Translation Issues

Shelly Jain, 20171008

Introduction

Translation involves decoding of a text in one language followed by encoding it into another. Divergence between languages is observed at several levels.

During machine translation, various problems are observed. These may be at decoding level, or at the level of cross lingual transfer of information (which must be resolved at encoding level).

Some Problems for Machine Translation

Barbara Cairns
Lund University, Dept. of Linguistics

Abstract

An analysis of the translation of a Swedish magazine article text into English. Special emphasis is on the type of semantic problems which could arise in machine translation, as envisaged by a human translator. Some syntactic problems of word order are also discussed.

Translation process

A general translation initially works top-down. Once potential meanings are reduced it works bottom-up. Good translation requires consistency, continuity and error-checking.

1. whole passage (in order to get the gist)
2. smaller chunks, e.g. paragraphs
3. sentences
4. phrases
5. words

Syntactic problems

Two languages such as English and Swedish with similar word order are fairly compatible as long as correspondence is consistent. But there are examples where the correspondence is much less consistent, hence less programmable.

The passive - One form is represented by the morpheme -s at the end of the verb, used frequently in Swedish. It presents two problems: passive is not often used in English and requires a Swedish paraphrase; even when a parallel exists, English structure is different and requires complete restructuring.

Syntactic problems

The infinitive - Swedish allows the insertion adverbial phrases between att+INF while English prefers it outside the sequence. It is not always found in the same contexts; the machine would need instructions to decide translations.

Negation - It follows different patterns in Swedish and English. The complex correspondence depends on syntax and meaning of the whole negated phrase.

Tense - Mostly direct correspondence, except that Swedish lacks progressive tense, and certain cases of future tense require English restructuring.

Semantic problems

Ambiguity - In cases of the usage of words which are homonymous or polysemous.

Compounds - Compound lexical items not formed by combining translation of the parts, where word-to-word translation may cause unintelligible mistakes.

Prepositions - Rarely any neat correspondence. Translation of prepositions between Swedish and English is troublesome due to many interpretations.

Semantic problems

Ellipsis - Situations where words in original text may be translatable, but are not required and are generally omitted from target language text.

Reference - Swedish has more inflectional morphology; referring expressions require more agreement-checking than in English. Swedish pronouns may be anaphora, cataphora, exophora, or possessives. There are also many reflexives.

Definite/indefinite - Correspondence between Swedish and English in their use of definiteness is not always consistent.

Semantic problems

Morpheme -s - It can also be reciprocal or be an active verb. Certain 'deponent' verbs only exist with the -s morpheme. Differences are only obvious in context.

Cultural differences - Texts are bound to a specific culture and every language makes reference to its own culture. Lexical gaps are also tied to the background culture of a language.

Idioms - Expressions which cannot be directly translated, because equivalent expression already exists as a given chunk in the target language.

Gender Issues in Machine Translation

Anke Frank, Christiane Hoffmann, Maria Strobel
Lingenio GmbH, Heidelberg

Abstract

Comparative evaluation of commercial MT systems; a discussion of gender-related phenomena like person references, anaphoric references, handling of definite noun phrases, aspects of productivity and lexicalization, scope and coverage of dictionaries, and rules in morphology, grammar and transfer; aspects of usability when the program is unable to decide, but leaves the decision to the user, as in German-French translations

Gender marking in languages

In German, references cannot be made without expressing the linguistic category gender. As per the economy principle, languages offer default interpretation of gender. In German, French and English this is masculine.

In English-German translations, certain contexts require transformation of a gender-neutral word into one which is gender-marked.

Gender-sensitive translations

For correct and gender-sensitive translations, the software must have certain prerequisites.

1. The grammar must provide analysis of texts.
2. The semantics should distinguish between female and male.
3. Dictionary coverage must be broad.
4. Discourse representation component for coreference resolution.
5. There must be either exhaustive dictionary or rules in morphology for analysis via derivations.

Comparative evaluation

In the paper, the most relevant commercial MT systems are evaluated. The choice of systems was influenced by language pairs offered, and availability of online translations.

