A Flexible Framework for Extracting Bilingual Dictionary from Comparable Corpus Without any Language-Specific Knowledge

Xiaodong Liu[1,a)] Kevin Duh[1,b)] Yuji Matsumoto[1,c)]

Abstract: We propose a flexible and effective framework for extracting bilingual dictionaries from comparable corpora without using any language-specific knowledge such as seeds or additional dictionaries. Our approach is based on a novel combination of topic modeling and word alignment techniques in a pipeline style: first, our approach converts a comparable *document*-aligned corpus into a parallel *topic*-aligned corpus using topic modeling techniques, then learns translation relationships between words using word alignment models such as IBM model I. Compared with previous work, our framework is advantageous in that it only uses the statistical information without requiring any language-specific knowledge for initialization. Furthermore, our framework is capable of handling polysemy: for example, it can extract distinct translations for the word "Apple" as a fruit or as a company. Experiments on a large-scale Wikipedia corpus, show that our framework reliably extracts high-precision word pairs on a wide variety of comparable data conditions.

1. Introduction

A bilingual dictionary or translation dictionary is a specialized dictionary used to translate words or phrases from one language to another which plays a fundamental role in both machine translation community and cross-lingual information retrieval community. In a machine translation community, a high-quality bilingual dictionary can be extremely helpful in improving the quality of translation in the domain of interest [6] in both rule-based and statistical machine translation systems. It is also used as an efficient means for query translation in cross-language information retrieval [10].

One approach for building a bilingual dictionary resource uses parallel sentence-aligned corpora. This is often done in the context of Statistical MT, using word alignment algorithms such as the IBM models [2], [3]. Parallel corpora are plentiful for a few high-resource language pairs such as English-Chinese or English-French, however, they are scarce or non-existent in the majority of the world's languages [26]. Furthermore even for high-resource language pairs, there may exist a shortage of parallel corpus for specific domains of interest (e.g., medical and microblog domains).

On the other hand, with the development of internationalization, comparable corpora are becoming increasingly available. Here, a comparable corpus is defined as collections of document pairs written in different languages while talking about the same topic [25]. Examples include VOA English-Chinese news and Wikipedia[*1]. The challenge with bilingual dictionary extraction from comparable corpus is that existing word alignment methods developed for parallel corpus cannot be directly applied. Thus, this has become an active research topic, with many proposed methods [23], [27]. Our approach follows the tradition set forth by the previous work.

Our framework is a novel combination of topic models and word alignment, as indicated in **Fig. 1**. Intuitively, our approach works by first converting a comparable *document*-aligned corpus into a parallel *topic*-aligned corpus, then apply word alignment methods to model co-occurence within topics. By employing topic models, we avoid the need for seed lexicon and operate purely in the realm of unsupervised learning. By using word alignment techniques on topic model results, we can easily model polysemy and extract topic-dependent lexicon pairs.

The remainder of this paper is organized as following. Section 2 describes related work in dictionary extraction. Section 3 provides background on the main sub-components of our framework, polylingual topic model and word alignment, and can be skipped for the knowledgeable reader. Our proposed framework is presented in detail in Section 4. Finally, we discuss the experiment results in Section 5 and give a conclusion of our work with future work in Section 6.

2. Related Work

There is a plethora of research on bilingual lexicon extraction from comparable corpora, starting with seminal works of

[1] Nara Institute of Science and Technology, Nara, 630-0192, Japan
[†1] Presently with Nara Institute of Science and Technology
[a)] xiaodong-l@is.naist.jp
[b)] kevinduh@is.naist.jp
[c)] matsu@is.naist.jp

[*1] Wikipedia is a free, collaboratively edited, and multilingual Internet encyclopedia.

