A RULE-BASED APPROACH FOR PRONOUN EXTRACTION AND PRONOUN MAPPING IN PRONOMINAL ANAPHORA RESOLUTION OF QURAN ENGLISH TRANSLATIONS

Arash A. Tabrizi[1], Rohana Mahmud[2], Norisma Idris[3], and Hossein Tohidi[4]

[1, 2, 3] Department of Artificial Intelligence,
Faculty of Computer Science and Information Technology
University of Malaya, Kuala Lumpur, Malaysia

[4]PyCademy (KHCC SDN. BHD.), Kuala Lumpur, Malaysia

Email: aminitabrizi@siswa.um.edu.my[1], rohanamahmud@um.edu.my[2], norisma@um.edu.my[3], hossein@pycademy.net[4]

ABSTRACT

The nature of the Quran and its translations as classic Arabic and English texts reduces the accuracy of ordinary natural language processing tools such as pronominal anaphora resolution systems. Pronominal anaphora resolution simply involves finding an antecedent for anaphoric pronouns as the referring expressions of discourse. The performance of a pronominal anaphora resolution system is vitally related to the efficiency of pre-processing tools that analyze and prepare the input data for feeding the resolution algorithm. This paper proposes a novel pre-processing approach for pronoun extraction and pronoun mapping in the pronominal anaphora resolution system of English translations of the Quran, which facilitates the anaphora resolution, specifically for the English pronouns without an explicit antecedent that contributes close to 50% of the anaphoric relations in the Quran. This approach uses the morphologic, statistic and anaphoric knowledge that is extracted from the Arabic corpus of the Quran. For evaluating the arrangement, 1% of an English translation was annotated with labeling for all anaphoric and non-anaphoric English pronouns. These pronouns were aligned to the equivalent Arabic pronouns and linked to the concepts in the Arabic text. Through statistical results, it was shown that our rule-based pre-processing tools perform well. The precision, recall, and accuracy of pronoun extraction stage are 96.38%, 100%, and 99.5%, respectively. The result of mapping algorithm is promising whereby we score 85.51% in precision, 96.32% in recall, and 82.81% in accuracy.

Keywords: Pronominal anaphora resolution, Anaphora resolution pre-processing, Pronoun resolution, Word alignment, Quran English translation, Rule-based approach, Natural language processing

1.0 INTRODUCTION

Pronominal anaphora resolution, which appears as pronoun resolution, simply involves finding an antecedent for anaphoric pronouns as the referring expressions of discourse. According to Mitkov [1, 2], anaphora resolution (AR) is known as a difficult problem in NLP. AR has attracted the attention of scholars over the last two decades, who have carried out a great deal of successful works [3-11]. Developing methods for AR is useful in applications, such as information extraction, text summarization, text segmentation, coherence studies, multimodal interfaces, machine translation, and question answering systems [12-16]. In addition, in the study of the theories of translation equivalency, pronominal AR is a reliable tool for investigating referential, contextual, and thematic cohesion (e.g. Bakers [17]).

Anaphora resolution systems rely on syntactic, semantic or pragmatic clues to identify the antecedent of different types of referring expression, e.g. definite noun phrases, pronouns, one-anaphora, and demonstratives [18]. In comparison with the other types of anaphora, pronominal is the most widespread type, where the referent is referred by an anaphoric pronoun (e.g. 'He' refers to 'Alex' in: *Alex* has bought a piano. *He* likes music).

Pronominal anaphora can be divided into four subclasses: nominative (e.g. he, she, it, and they), reflexive (e.g. himself, herself, itself, and themselves), possessive (e.g. his, her, its, and their), and objective (him, her, it, and them). However, the usage of "it" may cause ambiguous situations, because of its usage as a non-anaphoric expletive or pleonastic 'it' that has no antecedent in previously seen discourse (e.g. 'it' in 'it is raining') [7]. An anaphora resolution system should recognize the occurrences of this non-anaphoric pronoun. Barbu and Mitkov

[19] showed that their corpus of 28,272 words with 422 pronouns, contains 60 non-anaphoric pronouns (i.e. 14.2%).

The performance of a pronominal anaphora resolution system is vitally related to the efficiency of pre-processing tools that analyze and prepare the input data for feeding the resolution algorithm. The pre-processing stages that can differ for various AR systems include such hard pre-processing sub-systems as morphological analyzer / POS-tagger, named entity recognizer, unknown word recognizer, noun phrase extractor, text parser, and pleonastic pronoun identifier [20]. Improvement of these subsystems can increase the success rate of the anaphora resolution system. One of the most important pre-processing tasks is the pronominal extractor that is a look-up module to identify the pronominal pronouns in a text. In addition, utilizing pronoun mapper boosts the performance of the parallel text AR systems. The pronoun mapper aligns every pronoun to its equivalent pronoun and links it to a referent concept in another language. Mapping between pronouns causes to obtain more information about the detected pronoun to disambiguate potential referents. This study focuses on development of specialized pronominal extractor and pronoun mapper.

Anaphora resolution is a well-known topic in Arabic natural language processing [21, 22]. Arabic Quran script has been annotated with anaphoric resolution in recent studies [22, 23]. The pronominal AR of the Quran's English translations is extremely new field of study which addresses unique features of Quran translation as using classic vocabulary [24]. This research introduces a novel rule-based and corpus-based approach for the pre-processing stages which remarkably increases the AR accuracy in the Quran target language (TL) texts.

The rest of the paper is organized as follows: Section 2 describes the background study and importance of AR in the Quran and its translations, followed by the description of the corpus in Section 3. Section 4 describes the novel rule-based algorithms for pre-processing stages to overcome the challenges of pronominal AR of Quran English translations. Section 5 shows the experimental results and comparison. Finally, Section 6 concludes this paper with a brief summary and future work.

2.0 BACKGROUND STUDIES

Pronouns have an important role in the semantic and structural connection of discourse segments. Finding referents and antecedents of pronouns helps to reveal the structure of discourse. Increasing the number of pronouns in a text means that more referential relations should be resolved during the study of discourse structure.

