Hierarchical Statistical Machine Translation Using Sampling Based Alignment

Tongxu Liu & Yves Lepage
IPS, Waseda University, Kitakyushu, Japan
liutongxu0507@fuji.waseda.jp & yves.lepage@aoni.waseda.jp

Abstract

We propose an extension of the sampling-based alignment technique to implement the hierarchical phrases-based model. The proposed technique outputs a hierarchical phrase based rule table. It is learned from a bi-text without any syntactic information or any linguistic commitment. We expect the following advantages from our rule tables: 1. better reordering, 2. better translation of discontinuous phrases.

1 Introduction

Sub-sentential alignment plays an important role in the process of building a machine translation system. The quality of the sub-sentential alignments, which identify the relations between words or phrases in the source language and those in the target language, is crucial for the final result, [6]. There are mainly two models being proposed and implemented to solve the problem of alignment: standard phrase-based model [8] and the hierarchical phrase-based model [1]. There are also various techniques associated with these models (e.g. sampling-based alignment technique [5] for the phrase-based model).

Hierarchical phrase-based models first proposed by [1] are expected to contribute in reordering. For example, Japanese PPs almost always modify VP on the left, whereas English PPs usually modify VP on the right (e.g. I have relationship with him :: 私は彼と関係を持つ.) Using hierarchical phrases should allow to better capture reordering phenomena between different languages than using standard phrase-based models.

Following our use of the sampling based method for word alignment,[1] [5], we modify its implementation to output hierarchical phrase-based rule tables so that our translation system will have the two following advantages that rule tables are supposed to have: 1. better reordering, 2. better translation of discontinuous phrases.

The rest of the paper is divided into two main parts. The first part introduces the basic notions used. Section 2 describes the notion of hierarchical phrase and its advantages. Section 3 describes the sampling-based alignment method.

In the second part, the production of rule tables is described in three sections according to the way we produce our rule tables. Section 4 shows how to extract contexts of alignment pairs from a parallel corpus by using the sampling-based alignment method. Section 5 shows how to set up a filter by the number of place holders. Section 6 shows how to calculate the correspondence between place holders to produce the final rule tables.

Section 7 reports our experiments. A conclusion is given in Section 8.

2 Hierarchical phrase-based model

2.1 Hierarchical Phrases

The hierarchical phrase based model uses hierarchical phrases, i.e., phrases that contain subphrases as their basic translation units [1].

Table 1: A rule table.

Source language	Target language	Feature scores				
French	English	correspondence	$\phi(f\|e)$	$lex(f\|e)$	$\phi(e\|f)$	$lex(e\|f)$
Merci [x].	Thank you [x].	1-2	0.33	0.04	0.33	0.22
[x] et [x]	[x] and [x]	0-0 2-2	0.61	0.50	0.88	0.61

- In Table 1, ”[x]” is a place holder which represents a subphrase of the hierarchical phrase.
- Correspondences between place holders(third colum in Table 1) show which subphrase corresponds to which subphrase in the sourse language and the target language.

2.2 Advantages of hierarchical phrase-based models

There are two main advantages in using hierarchical phrase-based model.

1. Reordering: the rule tables output by hierarchical phrases based model would capture different orders between different languages.

[1] Anymalign is an implementation of the sampling-based alignment technique.

2. Discontinuous phrases: this model would also allow discontinuous sequences alignment (like “put on” in “put it on”).

Consider the following Japanese example and its English translation:

彼はこの問題との関係を持っている
He has a relationship with this matter

(1) [x] との [x] を持っている，has [x] with [x] 0-3,2-1

would capture Japanese PPs “この問題との関係” modify VP “持っている” on the left. whereas English PPs “with this matter” modify VP “has” on the right very well.
Another example:

彼らはそれをあきらめる。
They give it up.

(2) [x] をあきらめる， give [x] up 0-1

allows to capture the relation between the verb ’give’ and its associated separable particle ’up’ which forms the meaning ’to abandon, to stop’.