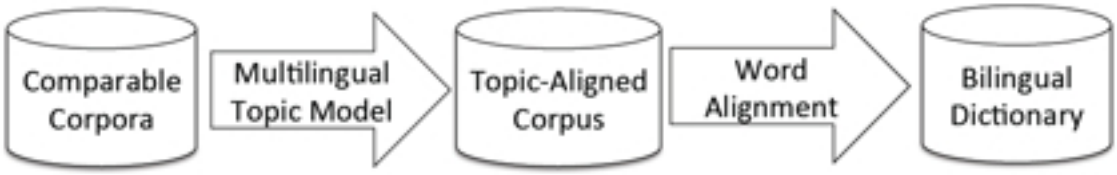

Fig. 1 Proposed Framework for Bilingual Dictionary Extraction

[13], [19]. The assumption employed by most works that translation pairs will have similar contexts, i.e., the *distributional hypothesis*. The basic extraction algorithm consists of 3 steps: (1) identify context windows around words, (2) translate context words using a seed bilingual dictionary, and (3) extract pairs that have high resulting similarity. Methods differ in how the seed dictionary is acquired [14], [30] and how similarity is defined by [21], [31]. Alternative projection-based approaches have also been proposed, though they can be shown to be related to the aforementioned distributional approaches [15]. For example, the paper [23] uses canonical component analysis to map vectors in different languages into the same latent space. The paper [17] presents a good summary.

The paper [27] pioneered a new approach to bilingual dictionary extraction based on topic modeling approach. Compared to the above works, the topic modeling approach requires no seed dictionary. While our approach is motivated by [27], we exploit the topic model in a very different way (explained in Section 4.2). They do not use word alignments like we do and thus cannot model polysemy. Further, their approach requires training topic models with a large number of topics, which may limit the scalability of the approach.

In the machine learning community, there has been a surge of interest in developing multilingual topic models [22], [29], [32], [33]. Many of these models give $p(t|e)$ and $p(t|f)$, but stop short of extracting a bilingual lexicon. Although topic models can group related e and f in the same topic cluster, the extraction of a high-precision dictionary requires additional effort. One of our contributions here is that we demonstrate a flexible way to exploit the advances of topic modeling research by running word alignment on top of the topic results. Although we experiment with the topic model of [29], our framework can plug-in any other multilingual topic model.

Our work is much inspired by the work of [28], which proposed a word alignment model that incorporates topic models. That model learns $p(e|f,t)$ given parallel text. One of the main contribution is to extend this concept of topic-dependent bilingual lexicon to comparable corpora.

3. System Components

Now let us describe the main components used in our framework: Multilingual Topic Models and Word Alignment algorithms. We take advantage of existing techniques–the knowledgeable reader may wish to only skim this section before Section 4, which describes our contribution.

3.1 Multilingual Topic Model

Any multilingual topic model may be used with our framework. Here we present a popular one by [29], which extends the monolingual Latent Dirichlet Allocation [1] to multilingual corpora in a straightforward fashion. Given a comparable corpus E in English and F in a foreign language, we assume that the document pair boundaries are known. I.e, for each document pair $d_i = [d_i^e, d_i^f]$ consisting of English document d_i^e and Foreign document d_i^f (where $i = 1, \ldots, D$, D is number of document pairs), we know that d_i^e and d_i^f talk about the same contents. We do not assume sentence-level alignments between d_i^e and d_i^f.

for *each topic pair k* **do**
  for $l \in [e, f]$ **do**
    sample $\varphi_k^l \sim Dirichlet(\beta^l)$;
  end
end
for *each document pair* d_i **do**
  sample $\theta_i \sim Dirichlet(\alpha)$;
  for $d_i^l \in [d_i^e, d_i^f]$ **do**
    sample $z^l \sim Multinomial(\theta_i)$;
    for *each word* w^l *in* d_i^l **do**
      sample $w^l \sim p(w^l|z^l, \varphi^l)$;
    end
  end
end

Algorithm 1: Generative procedure for [29]. θ_i is the topic prior shared in a document pair d_i (This is the comparable corpora assumption). Words w^l are drawn from language-specific distributions $p(w^l|z^l, \varphi^l)$, where l is shorthand to indicate whether the language e or f. While monolingual topic models use a single set of topics, here there are pairs of language-specific topics φ^l each drawn from its own language-specific Dirichlet distribution with prior β^l.

