The Deficiencies of Machine Translation of Proverbs and Poetry: Google and Systran Translations as a case study

BOUHADIBA Malika
Université d'Oran1

Résumé

This paper attempts to highlight the deficiencies of proverb and poetic machine translation. It particularly focuses on Google and Systran's translations. The paper is in two parts. The first part deals with proverb machine translation and the stylistic, syntactic and semantic errors of this translation. The second part deals with poetic machine translation and points out its cultural misrepresentation of the source language. In an attempt to demonstrate the stylistic failures of the machine and its inability to handle cultural issues, we selected some proverbs and poems from Arabic. In the practical section of this paper, we try to highlight, not only the unfaithfulness of machine translation, but its cultural misrepresentation of the Source Language.

> *"We will only have adequate machine translation when the machine can understand what it is translating and this will be a very difficult task indeed ".*

> (Victor Yngve)[1]

Machine translation is not error-free. The mistakes made by the machine range from lexical, to syntactic, and from semantic to stylistic ones. These mistakes often occur when the machine

---

[1]. *Yngve Victor H. (1964) "Implications of Mechanical translation Research" Proceedings of the American Philosophical Society, Vol. 108, No. 4 (Aug. 27) pp. 275-281.*

translates ordinary text. They, however, abound in texts that are culturally bound such as idioms, proverbs or poetry. The cultural context of proverbs and poetry is of major importance, and since this context is out of reach for the machine, the outcome is stylistically ambiguous and culturally inappropriate translations. When the source language and the target language belong to two different families, like Arabic and English, the outcome is what Gellerstram called "Translationese " i.e., awkwardness and ungrammicality.

1. The syntactic and stylistic awkwardness of machine translation of proverbs

The major challenge to the translation of proverbs is their syntactic structure since it varies from that of common language. As Norrick notes, proverbs are "ungrammatical".[2] It is evident that stylistic rules are to be respected to provide meaningful sentences, and grammar is the backbone of any stylistically elaborated piece of writing. In fact, it is easier for the machine to provide syntactically correct translations than stylistic ones, for style also depends on structural constructions and word-order. Similarly, Nida and Taber maintain, that a translation whose style is not as correct as that of the source language, cannot be faithful.[3]

Stylistic ambiguities often occur when the machine faces lexical variations, i .e., when a word has different

---

[2]*Neal R. Norrick (1985) How Proverbs Mean: Semantic Studies in English Proverbs. Berlin: Mouton Publishers,   p.170.*
[3] *Eugene Albert Nida, Charles Russell Taber (2003) The Theory and Practice of Translation, Boston: Brill, p. 12.*

meanings in two languages. In such a case, the machine ends up providing lexical mismatches. Since it is unable to identify context, it selects any equivalent from the ones stored in its database. And when the chosen lexical item is out of context, this results in semantic ambiguity. This ambiguity also occurs when the machine does not provide the correct punctuation. It is evident that punctuation, which varies from one language to the other, is an important aspect of stylistic correctedness.

As regards cultural misrepresentation in machine translation, one might say that this is due to the fact that the machine is unable to deal with the extra-linguistic components of the source language, notably culture and context. As it is well-known, language is culture-bound and depends heavily on context, and both culture and context are not accessible to "artificial intelligence", i.e., the machine. In fact, it is rather difficult for a human translator, who is not versed into another culture to decipher the cultural connotation of a source language, least of all a machine. The latter is unable to provide a culturally appropriate translation of poetic texts, idioms or proverbs. It merely operates a lexical substitution and that ends up in "Translationese ". This is particularly due to the fact that the machine lacks the cognitive power that the human translator possesses to make sense of the different connotations, metaphors, or collocations that are embedded into poetic texts or proverbs. The machine has, hence, problems translating figurative proverbs. In fact, the machine is unable to provide the figurative meaning of proverbs since it relies on literal translation and at times, a literal translation of these proverbs

results in semantic and stylistic ambiguities, as it will be demonstrated through the following examples.

