

Amharic-English Speech Translation in Tourism Domain

Michael Melese Woldeyohannis

Addis Ababa University, Addis Ababa, Ethiopia
michael.melese@aau.edu.et

Laurent Besacier

LIG Laboratory, UJF, BP53,
38041 Grenoble Cedex 9, France
laurent.besacier@imag.fr

Million Meshesha

Addis Ababa University,
Addis Ababa, Ethiopia
michael.melese@aau.edu.et

Abstract

This paper describes speech translation from Amharic-to-English, particularly Automatic Speech Recognition (ASR) with post-editing feature and Amharic-English Statistical Machine Translation (SMT). ASR experiment is conducted using morpheme language model (LM) and phoneme acoustic model (AM). Likewise, SMT conducted using word and morpheme as unit.

Morpheme based translation shows a 6.29 BLEU score at a 76.4% of recognition accuracy while word based translation shows a 12.83 BLEU score using 77.4% word recognition accuracy. Further, after post-edit on Amharic ASR using corpus based n-gram, the word recognition accuracy increased by 1.42%. Since post-edit approach reduces error propagation, the word based translation accuracy improved by 0.25 (1.95%) BLEU score.

We are now working towards further improving propagated errors through different algorithms at each unit of speech translation cascading component.

1 Introduction

Speech is one of the most natural form of communication for humankind (Honda, 2003). Computer with the ability to understand natural language promoted the development of man-machine interface. This can be extended through different digital platforms such as radio, mobile, TV, CD and others. Through these, speech translation facilitates communication between the people who speak different languages.

Speech translation is the process by which spoken source phrases are translated to a target lan-

guage using a computer (Gao et al., 2006). Speech translation research for major and technological supported languages like English, European languages (like French and Spanish) and Asian languages (like Japanese and Chinese) has been conducted since the 1980s by NEC Corporation (Kurematsu, 1996). The advancement of speech translation captivates the communication between people who do not share the same language.

The state-of-the-art of speech translation system can be seen as the integration of three major cascading components (Gao et al., 2006; Jurafsky and Martin, 2008); Automatic Speech Recognition (ASR), Machine Translation (MT) and Text-To-Speech (TTS) synthesis.

ASR is the process by which a machine infers spoken words, by means of talking to computer, and having it correctly understand a recorded audio signal. Beside ASR, MT is the process by which a machine is used to translate a text from one source language to another target language. Finally, TTS creates a spoken version from the text of electronic document such as text file and web document.

As one major component of speech translation, Amharic ASR started in 2001 (Melese et al., 2016). A number of attempts have been made for Amharic ASR using different methods and techniques towards designing speaker independent, large vocabulary, continuous speech and spontaneous speech recognition.

In addition to ASR, a preliminary English-Amharic machine translation experiments was conducted using phonemic transcription on the Amharic corpus (Teshome et al., 2015). The result obtained from the experiment shows that, it is possible to design English-Amharic machine translation using statistical method.

As the last component of speech translation, a number of TTS research have been attempted

using different techniques and methods as discussed by (Anberbir and Takara, 2009). Among these, concatenative, cepstral, formant and a syllable based speech synthesizers were the main methods and techniques applied.

All the above research works were conducted using different methods and techniques beside data difference and integration as a cascading component. Moreover, dataset and tools used in the above research are not accessible which makes difficult to evaluate the advancement of research in speech technology for local languages.

However, there is no attempt to integrate ASR, SMT and TTS to come up with speech translation system for Amharic language. Thus, the main aim of this study is to investigate the possibility to design Amharic-English speech translation system that controls recognition errors propagating through cascading components.

2 Amharic Language

Amharic is a Semitic language derived from Ge'ez with the second largest speaker in the world next to Arabic (Simons and Fennig, 2017). The name Amharic (አማርኛ) comes from the district of Amhara (አማራ) in northern Ethiopia, which is thought to be the historic, classical and ecclesiastical language of Ethiopia. Moreover, the language Amharic has five dialectical variations spoken named as: Addis Ababa, Gojam, Gonder, Wollo and Menz.

