IEEJ TRANSACTIONS ON ELECTRICAL AND ELECTRONIC ENGINEERING
IEEJ Trans 2016; **11**: 325–330
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI:10.1002/tee.22221

Paper

A Method of Generating Translations of Unseen n-grams by Using Proportional Analogy

Juan Luo[a], Non-member
Yves Lepage, Non-member

In recent years, statistical machine translation has gained much attention. The phrase-based statistical machine translation model has made significant advancement in translation quality over the word-based model. In this paper, we attempt to apply the technique of proportional analogy to statistical machine translation systems. We propose a novel approach to apply proportional analogy to generate translations of unseen n-grams from the phrase table for phrase-based statistical machine translation. Experiments are conducted with two datasets of different sizes. We also investigate two methods to integrate n-grams translations produced by proportional analogy into the state-of-the-art statistical machine translation system, Moses.[1] The experimental results show that unseen n-grams translations generated using the technique of proportional analogy are rewarding for statistical machine translation systems with small datasets. © 2016 Institute of Electrical Engineers of Japan. Published by John Wiley & Sons, Inc.

Keywords: statistical machine translation, proportional analogy, unseen n-grams, phrase table

Received 29 July 2014; Revised 7 June 2015

1. Introduction

Machine translation has achieved significant advancement over the years. Recently, statistical machine translation has gained much attention in both academic studies and commercial usage because of the advances in the field. Phrase-based statistical machine translation systems rely on parallel corpora for learning translation knowledge and translation rules, which are stored in the so-called phrase tables. The quality of the phrase table is crucial to the translation quality of machine translation systems. Thus, the phrase table is the fundamental and vital component in the translation process. A phrase table consists of sequences of words in the source language and sequences of words in the target language, as well as feature scores showing how likely these two sequences are translations of each other. It is usually constructed in two steps: first, the generation of source-to-target and target-to-source word alignments, and, second, extraction of bilingual phrase pairs from these alignments through heuristic combination of both directions.

In recent years, some schemes have been proposed to deal with phrase tables in statistical machine translation systems. Research trying to acquire additional data to increase translation coverage has focused on introducing paraphrases [1,7,18], n-grams [9,17], and multiword units [23].

In Ref. [1], paraphrases of unseen source phrases are incorporated into phrase tables to improve the coverage and translation quality. However, their method is particularly pertinent to small corpus and out-of-vocabulary words. Augmenting phrase tables via paraphrasing is also investigated in Refs [7,18]. A method of enlarging the n-grams in phrase tables has been reported in Ref. [17], in which 'word packing' is used to obtain 1-to-n alignments based on co-occurrence frequencies. They evaluated the performance on Chinese-to-English machine translation task and reported significant improvements. In Ref. [9], collocation segmentation is performed on bilingual corpus to extract n-to-m alignments, which are used to augment phrase tables. However, experimental results showed no difference in the evaluation metric scores. Ref. [23] proposed a strategy to extract domain bilingual multiword expressions and investigated methods to integrate these multiword units to phrase tables.

Proportional analogy has been researched and applied to address problems in various domains, for instance, machine transliteration [4], machine translation [15], handling unknown words [5,13], handwritten character recognition [19], and so on.

In Ref. [4], methods of applying analogical learning over strings to the transliteration tasks were proposed. The authors showed that a combination of proportional analogy and statistical machine translation engine could lead to improvements over individual transliteration systems. Ref. [15] proposed an example-based machine translation system that is built upon proportional analogy. Their machine translation system works well on short sentences. Proportional analogy is also applied at the character level to translate unknown words, which was reported in Ref. [5] on Japanese-to-English tasks and Ref. [13] on language pairs with close morphological structure. In Ref. [19], an analogy-based sequence generation is applied, which enables a handwritten character recognition system to be rapidly adapted to a new writer.

Given a test sentence, those words that cannot be found in phrase table thus result in the unknown sequences for a machine translation system. In this paper, we attempt to address the problem of unseen n-grams. Here, we propose to use proportional analogy to generate translations of unseen n-grams from phrase tables for statistical machine translation systems. We show that this method is useful for systems with small data. To the best of our knowledge, this paper is the first attempt at associating proportional analogy with phrase tables to generate strings that go beyond words, i.e., n-grams.

The remainder of this paper is organized as follows: In Section 2, we briefly introduce the notion of proportional analogy. In Section 3, we describe our proposed method that is based on proportional analogy. In Section 4, experiments are reported and evaluation results are analyzed. Finally, we conclude in Section 5.

