Pivot-based Triangulation for Low-Resource Languages

Rohit Dholakia rdholaki@cs.sfu.ca
Anoop Sarkar anoop@cs.sfu.ca
School of Computing Science,
Simon Fraser University,
Burnaby, V5A 1S6, Canada

Abstract

This paper conducts a comprehensive study on the use of triangulation for four very low-resource languages: Mawukakan and Maninkakan, Haitian Kreyol and Malagasy. To the best of our knowledge, ours is the first effective translation system for the first two of these languages. We improve translation quality by adding data using pivot languages and experimentally compare previously proposed triangulation design options. Furthermore, since the low-resource language pair and pivot language pair data typically come from very different domains, we use insights from domain adaptation to tune the weighted mixture of direct and pivot based phrase pairs to improve translation quality.

1 Introduction

Triangulation for phrase-based statistical machine translation (SMT) [Utiyama and Isahara, 2007, Cohn and Lapata, 2007, Wu and Wang, 2007] refers to the use of a pivot language when translating from a source language to a target language. Previous research into triangulation for machine translation either used Europarl [Cohn and Lapata, 2007, Utiyama and Isahara, 2007, Huck and Ney, 2012] or languages with large corpora in same domain [Gispert and Mario, 2006] or assumed presence of languages that are very closely related [Nakov and Ng, 2012, Wang et al., 2012]. However, low resource languages are quite different when it comes to the kind and size of parallel data that is available. This paper considers machine translation into English from four diverse low-resource languages: Mawukakan and Maninkakan, which are West African languages; Haitian Kreyol, in the domain of short messages sent in the aftermath of the Haiti earthquake in 2010; and Malagasy, an Austronesian language from Madagascar. This is the first comprehensive study of triangulation for these four languages, and to our best knowledge, Mawukakan and Maninkakan have not been studied before in the SMT literature.

Faced with a low-resource language pair, several questions arise when trying to use the approach of triangulation:

- [Utiyama and Isahara, 2007] use a different way of computing lexical scores from [Cohn and Lapata, 2007]. Which one is better suited for triangulation in a resource-poor scenario?
- In [Utiyama and Isahara, 2007, Cohn and Lapata, 2007, Wang et al., 2012, Wu and Wang, 2007] many different feature functions are provided for the log-linear model over triangulated phrase pairs. We conduct extensive experiments to show which features should be used for real world low-resource languages based on the data settings for each language pair.
- In [Cohn and Lapata, 2007] a mixture of the direct system and the triangulated system is shown to work better. However, they used uniform weights. In [Wang et al., 2012] a few different weights were selected heuristically while in [Wu and Wang, 2007] 0.9 is assumed for the baseline. We provide an algorithm that combines grid search for learning the mixture weights and minimum error rate training of the direct and triangulated log-linear models.

Setting	Direct	Src-Pivot	Pivot-Tgt	Domains
[Utiyama and Isahara, 2007]	560K	560K	560K	Multi-Parallel
[Cohn and Lapata, 2007]	700K	700K	700K	Multi-Parallel
[Cohn and Lapata, 2007]	10K	10K	10K	Multi-Parallel
Mawukakan	3K	3K	2M	Different
Maninkakan	4K	4K	2M	Different
Haitian Kreyol	120K	30K	2M	Different
Malagasy	88K	30K	2M	Different

Table 1: Comparison of our data settings (last four rows) with previous work. Haitian Kreyol data are short messages sent after earthquake. Malagasy data is automatically aligned news articles in Malagasy). For these two languages we use the Bible as our source-pivot bitext as they have no parallel data source with French, our pivot language. Mawukakan and Mawukakan have a very small source-pivot and source-target bi-texts, but the source-pivot corpus has common sentences with the source-target corpus. We use French as the pivot language to keep the same experimental setting for all our source languages.

