Interlingual Machine Translation: Prospects and Setbacks

Firat Açıklöz and Olcay Sert,
TOBB ETU—University of Economics and Technology
Hacettepe University, Ankara, Turkey

Abstract

This study, in an attempt to rise above the intricacy of 'being informed on the verge of globalization,' is focused on the premise that Machine Translation (MT) applications searching for an ideal key to find a universal foundation for all natural languages have a restricted say over the translation process at various discourse levels. Our paper favors not judging against the superiority of human translation vs. machine translation or automated translation in non-English speaking settings, but rather referring to the inadequacies and adequacies of MT at certain pragmatic levels, lacking the right sense and dynamic equivalence, but producing syntactically well-formed or meaning-extractable outputs in restricted settings. Reasoning in this way, the present study supports MT before, during, and after translation. It aims at making translators understand that they could cooperate with the software to obtain a synergistic effect. In other words, they could have a say and have an essential part to play in a semi-automated translation process (Rodrigo, 2001). In this respect, semi-automated translation or MT courses should be included in the curricula of translation departments worldwide to keep track of the state of the art as well as make potential translators aware of future trends.

1. Introduction

It is now widely accepted that global communications must be accessible and transferable, in a timely manner, in as many languages as feasible. Given that any field in which human beings are actively involved requires the knowledge of another field, MT, having a history almost as old as the modern digital computer, emerged as an attempt to overcome the intricacy of 'being informed' in a group of offers to sustain communication. In doing this, MT, much advanced since then, is a key mean for the human translator, although not without its problems. MT applications, for some, have long been challenging human translators. For others, despite MT researchers' arguments, it cannot aim at replacing the human mind. Moreover, MT designers, taking a simplistic view of language translation, have also long been searching for an idealistic key to find a universal foundation for all natural languages. For instance, Arnold et al. (1995) suggest a comprehensive assessment of the issues behind MT and popular misconceptions. For them, the types of knowledge an automated translation system should have are: a) linguistic knowledge independent of context (semantics), b) linguistic knowledge that relates to context (pragmatics), and c) common sense / real world knowledge (non-linguistic knowledge).

Although this attempt to refine the issue had proposed ideas valid for the translation itself on a hypothetical basis, it failed to see the phenomenon as a cumulative process and internalize all other elements that make a piece of writing representative of a specific culture and a language, which can evidently be observed as a neglected dimension in MT pioneer Weaver's (1949) words: "I have a text in front of me which is written in Russian but [...] pretend that it is really written in English [...] All I need do is strip off the code in order to retrieve the information contained in the text." In Warren Weaver Memorandum, solely organized for the purpose of selling the idea that machine translation works, he called attention to his four proposals mentioning that:

a) The problem of multiple meanings might be tackled by the examination of the immediate context, (only accomplished in particularly restricted terminology, e.g. discourse of law, discourse of medicine etc.,

b) there are logical elements in language. He drew attention to a theory that "a robot (or computer) is able to deduce any legitimate conclusion from a finite set of premises." Since written language is logical in nature, the problem of translation is formally solvable

c) frequencies of letters, letter combinations, intervals between letters and letter combinations, letter patterns, etc. which to some significant degree are independent of the language used,
d) and lastly, he asserted his belief in the existence and applicability of language universals.

However, this task is not as easy and clear-cut as suggested, and may not even be attainable in some contexts. It has proved impracticable in view of the fact that compressing natural languages into a universal set of symbols denies the very nature and cultural value of language. Hutchins (2003), on account of a chronological study of MT systems, recommended "that different types of MT systems are required to meet widely reverse translation needs" in which he implies that MT does not have the same success with varied genres, registers and discourse levels, or those involving a high level of abstraction and culture-bound items.

Most authors of the literature on MT systems have focused on a particular problem and stated the limitations of their studies in abstracts and conclusions. E.g. the study entitled 'The Automatic Translation of Idioms—Translation Memory Systems vs. Machine Translation' by Volk (1998) agrees on the impracticability of translating idioms due to syntactic setbacks, but proposes an untested and questionable approach, whereas the truth that language cannot be separated from its segments either culturally or linguistically remains disregarded. In this sense, Thriveni (2002) lays emphasis on the idea: "One language cannot express the meanings of another; [...] different languages predispose their speakers to think differently [...]," which could be taken as the very statement that inspired this work.

