



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/112950/>

Version: Accepted Version

Proceedings Paper:

Alasmari, J, Watson, JCE and Atwell, ES (2016) A comparative analysis of the Arabic and English verb systems using a Quranic Arabic corpus. In: IMAN'2016 4th International Conference on Islamic Applications in Computer Science and Technologies. IMAN'2016 4th International Conference on Islamic Applications in Computer Science and Technologies, 20-22 Dec 2016, Khartoum, Sudan.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

A comparative analysis of the Arabic and English verb systems using a Quranic Arabic corpus

A CORPUS-BASED APPROACH

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

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

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

ABSTRACT

The Quranic Arabic corpus is one of the most important computational tools that has been produced in Arabic language service. Therefore, the main purpose of this papers is to provide some details of morphological and syntactic structures of Arabic and English verbs through deep computing studies of the Quran. The paper will also highlight some investigations into the use of a sub-verb corpus, along with translations, in order to consider how Quranic contexts employ verb forms to indicate time and how Arabic verbs are rendered into English.

Keywords: Quranic Arabic corpus, Arabic verb, English translations, Sub-corpus verb.

1. Introduction

Corpus is ‘a collection of a single writer's work, or of writing about a particular subject (Cambridge Dictionary, 2016). Nowadays corpora have become one of the main computational tools used in teaching and learning languages. The importance of corpora in language studies is that corpora can contain a large amount of information about all areas of language. This tool can be used in some linguistics research areas such as language teaching and learning, applied linguistics, lexicography, etc. As well as for other purposes such as morphology analysis, grammar analyzing, lemmatization, parsing and it also can be used as a tool of bilingual concordance (Alfaifi, 2014). The present paper aims to explain how Quranic Arabic corpus can be used to explore the verb systems of both languages, paying attention to the similarities and differences between them, and provide an understanding of morphology and forms of Arabic and English verbs in their syntactic context.

2. Literature review

2.1-The importance of using a corpus of Arabic language texts in language studies

In using empirical data, the corpus is a tool “that enables linguists to make objective statements rather than those which are subjective” (Alansary & eta, 2014). The corpus allows empirical analysis of a large number of texts related to linguistics or language such as Grammar, lexicography, semantics, natural language processing and other language studies that could not be done in any other way (ibid). The corpus allows students who are studying a foreign language to be able identify more possible contexts in which to respond to the word. Though, currently researchers who are interested in the English language use a corpus as a powerful tool for

learning and teaching or the development of machine translation, the use of a corpus in Arabic language studies has not received enough attention (Alfaifi, 2014).

Arabic words can be changed according to their moods (nominative الرفع, accusative النصب or genitive الجر). Using a corpus can help the researchers to discover the changes that occur to a word. For example, the differences between the word "الطالبان" and "الطالبين" happen according to the differences in its case moods. Alansary & eta (2014) make a significant point about the benefit of using a corpus-based approach using empirical methods that the theoretical method usually used to study language issues. Grammarians have discovered a corpus to be a useful resource in investigating the grammar (syntax) or (semantics), of a language. Therefore, the empirical data and the representative quantification language variety make corpora a useful tool for syntactical research.

2.2 The need of using Corpus in Grammar

A-Investigating morphological characteristics

Using a corpus for morphological analysis allows the users to search for prefixes and suffixes and infixes that can be added to the word. For example: the Arabic verb كتب kataba may give various tenses by adding different prefixes or suffixes, كتَبَ kataba, يكتب yaktubu, سيكتب sa yaktubu.

B-The distribution and function of a syntactic construction:

The investigation of the distribution of words, can enable grammarians to carry out the rules of the language's syntax restrictions (Alansary & eta, 2014). For example: in English, the Present Perfect can be used with unspecific expressions such as: ever, never, once, many times, several times, before, so far, already, yet, etc .

3.3 The need for using parallel corpora for the English–Arabic languages:

As mentioned above, the Arab world's lack of using a parallel corpus leads to uncertainty and doubts of the significance of lexical data in bilingual dictionaries, grammar (syntax) and parallel translations in English. The effectiveness of developing a parallel corpus for English-Arabic has not been given enough attention in the Arab world (Al-Ajmi, 2004). Many linguistic corpora have been created by those more interested in computational usage or the development of machine translation systems than in actual linguistics research (Alansary & eta, 2014). Therefore, the corpora they use are designed differently from the corpora designed to work with grammatical information or characteristics observable in syntactic contexts that in the past might be obtained only by manual analysis of the characteristic structures of English and Arabic. Making a comparison using a parallel corpus such as a Quran corpus which includes several translations in English can help us to understand the differences in meaning and grammar in different contexts and the technical problems faced in translation (Al-Ajmi, 2004) Therefore, Using Quran corpus in my research, will show us all the examples of a verb in context, then we can find all the situations in which different tenses and aspects of the verb and its phrases and how Arabic verbs are translated into English.

