The Interaction Between Morphosyntactic Features and the Performance of Machine Translation Tools: The Case of Google Translate, Systran, and Microsoft Bing in English-Arabic Translation

Amer M Th Ahmed

Department of English Language and Literature, Dhofar University, Salalah, Oman

Iryna Lenchuk

Department of English Language and Literature, Dhofar University, Salalah, Oman

Abstract—The present study compares the performance of three machine translation tools in English-Arabic translation to answer the questions of (a) whether the three machine translation tools, Google Translate, Systran and Microsoft Bing can be ordered in a hierarchy of performance, and (b) whether they can handle lexically and structurally ambiguous sentences and garden path sentences. Using a number of constructed and selected English sentences, the morphosyntactic features of number, gender, case, definiteness, and humanness, agreement between cardinal numerals and their head nouns, lexically and structurally ambiguous sentences and garden path sentences are used to test the three machine translation tools for performance. The results show that (a) as far as handling the morphosyntactic features of subject-verb agreement in Standard Arabic, all three machine translation tools perform equally well, and no machine translation tool seems to perform significantly better than the other two (b) some marked features (e.g. dual number and humanness) of SA seem to pose some problems for machine translation tools, and (c) lexically and structurally ambiguous sentences and garden path sentences seem to be the most challenging sentences for the three machine translation tools.

Index Terms—machine translation tools, Google Translate, Systran Translate, Microsoft Bing Translator, English-Arabic translation

I. INTRODUCTION

Machine translation (MT) refers to the use of computer programs to translate texts from a source language (henceforth SL) to a target language (henceforth TL) with or without the help of human translators with the ultimate goal of a full automation of the process (Huchins, 1995). Work on MT tools first began in the fifties of the twentieth century with two opposing approaches to development, one emphasizing the use of statistical models to develop MT tools, and the other using theoretical linguistic models to develop the tools (Huchins, 1995). Beaver (1955) is credited as the first person to suggest the use of cryptography, as was used in World War II and information theory, as discussed in Claude Shannon for natural language processing (Hovy, 1999, pp. 498-499). Neural networks or artificial neural networks are designed to mimic the way human neurons connect to one another. Each network is made up of an input node layer, one or more hidden node layers, and an output node layer. If the output node has a value above a given threshold, the node is activated and the data in that node are passed on to the next layer. If, on the other hand, the output of a given node is less than the given threshold, then the node is not activated, and the data in that node are not passed on to the next node in the network (IBM Education, 2020).

The present study is an attempt to evaluate the quality of English-Arabic translation as performed by three free online MT tools, namely Google Translate (henceforth GT), Systran Translate (henceforth Systran), and Microsoft Bing (henceforth MB). The goal of the study is to investigate the performance of these three MT tools by focusing on the complex morphosyntactic features of subject-verb agreement in the features of gender, number, case, humanness and definiteness, targeting cases of lexical and structural ambiguity, and investigating how these machines handle garden path sentences. The choice of the features of number, gender, case, humanness and definiteness is dictated by the fact that Standard Arabic (henceforth SA) exhibits complex and sometimes idiosyncratic patterns of agreement in these features; therefore, it is the belief of the authors of this paper that no proper evaluation of English-Arabic MT tools can be made if the evaluation process does not take into account the morphosyntactic features targeted in this study.

II. Research Questions

The present study aims to answer the following research questions:

(1) Can the three MT tools investigated in this study be ordered in a hierarchy of performance?

(2) Do the three MT tools encounter problems with lexically and structurally ambiguous sentences and garden path sentences?

This paper is organized as follows. Section II includes the research questions of the study. Section III is the literature review. Section IV reports on the methodology and instruments used. Section V reports on the results of the study. Section VI is a discussion of the results, and section VII concludes the paper.

III. Literature Review

The literature on English-Arabic-English MT tools falls into three groups. The first group of studies evaluates the performance of one MT tool, namely GT with a focus on the types of errors that the tool is making, be they lexical, syntactic or semantic. For example, Al-Samawi (2014) considers the errors made by GT in the English-Arabic translation of 10 texts representing different academic disciplines taken from Wikipedia, and concludes by identifying an average of 3.66 syntactic and semantic errors per sentence. Jabak (2019) feeds GT with sentences taken from the book called Thinking Arabic-English translation by Dickins, Hervey and Higgins (as cited in Jabak, 2019) and compares the output of those sentences to the human translations offered in the book. He concludes that GT makes some lexical and syntactic errors that negatively impacted the quality of the translations. Abdelaal and Alazzawie (2020) evaluate the performance of GT in the translation of informative texts from Arabic to English, and conclude that the most common types of errors are lexical errors (e.g. omission) and semantic errors (e.g. the use of inappropriate lexical items as equivalents). Aljerf (2021) evaluates the accuracy of GT in the English-Arabic translation of technical terms, and concludes that GT faces problems with words with varying prefixes, roots combined with the same suffix, compounds and blends. Sabtan et al. (2021) use comments made on social media to create a corpus of Egyptian Arabic. Using a number of sentences from that corpus, the authors feed a number of sentences into GT and evaluate the output of GT English-Arabic translation. They conclude that GT makes a number of mistakes having mostly to do with lexical addition (i.e., adding words that are not in the SL text) or omission (i.e., omitting words that are in the SL text). All of the above studies show the inadequacy of MT compared to that of human translation.

The second group of studies on English-Arabic-English MT compares the performance of GT against that of another MT tool. Al-Kabi et al. (2013) test the performance of GT against that of Babylon in the English-Arabic translation of some well-known English sayings as well as some sentences taken from websites. The authors use the automatic evaluation metric called the Bilingual Evaluation understudy (BLEU), and conclude that GT does a better job than Babylon in terms of the precision of translation. A similar conclusion has been reached in Almahasees's (2018) study where the performance of GT is tested against that of MB in the English-Arabic translation of 25 sentences taken from Petra News Agency of Jordan using BLEU as an evaluation metric. The study concludes that GT does a better job than MB. Similar studies on other language pairs can be found in Seljan et al. (2012).

