International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 11 | Nov 2019 www.irjet.net p-ISSN: 2395-0072

KINYARWANDA SPEECH RECOGNITION IN AN AUTOMATIC DICTATION SYSTEM FOR TRANSLATORS: THE TKTALK PROJECT

Kayitare Philbert[1], Dr Niyigena Papias[2]

[1]University of lay Adventist of Kigali, Adventist University of Central Africa
[2]Master, Dept. of Computer and Information Sciences, University of lay Adventist of Kigali, Kigali, Rwanda

---

Abstract - *Machine translation (MT) plays an important role in benefiting linguists, sociologists, computer scientists, etc. by processing natural language to translate it into some other natural language. And this demand has grown exponentially over past couple of years, considering the enormous exchange of information between different regions with different regional languages. Machine Translation poses numerous challenges, some of which are: a) Not all words in one language has equivalent word in another language b) Two given languages may have completely different structures c) Words can have more than one meaning. Owing to these challenges, along with many others, MT has been active area of research for more than five decades. Numerous methods have been proposed in the past which either aim at improving the quality of the translations generated by them, or study the robustness of these systems by measuring their performance on many different languages. In this literature review, we discuss statistical approaches (in particular word-based and phrase-based) and neural approaches which have gained widespread prominence owing to their state-of-the-art results across multiple major languages.*

This paper describes a system designed for use by translators that enables them to dictate their translation. Because the speech recognizer has access to the source text as well as the spoken translation, a statistical translation model can guide recognition. This can be done in many different ways. We discuss the experiments that led to integration of the translation model in a way that improves both speed and performance.

***Key Words*: TKTALK, Kinyarwanda, ubusa, hano, nonaha**

1. INTRODUCTION

The TKTALK project attempts to integrate speech recognition and machine translation in a way that makes maximal use of their complementary strengths. Translators often dictate their translations first and have them typed afterwards. If they dictate to a speech recognition system instead, and if that system has access to the source language text, it can use probabilistic translation models to aid recognition. For instance, if the speech recognition system is deciding between the acoustically similar Kinyarwanda words ubusa (nothing) and ubusa (nakedness), the presence of the word nakedness in the English source text will guide it to the correct choice.

We have implemented a prototype of TKTALK that takes as input an English text and a spoken Kinyarwanda translation, and yields Kinyarwanda text. This paper will focus on the unique problems of large vocabulary speech recognition in the Kinyarwanda language, and features of our system designed to deal with these problems, as well as presenting experiment

2. KINYARWANDA SPEECH RECOGNITION

In the system's 20,000-word Kinyarwanda words each entry has a phonetic representation based on 46 phones including 5 vowels and 19 consonants. To enhance the effect of the translation module, certain frequent expressions such hano or nonaha are entered as units. The base pronunciations were obtained automatically, using a set of grapheme-to-phoneme rules which take into account phonetic idiosyncrasies of the Kinyarwanda spoken in Rwanda (such as assibilation and vowel), and were then manually varied.

Other idiosyncrasies of the Kinyarwanda language have to be taken into account when building a speech recognition system. Dictionary explosion due to elision and homophones is one of them. Elision causes the last vowel of some function words to be omitted when the following word begins with any vowel. Our system handles these contracted units as separate words.

Homophones are more frequent in Kinyarwanda than in English, and often cause word errors in Kinyarwanda speech recognition. In our 20,000 word vocabulary there are 6,020 words such that for each of them, at least one homophone exists. Thus, 30% of our words is made up of homophones, compared to 5% for the 19,977 WSJ word words [4]. Where only one member of a set of homophones is predicted by the English source sentence, the translation module will be particularly useful a major challenge that must be overcome in Kinyarwanda speech recognition is liaison. Liaison appears in continuous speech and occurs when a consonant at the end of a word, orthographically present but not pronounced in

the isolated word, is pronounced in the presence of a vowel at the beginning of the next word. Whether or not this consonant is pronounced is di cult to predict it depends on a complex interaction between orthography, syntax, semantics, and other factors.

We have tried several methods for handling liaison during acoustic testing. Currently, possible liaisons are derived automatically in the words by a set of simple rules.

During training and recognition this liaison will be active if the following word begins with a vowel. However, even when the liaison is active, the acoustic realization of liaison is optional; if it occurs, an insertion penalty is imposed. Four consonants can participate in liaison: y, n, m, p, k/.

3. Applying Translation Models

Given a source sentence in English and a spoken Kinyarwanda translation of that sentence, our system must try to generate the transcription of the Kinyarwanda sentence. A variety of translation models, all requiring bilingual training data, will yield probability estimates for Kinyarwanda words, given the English source sentence. The models mainly in the extent to which they take into account the position of given words in the English and in the Kinyarwanda sentence. For the purposes of this paper, one need only understand that some of these models are crude and computationally cheap, while others are sophisticated and expensive. What models should be chosen, and how can the probabilities they generate be applied to constrain speech recognition?

4. CONCLUSION

We considered two problems in the context of the English-Kinyarwanda. First, can we tell the difference between an original and translated word, and to what level of accuracy? Second, is the knowledge of the translation direction useful for machine translation, and if so, is the classification performance sufficient? Using various speech representations, we found that we could detect original word vs. translation using SVMs with high accuracy: 90+% using words

REFERENCES

1. Marc Dymetman, Julie Brousseau, George Foster, Pierre Isabelle, Yves Normandin, and Pierre Plamondon. Towards an automatic dictation system for translators: In Proceedings, ICSLP 94, volume 2, pages 691{694, Yokohama, Japan, September 1994.

2. J.L. Gauvain and L.F. Lamel et al. Speaker-independent continuous speech dictation. Speech Communication, 15:21{37, 1994.

3. M. Lennig. 3 listes de 10 phrases phonetiquement equilibrees. Revue d'acoustique, 1(56):39{42,1981.

4. Y. Normandin and D. Bowness et al. CRIM's November 94 continuous speech recognition system. In Proceedings of the Spoken Language Systems Technology Workshop, Austin, Texas, January 1