Table 1 and Fig. 1 illustrate the word frequency of the Wall Street Journal portion of Penn Treebank corpus (PTB III) [25, 26], Brown [27], and Quran [23] Corpora. The results show that pronouns have a vital contribution in the Quran's Arabic text. While the average word frequency in the WSJ and Brown are respectively four times and three times higher than the Quran, the rate of pronoun words in the Quran is considerably higher than these corpora (i.e. 23%, in comparison with 3% and 6%, respectively). English translations of the Quran that follow the original text, have the same situation in their word frequencies. Quran translators usually concentrate on the similarity of the TL with the source language (SL) texts in meaning and structure (e.g. grammatical units, word order, and segmentation of text) [28]. Section 4 shows that, on average, the pronominal pronoun group contributes close to 15% of the words in Quran translations. This means that the structure of Arabic text affects the degree of pronoun usage in TLs. In comparison with WSJ and Brown, it is remarkably higher, and the effect of this pronoun frequency cannot be neglected.

Table 1. Comparison of WRJ, Brown and Quran Arabic corpora.

Corpus	Word count	Word type	Pronoun count
WSJ of PTB	1253013	44377	34704
Brown	1161192	56057	68968
Quran	127795	19287	29494

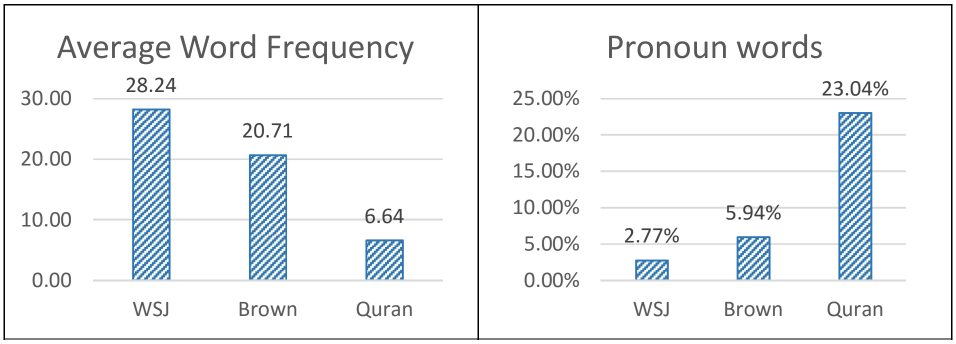

Fig. 1. The average word frequency and the pronoun word contribution of WSJ, Brown, and Quran.

However, from the natural language processing viewpoint, the AR of Quran translations is more complicated than a general AR. The first issue arises from the various styles used in the Quran text (i.e. historical narrative, didactic, argumentative, literary, and persuasive). Style shifts between diverse chapters, and also inside chapters (e.g. chapter 2), make it difficult to use a unique AR method for the whole of each translation. In addition, the characteristics of the Quran script, such as different length of the chapters (3 verses in chapter 108 to 286 verses in chapter 2), different length of the verses (varies from one word to more than 200), and the diverse number of sentences in a verse (varies from a clause to several sentences in a verse) cause instabilities in the automated AR of translations. Finally, the frequent usage of the ellipsis, redundant, and extra-positive constructions in the Arabic text, produces unusual structures in TL texts.

Although linguistic study of the Quran has a strong background over the last fifty years [29, 30, 31, 32], it seems that the human annotation of Sharaf and Atwell (i.e. QurAna) is the first complete pronominal AR of Quran [23]. In the QurAna research, they used the knowledge of previous scholars on the Quran for annotation of the Arabic text. They described that of 24679 annotated pronouns in the Quran, only 0.3% (i.e. 90 instances) are cataphoric. In addition, 46.7% of pronouns have no explicit antecedent in the text, which shows the importance of human knowledge, extracted from Islamic resources for pronoun resolution of the Quran. For modeling both explicit and implicit referents, they produced an ontology for concepts out of pronoun antecedents with more than 1000 entries. This ontology has an Arabic and English clause for every concept. Since the extraction of these concepts and finding their anaphoric role use the knowledge of interpretation of the Quran and the history of Islam, this study uses the QurAna as a benchmark for AR of Quran translations.

Quran Arabic Corpus (QAC) project is another study on Quran script, which produced a corpus with morphological, part-of-speech, and syntactic annotation of the Quran [33]. In addition, this corpus, with its included Quranic Arabic Dependency Treebank (QADT), shows the implicit hidden pronouns, and linked the verses to a word by word English translation [34].

In a recent research, Seddik, Farghaly, and Fahmy, reported an annotation of Quran with anaphoric information [22]. They annotated about 24,653 personal pronouns with their antecedent and anaphoric information as anaphora-antecedent distance, and pronoun features as gender, number, and person [22]. They used the QurAna and QAC as their input corpora, and tried to solve the mismatches and mistakes of their human annotations.

The linguistic study of Quran translations is an interesting topic for scholars [28, 29, 35, 36]. However, the computational study of anaphora and pronoun resolution in the TL texts of the Quran is a new challenging area. In [24], Tabrizi and Rahman describe some challenges in the automated pronoun resolution of translations, but they do not implement a practical solution for the challenges of PR or AR in English texts. This study enhances pronominal AR in the Quran TL texts using two specialized pre-processing tools: pronoun mapper and extractor. Our pre-processing tools are designed based on corpus-based knowledge (extracted from both SL and TL corpora) and rule-based approaches.

Since there is no published study on English-Arabic Quran pronoun mapping, our approach is compared with a state-of-the-art English-Czech pronoun alignment method. Novak and Zabokrtsky [37] present a supervised pronoun aligner that is a discriminative log-linear model which trained in a one-against-all and cost-sensitive strategy. 471 occurrences of personal pronouns is aligned, which account for over 50% of all occurrences. Then, they apply a simple heuristic to find the corresponding English personal pronouns. Their aligner surpasses a GIZA++ baseline in terms of both intrinsic and extrinsic evaluation.

3.0 QURAN CORPUS

The Holy Quran is a classic Arabic corpus that contains 127795 word segments, 77784 words, and 19287 word types [23, 24]. The book has been structured in 6236 verses with obvious boundaries. The verses are categorized into 114 chapters. The length of the verses, length of chapters, the number of sentences in every verse and the style of the text change throughout the Quran text. Translation of Quran to other languages has created a multilingual corpora in about 100 different languages. As for English, native Arabic-speakers and non-native translators have published more than 80 translations over the last nine centuries. Almost all of the translators have followed the structure of the Arabic text and have translated every chapter and verse in their invariant positions, thus a comparison of translations can be done by comparing the translations of each verse.