While the monolingual topic model lets each document have its own so-called document-specific distribution over topics, the multilingual topic model assumes that documents in each tuple share the same topic prior (thus the comparable corpora assumption) and each topic consists of several language-specific word distributions. For detailed explanations, please see the generative story in Algorithm 1 or its graphic model representation (**Fig. 2**).

$$p(z_n^l|w, z_{-l,n}, \Phi^1, ..., \Phi^L, \alpha m) \propto \phi^l_{w_n^l|t} \frac{(N_t)_{-l,n} + \alpha m}{\sum_t N_t - 1 + \alpha} \qquad (1)$$

3.2 Statistical Word Alignment

Word alignment is one of the basic components of a statistical machine translation system [25]. Given a sentence-aligned

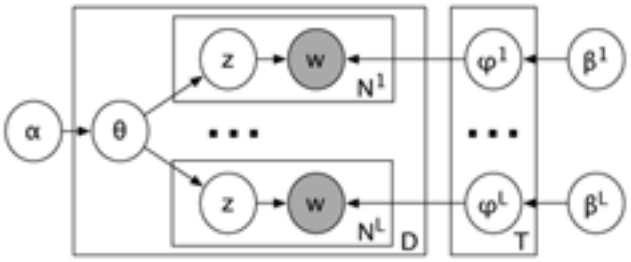

Fig. 2

corpora, the goal is to find the word-by-word correspondence between each English w^e and foreign word w^f in each sentence pair. We will be using the standard IBM Model 1, which we describe below. Refer to [2], [3] for details.

For a given sentence-pair $(\boldsymbol{e}, \boldsymbol{f})$, let $\boldsymbol{e} = [w_1^e, w_2^e, \ldots w_{|\boldsymbol{e}|}^e]$ be the English sentence with $|\boldsymbol{e}|$ words and $\boldsymbol{f} = [w_1^f, w_2^f, \ldots w_{|\boldsymbol{f}|}^f]$ be the foreign sentence with $|\boldsymbol{f}|$ words. For notation, we will index English words by i and foreign words by j. The goal of word alignment is to find an alignment function $a : i \rightarrow j$ mapping words in $\boldsymbol{e}$ to words in $\boldsymbol{f}$ (and vice versa). IBM Model 1 proposes the following probabilistic model for alignment:

$$p(\boldsymbol{e}, a, |\boldsymbol{f}) = \epsilon \prod_{i=1}^{|\boldsymbol{e}|} p(w_i^e | w_{a(i)}^f) \tag{2}$$

Here, $p(w_i^e | w_{a(i)}^f)$ captures the probability of translation the English word at position i from the foreign word at position $j = a(i)$, where the actual alignment a is a hidden variable. The term ϵ is a normalization constant to ensure a valid probability distribution. Since the alignments are originally hidden, the EM algorithm is used to learn the parameters (i.e. word translation probability table). Finally, given the parameters, one could infer the most likely alignments given any sentence pair.

The above model does not use any linguistic knowledge other than the fact that each English word has only one foreign word as translation. More advanced models have been proposed, e.g. those that incorporate multi-word translations and word order biases [2], [3], though for our purposes this simple model is sufficient. The central point to remember is that word alignment techniques enable us to *find correspondence between distinct objects from paired sets*. In the case of machine translation, the distinct objects are words from different languages while the paired sets are sentence-aligned corpora. In our case, our distinct objects are also words from distinct languages but our paired sets will be *topic*-aligned corpora.

4. Proposed Framework for Bilingual Dictionary Extraction

The general idea of our proposed framework was already mentioned in Fig. 1: First, we run a multilingual topic model to convert the comparable corpora to topic-aligned corpora. Second, we run a word alignment algorithm on the topic-aligned corpora in order to extract translation pairs. The main innovation is in how this topic-aligned corpora is defined and constructed, the link between the two stages. We describe how this is done in Section 4.1 and show how existing approaches are subsumed in our general framework in Section 4.2.

4.1 Topic-Aligned Corpora

Suppose the original comparable corpus has D document pairs $[d_i^e, d_i^f]_{i=1,\ldots,D}$. We then run a multilingual topic model with K topics, where K is user-defined (Section 3.1). The topic-aligned corpora is defined hierarchically as a *set of sets*: On the first level, we have a set of K topics, $\{t_1, \ldots, t_k, \ldots, t_K\}$. On the second level, for each topic t_k, we have a set of D "word collections" $\{C_{k,1}, \ldots, C_{k,i}, \ldots, C_{k,D}\}$. Each word collection $C_{k,i}$ represents the English and foreign words that occur simultaneously in topic t_k and document d_i.

For clarity, let us describe the topic-aligned corpora construction process step-by-step together with a flow chart in Figure 3:

(1) Train a multilingual topic model.

(2) Infer a topic assignment for each word in the comparable corpora, and generate a list of word collections $C_{k,i}$ occurring under a given topic.