For English-German, translations largely fail; for German-English the results are unclear. Most systems show severe shortcomings in their handling of gender-appropriate translations and in the treatment of gender-neutral language.

Gender issues in French

French has a moderate use of morphology to express syntactic relationships, for marking number, person and gender. The marking is clear when nouns show clear membership to a gender category. But more information can be deduced from context, especially in unmarked cases.

For French-German translations, accuracy of gender marking in unclear cases is dependent on accuracy of the discourse analysis. Comparison shows that all systems successfully handle agreement rules in French concerning participles and predicate adjectives.

User issues and feedback

In some cases, inferring gender information from context is difficult. The form of French verbs varies with gender and number of the speaker or addressee. Here, machine translation requires additional gender information about speaker and/or addressee from the user.

There was no consciousness for gender-appropriate translations observed in comparative analyses of translation software in the relevant media.

Machine Translation Approaches: Issues and Challenges

M. D. Okpor

Department of Computer Science, Delta State Polytechnic, Ozoro Delta State, Nigeria

Abstract

The demand for translation has become more in recent years due to increase in the exchange of information between various regions using different regional languages. Many approaches have been used in the recent times to develop an MT system. Each of these approaches has its own advantages and challenges. This paper takes a look at these approaches with the few of identifying their individual features, challenges and the best domain they are best suited to.

Machine Translation

A machine translation (MT) system first analyses the source language input and creates an internal representation. This representation is manipulated and transferred to a form suitable for the target language. Then at last output is generated in the target language.
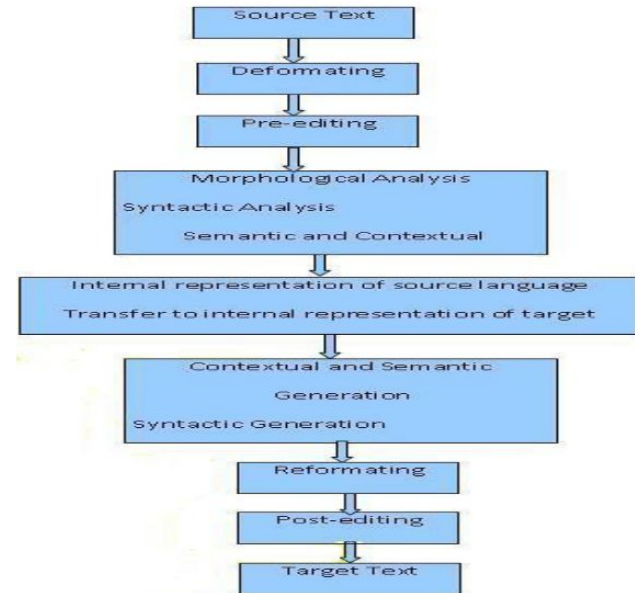


Fig. 1   A Typical Machine Translation Process

Machine Translation

In rule-based approach, humans specify a set of rules to describe the translation. This requires enormous amounts of input.

In corpus-based approach the knowledge is automatically extracted by analysing translation examples from a parallel corpus.

Combining the features of the two major classifications results in the hybrid approach.

