Utilizing Lexical Similarity between Related, Low-resource Languages for Pivot-based SMT

Anoop Kunchukuttan, Maulik Shah
Pradyot Prakash, Pushpak Bhattacharyya
Department of Computer Science and Engineering
Indian Institute of Technology Bombay
{anoopk,maulik.shah,pradyot,pb}@cse.iitb.ac.in

Abstract

We investigate pivot-based translation between related languages in a low resource, phrase-based SMT setting. We show that a subword-level pivot-based SMT model using a related pivot language is substantially better than word and morpheme-level pivot models. It is also highly competitive with the best direct translation model, which is encouraging as no direct source-target training corpus is used. We also show that combining multiple related language pivot models can rival a direct translation model. Thus, *the use of subwords as translation units coupled with multiple related pivot languages can compensate for the lack of a direct parallel corpus*.

1 Introduction

Related languages are those that exhibit lexical and structural similarities on account of sharing a **common ancestry** or being in **contact for a long period of time** (Bhattacharyya et al., 2016). Machine Translation between related languages is a major requirement since there is substantial government, commercial and cultural communication among people speaking related languages *e.g.*, Europe, India and South-East Asia. These constitute some of the most widely spoken languages in the world, but many of these language pairs have few or no parallel corpora. We address the scenario when no direct corpus exists between related source and target languages, but they share limited parallel corpora with a third related language.

Modelling **lexical similarity** among related languages is the key to building good-quality SMT systems with limited parallel corpora. *Lexical similarity* means that the languages share many words with similar form (spelling and pronunciation) and meaning *viz.* cognates, lateral borrowings or loan words from other languages *e.g.*, `blindness` is `andhapana` in Hindi, `aandhaLepaNaa` in Marathi.

For translation, lexical similarity can be utilized by transliteration of untranslated words while decoding (Durrani et al., 2010) or post-processing (Nakov and Tiedemann, 2012; Kunchukuttan et al., 2014). An alternative approach involves the use of subwords as basic translation units. Subword units like character (Vilar et al., 2007; Tiedemann, 2009), orthographic syllables (Kunchukuttan and Bhattacharyya, 2016b) and byte pair encoded units (Kunchukuttan and Bhattacharyya, 2017) have been used with varying degrees of success.

On the other hand, if no parallel corpus is available between two languages, pivot-based SMT (Gispert and Marino, 2006; Utiyama and Isahara, 2007) provides a systematic way of using an intermediate language, called the *pivot language*, to build the source-target translation system. The pivot approach makes no assumptions about source, pivot, and target language relatedness.

Our work **brings together subword-level translation and pivot-based SMT in low resource scenarios**. We refer to orthographic syllables and byte pair encoded units as subwords. We show that using a pivot language related to both the source and target languages along with subword-level translation (i) significantly outperforms morpheme and word-level pivot translation, and (ii) is very competitive with subword-level direct translation. We also show that combining multiple pivot models using different related pivot languages can rival a direct parallel corpora trained model. To the best of our knowledge, ours is the first work that shows that **a pivot system can be very competitive with a direct system (in**

Proceedings of the The 8th International Joint Conference on Natural Language Processing, pages 283–289,
Taipei, Taiwan, November 27 – December 1, 2017

the restricted case of related languages). Previous work on morpheme and word-level pivot models with multiple pivot languages have reported lower translation scores than the direct model (More et al., 2015; Dabre et al., 2015). Tiedemann (2012)'s work uses a character-level model in just one language pair of the triple (source-pivot or pivot-target) when the pivot is related to either the source or target (but not both).

2 Proposed Solution

We first train phrase-based SMT models between source-pivot (S-P) and pivot-target (P-T) language pairs using subword units, where the pivot is related to the source and target. We create a pivot translation system by combining the S-P and P-T models. If multiple pivot languages are available, linear interpolation is used to combine pivot translation models. In this section, we describe each component of our system and the design choices.

Subword translation units: We explore *orthographic syllable (OS)* and *Byte Pair Encoded unit (BPE)* as subword units.