The third group of studies on English-Arabic-English MT focuses on the attitudes of a specific population, namely university students in the Arab World towards the use of MT tools. Alhaisoni, E. and Alhaysony (2017) find that the majority of their Saudi university students use GT to help them read English textbooks, look up the meaning of a word, or do their writing assignments.

The present study differs from all of the literature cited above in a number of significant ways. First, unlike all previous studies on MT tools, the present study compares the performance of GT to two other MT tools, namely MB and Systran. Another important aspect which makes the present study stand out is that rather than feeding the MT tools with sentences or texts and then classifying errors into types such as lexical, syntactic and semantic, a number of structures and complex morphosyntactic features are tested by constructing sentences that target those features. For example, gender as a linguistic feature may not appear as a problematic aspect of MT between English and Arabic when random sentences are fed into the MT tools. However, it might turn out to be problematic once we take into account areas of SA morphosyntax where gender agreement is quite complex, as in the case of gender agreement between the verbs and subjects that refer to nonhumans. Complex patterns of agreement also arise when one considers gender concord between cardinal numerals and the head noun, an area which is known in the literature on Arabic linguistics to be quite complex (Ryding, 2005). A third aspect in which the present study is unique is that it includes for the first time, to the best of our knowledge, sentences that are structurally and lexically ambiguous as well as sentences that cause initial temporary parsing problems for humans, known in the literature on psycholinguistics as garden path sentences.

IV. Method

In this study, three MT tools are tested for performance against one another, and these are GT, Systran and MB. The choice of these three MT tools is dictated by two important factors. The first is that some of these tools such as GT is the most widely used MT tool by individuals around the world (Greene, 2016) and others such as Systran are widely used by corporations and organizations such as the European Commission (Wilks, 2009). The second factor that motivated the choice of these MT tools is that all three are free of charge.

GT is one of the most widely used MT tools used to employ statistical models for the purpose of translation, but since 2016 it moved to neural machine translation (Alkhawaja et al., 2020; Abdelaal & Alazzawie, 2020), when fed with a word or a phrase, it searches the Internet from the billions of documents that are translated by human expert translators in order to find a translation for the source language word or phrase. To find a good translation candidate, GT searches specifically for pages on the Internet that have identical domains with different endings such as /en or /fr. The software has then to make a decision; either that the candidate is a good candidate, and therefore it is in, or that the candidate is not a good candidate, and therefore, it is out; in other words, GT is a binary decision- making system rather than one that assigns a given weighting to a given pair. This software started as a statistical-based system, but later developed into a hybrid statistical-based system and a rule-based system (Greene, 2016). The software offers translations in more than 103 languages (Almahasees, 2018).

Systran is the oldest MT tool that initially provided translation services to the US Ministry of Defense and the European Community. It started as a rule-based MT tool, and later developed into a hybrid rule-based system and a statistical-based system (de olivera & Anastasiou, 2011; Zughoul & Abu-Alshaar, 2005).

MB started as a statistical-based system, but has recently developed into a neural system. It offers translation for more than 60 languages (Almahasees, 2018).

The following morphosyntactic features are targeted in the present study.

A. Number, Gender, Case, and Definiteness

SA is a language where adjectives agree with the head nouns in gender, number, case and definiteness (Ryding, 2005). To illustrate, consider the following examples.

(1)   a.   Laila   Taalibat-u-n   mujidda-t-u-n[1]
            L.   student.**FSG-NOM-INDEF**   hard.working-**FSG-NOM-INDEF**[2]
            'Laila is a hardworking student.'
      b.   *Laila Taalibat-u-n   mujidd-u-n
            L. student.**FSG-NOM-INDEF**   hard.working.**MSG-NOM-INDEF**
      c.   *Laila Taalibat-u-n   mujidda-t-aan
            L. student.**FSG-NOM-INDEF**   hard.working.F-**DU**.NOM.INDEF
      d.   *Laila Taalibat-u-n   mujidda-t-a-n
            L.student.FSG-**NOM**-INDEF   hard.working-FSG-**ACC**-INDEF
      e.   *Laila Taalibat-u-n   al-mujidda-t-u
            L.   student.FSG-NOM-**INDEF**   **DEF**-hard.working-FSG-NOM

The examples in (1) are all ungrammatical except for the one in (1a). The reason for the ungrammatical status of (1b-e) is the mismatch in one of the features of gender, number, case or definiteness between the nominal predicate *Taaliba* 'student' and the adjective *mujiddat* 'hardworking' modifying the nominal predicate.

B. Subject-Verb Agreement

SA is a language which exhibits two different types of subject-verb agreement depending on the order of the subject (S) and the verb (V). In the V-S order, the verb is always singular regardless of what number feature the subject has. This rule explains why (2a) is grammatical, but (2b) is not. In the S-V order, the number feature on the verb must match the number feature of the subject; otherwise, the sentence is ungrammatical, as can be shown in the contrast between (3a) and (3b).

