

# Augmenting Performance of SMT Models by Deploying Fine Tokenization of the Text and Part-of-Speech Tag

Abraham Tesso Nedjo<sup>1</sup> & Huang Degen<sup>1</sup>

<sup>1</sup> School of Computer Science and Technology, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian, Liaoning Province, 116024, China

Correspondence: Abraham Tesso Nedjo, #2 Linggong Road, Ganjingzi District, Dalian University of Technology, School of Computer Science and Technology, Dalian 116024, Liaoning Province, China. E-mail: abratesso@gmail.com

Received: November 11, 2014      Accepted: November 27, 2014      Online Published: January 28, 2015

doi:10.5539/cis.v8n1p119      URL: <http://dx.doi.org/10.5539/cis.v8n1p119>

*The research is financed by the National Natural Science Foundation of China. (Project No 61173100, 61173101, and 61272375).*

## Abstract

This paper presents our study of exploiting the languages' word class information augmented with some rule-based processing for phrase-based Statistical Machine Translation (SMT). In statistical machine translation, estimating word-to-word alignment probabilities for the translation model can be difficult due to the problem of sparse data: most words in a given corpus occur at most a handful of times. With a highly inflected language such as Oromo, this problem can be particularly severe. In addition, there is variant nature or use of different symbols for 'hudhaa' (the diacritical marker) in Oromo language which intrudes another severe data sparsity problem. In this work, we show that using fine tokenization of words considering intra-word behavior of words consisting *hudhaa*, and POS tag to modify the Oromo input and see how it improves Oromo-English machine translation system. The models were trained on a very small parallel corpus of data set (usually unacceptable for normal SMT system) and also the quality of the parallel corpus both in translation and spelling errors were not so good. Yet, our final system achieves a BLEU score of 2.88, as compared to 2.56 for the baseline system.

**Keywords:** fine tokenization, Oromo language, POS, statistical machine translation

## 1. Introduction

Several phrase-based Statistical Machine Translation (SMT) models such as that of Och and Ney (1999, 2004), Marcu and Wang (2002), and Koehn et al. (2003) have achieved a promising performance. These phrase based models have shown a number of advantages when compared with the pioneer models of Brown et al. (1993) which were actually based on word level expression recognition and local restructuring. The advantages gained from such phrase-based SMT are the result of moving from words to phrases as the basic unit of translation.

There is no question that these phrase-based models are much better statistical translation methods. But, although these systems have been successful, there are some potential drawbacks when it comes to modeling word-order differences between languages, because of the fact that phrase-based systems does not make use of little or only indirect use of the languages' syntactic information. That is the reason why they are still called "non-linguistic", as Nguyen and Shimazu (2006) quotes. That means, in phrase-based systems, the treatment of tokens are just words and phrases can be any sequence of tokens which may not necessarily be phrases in any sense. And the reordering models of these systems are based only on movement of distances, but not on the phrase content as Och and Ney (2004) and Koehn et al. (2003) argue. Another drawback of these phrase-based SMT without considering the language information is the sparse data problem, since obtaining huge parallel corpora is difficult and expensive for unexplored languages like Oromo-English.

In literal phrase-based SMT, the problem prevails since the inflected forms of the same word are often treated as different words. This problem is more serious and apparent when one or both of the source and target language is morphologically rich and an inflectional one. In our case, Oromo is an inflectional language and scarce resource which needs special treatment to get a better phrase-based SMT system. In addition, the variant nature or use of

different symbols for ‘hudhaa’ (the diacritical marker) in Oromo language intrudes another severe data sparsity problem. In this paper, we describe our approach to improve statistical machine translation by augmenting linguistic annotation of the language together with deeper in-word processing tokenization of Oromo in order to address the aforementioned data sparsity problems of phrase-based SMT.