[a] Correspondence to: Juan Luo. E-mail: juan.luo@suou.waseda.jp

Graduate School of Information, Production and Systems, Waseda University, 2-7 Hibikino, Wakamatsu-ku, Kitakyushu, Fukuoka 808-0135, Japan

[1] Moses [12]: http://www.statmt.org/moses/

2. Proportional Analogy

Proportional analogy is defined as a general relationship between four objects, i.e. four strings in this work. It is noted as $A : B :: C : D$, which is to be read as 'A is to B as C is to D'. Analogy can be seen at the semantic level or at the formal level. Here we work on the formal level only to the possible detriment of meaning.

Ref. [14] proposed a formalization of analogies between strings. This formalization reduces to the counting of the number of symbol occurrences and the computation of edit distances. The four strings, A, B, C, and D, form an analogy only if

$$A : B :: C : D \Rightarrow \begin{cases} |A|_a - |B|_a = |C|_a - |D|_a, \forall a \\ \\ \delta(A,B) = \delta(C,D) \end{cases} \quad (1)$$

where $|A|_a$ stands for the number of occurrences of character a in string A, and δ is the edit distance that involves only insertion and deletion with equal weights. $\delta(A,B)$ stands for the edit distance between strings A and B. As B and C may be exchanged in an analogy, the constraint on edit distance has to be verified in addition for $A : C :: B : D$, i.e. $\delta(A,C) = \delta(B,D)$. There is no need to verify the first constraint for $A : C :: B : D$, because trivially $|A|_a - |B|_a = |C|_a - |D|_a \Leftrightarrow |A|_a - |C|_a = |B|_a - |D|_a$.

As analyzed in Ref. [16], proportional analogies can be written between words (2), chunks (3), or sentences (4) (examples from Ref. [16]):

$$\text{abundant} : \text{abundance} :: \text{present} : \text{presence} \quad (2)$$

$$\begin{matrix} \text{my room} \\ \text{key} \end{matrix} : \begin{matrix} \text{the room} \\ \text{key} \end{matrix} :: \begin{matrix} \text{my first} \\ \text{visit} \end{matrix} : \begin{matrix} \text{the first} \\ \text{visit} \end{matrix} \quad (3)$$

$$\begin{matrix} \text{Do you} \\ \text{like} \\ \text{music?} \end{matrix} : \begin{matrix} \text{Do you go} \\ \text{to conserts} \\ \text{often?} \end{matrix} :: \begin{matrix} \text{Do you like} \\ \text{classical} \\ \text{music?} \end{matrix} : \begin{matrix} \text{Do you go} \\ \text{to classical} \\ \text{concerts often?} \end{matrix} \quad (4)$$

In this work, we focus on proportional analogies between sub-sentence strings, i.e. n-grams in phrase tables.

3. Generation of Unseen n-gram Translation Using Proportional Analogy

In this section, we present our proposed method of generating entries by applying proportional analogical learning of unseen source n-grams in the test sentences. Instead of adding generated *analogy* n-gram as new entries to the baseline phrase table, we collect these entries to form an additional *analogy* phrase table.

The method comprises three stages:

1. Producing unseen n-grams:
   In this stage, given a test sentence, we first segment it into n-grams. These n-grams are then searched in the baseline phrase table. Finally, a list of unseen n-grams (i.e. that are not found in baseline phrase table) are extracted and produced.
2. Searching analogical candidates in baseline phrase table:
   In this stage, given an unseen source n-gram, three candidate n-grams that should form analogical relationship with this unseen n-gram are searched in the source part of the baseline phrase table. After searching three source candidates, we thus obtain their corresponding n-grams in the target language.

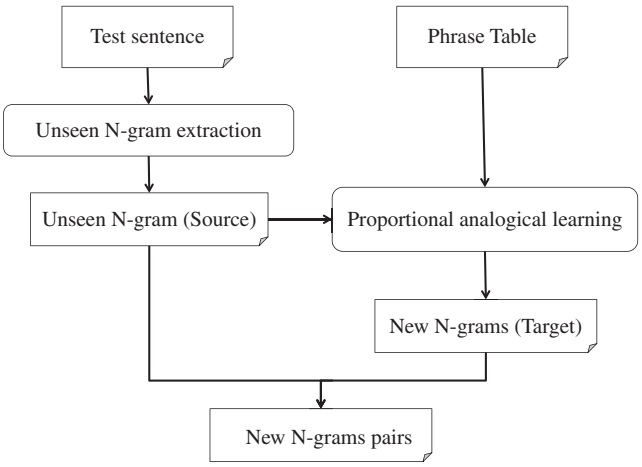

Fig. 1. Process of analogical learning of unseen n-grams from the phrase table

3. Producing analogies of n-grams to form an *analogy* phrase table:
   In this stage, three target candidate n-grams are used to generate a new n-gram in target language by proportional analogical learning. Finally, the newly generated n-gram pair is added as an entry to form an *analogy* phrase table.