(2)   a.   Hadhar-at   l-fataat-aani   l-hiSSat-a
            attend.Pst-F.**SG** DEF-girl.F-**DU**.NOM   DEF-class-ACC
            'the two girls attended the class.'
      b.   *Hadhar-ataa   l-fataat-aani   l-HiSSat-a[3]
            attend.Pst-F.**DU**   DEF-girl.F-**DU**.NOM   DEF-class-ACC
(3)   a.   Ɂal-fataat-aani   hadhar-ataani   l-HiSSat-a
            DEF-girl-**DU**.NOM   attend.Pst-F.**DU** DEF-class-ACC
            'The two girls attended the class.'
      b.   *Ɂal-fataat-aani   hadhar-at   l-HiSSat-a
            DEF-girl-**DU**.NOM   attend.Pst-F.**SG**   DEF-class-ACC

C. Subject-Verb Agreement in the Feature of Humanness

SA has an idiosyncratic feature, where the number and gender features of the verb in both the V-S and the S-V orders are always singular feminine whenever the subject is plural non-human (Ryding, 2005). This is shown by the contrast between (4) and (5).

---

[1] The following abbreviations are used in the study. ACC = accusative case; DEF = definite; DU = dual number; F = feminine; GEN = genitive case; INDEF = indefinite; M = masculine; NOM = nominative case; PL = plural number; Pst = past tense; SG = singular
[2] Targeted morphosyntactic features are bolded.
[3] A star * before a sentence is the standard notation used in the linguistic literature to indicate that a sentence is ungrammatical.

(4)   a.   nahaq-at   l-Hamiir-u
               bray.Pst-**F.SG**   DEF-donkey.**M.PL**.-NOM
               'The donkeys brayed.'
        b.   *nahaq-a   l-Hamiir-u
               bray.Pst-**M.SG**   DEF-donkey.**M.PL**.-NOM
(5)   a.   ʔal-Hamiir-u   nahaq-at
               DEF-donkey.**M.PL**.-NOM   bray.Pst-**F.SG**
               'The donkeys brayed.'
        b.   *ʔal-Hamiir-u   nahaq-uu
               DEF-donkey.**M.PL**.-NOM   bray.Pst-**M.PL**

D. Agreement Between Cardinal Numerals and Head Nouns

Another idiosyncratic feature of the agreement system of SA relates to the type of agreement between cardinal numerals and head nouns. The type of agreement observed is known in the Arabic linguistic literature as gender polarity or morpheme polarity (Al-qarni, 2021). For example, when the numeral is between 3 and 10, the numeral and the head noun must show inverse agreement in gender rather than normal agreement, as is shown in (6).

(6)   a.   qaraʔ-at   Laila thalaath-at-a   kutub-in
               read.Pst-F.SG L.   three-**F**-ACC   book.**M**.PL-GEN
               'Laila read three books.'
        b.   *qaraʔ-at   Laila   thalaath-a   kutub-in
               read.Pst-F.SG L.   three.**M**-ACC   book.**M**.PL-GEN
               'Laila read three books.'

E. Lexically and Structurally Ambiguous Sentences

One of the characteristic features of natural languages is ambiguity in the sense that sentences can sometimes be lexically or structurally ambiguous (Hillert, 1997; Hutchins & Somers, 1992). To illustrate, consider the following example from English.

(7)   The man put the book in the box on the table.

The sentence in (7) is structurally ambiguous, as it can mean either (7a) or (7b).

        a.   There is a box, which is on the table, and the man put the book in the box.
        b.   The man put the book in a box and put the box on the table.

Interestingly, the equivalents of structurally ambiguous sentences such as (7) are not structurally ambiguous sentences in SA. For example, (7) can be rendered into SA either as (7c) or as (7d) neither of which is structurally ambiguous.

(7)   c.   waDaʕ-a   r-rajul-u   l-kitaab-a
               put.Pst-3.M.SG   DEF-man-NOM   DEF-book-ACC
               llaðii   fii l-Sunduuq-i   ʕala
               which   in   DEF-box-GEN   on
               T-Taawilat-i
               DEF-table-GEN
               'The man put the book which is in the box on the table.'
        d.   waDaʕ-a   r-rajul-u   l-kitaab-a
               put.Pst-3.M.SG DEF-man-NOM   DEF-book-ACC
               fii l-Sunduuq-i   wa   waDaʕ-a   l- Sunduuq-a
               in DEF-box-GEN and   put.Pst-3.M.SG   DEF-box-ACC
               ʕala T-Taawilat-i
               on   DEF-table-GEN
               'The man put the book in the box and put the box on the table.'

Lexical ambiguity can be illustrated with the following example.

(8)   Give me a ring.

The lexical item ring in (8) is lexically ambiguous in that it could mean either a ring that one can wear in one of the fingers, or it could mean a phone call. Especially important in this respect is the fact that MT tools, when faced with examples such as (8), cannot use the syntactic parsing inbuilt now in most MT tools, as way of resolving the ambiguity, as the word ring in both meanings has the same syntactic category, namely a noun.

F. Garden Path Sentences

Another type of structures which is also tested in this study is the type of sentences known in the psycholinguistic literature as garden path sentences (Osterhout et al., 1994). These are sentences that initially cause parsing problems for humans in that they initially lead the hearer/reader to a wrong syntactic representation (i.e., down a garden path) before they are later assigned the right syntactic representation. Parsing is the cognitive mechanism responsible for assigning a

real-time syntactic representation to an incoming linguistic input (Gorrell, 1999). To illustrate, consider the following example from English.

(9)   The florist sent the flowers was pleased.

When speakers of English initially hear or read the sentence in (9), they are led into a garden path in that they initially assign the sentence the meaning wherein the florist is the one who sent something and the flowers are the thing that was sent by the florist. In other words, the sentence is at first parsed as ([The florist] [sent the flowers]). However, at the point where the hearers/readers encounter the rest of the sentence, i.e., [was pleased], they realize that they have assigned the sentence an initial wrong analysis, and they reassign the sentence the right analysis, where the florist is the receiver of the flowers that were sent to the florist by someone. That is, they reassign the sentence the analysis ([The florist sent the flowers] [was pleased]).