The *orthographic syllable*, a **linguistically motivated unit**, is a sequence of one or more consonants followed by a vowel, *i.e.* a C^+V unit (*e.g. spacious* would be segmented as *spa ciou s*). Note that the vowel character sequence *iou* represents a single vowel.

On the other hand, the *BPE unit* is motivated by **statistical properties of text** and represents stable, frequent character sequences in the text (possibly linguistic units like syllables, morphemes, affixes). Given monolingual corpora, BPE units can be learnt using the Byte Pair Encoding text compression algorithm (Gage, 1994).

Both OS and BPE units are variable length units which provide appropriate context for translation between related languages. Since their vocabularies are much smaller than the morpheme and word-level models, data sparsity is not a problem. OS and BPE units have outperformed character n-gram, word and morpheme-level models for SMT between related languages (Kunchukuttan and Bhattacharyya, 2016b, 2017).

While OS units are approximate syllables, BPE units are highly frequent character sequences, some of them representing different linguistic units like syllables, morphemes and affixes. While orthographic syllabification applies to writing systems which represent vowels (alphabets and abugidas), BPE can be applied to text in any writing system.

Training subword-level models: We segment the data into subwords during pre-processing and indicate word boundaries by a boundary marker (_) as shown in the example for OS below:

word: `Childhood means simplicity .`

subword: `Chi ldhoo d _ mea ns _ si mpli ci ty _ .`

For building subword-level phrase-based models, we use (a) monotonic decoding since related languages have similar word order, (b) higher order language models (10-gram) since data sparsity is a lesser concern due to small vocabulary size (Vilar et al., 2007), and (c) word-level tuning (by post-processing the decoder output during tuning) to optimize the correct translation metric (Nakov and Tiedemann, 2012). After decoding, we regenerate words from subwords (desegmentation) by concatenating subwords between consecutive occurrences of the boundary markers.

Pivoting using related language: We use a language related to both the source and target language as the pivot language. We explore two widely used pivoting techniques: phrase-table triangulation and pipelining.

Triangulation (Utiyama and Isahara, 2007; Wu and Wang, 2007; Cohn and Lapata, 2007) "joins" the source-pivot and pivot-target subword-level phrase-tables on the common phrases in the pivot columns, generating the pivot model's phrase-table. It recomputes the probabilities in the new source-target phrase-table, after making a few independence assumptions, as shown below:

$$P(\bar{t}|\bar{s}) = \sum_{\bar{p}} P(\bar{t}|\bar{p})P(\bar{p}|\bar{s}) \quad (1)$$

where, $\bar{s}, \bar{p}$ and $\bar{t}$ are source, pivot and target phrases respectively.

In the **pipelining/transfer** approach (Utiyama and Isahara, 2007), a source sentence is first translated into the pivot language, and the pivot language translation is further translated into the target language using the S-P and P-T translation models respectively. To reduce cascading errors due to pipelining, we consider the top-k source-pivot translations in the second stage of the pipeline (an approximation to expectation over all translation candidates). We used $k = 20$ in our experiments. The translation candidates are scored

as shown below:

$$P(\mathbf{t}|\mathbf{s}) = \sum_{i=1}^{k} P(\mathbf{t}|\mathbf{p}^i)P(\mathbf{p}^i|\mathbf{s}) \quad (2)$$

where, $\mathbf{s}$, $\mathbf{p}^i$ and $\mathbf{t}$ are the source, i^{th} best source-pivot translation and target sentence respectively.

Using Multiple Pivot Languages : We use multiple pivot languages by combining triangulated models corresponding to different pivot languages. Linear interpolation is used (Bisazza et al., 2011) for model combination. Interpolation weights are assigned to each phrase-table and the feature values for each phrase pair are interpolated using these weights as shown below:

$$f^j(\bar{s},\bar{t}) = \sum_i \alpha_i f_i^j(\bar{s},\bar{t}) \quad (3)$$

$$\text{s.t} \quad \sum_i \alpha_i = 1, \quad \alpha_i \geq 0$$

where, f^j is feature j defined on the phrase pair $(\bar{s},\bar{t})$, α_i is the interpolation weight for phrase-table i. Phrase-table i corresponds to the triangulated phrase-table using language i as a pivot.