(3) Re-arrange the word collections such that $C_{k,i}$ belonging to the same topic are grouped together. We call this resulting set of sets a topic-aligned corpora, since it represents word collections linked by the same topics.

(4) For each topic t_k, we run statistical word alignment on $\{C_{k,1}, \ldots, C_{k,i}, \ldots, C_{k,D}\}$. In analogy to statistical machine translation, we can think of this dataset as a parallel corpus of D "sentence pairs", where each "sentence pair" contains the English and foreign words that co-occur under the same topic and same document. Note that word alignment is run independently for each topic, resulting in K topic-dependent lexicons $p(w^e | w^f, t_k)$.

(5) To extract a bilingual dictionary, we find pairs (w^e, w^f) with high probability under the model:

$$p(w^e | w^f) = \sum_k p(w^e | w^f, t_k) p(t_k | w^f) \tag{3}$$

The first term is the topic-dependent bilingual lexicon from Step 4, while the second term is the topic posterior of a word defined by the topic model in Step 1.

In practice, we will compute the probabilities of Equation 3 in both directions: $p(w^e | w^f)$ as in Eq. 3 and $p(w^f | w^e) = \sum_k p(w^f | w^e, t_k) p(t_k | w^e)$. The bilingual dictionary can then be extracted based on a probabilities threshold or some bidirectional constraint. We choose to use a bidirectional constraint because it gives very high-precision dictionaries and avoid the need to tune probability thresholds. In particular, a pair $(\tilde{e}, \tilde{f})$ is extracted if the following holds:

$$\tilde{e} = \arg\max_e p(e | f = \tilde{f}) \tag{4}$$

$$\tilde{f} = \arg\max_f p(f | e = \tilde{e}) \tag{5}$$

To summarize, the main innovation of our approach is that we allow for polysemy as topic-dependent translation explicitly in

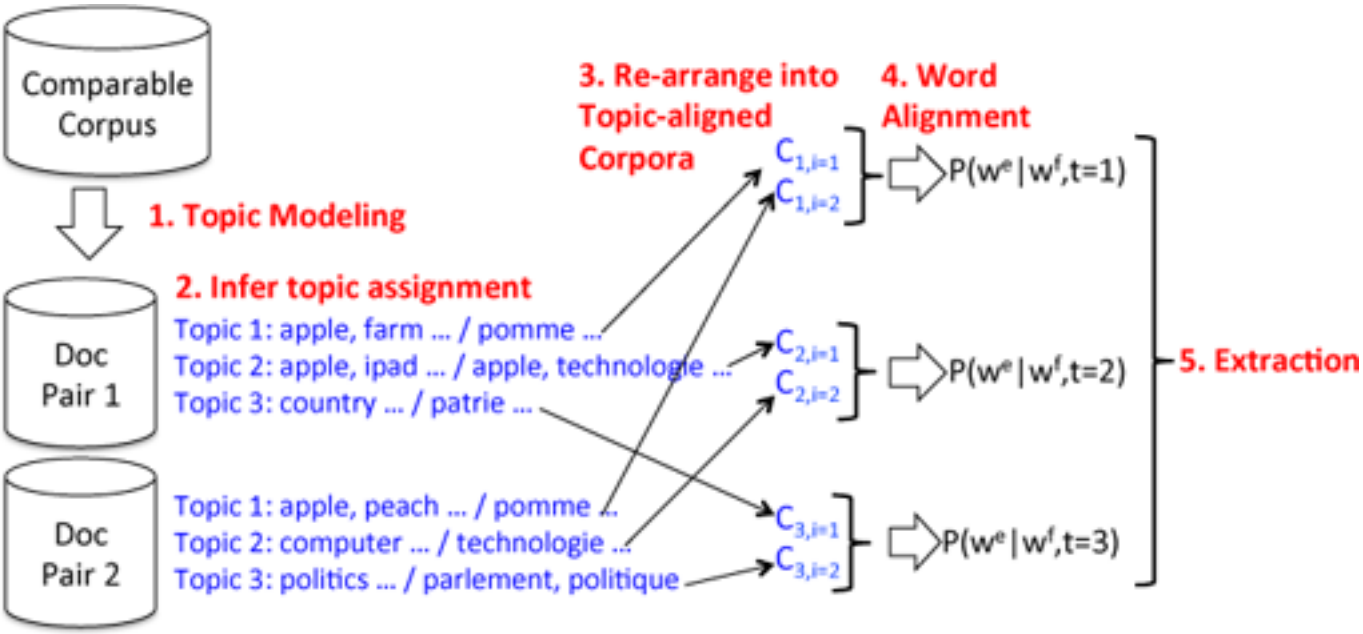

Fig. 3 Construction of topic-aligned corpora.