To answer some of the above questions, we study the effectiveness of pivot-based triangulation for languages with insufficient resources, Mawukakan, Maninkakan, Malagasy and Haitian Kreyol. Table 1 compares our data settings with previous research into triangulation. Note that we are using all the available data for each language pair. In most cases there was only one possible choice for each source-pivot or pivot-target parallel corpus. Mawukakan and Maninkakan are two languages from the Mandekan family, spoken by almost 3.5 million people in West Africa. The Mandekan languages are a part of the Niger-Congo language family. Maninkakan and Mawukakan have little writing tradition, are written using multiple alphabets[1] and have very little

[1]The data we have used has Latin script, obtained via LDC. See Table 5.

Language	Source/Target
Mawu	*à à ƒ á nè kò búlàmá mùsò kwáò à yá wɛ́ɛ́ó lé é à mátá à* people say that she performs magic
Manin	*àlù bárá álámándɨ́ bèn à kàn , kà à másɽ̀ sɛ́bɛ́ tɛ́ à lá mòrɨ̀fà lá* they fined him because his gun is not legally registered
Ht	*j' aimerais avoir quelques informations svp , concernant ce numero 4636 en quoi je peux l' utiliser* i would like to have information regarding the number 4636. how do i use it
Mlg	*takelaa facebook ho an ' i laura sy euna efa manana mpikambana maherin ' ny dumy arivo sahady* a facebook page for laura and euna already has more than five thousand members

Table 2: An example for each language: Mawu = Mawukakan, Manin = Maninkakan, Ht = Haitian Kreyol, Mlg = Malagasy

resources for machine translation. Malagasy is the national language of Madagascar, spoken by 18 million people worldwide. Haitian Kreyol is the national language of the Republic of Haiti and data used is from the Sixth Workshop on Machine Translation, 2011 [Callison-Burch et al., 2011]. It comprises short messages sent to the number 4636 after the devastating earthquake in January, 2010. Although nine systems participated in the workshop on Haitian Kreyol, the approach of triangulation was not used.

In the aftermath of the earthquake in Haiti in January, 2010, Mission 4636 set up a service where anyone in Haiti could send a message for free to a phone number 4636[2]. A group of volunteers translated the messages into English and helped the relief organizations provide swift help to the affected masses. Microsoft Research released a translation system to the public, for Haitian Creole, 5 days after the devastating earthquake [Lewis et al., 2011]. The fast turnaround time[3] and the usefulness of machine translation in the time of crisis inspired the featured task in the 6th Workshop on Statistical Machine Translation.

Malagasy is an Austronesian language and the national language of Madagascar, spoken by 18 million around the world. Although it shares several words with Ma'anyan, it has influences from Arabic, French, Swahili and Bantu. Characters can have diacritics but not always. Numbers are written right-to-left like Arabic, while some words are in common with French. It follows the Latin alphabet but with 21 characters. Finally, the dataset we have is real-world news articles translated by volunteers across the world[4] and aligned using a sentence aligner, thus, introducing some inconsistencies.

[2] *http://www.mission4636.org*

[3] *To know the exact timeline, refer to http://languagelog.ldc.upenn.edu/nll/?p=2068*

[4] *http://www.ark.cs.cmu.edu/global-voices/*

2 Related Work

Consider a source language *s*, pivot language *p* and target language *t*. When using the *cascading* approach, we build two systems, between *s* and *p* and between *p* and *t*. In this paper, we do not discuss the approach of *cascading*, which would translate sentences in *s* to *p* and use the n-best list to translate the sentence into *t*. It was shown previously [Utiyama and Isahara, 2007, Gispert and Mario, 2006] that *cascading* was not the best approach.