3 Experimental Setup

Languages: We experimented with multiple languages from the two major language families of the Indian subcontinent: *Indo-Aryan* branch of the Indo-European language family (Bengali, Gujarati, Hindi, Marathi, Urdu) and *Dravidian* (Malayalam, Telugu, Tamil). These languages have a substantial overlap between their vocabularies due to contact over a long period (Emeneau, 1956; Subbarao, 2012).

Dataset: We used the *Indian Language Corpora Initiative (ILCI) corpus*[1] for our experiments (Jha, 2012). The data split is as follows – **training: 44,777, tuning: 1K, test: 2K** sentences. Language models for word-level systems were trained on the target side of training corpora plus monolingual corpora from various sources [hin: 10M (Bojar et al., 2014), urd: 5M (Jawaid et al., 2014), tam: 1M (Ramasamy et al., 2012), mar: 1.8M (news websites), mal: 200K, ben: 400K, pan: 100K, guj:400K, tel: 600K (Quasthoff et al., 2006) sentences]. We used the target side of parallel corpora for morpheme, OS, BPE and character-level LMs.

System details: We trained PBSMT systems for all translation units using *Moses* (Koehn et al., 2007) with *grow-diag-final-and* heuristic for symmetrization of alignments, and Batch MIRA (Cherry and Foster, 2012) for tuning. Subword-level representation of sentences is long, hence we speed up decoding by using cube pruning with a smaller beam size (pop-limit=1000) for OS and BPE-level models. This setting has been shown to have minimal impact on translation quality (Kunchukuttan and Bhattacharyya, 2016a).

We trained 5-gram LMs with Kneser-Ney smoothing for word and morpheme-level models, and 10-gram LMs for OS, BPE, character-level models. We used the *Indic NLP library*[2] for orthographic syllabification, the *subword-nmt library*[3] for training BPE models and *Morfessor* (Virpioja et al., 2013) for morphological segmentation. These unsupervised morphological analyzers for Indian languages, described in Kunchukuttan et al. (2014), are trained on the ILCI corpus and the Leipzig corpus (Quasthoff et al., 2006). The BPE vocabulary size was chosen to match OS vocab size. We use *tmtriangulate*[4] for phrase-table triangulation and *combine-ptables* (Bisazza et al., 2011) for linear interpolation of phrase-tables.

Evaluation: The primary evaluation metric is word-level BLEU (Papineni et al., 2002). We also report LeBLEU (Virpioja and Grönroos, 2015) scores in the appendix. LeBLEU is a variant of BLEU that does soft-matching of words and has been shown to be better for morphologically rich languages. We use bootstrap resampling for testing statistical significance (Koehn, 2004).

4 Results and Discussion

In this section, we discuss and analyze the results of our experiments.

4.1 Comparison of Different Subword Units

Table 1 compares pivot-based SMT systems built with different units. We observe that *the OS and BPE-level pivot models significantly outperform word, morpheme and character-level pivot models* (average improvements above 55% over word-level and 14% over morpheme-level). The greatest improvement is observed when the source and target languages belong to different families (though they have a contact relationship), showing that subword-level models can utilize the lex-

[1] available on request from tdil-dc.in

[2] http://anoopkunchukuttan.github.io/indic_nlp_library

[3] https://github.com/rsennrich/subword-nmt

[4] github.com/tamhd/MultiMT

Lang Triple	Word	Morph	BPE	OS	Char
mar-guj-hin	30.23	36.49	39.05	**39.81**†	34.32
mar-hin-ben	16.63	21.04	22.46	**22.92**†	17.00
mal-tel-tam	4.55	6.19	**7.69**†	7.19	3.51
tel-mal-tam	5.13	8.29	**9.84**†	8.39	4.26
hin-tel-mal	5.29	8.32	9.57	**9.67**	6.24
mal-tel-hin	10.03	13.06	**17.68**	17.26	9.12
mal-urd-hin	7.70	11.29	**16.40**	NA	7.46
urd-hin-mal	5.58	6.64	**7.58**	NA	4.07
average % change w.r.t (BPE,OS)	*(+66,+57)%*	*(+21,+14)%*			*(+81,+66)%*