3.1. An example Let us illustrate the method more clearly with an example (see Fig. 1).

Assume a test sentence:

The international conference will be held next week in Iceland.

We segment it into n-grams, for example, 3-gram:

the international conference
international conference will
conference will be
.....

These n-grams are searched in the baseline phrase table. N-grams that are not found in the phrase table are extracted. Here we obtain an unseen English source n-gram D_s:

D_s = *the international conference*

In order to form an analogical relationship

$A_s : B_s :: C_s : D_s$

Three candidates are searched in the source part of the baseline phrase table: they are

A_s = *national*
B_s = *the national conference*
C_s = *international*

Their corresponding French target n-grams in phrase table thus are

A_t = *nationale*
B_t = *la conférence nationale*
C_t = *internationale*

A new n-gram D_t can be generated by proportional analogical learning:

$A_t : B_t :: C_t : D_t$
D_t = *la conférence internationale*

Finally, we obtain a new n-gram pair (D_s, D_t), which can be added as an entry to *analogy* phrase table:

D_s = *the international conference*
D_t = *la conférence internationale*

Table I. Evaluation results

		German–English				Polish–English			
		BLEU	NIST	WER	TER	BLEU	NIST	WER	TER
Train ≈ 350 k	Baseline	25.30	6.6193	55.27	59.65	34.26	7.5076	44.26	47.70
	Multiple PT	24.68	6.4787	56.55	61.09	33.91	7.4041	45.01	48.56
	Backoff model (1-g)	25.16	6.5702	55.91	60.25	34.03	7.4790	44.29	47.73
	Backoff model (2-g)	24.90	6.5557	55.71	60.20	34.06	7.4563	44.46	47.94
	Backoff model (3-g)	24.96	6.5860	55.47	60.01	34.11	7.4695	44.60	48.07
	Backoff model (4-g)	24.90	6.5527	55.93	60.56	34.01	7.4519	44.61	48.16
	Backoff model (5-g)	24.95	6.5383	55.99	60.58	33.67	7.4202	44.92	48.34
	Backoff model (6-g)	24.48	6.5040	56.14	60.77	33.88	7.4162	44.78	48.43
	Backoff model (7-g)	24.54	6.4809	56.54	60.97	33.81	7.4264	44.80	48.42
Train=10 k	Baseline	20.69	5.8867	58.76	63.73	19.35	5.6148	55.81	60.03
	Multiple PT	20.23	5.8173	59.61	64.87	**19.85**	**5.7568**	**55.64**	**59.91**
	Backoff model (1-g)	**20.83**	**5.9339**	59.10	63.95	**19.94**	**5.8243**	**54.84**	**59.15**
	Backoff model (2-g)	**20.92**	**5.9676**	**58.76**	**63.73**	**19.95**	**5.8305**	**54.91**	**59.11**
	Backoff model (3-g)	**20.84**	**5.9205**	**58.64**	63.80	**19.98**	**5.8115**	**54.92**	**59.28**
	Backoff model (4-g)	20.38	5.8712	58.89	64.11	**19.76**	**5.7564**	**55.64**	**60.02**
	Backoff model (5-g)	20.26	5.8104	59.51	64.79	**19.71**	**5.7502**	**55.61**	**59.78**
	Backoff model (6-g)	20.27	5.8211	59.68	64.67	**19.68**	**5.7556**	**55.53**	**59.84**
	Backoff model (7-g)	20.07	5.8149	59.65	64.93	**19.74**	**5.7541**	**55.67**	**59.87**

The numbers in boldface indicate that they are higher than the baseline.

3.2. Calculation of feature scores In the default phrase table of a standard statistical machine translation system, there are five feature scores: two translation probabilities and two lexical weights as proposed by Ref. [11], as well as the commonly used phrase penalty. Here, we calculate the feature scores of an *analogy* n-gram pair by two steps. In the first step, given three candidate n-gram pairs (A_s, A_t), (B_s, B_t), and (C_s, C_t), a feature score f of the *analogy* n-gram pair (D_s, D_t) is calculated by geometric mean of the feature scores of the candidate n-gram pairs:

$$f(D_s, D_t) = \sqrt[3]{f(A_s, A_t) \times f(B_s, B_t) \times f(C_s, C_t)} \quad (5)$$

For such an *analogy* n-gram pair, sets of analogies can be found in the phrase table. Thus, in the second step, it is calculated as

$$f(D_s, D_t) = \frac{1}{n} \sum_{i=1}^{n} \sqrt[3]{f(A_s, A_t) \times f(B_s, B_t) \times f(C_s, C_t)} \quad (6)$$

Here, in this investigation, we compute lexical weights as the above equations. We will compare the differences on the translation quality by using this computation and the one that was originally proposed by [11] later.