C- Using Quranic Arabic corpus of the verb system studies

The Arabic language is a diverse language in nature and it uses a unique 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 (Eiesel, 1990; Zollmann et al., 2006). However, in spite of this, the Arabic language has not been widely studied in computational terms (Eiesel, 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 (Gadalla & Abdel-Hamid, 2002).

In Arabic, a verb is formed by the insertion of three to four consonant roots into one of numerous verb patterns. The verb suffixes and prefixes are then affixed in these templates to locate positioning in these templates to locate positioning in the linear structure in relation to: number, person, gender, tense etc. The active/passive voice and perfective tense/aspect is used for the dummy root (Truck, 2010)

The understanding of the correspondence between the verb-form and the concept of time whether past, present or future is one of the important aspects of any language. Both grammarians of Arabic and English languages have made many attempts to describe the correspondences (Reishaan & Ja'far, 2008). A discussion of this topic is intended to provide understanding of the differences between the completeness and incompleteness of actions, their points and periods, simultaneous and successive events, and the similarities between then (past) and then (future)" (ibid).

In fact, one of the main purpose that make both grammarians of Arabic and English languages interested in this topic is to find a resolution to the greatest challenges to reliable translations between Arabic and 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. . For example: "yaktubu يكتُبُ, taktubu تكتبُ, aktubu أكتبُ , Dual/Pl. yakutba يكتبا and naktubu نكتبُ" are used to express present tense verbs and numbers with the same consonantal root conveying related meaning, whereas the suffixes(*tu* and *-nā*) "katbtu كتبتُ, ktabna كتبنا" indicate past tense verbs. The passive voice from the root (K-T-B كتب) is "kutiba كُتِبَ" which means 'it was written' (Bahloul, 1994).

The understanding of the morphemic composition and forms of Arabic and English verbs, as well as examining discharges and conjugates in their syntactic context using a parallel corpus is a vital step in order to reveal details in Arabic texts translations (Shamaa, 1978).

3. Methodology

In the Quranic Arabic corpus, there are 1,475 verbs with a large number of contexts; for example, there are 1,618 contexts of the pos:v (i) root "قول". Sometimes, verbs have a different syntactic and morphological analysis, whereas at other times they use the same format and meaning in context.

At first stage, all the context for most common verbs "qaala/قال" and "kaana/كان," in the verses will be considered by building a specific corpus of these verbs with their translations. They will then be compared with their equivalents in the target translation and analysed in terms of syntactic and morphological features. The study will consider the main translation errors that arise. These sentences will be analysed in accordance with the structure of TAM markers, vowels, gender and person etc. A frequency count of the different verb constructions in the two languages will be performed to explain ways in which Arabic verbs can be rendered into English. A sup-corpus of the verbs with their contexts will be randomly chosen from the Quranic Arabic

| Sanin international | Pickthall | Yusuf Ali | Shakir | Munammad Sarwar | Mohsin Khan | Arberry |
|---------------------|-----------|----------------|-----------|---|----------------|------------|
| says | saith | says | says | explained, "It must be neither too old nor to | says | says |
| says | saith | says | says | says | says | says |
| says | saith | says | says | says | says | says |
| says | saith | saith | says | commands | says | says |
| says | saith | say | say | say | say | say |
| says | saith | say | say | pray | say | say |
| says | saith | saith | says | orders | says | say |
| say | say | say | say | say | say | say |
| say | say | say | say | say | say | saying |
| says | saith | saith | saith | commands | say | says |
| says | saith | says | says | ask | says | say |
| say | say | say | say | ask | say | say |
| say | say | say | say | ask | say | say |
| say | say | say | say | say | say | say |
| say | say | say | say | say | say | say |
| will say | will say | will say | will say | will ask | will say | saying |
| say | say | say | say | say | say | say |
| says | saith | says | says | says | says | says |
| says | saith | says | says | his words | says | says |
| mentions | speak | talks | says | speaks | talks | says |
| will say | saith say | will say | shall say | will say | will say | shall say |
| will say | will say | will say | will say | will ask | will say | shall say |
| say | say | say | say | ask | say | shall say |
| say | saith | say | says | say | say | say |
| will say | will say | will say | will say | say | will say | will say |
| will say | saith | will ask | will say | will ask | will say | shall say |
| will say | will say | will say | will say | saying | will say | shall say |
| said | said | say | said | say | said | said |
| said | said | said | said | telling | said | said |
| will say | will say | will say | shall say | will say | will say | will say |
| will say | will say | say | will say | will say | will say | shall say |
| said | to say | could only say | said | said | could only say | was saying |
| will say | will say | will say | shall say | asks | will say | shall say |
| will say | say | will say | say | will say | will say | will say |