3 The sampling based alignment method

In sampling-based alignment, only those sequences of words sharing the exact same distribution (i.e., they appear exactly in the same sentences of the corpus) are considered for alignment [6], The key idea is to make more words share the same distribution by artificially reducing their frequency in multiple random subcorpora obtained by sampling. Indeed, the smaller a subcorpus, the less frequent its words, and the more likely they are to share the same distribution. Hence the higher the proportion of words aligned in this subcorpus [4]. The subcorpus selection process is guided by a probability distribution which ensures a proper coverage of the input parallel corpus by giving much more credit to small subcorpora, which happen to be the most productive [4]. From each subcorpus, sequences of words that share the same distribution are extracted to constitute alignments along with the number of times they were aligned. Eventually, the list of alignments is turned into a full-fledged phrase translation table by calculating various features for each alignment. In the following, we use two translation probabilities and two lexical weights as proposed by [2], as well as the commonly used phrase penalty, for a total of five features.

One important feature of the sampling-based alignment method is that it extracts phrase alignments and the context of these alignments in the corpus at the same time. For example:

If we have the following toy parallel corpus:

⟨ ありがとう先生。 Thank you professor. ⟩
⟨ ありがとう劉さん。 Thank you Liu. ⟩

By using the sampling-based alignment method, we will get the following alignments:

⟨ 先生 ⟺ professor ⟩
⟨ 劉さん ⟺ Liu ⟩

and the following context alignment:

⟨ ありがとう [x]。⟺Thank you [x].⟩

as well. This latter alignment can be regarded as a sub-sentential alignment. In this work, we use this kind of alignments as our first level rule tables.

4 Extracting contexts of alignment pairs

We use an implementation of the sampling-based method Anymalign to get the contexts of alignment pairs, as Table 2 shows.

Table 2: Numbers of first level rule tables output by Anymalign

# of place holders (source-target) source language side target language side	# of rules
0 : 0	983,432
1 : 1	1,278,721
2 : 2	966,368
0 : 1	1,466,185
0 : 2	935,145
1 : 0	1,191,180
1 : 2	1,399,600
2 : 1	971,373
3 : 4	1,023,718
5 : 5	317,413
.	.
.	.
.	.
total rules	24.1million

5 Filtering rule table entries

We filter discontinuous entries according to the principles suggested in [1]:

1. Initial phrases are limited to a length of 10, rules are limited to a length of 5 (place holders plus words) on the French side.

2. Rules can have at most two place holders, which simplifies the decoder implementation. Moreover, we prohibit place holders that are adjacent.

3. A rule must have at least one pair of aligned words, making translation decisions always based on some lexical evidence.

After filtering our tables, the number of rules that we get are shown in Table 3 and Table 4.

Table 3: Number of rules (source-target) after filtering

Number of place holders (source-target)	Number of rule tables
0-0	983,432
1-1	1,186,729
2-2	561,882
total	2.7 million

Table 4: Examples of the rules after filtering

source(French)	target(English)	correspondence
conseil	council	none
merci monsieur [x] .	thank you mr [x] .	?-?
le [x] de [x] .	the [x] of the [x] .	?-?, ?-?

6 Computing the correspondences between place holders

For rules which include one place holder on both sides, we just find the position of each place holder to write the correspondence between place holders as was shown in Table 1. For rule tables which include two place holders on both sides, we proceed as follows:

1. Firstly, we find two mid-terms, i.e., terms between two place holders in the rule tables on both sides.
2. We then find every line from the parallel corpus which include the two mid-terms.
3. We divide every line into four parts by the two mid-terms. We call them A, B, $\bar{A}$ and $\bar{B}$. They represent two parts for each of the sentences (see Section 6.3 for an example).
4. In order to be able to judge whether the correspondence is *monotonous* (as Figure 1 shows) or *crossing* (as Figure 2 shows), we compute the lexical weights of the four possible correspondences relying on the values of the place holders.

6.1 Part of everyline

Not only in order to simplify the calculation of lexical weight but also in order to get a more accurate result of correspondence between place holders we do not use the whole sentence in everyline which include the ”mid-term”, we only use a part of them, precisely the 3 words before and after the ”mid-term”.

