Design and Development of a Frame Based MT System for English-to-ISL

Anuja.K, Suryapriya.S, Sumam Mary Idicula
Department of Computer Science
Cochin University of Science & Technology
Cochin – 682 022, Kerala, India
anujakarthikeyan@gmail.com,suryasumathy@gmail.com,sumam@cusat.ac.in

Abstract—**This paper presents the design and development of a frame based approach for speech to sign language machine translation system in the domain of railways and banking. This work aims to utilize the capability of Artificial intelligence for the improvement of physically challenged, deaf-mute people. Our work concentrates on the sign language used by the deaf community of Indian subcontinent which is called Indian Sign Language (ISL). Input to the system is the clerk's speech and the output of this system is a 3D virtual human character playing the signs for the uttered phrases. The system builds up 3D animation from pre-recorded motion capture data. Our work proposes to build a Malayalam to ISL**

Keywords-machine translation; Indian sign language; language processing; motioncapture; 3D virtual human

I. INTRODUCTION

Deaf people commonly use sign language as their communication medium and other natural languages are secondary for them. Sign language used by Indian deaf community are collectively called Indian Sign Language (ISL). Different dialects of ISL exist in different parts of India like Delhi dialect, Madras dialect etc. In the 1981 census, the hearing disabled age 5+ population was estimated to be 6,315,761 in India, which is 0.92% of the total population [9]. Later census did not include deaf population and hence no accurate information is available. Gopinath estimated the 1991 deaf population as 7,770,753 by extrapolating from the 1981 census deaf population and 1991 total census population [10]. The 2008 total population is estimated to be 1.147 billion (CIA the world fact book) and the deaf population approximately 10.55 million. In 2005 Zeshan et al. estimated the complete deaf population as over one million and hard of hearing population approximately 10 million. This figure was extrapolated from the number of deaf and hard of hearing people in western nations [11]. According to this calculation the figure goes still high up.

It is difficult for deaf people to access the information because of their language problem and hence this hinders their normal social life. Some efforts have been started in this area to aid the deaf people for communication in public places. In India, linguistic studies on sign language started in 1978 and the studies concluded ISL is a language on its own with specific syntax, grammar, phonology and morphology [1][2].

Several works have been done on sign languages like American Sign Language (ASL), British Sign Language (BSL) [12][13][14]. But in ISL, only limited works have been done in this area [7][8]. Most of the Machine Translation (MT) systems use one of the following three basic approaches. The Fig. 1 shows the classic MT pyramid [15].

A. Direct translation

This MT system is designed specifically for a particular pair of languages. This ignores the syntax, semantics, and morphology of the language. The out put may not be desired one. This produces an out put with direct replacement for corresponding words.

B. Transfer-based Architecture

This system is based on the specification of transfer rules for morphology, syntax, lexical selection, semantics analysis and generation. In transfer-based MT, it has some dependence on the language particular pair which needs the addition of new transfer rules for the new source or target languages.

C. Interlingua-based Architecture

In this approach, the source language is transformed into an interlingua, an abstract language-independent representation. The target language is then generated from the interlingua. This saves work with respect to direct translation/transfer based if there are several target languages whose modules are well enough developed.
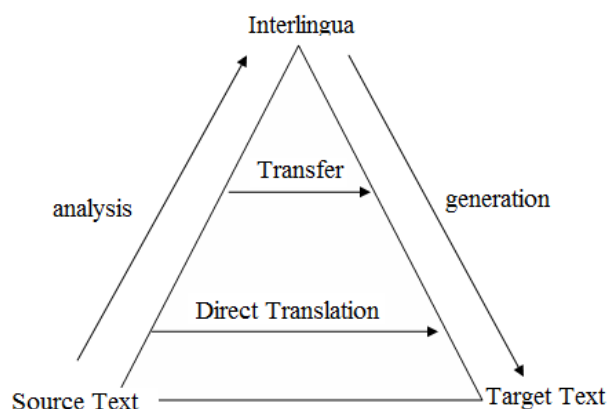

Figure 1. MT Pyramid

Deaf people find it difficult to interact with people in common place as their medium of communication is different. Our system aims to aid the deaf people in such situations. A system to translate the input speech to ISL which is displayed with the help of a 3D virtual human-avatar has been proposed. System uses a limited domain of railways and banking, since in these places the amount of interaction with common people is considerably large.

The organization of the paper is as follows: Section 2 presents an outline of ISL Linguistic issues, Section 3 presents an outline of the proposed system, input and output of the system and rules identified for the system, Section 4 presents the architecture of the proposed system, Section 5 presents conclusion of our work and Section 6 draws out the scope of future work in this area.