2. Systran translation of selected Arabic proverbs

<div dir="rtl">من حسنت سياسته دامت رياسته</div>

"From his politics improved his presidency lasted"

Here there is a syntactic and a stylistic error:

1) the preposition at the beginning of the sentence is needless.

2) we cannot use "lasted" after improved + preposition +noun.

<div dir="rtl">الطيور على أشكالها تقع</div>

"Birds of a feather flock together"

For this proverb, both Google and Systran give the equivalent proverb in English, rather than a literal translation as they often do.

<div dir="rtl">اتق شر الحليم اذا غضب</div>

"atq evil patient if anger"

This is an example of lexical, syntactic, stylistic and semantic ambiguity. The machine invents words for which it has no equivalents. Here it gives a phonetic reproduction of the Arabic word "اتق". The translated statement is nonsensical. An example of lexical error is: the word "الحليم". Here the cultural context is not taken into account. For us the word "الحليم" has religious connotations, since it is associated with ALLAH, so it has a positive dimension. In the machine translation we

have the noun "patient", so it has a negative dimension since it refers to an ill person. The Arabic term "الحليم" in this proverb refers to a calm and sensible person, so an adjective is required. Stylistically, the sentence is open-ended, we expect a clause or a sentence rather than a word after "if". Grammar mistakes are also due to omissions of articles, prepositions and pronouns. A more appropriate translation would be "fear the evil of the sensible and calm man if he gets angry or if he is angered".

اتّق شرّ من أحسنت إليه

"Beware the man who has received charity from you"

The translation of this proverb is more or less accurate.

تجري الرياح بما لا تشتهي السفن

"The winds occur in what the ships do not desire".

Here there is a lexical mismatch: "تجري" is translated as "occur", which means happen. We do not say winds happen, but winds blow, the syntactic error here is the wrong usage of the preposition "in".

لا تأكل خبزك على مائدة غيرك

"Erosion of bread you on table other you"

The mistakes, here, are syntactic and stylistic and are the cause of the semantic ambiguity of the statement: we do not say erosion of bread, and "other" should not be followed by a pronoun but by the conjunction "than". The word "eat" is not translated. A literal translation would be "do not eat your bread on others' table".

جولة الباطل ساعة وجولة الحق إلى قيام الساعة

"The invalid hour and tour of the truth to rising the hour".

This is an example of cultural misrepresentation. "قيام الساعة" refers to doomsday. It has a religious connotation. It relates to the language of the Holy book The Koran. It is translated as "rising the hour". The first part of the sentence is a series of lexical items that do not match those of the Source Language. Again, there is a grammar mistake: it should be till, instead of to, and the article is in the wrong place: grammatically it should be "the rising hour" and not "rising the hour".

The best example of syntactic, stylistic and semantic ambiguity or "Translationese" is the translation of the following Arabic proverb:

احذر عدوك مرة وصديقك ألف مرة فإن انقلب الصديق فهو أعلم بالمضرة

"Your enemy warns bitter and your friend bitter thousand the friend overturned so he informs in harmful".

3. Google translation of selected Arabic proverbs

لا تأكل خبزك على مائدة غيرك

"Do not eat your bread on the table, others".

Here, there is a semantic ambiguity due to wrong punctuation, the comma is in the wrong place and disrupts the meaning of the sentence, and wrong word order. Instead of using the possessive phrase: "others' table", there is "table others".

227

جولة الباطل ساعة وجولة الحق إلى قيام الساعة

"Tour falsehood hour and round the right time to do".

This translation is inappropriate and overlooks the cultural connotation of the proverb.

احذر عدوك مرة وصديقك ألف مرة فإن انقلب الصديق فهو أعلم بالمضرة

 "Be careful of your enemy is your friend once and a thosand times the capsized friend he knew Palmdharh".

The Google translation of this proverb is even worse than that of Systran. The word "capsized" is a lexical mismatch, since this verb is used for boats, and not for human beings.

