TAMIL TEXT TO ENGLISH SPEECH CONVERSION WITH UNIVERSAL NETWORKING LANGUAGE FOR TRANSLATION

INTRODUCTION

This project in the field of Natural Language Processing aims at translating an input Tamil sentence into an equivalent ***spoken*** *English translation of the sentence. This brings together two major domains in NLP,i.e, machine translation and TTS(Text to Speech). Natural Language Processing is the field of computer science that deals with interactions between computer and human languages. One of the aim of NLP is "Natural Language Understanding", i.e., to enable computers to derive meaning from human language input. Computational linguistics is a sub domain under NLP, which focuses specifically on language related work such as language modelling, representation. Machine Translation is one such field of computational linguistics. Machine Translation involves the development of software for translation of given text from a source language to the target language.*

PROBLEM STATEMENT

Given a Tamil sentence as input, the project will translate it into its corresponding English sentence with no loss in meaning while preserving the semantic structure of English.The translated sentence would be voiced over with a humanised robotic voice. Given a batch of sentences (document input), each sentence would be processed and translated and synthesised into speech.

PROJECT DESCRIPTION

The first phase of the project i.e,conversion of Tamil input text into English is done by converting Tamil text into the Universal Networking Language representation and then converting from UNL into the equivalent English form. Here UNL is used as the pivot language in between the Tamil and English languages. Since UNL is language-independent, it offers a very flexible platform for the representation of knowledge and thus gives great scope for research, in conjunction with the rich language of Tamil, that we have undertaken in our project.

The second phase of our project takes the converted English text and processes it with a Text-To-Speech Engine, to convert it into speech. This TTS system is modified at signal level so as to incorporate emotion, to make the robotic speech more humanised. The given text is transformed into speech with the necessary prosody,stress and intonation marked so as to add emotion to the text.

RELATED WORKS

1.TRANSLATION SYSTEM

Translation is one domain of computational linguistics where many different approaches have been formulated. Some of these approaches are :

- ➔ *Rule based translation*
- ➔ *Phrase based translation*
- ➔ *Example based translation*
- ➔ *Context based Translation*
- ➔ *UNL approach*

Rule-Based Machine Translation (RBMT):

RBMT is based on rules from source and target languages and it depends on the morphology and grammar of those languages for translation. *{ Dugast, J. Senellart, and P. Koehn. 2007. Statistical Post-Edition on SYSTRAN Rule-Based Translation System, In Proceedings of the Second Workshop on Statistical Machine Translation, Prague, Czech Republic.}*

The main steps in rule based translation are

1) *Analyse the input sentence grammatically (Tokenization and parsing)*

2) *Generating equivalent sentence in target language in accordance with target grammar. (Word mapping and Semantics)*

Components of a rule based translator are:

1) *Source language analyser*
2) *Source language parser*
3) *Translator (lookup between source and target language)*
4) *Target language generator*
5) *Target language parser.*

SYSTRAN, one of the very old and successful machine translation system works on rule based translation. *{ MBA Salai Aaviyamma and K. Kathiravan, **"Problems related to Eng-Tam Translation"**, In Proceedings of the International Forum for Information Technology in Tamil, pp. 169–172, 200 9.[2]}*

In general RBMT is not complete by itself as its a very rigid system which relies completely on the knowledge base(semantics of source and

target language). The system has to be trained to increase its efficiency.

PHRASE BASED TRANSLATION

In phrase based translation, the input is translated by grouping together words into phrases. So phrases are mapped from the source language to the target language, which is an improvement over the traditional word to word mapping.

*This paper {Philip Koehn, Franz Josef Och, Daniel Marcu –**"Statistical Phrase-based translation"** NAACL '03 Volume 1[3]} discusses the statistical approach to phrase based translation. This paper shows that extremely high performance can be achieved by heuristic learning and lexical weighting of phrase translations.*

*{Shrikanth Narayanan Sridhar, Vivek Kumar Rangarajan and Srinivas Bangalore, **"Enriching spoken language translation with dialog acts"**, In Proceedings of Association for Computational Linguistics, Short Papers (Companion Volume), pp. 225–228, 2008.[4] }*

The performance of a phrase based translation system can be further enhanced by incorporating dialog acts. The dialog act is primarily used to characterize the nature of the sentence. It is capable of giving the additional information of how some data has been presented, and not just the contents of the data. In this case, every utterance is tagged with a DA label (made possible by a DA Tagger). After having completed the phrase based translation, these Dialog Acts are used to further enhance the performance. Farsi English, Japanese English and Chinese English have been translated using this approach. BLEU scores were improved from +1 to +4 upon addition of Dialog Acts.

