
Mohammad Ghoniem
Modern Sciences & Arts Univ.
Computer Science
6th of October City, Egypt
mghoniem@msa.eun.eg

Khaled Tokal
French Univ. in Egypt
Applied Languages
Shourouq City, Egypt
khaled.tokal@ufe.edu.eg

Ahmed Y. Tawfik
Modern Sciences & Arts Univ.
Computer Science
6th of October City, Egypt
atawfik@msa.eun.eg

Abstract

This paper presents an analysis of English Arabic Statistical Machine Translation of research texts in Artificial Intelligence. We provide a typology of Machine Translation errors encountered in this analysis in order to help improve the output.

1 Introduction

The translation of terminology presents a challenge to machine translation as it requires domain specific knowledge in addition to a linguistic mastery of both the source and target languages. The focus in this paper is on the translation of Artificial Intelligence (AI) terminology from English to Arabic. A terminology that borrows from natural languages terms like perception, memory, learning, and reasoning. AI has also inherited terminology from philosophy, mathematics, statistics, medicine, and linguistics to reflect its origins. AI has also introduced its own new words and abbreviations with terms like LISP, Prolog, SOAR, and many compound and complex terms such as multiagents, semantic networks, Bayesian networks, case-based reasoning, etc.

The machine translation (MT) of technical terminology has been the focus of some recent research. Itagaki et al. (2007) examine the consistency of translating computing terminology. Shaalan et al. (2004) introduce a transfer tree methodology to map computing theses titles between Arabic and English through a series of syntactic and lexical matching. The work also identifies eight sources of difficulty in mapping terminology between the two languages.

Recently, online MT services have become popular and the number of languages supported by these services has steadily increased. The availability of parallel translations of large corpora and the development of statistical machine translation (SMT) algorithms have enabled the rapid development of some of these services (Lopez, 2008). However, the quality of the translations obtained remains an issue. Nonetheless, a recent study (Kit and Wong, 2008) has shown that a SMT online service, Google Translate, has out-performed all other systems in English to Arabic translation of legal documents.

Automatic evaluation of MT is an active research area in itself. Metrics such as BLEU (Papineni et al., 2002), METEOR (Lavie and Denkowski, 2009), and HTER (Snover et al., 2006) have been introduced to measure the degree of agreement between a machine generated hypothesis and a reference translation. However, each of these metrics introduces a different measure of distance between the hypothesis and the reference translation. BLEU measures the agreement at the level of unigrams (word) and higher order n-gram sequences. Meteor combines recall, precision, and the number of common chunks to calculate a score. Meteor also combines different matching criteria including exact matches, synonyms, and stem-level matching. HTER uses the minimum number of edits necessary to convert a given machine translation to a reference translation as a distance metric. Error typologies have been proposed for SMT. For example, Vilar et al. (2006) provides a hierarchical typology of errors for Chinese to English and Spanish to English. Here, we focus on English to Arabic translation of AI texts.
2 Evaluation of Translation Workflow

While terminology-rich text tends to be rhetorically simple, the terminology itself often extends the meaning of words in one language than the corresponding translation in the other language would allow. Therefore, the MT of terminology will have to rely on the context to identify the proper translation.

As a new tool MT has the potential of changing the workflow of the translator such that the translation process begins with the original document and a machine translated draft. The translator can then focus on modifying the draft to better reflect the intended meaning of the original document. The performance metrics of MT relevant to this new workflow are the amount of text that did not require modification, and the number of edits that were necessary by the human translator. Meteor’s exact matching module reports the number of matching words and chunks in two translations. HTER measures the number of edits needed to reach the final document starting with a machine translation.

To evaluate this workflow, we use a source collection consisting of the abstracts taken from 20 AI papers written in English. This collection has been machine translated from English to Arabic using “Google Translate”, as it is currently considered the leading online English to Arabic translation service. The machine translated collection has subsequently undergone three rounds of corrections, the first two by native Arabic speakers specialized in computer science in general and AI in particular and the third by an Arabic language specialist. The collection of corrected Arabic language abstracts serves as reference translation for reported metrics.

The percentage of words that were unchanged was slightly below half at 49.4% and the average unchanged sequence size was 5.24 words. The number of edits was dominated by substitutions (69%), followed by insertions and deletions (about 12% each), and the remaining operations were shifts. The high number of substitutions includes a fair amount of minor modifications of the words such as modifying the article attached to the word or changing the last letter to reflect the correct gender. On the level of individual abstracts, the level of agreement ranged from 10% to 80%. These results prompted us to further investigate the sources of disagreement between the two translations.