Table 2 shows the specification of the dataset [38-48]. The dataset covers almost all the non-literal Quran translations that have been converted to TXT and XML formats. Each new translation can be added to the dataset. All of the translations were published after 1900.

Table 2. Corpora of Quran English translations.

Translator	Name	Number of words (Approximately)	Word terms
Muhammad Ali [38]	The Holy Qur'an	151852	5307
M. Pickthall [39]	The Meaning of the Glorious Qur'an	155257	6165
A. Arberry [40]	The Koran Interpreted: A Translation	146173	5438
M. H. Shakir [41]	Holy Qur'an	160563	5317
M. T. Sarwar [42]	The Holy Qur'an	154003	5575
A. Daryabadi [43]	Holy Qur'an	155384	6764
Ahmad Ali [44]	The Qur'an: A Contemporary Translation	149417	6809
A. Nooruddin [45]	The Holy Qur'an, Arabic Text - English Translation	209682	8902
Ali Unal [46]	The Qur'an with Annotated Interpretation in Modern English	220560	7949
Faridul Haque [47]	The Holy Quran, from Kanzul Iman	169270	6582
Hamid S. Aziz [48]	The Meaning of The Holy Quran, with Explanatory Note	161956	6718

In addition to English translations, two Arabic human annotated corpora are used in this study. The Quranic Arabic Corpus (QAC) project [33, 34], which has morphological, part-of-speech, and syntactic annotation of the Quran, and QurAna [23], which is an Arabic anaphoric resolution with a parallel text ontological concept reference in the Quran. Dukes and Atwell [33], in the morphological layer of QAC, annotated 24685 pronouns in the Quran. Sharaf and Atwell [23] described that the Quran contains 29494 pronouns, including 24679 anaphoric pronouns (i.e. the remainder, over 15%, is demonstrative or relative pronouns).

Evaluation of the algorithms requires a corpus as benchmark data to analyze the precision, recall, F-measure, and accuracy of the results. For this purpose, the 30th chapter of the Pickthall translation [39] was selected for human annotation. Table 3 shows the specification of this chapter, which contains 1% of the Quran text. In the annotation of the pronouns, the data of the SL Quran text, extracted from QurAna and QAC, were considered as accurate data, and the annotators only extract the English pronouns and map them to the Arabic equivalents. From 221 annotated English pronouns, 213 items have an equivalent in the Arabic text, and eight reminder pronouns are cataphoric or pleonastic. The annotated data (with 213 anaphoric pronouns in 1610 words) are comparable with the Mitkov dataset with 362 annotated anaphoric pronouns in 28272 words [19].

Table 3. Statistics of human annotated chapter in Pickthall translation.

Pickthall Translation		Arabic Text	
Number of verses	60	Number of verses	60
Length of verses	3 to 65	Length of verses	3 to 62
Word count	1610	Word count	1341
Number of pronouns	221	Number of pronouns	241

4.0 PRE-PROCESSING ALGORITHMS FOR PRONOMINAL AR

This section describes the novel algorithms for pre-processing stages of pronominal anaphora resolution of Quran translations. These algorithms have been customized for the English language, however, with some minor changes, they can be used in any TL. The structure of the algorithms is based on the nature of the Quran discourse and the method of translation. The algorithms use the advantages of syntactic, semantic and structural similarities between translations and the Arabic text:

- Every chapter is translated as a separate text, so there is no anaphoric relation between chapters.
- Every verse is translated as a separate text segment, so there is a clear boundary for finding candidate antecedents of every pronoun. If an English pronoun is equivalent to an Arabic pronoun in SL text, the antecedents will be in the same SL and TL verses.

Fig. 2 illustrates the structure of a cross-lingual pronominal anaphora resolution system. The pronominal extractor and pronoun mapper are the pre-processing stages for this type of the pronominal AR. The pronominal extractor identifies the pronominal pronouns of the English text. The second component, the pronoun mapper, aligns every English pronoun to its equivalent in the Arabic text, and identifies the reference concept of the pronoun. The main system is the pronominal anaphora resolution engine that can be assumed as a pronoun resolver. The AR system produces a candidate antecedent set for every English pronoun, and with a pronoun resolution engine, emphasizes the antecedent with the highest preference. Pre-processing components produce the features that are used in the pronominal AR engine. This section describes our design of the first and the second components.

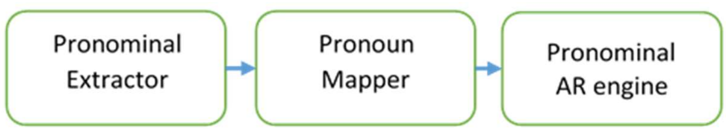

Fig. 2. The pre-processing components and the engine of Quran pronominal AR.

And We said: O Adam! Dwell thou and thy wife in the Garden, and eat ye freely Pickthall [39], chapter 2, Verse 35
('And', 'CC'), **('We', 'PRP')**, ('said', 'VBD'), ('O', 'NNP'), ('Adam', 'NNP'), ('Dwell', 'NNP'), **('thou', 'NN')**, ('and', 'CC'), **('thy', 'JJ')**, ('wife', 'NN'), ('in', 'IN'), ('the', 'DT'), ('Garden', 'NNP'), ('and', 'CC'), ('eat', 'VB'), **('ye', 'NN')**, ('freely', 'RB') Results of default POS-tagger of Natural Language Toolkit [49]
('And', 'CC'), **('We', 'PPSS')**, ('said', 'VBD'), ('O', 'UH'), ('Adam', 'NP'), ('Dwell', ''), **('thou', ' PPSS ')**, ('and', 'CC'), **('thy', 'PP$')**, ('wife', 'NN'), ('in', 'IN'), ('the', 'AT'), ('Garden', ''), ('and', 'CC'), ('eat', 'VB'), **('ye', '')**, ('freely', 'RB') Unigram POS-tagger with religion corpus using Natural Language Toolkit [49]

Fig. 3. Part of speech tagging for verse 35, chapter 2 of Pickthall translation.