Equation 3, and use a novel combination of topic modeling and word alignment techniques to compute the term $p(w^e|w^f, t_k)$ in an unsupervised fashion.

4.2 Alternative Approaches

We can view the alternative approaches proposed by [27] as a simplification of Equation 3. To the best of our knowledge, [27] is the only work that uses topic models for bilingual lexicon extraction like ours, but they exploit the topic model results in a different way and do not utilize word alignment techniques. Their so-called "Cue Method" computes:

$$p(w^e|w^f) = \sum_k p(w^e|t_k)p(t_k|w^f) \qquad (6)$$

Equation 6 is similar to our Equation 3 but assumes that the probability of generating w^e is independent of w^j given topic, i.e. $p(w^e|t_k, w^j) = p(w^e|t_k)$. Another variant is the so-called Kullback-Liebler (KL) method, which scores translation pairs by $-\sum_k p(t_k|w^e) \log p(t_k|w^e)/p(t_k|w^f)$. In either case, their contribution is the use of topic-word distributions like $p(t_k|w^f)$ or $p(w^f|t_k)$ to compute translation probabilities.[*2] The main difference with our method is that theirs assumes a conditional independence assumption where w^e is independent of w^f given t_k; we do not make this assumption and focus on estimating $p(w^e|w^f, t_k)$.

5. Experimental Setup

5.1 Data Set

We perform experiments on the "Japanese-English Bilingual Corpus of Wikipedia's Kyoto Articles" (Kyoto Wiki Corpus) [*3]. We chose this corpus for several reasons:

(1) It is a *parallel* corpus, where the Japanese is translated manually into English sentence-by-sentence. This allows us to control the experimental conditions carefully. We do *not* use the sentence-alignment information in our extraction algorithms, but we use it help create gold-standard translation pairs for evaluation. Furthermore, by artificially deleting parts of the corpus, we can simulate successively more or less comparability in the corpora.

(2) It is a large-scale corpus with around 14,000 documents and 472,000 sentences, covering a variety of topics concerning Kyoto tourism, Japanese history, culture, religion, literature. This enables us to test the scalability of our approach with large datasets and high vocabulary.

We then prepared several versions of the data, as shown in Table **Table 1**. **Parallel** is the original sentence-aligned Kyoto Wiki Corpus. **Comp%100** is a comparable version of **Parallel** that deletes all the sentence alignments but otherwise keeps all content on both Japanese and English sides. In other words, it is a corpus comparable on the *document* level, where we know for sure both sides talk about the same topics. To simulate lower degrees of comparability, we then randomly delete English sentences from **Comp100%**. **Comp%50** is a version that deletes half of all sentences, whereas **Comp%20** is a more extreme version that deletes 80% of sentences, leaving only 20% of "comparable information".

Some researchers have stressed the importance of comparable data *quality* [4], and advocated for techniques to clean up comparable corpus prior to bilingual dictionary extraction [18]. Unfortunately, in bilingual dictionary extraction research, it is currently not common practice to evaluate on a wide range of datasets of different degrees of comparability. We hope our simple method of simulating various versions of comparable corpora will encourage more researchers to examine the question of sensitivity to data conditions.

[*2] A third variant uses TF-IDF weighting, but conceptually all are similar. The Cue Method (and it's TF-IDF variant) are reported to have best results.

[*3] Available at: *http : //alaginrc.nict.go.jp/WikiCorpus/index$_E$.html*

Dataset	Comments	#sent(e)	#sent(j)	voc(e)	voc(j)
Parallel	original, sentence-aligned	472k	472k	152k	116k
Comp100%	comparable, alignment deleted	472k	472k	152k	116k
Comp50%	Comp100% with 50% content	236k	472k	100k	116k
Comp20%	Comp100% with 20% content	94k	472k	62k	116k

Table 1 Datasets used. The number of sentences (#sent) and vocabulary size (voc) of English (e) and Japanese (j) for each dataset. For pre-processing, we did word segmentation on Japanese using Kytea [11] and Porter stemming on English. A TF-IDF based stop-word lists of 1200 in each language is applied.