The second approach is the pivot-based approach where a triangulated phrase table is generated between the source and target, by using the common pivot phrases between the source pivot and pivot target tables [Utiyama and Isahara, 2007, Cohn and Lapata, 2007, Wu and Wang, 2007]. [Utiyama and Isahara, 2007] observed that the triangulated table was able to achieve comparable BLEU scores to the direct system for French, German and Spanish. This could be owing to the fact that the data comprised multi-parallel 560K sentences. [Cohn and Lapata, 2007] observe that multiple pivot languages lead to more fluent translations compared to one pivot language. Multiple pivot language lead to multiple alternative translations, thus, increasing phrase coverage and rewarding the more appropriate translations and reducing out-of-vocabulary words further. They also propose a systematic way of combining the triangulated translation model with the direct model using linear interpolation and log-linear interpolation, although they assume the same weight for all the models. To "simulate" a low-resource scenario, the top 10K multi-parallel sentences are considered for source pivot, pivot target and source target systems. [Sennrich, 2012] compared various methods of using linear interpolation in a domain adaptation setting. [Wu and Wang, 2007] also approach triangulation in a similar way to [Cohn and Lapata, 2007] but use different methods to compute lexical weights. [Nakov and Ng, 2012] propose a language-independent approach to improving translation for low-resource languages, but the approach assumes the presence of a resource-rich language that bears similarity to the low-resource language, the similarities helping in creating a large triangulated phrase table. In [Wang et al., 2012], the resource-rich language is adapted to be more like the resource-poor one. Notice that this also assumes both are very similar. Results are reported using both Malay-Indonesian and Bulgarian-Macedonian, the third language being English in both cases. [Gispert and Mario, 2006] translate Catalan to Spanish via English by using news corpora on both source pivot and pivot target side. [Huck and Ney, 2012] report on BLEU score improvements by using 10^9 parallel sentence between German and French.

[Kholy et al., 2013] observe that using categories for source target pairs when combining the direct and triangulated models helped in improving the BLEU score. In other words, a source target pair can be in both the direct and triangulated phrase tables, or only one of them could be in both. They enumerate the different possibilities and use them as separate decoding paths. [Zhu et al., 2013] try to address the problem of missing translations in triangulation (as pivot phrases are not always in

both tables) by using a random walk approach. The initial triangulated phrase table is extended by treating the table as a graph and using a random walk to obtain more translations. [Crego et al., 2010] focus on improving one system (German-English) by using a dynamically build model from auxiliary sources. In other words, they translate the source sentence using various models and then use a framework to combine the different outputs. [Bertoldi et al., 2008] suggested using alternative decoding paths when having different translation models. In our experiments, we found that alternative decoding paths did not work so well. This could be partly be because there are not that many alternatives when having two translation models of very different sizes and from different domains. When we do have alternative paths, they may not always be useful.

A common thread that binds the previous work using the approach of triangulation is the usage of resource-rich languages. The fundamental reason behind the effectiveness of triangulation is the reduction in the number of OOVs when using the pivot language(s). All the Europarl languages are based on parliamentary proceedings and have minimal noise. Hence, the improvements using triangulation over the direct systems cannot be generalized for systems for low-resource languages.

3 Design choices for Triangulation

Given a source language *s*, pivot language *p* and target language *t*, pivot-based triangulation uses common pivot phrases between the source-pivot phrase table p_{sp} and pivot-target p_{pt} to generate a new phrase table between source and target. As the triangulated table is generated using common phrases, the feature values cannot be computed using the alignments and co-occurrence counts. We discuss two ways of computing phrase scores in section 3.1 and two ways of computing lexical scores 3.2. Following [Cohn and Lapata, 2007] we build a mixture model of the direct source-target system and the triangulated source-pivot-target system. In Section 3.4, we propose a new iterative method to find the mixture weights.