To target all of the above morphosyntactic features and structures, we constructed some sentences and selected a number of other sentences from the internet. Table 1 below is a summary of each feature or structure and the number of sentences constructed or selected for the purpose of the present study.

TABLE 1
A SUMMARY OF TARGETED FEATURES AND STRUCTURES

morphosyntactic feature or syntactic structure	total number of test sentences = 60
number, gender, case, and definiteness	10-constructed
subject-verb agreement	10- constructed
subject-verb agreement in the feature of humanness	10- constructed
agreement between cardinal numerals and head nouns	10- constructed
lexically and structurally ambiguous sentences	10 (5 lexically ambiguous sentences + 5 structurally ambiguous sentences-selected)
garden path sentences	10-selected

V. RESULTS

The findings of the study are offered in the following tables. The notation [✓] indicates that the MT tool offers a translation which conforms to the grammatical rules of SA, and is considered an acceptable translation of the SL sentence. The notation [X], on the other hand, indicates that the MT tool offers a translation where the grammatical rules of SA are not respected, and the translation offered is not an appropriate translation. Table 2 below offers the results obtained when the sentences constructed to target the morphosyntactic features of number, gender, case and humanness were fed to the three MT tools under study.

TABLE 2
NUMBER, GENDER, CASE AND DEFINITENESS

English (source)	Feature checked	GT	Systran	MB
1) Mary is an intelligent student.	Gender (feminine) agreement	ماري **طالبة** ذكية [✓]	ماري **طالبة** ذكية [✓]	ماري **طالبة** ذكية [✓]
2) Mary and Elizabeth bought tickets for the Opera	Gender (feminine) agreement	**اشترت ماري وإليزابيث** تذاكر للأوبرا [✓]	ماري وإليزابيث اشتريا تذاكر الأوبرا [X]	**اشترت ماري وإليزابيث** تذاكر للأوبرا [✓]
3) Jack is an intelligent student	Gender (masculine) agreement	**جاك طالب** ذكي [✓]	جاك طالب ذكي [✓]	**جاك طالب** ذكي [✓]
4) Mary and Elizabeth are intelligent students	Number (dual) agreement	**ماري وإليزابيث طلاب** أذكياء [X]	ماري وإليزابيث تلميذان أذكياء [X]	**ماري وإليزابيث طلاب** أذكياء [X]
5) The workers went on strike	Number (plural) agreement	وقام العمال بالإضراب [✓]	**العمال أضربوا** عن العمل [✓]	**أضرب العمال** [✓]
6) The worker went on strike	Number (singular) agreement	**دخل العامل** في إضراب [✓]	قام **العامل** بالإضراب [✓]	**أضرب العامل** [✓]
7) The two Deans rewarded the two intelligent students	Case (nominative) agreement	وقام **العميدان** بمكافأة الطالبتين الذكيتين [✓]	**لقد كافأ العمدان الطالبين الأذكياء** [✓]	كافأ **العميدان** الطالبين الذكيين [✓]
8) The two students passed the two tests	Case (nominative + accusative/genitive) agreement	وقد اجتاز **الطالبان الاختبارين** [✓]	**نجح الطالبان** في **الاختبارين** [✓]	اجتاز **الطالبان الاختبارين** [✓]
9) The writer won two important awards	agreement in indefiniteness	فاز الكاتب **بجائزتين مهمتين** [✓]	وقد فاز الكاتب **بجائزتين مهمتين** [✓]	فاز الكاتب **بجائزتين مهمتين** [✓]
10) The writer won the two prestigious awards	agreement in definiteness	فاز الكاتب **بالجائزتين المرموقتين** [✓]	فاز الكاتب **بجائزتين مرموقتين** [X]	فاز الكاتب **بالجائزتين المرموقتين** [✓]
Score out of 10		9	7	9

Of the 10 sentences considered in Table 2, GT has a score of 9 out of 10, MB 8 of 10; whereas Systran has a score of 7 out of 10. Table 3 below offers the results obtained when the sentences constructed to target subject-verb agreement were fed to the three MT tools investigated in this study.

TABLE 3
SUBJECT-VERB AGREEMENT

English (source language)	Feature checked	GT	Systran	MB
1) The girls attended the class	subject-verb agreement (in person, number and gender)	حضرت الفتيات الفصل [✔]	الفتيات حضرن الفصل [✔]	الفتيات حضرن الصف [✔]
2) The girls are listening to the Opera	subject-verb agreement (in person, number and gender)	الفتيات يستمعن إلى الأوبرا [✔]	الفتيات يستمعن إلى الأوبرا [✔]	الفتيات يستمعن إلى الأوبرا [✔]
3) The boys attended the class	subject-verb agreement (in person, number and gender)	حضر الأولاد الفصل [✔]	الأولاد حضروا الفصل [✔]	الأولاد حضروا الصف [✔]
4) The boys are listening to music	subject-verb agreement (in person, number and gender)	الأولاد يستمعون إلى الموسيقى [✔]	الأولاد يستمعون إلى الموسيقى [✔]	الأولاد يستمعون للموسيقى [✔]
5) The two girls attended the class	subject-verb agreement	حضرت الفتاتان الفصل [✔]	الفتاتان حضرتا الصف [✔]	حضرت الفتاتان الصف [✔]
6) The two girls are listening to music	subject-verb agreement (in person, number and gender)	الفتاتان تستمعان إلى الموسيقى [✔]	الفتاتان تستمعان إلى الموسيقى [✔]	الفتاتان تستمعان للموسيقى [✔]
7) The two boys attended the class	subject-verb agreement (in person, number and gender)	حضر الصبيان الفصل [✔]	الأولاد حضروا الصف [✔]	حضر الصبيان الصف [✔]
8) The two boys are listening to music	subject-verb agreement (in person, number and gender)	الصبيان يستمعان إلى الموسيقى [✔]	الولدان يستمعان إلى الموسيقى [✔]	الصبيان يستمعان للموسيقى [✔]
9) The girl attended the class	subject-verb agreement (in person, number and gender)	حضرت الفتاة الفصل [✔]	حضرت الفتاة الفصل [✔]	الفتاة حضرت الصف [✔]
10) The boy attended the class	subject-verb agreement (in person, number and gender)	حضر الصبي الفصل [✔]	حضر الصبي الفصل [✔]	حضر الصبي الصف [✔]
Score out of 10		10	10	10

Out of a total of 10 sentences, GT and MB each has a score of 10 out of 10; whereas Systran has a score of 09 out of 10.