5.2 Evaluation Methodology

The first step in evaluation is to prepare a "gold standard" bilingual lexicon. Our focus on large-scale datasets necessitates the use of automatic evaluation methods. In order to obtain "gold standard" lexicon, we exploit the fact that the original dataset, **Parallel**, is sentence-aligned. Therefore, we can obtain a bilingual lexicon automatically using developed techniques from statistical word alignment. In particular, we trained IBM Model 4 using GIZA++ [3] for both directions $p(e|f)$ and $p(f|e)$. Then, we extract word pair $(\tilde{e}, \tilde{f})$ as a "gold standard" bilingual lexicon if it satisfies Eq. 5. Due to the large size of the dataset and the strict bidirectional requirement imposed by Equation 5, these "gold standard" bilingual dictionary items are of high quality.[*4]

Given the "gold standard", we can compute the standard Precision metric, i.e. number of correct items (e, f) divided by total number of predicted items:

$$Precision = \frac{|\{Gold(e,f)\} \bigcap \{Predicted(e,f)\}|}{|\{Predicted(e,f)\}|} \qquad (7)$$

Another standard evaluation metric is #Extracted=$|\{Predicted(e,f)\}|$, or simply the number of extracted lexicon pairs. This raw number is preferred over Recall, i.e. $|\{Gold(e,f)\} \bigcap \{Predicted(e,f)\}|/|\{Gold(e,f)\}$, because the gold standard is optimized for high precision and we cannot guarantee that it contains all true translation pairs in the data.

Finally, to corroborate our automatic evaluation results, we also perform manual evaluation of precision in all our experiments (based on a bilingual speaker's evaluation of 100 randomly-drawn predictions).

5.3 Experimental Results

In our experiments, we want to check how does the proposed framework compare to previous work.

We would like to see how our proposed framework compares with other topic-modeling approaches from previous work, namely [27]. We therefore compared three methods:

- **Proposed**: The proposed method which exploits a combination of topic modeling and word alignment to incorporate topic-dependent translation probabilities (Equation 3).
- **Cue**: The Cue Method in [27], i.e. Equation 6.
- **JS**: The KL-Divergence method in [27] (Section 4.2). Symmetrizing KL by Jensen-Shannon (JS) divergence improves results, so we report this variant.

All methods use the same in-house implementation of polylingual topic models [29], with hyper-parameters set as $\alpha = 50/K$ and $\beta = 0.1$ following [27]. Both the **Proposed** and **Cue** methods generate probability distributions of the form $p(e|f)$ and $p(f|e)$, so to extract bilingual lexicons for evaluation we employ the constraint in Equation 5. The **JS** method is symmetric so given any f the e with lowest JS score is extracted.

Table 2 show the results of 4 different metrics on the **Comp50%** dataset. Observations are:

(1) For all K, **Proposed** is the best method in all Precision-based metrics. On this **Comp50%** which contains a significant amount of imbalance, the **Proposed** method can already achieve 88-91% SubsetPrecision on a large number of extracted pairs (6766-9076).

(2) The **JS** method suffers from extremely poor precision; the **Cue** method achieves reasonable precision, but suffers from insufficient #Extracted. Both methods improve as we increase K, and this is consistent with results by [27] which showed best results with $K > 2000$. However, training a topic model with such a large number of topics is computationally-demanding for a corpora as large as ours. In this regard, the **Proposed** method is much more *scalable*, achieving good results with low K, satisfying one of original desiderata.[*5]

K	Method	Precision	ManualPrecision	#Extracted
50	JS	0.013	0.010	115707
50	Cue	0.350	0.418	43
50	Proposed	**0.699**	**0.750**	9076
200	JS	0.031	0.05	106543
200	Cue	0.638	0.720	136
200	Proposed	**0.728**	**0.800**	7834
400	JS	0.035	0.060	93650
400	Cue	0.734	0.740	276
400	Proposed	**0.761**	**0.850**	6766

Table 2 Comparison of various topic-modeling approaches for bilingual dictionary extraction on the **Comp50%** dataset. For each K (number of topics), best results per metric are boldfaced.

