Software Proposal Document for Computer-aided Simultaneous Conference Interpretation

Wessam Ahmed, Mostafa Farhgaly Mahran, Ahmed Sameh, and Abdelrahman Ashraf

September 26, 2018

Abstract

Language interpreters rely on a number of techniques for real-time conference interpretation. A study performed on a number of interpreters showed that 16 percent of those polled associate the profession of language interpretation to that of an air traffic controller, further backed by other studies proving that physiological measures of stress levels and depression in language interpreters during both simultaneous and remote interpreting jumps to abnormal levels, often leading to physical side-effects, sometimes fatal. The project discussed herein explores the proposed method of automatic speech recognition and the subsequent transcription and translation of said speech through the use of a context-aware artificially intelligent computer system that makes use of textual entailment methods for semantic interpretation of the speech fed to the system in order to determine its context. Existing approaches to establish textual entailment achieve 90 percent accuracy, which this project aims to exceed in terms of the English language and at least match for the Arabic language.

1 Introduction

1.1 Background

In recent years the integration of common business needs with information technology has been both customary and necessary, though some markets, fields, and domains have refused to adapt. One such field is the language services industry, which currently relies on already-proven approaches to the task of language interpretation (translation) instead of moving on to faster technological approaches. Computer-assisted interpreting tools have been the main challenge posed to old-fashioned and archaic systems. IBM is one of the companies currently leading the charge to automate the field of language interpretation, with headphones that enable conference interpreters to focus on cognitive tasks rather than confuse themselves with laptops and tablets running search engines that enable them to quickly look up a phrase or term. While another proposed approach is remote video-interpreting. Both have been actively refused by the

language services industry, whose experts are firm in their belief that no such system can match the efficiency of being in the attendance of the speaker whose speech is undergoing the interpretation process.

The language services industry has seen the emergence of several computer-aided interpreting tools including Intragloss and InterpretBank, though those systems are typically restricted to their adaptive rather than automated nature in their use.

The use of computer-aided technologies in the context of language interpretation is currently constrained to booths, laptops equipped with a browser capable of running a search engine, and wired-in headphones that ease the cognitive processes that enable a human to accurately translate one language to another. However the significant distractions that these conveniences come with set interpreters up for failure, in that the difficulty of an interpreter's job is made considerably more complex because of the added hindrance of entering search queries into a search engine to display the meaning of a term or phrase. A mock conference held in May 1997 at the SSLMIT[1] (Advanced School of Modern Languages for Interpreters and Translators) surveyed the results of 45 participants, 30 of which were students with no professional interpreting experience, while the remaining 15 were free-lance interpreters with at least 8 years of experience. While the stress levels of all participants varied within the standard range of the general population[2] (Krug et al. 1976; Krug et al. 1976), students showed considerably higher levels of anxiety and stress, while interpreters in simultaneous interpreting situations were relatively unaffected by the tension involved with the task of interpretation. This however changed when the same experienced interpreters were placed in remote interpreting situations, in which they have very little experience as it is a newly introduced technique that requires that interpreters develop more strategies and coping techniques to deliver satisfactory results.

These preliminary studies show that more work is needed in the language services industry to prepare interpreters for situations they might not be used to, if they are to rely on existing methods for interpretation. Through the use of the software proposed in this document, the need for this training would be averted since the interpreter would only be left with the task of using their linguistic and extra-linguistic knowledge to make use of the automatically transcribed and translated speech, rather than using hardware that would only transfer acoustic signals to the ear and not intelligently assist the interpreter the way an artificially intelligent system would.

It is therefore proposed in this document that a suitable software is introduced to automate interpreters' work and save them the distraction of laptop screens and flashy gadgets, instead abstracting their use to a clear-text screen displaying only the information required by the interpreter with an approach that prioritizes academic value as well as marketability and ease-of-use.

1.2 Motivation

Language variability poses a significant challenge in the semantic interpretation of strings. Human semantic relationships are complex and cannot be derived from linguistic understanding alone, oftentimes one word may carry many meanings. This problem is exacerbated when the text to be processed is of a language with different structure than the English language, such as in the case of the Arabic language. In that regard, methods such as textual entailment can be used to obtain the semantic interpretation of a given text. In the context of the proposed system herein, this text would be fed to the system using supportive hardware, such as a microphone. The acoustic data will be passed through a semantic network for semantic processing, deep structure formation, synthetic processing, surface structure formation, and finally morphological processing to extract the underlying natural language sentence.