4. Experiments

In this section, we evaluate the performance of our proposed method empirically. First, we describe the experimental setup and present the evaluation results. Then, we investigate and analyze n-gram pairs in detail.

4.1. Experimental setup and datasets Standard phrase-based statistical machine translation systems were built by using the conventional pipeline: the Moses toolkit [12], Batch MIRA [3] to tune the parameters, the SRI Language Modeling (SRILM) toolkit [25] to build a 5-g target language model with Kneser–Ney smoothing, and GIZA++ [21] to generate word alignment. The maximum length of phrase pairs in phrase tables is set to 7 (the default phrase length in Moses).

The experiments were carried out using the Europarl parallel corpus [10]. We examined two language pairs: German-to-English and Polish-to-English. For each language pair, we tested two different sizes of training data. For the first setting, we used a training set of 347 614 and 350 000 sentence pairs, respectively. For the second setting, we used a small size dataset of 10 000 sentence pairs for both language pairs. We refer to this two settings as *Train≈350 k* and *Train=10 k* in the following sections. The development set was made up of 500 sentence pairs, and test set contained 1000 sentence pairs.

As for evaluation, four standard automatic evaluation metrics were used to assess the output of machine translation systems: BLEU [22], NIST [6], WER [20], and TER [24].

4.2. Experimental results Since Moses supports multiple phrase tables, here we investigate two methods to utilize *analogy* n-gram pairs: (i) Multiple PT, in which *either* phrase table is used for scoring; (ii) backoff model, in which the second phrase table is used as a backoff for unknown sequences. We used *analogy* phrase table as backoff table and experimented on limiting the length of n-grams that were used from the backoff table.

The results of experiments are shown in Table I. Intuitively, *analogy* n-gram pairs are useful to improve the performance of statistical machine translation. However, from the results it can be seen that, given a training parallel corpus of 350 k sentences, the evaluation scores decrease slightly by comparing with the baseline. In the case of a small training data size (10 000 sentences), we can observe improvements over the baseline in both language pairs. The multiple PT method achieves improvements in four evaluation metrics for Polish–English. However, this is not in consistent with the results obtained for language pair German–English, where a decrease in the translation quality is observed. The backoff model improves translation quality in both languages. By limiting the phrase length in the backoff table, i.e. *analogy* phrase table, the greatest increase in evaluation scores is obtained for 2-g or 3-g.

From the evaluation results in this table, we can conclude that n-gram pairs generated by proportional analogy are useful for translation systems with less training data.

4.3. Discussion In order to examine n-gram pairs in detail, we analyzed the number of unique unseen n-grams in test

Table II. Number of unique n-grams in test set. Unseen: the number of unique unseen n-grams

	Total	Unseen	Analogy (%)
Train≈350 k			
1-g	5 888	1 516	879 (58%)
2-g	18 602	10 860	8 009 (74%)
3-g	24 001	20 014	14 426 (72%)
4-g	24 614	23 190	16 152 (70%)
5-g	23 926	23 359	15 726 (67%)
6-g	23 019	22 772	14 715 (65%)
7-g	22 064	21 927	13 612 (62%)
Train = 10 k			
1-g	5,888	2 973	1 073 (36%)
2-g	18 602	14 946	9 073 (61%)
3-g	24 001	22 560	13 709 (61%)
4-g	24 614	24 142	13 704 (57%)
5-g	23 926	23 742	12 341 (52%)
6-g	23 019	22 943	10 846 (47%)
7-g	22 064	22 030	9 444 (43%)

Analogy: the number of unique unseen n-grams translations that are generated by proportional analogy (German–English)

Table III. Number of unique n-grams in test set (Polish–English)

	Total	Unseen	Analogy (%)
Train ≈ 350 k			
1-g	4 856	402	307 (76%)
2-g	14 402	7 444	5 327 (72%)
3-g	18 034	14 780	9 633 (65%)
4-g	18 988	17 859	10 729 (60%)
5-g	18 864	18 528	10 066 (54%)
6-g	18 385	18 291	8 888 (49%)
7-g	17 782	17 761	7 599 (43%)
Train = 10 k			
1-g	4 856	2 544	960 (38%)
2-g	14 402	12 708	5 990 (47%)
3-g	18 034	17 487	7 219 (41%)
4-g	18 988	18 865	6 413 (34%)
5-g	18 864	18 842	5 024 (27%)
6-g	18 385	18 379	3 719 (20%)
7-g	17 782	17 782	2 753 (15%)

sentences and those n-grams translations that can be generated by proportional analogy. We also investigated the distribution of phrase lengths used during decoding.