The observant reader may note **Cue** suffers from low recall (extremely few #Extract) since the bidirectional constraint of Eq. 5 is very strict. Conversely, **JS** suffers from low precision because it extracts to many items. We think these multiple metrics together imply conclusively that our proposed framework is a very effective method for bilingual dictionary extraction.

[*4] A manual check on 100 random items confirmed that the quality of this "gold standard" is around 94% precision.

[*5] We have a hypothesis as to why **Cue** and **JS** depend on large K. Eq. 3 is a valid expression for $p(w^e|w^f)$ that makes little assumptions. We can view Eq. 6 as simplifying the first term of Eq. 3 from $p(w^e|t_k, w^f)$ to $p(w^e|t_k)$. Both probability tables have the same output-space (w^e), so the same number of parameters is needed in reality to describe this distribution. By throwing out w^f, which has large cardinality, t_k needs to grow in cardinality in order to compensate for the loss of expressiveness.

6. Conclusion

We have proposed an effective way to extract bilingual dictionaries by a novel combination of topic modeling and word alignment techniques. The key innovation is the conversion of a comparable *document*-aligned corpus into a parallel *topic*-aligned corpus, which allows word alignment techniques to learn topic-dependent translation models of the form $p(w^e|w^f, t_k)$. While this kind of topic-dependent translation has been proposed for the parallel corpus [28], we are the first to enable it for comparable corpora. Our large-scale experiments demonstrated the the proposed framework outperforms existing baselines under both automatic metrics and manual evaluation. We further show that our topic-dependent translation models can capture some of the polysemy phenomenon important in dictionary construction.

There are several avenues for future work:

(1) Previous work [12], [20] have shown that language pivots can significantly improve the precision of distribution-based approaches for bilingual dictionary extraction. Since multilingual topic models can easily train on more than 3 languages, it would be interesting to examine how massively multilingual data can help the topic-modeling approaches.

(2) We are interested in exploring other topic models and word alignment techniques in our framework. In particular, we think hierarchical topic models [8] and bidirectional word alignment [9] can improve results further.

(3) From the experiments, it is very easy to find lots of candidates extracted either the JS method nor proposed approach while the precision is not very perfect. In the future, we also would like to develop some filter approaches to remove the bad translation candidates to increasing the precision.

References

[1] Blei, D., Ng, A, and Jordan, M.: MWEs and topic modelling: enhancing machine learning with linguistics, *Journal of Machine Learning Research*, (2003).

[2] Brown, P., Pietra, S.D., Pietra, V.D., and Mercer, R.: The Mathematics of Statistical Machine Translation: Parameter Estimation *Computational Linguistics*, Vol. 19, No. 2, (1993).

[3] Och, Franz Josef and Ney, Hermann: A systematic comparison of various statistical alignment models, *Comput. Linguist.*, Vol. 29, pp. 19–51, DOI: 10.1162/089120103321337421 (2003).

[4] Morin, Emmanuel and Dailles, Béatrice and Takeuchi, Koichi and Kageura, Kyo: Brains, not brawn: The use of smart comparable corpora in bilingual terminology mining, *ACM Trans. Speech Lang. Process.*, Vol. 7, No. 1, pp. 1:1–1:23, DOI: http://doi.acm.org/10.1145/1839478.1839479 (2008).

[5] Koehn, Philipp.: *Statistical Machine Translation*, Cambridge University Press, 1nd edition (2010).

[6] Daume III, Hal and Jagarlamudi, Jagadeesh: Domain Adaptation for Machine Translation by Mining Unseen Words, *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp.407–412 DOI: http://www.aclweb.org/anthology/P11-2071 (2011).

[7] Baldwin, Timothy: MWEs and topic modelling: enhancing machine learning with linguistics, *Proceedings of the Workshop on Multiword Expressions: from Parsing and Generation to the Real World*, DOI: http://dl.acm.org/citation.cfm?id=2021121.2021123 (2011).

[8] Haffari, G., and Teh, Y.W.,: Hierarchical Dirichlet Trees for information retrieval, *NAACL*, (2009).

[9] DeNero, J., and Macherey, K.,: Model-Based Aligner Combination Using Dual Decomposition, *Proceedings of the Association for Computational Linguistics (ACL)*, (2011).

[10] Resnik, P., Oard, D., and Levow G.,: Improved Cross-Language Retrieval using Backoff Translation, *Proceedings of the First International Conference on Human Language Technology*, (2011).