In a pronominal AR system, the pronoun extractor component should detect all the pronoun words. The extractors usually use a part-of-speech tagger for pronoun annotation. However, because of the word domain of translations, using a general part-of-speech tagger for pronoun detection has low accuracy. Fig. 3 demonstrates the results of the default and the customized POS-taggers for verse 2-35 in the Pickthall translation. Using the default tagger [49], only the word ‘we’ has been extracted as a pronoun, and the words ‘thou’, ‘thy’, and ‘ye’, which are pronominal, have not been detected. The unigram tagger, which was trained using texts in the religion category of the Brown corpus [27], has better results, and only ‘ye’ has not been tagged correctly. This challenge can happen in every other TL language, thus instead of using machine learning methods to customize a more accurate parser, the tagger was replaced with a word extractor.

4.1 Pronominal Extraction

A word extractor requires a set of pronouns that can be obtained using an annotated corpus of the TL language. Since the dataset contains both new and old translations, it seems inevitable to use a pronoun set that covers pronouns in modern and classic English. To generate the set, the annotated words from the Brown portion of the Penn Treebank corpus (PTB III) [25] was used. The Brown corpus includes the modern and classic texts, and covers diverse genres [27]. First, a list of PRP and PRP$ tagged words (i.e. personal, reflexive, nominal possessive, and adjectival possessive) was collected. Then, the extracted pronoun list was compared with standard, non-standard, and archaic pronoun families in English grammar, and the words that are very un-formal were discarded (e.g. ‘ah’ which has been annotated in Brown as a pronoun). As the final step, the pronoun set was checked with a statistical search on the Quran. The results showed that from the non-standard pronoun family (e.g. ‘youse’, ‘y'all’, and ‘y'alls selves’), only ‘you all’ is used as a pronoun in translations. From the reflexive pronoun family the frequency of ‘hisself’, ‘themself’, ‘theirself’ and ‘theirselves’ is zero, so they were discarded from the list. Table 4 shows the final set of the selected pronouns for the extraction algorithm.

Table 4. Pronoun set for extraction component.

Person	Number	Pronouns	
First Person	Singular	I, me, myself, mine, my	
	Plural	we, us, ourselves, ourself, ours, our	
Second Person	Singular	you, yourself, yours, your, thou, thee, thyself, thine, thy, thine	
	Plural	you, yourselves, yours, your, ye, you all, youselves	
Third Person	Singular	Masculine	he, him, himself, his
		Feminine	she, her, herself, hers
		Neuter	it, itself, its
	Plural	they, them, themselves, theirs, their	

Table 5 shows the results of pronoun extraction for eleven translations. The advantage of this extraction method is its independence from the translation structure or style, because it is in the morphological layer. The results demonstrate that the total count of pronouns in the translations is higher than such corpora as PTB and Brown, and are close to the Quran Arabic results (with ±2000 difference in average).

Table 5. Pronoun extraction of eleven Quran translations.

	Pickthall	Nooruddin	Ahmad Ali	Shakir	Daryabadi	Muhammad Ali	Hamid Aziz	Sarwar	Arberry	Ali Ünal	Faridul Haque
I	662	791	646	713	535	662	692	625	666	726	582
my	487	553	409	486	464	485	477	393	478	573	448
me	428	429	389	447	435	430	421	364	425	490	365
mine	8	6	8	3	22	3	11	5	5	5	21
myself	8	22	9	19	15	16	13	14	12	22	18
we	1863	2111	1669	1920	1998	1881	1904	1686	1863	2025	1712

our	478	610	412	460	447	453	465	542	442	602	465
ours	8	13	5	2	7	3	7	4	5	6	8
us	449	488	422	449	462	457	465	424	457	530	423
ourselves	6	20	7	21	11	15	15	18	19	16	17
ourself	0	6	2	0	0	0	3	0	0	0	0
you	1509	5048	4180	4469	1487	3235	4584	4094	3336	5338	4230
you all	12	43	11	11	18	12	14	22	14	27	33
yourself	0	32	20	20	2	0	10	14	0	28	29
yours	19	33	16	17	13	13	12	14	11	24	24
your	773	1567	1163	1163	647	764	1205	1308	672	1599	1264
thou	758	0	0	125	1347	708	82	0	819	0	0
thee	634	0	0	45	611	615	53	0	627	0	0
thyself	11	0	0	2	11	20	9	0	17	0	0
thine	25	0	0	1	72	16	1	0	8	0	0
thy	440	0	1	32	386	438	33	0	456	0	0
yourselves	62	89	51	74	75	71	71	76	71	94	50
ye	1861	0	0	0	1851	7	13	0	1	0	0
he	2259	2536	1846	2453	2280	2385	2345	1503	2249	2416	1814
him	1322	1356	958	1236	1180	1210	1251	821	1180	1612	1391
himself	56	111	49	95	89	84	77	62	55	107	58
his	998	1406	959	947	964	930	1000	1087	943	1607	1010
she	80	87	66	86	81	80	80	53	82	67	66
her	101	113	87	99	143	93	106	78	94	114	89
herself	3	4	3	4	3	3	3	2	2	8	2
hers	2	0	0	0	3	1	0	0	0	0	0
it	1546	2275	1237	1922	1310	1774	1710	1178	1597	2243	1855
itself	7	18	8	9	10	9	11	7	5	26	12
its	121	236	149	169	75	177	160	143	167	296	201
they	2882	3203	2644	3209	3017	3019	2936	2495	2825	3431	2730
them	2322	2529	2039	2293	2221	2296	2264	2065	2353	2905	2269
themselves	100	208	139	121	176	116	128	176	140	252	123
theirs	87	35	23	15	81	31	20	5	37	24	5
their	1252	1924	1144	1217	1137	1231	1282	1573	1189	1924	1379
TOTAL	23639	27938	20771	24354	23686	23743	23934	20851	23322	29137	22693

Fig. 4 illustrates the gender, person, and number distribution of the English and Arabic pronouns. The English results are the average of eleven translations. For the person feature, the results of the English and Arabic are similar. However, the weakness of the English language to show the gender and number causes a high dissimilarity in the results. Nevertheless, the low frequency of feminine (in gender) and dual (in number) decrease the vagueness of the features.