The rest of this paper is structured as follows. Section II describes review of related works and frameworks that can leverage the task of MT in such scarce resources, to emphasize the importance and motivation of our research; and briefly reviews about the importance and how it figures prominently in text process. Section III presents an elucidation of Oromo Language features intended to further emphasize the problem of the research. In Section IV the experimental setups undergone for the research was discussed. Section V is the part in which we present the experimental results, discussion of results, and evaluation. Finally, Section VI concludes our paper and suggests avenues for future work.

## 2. Related Works

Researches in MT involving Oromo language is not very common, even not a single one has been tried to this date. The main reason perhaps is the lack of parallel corpora. Nevertheless there have been many models for other languages, especially translation between English and other languages. As far as our information is concerned, there has never been an effort for building Oromo and any other language parallel corpus. Therefore, one of our goals in this work is to perform experiments with corpus that we collected by ourselves from various web sources for the research.

There have been attempts on using different approaches in development of Statistical Machine Translation. Goldwater et al. (2005) and Durgar et al. (2006) presented some very preliminary results for problems in developing a statistical machine translation system mainly dealing with different morphological concepts. Applying such language contexts, like morphological processing to SMT is not a new, but has been done at different levels. The idea goes back to Lee (2004) and Isbihani et al. (2006) for Arabic-English translation; Nießen and Ney (2004) for German-English translation; Ramanathan et al. (2008, 2009), for their first attempt on English-Hindi translation. Nguyen and Shimazu (2006) also analyzed improvement of phrased-based SMT with morpho-syntactic analysis and transformation for English-Vietnamese language.

In SMT, correspondences between the words in the source and target language are learned from parallel corpora, and often little or no linguistic knowledge incorporated to structure the underlying models (Adam Lopez, 2008). In particular, existing statistical systems for machine translation often treat different inflected forms of the same lemma as if they were independent of one another. But the bilingual training data can be better exploited by explicitly taking into account the interdependences of related inflected forms. To improve translation quality in framework with scarce resources, Nießen et al. (2004) proposed the construction of hierarchical lexical models on the bases of equivalent classes of words. Some other publications like Al-Onaizan, et al. (2000) have also dealt with the problems of translation with less-resources language pairs. The report was on an experiment involving Tetun-to-English translation by different groups, including one using SMT.

## 3. Oromo Language Features

The important phenomena in Oromo language is morphology. Oromo is a highly inflected language. Even though both English and Oromo are inflected languages, Oromo is more inflectional one. Most Oromo inflected word forms can be translated into an English phrase.

In addition to these morphological, inflectional and word order problems, Oromo language has another difficulty that creates data sparsity. This is about the variation of symbols used to represent hudhaa in Oromo texts. Hudhaa is diacritical marker or glottal symbol in Oromo. Nedjo et al. (2014a) presented the problem of tokenization in natural language processing of Oromo language to find tokens in a text because of hudhaa or the Oromo diacritical marker. Nedjo emphasized that hudhaa intrudes data sparsity problem. Because, some people just use apostrophe (‘) or use of the right-quote character (’) or left-quote character (‘), grave character (`), or acute accent or just symbols to indicate the glottal somewhere in the middle of words or as component of words. This paper investigates on data sparsity problem of Oromo in development of MT, among the many other difficulties mentioned above, by using Oromo tokenizer and Oromo POS tagger from Nedjo (2014 a&b) to train the model.

## 4. Experimental Setups

### 4.1 Architecture of the System

Deploying a clear architecture forms the central role in making up statistical machine translation, as it is true for any system development. As depicted on Figure 1, the architecture shows the interfacing and coordination of all the translation system components like Language Model (LM), Translation Model <sup>TM</sup>, Pre-processing and Decoder (Modelling) which are needed for development of SMT. Language Model is prepared from the target language and the Decoder gives the probability of target sentence given the source sentence.