6.2 Lexical weights

The following equation gives the definition of lexical weights as stated in [2]. Given a phrase pair the target language t, the source language s and a word alignment a between the target word possitions $i = 1, \cdots, I$ and the source word positions $j = 0, 1, \cdots, J$, the lexical weight lex can be computed according to the following formula:

$$lex(t|s) = \prod_{i=1}^{n} \frac{1}{|\{j|(i,j) \in a\}|} \sum_{\forall(i,j)\in a} w(t_i|s_j) \qquad (1)$$

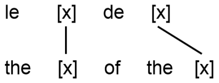

Figure 1: monotonous correspondence

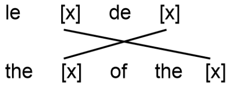

Figure 2: crossing correspondence

As many phrases have very low counts, simple phrase conditional probabilities are sparse and often do not provide reliable information about the correctness of the phrase pair, For this reason, we calculate lexical weights instead of conditional probabilities to tell whether two sequence of words are corresponding translation pairs or not [7].

6.3 Example

Suppose we want to calculate the correspondence of place holders in the following rule:

< la [x] de [x] , the [x] of the [x] >

French:	la couleur	de	la voiture
	A		B
English:	the color	of the	car
	$\bar{A}$		$\bar{B}$

lexical weight between monotonous sequences:

$$lex(A,\bar{A}) = \sqrt{max[lex(la|the), lex(couleur|the)] \times max[lex(la|color), lex(couleur|color)]}$$

$$lex(B,\bar{B}) = \sqrt{max[lex(la|car), lex(voiture|car)] \times max[lex(la|car), lex(voiture|car)]}$$

lexical weight between cross sequences:

$$lex(A,\bar{B}) = \sqrt{max[lex(la|the), lex(voiture|the)] \times max[lex(la|color), lex(voiture|color)]}$$

$$lex(B,\bar{A}) = \sqrt{max[lex(la|car), lex(couleur|car)] \times max[lex(la|car), lex(couleur|car)]}$$

7 Experiments

We use the Europarl corpus for our experiments. We use 347,614 lines for training, 500 lines for tuning and 38,123 lines for test. We use the same number of corresponding sentences in training, tuning and test for all the 11 languages of Europarl version 3. In this way, the experiments we performed over the 110 possible language pairs are really comparable.

7.1 BLEU score

- source sentence: tout le monde en est conscient , il est temps de mettre fin à ce jeu de cache-cache avec le gouvernement de khartoum .
- translation by our system: all is aware of this , it is time to an end to this game of hide-and-seek with the government in khartoum it .

Table 5: Hierarchical tables BLEU SCORE of Moses and Anymalign

	fr-en
Moses hierarchical	28.76
Anymalign hierarchical	25.86

Table 6: Hierarchical tables BLEU SCORE of other languages (%)

	da	de	el	en	es	fi	fr	it	nl	pt	sv
da	-	11.23	16.44	20.31	15.28	11.55	12.31	10.70	13.92	14.43	23.72
de	9.33	-	13.82	10.51	10.38	8.79	13.12	7.64	10.86	11.50	10.11
el	18.20	17.90	-	20.47	19.98	14.71	18.62	16.66	15.01	19.97	14.15
en	21.72	13.33	23.63	-	23.30	19.36	25.39	24.27	23.25	27.56	23.34
es	22.02	19.32	24.41	23.18	-	17.27	26.47	26.96	21.08	27.38	24.96
fi	9.21	7.63	9.70	9.26	5.49	-	4.76	7.62	6.97	5.46	8.90
fr	18.82	12.73	17.44	25.86	22.52	9.80	-	22.64	14.91	19.62	17.32
it	15.47	19.63	18.61	23.72	24.69	12.20	20.18	-	14.68	21.22	14.37
nl	17.96	13.61	13.23	20.02	19.03	12.92	14.97	11.45	-	20.66	18.85
pt	19.40	12.63	20.04	19.27	25.81	13.31	20.46	24.64	19.33	-	23.16
sv	23.21	13.67	17.10	20.12	15.82	13.25	15.36	12.36	15.12	19.73	-

Unfortunately, as Table 5 shows, we can not beat MOSES [3], but the translation example 7.1 shows that our system can translate the sentence pattern "it is time to .." very well. Table 6 shows other result of difference language pairs by our hierarchical system.