*This paper { David Chiang, **"A hierarchical phrase-based model for statistical machine translation"**, In Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, pp. 263–270, 2005.[5] } proposes a model that makes use of hierarchical phrases-phrases that contain sub phrases. In the traditional phrase based model, words are reordered but here phrases as a whole are logically reordered. This reordering can be realised by developing rules within the framework of a Context Free grammar. Mandarin to English translation was performed using this approach and the BLEU scores were found to have improved by 7.5% over the standard phrase based translation system.*

*The paper {Bhattacharyya P. Hegde J. Shah R. M. Ramanathan, A. and M. Sasikumar, **"Simple syntactic and morphological processing can help English-Hindi statistical machine translation"**, In Proceedings of International Joint Conference on Natural Language Processing, pp. 513–520, 2008.[6]} brings in few simple extensions to the phrase based translation system by integrating certain morphological and syntactic features. In the English to Hindi translation, additional processing is done as follows:*

a)The English input sentence is reordered according to the Hindi syntax by applying simple transformation rules on the English parse tree.(syntactic processing)

b)A suffix separation program is used for making use of Hindi suffixes.(morphological processing)
For a morphologically rich language like Hindi, this approach is suitable which is evident in the fact that translation fluency for English to Hindi translation is observed to have increased from 10% to 46%.

EXAMPLE BASED TRANSLATION SYSTEM

Humans learn to translate by way of examples. The example based translation approach is also modelled on this. It tries to draw similarity between parts of text at the word or stem level. Due to this mapping, an extensive bilingual or parallel corpus is required. As it is rather difficult to build such corpora, we make use of the world wide web that can be thought of a large corpus of data. The example based approach involves 3 steps:

MATCHING

The corpus is searched for a sequence of words in the source text and if there is a match, its score is incremented. i.e, greater the number of matches of the source text, greater will be its score. Those word sequences that have a score lesser than some threshold value are discarded from the corpus.

TRANSFER

Once the word sequence has been identified, the corresponding sequence is extracted from the target language. The total score is to be calculated as:

TOTAL= TRANSLATION SCORE X MATCH SCORE

RECOMBINATION

All translated word sequences are concatenated and the best N possibilities of recombinations which give the entire translated sentence are chosen.

*This paper {Y. Choueka Bar, Kfir and N. Dershowitz, "**An Arabic to English example - based translation system**", In Proceedings of Information and Communication Technologies International Symposium and Workshop on Arabic Natural Language Processing, pp. 355–359, 2007.[7]}*

discusses the application of the example based approach for Arabic to English translation. But this approach has few drawbacks as it gives way for many exceptions to arise, for which new rules need to be formulated.

{ Yves Lepage and Etienne Denoual, ***"Purest ever example-based machine translation: Detailed presentation and assessment",*** *Machine Translation, vol. 19, num. 3, pp. 251–282, 2005.[8] }*

Such a pure example based approach is feasible only by adopting proportional analogy.

UNL BASED APPROACH:

Universal networking language { Tarcisio Della Senta Hiroshi Uchida, Meiying Zhu, Universal Networking Language, UNDL Foundation, Tokyo, Japan, 2005.[9]} is a declarative formal language specifically designed to represent semantic data extracted from natural language texts. It can be used as a pivot language in interlingual machine translation systems or as a knowledge representation language in information retrieval applications.

UNL expresses information and knowledge in the form of semantic network. The semantic network of the UNL is a directed graph. Its nodes are called Universal Words.

There are 46 relations in the UNL, such as 'agt' (agent), 'gol' (goal/final state) , 'obj' (object), etc. They are used to connect every two UWs to construct the semantic networks of UNL Expressions.

Translation from a source language to target language involves converting source language to UNL (Enconversion) and then converting UNL structure to the target language (Deconversion)

*A UNL based MT system for conversion from Punjabi to HIndi { **"UNL Based Machine Translation System for Punjabi Language" ,** Prateek Kumar ,DCSE,Tharapur University, Patiala.} has achieved a BLEU score of 0.72*

BLEU (Bilingual Evaluation Understudy) is an algorithm for evaluating the quality of machine- translated text. It is used to determine how close the machine translated output is to that of a human translated one.

Context Based Machine Translation

This method is similar to example based approach but does not require a parellel corpus like the other one.It is effecient for languages , where translation is dependant on the context rather than rules. A bilingual dictionary, source language corpus and target language corpus are maintained.N-grams are identified from a given text and the words in each of these N-grams are converted to target language with the help of the bilingual dictionary.More than one n-gram can be generated in the target language for a single source language text.These N-grams are then ranked to find the best possible translation.Finally a decoder produces the translated output. Spanish to English, Arabic to English and Chinese to English translations have been experimented using this CBMT {Jaime G Carbonell, Steve Klein, David Miller, Mike Steinbaum,Tomer Grassiany, and Jochen Frey, "Context-based machine translation", The Association for Machine Translation in the Americas,pp. 19–28, 2006.} and they have attained a BLEU score of 0.6950.