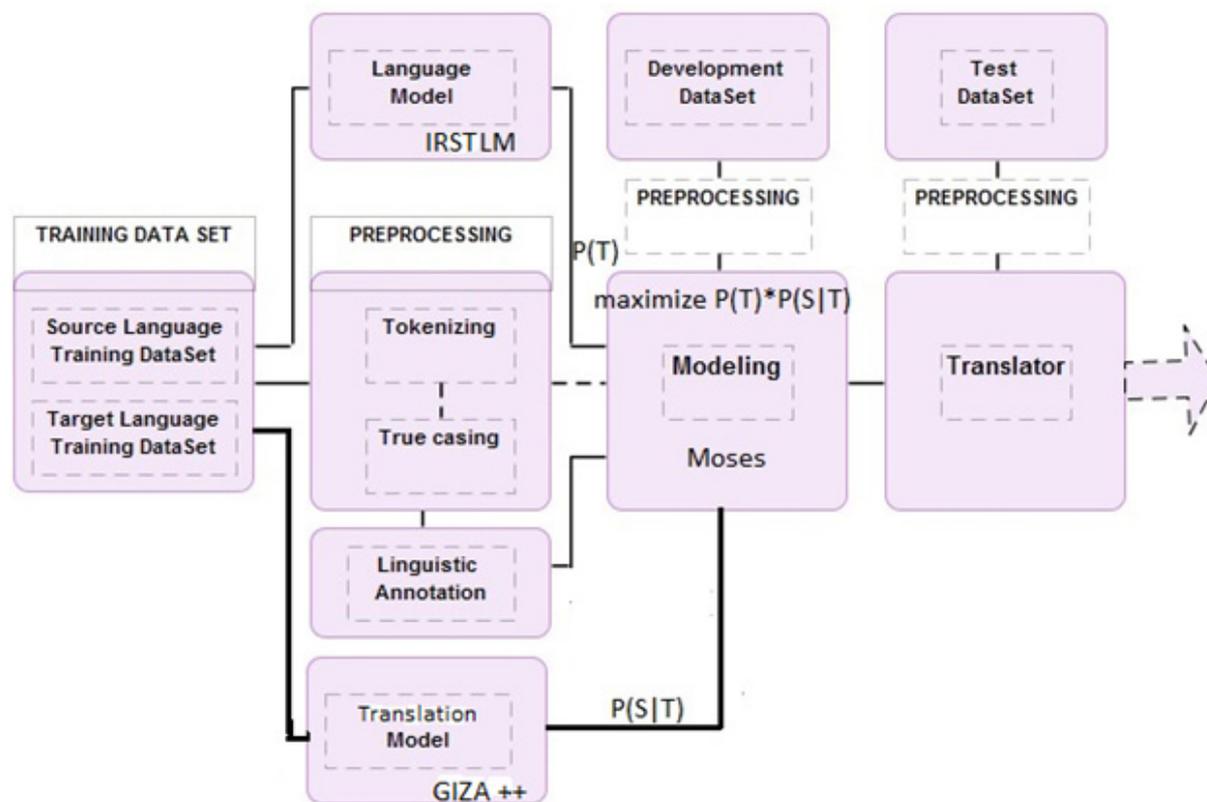


Figure 1. Detail Architecture of Oromo-English SMT System

#### 4.1.1 Language Model Module

The Language Model (LM) gives the probability of a sentence which depends on the probability of the individual words. This language model is used to ensure the fluency of the output. So it is built with the target language (i.e English in this case). If 'T' is the target language, the LM computes  $P(T)$  and feed this into the decoder software. IRSTLM of M. Federico et.al. (2008) was used for language modeling. IRSTLM is an open-source language modeling toolkit and is hosted on sourceforge. The IRSTLM builds an appropriate 3-gram language model, removing singletons, smoothing with improved Kneser-Ney, and adding sentence boundary symbols:

#### 4.1.2 Translation Model Module

The Translation Model <sup>TM</sup> computes the probability of source sentence 'S', for a given target sentence 'T'. Mathematically, the probability being computed by TM is given as,  $P(S|T)$  and the output of the TM is fed into the Moses decoder. The translation model for this system has been developed using open source software, GIZA++ of Och and Ney (2003). GIZA++ is also free and open source software hosted at Google Code (Note 2), and a mirror of the original documentation can be found here (Note 3).