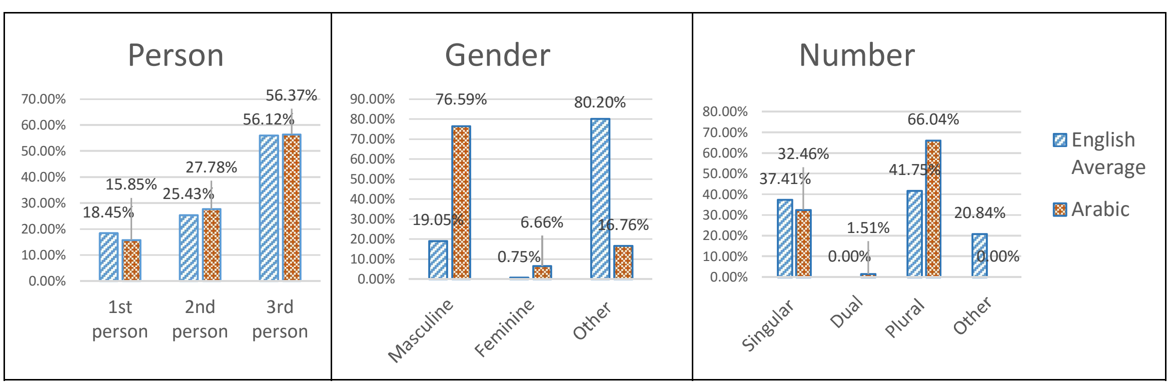

Fig. 4. Pronoun distribution in English and Arabic corpora.

The pronominal extractor component constructs an object for every verse in every translation with a vector of extracted English pronouns. For every pronoun, the detected features, such as position in the verse, number, gender, and person, have been added to the object. Table 6 shows the output of the component for verse 40 of chapter 30 in Pickthall translation [39].

Table 6. Example of output of pronominal extractor component.

Verse text analysis	Object structure of verse	Features	Abbreviations
Allah is [He]<3MS> Who created [you]<2O> and then sustained [you]<2O>, then causeth [you]<2O> to die, then giveth life to [you]<2O> again. Is there any of [your]<2O> (so-called) partners (of Allah) that doeth aught of that? Praised and Exalted be [He]<3MS> above what [they]<3P> associate (with [Him]<3MS>)! Pickthall [39], chapter 30, verse 40	verse_object: T= 'Pickthall' Chi = 30 Vi = 40 EngProVec = ([2,he,3MS],[5,you,2O],[9,you,2O], [12,you,2O],[19,you,2O],[25,your,2O], [39,he,3MS],[42,they,3P],[45,him,3MS])	Person	1: First 2: Second 3: Third
		Number	S: Single P: Plural O: None/other
		Gender	M: Masculine F: Feminine O: None/other

4.2 Pronominal Mapping

This component emphasizes that whether every English pronoun in the TL text is the equivalent of a particular Arabic pronoun in the Quran. While the smallest segments of the text are verses, the algorithm concentrates on the mapping of English pronouns in every verse (word alignment for pronouns). The mapping algorithm contains four steps:

1. Extracting the Arabic pronouns and their features as position (chapter, verse, and word), pronoun ID number, and the information of the antecedent (if it exists) from the QurAna corpus. In addition, the referent concept information of every pronoun is extracted from their concept ontology. There were cases of mistakes in human annotations [22] that were handled before utilizing the data.

2. Extracting essential features of every Arabic pronoun (i.e. number, gender, and person features) from the morphology layer of the Quran annotation [47]. Extracted data show mismatches between the morphology results and QurAna for 58 pronouns in 50 verses (i.e. in 28 verses QAC has annotated more pronouns than QurAna, while in 22 verses QurAna has more extracted pronouns). Although [22] illustrates a method to match the data of QurAna and QAC, in this study, these verses are discarded from the analysis. They contain 248 to 363 pronouns in different translations (i.e. less than 1.3%).

3. Finding a candidate equivalent set of Arabic pronouns, as the equivalent of every English pronoun in the verse. These sets include every possible equivalence between the English and Arabic pronouns. The rules of this component are similar to the gender-number agreement constraints of AR systems [1], based on the person, number, and gender features. Table 7 illustrates the candidate equivalence rules.

4. Aligning the English pronouns using a rule-based method. These rules are very simple and only use three features (i.e. number, person, and gender) for mapping.

Steps 3 and 4 are repeated for every translation.

Table 7 and Fig. 5 demonstrate the rules and algorithm for rule-based candidate equivalent set extraction and mapping.

Table 7. Candidate equivalence rules.

Equivalent candidates		**Feature Abbreviations**	
English	Arabic	Person	1: First
1S	1S		2: Second
1P	1P		3: Third
2S	2S,2MS,2FS	Number	S: Single
2P	2MD,2FD,2D,2MP,2FP,2MP		P: Plural
2O	2S,2MS,2FS,2MD,2FD,2D,2MP,2FP,2P		D: Dual
3MS	3MS	Gender	M: Masculine
3FS	3FS		F: Feminine
3S	3MS,3FS		O: None/other
3P	3MP,3FP,3D,3MD,3FD		

```
Candidate set production
For selected verse:
for every E pronoun in English pronoun list of verse:
    for every A pronoun in verse Arabic text:
        if A has no antecedent using QurAna:
            ant=0
        else:
            ant=1
        if (all features (f) are equal between E and A.
          or E.f is '2P' and A.f is in ('2MP','2FP','2D','2MD','2FD')
          or E.f is '2S' and A.f is in ('2MS','2FS')
          or E.f is '3P' and A.f is in ('3MP','3FP','3D','3MD','3FD')
          or E.f is'3S' and A.f is in ('3MS','3FS')
          or E.f is '2O' and A.f is in ('2P','2S','2MS','2FS','2MP','2FP','2D','2MD','2FD')):
             add (E.f, A.f, A.antecedent ID, ant) to candidate equivalent set of E
Alignment
For selected verse:
for every E pronoun in English pronoun list of verse:
  C= candidate equivalent list(E)
  if len(C)==0:
     Mapped and Has No Equivalent
  elif len(C)==1:
     Mapped and Has Equivalent
  else:
     If all items in C refer to the same concept:
        Mapped and Has Equivalent
     else:
        Not Mapped and is Ambiguous
```

Fig. 5. Steps 3 and 4 of mapping algorithm.