An analysis of the number of unique unseen n-grams in test set is shown in Tables II and III. From the tables, it can be seen that the percentage of the number of unique unseen n-grams translations that are generated by proportional analogy varies. The largest number of n-grams that can be produced by analogy are 2-g and 3-g in all cases for both language pairs.

Figure 2 shows the phrase lengths that are actually used during the translation process in all baseline systems. From the graph it can be seen that 50–80% of phrases are one-to-one translations. Further inspection shows that more than 90% of the phrases used in decoding are of length up to 3. In order to know how the phrase length differs from the baseline by using *analogy* n-gram pairs, we analyzed the distribution of phrases for all methods. The graphs are shown in Figs 3 and 4. For German–English, there is a slight difference in the distribution of phrase lengths between the baseline and the method of using two phrase tables. For Polish–English, the distributions are approximately the same. In general, the majority of phrases used in decoding are up to 3-g.

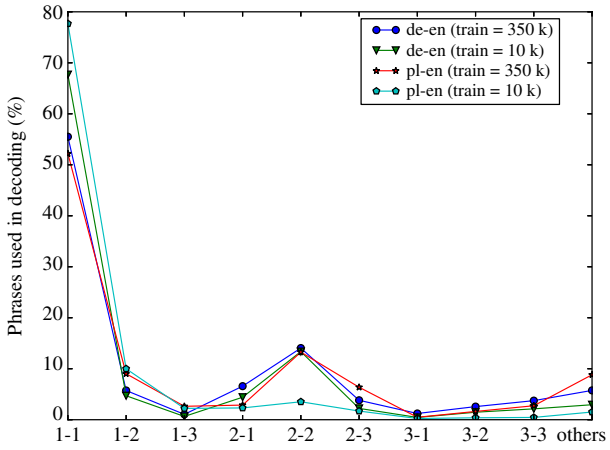

Fig. 2. Distribution of phrases used during decoding (baseline systems)

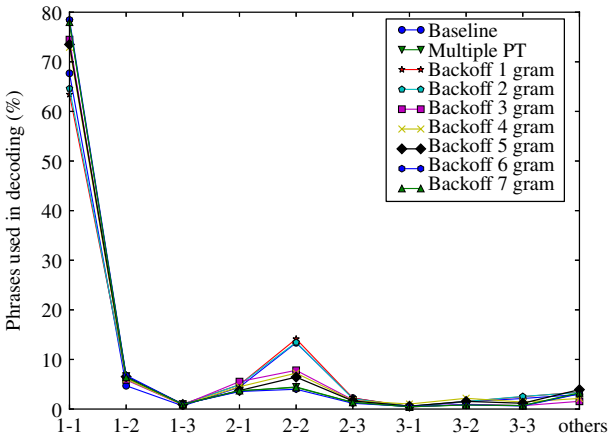

Fig. 3. Distribution of phrases used during decoding (German–English; Train=10 k)

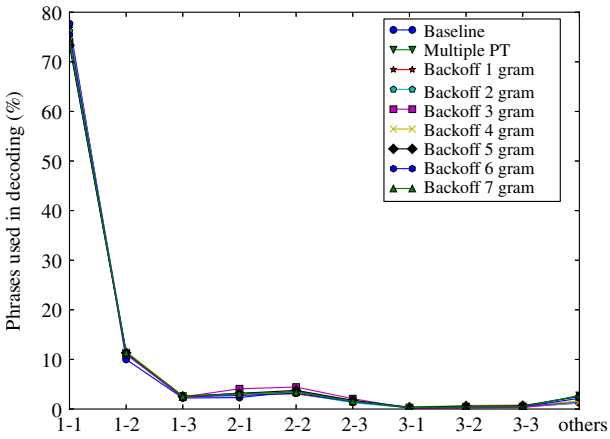

Fig. 4. Distribution of phrases used during decoding (Polish–English; Train=10 k)

