6.2 Machine Aided Translation Systems

6.2.1

- **Name of the Technology :** *AnglaHindi (AnglaBharti based Technology) Machine Translation System for English to Hindi*

- **Nature of Technology :** A machine-aided translation system for translation from English to Hindi

- **Level:** (Product / Technology / Sub-system) : Product and Technology for customization of domain.

- **Technical Description of the Technology / Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / Product:**

  Please visit http://anglahindi.iitk.ac.in for details

- **Representative Snapshot / screenshot of the Technology / Product:**

- **Scalability / Portability / Expandability:** The product is scalable with facility for dictionary augmentation and customization of domain. The system can be ported on Linux or Windows platform.

- **Readiness of Transfer of Technology (ToT):** Ready for Technology Transfer.

- **Availability of documentation:** Please see Reference manuals at http://anglahindi.iitk.ac.in

- **Testing of the Product / Technology:** The product is available on-line on web at http://anglahindi.iitk.ac.in for testing and feedback.

- **IPR / Open-source:** IPR belongs to IIT Kanpur

- **Potential beneficiaries:** All Govt./semi-Govt. organizations/Parliament/State Assemblies/ Judiciary where bi-lingual document preparation is mandatory; All commercial organizations (Banking, Pharma, Agro, Health, Tourism etc.) and educational institutions in Hindi belt. NRIs.

- **User-agency tie-up:** End Users feedback sought through on-line implementation of the system.

- **Name and address of the Resource Person:**
  Professor R.M.K. Sinha (rmk@iitk.ac.in)
  Dr. A. Jain(ajain@iitk.ac.in)
  Dept of Computer Science and Engineering
  I.I.I. Kanpur 208016.
  Web : http://ww.cse.iitk.ac.in/users/langtech

6.2.2

◗ **Name of the Technology :** AnglaUrdu (AnglaHindi based) **Machine Translation System for English to Urdu**

◗ **Nature of Technology :** A machine-aided translation system for translation from English to Urdu

◗ **Level: (Product / Technology / Sub-system) :** Product and Technology for customization of domain.

◗ **Technical Description of the Technology / Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / Product:**

Please see Reference manual.

◗ **Representative Snapshot / screenshot of the Technology / Product:**

◗ **Scalability / Portability / Expandability:** The product is scalable with facility for dictionary augmentation and customization of domain. The system can be ported on Linux or Windows platform.

◗ **Readiness of Transfer of Technology (ToT):** Ready for Technology Transfer.

◗ **Availability of documentation:** Please see Reference manuals

◗ **Testing of the Product / Technology:** As AnglaHindi is the core for AnglaUrdu, only additional testing that needs to be done is on correctness of display and Urdu equivalent vocabulary. Testing on both these counts are in progress.

◗ **IPR / Open-source:** IPR belongs to IIT Kanpur and Urdu display/word-processing belongs to CDAC, Pune.

◗ **Potential beneficiaries:** All Govt./semi-Govt. organizations/Parliament/State Assemblies/ Judiciary where Urdu document preparation is mandatory; All commercial organizations (Banking, Pharma, Agro, Health, Tourism etc.) and educational institutions in Urdu speaking belt. NRIs and residents of Pakistan and other countries with Urdu as their first/second language.

◗ **User-agency tie-up:** None

◗ **Name and address of the Resource Person:**
Professor R.M.K. Sinha (rmk@iitk.ac.in)
Dr. A. Jain(ajain@iitk.ac.in)
Dept of Computer Science and Engineering
I.I.I. Kanpur 208016.
*Web : http://www.cse.iitk.ac.in/users/langtech*

6.2.3

◗ **Name of the Technology :** AnuBharti Technology for Machine Translation among Indian Languages.

◗ **Nature of Technology :** A machine-aided translation technology for translation among Indian languages.

◗ **Level: (Product / Technology / Sub-system) :** Technology

◗ **Technical Description of the Technology / Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / Product:**

Please see http://www.cse.iitk.ac.in/users/langtech/anubharti.htm

◗ **Representative Snapshot / screenshot of the Technology / Product:**

Please see http://www.cse.iitk.ac.in/users/langtech/anubharti.htm

◗ **Scalability / Portability / Expandability:** Applicable to any Indian Language

◗ **Readiness of Transfer of Technology (ToT):** Ready for Know-How Transfer.