#### 4.1.3 The Preprocessing Module

The preprocessing module renders preprocessing solutions for Oromo-English SMT. Oromo is a morphologically rich language, which poses some problems for statistical machine translation approaches. Much research in statistical machine translation has shown the importance of POS tagging and morphological

preprocessing (aka, fine word-level tokenization) on translation quality. The common wisdom in the field is that such preprocessing helps, especially for morphologically rich languages, such as Oromo language, because it reduces model sparsity and increases source-target symmetry (particularly when the target is morphologically poor, as in English). For the target language—English, the preprocessing scripts accompanying the decoder software—Moses, have been directly used. For the source language—Oromo, the Oromo tokenizer and Oromo POS tagging, presented in Nedjo et al. (2014) and Nedjo et al. (2014), respectively, and some scripts from Moses have been used for implementation of this Oromo-English translation system preprocessing.

#### 4.1.4 The Decoder or Modelling Module

The modeling decoder maximizes the probability of the generated sentence. It makes use of the *argmax()* function to maximize the probability. Moses software which is freely available under open source licenses was used for modeling the decoder. This Moses software is compatible with IRSTLM and GIZA++ tools and accepts the source language text as input and generates the target language text. Therefore, the job of the decoder is to find the highest scoring sentence in the target language—English (according to the translation model) corresponding to a given source language—Oromo sentence. The probability files are accepted from the LM and the TM. Moses decoder can be set in interactive mode for translation (Koehn 2007). The software is free and obtained from here (Note 1). It is being developed as a reference implementation of state-of-the-art methods in statistical machine translation, as stated in the *Statistical Machine Translation* text book of Koehn (2010) and Moses manual.

#### 4.2 Data Sets

We conducted the experiments on parallel corpus of Oromo-English training, development, and test data sets. The statistics for these sets are shown in Table 1. Almost all of the data used were taken from the Holy Bible book and made available for the research, regardless of the quality of the translation for the reasons of the low quality as well as low quantity of the data.

Table 1. Data sets

		LANGUAGE	
		OROMO	ENGLISH
Training Sets	Sentences	13633	13633
	Words	230995	293510
Development Sets	Sentences	1001	1001
	Words	16433	21161
Test Sets	Sentences	360	360
	Words	5741	7505
Total Sentences		14994	14994
Total Words		253169	322176

#### 4.3 Training the Translator

To ensure the fluency of the output of the statistical machine translation system, language model was trained and built with the target language—English, and used to train the decoder. The language modeling software, IRSTLM, which had already been installed, was used to train the target language data set.

To train the translation model, which is the main event in developing of statistical machine translation, we ran word alignment (using GIZA++), for phrase extraction and scoring, create lexicalized reordering tables and create the Moses configuration file. First an appropriate directory was created. Then the training command was executed.

#### 4.4 Tuning for Quality and Speed

The parameters of the model were tuned by a small amount of parallel data, separate from the training data set, kept aside for development. The Moses software makes use of weights given in *moses.ini* file to translate texts. Default values are generated for these weights by the system during its training. So these weights are present in the configuration file of Moses, *moses.ini*. By tuning we set the model parameters in *moses.ini* file to improve the quality of the translation.

After taking considerable amount of time, the end result of tuning was *moses.ini* file with trained weights. The model was used to translate Oromo sentences into English by running the Moses decoder executable file. But it

was too slow to star up and run the translator. It was very slow process and memory consuming task. So, binarized phrase-table and lexicalized reordering models were used as prototype machine translation system of this research.

#### 4.5 Testing the Prototype

Having done with the binarized phrase-table and the lexicalized reordering table, we used another parallel data set (the test set) distinct from the ones we've used so far to evaluate the quality of the translated text, which helped to know how good the translation system is. The test set was similar in nature as the development-set used for development or tuning. It is taken from identical sources. In the same way as it was done for the training and development test, this test set was tokenized and truecased. Then the model that we've trained was filtered for this test set (in order to retain only the entries needed to translate the test set ) so that the translation is a lot faster. Automatic test on the translator was done by first translating the test set and then running the BLEU script on it. These test results were taken from both models, produced a BLEU scores that were used to compare the performance improvement.