إصلاح الموجود خير من انتظار المفقود

"Reform present welfare from wait lost".

Here there are lexical and syntactic mistakes: "**الموجود**" is translated as "present" whereas it should be the "existing one", or "the one at hand".

"**خير من**" means "better than". The syntactic error relates to the use the present rather than a gerund after a preposition. The other error is "wait lost".

To sum up, since cultural context is the essence of proverbs, and since the machine lacks cultural competence, it will not succeed in providing faithful translations.

4. The deficiencies of poetic machine translation

Poetry, like proverbs and idioms, is culture and context-dependent, and that is why translating it is a challenging task. The socio-cultural context of poetry is beyond the capacities of the machine, which often provides literal translations, full of lexical, semantic, structural, and stylistic mistakes. Among the other deficiencies of machine translation of poetry, there is the lack of attention to Rhyme and Meter. Yet, metrical constraints are a challenge to human translators. This is, particularly the case when the source and the target language, e.g., Arabic and English, have major structural differences. This is most probably the major cause of translation losses. As Robert Frost maintains: "poetry is what gets lost in translation". Vladimir Nabokov, on the other hand, contends that since we cannot reproduce both the meaning and content of poetry, it would be better to do without form.[4] Yet, since poetry is dependent on rhyme and meter, if form is overlooked, the outcome would be a "fade shadow" of the original text. In this connection, Nabokov maintains that "to reproduce the rhymes and yet translate the entire poem literally is mathematically impossible. But in losing its rhyme the poem loses its bloom, which neither marginal description nor the alchemy of scholium can replace". [5] This is particularly the case of Arabic poetry. In fact, sound and tone form

---

[4] *Vladimir Nabokov, (1964) Eugene Onegin: A Novel in Verse by Alexandr Pushkin, Translated from the Russian. Bollingen Foundation, IX.*
[5] *Ibid., IX.*

the core of Arabic poetry, which is marked by its strong musicality. Among the major deficiencies of poetic machine translation, is its inability to reproduce the sound, tone and mood of Arabic poetry.

Another deficiency of Arabic English poetic machine translation relates to metaphors. It is well-known that Arabic poetry is full of figurative and metaphoric language. This is, however, the case even for human translations. As Snell-Hornby notes: "The essential problem posed by metaphor in translation is that different cultures, hence different languages, conceptualize and create symbols in varying ways, and therefore the sense of the metaphor is frequently culture-specific".[6]

5. Google and Systran's unfaithful translations of selected Arabic poems

Poem: نَكِبَة دِمَشْق

أحمد شوقي)

GOOGLE Translation:

"Nakba Damascus"

SYSTRAN Translation:

"Calamity of Damascus"

سلامٌ من صَبا بَرَدى أَرَقّ

GOOGLE: "Peace of Saba Barada thinner"

---

[6] *Mary Snell Hornby, (1988) Translation Studies: An integrated Approach, Amsterdam: John Benjamin's Publishing Company. p.56.*

SYSTRAN: "Peace from reply of sleepless poured in"

Both Google and Systran do provide an accurate translation. Whereas Google uses borrowings, Systran uses words that have no equivalents in the source language.

<div dir="rtl">ودمعٌ لا يُكَفكَف يا دِمَشْقُ</div>

GOOGLE: "And tear does not Ikvkv O Damascus"

SYSTRAN: "Tear yukafkaf oh Damascus"

Both on line engines provide a sound reproduction of the terms they do not recognize.

