A comparative analysis of verb tense and aspect in Arabic and English using Google Translate

Jawharah Alasmari[1, a], Janet C.E. Watson [2, b], Eric Atwell[3, c]

[1] School of Languages Cultures and Societies, University of Leeds, UK
Princess Nourah bint Abdulrahman University. Riyadh, Saudi Arabia
[2] School of Languages Cultures and Societies, University of Leeds,
[3] School of Computing, University of Leeds, UK
[1]ml14jsna@leeds.ac.uk,[2] j.c.e.watson@leeds.ac.uk [3]E.S.Atwell@leeds.ac.uk

ABSTRACT
The aim of this study is to examine the challenges of handling verb tense and aspect in Arabic to English machine translation. A small corpus of selected Arabic sentences was submitted to Google Translate for a contrastive analysis of Arabic and English verb tense use. The main purpose of this study is to provide an understanding of morphology and forms of Arabic and English verbs in their syntactic context, in order to reveal details that can be used in current machine processing systems.
Keywords: **Google Translate, Arabic & English verbs, morphology, tense**

1. INTRODUCTION

The Arabic language has a verb system which differs from that used in English. Various different studies have been undertaken over the space of a thousand years to examine the differences between the verb systems used in Arabic and those used in other languages (Eisele, 1990; Zollmann et al., 2006). It is often classed as a highly inflectional, templatic language. On the basis of syntax, Arabic can be categorised as a theme pro-drop language that expresses person, number, and/or gender agreement, as well as tense, aspect, and modality, with the referent on the verb. The main purpose of this research is to provide an understanding of the morphemic composition of Arabic and English verbs in their syntactic context in order to see how this relates to tense and aspect, and to reveal details that can be used to build a corpus.

In order to examine the morphology that is encapsulated in the inflection of tense, aspect, and in the modality markers of verb systems so that Arabic can be translated into English using machine translation (MT), steps must be taken to deal with the order of words and their agreement. Agreement rules are vital for a comparison of various sentences with the target language. The rules must be proficient enough to deal with this challenge, and these rules must act as guidelines for the organisation of sentence elements (Gadalla & Abdel-Hamid, 2000).

In Arabic, a verb is formed by the insertion of three فَعَلَ *faʿala* to four فَعَّلَ *faʿlala* root consonants into one of a set of verb patterns. Suffixes and prefixes are then affixed in these templates to locate positioning in the linear structure to express number, person, and gender. The active/passive voice and perfective aspect is used for the dummy root (Truck, 2010).

Abdul-Halim et al. (2015, p. 139) define inflection as follows: "Inflectional affixes are those which are affixed to words to indicate grammatical function." They also note that "the Arabic language is one of the inflectional languages, whereas English shows only a few features of these languages".

One of the greatest challenges faced when designing a machine translation system that will translate between Arabic to English is that the Arabic language does not use a specified formula to construct the aspect of the verb in the same way English does. Reishaan and Ja'far (2008) explain that in the case of an absence of certain formal markers for the progressive and perfect aspects of the Arabic verb, Arab scholars denote them by using certain auxiliaries before the verb form, such as قد *qad, س sa-,* and "سوف", in terms of tense.

Arabic has only two verbal forms, a prefix conjugation and a suffix conjugation. The English language has sixteen tense forms (Gadalla, 2006). In Arabic verbs, derivational and inflectional morphology categories play an equal main role in the verb system. Tense and aspect in Arabic verbs are formed by using the rules of inflectional morphology (Al-Saleemi, 1987).

This study will explore the verb systems of both English and Arabic, paying attention to the similarities and differences between them. In order to do this, it is necessary to explain the morphemic composition of verb forms as they appear in modern utilization.

2. METHODOLOGY

The aim of this research is find the best method for analysing the two languages to deal with the translation of sentences. A list of sentences was created to begin the process of a contrastive analysis of the languages. Some of these sentences were suggested by Janet Watson, and some were chosen from short stories. No specific formula was used to select the sentences because one of the aims of this research is to compare Arabic sentences to their equivalents in the English language. After an analysis of these sentences was completed, an Access program was used to sort all seventeen examples to clarify the analysis. The examples recognised three fields: two fields for sentences in English and their translation into Arabic, and one other field for Arabic/English transliteration. The Arabic verbs were analysed and annotated as follows: verb form, derivation, meaning, inflection (prefixes, affixes), stem, root, person, number, gender, tense, verb mood, syntax, and morphology. Google Translate was used as an example of an existing machine translation system to show how words are employed and translated in the sentences.

3. ANALYSIS OF THE RESULTS

The perfect aspect verb in Arabic was recognised in English as the past tense, as follows:

Sentence

A.	*ḏahaba*	*al-walad-u*	*ilā*	*al-madrasat-i*
B.	*go. PERF.3MS*	*DEF-boy-NOM*	*to*	*DEF-school-GEN*

The boy <u>went</u> to the school.