Existing research in the area of language variability and ambiguity use mathematical approaches to obtain textual entailment. Current state-of-the-art systems achieve 90 percent accuracy [7], where a study found that humans were in agreement with the used dataset 95.25 percent of the time. The aim of this software is to delegate the process of domain specification to an artificially intelligent agent using pre-trained word vectors in both the English and Arabic language, with an algorithm that is capable of semantically interpreting the text to obtain the underlying meaning, to then be translated using a glossary collected by Alsun students who will be co-operating with the team responsible for implementing this project. Machine translation techniques will be used for the latter stage with the help of IBM's Watson API and relevant audio-capturing software.

This project aims to improve on the accuracy of obtained textual entailment for the purposes of measuring natural language understanding. We will explore existing methods such as surface syntax, lexical inference, and word embedding in terms of the English language, and attempt to replicate or at least match these results in term of the Arabic language, with due consideration to language discrepancies such as punctuation, diacritics, and word formation. Deep learning techniques will be used for deriving shallow semantic features for recognizing textual entailment, such as shallow word overlap.

1.3 Problem Definitions

1. The pressures and stress of conference interpretation highly affect interpreters in both simultaneous and remote interpreting situations

2. No existing techniques assist the interpreter in the task of interpretation but only serve as tools that ease the process of transferring either audio or video data feeds leaving translation and transcription up to the interpreter which allows room for human error

2 Project Description

2.1 Objective

Through the use of a structured and partitioned dictionary on the basis of contextual output in relation to a set of domains to be specified during the deployment phase of the application, the desired use of the software is the successful and accurate interpretation of conferences and what will be dubbed the "online chat" interaction between two sets of users, for the purposes of this document.

After the system is fed speech in the form of audio files transmitted to it via a speech recognition software adapter (or API) with the aid of computer hardware (a microphone) supporting this behavior, the output would be a string of text containing highlighted portions of the speech that match the stored dictionary with which the system interprets one language and translates it to another, or simply dissects it.

Contextual analysis will be achieved by achieving textual entailment through mathematical solutions, such as the cosine similarity approach used by Microsoft's Sent2vec API. Subsequently, sentence vectors will be compared with reference sentences to measure natural language understanding. The success of this stage will depend on further studies to measure human agreement with the obtained results.

2.2 Scope

This project explores domain-specific semantic understanding of text to obtain the underlying natural language meaning. It will make use of several machine learning techniques for tasks relating but not restricted to textual and logical entailment. The project will optionally expand to include audio refinement and pre-processing for speech-to-text interpretation and its subsequent translation.

2.3 Project Overview

The main components of this system are :

1. A computer or similar workstation that is preferably equipped with a monitor
    - This computer would be connected to hardware equipment (such as a cable) that enables it access to the Internet
    - The computer would be equipped with a microphone for the purposes of capturing audio
2. The software program required to perform the task of speech interpretation and transcription

- The software will carry out the task of speech recognition, transcription, and interpretation/translation with the use of custom language models and a context-aware agent
- The personal computer will be equipped with a helper software which will be responsible for speech recognition in the form of IBM's Watson computer system
- The software will carry out the task of domain specification using an artificially intelligent agent capable of semantic interpretation
- The resulting presentation layer will only be responsible for outputting ASCII characters corresponding to translated speech and will not be responsible for the task of translating/transcribing audio feeds itself

3. The computer or workstation using which the user will access the software program will optionally communicate with an online, cloud computing device (a server) that handles the process of text-lookup rather than processing it at high resource costs on the user's machine

3 Similar System Information

1. Evaluating different confirmation strategies for speech-to-speech translation systems.[3]

    - Q1: Motivation of the work?
    - Create efficient confirmation techniques for the translation done.
    - Q2: The main problem statement of the work?
    - The translated text may not be that accurate.
    - Q3: How the researchers contributed to solve the problem?
    - They allowed the used to judge the translation and the work of the system before carrying on.
    - Q4: What main results the researchers reach?
    - The results obtained, show that user-judged back-translation is a viable strategy for generating confirmation utterances.
    - Q5: How do you think this paper you read is important for you?
    - Provided different ideas and methodology to test the translations accuracy.