<div dir="rtl">ومعذِرة اليَراعةِ والقوافي</div>

GOOGLE: "Sorry for the Firefly and rhymes"

SYSTRAN: "Forgiveness aalyaraae'ti waalqwaafy"

Unlike Systran, Google uses the right equivalent for **"القوافي"**

<div dir="rtl">جلالُ الرُّزْءِ عن وَصْفٍ يَدِقُّ</div>

GOOGLE: "Jalal Alrze describe beats"

SYSTRAN: "Magnificence aalrruz'i about describing knocks"

Once again, Google relies on borrowings, and reproduces the sound of the word **"جلالُ الرُّزْءِ".**

<div dir="rtl">وبي مما رَمَتْكِ به الليالي</div>

<u>GOOGLE</u>:   "Whoopi which Rmtk by night"

SYSTRAN: "In me of which threw you in him the nights"

Systran provides a syntactically awkward and semantically ambiguous sentence.

<div dir="rtl">جراحاتٌ لها في القلب عُمْق</div>

GOOGLE: "Heart surgery in depth"

SYSTRAN: "Surgeries for her in the reversing depth"

Both translations use the word surgery for the word **"جراحاتٌ"**, whereas it should be "injuries". Systran does not even provide a correct translation of the word **"القلب"** that is "the heart".

<div dir="rtl">تكادُ لروعةِ الأحداثِ فيها</div>

GOOGLE: "Almost to the splendor of events"

SYSTRAN: "Magnificence of the events was about to for in her"

The Systran translation is a series of consecutive prepositions (that are not used in the original text), so the sentence is both syntactically and semantically wrong.

<div dir="rtl">تخال من الخُرافةِ وهْيَ صِدْق</div>

GOOGLE: "Abandonment of myth Believe It"

SYSTRAN: "txaal from the myth and it believed"

The term **"صِدْق"** means truth, whereas it is translated by both as "believe", the verb is used instead of the noun.

<div dir="rtl">وقيل: معالمُ التاريخ دُكَّ</div>

GOOGLE: "It was: history landmarks flattened"

SYSTRAN: "Chief: Guideposts of the history tear down"

The phrase **"وقيل"**, which should be translated as "It was said", is wrongly translated by both Google and Systran.

<p dir="rtl">ألستِ ـ دِمَشقُ ـ للإسلام ظِئْراً</p>

GOOGLE: "Are not Damascus Islam Zira"

SYSTRAN: "The six Damascus non-Islam DHiy'raaAA"

The Systran translation here is the worst. The term: **"ألستِ"**, which means: "Aren't You", is translated as "the six", so the message of the verse is totally wrong. Besides, there is a wrong religious reference: "non-Islam" **"للإسلام"** means "for Islam", and not "non-Islam". A totally contradictory translation of the meaning of the phrase.

<p dir="rtl">ومُرْضِعَةُ الأُبوَّة لا تُعَق</p>

GOOGLE: "And nursing paternity is not hindered"

SYSTRAN: "The suckling fatherhood tue'aq"

Here, both Google Systran's translations are semantically ambiguous. The term **"مُرْضِعَة"** means: "the one who breastfeeds".

<p dir="rtl">جزاكم ذو الجلال بني دِمَشق</p>

GOOGLE: "Praise Might built Damascus"

SYSTRAN:"Your cutting having the magnificence is

brown Damascus"

In the Google translation, the **"بني"** , which means the "sons", is translated as "built", the machine relied on the phonetics of the term and provided the wrong equivalent.
In the Systran translation, the term **"جزاكم ذو الجلال"**, which means "God will reward you", is wrongly translated. Here again, the machine confused two terms that sound alike: **"جزاكم"** and **"جزا ر"**, hence the use of the term "cutting".

<div dir="rtl">وعزُّ الشرق أَوَّلُهُ دِمَشْقُ</div>

GOOGLE: "Takes Middle ole Damascus"

SYSTRAN: "The east became strong first his Damascus"

The Google translation is lexically and semantically wrong.The Systran one is more or less appropriate.

<div dir="rtl">نصرتم يومَ مِحنتِهِ أَخاكم</div>

GOOGLE: "Nasrtm day ordeal brother"

<u>SYSTRAN</u>:"Day of ordeal Christianized him brother you"

The Google translation is a series of lexical unit, with no stylistic coherence.
In the Systran one, there is a semantic and syntactic incoherence. Here again, the machine confuses two terms that sound alike: The term **"نصرتم"** which means "you championed", is translated as "Christianized". The term is confused with **"النصارى"**, "The Christians".