Table 4 below offers the results obtained when the sentences targeting agreement between the subject and the verb in the feature of humanness are tested.

TABLE 4
SUBJECT-VERB AGREEMENT IN THE FEATURE OF HUMANNESS

English (source)	Feature checked	GT	Systran	MB
1) The girls are eating their food.	Agreement in humanness between the verb and the subject	**الفتيات يأكلن** طعامهن [✔]	**الفتيات يأكلن** طعامهن [✔]	**الفتيات يأكلن** طعامهن [✔]
2) The dogs are eating their food.	Agreement in humanness between the verb and the subject	**الكلاب تأكل** طعامها [✔]	**الكلاب تأكل** طعامها [✔]	**الكلاب تأكل** طعامها. [✔]
3) The boys are eating their food.	Agreement in humanness between the verb and the subject	**الأولاد يأكلون** طعامهم [✔]	**الأولاد يأكلون** طعامهم [✔]	**الأولاد يأكلون** طعامهم [✔]
4) The leaves of the tree are falling.	Agreement in humanness between the verb and the subject	**أوراق الشجرة تتساقط** [✔]	**أوراق الشجرة تتساقط** [✔]	**أوراق الشجرة تتساقط** [✔]
5) The girls ate their food.	Agreement in humanness between the verb and the subject	**الفتيات أكلن** طعامهن. [✔]	الفتيات أكلن طعامهن. [✔]	**أكلت الفتيات** طعامهن. [✔]
6) The boys ate their food.	Agreement in humanness between the verb and the subject	**أكل الأولاد** طعامهم. [✔]	كان الصبية يأكلون طعامهم. [✔]	**أكل الأولاد** طعامهم. [✔]
7) The dogs ate their food.	Agreement in humanness between the verb and the subject	**أكلت الكلاب** طعامها. [✔]	**تأكل الكلاب** طعامها. [✔]	**أكلت الكلاب** طعامهم. [✔]
8) The cats ate their food.	Agreement in humanness between the verb and the subject	**أكلت القطط** طعامها. [✔]	**أكلت القطط** طعامها. [✔]	**أكلت القطط** طعامها. [✔]
9) The airplanes have just taken off.	Agreement in humanness between the verb and the subject	لقد **أقلعت الطائرات** للتو. [✔]	لقد **أقلعت الطائرات** للتو. [✔]	لقد **أقلعت الطائرات** للتو. [✔]
10) The teenagers ate their food.	Agreement in humanness between the verb and the subject	**تناول المراهقون** طعامهم. [✔]	وكان **المراهقون يأكلون** طعامهم. [✔]	**أكل المراهقون** طعامهم. [✔]
Score out of 10		10	10	10

Out of a total of 10 sentences, all three MT have a score of 10 out of 10. Note that what is of interest is the agreement in the feature of humanness between the subject and the verb. Therefore, the translation offered by MB of sentence (7) above is considered correct even though the target translation is ungrammatical in SA, as the pronominal object هم– does not agree in humanness with the verb أكلت.

Table 5 below offers the results obtained when the agreement between cardinal numerals and their head nouns is targeted.

TABLE 5
AGREEMENT BETWEEN CARDINAL NUMERALS AND THEIR HEAD NOUNS IN GENDER

English (source)	Feature checked	GT	Systran	MB
1) I bought three chairs	Inverse gender agreement between the numeral (F) and the noun (M) in the noun phrase	اشتريت **ثلاثة كراسي** [✓]	اشتريت **ثلاثة كراسي** [✓]	اشتريت **ثلاثة كراسي** [✓]
2) I bought nine chairs	Inverse gender agreement between the numeral (F) and the noun (M) in the noun phrase	اشتريت **تسعة كراسي** [✓]	اشتريت **تسعة كراسي** [✓]	اشتريت **تسعة كراسي** [✓]
3) I bought ten chairs	Inverse gender agreement between the numeral (F) and the noun (M) in the noun phrase	اشتريت **عشرة كراسي** [✓]	اشتريت **عشرة كراسي** [✓]	اشتريت **عشرة كراسي** [✓]
4) I bought ten tables	Inverse gender agreement between the numeral (M) and the noun (F) in the noun phrase	اشتريت **عشرة طاولات**. [X]	اشتريت **عشرة موائد** [X]	اشتريت **عشرة طاولات** [X]
5) I bought thirteen chairs	Inverse agreement on the first part of the numeral (F) and head noun (M)	اشتريت **ثلاثة عشر كرسيًا** [✓]	اشتريت **ثلاثة عشر كرسيا** [✓]	اشتريت **ثلاثة عشر كرسيا** [✓]
6) I bought nineteen chairs	Inverse agreement on the first part of the numeral (F) and the head noun (M)	اشتريت **تسعة عشر كرسيًا** [✓]	اشتريت **تسعة عشر كرسي** [✓]	اشتريت **تسعة عشر كرسيا** [✓]
7) I bought sixteen forks	Inverse agreement on the first part of the numeral (M) and the head noun (F)	اشتريت **ستة عشر شوكة** [X]	اشتريت **ستة عشر شوكة** [X]	اشتريت **ستة عشر شوكة** [X]
8) I bought thirteen toothbrushes	Inverse agreement on the first part of the numeral (M) and the head noun (F)	اشتريت **ثلاثة عشر فرشاة** أسنان [X]	اشتريت **ثلاث عشرة فرشة** أسنان [✓]	اشتريت **ثلاثة عشر فرشاة** أسنان [X]
9) I bought nine paintings	Inverse agreement on the first part of the numeral (M) and the head noun (F)	اشتريت **تسع لوحات** [✓]	اشتريت **تسع لوحات** [✓]	اشتريت **تسع لوحات** [✓]
10) I bought four pencils	Inverse agreement on the first part of the numeral (F) and the head noun (M)	اشتريت **أربعة أقلام** رصاص [✓]	اشتريت **أربعة أقلام** رصاص [✓]	اشتريت **أربعة أقلام** رصاص [✓]
Score out of 10		7	8	7