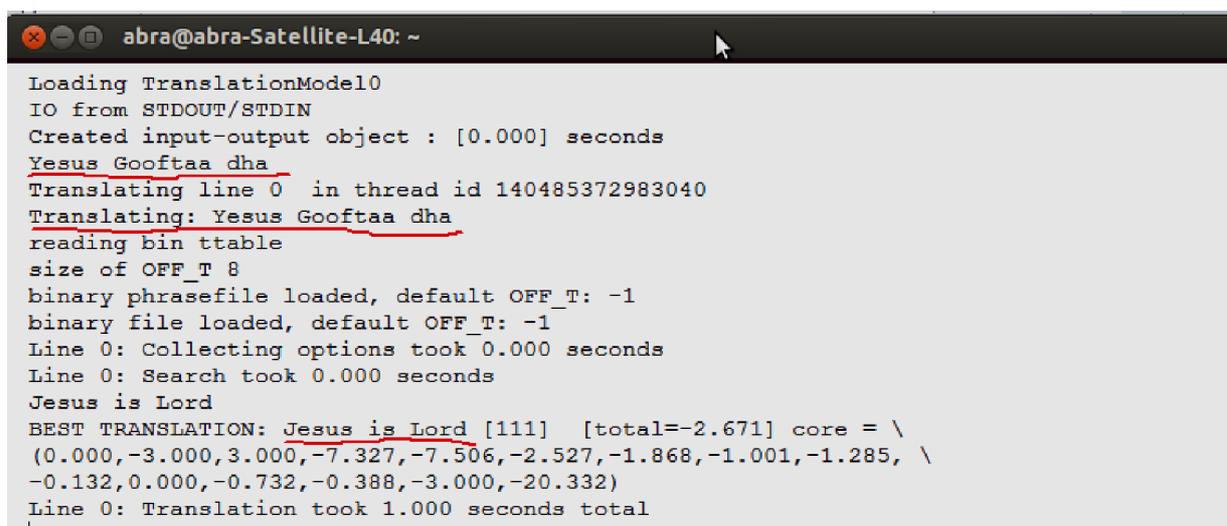
As mentioned earlier, this research was conducted on the same corpus for two different settings of experiment. The same data sets (same training set, development set, and test set) were used for the two experiments to compare the results. The first experiment was done to develop baseline translation system trained to translate Oromo sentences to English sentences. The second experiment was conducted to train a translator using the Oromo tokenizer that considers intra-word preprocessing for tokenization to handle the problem of diacritical marker-*hudhaa* (discussed in Nedjo et al. (2014)) and the Oromo POS tagger (discussed in Nedjo et al. (2014)). The results of the experiments are discussed in the next topic.

### 5. Results of the Experiment: Discussions and Evaluation

The proposed Oromo-English statistical machine translation system accepts Oromo language sentences as input and gives the corresponding English sentences as output. Then Oromo sentence “*Yesus Gooftaa dha*”, was supplied from the standard input. The model accepted this input in the interactive mode and produced acceptable English translation. One of the result of this translation output is shown in Figure 2.

In machine translation systems, there are two ways of evaluating performance of the model. One method is the automatic evaluation by a computer program and the other method is the manual evaluation by human translators. The evaluation reported in this paper has implemented both mechanisms for different purpose focusing on the evaluation of the output of the system, on performance or usability evaluation, in particular.