◗ **Availability of documentation:** Please see http://www.cse.iitk.ac.in/users/langtech/anubharti.htm

◗ **Testing of the Product / Technology:** It is a Know-how tested through a prototype implementation.

◗ **IPR / Open-source:** IPR belongs to IIT Kanpur

◗ **Potential beneficiaries:**

All Govt./semi-Govt. organizations/ Parliament/State Assemblies/Judiciary where bi-lingual/tri-lingual document preparation is mandatory; All commercial organizations (Banking, Pharma, Agro, Health, Tourism etc.) and educational institutions with need for localization. NRIs and countries in SAARC and South East Asia.

◗ **User-agency tie-up:** None.

◗ **Name and address of the Resource Person:**
Professor R.M.K. Sinha (rmk@iitk.ac.in)
Dr. A. Jain(ajain@iitk.ac.in) ,
Dept of Computer Science & Engg
I.I.I. Kanpur 208016.
Web : http://www.cse.iitk.ac.in/usres/langtech

6.2.4

◗ **Name of the Technology :** *HindiAngla (AnuBharti based Technology) Machine Translation Systems form Hindi to English*

◗ **Nature of Technology :** A machine-aided translation system for translation from Hindi to English for simple sentences.

◗ **Level:** (Product / Technology / Sub-system) : Product and Technology for customization of domain.

◗ **Technical Description of the Technology / Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / Product:**

Please see Reference manuals

◗ **Representative Snapshot / screenshot of the Technology / Product**
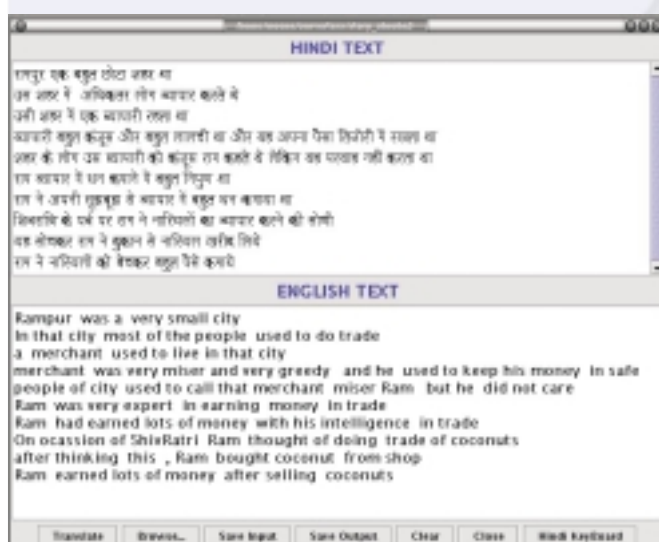
◗ **Scalability / Portability / Expandability:** The product is scalable with facility for dictionary augmentation, example-base augmentation and customization of domain. The system can be ported on Linux or Windows platform.

◗ **Readiness of Transfer of Technology (ToT) :** Ready for Technology Transfer.

◗ **Availability of documentation:** Please see Reference manuals

◗ **Testing of the Product / Technology:** Internal testing done for simple sentences. The correctness have also been verified with reverse translation through AnglaHindi (Please refer to snapshots).

◗ **IPR / Open-source:** IPR belongs to IIT Kanpur

◗ **Potential beneficiaries:** All Govt./semi-Govt. organizations/Parliament/State Assemblies/ Judiciary where bi-lingual document preparation is mandatory; All commercial organizations (Banking, Pharma, Agro, Health, Tourism etc.) and educational institutions in Hindi belt. NRIs.