Our scope in this paper is not to argue for the superiority of human translation versus machine translation or automated translation in non-English-speaking settings, but rather to touch upon the inadequacies and inadequacies of MT in certain contexts because of its failure to understand true meaning and dynamic equivalence, despite producing syntactically well-formed or meaning-extractable outputs in limited settings.

This work insists on the low probability of any future MT system meeting a wide range of translation needs unless computers are able to make judgments, decisions and choices consistent with non-linguistic knowledge that people frequently refer to in their daily lives. Yet, it is necessarily not a dogmatic claimer of MT's not being utilized in any areas. No doubt, it has certain possibilities, as Melby (1996) justly states: "On some texts, predominantly highly technical texts treating a very narrow topic in a rather dry and monotonous style, computers sometimes do quite well."

Within this framework, our work firmly favors two notions:

- The contemporary practical MT systems are designed mostly to render highly intricate, scientific and technical texts with a limited range of terminology—in other words, highly repetitive texts that are structure-bound in that all the target sentences are composed only from the syntactic structure of the source sentences such as web pages and electronic mails.

- The use of MT before, during, and after translation can be justified. We aim at making translators understand that they can cooperate with the software to obtain a synergistic effect. In other words, they could have a say and had an essential part to play in a semi-automated translation process (Rodrigo, 2001).

2. Synopsis—MT

MT is aimed at enabling a computer to transfer natural language utterances, or to process a natural language in terms of lexical, syntactic, and semantic dimensions (See Figure 1., Vauquois, 1968) in either text or speech from one language into another while preserving both explicit and implicit meaning. A distinction is originally to be made between human-aided MT (HAHT) and machine-aided human translation (MAHT). The latter uses computer-based translation tools which prop up translators by providing access to on-line dictionaries, remote terminology databanks, transmission and reception of texts, stores of previously translated texts ("translation memories"), and integrated resources, commonly referred to as translator workstations or translator workbenches. The term computer-aided translation (CAT) is sometimes used to cover all these computer-based translation systems.

The Machine Translation Pyramid (MTP) suggested by Vauquois (1968) specifies a way of processing comparable to that used by the human translator. The system begins on the left bottom by analyzing the source language; the analysis becomes more and more complicated as we ascend the pyramid to the semantic and syntactic levels. The term 'transfer' means a format suitable for the interpretation and generation of the target-language text.
Nida (1964:246-247) proposes the following nine steps to be employed by a competent translator with some steps being optional:

1. Reading over the entire document
2. Obtaining background information (culture-inclusive)
3. Comparing existing translations of the text [if they exist]
4. Making a first draft of sufficiently comprehensible units
5. Revising the first draft after a short lapse of time
6. Reading aloud for style and rhythm
7. Studying the reactions of receptors by the reading of the text by another person [omissible]
8. Submitting a translation to the scrutiny of other competent translators [omissible]
9. Revising the text for publication

Consider the word *bill*; it can have anumber of widely different meanings. (See Appendix II.). Subsequently, translating this particular word into other natural languages is not clear-cut. Due to this many-to-many mapping of words, word-for-word correspondence is deficient; hence, a lexical transfer system fails to achieve accurate translation even at this very morphological level.

The following representational stage is to do with Syntax. E.g. identifying whether *bill* is a noun or a verb provides a translator with a first element for translation. The categorization of words, cognitively, requires a rule-based process depending on the syntactic arrangement of the word used. The problem is still not vanished as multiple meanings for the noun *bill* and the verb *bill* are still available.