In order to measure the effect of using different training data sizes, we conducted experiments by increasing the training data size from 50 000 lines to 300 000 lines. The training data are extracted from the 350 k datasets. The development and test datasets are the same as those used in Table I. Here, we conducted experiments with the backoff model and the lengths of the *analogy* n-gram pairs were limited to 3. The results of the experiments are shown in Table IV. From the table, we can see that

Table IV. Evaluation results

		German-English				Polish-English			
		BLEU	NIST	WER	TER	BLEU	NIST	WER	TER
Train = 50 k	Baseline	24.16	6.4166	55.98	60.82	28.61	6.8731	48.55	52.08
	Backoff model (1-g)	24.10	**6.4887**	**55.84**	**60.40**	**28.63**	**6.9185**	**48.35**	**51.83**
	Backoff model (2-g)	24.02	**6.4792**	**55.96**	**60.37**	**28.67**	**6.9012**	**48.48**	**52.05**
	Backoff model (3-g)	23.95	**6.4404**	56.18	**60.62**	28.46	6.8372	48.85	52.43
Train = 100 k	Baseline	24.55	6.4516	56.07	60.71	30.71	7.1021	47.00	50.52
	Backoff model (1-g)	**24.58**	**6.4542**	56.25	60.87	**30.74**	7.0843	47.08	50.73
	Backoff model (2-g)	**24.64**	**6.5005**	**55.55**	**60.21**	30.44	7.0802	47.01	50.59
	Backoff model (3-g)	**24.69**	**6.4971**	**55.87**	**60.41**	30.56	7.0586	47.31	50.88
Train = 150 k	Baseline	24.84	6.5104	55.62	60.01	31.38	7.2739	45.72	49.19
	Backoff model (1-g)	24.63	**6.5356**	55.66	60.05	**31.41**	**7.2916**	**45.18**	**48.80**
	Backoff model (2-g)	24.75	**6.5377**	55.68	60.06	31.29	7.2646	**45.57**	**49.13**
	Backoff model (3-g)	24.67	**6.5221**	55.93	60.37	**31.40**	7.2597	**45.46**	**49.04**
Train = 200 k	Baseline	24.96	6.4895	55.93	60.22	32.83	7.3624	45.05	48.63
	Backoff model (1-g)	24.68	**6.5032**	56.03	60.45	**33.00**	**7.3851**	**44.94**	**48.47**
	Backoff model (2-g)	24.43	**6.4927**	56.07	60.40	32.69	**7.3730**	45.08	**48.55**
	Backoff model (3-g)	24.71	6.4526	56.54	60.88	**32.83**	7.3559	45.09	48.73
Train = 250 k	Baseline	24.89	6.4325	56.93	61.18	33.43	7.4779	44.22	47.88
	Backoff model (1-g)	24.74	**6.5292**	**56.05**	**60.41**	33.24	7.4684	44.37	47.93
	Backoff model (2-g)	24.58	**6.5191**	**56.05**	**60.43**	33.27	7.4482	44.39	48.09
	Backoff model (3-g)	24.67	**6.5107**	**55.99**	**60.52**	**33.68**	7.4595	44.50	48.14
Train = 300 k	Baseline	25.06	6.4940	55.98	60.61	33.87	7.5469	43.85	47.17
	Backoff model (1-g)	25.04	6.4809	56.27	60.70	33.72	7.5094	44.35	47.66
	Backoff model (2-g)	24.81	6.4596	56.01	60.71	**34.08**	7.5011	44.28	47.66
	Backoff model (3-g)	24.89	**6.5114**	56.01	60.61	33.88	7.5014	44.21	47.55

The numbers in boldface indicate that they are higher than the baseline.

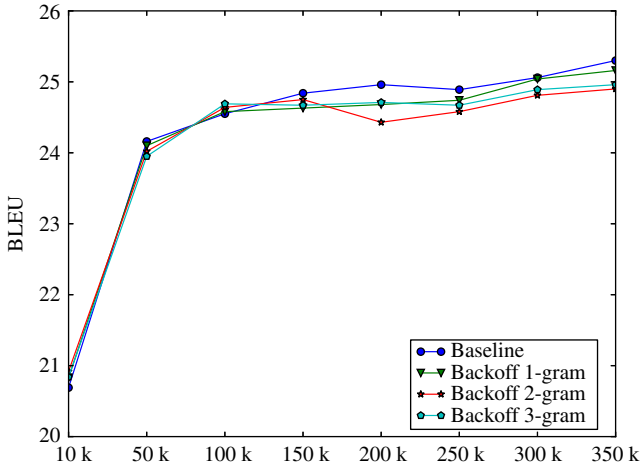

Fig. 5. BLEU scores for training data size range from 10 to 350 k (German–English)