Amharic is the official working language of government of Ethiopia among the 89 languages registered in the country with up to 200 different spoken dialects (Simons and Fennig, 2017; Thompson, 2016). Beside these, Amharic language is being used in governmental administration, public media and national commerce of some regional states of the country. This includes; Addis Ababa, Amhara, Diredawa and Southern Nations, Nationalities and People (SNNP).

Amharic language is spoken by more than 25 million with up to 22 million native speakers. The majority of Amharic speakers found in Ethiopia even though there are also speakers in a number of other countries, particularly Italy, Canada, the USA and Sweden.

Unlike other Semitic languages, such as Arabic and Hebrew, modern Amharic script has inherited its writing system from Ge'ez (ግዕዝ) (Yimam, 2000). Amharic language uses a grapheme

based writing system called fidel (ፊደል) written and read from left to right. Amharic graphemes are represented as a sequence of consonant vowel (CV) pairs, the basic shape determined by the consonant, which is modified for the vowel.

The Amharic writing system is composed of four distinct categories consisting of 276 different symbols; 33 core characters with 7 orders (አ, አ, አ, አ, አ, አ and አ), 4 labiovelars with 5 orders symbol (ቆ, ኅ, ከ and ነ), 18 labialized consonants with 1 order (ውአ) and 1 labiodental characters consisting 7 orders (አ, አ, አ, አ, አ, አ and አ).

In Amharic writing system, all the 276 distinct orthographic representation are indispensable due to their distinct orthographic representation.

However, as part of speech translation, speech recognition mainly deals with distinct sound. Among those, some of the graphemes generate same sound like (ሀ, ከ, ኅ and ከ) pronounced as h/ʈ/.

On the other hand, Machine translation emphasizes on orthographic representation which result the same meaning in different graphemes. As a result, normalization is required to minimize the graphemes variation which leads to better translation while minimizing the ASR model. Table 1 presents the Amharic character set before and after normalization.

	Unnormalized	Normalized	Difference
Core Character	33	27	6
Labiovelar	4	4	0
Labialized	18	18	0
Labiodental	1	1	0
Total	276	234	42

Table 1: Distribution of Amharic character set adopted and modified from (Melese et al., 2016)

As a result, graphemes that generate the same sound are normalized in to the seven order of core character. The normalization is based on the usage of most characters frequency in Amharic text document. This includes, normalization from (ሀ, ከ, ኅ and ከ) to ሀ, (አ, ዕ) to አ, (ሥ, ከ) to ከ and (ጸ, ዕ) to ጸ along with order.

3 Tourism in Ethiopia

Tourism is the activity of traveling to and staying in places outside their usual environment for not more than one year to create a direct contact between people and cultures (UNWTO, 2016). Ethiopia has much to offer for international

tourists¹ ranging from the peaks of the rugged Semien mountains to the lowest points on earth called Danakil Depression which is more than 400 feet below sea level.

In addition, tourism become a pleasing sustainable economic development that serves as an alternative source of foreign exchange for the counties like Ethiopia.

Moreover, The 2015 United Nations World Tourism report (UNWTO, 2016) and the World Bank² report indicate that, in 2015 a total of 864,000 non-resident tourists come to Ethiopia to visit different tourist attraction. These include; ancient, medieval cities and world heritages registered by UNESCO as tourist attraction. Since the year 2010 until 2015, the average number of tourist flow increase by 13.05% per year.

According to Walta Information Center³, citing Ethiopia Ministry of Culture and Tourism, Ethiopia has secured 872 million dollars in first quarter of its 2016/17 fiscal year from 223,032 international tourists. The revenue was mostly through conference tourism, research business and other activities. Majority of the tourists were from USA, England, Germany, France and Italy speaking foreign languages. Beside this, tourists express their ideas using different languages, the majority of the tourists can speak and communicate in English to exchange information about tourist attractions.