◗ **User-agency tie-up:** None

◗ **Name and address of the Resource Person:**
Professor R.M.K. Sinha (rmk@iitk.ac.in)
Dr. A. Jain(ajain@iitk.ac.in)
I.I.I. Kanpur 208016.
Web : http://www.cse.iitk.ac.in/users/langtech

6.2.5

◗ **Name of the Technology** : *AnglaBharti Technology for MAT from English to Indian Languages (Machine Translation Systems from English to Indian Languages)*

◗ **Nature of Technology** : A machine-aided translation technology for translation from English to all Indian languages

◗ **Level:** (Product / Technology / Sub-system) : Technology

◗ **Technical Description of the Technology / Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / Product:**

Please see http://www.cse.iitk.ac.in/users/ langtech/anglabharti.htm

◗ **Representative Snapshot / screenshot of the Technology / Product:**

Please see http://www.cse.iitk.ac.in/users/ langtech/anglabharti.htm

◗ **Scalability / Portability / Expandability:** Applicable to all Indian Languages

◗ **Readiness of Transfer of Technology (ToT):** Ready for Technology Transfer.

◗ **Availability of documentation:** Please see http://www.cse.iitk.ac.in/users/ langtech/anglabharti.htm

◗ **Testing of the Product / Technology:** A product based on this technology is available on-line on web at http://anglahindi.iitk.ac.in for testing and feedback.

◗ **IPR / Open-source:** IPR belongs to IIT Kanpur

◗ **Potential beneficiaries**: All Govt./semi-Govt. organizations/Parliament/State Assemblies/ Judiciary where bi-lingual/tri-lingual document preparation is mandatory; All commercial organizations (Banking, Pharma, Agro, Health, Tourism etc.) and educational institutions with

need for localization. NRIs and countries in SAARC and South East Asia.

◗ **User-agency tie-up:** None.

◗ **Name and address of the Resource Person**: *Professor R.M.K. Sinha (rmk@iitk.ac.in) Dr. A. Jain(ajain@iitk.ac.in) Dept of Computer Science and Engg. I.I.I. Kanpur 208016.*

6.2.6

◗ **Name of the Technology** : *AnuBharti Technology for MAT from Indian Languages to English.*

◗ **Nature of Technology** : A machine-aided translation technology for translation from any Indian language to English.

◗ **Level: (Product / Technology / Sub-system)** : Technology

◗ **Technical Description of the Technology / Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / Product:**

Please see http://www.cse.iitk.ac.in/users/langtech/anubharti.htm

◗ **Representative Snapshot / screenshot of the Technology / Product:**

Please see http://www.cse.iitk.ac.in/users/langtech/anubharti.htm

◗ **Scalability / Portability / Expandability:** Applicable to any Indian Language

◗ **Readiness of Transfer of Technology (ToT):** Ready for Technology Transfer.

◗ **Availability of documentation:** Please see http://www.cse.iitk.ac.in/users/langtech/anubharti.htm

◗ **Testing of the Product / Technology**: A product named HindiAngla (Hindi to English MAT) based on this technology for simple sentences is available for testing and feedback.

◗ **IPR / Open-source:** IPR belongs to IIT Kanpur

◗ **Potential beneficiaries:** All Govt./semi-Govt. organizations/Parliament/State Assemblies/Judiciary where bi-lingual/tri-lingual document preparation is mandatory; All commercial organizations (Banking, Pharma, Agro, Health, Tourism etc.) and educational institutions with need for localization. NRIs and countries in SAARC and South East Asia.

◗ **User-agency tie-up**: None.

◗ **Name and address of the Resource Person:** *Professor R.M.K. Sinha (rmk@iitk.ac.in) Dr. A. Jain(ajain@iitk.ac.in) , Dept of Computer Science and Engg I.I.I. Kanpur 208016.*

6.2.7

◗ **Name of technology : *MaTra: Human Aided Machine Translation System***

◗ **Nature of Technology:** Human aided Machine Translation.

◗ **Level:** Technology could be developed towards product.

◗ **Technical Description** : MaTra takes an innovative, pragmatic approach to the traditional problems of natural language analysis. This involves using an intuitive user interface and taking advantage of man-machine synergy to bypass some of the hard problems of NLP. Here, the human and the machine each do what they are best at, and the combination of the human and machine produces the translation, rather than just the human or just the machine.