2. Speech-to-Speech Translation in an Assisted Living Lab.[4]

    - Q1: Motivation of the work?
    - To make living in a safe, autonomous, independent, and convenient way in a domestic environment with multimodal human-machine interfaces possible.

- The main problem statement of the work?
- Users are more comfortable when speaking in mother tongue, so making most systems unilingual, and making life easier for monolinguals.
- How the researchers contributed to solve the problem?
- A user speaks in their mother tongue and the speech is transcribed into text through ASR software; then an MT system translates the existing grammar into the users natural language, and then it gives the output back in the users mother tongue through a TTS application.
- Q4: What main results the researchers reach?
- They managed to achieve multilingual human-robot interaction.
- Q5: How do you think this paper you read is important for you?
- It provides new ways to think about how multilingual human-robot interactions can be achieved and becomes more efficient.

3. ASR error detection in a conversational spoken language translation system.[5]

   - Q1: Motivation of the work?
   - To avoid propagation of errors in automated speech recognition.
   - Q2: The main problem statement of the work?
   - Error detection in automated speech recognition.
   - Q3: How the researchers contributed to solve the problem?
   - They ventured beyond traditional features obtained from the ASR decoder and hypothesized word sequence, and explore additional information streams provided by an error-robust CSLT system.
   - Q4: What main results the researchers reach?
   - They achieved the single biggest gain in error detection rate (1.7
   - Q5: How do you think this paper you read is important for you?
   - It helps in exploring more efficient ways in error detection during ASR.

4. Measuring human readability of machine generated text: Three case studies in speech recognition and speech translation.[6]

   - Q1: Motivation of the work?
   - Improving the readability of output text in translation systems.
   - Q2: The main problem statement of the work?
   - Some hinders and slowing downs take place when moving from one language to another.
   - Q3: How the researchers contributed to solve the problem?

- They improved the performance of the systems by using the gold standard data provided by humans.
- Q4: What main results the researchers reach?
- They managed to achieve progress in translating Arabic and Chinese into an English text that can read by a native speaker.
- Q5: How do you think this paper you read is important for you?
- It can be very helpful in knowing the current limitations in translating Arabic into a readable English text.

5. Comparative Study of Sentiment Classification for Automated Translated Latin Reviews into Arabic.[8]

    - Q1: Motivation of this work?
    - We believe automated translation may be considered as a viable option to produce rich and well represented datasets in order to develop robust and efficient Arabic sentiment analyzers.
    - Q2: The main problem statement of the work?
    - Arabic Text challenges: Standard language Dialects - Lack of resources names and figure of speech sentences.
    - Q3: How the researchers contributed to solve the problem?
    - They implemented two widely accepted supervised classifiers (SVM and Logistic). Both classifiers were trained on both balanced and unbalanced datasets.
    - Q4: What main results the researchers reach?
    - The results are comparable and competitive for the English datasets as well as the translated Arabic ones.
    - Q5: How do you think this paper you read is important for you?
    - The usage of multiple classifiers and achieving good results expands our options.

6. Automated Translation Machines: Challenges and a Proposed Solution.[9]

    - Q1: Motivation of this work?
    - Purpose of this paper is to discuss challenging issues in MT tool developments, the state of art of the MT tools and propose a framework for a semantic-based translation.
    - Q2: The main problem statement of the work?
    - The main reason why a robust tool cannot be developed is a natural language ambiguous. There are 4 types of ambiguity; lexical ambiguity, referential ambiguity, scope ambiguity and structural ambiguity.
    - Q3: How the researchers contributed to solve the problem?

- The ambiguity resolution technique can be resolved by integrating 4 techniques namely, NLP techniques, possibility theory, fuzzy sets and a knowledge based approach.
- Q4: What main results the researchers reach?
- The obtained precision rate is 85.7 percent and 80.3 percent for recall. The results indicate that the proposed technique is successful.
- Q5: How do you think this paper you read is important for you?
- They proposed a solution to a big issue in NLP which is the ambiguity of the language and this is massive for us.