Due to this, language barriers are a major problem for today's global communication (Nakamura, 2009). As a result, they look for an alternate option that lets them communicate with the surrounding.

Thus, speech translation system is one of the best technologies used to fill the communication gap between the people who speak different languages (Nakamura, 2009). This is especially true in overcoming language barriers of today's global communication besides supporting under-resourced language.

However, under-resourced languages such as Amharic, suffer from having a digital text and speech corpus to support speech translation. So, after collecting text and speech corpora, moving

¹http://www.investethiopia.gov.et/images/pdf/Investment_Brochure_to_Ethiopia.pdf

² <http://data.worldbank.org/indicator/ST.INT.ARVL?end=2015>

³<https://www.waltainfo.com/FeaturedArticles/detail?cid=28751>

one step further helps in solving language barriers problem.

Therefore, this study attempts to come up with an Amharic-English speech translation system taking tourism as a domain.

4 Data Preparation

Nowadays, Amharic language suffers from a lack of speech and text corpora for ASR and SMT. Beside these, collecting standardized and annotated corpora is one of the most challenging and expensive tasks when working with under resourced languages (Besacier et al., 2006; Gauthier et al., 2016).

For Amharic speech recognition training and development, 20 hours of read speech corpus prepared by Abate et. al (2005) were used. However, due to unavailability of standardized corpora for speech translation in tourism domain, a text corpus is acquired from resourced and technologically supported languages particularly English.

Accordingly, a parallel English-Arabic text data was acquired from the Basic Traveller Expression Corpus (BTEC) 2009 which is made available through International Workshop on Spoken Language Translation (IWSLT) (Kessler, 2010). A parallel Amharic-English corpus has been prepared by translating the English BTEC data using a bilingual speaker. This data is used for the development of speech translation cascading component such as, ASR and SMT.

The corpus has a total of 28,084 Amharic-English parallel sentences. To keep the dataset consistent, the text corpus has been further preprocessed, such as typing errors are corrected, abbreviations have been expanded, numbers have been textually transcribed and concatenated words have been separated.

Amharic speech recognition is conducted using words and morphemes as a language model with a phoneme-based acoustic model. Similarly word and morpheme have been used as a translation unit for Amharic in Amharic-English machine translation. Morpheme-based segmentation of training, development, testing obtained by segmenting word into sub-word unit using corpus-based, language independent and unsupervised segmentation for using morffessor 2.0 (Smit et al., 2014).

Once the Amharic-English BTEC corpus is prepared, it is divided into training, tuning and testing set with a proportion of 69.33% (19472 sen-

tences), 1.78%(500 sentences) and 28.88%(8112 sentences), respectively.

Then, the 8112 (28.38%) test set sentences are recorded under a normal office environment from eight (4 Male and 4 Female) native Amharic speakers using LIG-Aikuma, a smartphone based application tool (Blachon et al., 2016).

Accordingly, a total of 7.43 hours read speech corpus ranging from 1,020 ms to 14,633 ms with an average speech time of 3,297 ms has been collected from the tourism domain.

Moreover, as suggested by Melese et al., (2016), morphologically rich and under-resourced language like Amharic provides a better recognition accuracy using morpheme based language model with phoneme based acoustic model.

Similarly, language model data for Amharic speech recognition has been collected from different sources. A text corpus collected for Google project (Tachbelie and Abate, 2015) have been used in addition to BTEC SMT training data excluding the test data. Table 2 presents the training, development and language model data used for Amharic speech recognition.

	Train	Test	Language Model	
			Word	Morpheme
Sentence	10,875	8,112	261,620	261,620
Token	145,404	50,906	4,223,835	5,773,282
Type	24,653	4,035	328,615	141,851

Table 2: Distribution of Amharic data for ASR.