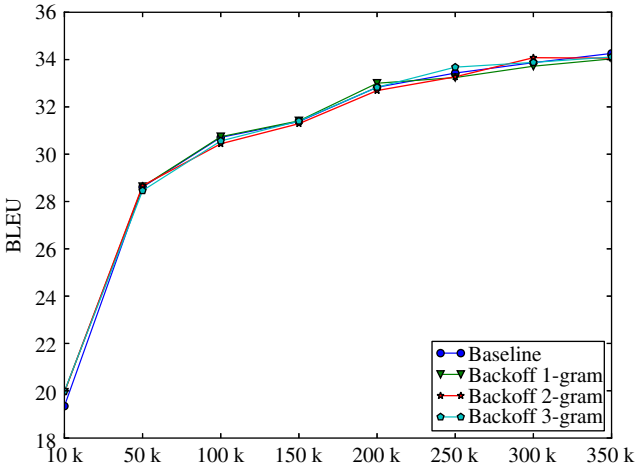

Fig. 6. BLEU scores for training data size range from 10 to 350 k (Polish–English)

the experimental results of the two language pairs are different from one another. For the language pair German–English, when we look at four evaluation metrics, we can see that improvements are obtained with a training data size of 100 k lines. As for the language pair Polish–English, when using 50 k, 150 k, and 200 k lines for training, we obtained improvements in four evaluation metrics. We also compared the BLEU scores between baseline and backoff model (1, 2, and 3 grams) and the training data size varies from 10 to 350 k lines. This is shown in Figs 5 and 6 for the language pair German–English and Polish–English, respectively. Undoubtedly, as the training data size increases, better evaluation results can be obtained. As for the relations between *analogy* n-gram pairs and training data size, it indicates that, for German–English, *analogy* n-gram pairs are useful for training data of less than 100 k lines, whereas for Polish–English they are useful for training data of less than 200 k lines.

As the experimental results show, *analogy* n-gram pairs are useful for translation systems with a small amount of training data. Since some language pairs do not have a large amount of data for training, populating phrase tables by proportional analogy can be rewarding for machine translation systems that are built for less resourced language pairs. However, it may require more time and more processing steps, as *analogy* n-gram pairs need to be generated.

5. Conclusion

In this paper, we investigated and proposed a novel method of applying the technique of proportional analogy on generating unseen n-grams translations from phrase tables for statistical machine translation systems. We conducted experiments on two different sizes of datasets. The evaluation results revealed that

populating phrase tables by proportional analogy is rewarding for machine translation systems with a small amount of data.

Further inspections will be conducted in the future. We will investigate the issue of fabricating new phrase pairs [2]. We will also identify which phrases may benefit from additional data and special processing [8]. Since the lexical weights of the *analogy* n-gram pairs are not computed in a conventional way in this paper, we would like to compare the effects on the translation quality by using different calculations in the future.

