work

Our proposed model is similar to the work of Cherry and Lin (2003). They use dependency tree structure for source side and construct a probabilistic model. The differences between their model and ours are following.

- Unit of alignment: Their model aligns words to words. Ours aligns syntactic phrases to phrases.
- Parser: They use a parser for source side only, and reproduce the target side dependency structure introducing a cohesion constraint. We use parsers for both source and target side, and do not require any constraints.
- Alignment constraint: They can only make one-to-one links. We can make even many-to-many links.

The third point is the most crucial for alignment, because there often occurs one-to-many or many-to-many correspondences. Their model was tested on English-French corpus, which are very similar languages, therefore even one-to-one model could achieve high accuracy.

8 Conclusion

In this paper, we have proposed a linguistically-motivated probabilistic phrase alignment model

based on dependency tree structure. Experimental results show that word-based statistical alignment model does not work well for linguistically different language pair, and it can be resolved by using syntactic information.

We have conducted the experiments only on Japanese-English corpus. To support firmly our allegation that syntactic information is important, it is necessary to do more investigation on other language pairs. We are mounting an experiment on Japanese-Chinese scientific paper corpus, whose characteristics are similar to the Japanese-English corpus we used in the experiment.

Proposed model handles content words and function words separately. This is harmful because function words in one side may appear as content words in the other side, and content words also may become function words. We need to construct more flexible model to solve this problem.

Moreover, one phrase often corresponds to more than one phrases in the other side. Currently we are handling such correspondence by symmetrization heuristics (growing). We are now trying to construct the framework which can model multiple of phrases. This enhanced model may be able to learn many-to-many alignment as one of the features of the model.

Most frequent alignment errors are derived from parsing errors. Since our method highly depends on the structural information, parsing errors easily make the alignment accuracy worse. Although the parsing accuracy is basically high for Japanese (around 90% for newspaper sentences), it sometimes outputs wrong dependency structure because there often appears technical, or unknown words in the scientific paper, and this is same to English. This problem is possible to be resolved by introducing parsing probabilities into our model as $\hat{a} = \mathrm{argmax}\, p(T_f|T_e, a) \cdot p(a|T_e) \cdot p(T_f) \cdot p(T_e)$ using parsing tools (KNP and nlparser) which can output n-best parsing with their parsing probabilities.

References

Necip Fazil Ayan and Bonnie J. Dorr. 2006. Going beyond AER: An extensive analysis of word alignments and their impact on MT. In *Proceedings of the 21st International Conference on Coling and 44th Annual Meeting of the ACL*, pages 9–16.

Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. 1993. The mathematics of statistical machine translation: Parameter estimation. *Association for Computational Linguistics*, 19(2):263–312.

Eugene Charniak and Mark Johnson. 2005. Coarse-to-fine n-best parsing and maxent discriminative reranking. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 173–180.

Colin Cherry and Dekang Lin. 2003. A probability model to improve word alignment. In *Proceedings of the 41st Annual Meeting of the Association of Computational Linguistics*, pages 88–95.

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

Brooke Cowan, Ivona Kučerová, and Michael Collins. 2006. A discriminative model for tree-to-tree translation. In *Proceedings of the 2006 Conference on EMNLP*, pages 232–241, Sydney, Australia, July. Association for Computational Linguistics.

Alexander Fraser and Daniel Marcu. 2007. Measuring word alignment quality for statistical machine translation. *Computational Linguistics*, 33(3):293–303.

Michel Galley, Jonathan Graehl, Kevin Knight, Daniel Marcu, Steve DeNeefe, Wei Wang, and Ignacio Thayer. 2006. Scalable inference and training of context-rich syntactic translation models. In *Proceedings of the 21st International Conference on Coling and 44th Annual Meeting of the ACL*, pages 961–968.

Daniel Gildea. 2003. Loosely tree-based alignment for machine translation. In *Proceedings of the 41st Annual Meeting on ACL*, pages 80–87.

Philipp Koehn, Franz J. Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *HLT-NAACL 2003: Main Proceedings*, pages 127–133.

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 *Annual Meeting of the Association for Computational Linguistics (ACL), demonstration session*.

Sadao Kurohashi and Makoto Nagao. 1994. A syntactic analysis method of long Japanese sentences based on the detection of conjunctive structures. *Computational Linguistics*, 20(4):507–534.

Sadao Kurohashi, Toshihisa Nakamura, Yuji Matsumoto, and Makoto Nagao. 1994. Improvements of Japanese morphological analyzer JUMAN. In *Proceedings of*

The International Workshop on Sharable Natural Language, pages 22–28.

Adam Lopez and Philip Resnik. 2006. Word-based alignment, phrase-based translation: What's the link? In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas*, pages 90–99.

Arul Menezes and Stephen D. Richardson. 2001. A best-first alignment algorithm for automatic extraction of transfer mappings from bilingual corpora. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics (ACL) Workshop on Data-Driven Machine Translation*, pages 39–46.

F. Och and H. Ney. 2003. A systematic comparison of various statistical alignment models.

Chris Quirk, Arul Menezes, and Colin Cherry. 2005. Dependency treelet translation: Syntactically informed phrasal SMT. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 271–279.

Masao Utiyama and Hitoshi Isahara. 2007. A japanese-english patent parallel corpus. In *MT summit XI*, pages 475–482.

Hideo Watanabe, Sadao Kurohashi, and Eiji Aramaki. 2000. Finding structural correspondences from bilingual parsed corpus for corpus-based translation. In *Proceedings of the 18th International Conference on Computational Linguistics*, pages 906–912.

Kenji Yamada and Kevin Knight. 2001. A syntax-based statistical translation model. In *Proceedings of 39th Annual Meeting of the ACL*, pages 523–530.