Like speech recognition, a total of 42,134 sentences (374,153 token of 8,678 type) English language model data have been used for Amharic-English machine translation. The data is collected from the same BTEC corpus excluding test data.

Consequently, corpus based and language independent segmentation have been applied on a training, development and test set of Amharic SMT data. Morfessor is used to segment words to a sub word units. Table 3 presents summary of the corpus used for Amharic-English machine translation using word and morpheme as a unit.

Likewise, the post-edit is conducted using a corpus based n-gram approach. Accordingly, a corpus containing 681,910 sentences (11,514,557 tokens) of 582,150 type data crawled from web including news and magazine.

Then, the data is further cleaned, preprocessed and normalized. From this data, a total of 5,057,112 bigram, 8,341,966 trigram, 9,276,600 quadrigram and 9,242,670 pentagram word se-

	Unit		Train	Dev	Test
			Amharic	Word	Sentence
		Token	107,049	2,795	37,288
		Type	18,650	1,470	4,168
	Morpheme	Sentence	19,472	500	8,172
		Token	145,419	3,828	50,906
		Type	15,679	1,621	4,035
English	Word	Sentence	19,472	500	8,172
		Token	157,550	4,024	55,062
		Type	10,544	1,227	3,775

Table 3: Distribution of Amharic-English SMT data.

quences have been extracted after expanding numbers and abbreviation.

5 System Architecture

As discussed in Section 1, the state-of-the-art of speech translation suggest to apply through the integration of cascading components to translate speech from source language (Amharic) to a target language (English).

As part of the cascading components, the output of a speech recognizer contains more and presents a variety of errors. These errors further propagates to the succeeding component of speech translation which results in low performance.

Hence, in this study we propose an Amharic ASR post-editing module that can detect an error, identify possible suggestion and finally correct based on the proposal. The correction is made using n-gram data store using minimum edit distance and perplexity before the error heads to statistical machine translation.

Figure 1 presents Amharic-English speech-to-speech translation (S2ST) architecture with and without considering ASR post-edit.

The post-edit process mainly consists of three different phases; error detection, correction proposal and finally suggest correction as depicted in Figure 2.

The first phase of post editing is to detect the error from ASR recognition output. Basically, to detect an error, recognized morpheme units are concatenated to form a word and its existence is checked in unigram Amharic dictionary.

Thus, a morpheme-based speech recognition output “የተ-ስጦታ እቃ -ተዘነጉ ተስፋ አደር+ -ጋለሁ”⁴ concatenated to form a phrase “የስጦታ እቃ -ተዘነጉ ተስፋ አደር,ጋለሁ”.

⁴“+” refers to morphemes followed by other morpheme while “-” refer to leading morpheme is there.

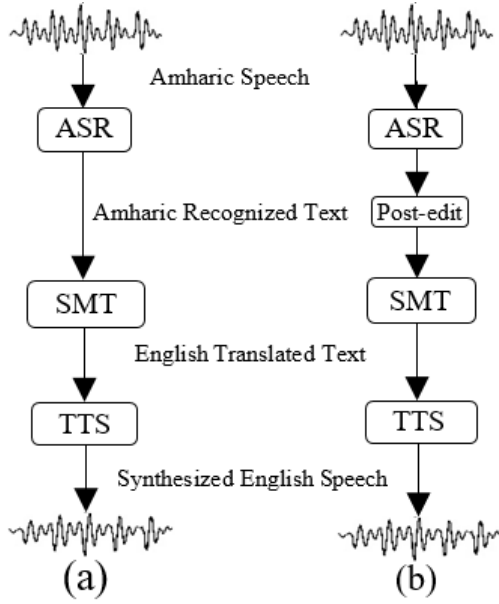


Figure 1: Amharic-English speech-to-speech translation architecture (a) without post-edit (b) with post-edit

If the word is not in the unigram Amharic dictionary, then the “word” is considered as an error and marked as error(“*”) then it is concatenated to the remaining words. Accordingly, each token checked in unigram dictionary and the word “-ተዘነገ” is not in dictionary which is marked as an error.