3.1 Phrase pair scores

3.1.1 Product Approach

[Utiyama and Isahara, 2007] computes feature values for the triangulation phrases by multiplying values from source-pivot and pivot-target phrase tables:

$$p_{tr}(t \mid s) = \sum_{p} p_{pt}(t \mid p) p_{sp}(p \mid s) \quad (1)$$

$$p_{tr}(s \mid t) = \sum_{p} p_{sp}(s \mid p) p_{pt}(p \mid t) \quad (2)$$

We are marginalizing over the pivot phrases (p), essentially making an independence assumption of the following form, as in [Cohn and Lapata, 2007]:

$$\begin{aligned} p_w(t \mid s) &= \sum_p p_w(t, p \mid s) \\ &= \sum_p p_w(t \mid p, s)\, p_w(p \mid s) \\ &\approx \sum_p p_w(t \mid p)\, p_w(p \mid s) \end{aligned}$$

3.1.2 Joint probability for triangulation scores

[Cohn and Lapata, 2007] propose using the joint probability $p_{tr}(s, t)$ to calculate the triangulated phrase scores $p_{tr}(t \mid s)$ and $p_{tr}(s \mid t)$. Since we do not have observed counts in the triangulated phrase table, counting the pairs after triangulation will not be a true reflection of the joint probability.

The joint probability of a phrase pair looks as shown in Equation 3, which decomposes to Equation 4. Making an independence assumption, shown in Equation 5, we compute the joint probability of a triangulated phrase pair as shown in Equation 6.

$$P(s, t) = \sum_p P(s, p, t) \tag{3}$$

$$P(s, p, t) = \sum_p P(p, t) P(s \mid p, t) \tag{4}$$

$$P(s \mid p, t) \approx P(s \mid p) \tag{5}$$

$$p_{tr}(s, t) = \sum_p p_{pt}(t) p_{pt}(p \mid t) p_{sp}(s \mid p) \tag{6}$$

The counts for the direct system are used to compute the joint and the conditional distributions in this equation.

3.2 Lexical Scores

3.2.1 Product Approach

Similar to phrase scores, we compute the triangulated lexical scores using the product of the lexical scores of the source-pivot and pivot-target tables.

$$p_{\text{lex}_{tr}}(t \mid s) = \sum_p p_{\text{lex}_{pt}}(t \mid p) p_{\text{lex}_{sp}}(p \mid s) \tag{7}$$

$$p_{\text{lex}_{tr}}(s \mid t) = \sum_p p_{\text{lex}_{sp}}(s \mid p) p_{\text{lex}_{pt}}(p \mid t) \tag{8}$$

3.2.2 IBM Model 1 Alignments

[Cohn and Lapata, 2007] propose an alternative way to compute the lexical score by using unsupervised alignments between source and target phrases in the triangulated phrase table. They use the IBM Model 1 [Brown et al., 1993] (Model 1, henceforth) score between the phrase pairs in the triangulated table. Treating the triangulated phrase table as a parallel corpus, we learn the Model 1 alignment scores in both directions using 5 iterations of the EM algorithm [Dempster et al., 1977]. Given a foreign sentence f = $f_1, \ldots, f_m$, english sentence e = $e_1, \ldots, e_l$, the IBM Model 1 score between the sentences is calculated as follows:

$$p(f, a \mid e) = \frac{\epsilon}{(l+1)^m} \prod_{j=1}^{m} t(f_j | e_{a(j)}) \tag{9}$$

Model 1 learns the likelihood of the alignment of the individual words, while also considering the fact that a triangulated table will have less number of source phrases translating into good and some noisy translations. Noisy translations will automatically get a lower Model 1 score, something less likely to happen when using the simpler approach of multiplying the lexical scores. This effect of noisy translations ending up as a viable translation during decoding is also because of the limited source-pivot training corpora available. Several translations have been only seen once and the phrase lengths are not very long either (85% of Mawu or Manin phrase table has ≤ 3 words).