3. Analysis of the Results

Table 3: the use of the verb yaqūlu/يقول/says to indicate multiple tenses/aspects in its several translations

| The verb form in its translations | Sahih | | | | | | | |
|--|---------------|-----------|-----------|--------|-----------------|-------------|---------|-------|
| | International | Pickthall | Yusuf Ali | Shakir | Muhammad Sarwar | Mohsin Khan | Arberry | Total |
| present simple | 30 | 35 | 33 | 29 | 32 | 31 | 28 | 218 |
| past | 4 | 2 | 1 | 4 | 2 | 3 | 4 | 20 |
| present perfect continuous | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 2 |
| future | 23 | 18 | 20 | 22 | 17 | 21 | 21 | 142 |
| passive | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 2 |
| perfect (would, used to, may, might, should) | 1 | 3 | 6 | 3 | 0 | 4 | 2 | 19 |
| infinitive | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 2 |
| noun | 0 | 0 | 0 | 1 | 5 | 0 | 4 | 10 |
| past continuous | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 2 |
| Comprehensive translation of meaning | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 3 |
| Conditional perfect | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Total | 60 | 60 | 60 | 60 | 60 | 60 | 60 | 420 |

Tables 1 and 2 indicate the following:

1. The present verb yaqūlu/يقول/says indicative mood was recognised in English as follows: says, saith, commands, mentions, speaks, talks.
2. In its translations, the verb yaqūlu/يقول/says was recognised in multiple tenses. Consider the following examples:

Table 4: the use of the verb *yaqūlu/يقول*/says to indicate multiple tenses in its several translations

| The verse | Sahih International | Pickthall | Yusuf Ali | Shakir | Muhammad Sarwar | Mohsin Khan | Arberry |
|---|---------------------|-----------|-----------|---------|-----------------|-------------|---------|
| إِذْ يَقُولُ الْمُنَافِقُونَ وَالَّذِينَ فِي قُلُوبِهِمْ مَرَضٌ غَرَّ هُوَلاءِ دِينُهُمْ | said | said | say | said | say | said | said |
| وَلِيَقُولَ الَّذِينَ فِي قُلُوبِهِمْ مَرَضٌ وَالْكَافِرُونَ مَاذَا أَرَادَ اللَّهُ بِهَذَا مَثَلًا | will say | may say | may say | may say | say | say | may say |

3. Recognition of the prefixes and affixes that are added to the verb forms in terms of number was dissimilar in all the sentences that provided the same tense. For example, the verb *yaqūlu/يقول* was employed in the singular form (say) but in its translations it was provided without the prefix (s) in the following example:

Table 5: the use of the verb *yaqūlu/يقول*/says to indicate the singular person, with its translations

| The verse | Sahih International | Pickthall | Yusuf Ali | Shakir | Muhammad Sarwar | Mohsin Khan | Arberry |
|--|---------------------|-----------|------------|------------|-----------------|-------------|---------|
| فَيَقُولُ رَبِّ لَوْلَا أَخَّرْتَنِي إِلَىٰ أَجَلٍ قَرِيبٍ | says | saith | should say | should say | say | says | says |

4. The verb corpus reveals that the present verb indicates the future tense through context (verbal clues and current time (القرائن اللفظية والحالية), not through form. Consider these examples:

Table 6: the use of the verb *yaqūlu/يقول*/says to indicate future tense, with its translations

| The verse | Sahih International | Pickthall | Yusuf Ali | Shakir | Muhammad Sarwar | Mohsin Khan | Arberry |
|---|---------------------|-----------|-----------|----------|-----------------|-------------|-----------|
| فَيَقُولُ الَّذِينَ ظَلَمُوا رَبَّنَا أَخْرْنَا إِلَىٰ أَجَلٍ قَرِيبٍ | will say | will say | say | will say | will say | will say | shall say |

5. An analysis of the whole context of a sentence is required by the translators in order to consider the agreement features of the text between the verb and other elements, such as subjects, pronouns, particular tools that change the tense of verb. For example: the Arabic imperfect form (فعل مضارع) can be constructed indicate the future tense if it comes in the context with specific particular tools such as: emphatic prefix لام التوكيد *lām*. In this case, it can be considered that the verbs can be rendered into English by the future tense:

Table 7: the use of the verb yaqūlu/يَقُولُ/says to indicate future tense, with its translations in the context with specific particular tools such as: emphatic prefix *lām* لام التوكيد

| The verse | Sahih International | Pickthall | Yusuf Ali | Shakir | Muhammad Sarwar | Mohsin Khan | Arberry |
|--|---------------------|-----------|---|------------------------|-----------------|-------------|----------|
| وَلَيَقُولَنَّ الَّذِينَ فِي قُلُوبِهِمْ مَرَضٌ وَالْكَافِرُونَ مَاذَا أَرَادَ اللَّهُ بِهَذَا مَثَلًا | will say | may say | may say they declare (with emphasis) | may say | say | say | may say |
| وَلَيْنَ سَأَلْتَهُمْ لَيَقُولُنَّ إِنَّمَا كُنَّا نَخُوضُ وَنَلْعَبُ | will say | will say | | would certainly say | say | declare | will say |

- Sometimes, in Arberry's translation, the noun (saying) was used to translate the verb yaqūlu/say يَقُولُ to indicate the present tense (Leech and Svartvik, 1975), and the Arabic language also has this ability to express time using nouns: (اسم الفاعل واسم المفعول/derivatives), roots and participles. However, in the selected examples, the verb yaqūlu/ يَقُولُ was used to indicate time using its inflectional morphology, and translating the verb form, as has been done in other translations, may help to clarify time with more accuracy.
- The study of the verb yaqūlu/ يَقُولُ may need to contain the following syntactic categories:
 - Grammatical relation (subject, direct object, indirect object, pronouns)
 - Grammatical category (person, number, tense, aspect, mood, gender, case, voice...)
- Semantic analysis of word contexts can help to provide an understanding of the context of what was said or to locate its time or place.

4. Conclusion

In a holistic view it can be concluded that the Quran corpus is one of the most important computational tool that has been produced in Arabic language service. It provides learners with what they need in the field of language, linguistics and computational studies. It also paves the way for researchers to study morphological and syntactic structures through deep computing studies of the Quran. In particular, a discussion of the differences and similarities between Arabic and English verb systems can help to provide some details which will be used to improve the machine translation of Arabic into English (MT).

References:

- Alfaifi, A., Atwell, E. And Hedaya, I. (2014). Arabic Learner Corpus (ALC) v2: A New Written and Spoken Corpus of Arabic Learners. In proceedings of the Learner Corpus Studies in Asia and the World (LCSAW) 2014, 31 May - 01 Jun 2014. Kobe, Japan.
- Alansary, S, Nagi, M, Adly. N. (2008) Building an International Corpus of Arabic (ICA): Progress of Compilation Stage. Bibliotheca Alexandrina, P.O. Box 138, 21526, El Shatby, Alexandria, Egypt.
- Al-Ajmi, H. (2004). A New English–Arabic Parallel Text Corpus for Lexicographic Applications. Lexikos 14 (AFRILEX-reeks/series: 14: 2004): 326-330.

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.

Tucker, M. A. (2010). The Morphosyntax of the Arabic Verb: Toward a Unified Syntax-Prosody. Santa Cruz, CA 95064-1077

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

The Cambridge Advanced Learner's Dictionary & Thesaurus © Cambridge University Press, 2016.

Shamaa, N. (1978): A Linguistic Analysis of Some problems of Arabic to English Translation, Ph.D. Dissertation, Linacre College, Oxford.

Nida, E. 1964 .Toward a Science of Translating, Leiden, Netherlands, E. J. Brill.

Leech, G and J. Svartvik .1975. A Communicative Grammar of English, London: Longman, (2nd and 3rd editions: 1994, 2002).

Bahloul, Maher. 1994. The Syntax and Semantics of Taxis, Aspect, Tense and Modality in Standard Arabic. Ithaca: Cornell University Department of Modern Languages and Linguistics.

Gadalla, h. & Abdel-Hamid, a. 2000. Genitive constructions in English and Arabic: a contrastive study. Bulletin of the faculty of arts, vol. 6, pp. 1-64. Assiut University, Egypt.

الكتب العربية

نصيرة غزالي. (2009). أزمنة الفعل في اللغة العربية وأثرها في التنوع الدلالي. رسالة دكتوراه.



السامرائي. (1983). إبراهيم الفعل زمانه وأبنيته. بغداد: مطبعة العاني.

أبو محمد عبد الله بن مسلم بن قتيبة (1393). تأويل مشكل القرآن الدينوري، السيد أحمد صقر، مكتبة دار التراث، ط2

قواقزة، محمد حسن. (2015). الدلالة الزمنية للأسماء في اللغة العربية في اللغة العربية : اسم الفاعل واسم المفعول. المجلد 42، العدد (1) .

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" |

| | |
|---|---|
| | <p>Bachelor's degree in arts and education are allocated, "Arabic language" from the Faculty of Education for Girls in Jeddah.</p> |
|  | <p>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</p> |
|  | <p>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; health informatics; autism</p> |