It is found that a program can identify small chunks of sentences, such as verb groups and noun phrases, fairly reliably, whereas it is more difficult to get the overall structure, mainly due to lexical and structural ambiguity. Humans, on the other hand, seem to intuitively know which part of the sentences are related and almost effortlessly resolve most of the lexical and structural ambiguity by using their common sense and world knowledge.

The core component of our approach is an interactive structure editor, which exploits this observation. This component is used to create an internal representation of the input sentence, in the form of a nested hierarchical slot-value structure (See Figure 2). The sentence-level-tree is composed of many clause-level-trees. The topmost slot of a clause is usually the main verb-group in the sentence, called a *pivot*. Each *Pivot* verb has an associated template, which specifies the mandatory and optional slots. For example, the pivot value *killed* could have sub-slots called *who-what* and *whom-what* (which roughly correspond to the subject and object in a traditional parse tree), and there may be additional *more-info* slots specifying the time and other optional information. Each of these slots may in turn have other sub-slots, which would correspond to post-modifiers in a traditional grammar. The value of any slot may either be a simple sentence fragment, called a *chunk*, or may in turn be another pivot (representing a clause or full sentence), thus allowing complex sentences to be represented using an intuitive, nested hierarchical structure. The topmost slot of a sentence is either a clause or an operator in case of compound sentences.

The basic architecture of the system is given in the Figure 1.

MaTra Versions:

MaTra is available in two versions

- MaTra Pro: The Professional Translator's Tool

- MaTra Lite: Automatic On-line Translator

MaTra Pro:

MaTra Pro is more appropriate for serious translation — it is a Translator's Tool, and allows the user to assist the system in generating more accurate translations. It uses a simple, intuitive GUI for interaction, and allows customisation of the lexicon to specific application domains. The aim is for the machine to try and translate the simpler or more routine texts, and free the human translator to focus on the more difficult and creative tasks.

MaTra Pro is currently available under non-exclusive license.

Features

• Auto, Semi-Auto and Manual Modes

• Intuitive GUI for disambiguation

• User-customisable lexicon

• More Accurate Translation

Ideal For

• Editors, content-providers

• Professional translators

MaTra Lite:

MaTra Lite expects no interaction from the user, and is therefore able to do only rough translation. The entire range of complexity of English inputs

is not handled. Also, due to some of the choices taken independently by the system, the translations may not be entirely correct.

A version of MaTra Lite is being offered as a free, experimental Web-based service hosted at CDAC Mumbai's (formerly NCST) website (http://www.ncst.ernet.in/matra/), and accessible through a browser.

Features

- Simple web-based interface
- Automatic, end-to-end operation
- General, approximate translation

Ideal For

- Non-Hindi speakers learning Hindi
- Web users unfamiliar with English

Current Status:

The prototype for assertive sentences with one verb group is now being extended to handle compound-complex sentences. This should be available by November.

◗ **Specification of the Technology / Product:** Hardware: Windows 2000 with South Asian Support.

Software: Java, Perl C++.

Our system is developed in Java, however for the word grouping we are using fnTBL, a freely available chunker, which uses Perl and C++.

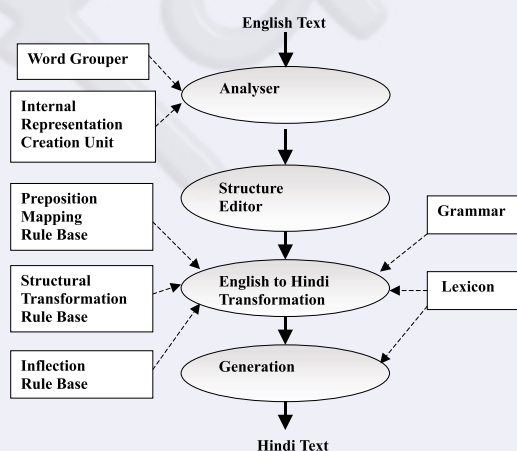Front-end / user interface: Java Swing
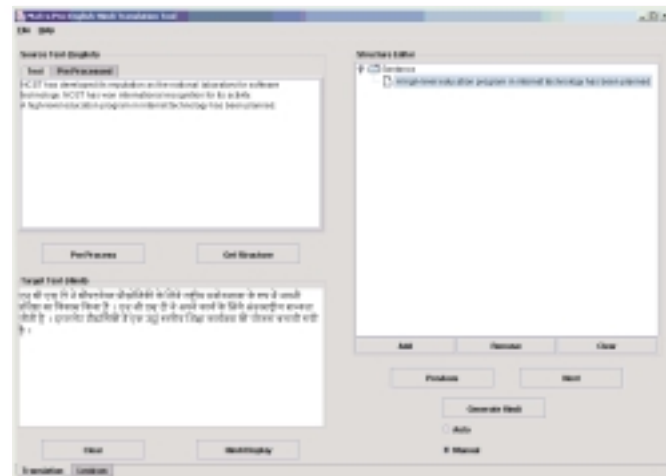