3 Qualitative Analysis of Terminology Errors

This section presents a typology of terminology errors encountered in this analysis.

3.1 Errors in Simple Terms

Errors in Acronyms: Acronyms present various problems in MT. Sometimes, the expansion of acronyms is attempted and results in irrelevant translations in the context of AI. Moreover, acronym translation presents consistency problems due to the use of synonyms. For example, “CRF” ("Conditional Random Field"), which is mistaken for “Common Report Form”, is translated as تدوين الإبلاغ الموحد namūḏaǧ al‘iblāġ almuwaḥḥad and استمرار الإبلاغ الموحد istīmārat al‘iblāġ almuwaḥḥad in the same document, where تدوين namūḏaǧ and استمرار istīmārah are two possible translations for ‘report’. When the acronym corresponds to an English word such as FLAME, the tool translates it literally to حمان ‘flame’. Besides expansion and translation, the tool has also resorted to transliteration of acronyms using Arabic alphabet such as IEEE/ACM instead of leaving it as is. Sometimes, the transliteration is partial. IEEE/ACM has been translated IEEE/ إيه سي إم.

Often, an acronym is expanded by the author at the first occurrence, then the abridged version is reused in the document. Unfortunately the tool failed to recall the translation it has suggested at the first occurrence farther in the document. For instance, QPN has been explained as “Qualitative Probabilistic Networks”, which was oddly translated as الرسوم البيانية ‘Graphical representations’. But this translation was not recalled for each occurrence of the acronym in the document.

Errors in Word Variants: The translation of English gerunds, e.g. the word “decreasing”, to an equivalent verbal noun form requires the MT to choose among several stems from the same three-letter root خفض. In this case, the MT chose نقصة tukhǧid instead of Form VII نقصة in-hişāf. Unexpectedly, words that can only be considered as verbs such as “discuss”, “provide” and “evaluates” are translated as verbal nouns مناقشة munāqašah ‘discussion’, توفر tawfīr ‘provision’,
and تقييم taqīm ‘evaluation’, respectively. Verbal nouns are not always the correct translation for English gerunds. In the phrase “Wireless sensor nodes forming a network”, the active participle الشكلةثلث forming/that forms must be used instead of the verbal noun تشكيل taškīl ‘formation’.

MT is also confused by English words that can be both verbs and nouns such as “guide”, “construct”, or “combine”. These words have been wrongly treated as nouns despite the presence of evidence to the contrary in the phrases “to guide” and “we construct”. Moreover, with no obvious reason MT does not always conform to the verb tense used in the source. Moreover, MT failed at times to decide whether a word ending with -ed is a preterit or an adjective. For instance, the adjective “dispersed” was translated as موزّف فارقاً الفارقات ‘dispersed’ (the perfective form of stem Form V) instead of موزّف متفرقة ‘dispersed’ (the active participle of stem Form V used as an adjective).

Some observed variations are due to discrepancy in grammatical number between the English reference text and the Arabic translation. For instance, the uncountable word ‘expertise’ is sometimes translated using the plural form in Arabic حوارات ﺡﺑﺮات, and sometimes using the singular form ﺡﺑرة ﺡﺑر.

3.2 Error in Complex Terms

Here, “complex terms” include both multiword phrases and compounds (single word compounds and hyphenated compounds).

Errors in Compounds: Failure to translate single word compounds is dealt with by MT in two ways: either by transliteration e.g. ميكروأري maykru’arī ‘microarray’ or by leaving the word untranslated, e.g. “multiagent”.

Surprisingly, Google Translate does not take hyphens into account while processing hyphenated compounds and does a word by word translation. For example, “context-awareness” is translated سياق الوعي siyāq alwa’y ‘context of awareness’ resulting in a detrimental word inversion in the genitive construct, instead of the correct translation الوعي بالسياق alwa’y bi alsiyāq ‘awareness of context’. Hyphenated compounds formed with the adjective -based such as “case-based”, “model-based” and “knowledge-based” have been translated as قائم qā’im, مستند mustand, and مرتكر murtakiz. In this example, the translation of the word case has varied between قضية (legal terminology as in a law case) which is inappropriate in the AI context and حالة ﺣالّة case or its plural form حالات ﺣلالات cases’.