II. INDIAN SIGN LANGUAGE LINGUISTICS

ISL is a visual /spatial language; it uses hands and arms, facial expressions, body postures, eye gaze etc to encode the linguistic information. To convey the meaning, a signer often makes use of a 3D space around him/her which is called Signing space. ISL consists of manual as well as non-manual components. The manual components include hand shape, hand location, orientation, movements and the non-manual components include facial expression, eye gaze and head/body posture.

There exist some signs which contain manual/non-manual components only. ISL incorporates non manual components in negation, affirmation, interrogatives, and conditional clauses. Such signs include the word "Yes" which is signaled by a vertical head movement and the word "No" which is signaled by a side to side head movement. The word ordering in ISL is different from English and most of other natural languages. ISL follows a participant first and predicate last rule. There are no temporal inflections in ISL. The signs for 'finish', 'later' etc. are used to show the te se. e.g.: 'I had dinner' will be treated as 'I night food finish'. Gender is shown by adding the signs for man and woman.

ISL have minimal paradigm for interrogative signs which is used along with other non-interrogative signs to convey the question. ISL differ from other sign languages with the position of wh-signs in the sentence. It can be placed in the beginning, at the end or both of the sentences i.e. when wh signs are doubled. ISL mainly uses single non-compositional wh-sign to indicate what, why, where, when and how are shown using same sign which is indicated in the Fig 2 [16].

Signer can place various objects in space and refer to these. ISL use indexical signs to realize pronouns, demonstrative and auxiliary by pointing at various locations in sign space. Pronouns are indicated by the extended index finger when it is used anaphorically. This is also used to for reference in sign space as a proximal demonstrative "here" and distal demonstrative "there" with a flat B hand shape. Auxiliary is shown by pointing at two locations and a transitional movement between the two [3][5][6]. This spatial nature of ISL is a particular challenge to the use of traditional computational linguistic approaches for ISL translation task.

Figure 2. Wh-ISL sign

III. SYSTEM ARCHITECTURE

The System Architecture shown in Fig 3 consists of the following modules.
- Speech Recognition Module
- Language Processing Module
- 3D Animation Module

The input to the system is the clerk's speech which is in English. The speech recognition module recognizes the speech and makes a text output. This text is subsequently passed to a parser module which tokenizes the string and tag the part-of-speech using a sample file. The output from the parser is given to an eliminator module which performs a reduction task by eliminating unwanted elements and further the root form of verbs are found using the stemmer module. The structural divergence of English and ISL is handled by a phrase reordering module using ISL dictionary and rules. This module generates ISL-gloss strings which can be played back through the 3D virtual human. A 3D animation module creates the animation from the motion captured data.

A. Input and Output of The System

The input to the System is the clerk's Speech which is English. The output is the corresponding ISL animation.

B. English-ISL structure conversion

ISL follows a different word order from English and this is illustrated in Fig 4. Therefore for the translation process, the words needs to be ordered based on the ISL syntax and grammar. This Structural divergence is analyzed by tagging the part of speech for English sentence and corresponding ISL gloss. Example is shown below which follow Penn Tree Bank tag notation [17]. The tags used in the examples are explained in the table below.