3.3 Connectivity Features:

The phrase pairs in the triangulated phrase table are contingent upon the common pivot phrases. As a result, we can have phrase pairs that map **"!"** to a target phrase **"and making the soup thick !"** in Haitian Kreyol to English triangulated phrase table. Due to the fan-out nature of triangulation, spurious phrase pairs like above get high enough feature values to end up as a translation during decoding. To reward phrase pairs that have more alignment links between and to penalize pairs that don't, we add two connectivity features to the phrase table, as proposed in [Kholy and Habash, 2013] for Persian to Arabic translation using English as the pivot language. For a source phrase p_s, target phrase p_t, and with the number of alignment links between them *N*, the strength feature is:

$$source_{strength} = \frac{\mathrm{N}}{\mathrm{S}}$$
$$target_{strength} = \frac{\mathrm{N}}{\mathrm{T}}$$

where S is the length of the source phrase p_s and T is the length of the target phrase p_t. To compute the connectivity strength feature, the alignments in the source-pivot phrase pair are intersected with the pivot-target phrase pair.

3.4 Translation Model Combination

[Cohn and Lapata, 2007] propose a mixture of the direct source-target system model p_d with the triangulated source-pivot-target model p_{tr}:

$$p_{interp}(s \mid t) = \lambda_d p_d(s \mid t) + (1 - \lambda_d) p_{tr}(s \mid t) \quad (10)$$

In their data setting, setting λ_d to 0.5 was a reasonable choice. [Nakov and Ng, 2012] try different heuristically selected values and re-learn the log-linear weights. However, both of these choices are unreasonable in our low-resource data setting because our datasets come from different domains (thus, using uniform weights would be unreasonable) and using weights that were set heuristically would not be a systematic search over the parameter space.

Grid search for interpolation: For Haitian Kreyol, we are trying to improve translations for real-world short messages using common phrases between Bible and parliamentary proceedings. For Malagasy, we are trying to do the same for news articles. To get the best of both worlds, we would want a λ_d in equation (10) which maximizes our BLEU score, where p_d represents the direct translation model while p_t represents the triangulated translation model.

Algorithm 1: Grid Search for Interpolation

Input: triangulated phrase table p_{tr},
direct phrase table p_d,
λ_d, $\lambda_{tr} = 1 - \lambda_d$, prev_{bleu} = 0,
minimum = e^{-2}
Output: best_{λ_d}

1: **while** $\delta_{bleu} >$ minimum **do**
2: interpolate p_d, p_{tr} to give p_{interp}
3: Run MERT using p_{interp} as translation model
4: find $\text{bleu}_{heldout}$
5: δ_{bleu} = $\text{bleu}_{heldout}$ - prev_{bleu}
6: prev_{bleu} = $\text{bleu}_{heldout}$
7: Based on δ_{bleu}, find newλ_d
8: **end while**

Algorithm 1 is an iterative method that learns the best λ_d using a publicly available toolkit, CONDOR [Berghen and Bersini, 2005], a direct optimizer based on Powell's algorithm, that does not require explicit gradient information for the objective function. The approach can be easily extended to multiple triangulated models or different co-efficients for each feature. Line 2 interpolates the two translation models using equation (10). We re-tune the log-linear weights using MERT for the interpolated feature values (on the same tuning data as the baseline) and use the tuned model to find BLEU score on the same heldout set. Based on the difference between the BLEU score obtained and the previous BLEU (Line 7), CONDOR searches for the new co-efficient in

Language	Best λ_d
Mawukakan	0.84
Maninkakan	0.75
Haitian Kreyol	0.95
Malagasy	0.82

Table 3: Different languages have different interpolation co-efficients that lead to the best system. Although we always start with 0.85, we iterate systematically over different values to obtain the best co-efficient.