Google Translate used the past simple to express past actions.
In English, to form the singular imperfect verb form (present simple tense), the suffix 's' is added to the base form of the verb. In the following example, the tense of the verb يقرأ *yaqra'u* 'he reads' is employed in Arabic using a free order verb-subject-object (VSO)

structuring. In Arabic, differences in meaning between a verb and noun in a sentence are related to context.

Sentence
  A. *yi-qra'-u*   *al-ṭālib-u*   *al-dars-a*
  B. *3MS.IMPF-read-IND*   *DEF-student-NOM*   *DEF-lesson-ACC*
The student <u>reads</u> the lesson.

Google Translate was not able to translate verbs by recognising the prefixes and affixes added to the verb forms in terms of number (dual), and, as Tucker (2010) notes, English tenses do not follow the same patterns as Arabic tenses. In Arabic, a derivative of a root may comprise suffixes and prefixes (Badawi et al., 2013). Therefore, manual correction was needed to add dual markers to the verb *ʕaraḏ-ā*. In the following example, the long vowel in the verb *ʕaraḏ-ā* and the suffixed pronoun *humā* relating to the subject of the sentence is not recognised in English.

Sentence
  A. *iṯnāni mina l-tullāb-i ʕaraḏ-ā   ʕarḏa-humā fi l-ḥalaqat   al-yawma*
  B. *Two of DEF-students   present-DU their work-DU   at DEF-seminar   today*
 Two of the students <u>presented</u> their work at the seminar today.

In the following example, the future tense in Arabic is formed using ـس *sa-*, سوف *sawfa*, whereas in English the future tense is formed using `will 'or `shall`. As Gadalla (2006) notes, "This occurs when that construction refers to future arrangements or planned events."

Sentence:
  A. *sawfa u-sāfir-u   ilā   sūrīya   al-ʔusbūʕ   al-qādim*
  B. *sawfa 1SM-travel-IND to Syria DEF-week DEF*next
  I <u>will</u> travel to Syria next week.

One of the challenges posed by machine translation when translating between Arabic and English is that, in English gender distinctions are not made using verbs; they are made using pronouns (Gadalla & Abdel-Hamid, 2000).Google Translate did not recognise the gender of an Arabic verb. The masculine is derived in the prefix conjugation verb using the prefix *ya-yaktubu* يكتبُ, whereas the feminine is derived by using the prefix *ta- taktubu* تكتبُ. In the following example, a manual correction was required to alter the gender of the verb *ta-ḥmil-u*. One lesson to be learned from this is that machine translation must be able to understand and distinguish the elements of a verb (Gadalla, 2006). In the following example, the present simple is used to denote the habitual verb action *ta-ḥmil-u* تحملُ- as part of a metaphorical description.

Sentence:
  A. *wa-baqāyā   nafḥāt al-baḳūr - allatī   ta-ḥmil-u   rāʾiḥa min ʿabq at-tārīx,*
  B. *And   the remnants of hints that   3SF-carry-IND   smell of fragrance DEF-history*
  And the remnants of hints that <u>carry</u> incense smell fragrant history.

Google Translate was not be able to recognise affixes that are added to the verb form to indicate the active voice/first singular person, and, therefore, an analysis of the whole

context of the sentence was required to recognise the number of subjects and the tense or aspect of the verb أتسلق *'atasallaqu.*

In Arabic, the prefix conjugation is indicated using the prefixes *ya, ta, na, and a`*. Examples of this rule are: يكتبُ *yaktubu 'he writes',* تكتبُ t*aktubu 'she writes',* أكتبُ *aktubu 'I am writing'.* The dual/Pl يكتبا *yaktuba* and نكتبُ *naktubu* are used to express present tense verbs and numbers with the same consonantal root to convey related meaning (Truck, 2010). In the following example, it is important to understand the whole meaning in context in order to recognise the aspect of the verb (imperfect verb progressive). Gadalla (2006) explains that, "the corpus shows that this translation is employed when the imperfect form denotes an activity that is occurring at or around the moment of speaking."

Sentence
  A. *'atasallaq-u   al-salālim-a   al-ḥajarīyah*
  B. *1S-climb-IND   DEF-stairs-ACC   DEF-stone*
I <u>am climbing</u> the stone stairs.

In the following example, Google Translate provided an inaccurate translation of the verb يقُولُونَ *yaqūlūna* 'they say'. The sentence uses present form to indicate present continuous tense of the verb يقُولُونَ *yaqūlūna* 'they say', and a pronoun *ū,* واو الجماعة is added to the verb to indicate plural/ number. Thus, in the following example, the Arabic verb *yaqūlūna*, is used to describe continuous present actions that can only be recognised by reading the whole sentence. Therefore, machine translation must be able to recognise the present tense and present progressive.