Errors in Multiword Phrases: Disagreement in gender between the adjective and the noun it modifies has been detected 19 times across the dataset. For the sake of brevity we quote one example only (words in bold have gender problems): “temporal reasoning” is translated الفنون الزمنية الألمنتiq almanṭiq almanṭiqaniyah instead of الفنون الزمنية الألمنتiq almanṭiq almanṭiqaniyah. Disagreement in gender also happens between the verb and its subject, like in ﺳﻮور ﺐﻴﻤﻞ نأسلوáb يوزّف تمّت ﺗرﻤﻴﻢ. Disagreement in gender also affects equational sentences (noun phrases) where the predicate and the subject have different genders. Disagreement in gender occurs as well when an adjective follows a genitive construction. Sometimes, the adjective describes the head of the annexation but MT relates it to the annexed word and vice versa.

Errors concerning definiteness occur frequently. Firstly, disagreement between the adjective and the noun it describes like in ﺟﻮوث ﻣﺘﻨﺴﻂ ORM al-Gamidah instead of ﺑﺴﻤﺤﺎ نأسلوáb يوزّف ﺗرﻤﻴﻢ. Secondly, disagreement between coordinated terms like in ﺟﻮوث و ﺩﻮاير al-Tanmiyah ‘research and development’ instead of ﺑﺴﻤﺤﺎ نأسلوáb يوزّف ﺗرﻤﻴﻢ. Thirdly, the definite article may be missing from the second term of a genitive construction, such as ﺑﻴﻤﻞ ﻣﻮﺷّﺮات مفاجئة mu’ašširat mu’afagah ‘indices of surprise’ instead of ﺑﻴﻤﻞ ﻣﻮﺷّﺮات مفاجئة mu’ašširat almu’afagah ‘indices of surprise’. Fourthly, the definite article may be missing from the first term of a noun sentence like in ﺣﻞﻴﻞ ﺗﻌﻤﺪ تاﻫﺮ ل ﻋﻤﺪ al-Talhili ya’tamid ‘an analysis that relies on’ instead of ﺣﻞﻴﻞ ﺗﻌﻤﺪ تاﻫﺮ ل ﻋﻤﺪ al-Talhili ya’tamid ‘the analysis relies on’. Fifthly, the definite article is added to the second term of a noun (equational) sentence transforming it into a noun-adjective sentence.

Word order is another major source of concern in MT. Inversions are often due to word by word translation from the source language to the target language, regardless of the compatibility of constituent word order in each language.
or lack thereof. In presence of two consecutive constructions, inversions may result in relating an adjective to the wrong noun. For example, “context evolution in conceptual spaces” is translated as تطور السياق الفاعميي في المساحات atawwur al-siyāq almafāhīmi fi almisāḥāt instead of تطور السياق الفاعميي في المساحات atawwur al-siyāq fi almisāḥāt almafāhīmiyyah. The two terms of the genitive construction are sometimes inverted as well. MT also fails at ordering the words in sentences with multiple adjectives. MT put the object before the verb while translating “assembles agents” as وکلاء يتجمع wukalā’ yataǧamma’ instead of يجمع الوكلاء yaǧammi’ alwukalā’. Erros in word order establishes wrong semantic relationships among the constituents of complex terms.

4 Discussion and Conclusions

While some problems remain difficult, it appears that SMT may be improved along the following lines. Firstly, a more consistent treatment of acronyms, whether transliteration or the retention of the original latin form, is desirable. When the expanded form is provided along with acronym, it is preferable to recall the translation of the expanded form. Agreement in number between the source and the target language seems easy to achieve, but may sometimes defy the usage in the target language.

Compounds and multiword terms are more challenging than uniwor terms as their translation involves the recognition of a variety of linguistic constructs such as noun sentences, verbal sentences, noun-adjective constructs, genitive constructs, etc. By all means, word by word translations of complex terms is unlikely to yield good results especially when the source and target languages do not have the same constituent word order. Part-of-speech tagging may be a valid complementary approach to tackle this problem, as it may help distinguish genitive construction from noun-adjective constructions, and so on.

Translation from English to Arabic has yet to solve many problems in order to provide better insights of English terminology-rich texts. This work has presented a typology of errors in the output of a widely used online SMT tool, Google Translate. Techniques such as light stemming (Kadri and Nie, 2006) or part-of-speech tagging are expected to improve the quality of the output.

References