7.2 Ratio of crossing rules

As we know, in rule tables according to the correspondence of the place holders, rules can be of two kinds of correspondence: monotonous or cross. Table 7 shows the ratio of crossing rules for each language pair used.

Table 7: Rate of cross rules (%)

	da	de	el	en	es	fi	fr	it	nl	pt	sv
da	-	6.3	8.7	6.3	12.3	**3.0**	15.4	10.5	5.8	16.4	3.7
de	6.7	-	7.3	9.8	8.0	5.6	10.3	8.8	9.9	6.5	8.1
el	8.6	7.7	-	6.8	8.2	5.9	7.2	12.6	4.2	13.2	10.4
en	6.7	10.3	6.8	-	7.9	7.5	5.3	10.1	13.2	12.8	5.2
es	6.5	9.2	7.1	8.1	-	7.2	7.5	9.6	10.3	14.2	9.1
fi	3.6	6.7	6.9	7.2	8.5	-	6.3	9.3	10.2	12.8	6.8
fr	14.6	9.7	7.1	8.1	6.9	6.5	-	9.6	10.9	10.2	6.1
it	10.8	8.4	11.3	9.8	9.2	9.1	5.7	-	3.2	14.5	9.1
nl	6.0	5.7	4.8	9.8	7.9	5.3	7.5	10.1	-	17.8	5.2
pt	19.2	18.6	14.0	13.2	15.8	13.3	17.4	18.6	**19.9**	-	17.2
sv	6.7	10.3	6.8	4.2	7.9	5.3	7.5	10.1	13.2	16.8	-

The ratio among language pairs varies from 3.0% (fi-da) to 19.9% (nl-pt). The average cross ratio of crossing rules between Portuguese and other languages are a little bit higher.

8 Conclusion

We have proposed a method to extract hierarchical phrase rule tables from a parallel corpora using the sampling-based alignment method. The proposed method is divided into three parts: 1. getting translation tables with place holders, 2. filtering and 3. calculating correspondences between place holders.

We have performed experiments on several different language pairs and assessed the translation quallity using the BLEU [9] metric.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number C 23500187.

謝辞

本研究はJSPS科研費 基盤C 23500187の助成を受けたものです。

References

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

[2] Philipp Koehn, Franz Josef Och, and Daniel Marcu. Statistical phrase-based translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, pages 48–54. Association for Computational Linguistics, 2003.

[3] Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, et al. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, pages 177–180. Association for Computational Linguistics, 2007.

[4] Adrien Lardilleux, Jonathan Chevelu, Yves Lepage, Ghislain Putois, Julien Gosme, et al. Lexicons or phrase tables? an investigation in sampling-based multilingual alignment. In *Proceedings of the 3rd Workshop on Example-Based Machine Translation*, pages 45–52, 2009.

[5] Adrien Lardilleux, Yves Lepage, et al. Sampling-based multilingual alignment. In *Proceedings of Recent Advances in Natural Language Processing*, pages 214–218, 2009.

[6] Juan Luo. Enhancing sampling-based alignment for statistical machine translation, (Waseda university Master thesis, 2012).

[7] Graham Neubig, Taro Watanabe and Shinsuke Mori. Inducing a discriminative parser to optimize machine translation reordering. In *Proceedings of the 2012*

Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pages 843–853. Association for Computational Linguistics, 2012.

[8] Franz Josef Och and Hermann Ney. The alignment template approach to statistical machine translation. *Computational linguistics*, 30(4):417–449, 2004.

[9] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics (ACL, 2002)*, pages 311–318. Association for Computational Linguistics, 2002.