This is a repository copy of *A comparative analysis between Arabic and English of the verbal system using google translate*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/112953/

Version: Accepted Version

A comparative analysis between Arabic and English of the verbal system using google translate

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

[1, 2] School of language, Culture and Societies, University of Leeds, The United Kingdom

[3] School of computing, University of Leeds, The United Kingdom

[1] ml14jsna@leeds.ac.uk,[2] j.c.e.watson@leeds.ac.uk [3] E.S.Atwell@leeds.ac.uk

ABSTRACT

The Arabic language has not been widely studied in computational terms. Therefore, 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 that differs from that of English. Various 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). Arabic is surrounded by myths. It is classed as an inflectional, derivational and 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 markers with the referent on the verb. Every particular inflection of the verb is quantified exclusively. The main purpose of this research is to provide an understanding of the morphemic composition and forms of Arabic and English verbs, as well as examining discharges and conjugates in their syntactic context in order to reveal details that can be used to build a corpus.

In order to examine the morphology that is encapsulated in the inflection of tenses, 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 فعلل (faʿlala) consonant roots into one of numerous verb patterns. The verb suffixes and prefixes are then affixed in these templates to locate positioning in the linear structure in relation to number, person and gender. The active/passive voice and perfective tense/aspect are 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"،"س" , and "سوف", in terms of tense.

Arabic has only two tense forms, while the English language has sixteen tense forms (Gadalla, 2006). In Arabic verbs, the 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).

The proposed study will explore the verb systems of both the English and Arabic languages, 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.

Question research:

- How are the different verb forms that express tenses used in the Arabic and English languages?

- Does the Arabic language have a perfect tense in the same way that English does?

- Do both languages express the form morphologically through definite forms, then assert or modify it contextually?

- What are the most common challenges facing translators when translating the verb tense from Arabic into English?
- Does morphological and syntax analysis of the whole context can help to improve the translation of google?

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 Professor Janet in our second meeting, and some were chosen from short stories previously read by myself. 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 recognised 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 in the sentences.

3. Analysis of the Results
The past simple verb (regular) in Arabic was recognised in English as follows:

Sentence:

    A. *ḏahab-a   al-walad-u   ilā   al-madrasat-i*

    B. *The boy   PERF.3SM-went-ACC to the school-NEG*

    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 yiqraʔu يقرأ 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. <u>ya-qra</u>'-u   *al-ṭāalib-u   al-dars-a*

    B. *DEF- student-MON IMPRF.3SM.read-IND DEF- book-ACC*

    The boy <u>reads</u> the lesson.

Google is not able to translate verbs by recognising the prefixes and affixes that are 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-a. In the following example, the long vowel iatna`ni relating to the subject alfa3l is not recognised in English.

Sentence:

    A. *iaṯna`ni mina al-tulab-ṭ "ʕaraḏ-a ʕarḏa-humā   fi halaqaat-i al-yawmi"*

    B. *Two of the students PERF.3SM-present-PORN.DUM   their work-PRON.DU at the seminar today*

    Two of the students <u>presented</u> their work at the seminar today.

In the following example, the present progressive tense in Arabic (future simple constructer) is formed using sa ـس, sawafaسوف whereas in English the present progressive 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 ua-sāfir-u   ilā   sūrīya   al-ʔusbūʕ-a   al-qādim-i*

    B. *Sawfa-IPERF.1SM-travel-IND to Syria DEF-next DEF-week*

    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 and Abdel-Hamid, 2000). Google Translate cannot recognise the gender of an Arabic verb. The masculine is derived through the 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-ḥml 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 -bqāyā   nafḥāti al-baḳawr-i - allatay   ta-ḥmil-u   rayʾiḥatann min ʿabaqi al-ta-taārayḳ-i,*
   B. *And   the remnants of hints that   IPERF-3SFcarry-INDI   incense smell   fragrant DEF-history-NEG*
   And the remnants of hints that <u>carry</u> incense smell fragrant history.

Google translate is 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 of the verb (atasalluqu/أتسلق).