As a part of our project we will be working with UNL, which is one of the foremost applications of the **Interlingual** approach. It falls under the **hybrid translation** from the source language to the UNL framework(called enconversion).

2. PROSODY GENERATION IN TEXT TO SPEECH SYSTEM

TEXT TO SPEECH SYSTEM

*A **text-to-speech (TTS)** system converts normal language text into speech. This synthesis of speech from text can be carried out by the following approaches:*

- *Concatenative synthesis*
- *Formant synthesis*
- *Articulatory synthesis*
- *HMM based synthesis*
- *Sinewave synthesis*

***Concatenative synthesis**: It is based on the concatenation (or stringing together) of segments of recorded speech.*

***Formant synthesis :** The synthesized speech output is created using additive synthesis and an acoustic model.Parameters such as fundamental frequency, voicing, and noise levels are varied over time to create a waveform of artificial speech.*

***Articulatory synthesis**: It refers to computational techniques for synthesizing speech based on models of the human vocal tract and the articulation processes occurring there.*

HMM-based synthesis *:HMM-based synthesis is a synthesis method based on hidden Markov models.*

Sinewave synthesis: *Synthesizes speech by replacing the formants (main bands of energy) with pure tone whistles.*

MaryTTS:

MaryTTS is an open-source, multilingual Text-to-Speech Synthesis platform written in Java. It was originally developed as a collaborative project of DFKI's Language Technology Lab and the Institute of Phonetics at Saarland University. It is now maintained by the Multimodal Speech Processing Group in the Cluster of Excellence MMCI and DFKI.

Flite:

*Flite (festival-lite) is a small, fast run-time synthesis engine developed at CMU and primarily designed for small embedded machines and/or large servers. { Min Chu, Yao Qian, **"Locating Boundaries for Prosodic Constituents in Unrestricted Mandarin Texts"**, In International Journal for Computational Linguistics and Chinese Language Processing, Vol. 6, No. 1,pp 61-82(2001) }*

A TTS can be more natural if prosodic features are also incorporated in it. Prosodic components have to be extracted from text and methods which use Hidden Markov models are found to be more successful.