TABLE I. TAG DESCRIPTION

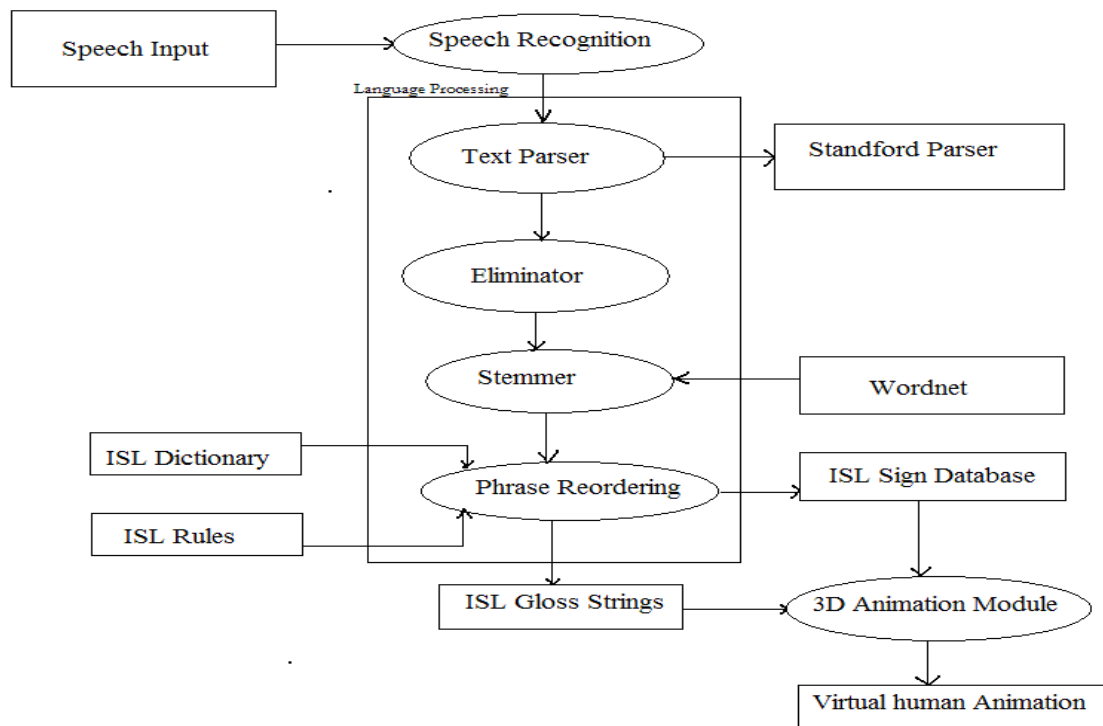| Key word	Abbreviation
Personal pronoun	PRP
Modal	MD
Verb, base form	VB
Determiner	DT
Noun, Singular or mass	NN
Wh-adverb	WRB
Verb, non-3rd person singular present	VBP
Verb, gerund or present participle	VBG
Verb, 3rd person singular present	VBZ
Verb, past participle	VBN

Figure 3. The System Architecture

I will give you a form - PRP MD VB PRP DT NN
I form give you -   PRP NN VB PRP

You fill the form – PRP VB DT NN
You form fill -   PRP NN VB

Where are you going – WRB VBP PRP VBG
You go where -   PRP VB WRB

Train is delayed – NN VBZ VBN
Train delay -   NN VB

Rules needed to generate the ISL equivalent sentence for texts are generated by analyzing the above tagged phrases and the ISL equivalent.
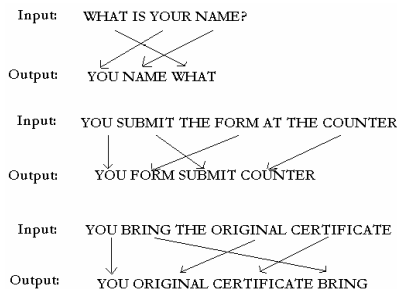

Figure 4. Analysis of Structural divergence between English and ISL

C. Motion capture

Domain related frequently used phrases and queries are identified and a video database of the signs for these phrases is created with the help of native ISL signers. This sign database has been used for motion capture unit where a motion capture tool, ShapeHandPlus is used.

Motion Capturing is a technology for digitally recording a real actor motions. The motion capture software records and calculates the positions, angles, velocities, accelerations and impulses, providing an accurate digital representation of the motion. For each motion sequence we capture, there are three different phases to be carried out: rehearse, record, and review. The motion capture device samples motion and records it.

IV. SYSTEM COMPONENTS

A. Speech Recognition Module

Input to this module is the clerk's speech and each utterance is converted to corresponding text. For this, QPointer, continuous speech recognition system, is used. This recognizes utterances formed by several words continuously spoken. The recognition system can generate one optimal word sequence from the given acoustic and language models. It is having voice data of the person, who has trained the software. The second source of knowledge included in a speech recognizer (besides the acoustic model)

is the language model. This model complements the acoustic knowledge with the information about the most probable sequences of words.

B. Language Processing Module

Language processing module consists of

- Input Text Parser
- Eliminator
- Stemmer
- Phrase Reordering & ISL gloss generation

1) Input Text Parser

System accepts transcribed spoken language strings as input which needed to be parsed. System uses an API for Stanford parser which is a full-fledged parser developed in java by Stanford University. This parser accepts the text as input and generates a typed dependency structure and phrase grammar structure. As an example System takes input:

Where are you going?