Up the pyramid, the semantic level probably makes the greatest contribution to the translation process when two other levels are taken for granted. Semantic disambiguation is indispensable for the proper translation of words represented with the same symbols but divergent meanings in various languages. To cite a more concrete example: Within the context of drug trafficking, "a mule" is a person who carries drugs across frontiers. In French, the term "le passeur" is used with the same meaning. However, in another contextual setting or via the help of surrounding words, it might be a term used by either the police or a drug dealer. What if human translators and computers are swapped? Could computers tell apart a drug dealer speaking and police interrogating? The rationale behind this depending fully on pragmatics—here it is the scene of the occurrence—drives a translator to strictly adhere to the decision-making process: the real world knowledge in which the most crucial step is "the capacities, which computers lack, to make real choices by exercising our innate faculty, choices for which we are responsible" (Melby, 1996)

3. Evidential theory towards praxis

Benjamin Whorf's field study in the Hopi speaking community so as to do an extensive research in this Native American Language came up with the findings that the Hopi creates or interprets the world in a way quite dissimilar from that of the ordinary speaker of any European language. He asserted that the way in which people see the 'real world' was based on the language habits of that specific community—a phenomenon known as linguistic relativity.

For instance, as cited in Bahar (2001), Bloor and Bloor (1995:246) emphasize "that the speakers of one language may describe two objects as, say *red* and *orange* (English) where those who speak another will describe them with the same term, say *kai* (Amharic)...". Likewise, the Hopi have only one word for most flying objects (Hayes et al., 1996). A *dragonfly*, an *airplane*, and a *pilot* are all defined via the same word founded on a specific-shared feature theory or otherwise called 'overgeneralization'. Whorf also proposed that in the Hopi language; a system of tenses is not available, but as a substitute, they talk about concepts on the subject of durations and the speaker's perceptions. Hence, in Hopi, the form of verbs changes to express whether the speaker and listener can currently witness the occurrence or whether they are predicting or remembering it—a phenomenon known as aspectual categorization.

As it is the case with mono-cultural words. An unhygienic round dessert sold widely in front of brothels in Turkey called 'halka tatlısı' which literally translates as 'round dessert,' also called 'brothel dessert,' (Çakgöz, 2005) is believed to be an aphrodisiac to a Turkis male, but means no more than an ordinary dessert to a foreigner. Correspondingly, in some local dialects of western Turkey, people originally make use of distinctive adverbs 'ösen', 'yalımart' meaning 'maybe, perhaps', whether it be in a literary piece written by a nationally known author or in local speech; it means nothing to an outsider of that district in Turkey unless an explanatory footnote is given at the bottom of the page.
Given that thought and languages have borrowed from one another so much and have developed interactively over time, societies now and then let the language be influential over thoughts, and at times thoughts over the language itself. The above-mentioned examples represent the former aspect, but there have been times ideologies and thoughts of societies shape the language in certain contexts. A good example of this was witnessed in 1950-60's and could be here given as a case of 'ideo-euphemism'. The language surrounding weapons were knowingly developed in harmony with the ideology of certain groups (cited in Carroll, 1956; Knight, 2003). Terms used at that juncture seem to be scientific or specialized and often have positive connotations or hide a sinister idea. An "Enhanced radiation weapon" that literally means a bomb that kills people leaving property intact, or, Demographic targeting that means killing the civilian population. It is unquestionably the translator's job to make a distinction between varied meanings of a word which may further break itself into numerous connotations, whereas such a distinction is beyond the capabilities a machine deprived of word which may further break itself into numerous connotations, whereas such a distinction is beyond the capabilities a machine deprived of functional-contextual aspects and cultural values. Furstenberg et al. (2001), in their cultural project, talks about how an identical word can carry totally opposite connotations in unlike cultures. The word the fetus of a goat locally named kutti pi as a food, for instance, is a prime example where highly positive connotations of words such as "delicacy," "rare; but tasty," in the Anglo-Indian culture, while in the Turkish culture it has negative undertones such as "nausea," "disgusting." (National Geographic-Online)

Gross (1992) notably speaks of two main setbacks of MT: They lack the contextual and pragmatic competence of humans. On the other hand, Canale lists four competencies of human translators: (a) grammatical, (b) sociolinguistic, (c) discourse, and (d) strategic competencies both in the source and target languages. Additionally, as cultural diversity doubles the richness of a language, machines are at a disadvantage since they lack culture-specific notions.