Setting	Mawu	Manin	Haitian-Creole	Malagasy
Baseline	7.08	9.41	33.6	18.8
Uniform	7.3	10.50	33.44	18.51
Product	7.39	10.91	33.84	19.17
Product + Strength	7.03	10.80	33.92	19.03
IBM Model 1	**7.64**	10.69	**34.00**	**19.20**
Joint	7.42	**11.06**	33.85	19.10

Table 4: All results for all languages. Product approach refers to using product of both phrase and lexical scores. Strength refers to adding connectivity features on top of Product. IBM Model 1 substitutes IBM Model 1 scores in place of product lexical score while Joint substitutes joint phrase scores in place of product phrase scores. Both Model 1 and joint approaches do not use the strength feature.

the corresponding direction. The search will culminate when consecutive BLEU scores show a marginal difference (Line 1). For instance, we start with a value of 0.85 for the direct system from Mawukakan to English we obtain a BLEU score of 9.10. If we use uniform weights for both the tables, we get BLEU scores on heldout as shown in Table 4. In three of four cases, we would not have out-performed our baseline. We can try 0.50, 0.60, 0.70 and 0.80 [Nakov and Ng, 2012] and run MERT for each choice. Although 0.70 would have given us our best BLEU for this pair, we observed that different languages led to different interpolation weights (Table 3), and this was different for different design choices for each language pair (Haitian Kreyol and Malagasy have disjoint systems). Our method automates grid search for the mixture weight and combines it with minimum error rate training of the log linear models for both direct and triangulated systems.

4 Experiments

4.1 Datasets

Table 5 contains the details about the data sets for the four language pairs. Data for Mawukakan[5] and Maninkakan[6] has been released by LDC. For Malagasy, the training

[5]http://catalog.ldc.upenn.edu/LDC2005L01
[6]http://catalog.ldc.upenn.edu/LDC2013L01

Language	Training	Dev	Heldout
Mawukakan	3076	1000	500
Maninkakan	3600	1000	500
Haitian Kreyol	121K	900	900
Malagasy	88.5K	1133	500

Table 5: Training, Development, Heldout (devtest-clean for Haitian Creole) for all four languages

and development sets have been used as-is[7]. As there is no separate heldout data, the top 500 sentences of the test set is used as heldout. All experiments in Haitian Kreyol use the same training, development, heldout and test sets as the WMT 2011 shared task. The training corpus for Haitian Kreyol comprises only 16% in-domain data, while the development, heldout and test sets comprise only real-world short messages. All but training data for Haitian Kreyol have *raw* and *clean* versions. The *clean* version has the same short messages as *raw*, but have been manually cleaned of misspellings and other errors e.g. **caf*** in *raw* has been changed to **cafe** in *clean*. The pivot language used in all our experiments is French because we can only use French for Mawukakan and Maninkakan. We experimented with additional pivot languages for Haitian Kreyol but they did not help. The pivot-target data set is the 1.9M sentence pair French-English EuroParl (v7) corpus.

4.2 Setup

All the experiments have been run using the Moses toolkit [Koehn et al., 2007], following the standard pipeline. After tokenizing and lowercasing and removing any empty lines, the alignments in both directions are generated using GIZA++ [Och and Ney, 2003], followed by the –grow-diag-final-and heuristic to extract phrases. Weights for the features are learnt based on Algorithm 1 outlined in section 3.4 (Both tuning and heldout sets used are same for all the results in Table 4.) All BLEU scores reported are case-insensitive. SRILM [Stolcke, 2002] was used to generate the language models. For Haitian Kreyol, an interpolated 5-gram language model, using the English side of WMT data and the English side of Europarl, is used. For the other three languages, the language model used is 5-gram Gigaword. We use KenLM [Heafield, 2011] for LM scoring during decoding.

4.3 Results

The BLEU scores for all languages are in Table 4. *Baseline* in Table 4 refers to translation model generated by just using the source-target parallel data for each of the language pairs. We use another baseline, *Uniform* that refers to using a triangulated translation model combined with *Baseline*, but by using uniform weights (0.5 each.) All the BLEU scores, including baseline scores, are reported on the held-out data (devtest-

[7] http://www.ark.cs.cmu.edu/global-voices/

Language	Category	Example translation
Haitian Kreyol	Before	*do we still have earth-shock for haiti ?*
	After	are there always earthquake in haiti ?
	Reference	are there any more earthquakes in haiti ?
Maninkakan	Before	*we will go there tɛ́nɛ́nlón*
	After	we will go there on monday .
	Reference	we will go there on monday .