Sentence
  A. *'in   lam   yatawaqqaf   al- ṣubīy-an   çammā yaqūl-ūna   sa-ynāluhum   çiqāb.*
  B. *CON   NEG   3SM.IMPRF-yatawaqaf-JUSS   DEF-ṣubīy-M.PL   çammā IMPRF-yaqūl-3PM.PRON   FUT-IMPRF-ynalahum-3PM.PRON   çiqāb*
  C. *If the boys don not stop what they are saying, they will get a punishment*

Arabic does not have the facility to distinguish between vowels and consonants in the same way that English does. Furthermore, the utilisation of diacritics (a small sign indicated on top or sometimes underneath letters such as كتبت *katabtu* 'I wrote', كتبتِ *katabti* 'you wrote' (F), كتبتَ *katabta* 'you wrote' (MS) acts as a guide to pronunciation. In Arabic, vowels cannot be spelled out in a text whereas in English they can be.

4. CONCLUSION

From the above analysis it can be seen that there are significant differences between Arabic and English in respect of morphological rules and word structure, particularly in respect of the verb systems in the two languages. The system of free word-order in the Arabic language makes it difficult to translate by machine translation, especially when the aim is to create logical sentence structure in English. Exact equivalents to English tenses are not present in the Arabic language. Both in the Arabic and English, the perfect aspect are utilised for finished actions in relation to the present moment, a moment in the future or a moment in the past, but there is no a specified formula to construct the aspect of the verb in the same way English does (Reishaan, 2008), and this adds extra complication to the translation process. These differences pose challenges for designers of translation machines. Furthermore, vital

comparison tools such as syntactic analysis and morphological analysis are needed in order to improve machine translation systems that seek to translate between Arabic to English. It could be argued that these differences create a gap in meaning and sense, which makes the translation process much more complicated.

In order to undertake the process of translation, human translators need to acquire three particular types of knowledge. The first is knowledge of the source language in terms of morphology, semantics and syntax in order to understand the source text. The second is knowledge of the target language in terms of morphology, semantics and syntax in order to produce an understandable, satisfactory and acknowledgeable target text. The third type is information relating to the subject matter. This permits the translator to recognize and comprehend the precise context of what is being translated from the source language so that it is possible to translate items and structures of syntax and make appropriate matches in the target language (Michalski, 2009). After completing a comparison between the two languages, it is possible to conclude the following from machine translation:

1. The perfect aspect in Arabic can be used for the past tense in English.
2. The imperfect aspect in Arabic can be used for the past continuous in English (when an action takes a long time to complete or has a continuous effect).
3. The imperfect aspect in Arabic can be used for an activity that is happening at the moment of speaking.
4. The imperfect aspect in Arabic with the future particle *sa-* or *sawfa* (future simple constructer) can be used to refer to planned events or future arrangements, translated into English using 'will' or 'shall'.
5. The imperfect aspect in Arabic can be employed to express a habitual action relating to the verb.

5. REFERENCES

Abdul-Halim, M., Shamsan, A., and Attayib, A. M. (2015). Inflectional morphology in Arabic and English: a contrastive study. International Journal of English Linguistics. 5, 139.

Abdul-Hussein, R, k.,and Ali, J, A.(2008). Time, Tense, and the verb Form in Arabic and English: a Contrastive Study.2-1 العددان ,مجلة القادسية في الآداب والعلوم التربوية.

Al-Saleemi, E. (1987). A contrastive study of the verb systems of English and Arabic. Masters Thesis, Durham University.

Badawi, E. S., Carter, M., and Gully, A. (2013). Modern written Arabic: a comprehensive grammar. Routledge. Printed in New York.

Eisele, J. (1990). Time Reference, Tense and Formal Aspect in Cairene Arabic. In E. Mushira (Ed.) Perspectives on Arabic Linguistics, Vol. 1. John Benjamins Publishing. Amsterdam. pp. 173-212

Reishaan,A. K. (2008). The Relationship between Competence and Performance: Towards a Comprehensive TG Grammar. آداب الكوفة,*1*(2).

Gadalla, H. A. H. (2006). Arabic imperfect verbs in translation: a corpus study of English renderings. Meta: Translators' Journal. 51, 51–71.

Gadalla, H. A. H., and Abdel-Hamid, M. E. (2000). Genitive constructions in English and Arabic: a contrastive study. Bulletin of the Faculty of Arts, vol. 6. Assiut University. Egypt. pp. 1–64.

Tucker, M. A. (2010). The Morphosyntax of the Arabic Verb: Toward a Unified Syntax-Prosody. Research report, University of Califrnia at Santa Cruz.

Zollmann, A., Venugopal, A., and Vogel, S. (2006). Bridging the inflection morphology gap for Arabic statistical machine translation. Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL, June 4-9, 2006, New York, pp. 201–204.

Michalski, M. (2009). Adjectives in hypotaxis: proposed dimensions for a Description of Syntagms in Modern Written Arabic Lingua Posnaniensis, DOI: 10.2478/v10122-009-0003-0 51(1)