Methods and challenges
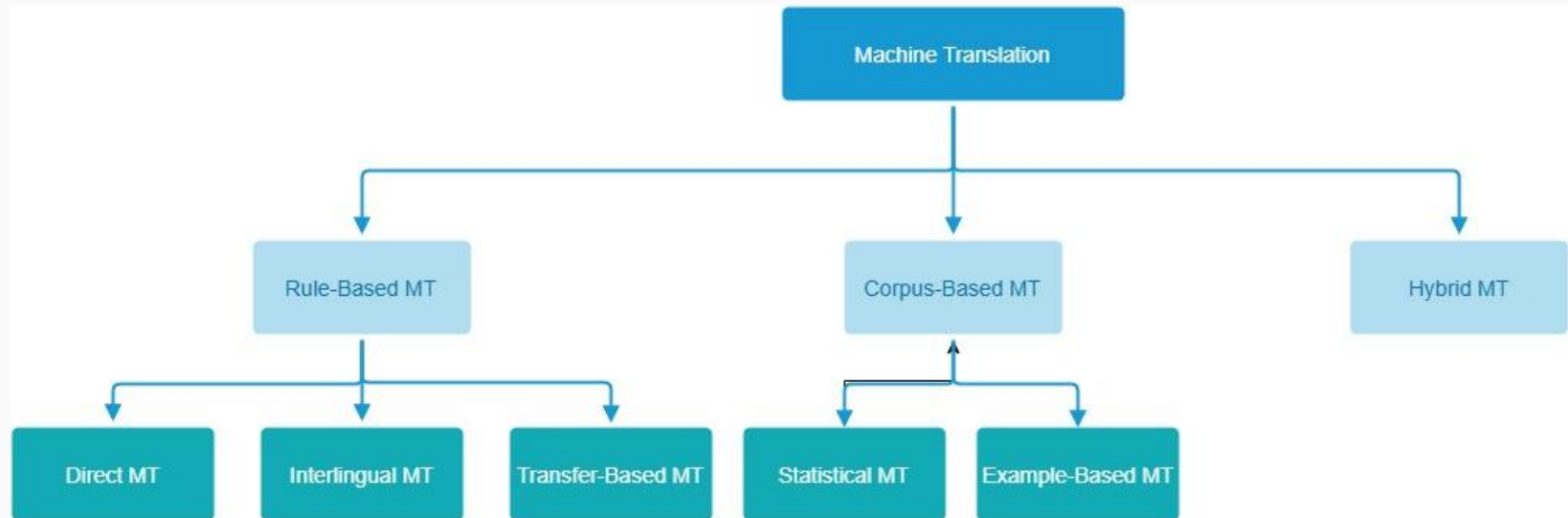
Methods and challenges

Direct MT (Rule-based)

1. 'Word-for-word' translation, hence frequent mistranslations at the lexical level and incorrect syntax structures which mirror the source language.
2. Linguistic and computational naivety; lacks analysis of internal structure of source text and no computational sophistication.

Interlingual MT (Rule-based)

1. Difficulty in defining an interlingua, even for closely related languages.
2. Difficult to extract meaning from texts in the original languages to create the intermediate representation.
3. Semantic differentiation is specific to target-language. Making such distinctions resembles lexical transfer and not all distinctions are needed.

Methods and challenges

Transfer-based MT (Rule-based)

1. Rules must be applied at every step. There are rules for source language analysis, rules for source-to-target transfer and rules for target language generation.
2. Difficult to work using reusable modules of analysis and synthesis.
3. It is difficult to keep transfer modules as simple as possible

Statistical MT (Corpus-based)

1. Corpus creation is costly for users with limited resources.
2. The results are unexpected. Superficial fluency can be deceiving.
3. Does not work well between languages with significantly different word orders.
4. Benefits are overemphasized for European languages.

Methods and challenges

Example-based MT (Corpus-based)

1. It avoids the need for manually derived rules, but requires analysis and generation modules to produce dependency trees needed for the examples database and for analysing the sentence.
2. Poor computational efficiency, especially for large databases, although parallel computation techniques may be applied.

Hybrid MT

1. Using a combination of rule-based and corpus-based MT allows flexibility, but devising the optimal method becomes tedious and requires lots of trial and error.

Other Issues faced in MT

1. Issues with datasets for translation
2. Miscellaneous linguistic issues
3. Possible errors

Quality of translation is also affected by the datasets, though it is not directly improved by strategies like more data or parallel datasets, or even morphological coherence.

The machine may face linguistic challenges like - sentence alignment, compound words, idioms, morphology, different word orders, syntax, out of vocabulary (OOV) words, statistical anomalies, data dilution.

Possible errors encountered in the translated text are - missing words, incorrect word order, unknown words, incorrect words.

Conclusion

Issues in MT can be caused by a variety of factors, from the beginning to the end of the process. It may be at data level, encoding level, intermediate levels, or decoding level. The issues may be syntactic or semantic. Erroneous results can take many forms.

Thank You