Stanford Parser produces an output for this in phrase grammar structure as the following

```
[ROOT](
    [SBARQ](
        [WHADVP](
                [WRB]([where])
        ) +
        [SQ](
            [VBP]([are])+
            [NP](
                )+  [PRP]([you])
            [VP](
                [VBG]([going])
            )
        )
    )
)
```

This is further refined in to a the following tagged form
Where<WRB> are<VBP> you<PRP> going<VBG>

2) Eliminator

Output from the parser is further given to an eliminator module. This module uses the ISL dictionary to remove the unwanted tokens from the parsed text .This is important as ISL does not have corresponding signs for all English words. The output from the eliminator module for the previous example is

Where you going

3) Stemmer

In the next step, the language processing module performs stemming on the identified tokens. This step is carried out to match another ISL linguistic property; ISL uses simple present form for the verbs. As there is no sign for the forms of verbs other than simple present, these are converted to the root form. Out put from this stage is as follows

[Where, you, go]

4) Phrase Reordering & ISL gloss generation

The last stage of language processing module involves phrase reordering. To match the structural divergence between the source and target language the rules are mapped into a slot structure and token are filled in to the right slots. This module generates ISL gloss strings or a pattern which in turn is displayed to the user as an animation. Output from the phrase reordering module is as follows.

You go where

A frame based approach is used for this translation procedure, where slot value is filled as showed in the Fig 5.

C. 3D Animation Procedure

ISL does not have a written representation. This ISL gloss has to be converted into an animation for the user. For the animation purpose, a character is created using Autodesk Inc's Maya software. The motion captured signs are mapped onto the character and the character is animated using the software. The output of the animation is in a video format which is played back to the user.

1) 3D Virtual Human Character

Character modeling is done using Autodesk Inc's Maya Software. First a model of character is created using this software. Then a character rig is built to move portions of a model-such as arms, legs. Rigging is used to give structure and continuity to a set of geometric shapes that would otherwise have no connection. For building a skeleton in Maya, two inbuilt structures- Bones and joints are used. Maya will connect the joints with a visual reference called a bone. Joints are the only element in a joint/bone pair that can be selected and manipulated.

For the skeleton hierarchy to affect geometry, the character is skinned by binding elements of the geometry to the joints in the skeleton.

2) Animation Procedure

The Trax Editor in Maya is used for working with motion capture data. A character based on the actor's body is created. A skeleton that has joints located between opposing sensor positions is constructed. A generic human actor is created in Maya and the motion capture files are imported into the scene. With the motion capture data in T-pose, the actor is matched approximately to the data by translating, rotating, and scaling. The character is rigged in such a way that copies the rotational values of the delivered skeleton onto the skeleton of created character

DT	PRP	JJ	NN	VB	PRP	RB	WRB	WP	CD
	You			Go			Where		
	Your		Name				What		
	You	New	A/C	Open					

Figure 5.   Slot Value Frame

V. Observations

The developed system should convey information in a way helpful and acceptable to deaf users. The extent to which Signing Assistant met this aim was assessed using evaluation of the coverage of phrases in the domain and quality of the signing. Evaluation was done over 250+ phrases identified for the limited domain of banking and railways, some of which were repetition. The evaluation was done with the help of native ISL users.

A. Coverage of Signs

From the identified phrases more than 50% were redundant. Out of 50 distinct phrases, 75 phrases were generated by incorporating days of the week, numbers and time of the day. Since the work has been carried out with a small corpus of sign units, coverage can be improved by adding more sign data.

B. Quality of signing

The quality of 'signing assistant' signing was measured in two ways: accuracy of signs and acceptability of signs to deaf users. 60% generated phrases were interpreted correctly as complete phrase unit while 21% had contained semantic error. 19% of the phrases were incomprehensible. Accuracy of the signs can be improved by adding more semantic rules to the system as well as by giving facial expression and lip movement to the animation. The table given below shows the evaluation rating.

TABLE II.   EVALUATION RESULT

Accuracy of translation	% of phrases
Correct translation as complete phrase unit	60%
Translation with semantic error	21%
Incomprehensible translation	19%

Current system takes a longer time to generate the translation, which should be reduced to a minimum acceptable level. The hand movement and finger positions can be made clearer to help the improved the acceptance.

VI. Conclusion

In ISL, only limited works have been done so far. This work aims to present a speech to ISL translation system which will aid the deaf community of India for interaction in public places like a railway reservation counter and bank. System used a frame based development approach which transfers the rules from English to ISL. System is realized for a limited domain of railways and banking with a small corpus. A large corpus of phrases and analysis will improve the over all performance of the system. System used hand and arm motion capture tool for animation purpose. Intelligibility and acceptance of the signs can be increased by incorporating the facial expressions, lip movement and body posture. This System can be extended to generate the ISL animation synthetically [4].