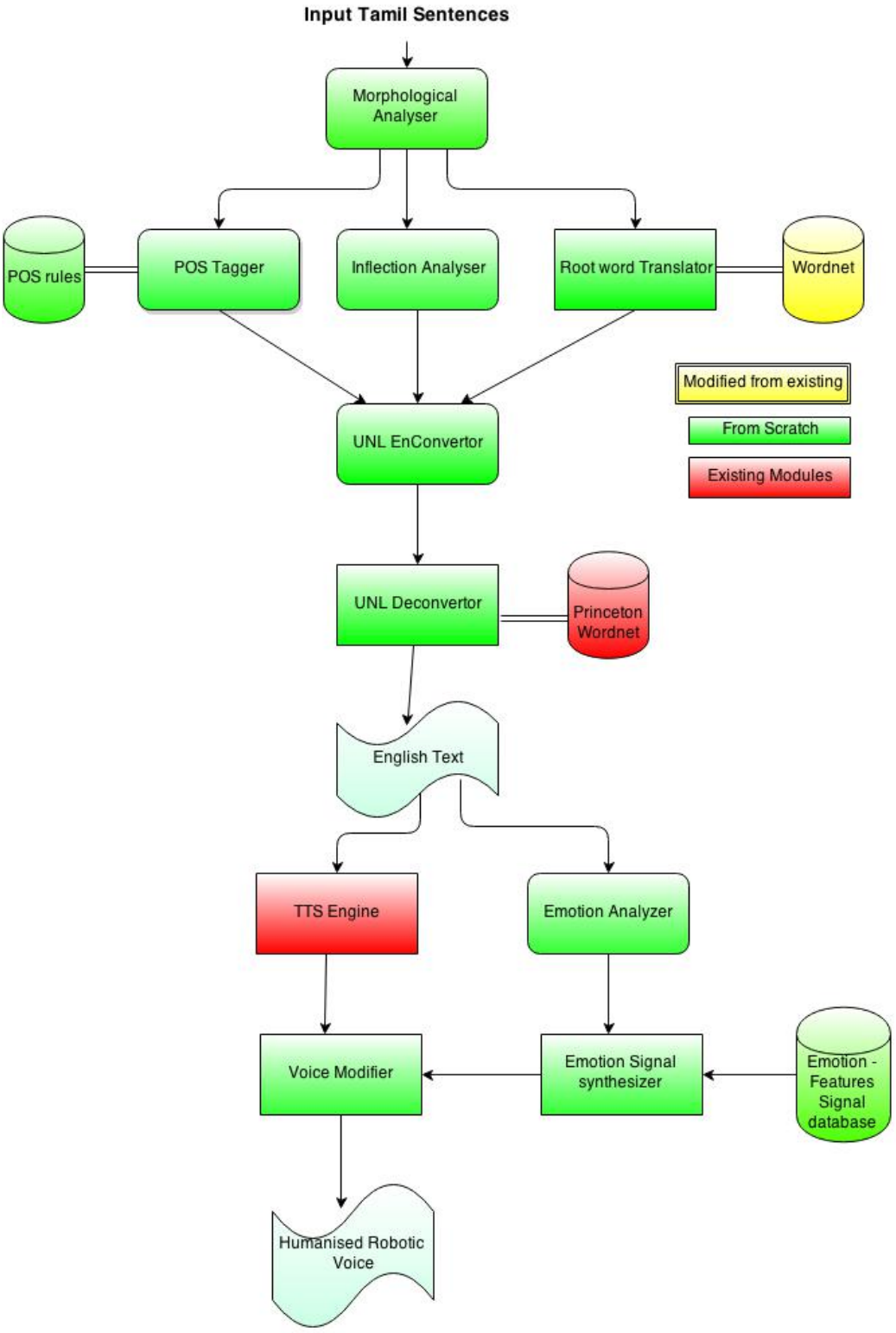
*This paper { Sudhakar Sangeetha, Sekar Jothilakshmi, "**Syllable based text to speech synthesis system using auto associative neural network prosody prediction**", In International Journal of Speech Technology, Vol 17, Issue 2, pp 91-98.[10] } describes a successful attempt to detect prosody from text using five layer auto associative neural network. The prosodic features extracted from the text will be used for the prosody generation in speech.*

Human speech varies depending on the variation of pressure in vocal chord. Different acoustic factors like pitch, amplitude affect the voice.

*The paper { Tristan Bowles, Sandra Pauletto, (2010). "**Emotions in the Voice: Humanising a Robotic Voice.**" In: Proceedings of the 7th Sound and Music Computing Conference, Barcelona, Spain.[12] } states states that for monosyllabic or short words, the duration of each word was greatly reduced in Anger and happiness. And sadness has considerably increased the duration of words. In terms of pitch, anger and happiness have high fundamental frequency peaks (225hZ+) and sadness has relatively low peak(140Hz). Similar results are obtained from amplitude analysis, anger and happiness topping with 100+dB and sadness with <95dB. So in general a person is more loud and speaks faster if he anger and very slow if he is sad.*

Proposed System

BLOCK DIAGRAM : TAMIL TEXT TO ENGLISH SPEECH CONVERSION USING UNL



MODULES:

OFFLINE MODULES

- *WORDNET CONSTRUCTION:*

The condensed wordnet is built with words from tamildict.com and an exisiting wordnet [13] organized based on the parts of speech of English words from Princeton database. A tree like data structure is used where the word is the root and edges named noun, verb leading to the respective forms with English translation.

- *EMOTION -FEATURE SIGNAL DATABASE*

Voice corpus from English films are split based on emotions and various features[14] like pitch, frequency and other mel-frequency ceptral components are tabulated for each input. These parameters for each emotion is determined.

ONLINE MODULES :

- *MORPHOLOGICAL ANALYSER:*

Tamil is a morphologically rich language. A number of words can be formed from a root word by inflections. But only the root word will be mapped for translation and features like tense, gender and plurality can be extracted from the inflections which will be used in Deconversion to target language.A morphological analyser has been developed for this purpose based on the rules[15].

- ***POS TAGGING:***

Tamil being a partial free word order language, the inflections are needed to find the part of speech[16] of that word. This tagging is essential for UNL graph conversion.For eg. Inflictions such as ‘ aye’ added to words , denote that they are used as objects in that sentence.

Eg. Kaiyai pidithaan

- ***UNL ENCONVERSION AND SUBGRAPH IDENTIFICATION:***

Based on the root word and its POS, the equivalent English word is decided and the tense previously determined by the analyser are used to construct the UNL graph.