In Arabic, the present tense is indicated using the prefixes ya, ta, na, and a`. Examples of this rule are: "yaktubu يكتبُ, taktubu نكتبُ, aktubu أكتبُ. The dual/Pl yakutba يكتبا and naktubu نكتبُ 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. *A'-tasallaq-u al-ssalālim-a al-hajaray-i*
   B. *PERF.1SM –Climb-INDI   DEF-stairs-ACC   of DEF-stone*
   I <u>am climbing</u> the stairs of stone.

In the following example, Google translate provides an inaccurate translation of the verb يقُولُونَ/yaqūlūna. The sentence uses present form to indicate present continuous tense of the verb يقُولُونَ/ yaqūlūna, 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 yatawaqaf-u al-ssuby-an ʿammā yaqūl-ūna   sa-yanālahum   ʿiqāb.*
   B. *In   lam IMPERF-yatawaqaf-3PM.PRON   DEF-ssuby-an ʿammā IMPERF-yaqūl-3PM.PRON. sa- IMPERF-ynalahum-3SM.PRON eiqab*
    If the boys don not stop what they are saying, they will get a punishment

After completing a comparison between the two languages, it is possible to achieve the following from machine translation:

1. Past simple tense in Arabic can be used for the equal tense in English.
2. Past tense progressive can be used to convey past continuous (when an action takes a long time to complete or has a continuous effect).

3. Present progressive can be used for an activity that is happening at the moment of speaking.
4. Present progressive tense (future simple constructer) can be used to refer to planned events or future arrangements, using 'will' or 'shall'.
5. Present simple can be employed to express a habitual action relating to the verb.

4. Conclusion

From the above analysis it can be seen that there are very significant differences between Arabic and English in respect of morphology rules and word structure, particularly in respect of verb systems in the two languages. The prolificacy of free word-order in the Arabic language makes it difficult to translate, especially when the aim is to create logical sentence structure in English.

In both Arabic and English, the perfect aspect is utilised for completed actions in relation to the present moment, a moment in the future or a moment in the past. The widespread supposition held by translators that Arabic verbs are narrowed to limited indications of past, present and future is not correct (Fayyad, 1997) .This is due to the inefficiency of objectionable methodology and the atmosphere of teaching. Poor knowledge, inadequate and undisciplined grammar practice, and an understanding of syntactical and morphological instruction must be coupled. Occasionally, the problem in the Arabic-English translation due to a lack of understanding of morphological conditions (infixation vs. prefixation), different meanings by individual form, and the recognition of particular tools contained in contexts. This adds extra complications to the translation process. These differences pose challenges for designers of translation machines. Furthermore, vital comparison tools are required, such as syntactic analysis and morphological analysis, in order to improve the machine translation systems that aim to translate between Arabic and English. It could be argued that these differences create a gap in meaning and sense, making 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 (Morphology `prefix (es)-stem-suffix (es), semantic, and also syntax) in order to understand the connotations of the relative text. The second is knowledge of the target language (Morphology 'prefix (es)-stem-suffix (es), semantic, and also syntax) in order to produce an understandable, satisfactory and acknowledgeable 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).

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 مجلة القادسية في الآداب والعلوم التربوية، العددان.

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. Printed in 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. Santa Cruz, CA 95064-1077

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)

Fayyad, S. (1997) Azminat Al-Fi'l Al-Arabi Al-Nahwiyyah: Arba'ata Ashara Zamanan [Grammatical Tenses of the Arabic Verb: Fourteen Tenses]. Beirut: Dar Al- Mustaqbal Al-Arabi. 80 pp.

Biodata

	Jawharah Alasmari
	PhD student
	(Arabic, Islamic and Middle Eastern Studies)
	School of language, Cultures and Societies. University of Leeds
	Backgrounds:
	Master's degree from the Faculty of Arts and humanities in the Arabic language and literature "morphology, Syntax and linguistics"
	Bachelor's degree in arts and education are allocated, "Arabic language" from the Faculty of Education for Girls in Jeddah.

	Professor Janet C.E. Watson Leadership Chair for Language at Leeds Areas of expertise: documentation of Modern South Arabian languages; Arabic dialectology; phonology; morphology; acoustic and instrumental phonetic
	Eric Atwell Associate Professor Areas of expertise: data mining; text analytics; computing; linguistics; language; semantics; ontology; Arabic; Quran; English; artificial intelligence; machine learning; knowledge based systems; heath informatics; autism