Figure 1: Basic Architecture of the MaTra System

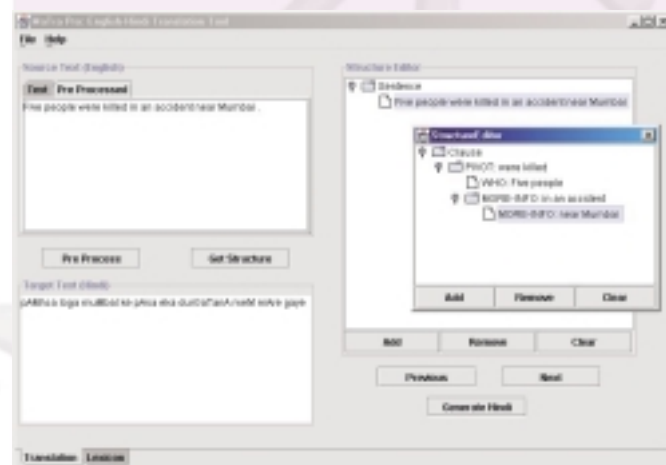Figure 2: Screen shot of MaTra system: Interactive UI

Figure 3: Screen shot of MaTra System: Interactive Structure Editor

◗ **Portability / Expandability: Expandability:** The system can be expanded to translate from English to other Indian languages with minimal changes to the core. The analysis component can be used as it is. The rule-bases (preposition mapping, inflection, and structural transformation), and the lexicon, of course, will need to be changed according to the target language.

Portability: Most of the MaTra system has been developed in Java except for the word grouping component, fnTBL. The source code of fnTBL, which is a mix of Perl and C++ code, is available freely under a license, which is very liberal even when compared to licences like the GPL.

Since Perl and C++ are also available easily for most popular OS, the system developed as such is portable.

◗ **Readiness of Transfer of Technology (ToT):** The system can be given to individuals/organisations under a non-exclusive license.

◗ **Availability of documentation:** Installation and usage guidelines are available.

◗ **Testing of the Product / Technology:** MaTra Pro1.0 has been used on NCST's annual report, news wire, and Linux phrases (for the localization effort) with good results. Large scale testing has not been carried out.

◗ **IPR / Open-source :** IPR lies with CDAC Mumbai. MaTra 1.5 (Pro/Lite) would be Open Source under a suitable licence.

◗ **Potential beneficiaries :** The professional version is beneficial for editors, content-providers, and professional translators. The system will also be useful for localisation applications. The Lite version is ideal for non-Hindi speakers to learn basic Hindi and for Web users unfamiliar with English.

◗ **User-agency tie-up :** The system is being used as the translation sub-component of the Cross Lingual Information Retrieval System, being developed for the Development Gateway Fund Project, World Bank. Discussions are in progress with other agencies.

◗ **Name and address of the Resource Person**
M Sasikumar, Kavitha Mohanraj
Centre for Development in Advanced Computing,
Raintree Marg, Near Bharati Vidyapeeth,
Opp. Kharghar Railway Station,
Sector 7, CBD Belapur,
Navi Mumbai 400614, India.
Phone: +91-22-27560013
Telefax : +91-22-27560004

6.2.8

◗ **Name of the Technology:** *English-Assamese Machine Translation System. (Machine Translation System from English to Assamese)*

◗ **Nature of Technology:** Human Machine Interface Systems.