If the error is detected during the first phase, then the correction proposal phase takes the sentence with error mark and creates $(w-n+1)$ n-grams after adding start “<s>” and end “</s>” symbol, where w is number of token in sentence and n specifies n-grams. Otherwise, the sentence is considered as correct.

Consequently, three pantagram word sequences are generated from the speech recognition of “<s> የሰጦታ ሊቃ -ተዘነገ ተስፋ አደርጋለሁ </s>” sentence. These are;

1. <s> የሰጦታ ሊቃ * ተስፋ
2. የሰጦታ ሊቃ * ተስፋ አደርጋለሁ
3. ሊቃ * ተስፋ አደርጋለሁ </s>

Subsequently, we select the n-grams with error marks and search in n-gram data store to select possible candidates for correction after removing the error mark. If there is no candidate in n-gram, then go for (n-1)-gram order until bigram.

Once the candidate identified, the suggestion is made taking the minimum edit distance between

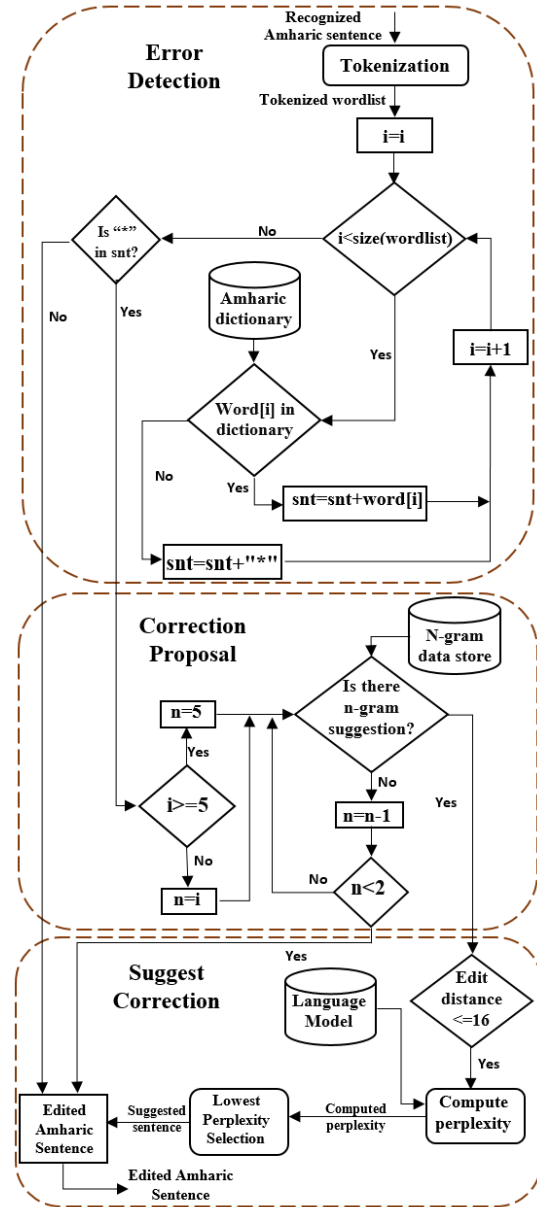


Figure 2: Amharic ASR post-edit algorithm

the error detected and suggestion selected. In this phase, the sum of maximum edit distance has been set experimentally to 16. The maximum edit distance 16 was selected to provide at least one suggestion per sentence and minimize the computation of perplexity. Table 4 depicts a sample of possible correction proposal for a sentence “የሰጦታ ሊቃ -ተዘነገ ተስፋ አደርጋለሁ”.