Out of a total of 10 sentences, GT and MB each has a score of 7 out of 10; whereas Systran has a score of 8 out of 10. Table 6 offers the results obtained when lexically and structurally ambiguous sentences are targeted.

TABLE 6
LEXICALLY AND STRUCTURALLY AMBIGUOUS SENTENCES

English (source)	Feature checked	GT	Systran	MB
1) The man put the book in the box on the table	Structural ambiguity	وضع الرجل الكتاب في الصندوق على الطاولة [X]	وضع الرجل الكتاب في الصندوق على الطاولة [X]	الرجل وضع الكتاب في الصندوق على الطاولة [X]
2) I killed the king with the knife	Structural ambiguity	لقد قتلت الملك بالسكين [X]	قتلت الملك بالسكين [X]	قتلت الملك بالسكين [X]
3) I killed the king with the red hair	Structural ambiguity	لقد قتلت الملك بالشعر الأحمر [X]	قتلت الملك بالشعر الأحمر [X]	قتلت الملك بشعر أحمر [X]
4) Chop the tree with the leaves	Structural ambiguity	اقطع الشجرة بالأوراق [X]	أقص الشجرة بأوراق الشجر [X]	فرم الشجرة مع الأوراق [X]
5) Draw the man with the sunglasses	Structural ambiguity	ارسم الرجل بالنظارات الشمسية [X]	أرسم الرجل مع النظارات الشمسية [X]	رسم الرجل مع النظارات الشمسية [X]
6) Give me a ring	Lexical ambiguity (homonymy)	أعطني الخاتم [X]	أعطني خاتما [X]	أعطني خاتما [X]
7) I bought a pen for my dog	Lexical ambiguity (polysemy)	اشتريت قلما لكلبي [X]	اشتريت قلم لكلبي [X]	اشتريت قلم لكلبي [X]
8) I am looking at the face of the clock	Lexical ambiguity (polysemy)	أنا أنظر إلى وجه الساعة [X]	أنا أنظر إلى وجه الساعة [X]	أنا أنظر إلى وجه الساعة [X]
9) The crane is picking a small fish	Lexical ambiguity (polysemy)	الرافعة تلتقط سمكة صغيرة [X]	الرافعة تقطف سمكة صغيرة [X]	الرافعة تلتقط سمكة صغيرة [X]
10) I bought ten rulers	Lexical ambiguity (polysemy)	اشتريت عشرة حكام [X]	اشتريت عشرة حكام [X]	اشتريت عشرة حكام [X]
Score out of 10		0	0	0

Out of a total of 10 sentences, all three MT tools have a score of zero out of 10.
Table 7 offers the results obtained when garden path sentences are targeted.

TABLE 7
GARDEN PATH SENTENCES

English (source)	Feature checked	GT	Systran	MB
1) The horse raced past the barn fell	Parsing issue	تسابق الحصان أمام سقط الحظيرة [X]	أسرع الحصان بعد سقوط الحظيرة [X]	الحصان تسابق عبر الحظيرة سقط [X]
2) The old man the boat	Parsing issue	العجوز القارب [X]	الرجل العجوز القارب [X]	الرجل العجوز القارب [X]
3) The florist sent the flowers was pleased	Parsing issue	أرسل بائع الزهور وسر [X]	أرسل بائع الأزهار الأزهار مسرورا [X]	بائع الزهور أرسل الزهور كان مسرورا [X]
4) The complex houses married and single soldiers and their families	Parsing issue	يضم المجمع منازل المتزوجين والجنود العزاب وعائلاتهم [X]	منازل المجمع متزوجة وعزباء وعوائل [X]	يضم المجمع الجنود المتزوجين والعاز بأسرهم [X]
5) The cotton clothing is made of grows in Mississippi	Parsing issue	الملابس القطنية مصنوعة من تنمو في ميسيسيبي [X]	الملابس القطنية مصنوعة من النباتات في الميسيسيبي [X]	يرصد الملابس القطنية من ينمو في ولاية ميسيسيبي [X]
6) The man who hunts ducks out on weekends	Parsing issue	الرجل الذي يصطاد البط في عطلات نهاية الأسبوع [X]	الرجل الذي يطارد البط في عطلة نهاية الأسبوع [X]	الرجل الذي يطارد البط في عطلة نهاية الأسبوع [X]
7) Fat people eat accumulates	Parsing issue	يتراكم الناس الدهون يأكلون [X]	الدهون تأكل التراكمات [X]	الناس الدهون أكل يتراكم [X]
8) Until the police arrest the drug dealers control the street.	Parsing issue	حتى تعتقل الشرطة تجار المخدرات يسيطرون على الشارع . [X]	حتى تعتقل الشرطة تجار المخدرات المتحكمين بالشارع. [X]	حتى تعتقل الشرطة تجار المخدرات يسيطرون على الشارع [X]
9) When Fred eats food gets thrown.	Parsing issue	عندما يأكل فريد يتم طرح الطعام . [X]	عندما يأكل فريد الطعام يرمى. [X]	عندما يأكل فريد الطعام يحصل ألقيت [X]
10) Mary gave the child the dog bit a bandaid	Parsing issue	أعطت ماري الطفل ضمادة على الكلب [X]	ماري أعطت الطفل الكلب قطعة بندية [X]	ماري أعطت الطفل الكلب عضة [X]
Score out of 10		0	0	0

Out of a total of 10 sentences, all three MT tools have a score of zero out of 10.