Table 1: Comparison of triangulation for various translation units (BLEU). Lang triple refers to the source-pivot-target languages. Scores in **bold** indicate highest values for the language triple. † means difference between OS and BPE scores is statistically significant ($p < 0.05$). NA: OS segmentations cannot be done for Urdu. The last row shows average change in BLEU scores for word, morpheme and character-level model compared to the OS and BPE-level models.

Lang Triple	BPE		OS	
	pip	*tri*	*pip*	*tri*
mar-guj-hin	38.25	**39.05**†	38.11	**39.81**†
mar-hin-ben	**22.50**	22.46	22.83	**22.92**
mal-tel-tam	**7.84**	7.69	6.94	**7.19**
tel-mal-tam	8.47	**9.84**†	7.96	**8.39**†
hin-tel-mal	9.31	**9.57**	9.31	**9.67**†
mal-tel-hin	17.39	**17.68**	16.96	**17.26**
mal-urd-hin	**16.93**†	16.40	NA	NA
urd-hin-mal	**8.83**†	7.58	NA	NA

Table 2: Comparison of pipelining (*pip*) and triangulation (*tri*) approaches for OS and BPE (BLEU). † means difference between *pip* and *tri* is statistically significant ($p < 0.05$)

ical similarity between languages. Translation between agglutinative Dravidian languages also shows a major improvement. The OS and BPE models are comparable in performance. However, unlike OS, BPE segmentation can also be applied to translations involving languages with non-alphabetic scripts (like Urdu) and show significant improvement in those cases also. Evaluation with LeBLEU (Virpioja and Grönroos, 2015), a metric suited for morphologically rich languages, shows similar trends (results in Appendix A). For brevity, we report BLEU scores in subsequent experiments.

Subword-level models outperform other units for the pipelining approach to pivoting too. Triangulation and pipelining approaches are comparable for BPE and OS models (See Table 2). Hence,

Lang Triple	Word	Morph	BPE	OS	Char
mar-guj-hin	0.64	1.39	1.74	2.33	3.04
mar-hin-ben	0.58	1.36	1.71	2.6	3.47
mal-tel-tam	0.61	2.32	3.27	4.19	2.58
tel-mal-tam	0.75	2.82	4.09	2.76	2.42
hin-tel-mal	0.56	2.08	2.86	2.97	2.25
mal-tel-hin	0.55	2.28	2.85	3.56	2.57
mal-urd-hin	0.25	1.16	1.84	NA	2.05
urd-hin-mal	0.42	0.79	1.62	NA	1.47

Table 3: Ratio of triangulated to component phrase-table sizes. We use the size of larger of the component phrase-tables to compute the ratio.

Lang Triple	Pivot	Direct				Pivot
	BPE	BPE	Word	Morph	OS	OS
mar-guj-hin	39.05	43.19	38.87	42.81	43.69	39.81
mar-hin-ben	22.46	24.13	21.13	23.96	23.53	22.92
mal-tel-tam	7.69	8.67	6.38	7.61	7.84	7.19
tel-mal-tam	9.84	11.61	9.58	10.61	10.52	8.39
hin-tel-mal	9.57	10.73	8.55	9.23	10.46	9.67
mal-tel-hin	17.68	20.54	15.18	17.08	18.44	17.26
mal-urd-hin	16.4	20.54	15.18	17.08	18.44	NA
urd-hin-mal	7.58	8.44	6.49	7.05	NA	NA

Table 4: Pivot *vs.* Direct translation (BLEU)

we report results for only the triangulation approach in subsequent experiments.