7. Semi Natural Language Algorithm to Programming Language Interpreter.[10]

   - Q1: Motivation of the work?
   - To make it easier for programmers to write algorithms without being held back by different programming languages illiteracy.
   - Q2: The main problem statement of the work?
   - Develop software that is capable of converting algorithms. written in natural language to a programming language.
   - Q3: How the researchers contributed to solve the problem?
   - They create a model that can take an English algorithm written in natural language and convert it to a specific programming language.
   - Q4: What main results the researchers reach?
   - They managed to achieve very promising results as they are from a very small group that tried to solve such problem.
   - Q5: How do you think this paper you read is important for you?
   - Looking on understanding natural language and processing it from a different perspective can be very helpful.

8. Semantic natural language translation based on ontologies combination.[11]

   - Q1: Motivation of the work?
   - Solve the different task in automatic translation of the diversity of the contextual interpretation between diverse languages.
   - Q2: The main problem statement of the work?
   - Translate text from a language to another without losing the true meaning or context of the text.
   - Q3: How the researchers contributed to solve the problem?
   - They created a model that uses ontologies combination to translate a given text from a source language to another.
   - Q4: What main results the researchers reach?

- They managed to achieve progress in training the model to detect different meanings of text and are working in the future to enhance accuracy and running time.
- Q5: How do you think this paper you read is important for you?
- Opens a possibility of using ontologies in our work to increase the accuracy of context detection and language semantics.

9. Machine Learning Algorithms for Natural Language Semantics and Cognitive Computing.[12]

    - Q1: Motivation of the work?
    - Improve learning of natural language semantics in natural language processing.
    - Q2: The main problem statement of the work?
    - Existing machine learning algorithms do no perform well in learning natural language semantics.
    - Q3: How the researchers contributed to solve the problem?
    - They developed some algorithms to calculate semantics in language.
    - Q4: What main results the researchers reach?
    - They developed an elegant algorithm that uses a new paradigm to learn semantics in an effective way.
    - Q5: How do you think this paper you read is important for you?
    - Learning semantics is a huge part of what we are trying to do so being exposed to such new and effective algorithms can be very helpful.

10. A Suite of Tools for Arabic Natural Language Processing: A UNL Approach.[13]

    - Q1: Motivation of the work?
    - To introduce a new UNL framework, that claims to be able to analyze automatically natural languages into their abstract semantic meanings.
    - Q2: The main problem statement of the work?
    - Most of the Arabic resources available handle the Arabic language morphologically and some syntactically, none are devoted to analyzing Arabic sentences semantically.
    - Q3: How the researchers contributed to solve the problem?
    - So they created some rules and stages to create a well formed sentence.
    - Q4: What main results the researchers reach?
    - They managed to create a platform that helps linguistic research and to do NLP tasks.
    - Q5: How do you think this paper you read is important for you?
    - It will be a very important reference material in Arabic language process to achieve an accurate semantic meaning of the word.

3.1 Similar System Description

Intragloss is a software that helps interpreters get prepared for conferences and events that they are interpreting. They input the documents they are interpreting and highlight the words that they might find it difficult to interpret and then the software search for the word, highlight it and enter it in a glossary that the user can access afterwards, the final output is a document with highlighted words and their meaning annotated above them.

InterpretBank is a similar software to IntraGloss where you can identify two words from two different documents (two different languages) and enter them in the glossary. Also it finds translations and look up terms and extracts terms from the monolingual preparation document. Lately, they developers of this software have started experimenting with speech recognition to assist the interpreter in translating terminology, numbers and entities, the software is still in development stages and is only available to organizations and institutions with dedicated IT infrastructure.

3.2 Comparison with Proposed System

Our software is an integrated note-taking and terminology management system that uses cutting-edge speech recognition technology to perform two tasks: it converts the audio speech into text that is displayed on a computer screen and extracts the used terms to provide their translations after checking them against a database of human translated terms. New utterances are highlighted to create a state of the art interface, moreover there is advanced options for background noise cancellation. Our system differs from the two previous systems as it contains latest technology for speech recognition which is not available in one of them and the other is still using it only for specific organizations and still not fully developed. Also our system will have the glossaries built in by students, there are more than 10,000 English or Arabic terms ready to be automatically translated in different fields from international law to medicine or economics etc...

Domain-specific tasks will be carried out by an artificially intelligent computer system for the purposes of contextual analysis.