Automatic evaluation was used to quantitatively measure and clearly compare the results of the two experiments. The output of the baseline system and the new system, designed to complement performance of SMT for less-resourced language pairs by deploying word-level tokenization or fine preprocessing and POS tag of the text, is measured by BLEU. The model achieved improvement of 12.5% as shown in Table 2. The BLEU evaluation code is included in Moses decoder package and running the script on the same test data for both models gave the following results.



```

abra@abra-Satellite-L40: ~
Loading TranslationModel0
IO from STDOUT/STDIN
Created input-output object : [0.000] seconds
Yesus Gooftaa dha
Translating line 0 in thread id 140485372983040
Translating: Yesus Gooftaa dha
reading bin ttable
size of OFF_T 8
binary phrasefile loaded, default OFF_T: -1
binary file loaded, default OFF_T: -1
Line 0: Collecting options took 0.000 seconds
Line 0: Search took 0.000 seconds
Jesus is Lord
BEST TRANSLATION: Jesus is Lord [111] [total=-2.671] core = \
(0.000,-3.000,3.000,-7.327,-7.506,-2.527,-1.868,-1.001,-1.285, \
-0.132,0.000,-0.732,-0.388,-3.000,-20.332)
Line 0: Translation took 1.000 seconds total

```

Figure 2. Result of sample Oromo sentence translation

Table 2. The baseline and the improved system translation outputs in BLEU

Baseline System output in BLEU	Improved System output in BLEU	Difference in BLEU
2.56	2.88	0.32

In these experiments, manual evaluation was also used to judge the extent to which the prototype would be acceptable by the users and tell how promising result was obtained so that a full-fledged Oromo-English MT system is optimistically foreseeable based on this research. By executing the translator, 100 Oromo sentences were translated into English language. But only ten pairs of sentences, just 10%, are randomly taken as shown in Appendix A. The table shows the Oromo sentences along with the corresponding translation output of the Oromo-English SMT system, and the human translation reference sentences and average evaluation results of five translators based on adequacy and fluency parameters. There are some Out Of Vocabulary (OOV) words seen in the output, generated by the system and tagged as UNK. It is based on these translation outputs that the human evaluators have judged the performance of this research.

In this manual evaluation method, the translation was evaluated on the parameters of adequacy and fluency. Adequacy is defined as the degree to which the reference sentence is conveyed in the translation; whereas, fluency refers to the grammatical accuracy of the translated text. So the translators used these adequacy and fluency since both parameters have different level for a given translation on which they can be evaluated. For this translation output, the level on which adequacy and fluency were evaluated is given in Table 3 and Table 4, respectively.

Table 3. Level of adequacy of the translator points

Parameter	Definition	Ranking
All	The translated sentence expresses all the meanings of the source sentence	5
Most	The translated sentence conveys most of the meanings of the source sentence	4
Some	The translated sentence conveys some of the meanings of the source sentence	3
a little	The translated sentence conveys few of the meanings of the source sentence	2
None	The translated sentence conveys none of the meanings of the source sentence	1

Table 4. Level of fluency of the translator points

Parameter	Definition	Ranking
Perfect	The translated sentence has very good grammar of the target language	5
Fair	The translated sentence is easy to understand but lack some grammar	4
Acceptable	The translated sentence is broken but is understandable with efforts	3
Nonsense	The translated sentence is not clear	2
Wrong	The translated sentence is understood wrongly	1

The above mentioned parameters (adequacy and fluency) were evaluated by five persons who are native Oromo and are fluent in both languages. The geometric mean of the test parameters was taken and the result is shown in Table 5. This geometric mean was preferred to normalize the ranges being averaged, so that no range dominates the weighting. The names of the translators were kept unmentioned and alias name 'Evaluator #' was used as name of the respondents.

Table 5. The prototype evaluation results by human translators

Evaluator	Adequacy		Fluency		Average Performance	
	The Baseline	The New Model	The Baseline	The New Model	The Baseline	The New Model
Evaluator #1	2.93	2.99	2.88	2.93	2.91	2.96
Evaluator #2	3.06	3.15	2.74	2.83	2.90	2.99
Evaluator #3	4.19	4.34	4.34	4.55	4.27	4.45
Evaluator #4	2.71	2.73	2.97	3.01	2.84	2.87
Evaluator #5	3.51	5.58	3.49	3.55	3.50	4.57
<b>Geometric Average</b>	<b>3.36</b>	<b>3.45</b>	<b>3.39</b>	<b>3.48</b>	<b>3.38</b>	<b>3.47</b>