<div dir="rtl">وكلُّ أَخٍ بنصر أَخيه حَقّ</div>

GOOGLE: "Each brother, his brother's victory right".

SYSTRAN: "All brother in   victory of   two brothers him encloses"

The Google translation makes little sense, that of Systran is nonsensical.

<div dir="rtl">

Poem: دِمَشْق

(أحمد شوقي)

آمنت بالله ، واستثنيتُ جَدَّتَهُ

</div>

GOOGLE: "I believed in God, and excluded committee"

SYSTRAN:"Allah believed in, and excluded become mad him"

Cultural loss is particularly obvious in religiously laden language. As it is the case in this translation. There is besides the wrong equivalents in both translations. The term **"جَدَّتَه"**, means "his paradise" and not "committee". In the Systran translation, this term is related to the sound-like other term: **"جنون"**, that is "madness"

<div dir="rtl">

دِمَشْقُ رَوْحٌ ، وجَدَّاتٌ ، ورَيْحان

</div>

GOOGLE: "Damascus, spirit, and gardens, basil"

SYSTRAN:"Damascus is spirit, and paradises, and basil"

Both translations are, to a certain extent, accurate because the verse is a series of nouns. That of Systran is, however more accurate than that of Google.

<div dir="rtl">

نصيحةٌ مِلْؤُها الإخلاصُ ، صادقةٌ

</div>

GOOGLE: "Sincerity filled advice, honest

SYSTRAN: "Advice milu'haa the loyalty, truthful"

Here the term **"صادقة"** means "alms" and not "honest". Whereas the Google translation is more or less appropriate, that of Systran is semantically ambiguous.

<div dir="rtl">

والنُّصحُ خالصُه دِينٌ وإيمان

</div>

GOOGLE: "And advice pure religion and faith"

SYSTRAN: "The advising is pure his debt and faith"

Both translations are syntactically incorrect.

<div dir="rtl">

والشِّعرُ ما لم يكن ذكرى وعاطفةً

</div>

GOOGLE: "Hair unless the memory and emotion"

SYSTRAN: "The hair if not memory and emotion cherishes"

Both translations confound the word **"الشِّعرُ"**, poetry with the word **"الشعر"** hair, because the two words are written almost similarly.

<div dir="rtl">

أو حِكمَةً ؛ فهو تقطيع وأَوزان

</div>

GOOGLE: "Or wisdom; he is chopping and weights"

SYSTRAN: "Or wisdom; So he is division and weights"

The two translations are lexically and semantically inappropriate.

<div dir="rtl">

ونحن في الشرق والفُصْحى بنو رَحِم

</div>

GOOGLE: "We are in the east and classical built womb"

SYSTRAN: "We in the east and fSHY sons of uterus"

Both translation are ambiguous, but the Systran one is semantically slightly better.

<div dir="rtl">ونحن في الجُرح والآلآم إخوان</div>

GOOGLE: "We in the wound and pain Brothers"

SYSTRAN:"We in the wounding waalEElEEm brothers"

The Google translation is semantically better than the Systran one.

Poem by Ibn Zouhayr

(Praising the Prophet Mohamed, Peace be upon Him) [7]

<div dir="rtl">مسحَ   النبيُّ جبينـهُ فله بياضٌ بالخـدود</div>

GOOGLE: "Wiping his forehead prophet hath whiteness Balkhaddod"

SYSTRAN:"The prophet wiped forehead him so for him whiteness in the cheeks"

The first part of the verse is correctly translated by Google, whereas the second part is not, inversely, Systran provides a rather correct translation of the second part, giving the right equivalent for **"الخـدود"**. But it there are syntactic errors.