The system also can work on both live or recorded speeches and can be used in consecutive interpretation and at sight translation thanks to the feature of automatic note-taking. Thus the interpreter can do things when using our software that cannot be done by similar software like Refer to the transcript of the audio speech to check any missed information (e.g. numbers and proper nouns) especially when having fast speakers, check the meaning of specialized terms instantly without wasting time in manual search. The software help achieves more accuracy and consistency in the translation of terminology among interpreters.

3.3 Screenshots from previous system

4 Project Management

4.1 Tasks and Time Plan

Task	From	To
Information gathering	25/7/2018	12/9/2018
Prototype delivery	12/9/2018	26/9/2018
Proposal evaluation	12/9/2018	26/9/2018
Survey paper	25/7/2018	20/10/2018
Dataset collection (Alsun)	27/9/2018	25/12/2018
Database design	14/9/2018	26/9/2018
Writing SRS document	1/10/2018	30/10/2018
SRS evaluation	30/10/2018	30/10/2018
Writing SDD document	1/1/2019	12/2/2019
SDD evaluation	12/2/2019	12/2/2019
System implementation	TBD	TBD
Implementation evaluation	TBD	TBD
Validation & testing	TBD	TBD
Preliminary preparations	TBD	TBD
Final presentation	5/5/2019	5/5/2019

Table 1: Time plan for Fall 2018/Spring 2019

The above table showcases our time table for the 2018/2018 academic year. Any dates that haven't yet been decided by the faculty board have a placeholder value.

4.2 Budget and Resource Costs

The use of IBM's Watson engine for speech-to-text transcription is subject to a paid premium plan amounting to $0.02 USD per thousand characters.

4.3 Supportive Documents

No surveys or reports have been carried out at the time of writing this text.

References

[1] World Congress on large dams, 26-30 May 1997; E.C.C.O. Congress - European Confederation of Conservator-Restorers' Organizations, 29-31 May

1997; sterreich, Italien, Slowenien. Drei Kulturen der ffentlichen Verwaltung im europischen Vergleich 24-26 April 1997.

[2] Krug S.E. Laughlin J.E. (1976): CDQ - IPAT Depression Scale - Institute for Personality and Ability Testing, Champaign, Ill.

[3] Stallard, David, et al. "Evaluating different confirmation strategies for speech-to-speech translation systems." Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on. IEEE, 2010.

[4] Anastasiou, Dimitra. "Speech-to-speech translation in an assisted living lab." Proceedings of the 4th International Conference on PErvasive Technologies Related to Assistive Environments. ACM, 2011.

[5] Chen, Wei, et al. "ASR error detection in a conversational spoken language translation system." Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on. IEEE, 2013.

[6] Jones, Douglas, et al. "Measuring human readability of machine generated text: three case studies in speech recognition and machine translation." Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on. Vol. 5. IEEE, 2005.

[7] Bos, Johan, and Katja Markert. "Recognising textual entailment with logical inference." Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2005.

[8] Elnagar, Ashraf, Omar Einea, and Leena Lulu. "Comparative Study of Sentiment Classification for Automated Translated Latin Reviews Into Arabic." 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA). IEEE, 2017.

[9] Jusoh, Shaidah, and Hejab M. Alfawareh. "Automated Translation Machines: Challenges and a Proposed Solution." 2011 Second International Conference on Intelligent Systems, Modelling and Simulation. IEEE, 2011.

[10] Nadkarni, Sharvari, et al. "Semi natural language algorithm to programming language interpreter." Advances in Human Machine Interaction (HMI), 2016 International Conference on. IEEE, 2016.

[11] Okba, Kazar, et al. "Semantic natural language translation based on ontologies combination." Information Technology (ICIT), 2017 8th International Conference on. IEEE, 2017.

[12] Khan, Emdad. "Machine Learning Algorithms for Natural Language Semantics and Cognitive Computing." Computational Science and Computational Intelligence (CSCI), 2016 International Conference on. IEEE, 2016.

[13] Alansary, Sameh, Magdy Nagi, and Noha Adly. "A suite of tools for Arabic natural language processing: A UNL approach." Communications, Signal Processing, and their Applications (ICCSPA), 2013 1st International Conference on. IEEE, 2013.