## 6. Conclusions and Final Remarks

With these experimental setups, discussed so far, the Oromo-English machine translation system produced the required translation output, even under such a scarce resource constraints. With the automatic evaluation method, a BLEU score difference of 0.32, which is about 12.5 % improvement, was achieved. This model was trained on a very small parallel corpus of data set (usually unacceptable for normal SMT system) and also the quality of the parallel corpus both in translation and spelling errors were terribly poor. Over all, the accuracy of the model seems very low, when compared with trillion data set SMT models of other languages. Yet, this measurement was enough to evaluate and contrast the performances of the two models and proved that the idea of complementing the performance of SMT even for such a very small parallel corpus (only about 15,000 sentences) by deploying fine tokenization and POS tag is achievable. By the feedbacks obtained from the human evaluators, this Oromo-English machine translation prototype was found to be a promising startup and a full-fledged MT system may be optimistically adopted from this model.

The translation of 100 sentences was evaluated using human evaluation method on the parameters of adequacy and fluency. For both parameters, a geometric average of 3.45 and 3.48, respectively, was achieved out of 5 maximum points. This human judgment also confirms that the intended system has performed well. It has to be recalled that the quality of the translated text highly depends on the size and quality of the corpus used to train and develop the system, which this research lacks, indeed.

In the future, a better machine translation system is probably achievable by combining this idea and morphological analyzer.

## References

- Adam, L. (2008). Statistical Machine Translation. *ACM Computing Surveys*, 40(3), Article 8. Retrieved from <http://dl.acm.org/citation.cfm?id=1380586>. <http://dx.doi.org/10.1145/1380584.1380586>
- Al-onaizan, Y., Germann, U., Hermjakob, U., Knight, K., Koehn, P., Marcu, D., & Yamada, K. (2000). Translating With Scarce Resources. In Proceedings of the 17<sup>th</sup> National Conference on Artificial Intelligence (AAAI), Pages 672-678, Austin, TX, August.
- Brown, P. F., Pietra, S. A. D., Pietra, V. J. D., & Mercer, R. L. (1993). The Mathematics of Statistical Machine Translation: parameter estimation. *ACL*, 22(1), 39-69.
- El Isbihani, Khadivi, A. S., Bender, O., & Ney, H. (2006). *Morpho-syntactic Arabic Preprocessing for Arabic-to-English SMT*. Proceedings of the Workshop on SMT, pages 15–22, New York City, ACL.
- El-Kahlout, I. D., & Oflazer, K. (2006). *Initial Explorations in English to Turkish Statistical Machine Translation*. Proceedings of the Workshop on SMT, pages 7–14, New York City, June 2006. ACL.
- Federico, M., Bertoldi, N., & Cettolo, M. (2008). *IRSTLM: an Open Source Toolkit for Handling Large Scale Language Models*. Proceedings of Interspeech, Brisbane, Australia.
- Goldwater, S., & McClosky, D. (2005). *Improving Statistical MT through Morphological Analysis*. Proceedings of the conference on Human Language Technology and Empirical Methods in NLP, Pages 676-683, ACL Stroudsburg, PA, USA. Retrieved from <http://dl.acm.org/citation.cfm?id=1220660>. <http://dx.doi.org/10.3115/1220575.1220660>
- Koehn, P., Och, F. J., & Marcu, D. (2003). *Statistical Phrase-based Translation*. In Proceedings of HLT-NAACL. Retrieved from <http://dl.acm.org/citation.cfm?id=1073462>. <http://dx.doi.org/10.3115/1073445.1073462>
- Lee, Y. S. (2004). *Morphological analysis for statistical machine translation*. In Proceedings of HLT-NAACL