Using rules in the [17] , the UNL graph is constructed.Based on punctuations and clause rules[18] , the subgraphs are identified to be processed together

- ***UNL DECONVERSION:***

Conversion from the UNL structural language to the final target language, i.e, English is called UNL deconversion.[19]

- ***EMOTION IDENTIFICATION:***

This phase involves the detection of emotion from the text [11]. The emotion extracted from the text will be used for the prosody generation in the speech.

- ***VOICE MODIFICATION and OUTPUT:***

Achieved by varying parameters like pitch, amplitude, word

speech rate of speech in accordance with values from the emotion-features database corresponding to the emotion identified in the previous step.[12] In general, happy phrases have higher frequency and amplitude than sad phrases.

Evaluation Parameters:

The translation subsystem and prosody subsystem will be evaluated independently.

For translation subsystem , BLEU score will be used for accuracy measure.And a manual survey will be used to judge the prosody subsystem.

REFERENCES:

[1]Dugast, J. Senellart, and P. Koehn.2007. ***"Statistical Post-Edition on SYSTRAN Rule-Based Translation System",*** *In Proceedings of the Second Workshop on Statistical Machine Translation, Prague, Czech Republic.*

[2]MBA Salai Aaviyamma and K. Kathiravan, ***"Problems related to Eng-Tam Translation",*** *In Proceedings of the International Forum for Information Technology in Tamil, pp. 169–172, 2009.*

*[3]Philip Koehn, Franz Josef Och, Daniel Marcu –"****Statistical Phrase-based translation****" NAACL '03 Volume 1*

[4]Shrikanth Narayanan Sridhar, Vivek Kumar Rangarajan and Srinivas Bangalore, ***"Enriching spoken language translation with dialog acts",*** *In Proceedings of Association for*

Computational Linguistics, Short Papers (Companion Volume), pp. 225–228, 2008.

[5]David Chiang, ***"A hierarchical phrase-based model for statistical machine translation"****, In Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, pp. 263–270, 2005.*

[6]Bhattacharyya P. Hegde J. Shah R. M. Ramanathan, A. and M. Sasikumar, ***"Simple syntactic and morphological processing can help English-Hindi statistical machine translation"****, In Proceedings of International Joint Conference on Natural Language Processing, pp. 513–520, 2008.*

[7]Y. Choueka Bar, Kfir and N. Dershowitz, ***"An Arabic to English example - based translation system"****, In Proceedings of Information and Communication Technologies International Symposium and Workshop on Arabic Natural Language Processing, pp. 355–359, 2007.*

[8]Yves Lepage and Etienne Denoual, ***"Purest ever example-based machine translation: Detailed presentation and assessment",*** *Machine Translation, vol. 19, num. 3, pp. 251–282, 2005.*

[9] Tarcisio Della Senta Hiroshi Uchida, Meiying Zhu, Universal Networking Language, UNDL Foundation, Tokyo, Japan, 2005.

[10] Prateek Kumar ,DCSE,Tharapur University, Patiala, ***UNL Based Machine Translation System for Punjabi Language.***

*[11]Sudhakar Sangeetha, Sekar Jothilakshmi, **"Syllable based text to speech synthesis system using auto associative neural network prosody prediction"**, In International Journal of Speech Technology, Vol 17, Issue 2, pp 91-98.*

*[12]Tristan Bowles, Sandra Pauletto, (2010). **"Emotions in the Voice: Humanising a Robotic Voice."** In: Proceedings of the 7th Sound and Music Computing Conference, Barcelona, Spain.*

*[13] Jayavasanth R ,Rajeswari Sridhar **"A hybrid approach to construction of Wordnet "** In : INFITT -2014 , Pondicherry.*

*[14]Pierre-Yves Oudeyer , **"The production and recognition of emotions in speech: features and algorithms" ,** In : Int. J. Human-Computer Studies 59 (2003) 157–183*

*[15] Anand Kumar M., Dhanalakshmi, Dr. S. Rajendran, Dr. K. P. Soman, **"A Novel Approach to Morphological Analysis for Tamil"***

*[16] Anand Kumar M., Dhanalakshmi, Dr. S. Rajendran, Dr. K. P. Soman, **"POS Tagger and Chunker for Tamil Language"**.*

*[17] J Balaji, T V Geetha, Ranjani Parthasarathi, Madhan Karky, **"Morpho-Semantic Features for Rule-based Tamil Enconversion"**.*

*[18] Balaji J, Geetha T V, Ranjani Parthasarathi, **"Semantic Parsing of Tamil Sentences"**.*

[19] Pelizzoni J M, Nunes M V “ Flexibility, ***Configurability and optimality in UNL deconversion via multiparadigm programming. Universal Network Language: Advances in Theory and Applications”****, 2005, Research on Computing Science: 175-194.*