References

(1) Callison-Burch C, Koehn P, Osborne M. Improved statistical machine translation using paraphrases. *Proceedings of the NAACL-HLT*, New York, 2006; 17–24.
(2) Chen B, Kuhn R, Foster G. Semantic smoothing and fabrication of phrase pairs for SMT. Proceedings of the 7th IWSLT, San Francisco, USA, 2011; 144–150.
(3) Cherry C, Foster G. Batch tuning strategies for statistical machine translation. Proceedings of the NAACL-HLT, Montréal, Canada, 2012; 427–436.
(4) Dandapat S, Morrissey S, Naskar SK, Somers H. Mitigating problems in analogy-based EBMT with SMT and vice versa: a case study with named entity transliteration. Proceedings of the 24th PACLIC, Sendai, Japan, 2010; 365–372.
(5) Denoual E. Analogical translation of unknown words in a statistical machine translation framework. Proceedings of MT Summit XI, Copenhagen, Denmark, 2007; 135–141.
(6) Doddington G. Automatic evaluation of machine translation quality using N-gram co-occurrence statistics. Proceedings of the 2nd HLT, San Diego, 2002; 138–145.
(7) Fujita A, Carpuat M. FUN-NRC: Paraphrase-augmented phrase-based SMT systems for NTCIR-10 PatentMT. Proceedings of the 10th NTCIR, 2013; 327–334.
(8) Haddow B, Koehn P. Analysing the effect of out-of-domain data on SMT systems. Proceedings of the 7th WMT, Montreal, Canada, 2012; 422–432.
(9) Henrquez QAC, Costa-jussá RM, Daudaravicius V, Banchs ER, Mariño BJ. Using collocation segmentation to augment the phrase table. Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR, Uppsala, Sweden, 2010; 98–102.
(10) Koehn P. Europarl: a parallel corpus for statistical machine translation. Proceedings of MT Summit X, Phuket, 2005; 79–86.
(11) Koehn P, Och FJ, Marcu D. Statistical phrase-based translation. Proceedings of the NAACL-HLT, Edmonton, 2003; 48–54.
(12) Koehn P, Hoang H, Birch A, Callison-Burch C, Federico M, Bertoldi N, Cowan B, Shen W, Moran C, Zens R, et al. Moses: open source toolkit for statistical machine translation. Proceedings of the 45th ACL, Prague, Czech Republic, 2007; 177–180.
(13) Langlais P, Patry A. Translating unknown words by analogical learning. Proceedings of EMNLP/CoNLL, Prague, Czech Republic, 2007; 877–886.
(14) Lepage Y. Analogy and formal languages. *Electronic Notes in Theoretical Computer Science* 2004; **53**:180–191.
(15) Lepage Y, Denoual E. ALEPH: an EBMT system based on the preservation of proportional analogies between sentences across languages. Proceedings of the 2nd IWSLT, Pittsburgh, PA, USA, 2005; 47–54.
(16) Lepage Y, Migeot J, Guillerm E. A measure of the number of true analogies between chunks in Japanese. Proceedings of the 3rd LTC, Poland, 2007; 154–164.
(17) Ma Y, Stroppa N, Way A. Bootstrapping word alignment via word packing. Proceedings of the 45th ACL, Prague, Czech Republic, 2007; 304–311.
(18) Marton Y, Callison-Burch C, Resnik P. Improved statistical machine translation using monolingually-derived paraphrases. Proceedings of the EMNLP, Singapore, 2009; 381–390.
(19) Miclet L, Bayroudh S, Delhay A. Analogical dissimilarity: definitions, algorithms and two experiments in machine learning. *Journal of Artificial Intelligence Research* 2008; **32**:793–824.
(20) Nießen S, Och FJ, Leusch G, Ney H. An evaluation tool for machine translation: Fast evaluation for machine translation research. Proceedings of the 2nd LREC, Athens, 2000; 39–45.
(21) Och FJ, Ney H. A systematic comparison of various statistical alignment models. *Computational Linguistics* 2003; **29**:19–51.
(22) Papineni K, Roukos S, Ward T, Zhu WJ. BLEU: a method for automatic evaluation of machine translation. Proceedings of the 40th ACL, Philadelphia, 2002; 311–318.
(23) Ren Z, Lü Y, Cao J, Liu Q, Huang Y. Improving statistical machine translation using domain bilingual multiword expressions. Proceedings of the Workshop on Multiword Expressions: Identification, Interpretation, Disambiguation and Applications, Suntec, Singapore, 2009; 47–54.
(24) Snover M, Dorr B, Schwartz R, Micciulla L, Makhoul J. A study of translation edit rate with targeted human annotation. Proceedings of the 7th AMTA, Cambridge, MA, 2006; 223–231.
(25) Stolcke A. SRILM-an extensible language modeling toolkit. Proceedings of the 7th ICSLP, Denver, CO, 2002; 901–904.

Juan Luo (Non-member) is currently pursuing the Ph.D. degree in example-based machine translation at the NLP laboratory, Waseda University. Her research interests include machine translation and natural language processing.

Yves Lepage (Non-member) received the D.E.A. and Ph.D. degrees in 1989 from GETA, Grenoble University, France, under the supervision of Prof. Vauquois and Prof. Boitet. After conducting post-doctoral research at ELSAP, University of Caen, and EDF, Paris, he joined ATR labs, Japan, where he worked as an Invited Researcher and a Senior Researcher until 2006. In 2003 he received the Habilitation for his thesis titled 'Of the kind of analogies that renders an account of commutations in linguistics'. In October 2006, he qualified for full professorship from the National Board of French Universities in both linguistics and computer science, and became full professor at the University of Caen Basse-Normandie, in October 2006. He joined the Graduate School of Information, Production and Systems, Waseda University, in April 2010. His research interests include natural language processing, machine translation, and, in particular, example-based machine translation. Lepage is a member of the Japanese Natural Language Processing Associations, a member of the French Natural Language Processing Association, ATALA, and an editor-in-chief of the French journal on Natural Language Processing, TAL.