Table 6: Examples of improvements in translations. These examples show how the pivot language can provide new useful candidate translations missing from the direct system.

Best system	Mawu	Manin	Kreyol	Malagasy
Our best vs Baseline	>0.05	<0.01	>0.05	>0.05
Our best vs Uniform	>0.05	>0.05	<0.03	<0.03

Table 7: Significance tests for our results. All use the same tuning and heldout set. (We used multeval [Clark et al., 2011] for the significance tests)

clean for Haitian Kreyol.) For Haitian Kreyol, our baseline BLEU score is +0.6 BLEU points higher than the best system from the 2011 Workshop on Haitian Kreyol. Despite using a disjoint and out-of-domain Bible as source-pivot, both Haitian Kreyol and Malagasy lead to a better BLEU score compared to both our baselines. For Mawukakan, we observed that triangulation was not causing a significantly better BLEU score while Maninkakan showed a BLEU score increase of 1.5 points. The significance results are shown in Table 7. Except in the case of Haitian Kreyol where it improves by a small margin, adding the two connectivity features reduces the BLEU score. This is likely due to source-pivot having a small intersection with cleaner Europarl alignments. In Mawukakan and Maninkakan, 60% and 66% phrase pairs have a source connectivity strength of more than 0.5 while 67% and 69% have more than 0.5 in the backward direction. Unsupervised alignments on the triangulated phrase table (using IBM Model 1) helps in the case of Malagasy and Haitian Kreyol. Adding the Joint and decomposed conditionals as features does well Maninkakan, leading to the best system for it, while IBM Model 1 lexical scores combined with the product of phrase scores works best for Haitian Kreyol and Malagasy. On the WMT 2011 Haitian Kreyol devtest-clean data, our system gets 34% BLEU score and a prominent web-based free translation system gets 16.72%. Example translations are shown in Table 6.

5 Conclusion

In this paper, we compared previously proposed and novel models, features and design choices in triangulation for low-resource languages. We show that in a noisy domain adaptation setting which we faced in Haitian Kreyol and Malagasy due to the Bible as a source-pivot corpus, the use of unsupervised alignments to compute the phrase

table feature scores led to a significantly higher BLEU score. We showed that the joint probability method [Cohn and Lapata, 2007] is better for languages with short, smaller sized phrase tables which is the case in Maninkakan. For interpolating the direct source-target system with the source-pivot-target system, we introduced an algorithm to automatically learn the mixture weights. Our algorithm provides better results across different low-resource language pairs.

References

[Berghen and Bersini, 2005] Berghen, F. V. and Bersini, H. (2005). CONDOR, a New Parallel, Constrained Extension of Powell's UOBYQA Algorithm: Experimental Results and Comparison with the DFO Algorithm. *J. Comput. Appl. Math.*, 181(1):157–175.

[Bertoldi et al., 2008] Bertoldi, N., Barbaiani, M., Federico, M., and Cattoni, R. (2008). Phrase-based statistical machine translation with pivot languages. In *IWSLT*, pages 143–149.

[Brown et al., 1993] Brown, P. F., Pietra, V. J. D., Pietra, S. A. D., and Mercer, R. L. (1993). The mathematics of statistical machine translation: Parameter estimation. *Comput. Linguist.*, 19(2):263–311.

[Callison-Burch et al., 2011] Callison-Burch, C., Koehn, P., Monz, C., and Zaidan, O. F. (2011). Findings of the 2011 Workshop on Statistical Machine Translation. In *Proceedings of the Sixth Workshop on Statistical Machine Translation and MetricsMATR*. Sixth Workshop on Statistical Machine Translation.