Finally, the suggestion is made primarily using minimum edit distance then by calculating the perplexity. The minimal edit distance is computed between the word “-ተዘነገ” and the underlined n-gram based possible suggestion from a sentence of Table 4.

Possible suggestion list	Distance
የስጦታ እቃ ተስፋ አደርጋለሁ ብለዋል	5
የስጦታ እቃ ብዩ ተስፋ አደርጋለሁ	5
የስጦታ እቃ አንደማይ ተስፋ አደርጋለሁ	5
የስጦታ እቃ ተስፋ አደርጋለሁ ይላሉ	5
የስጦታ እቃ ተስፋ አደርጋለሁ ለፊገግታ	5
የስጦታ እቃ ብዩ ተስፋ አደርጋለሁ	5
የስጦታ እቃ ተስፋ አደርጋለሁ ብሏል	5
የስጦታ እቃ ተስፋ አደርጋለሁ ይላሉም	5
ማንኛውም የስጦታ እቃ ተስፋ አደርጋለሁ	5
የስጦታ እቃ የሚሸጥበት ተስፋ አደርጋለሁ	6
የስጦታ እቃ አንደሚሆን ተስፋ አደርጋለሁ	6
የስጦታ እቃ አንደሚተካ ተስፋ አደርጋለሁ	6
የስጦታ እቃ አንደደረሰ ተስፋ አደርጋለሁ	6
የስጦታ እቃ አንደሚወጣ ተስፋ አደርጋለሁ	6
የስጦታ እቃ አንደሚመጣ ተስፋ አደርጋለሁ	6
የስጦታ እቃ አንደሚጠጣ ተስፋ አደርጋለሁ	6

Table 4: Sample n-gram based suggestion for a sentence “የስጦታ እቃ ብለዋል ተስፋ አደርጋለሁ”.

If the edit distance is the same as a different suggestion, then the decision is made by selecting the one that result lower perplexity.

Accordingly, the phrase “የስጦታ እቃ ተስፋ አደርጋለሁ ብለዋል” selected due to better perplexity of language model.

Similarly, Table 5 presents sample Amharic speech recognition output along with the corrected sentence using our post-edit technique.

No	Type	Sentence recognized and corrected
1	Raw	አራስን ጭምር ጉብኝት እገዛ ዋጋ ይኖረል
	Edited	አራስን ጭምር ጉብኝት በዛ ዋጋ ይኖረል
2	Raw	አባክዎን ተሸካሚ ተረፋት
	Edited	አባክዎን ተሸካሚውን ተረፋት
3	Raw	አባኩን ሌላ ፎጣ ማግኘት እችላለሁ
	Edited	አባክዎን ሌላ ፎጣ ማግኘት እችላለሁ
4	Raw	አባክዎን ሻንጣዎን ይክፈቱ+
	Edited	አባክዎን ሻንጣዎን ይክፈቱት
5	Raw	የስጦታ እቃ ተስፋ አደርጋለሁ ነው
	Edited	የስጦታ እቃ ለመግዛት አደርጋለሁ ነው
6	Raw	ይህ የጉዞ ላይ ህመም ይገጥም ይችላል
	Edited	ይህ የጉዞ ላይ ህመም ሊያገጥም ይችላል
7	Raw	-ሀ ባቡር ዶብር በስንት ሰአት ይደርሳል
	Edited	ይህ ባቡር ዶብር በስንት ሰአት ይደርሳል
8	Raw	ሻንጣዎን ይክፈቱ -ም
	Edited	ሻንጣዎን ይክፈቱ
9	Raw	አባክዎን ተሸካሚውን ተረፋት
	Edited	አባክዎን ተሸካሚውን ይተረፋት

Table 5: Sample corrected sentences of Amharic speech recognizer.

6 Experimental results

Speech translation experiments are conducted through cascading components of speech translation as discussed in Section 1. In speech recognition experiments, Kaldi (Povey et al., 2011), SRILM (Stolcke et al., 2002) and Morfessor 2.0 (Smit et al., 2014) have been used for Amharic speech recognition, language modeling and unsupervised segmentation, respectively.

