Extract the Semantic Meaning of Prepositions at Arabic Texts: An Exploratory Study

Mohammad Khaled A. Al-Maghasbeh[-1], Mohd Pouzi Bin Hamzah[-2]

[1, 2-]School of Informatics and Applied Mathematics, Universiti Malaysia Terengganu, Kuala Terengganu, Malaysia

Abstract

Arabic language considers the most complex language. It widely prevalent among the population and it contains a large vocabulary and grammar rather than other languages, so it has a lot of challenges; one of these challenges, it is how to predict of the correct meaning of the preposition in the word which represents as stop words, whereas it there are several of prepositions which have more than one meaning according its location in the sentence. The most information retrieval systems are facing many challenges when is applied in the Arabic Texts due to the large difference between Arabic languages with other language.

The preposition has a great importance in the Arabic language due to its effect at sentence meaning. To predict the suitable meaning of prepositions, there are also some rules must be followed to identify the correct meaning in the sentence. In this work, we proposed novel approaches for dealing with these prepositions.

Keywords: Proposition, Arabic propositions, Arabic language structure, Arabic stop words.

1. Introduction

More specialists' grammarians define that the preposition: What is indicated by the meaning of the other, being linked with any other speech. In fact, the characters in general, including the preposition indicates the meanings of both linked to other, or not linked, and the evidence that much of prepositions reported, including the intended meaning, whether they are starting "ابتداءً", or some of all "تبعيضاً, او بعض الشي", such as "من", or very like "حتى, الى" and there are also some structures in the Arabic language is proof that the meaning of the preposition benefit in itself [2].

Prepositions occur frequently in Arabic text. To aid information retrieval, they should be removed, so that variant forms of the same word are conflated to a single form. Separable prepositions can be easily detected and removed as stop words. However, inseparable prepositions are difficult to remove without inadvertently changing the meaning of other words in the text.The set of inseparable prepositions consists of the six letters: *lam, waw, faa, taa, baa, and kaf.* These letters differ in terms of their function in Arabic text and can be further divided into three different groups.

This paper is organized as follows. Section 2 briefly describes the related works in the area of Arabic texts. Section 3 describes Arabic prepositions, its types, and the meaning for each preposition. In section 4 shows the traditional techniques used in Arabic texts document. Section 5 presents the proposed method. In section 6 discuses the method, as well as the experiments carried out. Section 7 summarizes the work.

2. Related work

Study Arabic language and theirs complex morphology is more challenge because it contains a big amount of concepts; words, letters, and prepositions. Kardi, Nie in their study are showed and proposed a technique that contains sub other techniques to determine the stem of a word that represents the semantic core of this word according to Arabic morphology by using comparison steps with simple rules, after that the proposed method was more effective than the traditional methods[9]. While Al-Gharaibeh presented in his paper that tagged as "the searching on Quran text with the formal method", he mentioned the best way to create specification about Quran search system, and then applied to search about the concepts and word in Quran Texts [2].

There are many methods that used to search in Arabic text, in particular with the Holy Quran. A new approach has been presented a methodology to search about concepts, and keywords in the Holy Quran for improving the chances of finding all desired verses. The mentioned method included of applying three approaches; text-based, stem based, and synonyms-based system. These three

approaches were compared with respect to precision. The text-based approach was based on a full word, the synonyms-based approach was based on the synonyms of the word, and the stem-based approach was based on the stem of the word. Furthermore, A., Al-Taani A. and Gharaibeh in their study are mentioned the light stemming algorithm was applied to find the stem of the word [1].

According Mustafa study, he used the N-grams technique that common widely for a number of text processing and retrieval applications. In this article examined the performance of the diagram and trigram term conflation techniques in Arabic. The result of sufficient amount of Arabic text indicated that the diagram method offers a better performance than trigram with respect to conflation precision and conflation recall ratios. In either case, the N-gram approach didn't appear to provide an efficient conflation approach due to the peculiarities imposed by the Arabic infix structure that reduces the rate of correct N-gram matching [13].

Nurjannah in his study showed an application in learning, teaching of Arabic language, find out the similarities and the differences of Arabic Language, and Bahasa Indonesia prepositions and compare between them. The results revealed that the preposition in classical Arabic has various kinds and meanings depend on the certain spelling contexts of the preposition. It is concluded that the specification of the proposition has many meanings or even just one meaning. This article recommended that Indonesian students would find this article as solution in learning Arabic language and as a way in translating Arabic language to Bahasa Indonesia [10].