The latter of the two setbacks mentioned by Gross is to do with the functional aspect of languages. Machines' sole purpose is to convey meaning, whereas natural languages perform numerous functions such as humor, sharing emotions or feelings without needing to convey any actual information, establishing solidarity, etc. (Rodrigo, 2001). E.g. the single word 'yes' generally known as conveying confirmation, agreement and so forth may gain numerous divergent meanings depending on its function. Likewise, its function is uneven and situation-sensitive with context, stress, and intonation (See Appendix I). A further aspect deals with ambiguity, idioms, collocations, and structural and lexical differences between the source language and target language, which are highly valid concerns for Gross (1992, p.111). As seen, so many variables are dynamically engaged in this decision-making process, and it also requires a capacity of functional-contextual entry so as to account for multiple meanings of even a single word.

Pericliev (1994) in his article on structural ambiguity reached a conclusion that supports his hypothesis: Structural ambiguity in MT can be partially overcome by preserving the syntactical ambiguity of the source language into target language; in his case Bulgarian into English. However, a competent translator is able to make judgments, inferences, as well as deductions in the course of his/her decision-making process and of manipulating the ambiguous source sentence at ease following the nine steps of Nida. The solution to this problem is, as suggested by Pericliev (1994), that MT should not intend to explicate SL words. So, in this respect, "Perhiz ve Lahana Turusu" corresponds to "irrelevance of the topic under discussion with the current situation."

There exists a whole body of research literature for approaches to recognize and translate idioms (Volk, 1998). Consequently, the full treatment of idioms is considered a difficult problem, since it involves a flexible distinction between literal and non-literal interpretation. Moreover, Topçu (2004) rightly states that even a culture's tiny segment can get modified by time and experience changes in meaning consistent with context or setting. She further explains the intricacy of negative or positive connotative meanings of each idiom in a specific culture. Consider the following example extracted from Turkish newspapers. "Perhiz ve Lahana Turusu" (cited in Topçu, 2004, Cum., 28-09-1999): This ordinary use of idiom can frequently be encountered in daily lives of Turks, thus can be interpreted without much effort.-"Diet and pickled cabbage" when literally translated. Yet its meaning is independent of the literal meaning of the individual words. So, in this respect, "Perhiz ve Lahana Turusu" corresponds to "irrelevance of the topic under discussion with the current situation."

4. Impediments to example-based MT evaluation

Example-based translation (EBMT) is, to all intents and purposes, translation by analogy. An EBMT system stores a set of sentences in the source language and their corresponding translations in the target language, and uses those examples to translate other identical or similar sentences. The crucial argument is that, if a previously translated sentence re-occurs, the same translation is likely to be accurate again. Reasoning in this way, using separate examples from three different translation engines—Google and Alta Vista, Proceviri—in French, German, English, and Turkish in our efforts to support the notions offered in this study (See introduction), the principle is demonstrated with reference to different cultures, contexts, genres, and discourses.

The core motive of offering sample uses from actual, rather than invented, texts as did some scholars, is that natural language should be the only and absolute source for analyzing any sort of language produced by its users. Therefore, the data evaluated in this
article is adequate for the purpose.

It should not be neglected that there exist some setbacks. Initially, the texts are not complete texts because of technical limitations. Drawing upon this, it may, at times, not be so easy to deduce the intended meaning because of the lack of context and co-text of the extracts. Additionally, the data is randomly selected, and therefore the examples (See Appendix III.) may not be well-systemized. However, it is worth mentioning that random selection leads to objective evaluation, as systematic elimination of any sample might give the impression of subjective bias in selecting the entries.

These impediments mentioned are certainly subject to change depending on the nature of the study given that existing samples in Appendix III fulfill the necessary requirements the two principles require. In this sense, it may possibly be claimed that the samples under examination adequately explicate and support what is intended throughout this study. Within this framework, the words or phrases in *italics* revealing the inequivalences stemming from lexical, semantic, syntactic, cultural, and discourse-based incorrespondences among languages reflect the ideas behind all the above-mentioned theories of previous scholarly works as well as our own approach to MT.