4.2 Why is Subword-level Pivot SMT better?

Subword-level pivot models are better than other units for two reasons. One, *the underlying S-P and P-T translation models are better* (*e.g.* 16% and 3% average improvement over word and morpheme-level models for OS). Two, the triangulation process involves an *inner join* on pivot language phrases common to the S-P and P-T phrase-tables. This causes data sparsity issues due to the large word and morpheme phrase-table vocabulary (Dabre et al., 2015; More et al., 2015). On the other hand, *the OS and BPE phrase-table vocabularies are smaller, so the impact of sparsity is limited*. This effect can be observed by comparing the ratio of the triangulated phrase-table (S-P-T) with the component phrase-tables (S-P and P-T). The size of the triangulated phrase-table is less than the size of the underlying tables at the word-level, while it increases by a few multiples for subword-level models (see Table 3).

4.3 Comparison of Pivot & Direct Models

We compared the OS and BPE-level models with direct models trained on different translation units

Model	mar-ben		mal-hin	
	OS	BPE	OS	BPE
best pivot	22.92 (*hin*)	22.46 (*hin*)	17.52 (*tel*)	18.47 (*guj*)
direct	23.53	24.13	18.44	20.54
all pivots	23.69	23.20†	19.12†	20.28
direct+all pivots	24.41‡	**24.49**‡	19.44‡	**20.93**‡

Table 5: Combination of multiple pivots (BLEU). **Pivots used for** (i) mar-ben: guj, hin, pan (ii) mal-hin: tel, mar, guj. Best pivot language indicated in brackets. Statistically significant difference from *direct* is indicated for: *all pivots*(†) and *direct+all pivots*(‡) ($p < 0.05$).

(see Table 4). These subword-level pivot models outperform word-level direct models by 5-10%, which is encouraging. Remarkably, the subword-level pivot model is competitive with the morpheme-level models (about 95% of the morpheme BLEU score). The subword-level pivot models are competitive with the best performing direct counterparts too (about 90% of the direct system BLEU score). To put this fact in perspective, the BLEU scores of morpheme and word-level pivot systems are far below their corresponding direct systems (about 15% and 35% respectively). *These observations strongly suggest that pivoting at the subword-level can better reconstruct the direct translation system than word and morpheme-level pivot systems.*

4.4 Multiple Pivot Languages

We investigated if combining multiple pivot translation models can be a substitute for the direct translation model. *Direct model* refers to translation system built using the source-target parallel corpus. Using linear interpolation with *equal weights*, we combined pivot translation models trained on different pivot languages. Table 5 shows that *the combination of multiple pivot language models outperformed the individual pivot models, and is comparable to the direct translation system*. Previous studies have shown that word and morpheme-level multiple pivot systems were not competitive with the direct system, possibly due to the effect of sparsity on triangulation (More et al., 2015; Dabre et al., 2015). Our results show that once the ill-effects of data sparsity are reduced due to the use of subword models, multiple pivot languages can maximize translation performance because: (i) they bring in more translation options, and (ii) they improve the estimates of feature values with evidence from multiple languages. Linear interpolation of the direct system with all the pivot systems with equal interpolation weights also benefitted the translation system. Thus, *multilinguality helps overcome the lack of parallel corpora between the two languages*.

Lang Triple	Pivot			Direct		
	Morph	OS	BPE	Morph	OS	BPE
hin-tel-mal	4.72	5.96	6.00	5.99	6.26	6.37
mal-tel-hin	8.29	11.33	10.94	11.12	13.32	14.45
mal-tel-tam	4.41	5.82	5.85	5.84	5.88	6.75

Table 6: Cross domain translation (BLEU)

4.5 Cross-Domain Translation

We also investigated if the OS and BPE-level pivot models are robust to domain change by evaluating the pivot and direct translation models trained on tourism and health domains on an agriculture domain test set of 1000 sentences (results in Table 6). For cross-domain translation too, the subword-level pivot models outperform morpheme-level pivot models and are comparable to a direct morpheme-level model. The OS and BPE-level models systems experience much lesser drop in BLEU scores *vis-a-vis* direct models, in contrast to the morpheme-level models. Since morpheme-level pivot models encounter unknown vocabulary in a new domain, they are less resistant to domain change than subword-level models.