[Clark et al., 2011] Clark, J. H., Dyer, C., Lavie, A., and Smith, N. A. (2011). Better hypothesis testing for statistical machine translation: Controlling for optimizer instability. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers - Volume 2*, HLT '11, pages 176–181, Stroudsburg, PA, USA. Association for Computational Linguistics.

[Cohn and Lapata, 2007] Cohn, T. and Lapata, M. (2007). Machine translation by triangulation: Making effective use of multi-parallel corpora. In *Proceedings of 45th Annual Meeting of the Association of Computational Linguistics*, pages 728–735, Prague, Czech Republic. Association of Computational Linguistics.

[Crego et al., 2010] Crego, J. M., Max, A., and Yvon, F. (2010). Local lexical adaptation in machine translation through triangulation: Smt helping smt. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 232–240, Beijing, China. Coling 2010 Organizing Committee.

[Dempster et al., 1977] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38.

[Gispert and Mario, 2006] Gispert, A. D. and Mario, J. B. (2006). Statistical machine translation without parallel corpus: bridging through Spanish. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation*, pages 65–68.

[Heafield, 2011] Heafield, K. (2011). Kenlm: Faster and smaller language model queries. In *In Proc. of the Sixth Workshop on Statistical Machine Translation*.

[Huck and Ney, 2012] Huck, M. and Ney, H. (2012). Pivot lightly-supervised training for statistical machine translation. In *Conference of the Association for Machine Translation in the Americas 2012*, San Diego, California, USA.

[Kholy and Habash, 2013] Kholy, A. E. and Habash, N. (2013). Language Independent Connectivity Strength Features for Pivot Translation .

[Kholy et al., 2013] Kholy, A. E., Habash, N., Leusch, G., Matusov, E., and Sawaf, H. (2013). Selective combination of pivot and direct statistical machine translation models. In *International Joint Conference on Natural Language Processing*, Stroudsburg, PA, USA. Association of Computational Linguistics.

[Koehn et al., 2007] Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A., and Herbst, E. (2007). Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, ACL '07, pages 177–180, Stroudsburg, PA, USA. Association for Computational Linguistics.

[Lewis et al., 2011] Lewis, W., Munro, R., and Vogel, S. (2011). Crisis mt: Developing a cookbook for mt in crisis situations. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*. Association for Computational Linguistics.

[Nakov and Ng, 2012] Nakov, P. and Ng, H. T. (2012). Improving statistical machine translation for a resource-poor language using related resource-rich languages. *J. Artif. Int. Res.*, 44(1):179–222.

[Och and Ney, 2003] Och, F. J. and Ney, H. (2003). A systematic comparison of various statistical alignment models. *Comput. Linguist.*, 29(1):19–51.

[Sennrich, 2012] Sennrich, R. (2012). Perplexity minimization for translation model domain adaptation in statistical machine translation. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, EACL '12, pages 539–549, Stroudsburg, PA, USA. Association for Computational Linguistics.

[Stolcke, 2002] Stolcke, A. (2002). Srilm - an extensible language modeling toolkit. In *INTERSPEECH*.

[Utiyama and Isahara, 2007] Utiyama, M. and Isahara, H. (2007). A comparison of pivot methods for phrase-based statistical machine translation. In *in Proceedings of the conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics (NAACL-HLT*, pages 484–491.

[Wang et al., 2012] Wang, P., Nakov, P., and Tou Ng, H. (2012). Source language adaptation for resource-poor machine translation. In *EMNLP-CoNLL*, pages 286–296.

[Wu and Wang, 2007] Wu, H. and Wang, H. (2007). Pivot language approach for phrase-based statistical machine translation. In *Proceedings of the 45th Annual Meeting of Computational Linguistics*, Stroudsburg, PA, USA. Association of Computational Linguistics.

[Zhu et al., 2013] Zhu, X., He, Z., Wu, H., Wang, H., Zhu, C., and Zhao, T. (2013). Improving pivot-based statistical machine translation using random walk. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 524–534, Seattle, Washington, USA. Association for Computational Linguistics.