Fig. 6 shows the results of pronoun mapping for the dataset. The non-equivalent category shows the English pronouns that have no candidate in Arabic, and the ambiguous situation means that in the Arabic text, there is more than one acceptable referent concept for the pronoun. As the best estimation, in the Yusuf Ali translation, 71% of pronouns have equivalent English pronouns, while Hamid Aziz and Mohammad Ali have the lowest mapping results. For finding the cause for the low percentage, the structure of translations has been analyzed.

The obvious reason is segmentation mismatching between the translations and the main text. If a translator uses a verse numbering and segmentation that is differ from the original verse structure in the Arabic text, the mapping algorithm compares two distinct pronoun sequences between the Arabic and English, which causes a moderate accuracy (e.g. Mohammad Ali has translated verse 43 in chapter 2 of the Arabic text to 2-44, or Hamid Aziz has translated verses 6 and 7 in chapter 1 of Arabic text to 1-5, 1-6, and 1-7 in his translation). In future works, the TL text needs to be cleaned up and arranged correctly.

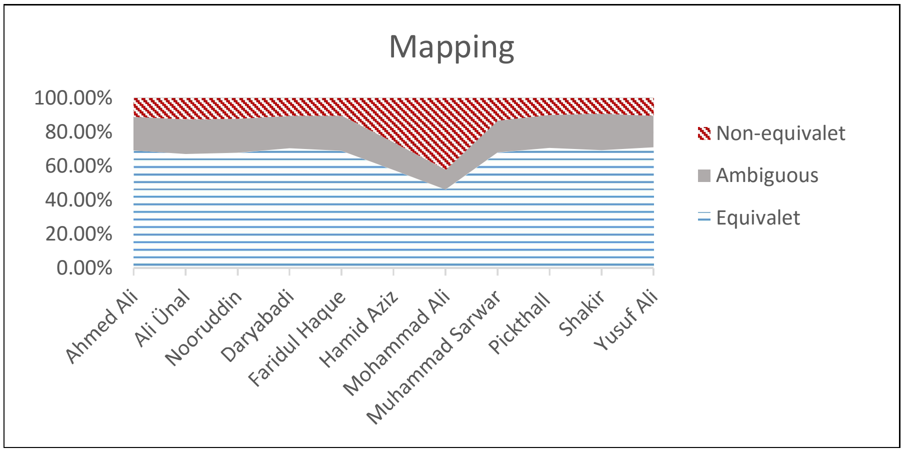

Fig. 6. The results of the baseline pronoun mapping.

4.2.1 Mapping of Ambiguous Pronominal

Due to ambiguous situations, the algorithm suffers from vagueness. Fig. 6 indicates that the method, on average, has no definite answer for over 18% of pronouns because there are two or more concepts in the Arabic as the candidate for mapping. For solving the problem, a concept frequency feature has been added to every Arabic pronoun in every chapter. The algorithm calculates the frequency distribution for every referred concept in the chapter, and, in ambiguous situations, the concept with higher frequency is selected as the mapping of the pronoun. Based on the Quran structure, every chapter is a separate text, and its semantics and discourse differ from the other chapters. Fig. 7 illustrates the results with the new feature set.

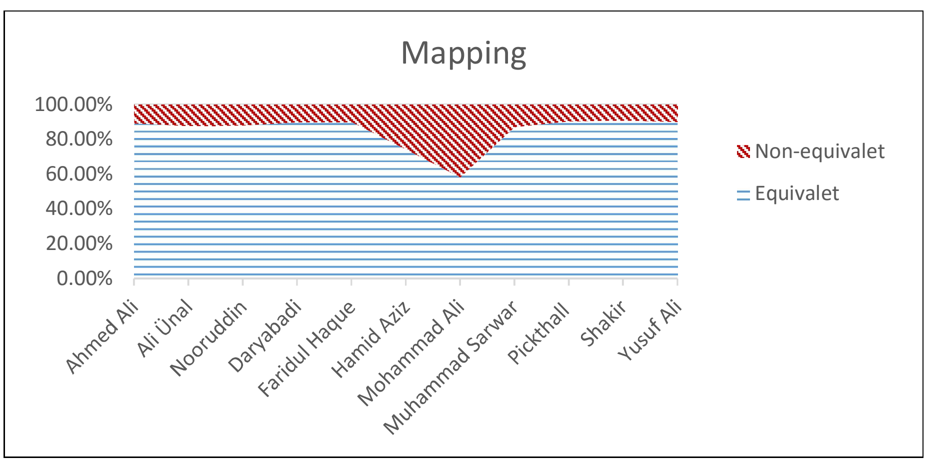

Fig. 7. The results of the pronoun mapping with concept frequency feature.

4.2.2 Mapping of Non-Equivalent Pronominal

If the algorithm emphasizes a reference concept for the pronouns that have no equivalent, then every pronominal in corpus will have a referent concept, and, as a complementary solution, will produce consistent output for this stage. This consistency is very useful for the pronominal AR engine, because the concept of every pronominal can be used as a feature for generating the candidate antecedent list. The complementary solution is a new statistical feature that was added to the algorithm. For every group of pronouns in Arabic, the algorithm generates the frequency, then, for each English pronoun that has no equivalent, the most frequent concept is chosen as the matched concept. Table 8 describes the statistic results for this matching. In the selected translation chapter (Pickthall chapter 30), the matching rule determines the concept ID 1 for all non-mapped pronouns, which is correct for nine pronouns (i.e. 60% in small dataset). After adding this rule, the algorithm has a unique referent concept for every pronoun with high accuracy (i.e. P, R, A, and F are 85.51%, 96.32%, 82.81%, and 90.59, respectively, for chapter 30 of the Pickthall translation). Fig. 8 illustrates the final version of the algorithm.

Table 8. Statistical concept frequency feature for mapping non-equivalent pronouns.