Table 8 below is a summary of the scores achieved by all three MT tools on the 60 test sentences used in this study, and their overall performance given in percentages.

TABLE 8
OVERALL PERFORMANCE OF THE THREE MT TOOLS

MT	Overall Performance
GT	36/60 = 60%
Systran	35/60 = 58.33%
MB	36/60 = 60%

Table 9 below shows the overall performance of all three MT tools when the problematic lexically and structurally ambiguous sentences and garden path sentences are excluded from the count.

TABLE 9
THE OVERALL PERFORMANCE OF THE THREE MT TOOLS (PROBLEMATIC STRUCTURES EXCLUDED)

MT	Overall Performance
GT	36/40 = 90%
Systran	35/40 = 87.5%
MB	36/40 = 90%

VI. DISCUSSION

The present study was an attempt to answer the following research questions:

(1) Can the three MT tools investigated in this study be ordered in a hierarchy of performance?

(2) Do the three MT tools encounter problems with lexically and structurally ambiguous sentences and garden path sentences?

To answer the first research question, it is helpful to consider Tables 8 and 9, which provide a summary of the overall performance of all three MT tools considered in this study. When we consider how each of the three MT tools fared in terms of their observance of the morphosyntactic features of SA, all three MT tools show evidence of sensitivity to the morpho-syntactic features of SA with an overall score of 36/40 for GT and MB and 35/40 for Systran. Based on these figures, we can answer the first research question by claiming that all three |MT tools considered in this study seem to perform well in English-SA translation. This means that as far as observance of the morphosyntatic features of subject-verb agreement in SA are concerned, there does not seem to be a difference in performance among the three MT tools considered in this study. In other words, all three MTs fare well in this respect, and not one of them seems to perform better than the other two. Our conclusion then is that the first research question receives a negative answer, as no hierarchy of performance can be provided among the three MT tools.

Notwithstanding the above, there does seem to be some evidence to suggest that dual number, which is a cross-linguistically marked number, and the interaction between the dual number and the gender feature on the agreeing verb pose a problem for MT tools (see for example the performance of all three MT tools on sentence 4 of table 2 above, and the performance of Systran on sentence 2 of table 2 above). This might explain the errors made by MT tools in table 2 above. There also seem to be evidence to indicate that another cross-linguistically feature, namely that of humanness, pose another problem for all three MT tools. That explains why all three MT tools made mistakes in table 5 above. Of particular interest in this regard is the observation that agreement between compound cardinal numerals (11-19, for instance) and their head nouns seem to be more challenging for all three MT tools than single numerals (see for example examples 7 and 8 of table 5 above).

When we, on the other hand, consider how the three MT tools fare with lexically and structurally ambiguous sentences and garden path sentences, we notice a consistent failure of all three tools in handling such sentences. Thus, the second research question posed in this study is positively answered, as these structures pose a serious problem for all three MT tools. One reason why the lexically ambiguous structures considered in this study are problematic for MT tools could be that the source of the polysemy cannot be resolved by recourse to the syntactic category difference between the target term and its polysemous term, as the target term in all the five lexically ambiguous sentences shares the same syntactic category with its polysemous word. For example, the term *ring* in *give me a ring* (= *give me a call*) and in *give me a ring* (= *give me a golden or silver ring to wear*) is a noun. The same applies to the other lexically ambiguous terms used in the test sentences of table 6 above. As for garden path sentences, it seems that unlike humans who, after an initial failure in assigning the right syntactic representation to such sentences, seem to be able to overcome their initial failure, MT tools seem at the time of writing this paper doomed to failure when dealing with such sentences. This, together with the consistent failure of MT tools to provide two translations for the two interpretations of syntactically ambiguous sentences suggests that syntactic ambiguity is a very thorny area for MT tools to grapple with.

VII. CONCLUSION

One of the major conclusions of this study is that no hierarchy of performance seems possible when we pit the three MT tools considered in this study against one another. This is because all three MT tools seem to be doing well when it

comes to SA subject-verb agreement in a number of features. That said, there does seem to be some evidence to indicate that marked cross-linguistically marked morphosyntactic features such as those related to agreement with a subject which is dual in number or agreement in humanness with a subject in SA are more problematic for MT tools than unmarked features.

Another major conclusion of this study is that lexically ambiguous sentences where the target word and its polysemous term share the same syntactic category seem to be challenging for MT tools.

A third major conclusion of this study is that syntactic ambiguity in the form of a sentence with two ambiguous interpretations or of the form known as garden path sentences in the psycholinguistic literature is a serious problem for MT tools.

<div style="text-align:center">REFERENCES</div>

[1] Abdelaal, N. & Alazzawie, A. (2020). Machine Translation: The Case of Arabic English Translation of News Texts. *Theory and Practice in Language Studies*, *10*(4), 408-418. http://doi.org/10.17507/tpls.1004.09

[2] Alhaisoni, E. and Alhaysony, M. (2017). An Investigation of Saudi EFL University Students' Attitudes towards the Use of Google Translate. *International Journal of English Language Education*, *5*(1), 72-82. http://doi.org/10.5296/ijele.1 0696

[3] Aljerf, R. (2021). An Investigation of Google's English-Arabic Translation of Technical Terms. *Eurasian Arabic Studies*, *14*, 16-37.