5 Conclusion and Future Work

We show that pivot translation between related languages can be competitive with direct translation if a *related pivot language* is used and *subword units* are used to represent the data. Subword units make pivot models competitive by (i) utilizing lexical similarity to improve the underlying S-P and P-T translation models, and (ii) reducing losses in pivoting (owing to small vocabulary). Combining multiple related pivot models can further improve translation. Our SMT pivot translation work is useful for low resource settings, while current NMT systems require large-scale resources for good performance. We plan to explore multilingual NMT in conjunction with subword representation between related languages with a focus on reducing corpus requirements. Currently, these ideas are being actively explored in the research community in a general setting.

References

Pushpak Bhattacharyya, Mitesh Khapra, and Anoop Kunchukuttan. 2016. Statistical Machine Translation Between Related Languages. www.cfilt.iitb.ac.in/publications/naacl-2016-tutorial.pdf. Annual Conference of the North American Chapter of the Association for Computational Linguistics: Tutorials.

Arianna Bisazza, Nick Ruiz, Marcello Federico, and Bruno Kessler. 2011. Fill-up versus interpolation methods for phrase-based SMT adaptation. In *International Workshop on Spoken Language Translation*.

Ondřej Bojar, Vojtěch Diatka, Pavel Rychlý, Pavel Straňák, Vít Suchomel, Aleš Tamchyna, and Daniel Zeman. 2014. HindEnCorp – Hindi-English and Hindi-only Corpus for Machine Translation. In *Proceedings of the 9th International Conference on Language Resources and Evaluation*.

Colin Cherry and George Foster. 2012. Batch tuning strategies for Statistical Machine Translation. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

Trevor Cohn and Mirella Lapata. 2007. Machine translation by triangulation: Making effective use of multi-parallel corpora. In *Proceedings of the Annual Meeting on Association for Computational Linguistics*.

Raj Dabre, Fabien Cromieres, Sadao Kurohashi, and Pushpak Bhattacharyya. 2015. Leveraging Small Multilingual Corpora for SMT Using Many Pivot Languages. In *Annual Conference of the North American Chapter of the Association for Computational Linguistics*.

Nadir Durrani, Hassan Sajjad, Alexander Fraser, and Helmut Schmid. 2010. Hindi-to-Urdu machine translation through transliteration. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*.

Murray B Emeneau. 1956. India as a lingustic area. *Language* .

Philip Gage. 1994. A New Algorithm for Data Compression. *The C Users Journal* .

Adri'a De Gispert and Jose B Marino. 2006. Catalan-English Statistical Machine Translation without parallel corpus: Bridging through Spanish. In *In Proc. of 5th International Conference on Language Resources and Evaluation (LREC)*.

Bushra Jawaid, Amir Kamran, and Ondřej Bojar. 2014. Urdu monolingual corpus. LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics, Charles University in Prague. http://hdl.handle.net/11858/00-097C-0000-0023-65A9-5.

Girish Nath Jha. 2012. The TDIL program and the Indian Language Corpora Initiative. In *Language Resources and Evaluation Conference*.

Philipp Koehn. 2004. Statistical Significance Tests for Machine Translation Evaluation. In *Conference on Empirical Methods in Natural Language Processing*.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, et al. 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*.

Anoop Kunchukuttan and Pushpak Bhattacharyya. 2016a. Faster decoding for subword level Phrase-based SMT between related languages. In *Third Workshop on NLP for Similar Languages, Varieties and Dialects*.

Anoop Kunchukuttan and Pushpak Bhattacharyya. 2016b. Orthographic Syllable as basic unit for SMT between Related Languages. In *Empirical Methods in Natural Language Processing*.

Anoop Kunchukuttan and Pushpak Bhattacharyya. 2017. Learning variable length units for SMT between related languages via Byte Pair Encoding. In *First Workshop on Subword and Character level models in NLP*.

Anoop Kunchukuttan, Ratish Pudupully, Rajen Chatterjee, Abhijit Mishra, and Pushpak Bhattacharyya. 2014. The IIT Bombay SMT System for ICON 2014 Tools Contest. In *NLP Tools Contest at ICON 2014*.