While Sa'eed mentioned a description about the problem that occurs when translates verbs-prepositions collection from Arabic into English language [12]. Indicates in both Arabic and English have many verbs to collocate with certain prepositions. Where's there are many English verbs do not collocate with proposition. In Yaary study the problem that appeared when we try to translate the texts from English to Arabic language explained, whereas he indicated the accurate mappings between English-Arabic prepositions are sometimes very difficult to determine by Arab learners. This study was designed to identify the

difficulties of translating the English prepositions at, in and on, which Saudi EFL students may face when translating them into Arabic. The purpose was to determine which type of the two sexes (males or females) can translate these three prepositions better [14].

3. Arabic Prepositions

In Arabic language the propositions are read so " حروف الجر Huruwf-ul-jarr", and it contains two parts, in the first one "Huruwf" its represent plural of "Hurf". In Arabic language, also the word "حرفHurf" refers about any word has meaning but it not noun, or verb. In the second word "الجرUl-Jarr" it's refer about the action of character, in other word its means to drag a work of the verb to after that character.

- o In Arabic language, prepositions have been added before nouns. There are a lot of prepositions, some of these are usually attached to the beginning of a word. These preposition known as inseparable prepositions which can't be separate such as لـ (lam) و (waw), كـ (kaf), ب (baa), and ت (taa).

- o Other group of preposition which can be used before noun only such as waw ,kaf, taa, and baa. But the preposition "ت"taa is also used as a prefix for verbs in the present simple tense.

- o In second hand the preposition "اللام"lam, that means its purpose as a preposition like"الكاف" kaf and"الباء" baa, can also be used with verbs as the " لام الامر lam of command". wheares, it is usually prefixed to the third person to give it an imperative sense, for example (لتفعلها) "do it". It is also used to indicate the purpose for which an action is performed.

3.1 Types of Preposition in Arabic language

Prepositions: divided into three types:

1- **Original prepositions**: These are characters that can't be dispensed with in a speech such as " من, إالى, عن, في, الباء, الكاف, اللام, منذ, حتى " and as well as from the original prepositions is called in Arabic "أدوات القسم" **oath or swear by Allah tools** such as أدوات "والله, تالله, بالله", and **the exception tools** "الاستثناء" such as "عدا, خلا ,حاشا".

2- **Prepositions excess**: that can be dropped in speech, some of these prepositions call to emphasize characters, and some of them call the relevant characters such as " الباء, اللام, الكاف"

3- **Prepositions semi excess**: it uses for indication something such as reducing or increasing something such as "**Rubba", " رُبَّ**" these word can be used to reduce.

3.2 The potential meaning for each preposition

Regardless of nouns (things) and verbs (actions), English and Arabic also contain prepositions– words like "in", "at", "from", and so on. In Arabic langauge; the prepositions are known as Huruwf-ul-Jarr; (حُرُوف الجَرّ singular is Harf-ul-Jarr).In grammars of Arabic langauge, the word that follows the preposition immediately becomes majruwr "مجرور!" For example, you can say اللعبة : (al-lubaa 'ala al-tawelah)–"The game is عَلَى الطاولةِ on the table." in this example, we see the preposition 'ala (on), and tawelah (table), the word that follows 'ala, becomes majruwr. in other example, you can say (ahmad fiy احمد في بَيتِكم baytikum)–"Ahmad is in your house." The preposition, here is fiy (in), and your house (baytikum) is majruwr. The table below lists all the prepositions "huruwf-ul-jarr" and their meanings of each one, the meaning really depends on the context, or location in their statements. In other words, the meaning of preposition may be changes depending on where and how the preposition " harf" was used.

Arabic	Translation	Meaning in English
في	Fiy	In
إلَى	Ila	to (eg. travel to)
عَلَى	'ala	on (top of)
مِن	Min	From
لِ	Li	for (eg. for you)
مَعَ	ma'a	With
بِ	Bi	With
حَتَّى	Hatta	Until
مذ,مُنذُ	Mundhu,Mudh	Since
عَن	Aan	About
كَ	Ka	like/as
ت, واو القسم	Taa,Wow	…..
رُبَّ	Rubba	Few, or More

Table 1: The Table of common prepositions

4. Traditional techniques used in Arabic Text

Many stemmers have been developed for Arabic [5]. However, none deals with the removal of all particles. Some particles, such as waw, are removed by all existing stemmers; other particles, such as kaf, have never been considered on their own in existing stemming approaches.