5. Conclusion

Put it simply, this paper does not claim the superiority of human translation versus machine translation in a comparative manner, but rather points out the impediments of MT in certain contexts although it, at times, may produce syntactically well-formed or meaning-extractable outputs. To replace this gap, it puts forward and suggests machine-aided human translation in that it is more efficient, effective, economic, and less time-consuming in comparison to solely machine translation. The post-editing process should follow a route, not through direct implementation but sensible integration of MT into pre-translation and actual translation process. Experience recommends a semi-automated translation process to obtain the desired synergistic effect.

Appendix I.

<table>
<thead>
<tr>
<th>Word</th>
<th>Symbols</th>
<th>Tones/words</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>↑</td>
<td>Levél</td>
<td>Neutral, uninterested 'Yes'</td>
</tr>
<tr>
<td>Yes</td>
<td>✔️</td>
<td>Fall</td>
<td>Negative response 'I agree'</td>
</tr>
<tr>
<td>Yes</td>
<td>✔️</td>
<td>Low-rise</td>
<td>Question, eliciting a response</td>
</tr>
<tr>
<td>Yes</td>
<td>✔️</td>
<td>High-rise</td>
<td>Disappointment: 'Are you saying Yes'</td>
</tr>
<tr>
<td>Yes</td>
<td>✔️</td>
<td>Fall-rise</td>
<td>'Carry on, I'm listening'</td>
</tr>
<tr>
<td>Yes</td>
<td>✔️</td>
<td>Rise-fall</td>
<td>Reserved, indicating doubt</td>
</tr>
</tbody>
</table>


Appendix II.

bill ¹ n.
1. An itemized list or statement of fees or charges.
2. A statement or list of particulars, such as a theater program or menu.
3. The entertainment offered by a theater.
4. A public notice, such as an advertising poster.
   b. Slang One hundred dollars.
   b. Obsolete A promissory note.
7. a. A draft of a proposed law presented for approval to a legislative body.
   b. The law enacted from such a draft: a bottle bill in effect in three states; the GI Bill.
8. Law A document presented to a court and containing a formal statement of a case, complaint, or petition.
tr.v. billed, bill-ing, bills
1. To present a statement of costs or charges to.
2. To enter on a statement of costs or on a particularized list.
3. 
   a. To advertise or schedule by public notice or as part of a program.
   b. To declare or describe officially; proclaim: a policy that was billed as an important departure for the administration.

bill ² n.
1. The horny part of the jaws of a bird; a beak.
2. A beaklike mouth part, such as that of a turtle.
3. The visor of a cap.
4. Nautical The tip of the fluke of an anchor.
intr.v. billed, bill-ing, bills
To touch beaks together.

Idiom:
bill and coo
To kiss or caress and murmur endearments.

[Middle English, from Old English bile.]

bill ³ n.
1. A billhook.
2. A halberd or similar weapon with a hooked blade and a long handle.

(www.thefreedictionary.com)

Appendix III.
Entry 1#

Source input: To be completely honest, I'm not a big fan of mystery endings.

Target output: Um vollständig ehrlich zu sein, bin ich nicht ein großer Ventilator der Geheimnisenden. (Google translation engine-GTE )

Target output: Um vollständig ehrlich zu sein, bin ich nicht ein großer Ventilator der Geheimnisenden. (Altavista translation engine-ATE )

deprived of a functional-contextual filter at the lexical level.

Entry 2#

Source input: After the divorce, my wife said she didn't know who or what she wanted to be. When I heard that she had become a toaster, I felt vindicated. A toaster! Was that all she could be without me? And she wasn't even good at it. She could only do two slices at a time, and they came out charred on one side and white on the other. Obviously, she was the one with inadequacies. -http://www.shortshortshort.com/

Target output: Nach der Scheidung sagte meine Frau, daß sie nicht wußte, wem oder, was sie sein wünschte. Als ich hörte, daß sie ein Toaster geworden war, glaubte ich vindicated. Ein Toaster! War das alles, das sie ohne mich sein könnte? Und sie war nicht an ihm sogar gut. Sie könnte zwei Scheiben nur hintereinander tun, und sie kamen heraus ver Kohlt auf einer Seite und Weiß auf dem anderen. Offensichtlich war sie die mit Unzulänglichkeit.

