Dialogue or Interactive Based MT

CS460 Natural Language Processing

By
Pranay Bhatia - 07005005
Akash Singhal - 07005016
Aditya Gandhi - 07d05007
Background
Limitations of Current MT

- LBMT gives raw translations, which might be useful only to an expert. These sentences cannot be used without revision, and there must be as many revisions as there are target languages.

- KBMT would be rejected as inexact paraphrase. It needs not only linguistic knowledge of some degree, but also a world knowledge sufficient to represent the considered domain of discourse.

- Also, ontologies used in KBMT are very expensive to create & maintain.
Motivation

- The main motivations for DBMT are
  - the limitations of current MT paradigms,
  - increasing importance of national language in the global context of internationalization and
  - the technological advances which result in individual users having the need to translate
Introduction
Dialogue-Based MT

• The goal of DBMT is to create a single MT system for all users to translate ‘clean text’ from one or more languages to many others.

• The user is interrupted with prompts when ambiguity in the data is encountered. The input preferences are then used to disambiguate the remainder of the data.
Advantages of DBMT

- Translations are of high quality, and require little or no review.
- It supports multilingual translation, i.e., one-to-many and many-to-many language translations.
- DBMT can translate without the language being controlled.
- Interaction with the user ensures correct disambiguation of the word.
Approach
Suppose we wish to translate a large English document into Hindi, French and German.

Considering the sentence ‘Ram walks to the bank’. In the following slides we shall see how DBMT would translate this sentence, and by extension, the document as well.
The MLDB

- The underlying multilingual lexical database (MLDB) contains one monolingual dictionary for each language, and one interlingua dictionary for the interlingua acceptions.

- Thus, the English monolingual dictionary would have the words ‘walks’ ‘to’ ‘the’ & ‘bank’, and the acceptions dictionary for each pair of languages would have the corresponding words in the foreign languages.
Interlingua Acceptions

- Items of the monolingual dictionaries (monolingual acceptions) are generally accepted meanings of words or expressions.
- In a MLDB composed of n monolingual dictionaries, the set of interlingual acceptions is equal to the union of the sets of monolingual acceptions of the n dictionaries.
- Two monolingual acceptions of different languages correspond to a unique interlingual acception if, and only if, they have the same meaning.

A 'rivière' is a rather small river flowing into another river. A 'fleuve' is a large river flowing into the sea.
Lexical Preferences

- When the system encounters an ambiguous word, it prompts the user for its sense disambiguation.

- In our sentence, when the system reads the word ‘bank’ it finds two possible meanings, **financial bank** and **river bank** in the sentence ‘Ram walks to the bank’.

- The user is then presented both the options along with their meanings (using the monolingual dictionary) and the user selects the correct meaning of the word in the document.
Lexical Preferences

* The important point to note here is that the user does not need to know any of the target languages (Hindi, German or French) to ensure that the correct meaning of the word is reflected in the translation.

* Also, since the user is adept in the source language (English in this case), taking his input for disambiguation greatly removes the possibility of erroneous translations, and therefore, the need to review the same.
Text Genres

- A given text type can be defined by a set of lexical weights relative to MLDB as belonging to a text genre with some numerical restrictions.

- For subsequent similar inputs, this can considerably reduce the number of ambiguities produced by the analyzer and consequently the amount of interactive clarification required from the user.

- In the example under consideration, if the user selects the financial bank sense of the word, subsequent ambiguities would assume higher weightage to finance-related senses of ambiguous words.
Flow Chart

INPUT SENTENCE

IS AMBIGUOUS?

IF YES

PROMPT USER FOR INPUT

IF NO

ASSIGN WEIGHTS TO THE LEXICAL PREFERENCE

IDENTIFY TEXT GENRE FOR NEXT TRANSLATION

TRANSLATE

Monday 18 April 2011
The LIDIA Project
Introduction

- It is a first running mock-up of this concept, used for translations from French into German, Russian and English.
- It incorporates an interactive user interface to provide translations using DBMT.
Translation

- The analyzer produces various possible trees for the source sentence
- The user is then prompted to resolve ambiguity in the structure of the sentence.
Translation

- Ambiguous phrases are prompted to the user for feedback
- “The captain brought back a vase from China”.
- The first paraphrase translates as “from China, the captain brought back a vase”
- Second one translates as “the captain brought back (a vase from China)
Translation

- Word sense disambiguation is also done by user input.
- In this case, the ambiguous word is captain, and the possible meanings are related to:
  - the military
  - the navy/shipping
  - the team

![Image](image-url)  
*Fig. 10. Word sense disambiguation (story 2).*
Translation

- Once the source text is disambiguated, the machine creates a corresponding structure for the target language.
- The disambiguated source words are then translated into the target language, and arranged into a sentence.
Ongoing/Future Work

Self Explaining Documents

These documents preserve the ‘author’s intention’ that went into the translation by way of ‘annotations’. By making these annotations visible, documents can be transformed into ‘Self Explaining Documents’ (SED).

An SED, can be translated from one language to the other without human intervention, provided an analyzer and transformer for that language are available.
Ongoing/Future Work

Personal Machine Translation

PMT is a new concept currently being developed in the LIDIA project. Ideally, a PMT system should be able to run on a PC and be accessible to everybody.

The major challenge here is the user not having any knowledge, linguistic or translation, about the target language(s).

PMT is expected to go a long way in the promotion of Local and National languages.
Conclusions
The Dialogue-Based MT can provide a viable solution to multi-lingual translations, without expertise in all the concerned languages.

The system evolves taking input preferences of the user, thus generating reliable final translations.

The incorporation of acceptions in the multilingual lexical database removes the need for language independent ontologies for all the languages. Since such ontologies are very difficult to create, the DBMT approach is a viable approach to global translations.
References


Thank You!