[4] Al-Kabi, M., Hailat, T., Al-Shawakfa, E. & Alsmadi, I. (2013). Evaluating English to Arabic Machine Translation Using BLEU. *International Journal of Advanced Computer Science and Applications*, *4*(1), 66-73.

[5] Alkhawaja, L., Ibrahim, H., Ghnaim, F. & Awwad, S. (2020). Neural Machine Translation: Fine-Grained Evaluation of Google Translate Output for English-to-Arabic Translation. *International Journal of English Linguistics*, *10*(4), 43-60. http://doi.org/10.5539/ijel.v10n4p43

[6] Almahasees, Z. (2018). Assessment of Google and Microsoft Bing Translation of Journalistic Texts. *International Journal of Languages, Literature and Linguistics*, *4*(3), 231-235. http://doi.org/10.18178/ijlll.2017.3.1.100

[7] Alqarni, M. (2021). No gender polarity in Arabic numeral phrases. *Linguistic Inquiry*, *52*(3), 441-472. https://doi.org/10.1162/ling_a_00385

[8] Al-Samawi, A. (2014). Language Errors in Machine Translation of Encyclopedic Texts from English into Arabic. *Arab World English Journal*, *3*, 182-211.

[9] Beaver, W. (1955). Translation. In W. N. Locke and A. D. Booth, (Eds.), *Machine Translation of Languages*. MIT Press.

[10] de olivera R. & Anastasiou, D. (2011). Comparison of SYSTRAN and Google Translate for English→ Portuguese. *revista. tradumatica technologie de la traduccio*, 118-136.

[11] Deng, L. & Yang, L. (2018). *Deep Learning in Natural Language Processing*. Springer. https://doi.org/10.1007/978-981-10-5209-5

[12] Gorrell, P. (1999). Sentence processing. In Wilson, Robert A. and Keil, Frank C., (Eds.), *The MIT encyclopedia of the cognitive sciences*, (pp. 748-751). MIT.

[13] Greene, L. (2016). *Everything you ever wanted to know about Google Translate, and finally got the chance to ask*. Retrieved October 10, 2023 from https://www.taus.net/insights/reports/everything-you-ever-wanted-to-know-about-google-translate-and-finally-got-the-chance-to-ask.

[14] Hillert, D. (1997). Language in Time: Lexical and Structural Ambiguity Resolution, in M. Stamenov (Ed.), *Approaches to Language and Consciousness*. John Benjamins.

[15] Hovy, E. (1999). Machine translation. In Wilson, Robert A. and Keil, Frank C., (Eds.), *The MIT encyclopedia of the cognitive sciences*, (pp. 498-499). MIT.

[16] Huchins, W. (1995). Machine translation: A brief history. In E.F.K. Koemer and R. E. Asher (Eds.) *Concise history of the language sciences: from the Sumerians to the cognitivists*, (pp. 431-445). Pergamon Press. https://doi.org/10.1016/C2009-1-28311-0

[17] Hutchins, W. & Somers, H. (1992). *An introduction to machine translation*. Academic Press.

[18] IBM Education. (2020). *Neural networks*. Retrieved September 15, 2023 from https://www.ibm.com/cloud/learn/neural-networks.

[19] Jabak, O. (2019). Assessment of Arabic-English translation produced by Google Translate. *International Journal of Linguistics, Literature and Translation* (IJLLT), 238-247. http://doi.org/10.32996/ijllt.2019.2.4.24

[20] Osterhout, L. Holcomb, P. & Swinney, D. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20: 786–803. http://doi.org/10.1037//0278-7393.20.4.786

[21] Ryding, K. (2005). *A Reference Grammar of Modern Standard Arabic*. CUP.

[22] Sabtan, Y, Hussein, M., Ethelb, H. & Omar, A. (2021). An Evaluation of the Accuracy of the Machine Translation Systems of Social Media Language: Google's Arabic-English Translation as an Example. *International Journal of Advanced Computer Science and Applications*, *12*(7), 406-415.

[23] Seljan, S., Vičić, T & Brkić, M. (2012). BLEU Evaluation of Machine-Translated English-Croatian Legislation. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation – LREC* (pp. 2143-2148). http://doi.org/10.13140/RG.2.1.4374.3204

[24] Wilks, Y. (2009). *Machine translation: its cope and limits*. Springer. http://doi.org/10.1007/978-0-387-72774-5

[25] Zughoul, M. & Abu-Alshaar, A. (2005). English/Arabic/English machine translation: A historical perspective. *Meta*, *50*(3), 1022-1041. http://doi.org/10.7202/011612ar

Amer M Th Ahmed holds a PhD in Linguistics and Applied Linguistics from York University, Toronto, Canada.

Before joining Dhofar University, he taught ESL in federally-funded programs and private institutions, and taught courses in linguistics in universities in Toronto, Canada. Currently, he holds the position of Associate professor at the Department of English Language and Literature, Dhofar University, Oman.

He has been published in a number of internationally refereed journals.

Iryna Lenchuk holds a PhD in Linguistics and Applied Linguistics from York University, Toronto, Canada.

Before joining Dhofar University, she taught ESL in federally-funded programs, colleges and private institutions, and taught ESL/TESL and courses in linguistics in colleges and universities in Toronto, Canada. Currently, she holds the position of Associate professor at the Department of English Language and Literature, Dhofar University, Oman.

She has been published in a number of internationally refereed journals.