VII. Future Work

The current system can be used as a baseline for the development of Malayalam-to-ISL MT system. As people in Kerala use Malayalam as first language for communication, development of such a system will enhance the communication.

ISL has almost similar word ordering of Malayalam. But this differs in syntax and morphology. As the first step a study is to be conducted to analyze the degree of structural divergence between these two languages. A unified mapping rule for Malayalam-to-ISL has to be identified.

A Malayalam language processor has to be developed which incorporates

- Lexical Database for Malayalam & ISL
- Morphological analyzer
- Syntax analyzer
- Word sense disambiguation
- WordNet
- Phrase Reordering

A lexical database has to be created for Malayalam which contains all noun and verb root forms and synonyms. For ISL, the lexical database used for the current system can be reused.

A Morphological analyzer has to be developed, which will help the words to be stemmed. It will identify the parts of speech and splits the word if necessary. Syntactic analyzer identifies the noun phrases, verb phrase etc. A word sense disambiguation module has to be incorporated in order to resolve the ambiguous words; this module will refer to a Malayalam WordNet for processing. Finally a phrase reordering module will be used to map between Malayalam to ISL gloss by using already identified Malayalam and ISL rules. This ISL gloss will be converted to a user friendly virtual human animation.

References

[1] M. Vasishta, J. Woodward and S. DeSantis. An Introduction to Indian Sign Language, All India Federation of the Deaf (Third Edition).1998

[2] M. Vasishta, J. Woodward and K. Wilson. Sign language in India: regional variation within the deaf population. Indian Journal of Applied Linguistics 4(2): 66-74,1978.

[3] U. Zeshan. Indo-Pakistani Sign Language Grammar: A Typological Outline. Sign Language Studies-Volume 3, Number2 157-212,2003.

[4] S. Prillwitz, R. Leven, H. Zienert, T. Hamke, and J. Henning. HamNoSys Version 2.0: Hamburg Notation System for Sign Languages: An Introductory Guide, volume 5 of International Studies on Sign Language and Communication of the Deaf. Signum Press, Hamburg, Germany, 1989.

[5] U. Zeshan. Interrogative Constructions in Signed Languages. Crosslinguistic Perspectives Language - Volume 80, Number 1, pp. 7-39, 2004.

[6] U. Zeshan, Sign Language in Indopakistan: A Description of a Signed Language, Amsterdam: John Benjamins, 2000.

[7] P. Kar, M. Reddy, A. Mukherjee, A. M. Raina. INGIT: Limited Domain Formulaic Translation from Hindi Strings to Indian Sign Language. ICON, 2007

[8] T. DasGupta, S. Dandpat, A. Basu. Prototype Machine Translation System From Text-To-Indian Sign Language, 2008.

[9] Government of India, Ministry of Social Welfare: A Summary of the Report on Survey of Disabled Persons, 1981.

[10] C. Gopinath 1998. Sexuality, STD, AIDS and Deaf in India. *Signpost* 2:1–4

[11] U. Zeshan, M. Vasishta, M. Sethna. 2004. Implementation of Indian sign language in educational settings- Volume 15, Number 2, *Asia Pacific Disability Rehabilitation Journal*, pp. 15-35

[12] J. A. Bangham, S. J. Cox, R. Elliot, J. R. W. Glauert, I. Marshall, S. Rankov, & M. Wells. Virtual signing: Capture, animation, storage and transmission – An overview of the ViSiCAST project. IEEE Seminar on Speech and language processing for disabled and elderly people, 2000

[13] T. Veale, A. Conway and B. Collins. The challenges of cross-modal translation: English to sign language translation in the ZARDOZ system. In Machine Translation, Volume 13, pp. 81-106, 1998.

[14] Cox, M. Lincoln, J. Tryggvason, M. Nakisa, M .Wells, M. Tutt, S. Abbott. 2002. Tessa, a system to aid communication with deaf people.*Fifth international ACM conference on Assistive technologies*.

[15] Dorr, P. Jordan and J. Benoit. 1998. A Survey of Current Paradigms in MachineTranslation. Technical Report: LAMP-TR-027/UMIACS-TR-98-72/CS-TR-3961, University of Maryland, College Park, December 1998.

[16] P. Roland and U. Zeshan.Wh-movement and wh-split in Indo-Pakistani Sign Language, *South Asian Languages Analysis XXIII, University of Texas at Austin, Oct. 11th, 2003*

[17] PennTreebank II Tags, http://bulba.sdsu.edu/jeanette/thesis/PennTags.html