[11] Neubig, G., Nakata, Y., and Mori S.,: Pointwise Prediction for Robust, Adaptable Japanese Morphological Analysis, *The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT) Short Paper Track*, (2011).

[12] Mausam and Stephen Soderland and Oren Etzioni and Daniel S. Weld and Michael Skinner and Jeff Bilme: Compiling a massive, multilingual dictionary via probabilistic inference, *ACL*, (2009).

[13] Rapp, R.,: dentifying word translations in non-parallel texts, *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, (1995).

[14] Djean, H., and Gaussier, E., and Sadat, F.,: An approach based on multilingual thesauri and model combination for bilingual lexicon extraction, *Proceedings of the 19th international conference on Computational linguistics - Volume 1*, (2002).

[15] Gaussier, E., Renders, J.M., and Matveeva, I., Goutte, C., and Dejean, H.,: A Geometric View on Bilingual Lexicon Extraction from Comparable Corpora, *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL'04), Main Volume*, pp. 526–533 (2011).

[16] Dyer, Chris and Clark, Jonathan H. and Lavie, Alon and Smith, Noah, A.,: Unsupervised Word Alignment with Arbitrary Features, *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp.409–419 (2011)

[17] Laroche, A., and Langlais, P.,: Revisiting Context-based Projection Methods for Term-Translation Spotting in Comparable Corpora, *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pp.617–625 (2010)

[18] Li, B., and Gaussier, E.,: Improving corpus comparability for bilingual lexicon extraction from comparable corpora, *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, (2010)

[19] Fung, P., and Lo, Y.Y.,: Translating Unknown Words Using Nonparallel, Comparable Texts, *Compiling a massive, multilingual dictionary via probabilistic inference booktitle*, (2009)

[20] Shezaf, D., and Rappoport, A.,: Bilingual lexicon generation using non-aligned signatures, *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, DOI: http://dl.acm.org/citation.cfm?id=2021121.2021123 pp, 98–107, (2010)

[21] Tamura, A., and Watanabe, T., and Sumita, E.,: Bilingual Lexicon Extraction from Comparable Corpora Using Label Propagation, *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, DOI: http://www.aclweb.org/anthology/D/D12/D12-1003 pp, 24–36, (2012)

[22] Jagarlamudi, J., and Daume, H.,: Extracting Multilingual Topics from Unaligned Comparable Corpora, *ECIR*, (2010)

[23] Haghighi, A., and Liang, P., Berg-Kirkpatrick, T., and Klein, D.,: Learning Bilingual Lexicons from Monolingual Corpora, *Proceedings of ACL-08: HLT*, DOI: http://www.aclweb.org/anthology/P/P08/P08-1088 pp, 771–779, (2008)

[24] Heinrich, G.,: Parameter estimation for text analysis (2004).

[25] Koehn, P.,: Statistical Machine Translation, Cambridge University Press, (2010).

[26] Schafer, C., and Smith, D.,: An Overview of Statistical Machine Translation *AMTA Tutorials*, (2006)

[27] Vulić, I., De Smet, W., and Moens, M.,: Identifying Word Translations from Comparable Corpora Using Latent Topic Models, *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, DOI: http://www.aclweb.org/anthology/P11-2084 pp, 479–484, (2011)

[28] Zhao, B., and Xing, E.,: HM-BiTAM: Bilingual Topic Exploration, Word Alignment, and Translation, *NIPS*, DOI: http://books.nips.cc/papers/files/nips20/NIPS2007_0188.pdf (2007)

[29] Mimno, D., Wallach, H., Naradowsky, N., Smith, D., and McCallum, A.,: Polylingual topic models, *EMNLP*, pp, 479–484, (2009)

[30] Koehn, P., and Knight, K.,: Learning a translation lexicon from monolingual corpora, *Proceedings of ACL Workshop on Unsupervised Lexical Acquisition*, (2002)

[31] Fung, P., and Cheung, P.,: Mining verynon-parallel corpora: Parallel sentence and lexicon extraction via bootstrapping and EM., *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, (2004)

[32] Boyd-Graber, J., and Blei, M.D.,: Multilingual Topic Models for Unaligned Text, *UAI*, (2009)

[33] Ni, X., Sun, J.T., Hu, J., and Chen, Z.,: Mining Multilingual Topics from Wikipedia, *WWW*, (2009)