<div dir="rtl">وبوجهـه ديباجـة كرم الذُّبُوةِ والجُدُودِ</div>

GOOGLE: "And face the preamble to the generosity of prophecy and grandparents"

---

[7] *See* **ديوان زهير بن كعب**

*availableat :http://al-hakawati.net/arabic/civilizations
/diwanindex2a3.pdf.*

SYSTRAN: "In the face of him preamble generosity of the prophecy and the grandfathers and the grandfathers"

Both translations are lexically and semantically wrong. In fact, the output of machine translation of poetry (for example that of Google or Systran), fails to reproduce both form and meaning, and consequently, it fails to convey the message of the poet. Not only does the machine provide mistranslations, using foreignization, and borrowing, i.e., it uses the words of the source language without translating them, but it contributes to the devaluation of the aesthetic quality of the poetry it translates.

As the examples above reveal, poetic machine translation is nothing but "Translationese", to borrow Gellerstram's expression . Sam Leith maintains that Google's program to translate poetry is going to "be a fascinating failure".[8] Having used Google to translate poetry from Arabic into English, and taking into account the mess it makes of such poetry, I totally adhere to such an opinion. I deem that as with the translation of other literary genres, e.g., fiction, the translator should be acquainted with literary techniques, and should, in the case of poetry translation, have a poetic feeling. This feeling should be a prerequisite for the translation of Arabic poetry. Since the machine lacks such a feeling, it will never succeed in providing translations that are culturally accurate and aesthetically pleasing. The issue of the translatability of

---

[8] *Sam Leith (2010) "Translating Poetry Might be Beyond Google. But We'll Have Fun Watching it Try", The Guardian, 24 October .*

poetry has led, and will lead to debates between scholars. Whereas Peter Robinson (2010) considers poetry and translation as: "The Art of the Impossible", [9] as the title of his book suggests, Charles Hartman, (1996) suggests that artificial intelligence will eventually succeed in writing and translating poems.[10]

To conclude following Yehoshua Bar-Hillel's position, *"Fully automatic, high quality machine translation is impossible".[11]* This is particularly the case when the source and the target languages have different cultural and structural patterns as it is the case of Arabic and English.

References

BAR-HILLEL 1960. "The Present Status of Automatic Translation of Languages," in F. L. Alt (ed.), *Advances in Computers*, New York, London, pp. 91-163.
BURGESS, A. 1984 "Is Translation Possible?" *Translation: The Journal of Literary Translation*, XII,
3-7.
HARTMAN, C. 1996. *Virtual Muse: Experiments in Computer Poetry*, London: Wesleyan University press.

---

[9] *Peter Robinson, (2010) Poetry and Translation: The Art of the Impossible, Liverpool: Liverpool University Press.*
[10] *Charles Hartman, (1996) Virtual Muse: Experiments in Computer Poetry, London: Wesleyan University press.*
[11] *Yehoshua Bar-Hillel (1960) "The Present Status of Automatic Translation of Languages," in F. L. Alt (ed.), Advances in Computers, New York, London, pp. 91- 163*

LAKOFF, G. 1987. Women, Fire and Dangerous Things: What Categories Reveal about the Mind. Chicago: University of Chicago Press.

LEITH, S. 2010. "Translating Poetry Might be Beyond Google. But We'll Have Fun Watching it Try", *The Guardian*, 24 October .

NIDA, E. A., TABER, C. R. 2003. *The Theory and Practice of Translation,* Boston: Brill.

NORRICK,Neal R. 1985. *How Proverbs Mean: Semantic Studies in English Proverbs*. Berlin: Mouton Publishers.

 NABOKOV, V. 1964. *Eugene Onegin: A Novel in Verse by Alexandr Pushkin*, Translated from the Russian.New York: Bollingen Foundation,

ROBINSON, P. 2010. *Poetry and Translation: The Art of the Impossible*, Liverpool: Liverpool University Press.

SNELL Hornby, M. 1988. *Translation Studies: An integrated Approach*, Amsterdam: John Benjamin's Publishing Company.