- 2004: Short Papers, HLT-NAACL-Short '04, pages 57–60, Stroudsburg, PA, USA. ACL.
- Marcu, D., & Wang, W. (2002). *A Phrase-based, Joint Probability Model for SMT*. In Proceedings of EMNLP.
- Nedjo, A. T., Huang, D., & Liu, X. (2014a). *Challenges of Diacritical Marker or Hudhaa Character in Tokenization of Oromo Text*. *Journal of Software*, 9(7), 1818-1826. <http://dx.doi.org/10.4304/jsw.9.7.1818-1826>
- Nedjo, A. T., Huang, D., & Liu, X. (2014b). *Automatic Part-Of-Speech Tagging for Oromo Language Using Maximum Entropy Markov Model (MEMM)*. *Journal of Information and Computational Science*, 11(10), 3319-3334. <http://dx.doi.org/10.12733/jics20103906>
- Nguyen, T. P., & Shimazu, A. (2006). *Improving Phrase-based SMT with Morpho-Syntactic Analysis and Transformation*. In Proceedings of the 7<sup>th</sup> Conference of the Association for MT in the North Americas, Pages 138-147. The Association for MT in the Americas.
- Nießen, S., & Ney, H. (2004). Statistical Machine Translation with Scarce Resources Using Morpho-syntactic Information. *CL*, 30(2), 181–204.
- Och, F. J., & Hermann, N. (2003). A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics*, 29(1), 19-51.
- Och, F. J., & Ney, H. (2004). The Alignment Template Approach to Statistical Machine Translation. *Computational Linguistics*, 30, 417-449.
- Och, F. J., Tillmann, C., & Ney, H. (1999). *Improved Alignment Models for SMT*. In proceedings of Joint SIGDAT Conf. on Empirical Methods in Natural Language Processing and Very Large Corpora, pages 20-28. University of Maryland, College Park, MD.
- Philipp, K. (2007). *MOSES : Open Source Toolkit for Statistical Machine Translation*, Proceedings of the ACL 2007 Demo and Poster Sessions, pp. 177-180, Prague.
- Philipp, K. (2010). *Statistical Machine Translation*. Cambridge University Press, USA, New York.
- Ramanathan, A., Bhattacharyya, P., Hegde, J., Shah, R. M., & M, S. (2008). *Simple syntactic and morphological processing can help English-Hindi statistical machine translation*. In Proceedings of International Joint Conference on NLP (IJCNLP).
- Ramanathan, A., Choudhary, H., Ghosh, A., & Bhattacharyya, P. (2009). *Case Markers and morphology: addressing the crux of the fluency problem in English-Hindi SMT*. In Proc. of the Joint Conf. of the 47th Annual Meeting of the ACL and the 4th Inter. Joint Conf. on NLP of the AFNLP: Volume 2, pages 800–808, Stroudsburg, PA, USA. ACL.

## Appendix A

Sample translation outputs evaluated by human translators comparing the system output and reference sentences, based on adequacy and fluency parameters.

No	Oromo Sentence	Input	English System Output	Sentence	Human Reference Sentence	Translation	Adequacy	Fluency
1	ganda warra gama sanaa		the other side of the town		The village of the people on the other side		3	4
2	namni cubbuun dhiifameef jireenya baraa baraa argata		He will have forgiveness for their sins and for eternal life		The forgiven ones have eternal life		2	3
3	Waaqayyo gadda garaa namaa beeka		The Lord has knowledge of sorrow is the tester of hearts		God understands man's sorrowful heart		3	3
4	nama du'ee fayye beektuu?		dead men was made well beektuu? UNK		Do you know a person risen from death		3	3
5	jaalalaa fi jibba hin beeku		i have no knowledge of hate of love and		He knows no love nor hatred		5	5

---

6	kun hojii gaarii dha	This is the good work	This is good work	4	4
7	hojiin kun waan gaarii dha	This is good work	This work is a good thing	3	3
8	kun hojii dinqisiisaa Yesus hojjete dha	This is the things which Jesus did the great works	This is amazing works of Jesus	5	5
9	kun jaalala guddaa dha	This is a great love	This is great love	4	4
10	mucaa bareedduu qabdi	has a child bareedduu UNK	She has a beautiful baby	3	4

---

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).