◗ **Level: (Product / Technology / Sub-system):** Sub-System

◗ **Technical Description of the Technology / Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / Product:**

◗ **Technical Description**: This MT system developed is basically a rule based one and relies on a bilingual English to Assamese dictionary. The dictionary-supported generation of Assamese text from English text is a major stage in this machine translation. A rule-based system is an effective way to implement a machine translation system because of its extensibility and maintainability. Each entry in the dictionary is supplied with inflectional information about the English lexeme and all of its Assamese equivalents. The dictionary is annotated for morphological, syntactic and partially semantic information.

It can currently handle translation of simple sentences from English to Assamese. The Dictionary contains around 5000 root words. The system simply translates source language texts to the corresponding target language texts phrase to phrase by means of the Bilingual dictionary lookup. The resulting target language words are re-organized according to the target language sentence format in the tree structure. In order to improve the output quality, the system performs morphological analysis before proceeding to the bilingual dictionary lookup. The basic translation method is from English phrase tree to Assamese phrase tree.

Currently it can handle general-purpose simple translation. It is being upgraded to handle complex sentences. Efficiency of the system can be improved by selecting a specific domain. The

domain is taken from the web site www.iitg.ernet.in. Currently the rule based Machine Translation system contains 22 rules An example of Machine translation from English to Assamese is depicted in figure below.

Machine translation system developed here consists of

1. A process of analyzing input sentences (morphological, syntactic and/or semantic analysis)

2. A process of translating source language texts (English) to the corresponding target language text (Assamese) word –for-word and phrase-to-phrase.

3. A process of re-organizing target language words according to the target language sentence format.

The dictionary and the rules control the steps of each process. As the accuracy of translation by the system is the product of the accuracies of each process, it is necessary to enlarge the magnitude of the existing dictionary and the number of rules.

Basic block diagram:

Basic block diagram:
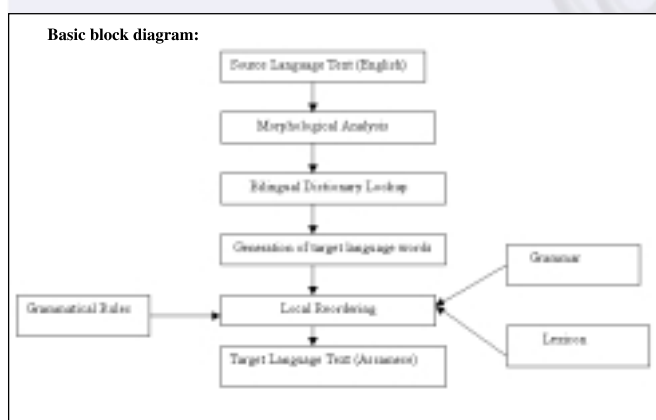
ALGORITHM

Some algorithms are developed to translate English sentences to Assamese sentences based on rules and dictionary.

The basic algorithm is

1. Splitting up of a sentence into subject, object and verb:

Verb offers a lot of grammatical information about the subject, object as well as tense and aspect. The part before verb is assigned as subject and the part after verb is assigned as object. Two types of verbs are there.

a) I am fine.

b) I am coming

When a sentence is entered by the user, the translation system processes the sentence from left to right. The basic algorithm used to split a sentence is given below:-

1.1 Identify the verb.

1.2 Once the verb is identified, the following word is identified. If this word is not the verb, the sentence is of type (**a**). If this word is also verb, the sentence is of type (**b**). Morphological analysis is performed if needed.

1.3 After identification of verb, the whole sentence is splitted into subject (part before verb), verb and object (part after verb).

Steps (1.1) and (1.2) are implemented using bilingual dictionary look up table.

2. To determine Assamese meaning of a English word from the Dictionary:

Algorithm:

2.1 If the English word is root, find the position of the row in the bilingual dictionary in which the root word is present.