Rohit More, Anoop Kunchukuttan, Raj Dabre, and Pushpak Bhattacharyya. 2015. Augmenting Pivot based SMT with word segmentation. In *International Conference on Natural Language Processing*.

Preslav Nakov and Jörg Tiedemann. 2012. Combining word-level and character-level models for machine translation between closely-related languages. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2*.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: A method for automatic evaluation of machine translation. In *Association for Computational Linguistics*.

Uwe Quasthoff, Matthias Richter, and Christian Biemann. 2006. Corpus portal for search in monolingual corpora. In *Proceedings of the fifth international conference on language resources and evaluation*.

Loganathan Ramasamy, Ondřej Bojar, and Zdeněk Žabokrtský. 2012. Morphological Processing for English-Tamil Statistical Machine Translation. In *Proceedings of the Workshop on Machine Translation and Parsing in Indian Languages*.

Karumuri Subbarao. 2012. *South Asian Languages: A Syntactic Typology*. Cambridge University Press.

Jörg Tiedemann. 2009. Character-based PBSMT for closely related languages. In *Proceedings of the 13th Conference of the European Association for Machine Translation*.

Jörg Tiedemann. 2012. Character-based pivot translation for under-resourced languages and domains. In *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics*.

Masao Utiyama and Hitoshi Isahara. 2007. A Comparison of Pivot Methods for Phrase-Based Statistical Machine Translation. In *Annual Conference of the North American Chapter of the Association for Computational Linguistics*.

David Vilar, Jan-T Peter, and Hermann Ney. 2007. Can we translate letters? In *Proceedings of the Second Workshop on Statistical Machine Translation*.

Sami Virpioja and Stig-Arne Grönroos. 2015. LeBLEU: N-gram-based Translation Evaluation Score for Morphologically Complex Languages. In *Workshop on Machine Translation*.

Sami Virpioja, Peter Smit, Stig-Arne Grönroos, Mikko Kurimo, et al. 2013. Morfessor 2.0: Python implementation and extensions for Morfessor Baseline. Technical report, Aalto University.

Hua Wu and Haifeng Wang. 2007. Pivot language approach for phrase-based Statistical Machine Translation. *Machine Translation* .

A LeBLEU Scores

Table 7 shows LeBLEU scores for the experiments using phrase-triangulation. We observe that the same trends hold as with BLEU scores.

Lang Triple	Word	Morph	BPE	OS	Char
mar-guj-hin	0.692	0.725	0.737	**0.747**	0.713
mar-hin-ben	0.505	0.616	0.638	**0.646**	0.577
mal-tel-tam	0.247	0.364	**0.426**	0.407	0.213
tel-mal-tam	0.242	0.433	**0.485**	0.441	0.392
hin-tel-mal	0.291	0.376	0.420	**0.432**	0.306
mal-tel-hin	0.247	0.364	**0.426**	0.404	0.213
mal-urd-hin	0.328	0.436	**0.501**	NA	0.377
urd-hin-mal	0.313	0.353	**0.420**	NA	0.323
average % change w.r.t (BPE,OS)	*(+51,+49)%*	*(+12,+8)%*			*(+42,+42)%*

(a) Comparison of phrase-triangulation for various subwords

Lang Triple	Pivot	Direct				Pivot
	BPE	BPE	Word	Morph	OS	OS
mar-guj-hin	0.737	0.766	0.746	0.767	0.766	0.747
mar-hin-ben	0.638	0.653	0.568	0.645	0.656	0.646
mal-tel-tam	0.426	0.465	0.314	0.409	0.447	0.407
tel-mal-tam	0.485	0.530	0.410	0.511	0.534	0.441
hin-tel-mal	0.420	0.468	0.393	0.436	0.477	0.432
mal-tel-hin	0.426	0.565	0.460	0.528	0.551	0.404
mal-urd-hin	0.501	0.565	0.460	0.528	0.551	NA
urd-hin-mal	0.420	0.416	0.350	0.379	NA	NA

(b) Pivot *vs.* direct translation

Table 7: LeBLEU Scores