Matching candidate		**Most frequent concept**	
English	Arabic	Concept ID	Frequency
1S	1S ('1',264)	1	264
1P	1P ('1',1437)	1	1437
2S	2MS ('4', 821) ,2FS ('302', 24)	4	821
2P	2MD ('612', 6), 2FD ('397', 1), 2D ('457', 64), 2MP ('97', 919), 2FP ('507', 5), 2MP ('97', 919)	97	919
2O	2MS ('4', 821), 2FS ('302', 24), 2MD ('612', 6), 2FD ('397', 1), 2D ('457', 64), 2MP ('97', 919), 2FP ('507', 5)	97	919
3MS	3MS ('1', 1154)	1	1154
3FS	3FS ('12', 137)	12	137
3S	3MS ('1', 1154), 3FS ('12', 137)	1	1154
3P	3MP ('5', 785), 3FP ('105', 55), 3D ('182', 25), 3MD ('22', 8), 3FD ('182', 4)	5	785

```
Alignment
For selected verse:
for every E pronoun in English pronoun list of verse:
  C= candidate list(E)
  if len(C)==0:
      Map to the most frequent concept that is matched with E.f
      Has equivalent
  elif len(C)==1:
      Mapped and Has Equivalent
  else:
      If all items in C refer to the same concept:
          Mapped and Has Equivalent
      else:
          In items of C:
              Map to the candidate that is most frequent in the chapter
          Mapped and Has Equivalent
```

Fig. 8. Final version of steps 3 and 4 of the mapping algorithm.

Table 9 illustrates the final results of the mapping stage for verse 40 of chapter 30 in the Pickthall translation. For every pronominal, the result of the extractor and mapper stages has been added to the verse structure. In addition, an object has been created for each verse, which contains the translation name, position of pronoun (i.e. chapter, verse, and position inside verse), pronoun word, features, antecedent position in Arabic, and equivalent concept ID.

Table 9. Example of output of pronominal mapping component.

Verse text analysis	Object structure of verse
Say (O Muhammad, to the disbelievers): Travel in the land, and see the nature of the consequence for those who were before [you]<2O><ant='0 0' con='202'>! Most of [them]<3P><ant='87006 87008' con='429'> were idolaters. Pickthall [39], chapter 30, verse 42	verse_object: T= 'Pickthall' Chi = 30 Vi = 41 EngProVec = ([position:22,p:you,f:2O,ant:(0,0),conID:202] [position:25,p:them,f:3P,ant:(87006,87008),conID:429])

5.0 EXPERIMENTAL RESULTS AND CAMPARISON

The description of every pre-processing component contains the algorithm and the statistical evidences that proves the word lookup extraction and word alignment methods can overcome some of the challenges in the pronominal AR of Quran translations.

5.1 Evaluation of Pronominal Extraction

For the evaluation of the first step, all the extracted pronominal pronoun words in English text were checked with the TL text. The results show that all the pronoun words have been correctly emphasized. However, the results contain a group of non-anaphoric pronouns, which were considered as the wrong-positive data. Using a pleonastic and cataphoric pronoun identifier that discards non-anaphoric pronouns, the accuracy and precision of the component can increase to 100%. Table 10 illustrates the results of the extraction algorithm for the 1% annotated accurate data of the Pickthall translation, and compares it with the results of unigram POS-tagger that uses the texts of religion category in the Brown corpus for training.

Table 10. Results of the pronominal extraction stage, compared with the unigram POS-tagger.

Pronominal extraction algorithm			**Unigram POS-tagger**		
Pronoun category	Pronoun count	Results	Pronoun category	Pronoun count	Results
Pronominal (Total)	221	**P 96.38%**	Pronominal (Total)	221	P 91.55%
Anaphoric	213	**R 100%**	Anaphoric	213	R 91.98%
Non-Anaphoric	8	**F 98.23**	Non-Anaphoric	8	F 91.76
Extracted pronouns	221	**A 99.5%**	Extracted pronouns	212	A 98.38%

5.2 Evaluation of Pronominal Mapping

The evaluation of pronominal mapping stage demonstrates to what degree the algorithm has succeeded in aligning the pronouns correctly. The results have been calculated before and after adding the candidate frequency and concept frequency features to compare the effect of using statistical features (Table 11).

The accuracy of the baseline algorithm, which only uses person, gender, and number features, is close to 65%, which is acceptable for an automated rule-based solution. The difference between precision and recall demonstrates the weakness of the algorithm in solving ambiguous situations (i.e. having two or more acceptable referent concepts). In the second version, which solves the ambiguous mappings, the recall and the accuracy rise dramatically. The final version, which has a marginal increase in the results, generates the consistent output with a referent concept for every English pronominal.

Table 11. Evaluation of pronominal mapping stage.

First version (baseline)			
Pronoun category	Pronoun count		Results
Pronominal (Total)	221	P	94.44%
Mapped, No equivalent	15	R	66.02%
Mapped, Has equivalent	144	F	77.71
Not mapped	62	A	64.71%
Second version (solving ambiguity situations)			
Pronoun category	Pronoun count		Results
Pronominal (Total)	221	P	84.88%
Mapped, No equivalent	15	R	94.57%
Mapped, Has equivalent	206	F	89.46
Not mapped	0	A	81.45%
Third version (solving non-equivalent situations)			
Pronoun category	Pronoun count		Results
Pronominal (Total)	221	**P**	**85.51%,**
Mapped, No equivalent	0	**R**	**96.32%**
Mapped, Has equivalent	221	**F**	**90.59**
Not mapped	0	**A**	**82.81%**

[37] shows the result of supervised personal pronoun alignment for English-Czech bilingual text. The precision, recall, accuracy, and F-measure of their method respectively are 88.52%, 86.40%, 84.50%, and 87.45. The comparison between the results of our method and the supervised alignment method proves that the performance of our rule-based system is on a par with the supervised alignment (Table 12).

Table 12. Pronoun mapping comparison.

Evaluation	Our approach	Supervised alignment
Precision	85.51%	88.52%
Recall	96.32%	86.40%
F-measure	90.59	87.45
Accuracy	82.81%	84.50%

5.3 Extrinsic Evaluation

After generating and collecting information during the pre-processing stages, it is essential to describe how this data can be useful for the pronominal AR algorithm of the Quran and overcome the challenges that arise from the nature of the Quranic text. A pronominal cross-lingual anaphora resolution is explained and a test case is elaborated. Fig. 9 illustrates the general structure of the customized AR algorithm. The algorithm is not implemented in the current research and is left for future works. The stages of customized AR algorithm is:

1. Extract the key terms from the collected data of pre-processing to find the antecedent of the pronouns:

    - English term in the concept of every pronoun (e.g. 'those who believed and did righteous deeds' and 'Allah')

    - Translation of antecedent of every pronoun, if exist (e.g. 'those who believed and worked righteous works' and 'their lord')

2. Generate a candidate antecedent set for every pronoun by searching the specific verse or verses, based on the position of antecedent. This step uses a text snippet similarity search with semantic, structural and syntactic features.