2.2 The corresponding Assamese of the root English word is found from corresponding row's Assamese column at the first position. In Assamese column the first position is always kept for the Assamese meaning of the root English word.

2.3 If the word is inflectional, then morphological analysis is performed to find out the root English word.

2.4 Find the position of the row in the dictionary where the root word obtained in step 3 is present.

2.5 Variation column contains all the inflectional forms of the root word. If the inflectional word

is present at position pos1 in the variations column, then corresponding Assamese of the word is found at position pos1+1 in the Assamese column.

3. Re arranging technique: Assamese language has the subject-object-verb (SOV) grammatical structure unlike the English language, which has subject-verb-object (SVO) sentence structure. The basic technique of re arranging from phrase tree (English) to phrase tree (Assamese)
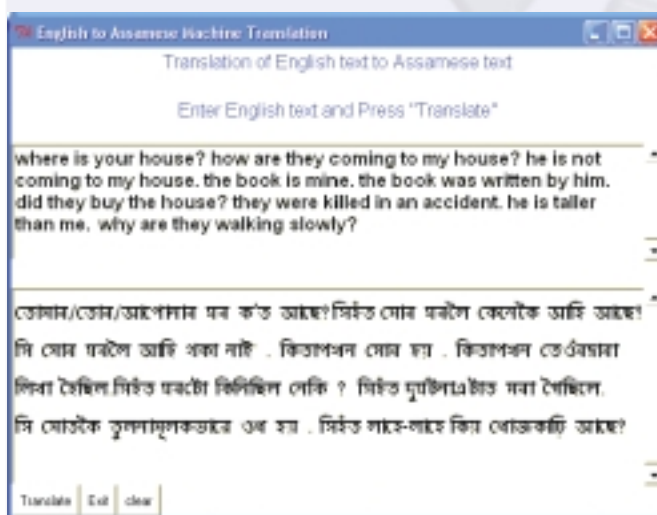
Operating System used: Windows 98, Windows XP, Windows2000

Front End/User Interface: User Interface developed using Perl/Tk.

Specifications:

• Dictionary and Rule based system.

• Currently supports simple sentences.

• User Friendly GUI available.

• Standalone system.

◗ **Representative Snapshot / screenshot of the Technology / Product:**

◗ **Scalability / Portability / Expandability:** Scalable and Portable.

◗ **Readiness of Transfer of Technology (ToT):** Installation guidelines and User's guide to use and configure the system, are under preparation.

◗ **Availability of documentation** : Under preparation.

◗ **Testing of the Product / Technology :** In progress

◗ **IPR / Open-source:** Open-Source

◗ **Potential beneficiaries** : Translators, Linguists, Govt. Offices, Research Organizations.

◗ **User-agency tie-up** : None

◗ **Name and address of the Resource Person:**
N.Deka   nilima_deka@iitg.ernet.in
P.K.Das   pkdas@iitg.ernet.in
Department of Computer Science & Engineering,
Indian Institute of Technology Guwahati
Assam – 781039 (India)

6.2.9

- **Name of the Technology :** *Shakti Machine Translation system*

- **Nature of Technology :** Human Machine Interface System (Machine translation from English to Indian languages)

- **Level: (Product / Technology / Sub-system):** Technology

- **Technical Description of the Technology / Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / Product:** Major goal of machine translation (MT) effort is to make English content on the web and other documents to be accessible to Indian readers not proficient in English. An alpha version of MT system called Shakti has been released over the web on 26th January 2003 for experimental use. Shakti combines linguistic analysis with statistical and example based methods (URL: http://gdit.iiit.net/~mt/shakti/index.html).

Shakti is a modular machine translation system which, broadly speaking, has the following steps in its processing

- source sentence analysis

- transfer grammar application

- target sentence generation

For source analysis the system uses a shallow parser and a morphological nalyser to get the POS, chunk and morph information about each word in he sentence, thus preparing it for further processing. Source analysis lso includes a number of other steps such as forming a PP, verb+adverb grouping, identifying phrases, word sense disambiguation etc. The tansfer grammar component uses a locally developed rule-based parser hich gives the dependency relations of the English sentences and marks ubject, object information. It then applies the re-order rules. Once the word-order of the target is achieved it applies the target rammar rules (such as agreement rules, vibhakti insertion etc) and finally generates the target language sentence.