Morfessor based segmentation has been applied to segment training, testing and language model data for Amharic. In addition to this, Moses and MGIZA++ for implementing a phrase based statistical machine translation and Python is used for implementing the post-edit algorithm and to integrate ASR and SMT under the Linux platform.

The entire ASR experiment is conducted using a morpheme-based language model with phoneme-based acoustic model. Accordingly, the experimental result is computed using NIST Scoring Toolkit (SCTK)⁵ and presented in terms of word recognition accuracy (WRA⁶) and morph recognition accuracy (MRA).

Thus, the Amharic speech recognition experiment shows a 76.4% for the morpheme-based. Then, after the concatenation of morphemes to words, a 77.4% word-based recognition accuracy have been achieved.

Consequently, Amharic-English SMT experiment have been conducted with and without considering Amharic ASR result.

The first two experiments were conducted without considering ASR. Accordingly, a word-word system resulted in a BLEU score of 14.72 while morpheme-word brings about 11.24 BLEU. Combining Amharic ASR with Amharic-English SMT as cascading component resulted in a 6.29 BLEU score through 76.4% of recognition accuracy for Amharic morpheme and English word based translation.

Similarly, Amharic word with English word based translation shows a 12.83 BLEU score using 77.4% recognition accuracy without using ASR post-edit. The result achieved by ASR can further be improved by applying post-edit on Amharic speech recognition.

⁵evaluation toolkit available at <http://my.fit.edu/~vkepuska/ece5527/sctk-2.3-rc1/doc/sctk.htm>

⁶WRA is obtained by concatenating the result obtained by MRA

Table 6 depicts Amharic-English speech translation before and after Amharic ASR post-edit.

	Before		After
	Morpheme	Word	Word
ASR (%)	76.4	77.4	78.5
SMT (in BLEU)	6.29	12.83	13.08

Table 6: Amharic-English Speech Translation result.

Accordingly, the morpheme based recognition followed by post-edit resulted in a BLEU score of 13.08 at 78.5% of word recognition accuracy speech translation.

The result obtained from the n-gram post-edit experiment shows an absolute advance by 1.42% from word recognition accuracy of 77.4% obtained by concatenating a 76.4% morpheme based recognition. Similarly, BLEU score evaluation advanced by 1.95% (from 12.83 to 13.08).

7 Conclusion and Future work

Speech translation research has been studied for more than a decade for resourced and technological supported languages like English, European and Asian. On the contrary, attempts for under resourced languages, not yet started, particularly for Amharic. This paper presents the first Amharic speech to English text translation using the cascading components of speech translation.

For ASR, a 20 hours of training and 7.43 hours of testing speech were used consuming a morpheme-based language model with a phonemic acoustic model. Whereas for SMT, 19,472 sentence for training and 8112 sentences for testing used. Similarly to apply ASR post-edit using n-gram approach, a corpus consisting 681,910 sentences were used.

Accordingly, speech translation through ASR post-editing resulted a 0.25 (1.95%) BLEU score enhancement from the word-based SMT. The enhancement seemed as a result of improving ASR by 1.42% using a corpus based n-gram post-edit.

The current study shows the possibility of enhancing the performance of speech translation by controlling speech recognition error propagation using post-editing algorithm.

Further works need to be done to apply post-editing both at the recognition and the translation stages of speech translation.