( GTE )

Target output: Nach der Scheidung sagte meine Frau, daß sie nicht wußte, wem oder, was sie sein wünschte. Als ich hörte, daß sie ein Toaster geworden war, glaubte ich vindicated. Ein Toaster! War das alles, das sie ohne mich sein konnte? Und sie war nicht an ihm sogar gut. Sie könnte zwei Scheiben nur hintereinander tun, und sie kamen
heraus verkohl auf einer Seite und Weiß auf dem anderen. Offensichtlich war sie die mit Unzulänglichkeit.

**Entry 3**

**Source input:** There is nothing more precious than being loved.

**Target Output:** Hiçbir şey yoktur, sevgili daha çok, sevilen ol. (Proçeviri)

**Possible well-formed output:** Sevilmek den daha değerli hiçbir şey yoktur.

*Simply making the distinction between a clause and sentence is impracticable for MT due to syntactic differences between Turkish and English.*

**Entry 4**

**Source input:** Nikolai Goryunov has been named coach of FC Gomel after Aleksandr Kuznetsov resigned following a run of poor results for the 2003 Belarussian champions. The club lies eighth after four defeats in six matches and 51-year-old Goryunov, who returns for his second spell at the helm having led the side from 1994-96, has ten games to turn things around.


*The same word has different connotations in German and English.*

**Entry 5**

**Source input:** He is a gay and he is happy with that.

**Target output:** O, bir mutludur, ve o, onla mutludur. (Proçeviri)

**Possible well-formed output:** O bir eşcinsel(dir) ve bununla mutlu(dur).

*MT is deprived of different functions of the same word at various discourse levels in Turkish and English.*

**Entry 6**

**Source input:** If I were in your shoes, I would beat him without saying a word.

**Target output:** Wenn ich in Ihren Schuhen war, würde ich ihn schlagen, ohne ein Wort zu sagen. (ATE)

**Target output:** Wenn ich in Ihren Schuhen war, würde ich ihn schlagen, ohne ein Wort zu sagen. (GTE)

*Idiomatic usage varies between languages.*

**Entry 7**

**Source input:** He only choked the chicken for years, because there was no female for sexual intercourse. This is not his fault!
**Target output:** Er erdrosselte nur das Huhn für Jahre, weil es keine Frau für sexuellen Verkehr gab. Dieses ist nicht seine Störung! (GTE)

**Target output:** Er erdrosselte nur das Huhn für Jahre, weil es keine Frau für sexuellen Verkehr gab. Dieses ist nicht seine Störung! (ATE)

*Idiomatic and colloquial usage varies between languages.*

### Entry 8#

**Source input:** He passed away in peace.

**Target output:** Er überschritt weg in Frieden. (GTE)

**Target output:** Er überschritt weg in Frieden. (ATE)

*Idiomatic usages vary between languages.*

### Entry 9#

**Source input:** How do you feel?

**Target output:** Comment vous sentez-vous? (GTE)

A class discussion on the topic brought to the surface the fact that there is no exact French equivalent to the American "How do you feel?" since the phrase "Comment tu te sens?" relates solely to the physical well-being of someone, and "How do you feel?" is usually expressed as "Qu'est-ce que tu penses?"

---

### Works cited:

- [http://www.nationalgeographic.com](http://www.nationalgeographic.com)


• Thriveni, C. (2002) Cultural elements in translation: The Indian perspective. Translation Journal. 6/1


Authors’ Bio-data:

Fırat Açıklı is currently an instructor working for TOBB University of Economics and Technology. He also graduated from Interpretation and Translation Department of the same affiliation in 2003. He has worked at several private founding language
schools as an English teacher, and is studying on his M.A entitled 'Teaching English to the Young Blind, at the department of ELT at Hacettepe University. He is interested in, ELT research, translation, special education, material development and discourse. erciyesfirat@yahoo.com

Olcay Sert, R.A is a linguist and ELT instructor at Hacettepe University, English Language Teaching Department. His research interests include sociolinguistics, discourse analysis and critical discourse analysis, Neuro-linguistic Programming, educational linguistics and their relations to foreign language teaching and learning.

© Copyright Translation Journal and the Authors 2006
URL: http://accurapid.com/journal/37MT.htm
Last updated on: 09/16/2010 13:49:30