- **Light stemming**

In Light stemming is statistically motivated. It is similar to the commonly used light stemmers. It truncates a word at the two ends. The decision to truncate or not a segment of a word is made according to some rules and statistics on the corpus [9]. Light stemming refers to a process of stripping off a small set of prefixes and/or suffixes, without trying to deal with infixes, or recognize patterns and find roots.

- **Text-based system specification**

In this approach "text-based system" represented a search method that based on the full word, each word in the user query is matched to the same word in the index table then the system retrieves the sentences that have the same word. The specification of the text-based system is expressed by using Z schema.

5. The proposed method

The new approach is proposed. It contains some of steps, in the first step; it read the Arabic text document, after that search about the prepositions to extract. The third step includes divide the statement that contains the preposition into parts. The first part known as a pre-part that comes before the preposition, the second part is the part that contains the preposition, and then the last part is the complement that comes after the preposition, if the statement contained more than one preposition, we fragment the statement into sub statement, and then we applied the previous steps for each sub statement. The important phase in our approach is being in last part or complement part of statement, in this phase we check this complement that comes after the "Hurfeljar" or prepositions to translate or exegetics the meaning of this part in Arabic language to obtain the best understanding of it, and therefore we try to find relationship between the complement and all parts of statement and taking into consideration the time and place and the reason for the event, and thus locate the nearest meaning of the prepositions with sentence. The mechanism of the approach has explained at the pseudo code that mentioned as follows.

Begin
read Arabic text document;
extract the prepositions;
determine all preposition
location;

divide the document into parts;
check the relation between the preposition with other parts
tokenization
take a preposition
search for the proposition in dictionary
if found then
return the relative meaning
else
check the relation between the preposition with other complement part again
En

This method allows the machine to expect about the best meaning of prepositions in the Arabic texts. Figure 1 shows the architecture of this approach. Firstly; the Arabic text is analyzed its preposition are represented in a database dictionary file. Secondly; extract the prepositions from Arabic texts. Finally; the relation between the sentence parts has been analyzed to expect the relative meaning that closed with the meaning of the sentence.
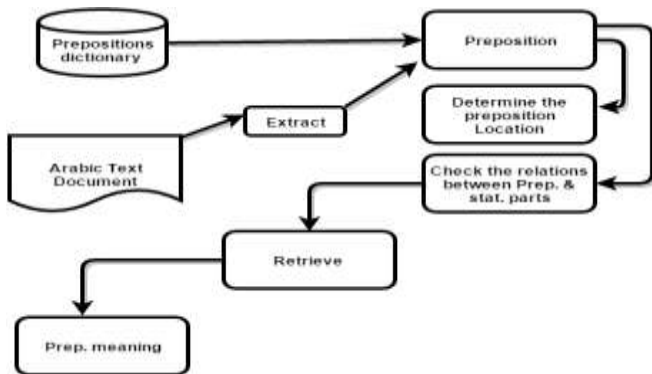

Figure 1 : Propsed Method Archtichure

6. **Discussion**

We will try to apply our proposed method in some example to evaluate the performance of this method through using some sentences from Arabic texts: In the first example "*kharaj Mohammad m'n albiat mobakiran* خرج محمد من البيت باكرا", or Mohammad left the house early, the proposed method is doing as the following:

Remove other stop word if found	Pre-Part	The specific part	The complement part
Has no other stop words	خرج محمد	من	البيت باكرا
It means, the statement doesn't contain any word or letter connect with prepositions	It means Mohammad came out…..	It has more than one meaning such as: from, enter in certain place, enter into certain time…etc	It means Mohammad went from home at time early.
The best meaning of the preposition"من" in this statement its" came out or out of place", in Arabic "الغاية المكانية"that means of the preposition related with place not related with time or other.			

.

In other example from the Holy Quran:

يوم ندعو كل اناس بايمامهم فمن اوتى كتابه بيمينه فاولئك "
"يقرءون كتابهم ولا يظلمون فتيلا

This statement" verse" when we apply our proposed method will be dividing into two sub statement. The 1st statement it's the same original statement, and the other 2nd statement is the part came after 1st preposition, or the complement of the 1st statement. The proposed method is done as follows:

The 1st statement: " **يوم ندعو كل اناس بايمامهم فمن اوتى** or **كتابه بيمينه فاولئك يقرءون كتابهم ولا يظلمون فتيلا** *"yawm nadou kol unaas bimanihm f-m'n autia'a kitabaho bi-yameenaho; f-aula'ika y'qrawn kitabhom wala yudhlamoon fatila.*

the 2nd statement which represents a part of the original statement: " **فمن اوتى كتابه بيمينه فاولئك يقرءون** *" or* **كتابهم ولا يظلمون فتيلا** *f-m'n autia'a kitabaho bi-yameenaho; f-aula'ika y'qrawn kitabhom wala yudhamoon fatila.*

The possess will applied as follows:

	Remove other stop word if found	Pre-Part	The specific part	The complement part
1st Stat		يوم ندعو كل اناس	حرف الباء المتصل بكلمة "ايمانهم"	ايمامهم فمن اوتى كتابه بيمينه فاولئك يقرءون كتابهم و لا يظلمون فتيلا
	It has no other stop words	The versus means Allah says We call all of people	Letter "Baa" it connects with word"إيمانهم", whereas it means: reason, by,…etc	It's a complement that also, contains the other preposition
	The best meaning of preposition at the first statement "الباء"indicate for the way or method that used when the Allah calls the all people, then the complement became as the second statement, because it contains other preposition "من"			
2nd Stat	Letter faa"الفاء"		Letter M'n" "مَن	اوتى كتابه بيمينه فاولئك يقرءون كتابهم و لا يظلمون فتيلا
	It may represent the excess letter connect with preposition for some aims	It has no pre-parts	The preposition "مَن"it represents as relative noun "الذي"	It means. The person who gives his/her book in your hand, that means he/she is wan at "Jannah", or paradise.
	So, by using the proposed method, we'll expect the meanings of prepositions of all the verses.			

7. Conclusion

In this work, we have been presented new suggested method for the extraction the meaning of prepositions in Arabic text using a well-known collection of Arabic documents. It was the first study in the expected meaning of preposition in Arabic text.

References

1. A., A.-T. A. (2011.). Searching Concepts and Keywords in the Holy Quran. *The International Arab Conference on Information Technology (ACIT).*
2. Abdusalam F.A. Nwesri, S. T. (2005.). Stemming Arabic Conjunctions and Prepositions. *Springer.*
3. Al-Gharaibeh, A. A.-T. (2011). The Usage of Formal Methods in Quran Search System.. *Proceedings of International Conference on Information and Communication Systems.*, pp. 22-24.
4. Aljlayl, M. &. (2002). On Arabic search: improving the retrieval effectiveness via a light stemming approach. *In Proceedings of the eleventh international conference on Information and knowledge management (pp. 340-347). ACM.*.
5. Al-Kharashi, I. A.-S. (2004.). Arabic morphological analysis techniques:A comprehensive survey. *Journal of the American Society for Information Scienceand Technology.*, 55(3):189–213.
6. Chew, P. A. (2008.). Latent morpho-semantic analysis: multilingual information retrieval with character n-grams and mutual information. *In Proceedings of the 22nd International Conference on Computational Linguistics-*.
7. El-Beltagy, S. R. (2011.). An accuracy-enhanced light stemmer for arabic text. *ACM Transactions on Speech and Language Processing (TSLP), 7(2), 2.*, .
8. Helou, M. A. (n.d.). Towards Building Lexical Ontology via Cross-Language Matching. .
9. Kadri. Y., J.-Y. N. (2006). Effective Steming for Arabic Information Retrieval. *Conference:THE CHALLENGE OF ARABIC FOR NLP/MT.*, 68.
10. Nurjannah, R. A. (2014.). PREPOSITION IN ARABIC AND BAHASA INDONESIA:A COMPARATIVE STUDY. *Journal of Arts, Science & Commerce.*, E-ISSN2229-4686 - ISSN2231-4172.
11. Rafe, V. &. (2014). An Efficient Indexing Approach to Find Quranic Symbols in Large Texts. *Indian Journal of Science and Technology, ISSN 0974-5645*, Vol 7(10), 1643–1649.
12. Sa'eed, M. A.-A.-J. (2011.). The Problems of Translating Verb-Preposition-Collocation from Arabic into English(A Didactic Study). *ADAB AL-RAFIDAYN,VOL.(65).*
13. SH Mustafa, Q. A.-R. (2004.). Using N-Grams for Arabic Text Searching. *Journal of the American Society.*
14. Yaari, S. A. (2013.). The Problem of Translating the Prepositions at, in and on into Arabic: An Applied Linguistic Approach. *Journal for the Study of English Linguistics.*, ISSN 2329-7034, Vol. 1, No. 2.