References

- Solomon Teferra Abate, Wolfgang Menzel, Bairu Tafila, et al. 2005. An amharic speech corpus for large vocabulary continuous speech recognition. In *INTERSPEECH*, pages 1601–1604.
- Tadesse Anberbir and Tomio Takara. 2009. Development of an amharic text-to-speech system using cepstral method. In *Proceedings of the First Workshop on Language Technologies for African Languages*, pages 46–52. Association for Computational Linguistics.
- Laurent Besacier, V-B Le, Christian Boitet, and Vincent Berment. 2006. Asr and translation for under-resourced languages. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, volume 5, pages V–V. IEEE.
- David Blachon, Elodie Gauthier, Laurent Besacier, Guy-Noël Kouarata, Martine Adda-Decker, and Annie Rialland. 2016. Parallel speech collection for under-resourced language studies using the ligaikuma mobile device app. *Procedia Computer Science*, 81:61–66.
- Yuqing Gao, Liang Gu, Bowen Zhou, Ruhi Sarikaya, Mohamed Afify, Hong-Kwang Kuo, Wei-zhong Zhu, Yonggang Deng, Charles Prosser, Wei Zhang, et al. 2006. Ibm mastor system: Multilingual automatic speech-to-speech translator. In *Proceedings of the Workshop on Medical Speech Translation*, pages 53–56. Association for Computational Linguistics.
- Elodie Gauthier, Laurent Besacier, Sylvie Voisin, Michael Melese, and Uriel Pascal Elingui. 2016. [Collecting resources in sub-saharan african languages for automatic speech recognition: a case study of wolof](#). In *Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2016, Portorož, Slovenia, May 23-28, 2016*.
- Masaaki Honda. 2003. Human speech production mechanisms. *NTT Technical Review*, 1(2):24–29.
- Daniel Jurafsky and James H Martin. 2008. *Speech and language processing (prentice hall series in artificial intelligence)*. Prentice Hall.
- Fondazione Bruno Kessler. 2010. A generic weaver for supporting product lines. In *International Workshop on Spoken Language Translation*, pages 11–18. ACM.
- Akira Kurematsu. 1996. *Automatic Speech Translation*, volume 28. CRC Press.
- Michael Melese, Laurent Besacier, and Million Mesheha. 2016. Amharic speech recognition for speech translation. In *Atelier Traitement Automatique des Langues Africaines (TALAF). JEP-TALN 2016*.

- Satoshi Nakamura. 2009. Overcoming the language barrier with speech translation technology. *Science & Technology Trends-Quarterly Review*, (31).
- Daniel Povey, Arnab Ghoshal, Gilles Boulianne, Lukas Burget, Ondrej Glembek, Nagendra Goel, Mirko Hannemann, Petr Motlicek, Yanmin Qian, Petr Schwarz, et al. 2011. The kaldi speech recognition toolkit. In *IEEE 2011 workshop on automatic speech recognition and understanding*, EPFL-CONF-192584. IEEE Signal Processing Society.
- Gary F. Simons and Charles D. Fennig. 2017. *Ethnologue: Languages of the World*. SIL, Dallas, Texas.
- Peter Smit, Sami Virpioja, Stig-Arne Gronroos, and Mikko Kurimo. 2014. Morfessor 2.0: Toolkit for statistical morphological segmentation. In *14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 21–24. European Chapter of the Association for Computational Linguistics, EACL.
- Andreas Stolcke et al. 2002. Srilm-an extensible language modeling toolkit. Interspeech.
- Martha Yifiru Tachbelie and Solomon Teferra Abate. 2015. Effect of language resources on automatic speech recognition for amharic. In *AFRICON, 2015*, pages 1–5. IEEE.
- Mulu Gebreegziabher Teshome, Laurent Besacier, Girma Taye, and Dereje Teferi. 2015. Phoneme-based english-amharic statistical machine translation. In *AFRICON, 2015*, pages 1–5. IEEE.
- Irene Thompson. 2016. [About world language](#). Accessed: 2017-05-26.
- UNWTO. 2016. World tourism organization annual report 2015. Technical report, United Nation, Madrid, Spain.
- Baye Yimam. 2000. *Yeamarigna sewasew (Amharic version)*. EMPDA, Addis Ababa, Ethiopia.