Development of English to Indian language translators will allow English documents to be translated into Indian languages, so that they are available to people at large. The difficult requirement placed on the technology is that it should work in any subject domain, as opposed to narrow subject domains. Some of the technology was developed (esp. for English language processing units) in collaboration with CMU and University of Pennsylvania. Work done in this area will be made "freely" available.

The Shakti machine translation system is highly modular, implying that we can replace any of the modules (for example, morphology unit, chunker or a parser) with newer modules if we find better technology or improved algorithms.

As the team at LTRC, IIIT - Hyderabad has developed the technology and the toolkits that can be used by other groups working on Indian languages, we invite others to join in developing systems for all Indian languages. The tools and technology will be released as MT Shakti kit.

- **Representative Snapshot / screenshot of the Technology / Product:**

Please visit the web sitehttp://gdit.iiit.net/~mt/shakti

IIIT-Hyderabad Machine Translation System (Experimental) Shakti (Version 0.57)

Please send your suggestions and comments to mt@students.iiit.net

- **Scalability / Portability / Expandability:** The system is a large scale system geared up to handle any kind ofEnglish text. It is highly modular and hence portability becomes easier. With the toolkits and technology framework provided, it is easy to expand and come up with translation systems from English to other Indian languages.

- **Readiness of Transfer of Technology (ToT):** The technology is not yet ready to transfer to the industry - however, the system is demonstration ready.

- **Availability of documentation:** The documentation is available on the website

- **Testing of the Product / Technology:** This system is being evaluated very rigorously every day by a large setof user groups regularly with new set of test sentences. Rigorous validation modules allow the technology to be tested and upgraded regularly.

- **IPR / Open-source:** IPR: IIIT Hyderabad and MCIT. (Also to be available as open system)

- **Potential beneficiaries:**

  a) General Indian population who needs to read English documents.

  b) Web sites, search engines and portal that would need translation services from English to Indian languages

  c) Government and non-government organizations that require automated translation services from English to Indian languages.

- **User-agency tie-up:** NA

- **Name and address of the Resource Person:**

  Prof. Rajeev Sangal
  Language Technologies Research Centre
  International Institute of Information Technology
  Gachibowli
  Hyderabad
  India 500 019
  Phone: (91) (40) 2300 1412, 2300 1967 x 144
  Fax: (91) (40) 2300 1413
  Email: ltrc@iiit.net

6.3 Text to Speech System

6.3.1

- **Name of the Technology**: Bengali Text to Speech Synthesis System

- **Nature of Technology**: Human Machine Interface System (It converts the given written Bengali text (in ISSCI) into unintonated phonetically clear speech)

- **Level**: Technology

- **Technical Description:** Speech generation is the process, which allows the transformation of a string of phonetic and prosodic symbols into a synthetic speech signal. The quality of the result is a function of the quality of the string, as well as of the quality of the generation process itself

  In the past few decades, various Researchers have worked in the area of Speech Synthesis and Recognition and have developed different algorithms and methodologies for different speech technology development. In area of Speech Synthesis there are a number of different methodologies like Formant, Articulatory, Sinusoidal and Concatenation Synthesis.

  In the last decade there has been a significant trend for development of speech synthesizers using Concatenative based Synthesis techniques. There are a number of different methodologies for Concatenative Synthesis like TDPSOLA, PSOLA, MBROLA and **Epoch Synchronous Non-Over Lapping Add** (**ESNOLA**).

  In TDPSOLA method based di-phone concatenative technique has inherent problem in introducing intonation and prosody. In manipulating pitch for introduction of intonation the phonetic quality is often seriously compromised as because it is only pitch synchronous. Only pitch synchrony does not guaranty the preservation of phonetic quality.

  MBROLA has the same problems. Along with those, to accommodate intonation a large multiplicity of the diphones is required. This is a major problem in building all necessary elements of the signal dictionary.