    - If pronoun has an antecedent in Arabic, the algorithm searches the verse for the Arabic antecedent

    - In all cases, it searches the current verse. Sometimes the translator repeats the antecedent in the current verse or adds a new antecedent for some null cases.

```
GENERATING SEARCH PHRASE SET
For selected verse:
for every E pronoun in verse_object:
    Add English translation of the referent concept to the search phrase set
    If pronoun has antecedent:
        Generate its translation and add it to the search phrase set
TEXT SNIPPET MATCHING
    Using semantic, syntactic and structural features:
            Find similar text snippets in the verse and add to the candidate antecedent set
    C= candidate antecedent set(E)
    if len(C) =1
        set the antecedent and finish
    else if len(C)>1
        use preference algorithm to select the antecedent
    else:
        use an ordinary AR algorithm with noun phrase extraction
```

Fig. 9. Customized AR algorithm.

Table 13 illustrates an example of the complicated AR circumstances that demonstrate the effect of pre-processing. The antecedent of the first and second pronouns is a sentence. By conventional AR and PR algorithms, addressing this type of antecedent is difficult. While the Quran's Arabic text includes more than 2000 antecedents that their length is more than four words (i.e. almost all of them are sentences), the AR engine should be customized for resolving this complicated situations. The third pronoun shows that this approach is useful for complicated resolutions, and can also resolve the noun phrase antecedents. The third advantage of this approach is resolving the English equivalent of the pronominal anaphora that are without an explicit antecedent in Arabic. Although some of these pronominal pronouns have a referent in the English text, the result of the pre-processing stages is their default AR prediction.

Table 13. Example of output of pronominal mapping component.

Verse text analysis	AR results for the verse
Then, as for *those who believed and worked righteous works*, [their]<3P><ant=106776 106783 con=11> Lord will cause [them]<3P><ant=106776 106783 con=11> to enter into [His]<3MS><ant=106787 106788 con=1> mercy: that is a manifest achievement. Daryabadi [43], Chapter 45, verse 30	[their]<3P><ant=106776 106783 con=11> [them]<3P><ant=106776 106783 con=11> Con[11]: those who believed and did righteous deeds Google translation of antecedent: Those who believe and do good deeds Similar text in verse: *those who believed and worked righteous works* [His]<3MS><ant=106787 106788 con=1> Con[1]: Allah Google translation of antecedent: Their lord Similar text in verse: *their Lord*

6.0 CONCLUSION AND FUTURE WORKS

Improvement in pre-processing steps such as pronoun extraction and pronoun mapping boosts the accuracy of Pronominal AR. Pronominal AR of English translations of Quran requires specialized pre-processing that are designed based on extracted knowledge from the Arabic script.

This paper described novel algorithms for pronoun extraction and pronoun mapping tools as the pre-processing steps of pronominal anaphora resolution in the English translations of the Quran. The pronominal extractor tool detects all the pronominal pronouns in the translations using a rule-based algorithm at the word level, with the person, gender, and number features. This extractor is implemented using a word finder to detect the pronoun words instead of parsing and POS tagging the text. The advantage of this extraction method is its independence from the translation structure or style, because it is in the morphological layer. The word extractor detects the pronoun words and counts their frequency in the translations, then discards the words with zero frequency. The final set of pronouns includes 39 words. The pronominal mapper component uses a rule-based algorithm to align the English and Arabic pronouns in the SL and TL texts. The mapping provides every English pronominal with one referent concept and an antecedent in the discourse, if it exists. The baseline mapper uses the person, gender, and number features. With employing the concept frequency and the antecedent frequency calculation functions, the result of baseline mapper is improved.

For evaluating the tools, 1% of Pickthall translation [39] with 221 human annotated pronouns is used. The precision, recall, and accuracy of the pronoun extraction tool are 96.38%, 100%, and 99.5%, respectively. The result is correct for all 213 anaphoric pronominal in the corpus, and eight non-anaphoric pronouns were marked as false-positive errors. Thus, this tool requires a complementary step for discarding non-anaphoric pronominal from the results. The baseline mapper obtains 94.44%, 66.02%, and 64.71% for precision, recall and accuracy, respectively. With employing the concept frequency and the antecedent frequency calculation functions, the result is accurate for 183 of 213 anaphoric pronominal in the annotated data (i.e. 85.91% success rate, 85.51% precision, 96.32% recall, and 82.81% accuracy). The results of this rule-based method is on par with supervised pronoun alignment [37]. In Total, pre-processing pronominal extraction and mapping sub-systems succeeded in 213 of 221 (96.4%) and 183 of 221 (82.81%) human annotated pronouns, respectively.

The description of the customized algorithm for the pronominal AR system explained how the algorithm can resolve complicated AR situations, such as the reference to a sentence. As a future work, the pronominal AR algorithm should be implemented for corpus and be evaluated using the annotated data. Annotating a bigger dataset can increase the reliability of evolution and may also highlight more exceptional situations that should be handled. While the majority of errors in the mapping algorithm arise from the ambiguous situations, the effect of new features, such as pronoun order and the accuracy should be studied. In addition, the long verses that include several pronouns needs internal segmentation in Arabic and English to increase the success rate of the pronominal anaphora resolution algorithm.

ACKNOWLEDGEMENT

The authors acknowledge the financial support under the University Malaya Research Grant (Grant No.: UM.TNC2/IPPP/UPGP/628 (UMRG) RG087/12ICT) from the University of Malaya, Malaysia.

REFERENCES

[1] R. Mitkov, "Anaphora resolution: the state of the art", *School of Languages and European Studies, University of Wolverhampton*, 1999.

[2] R. Mitkov et al., "A new, fully automatic version of Mitkov's knowledge-poor pronoun resolution method". in *Computational Linguistics and Intelligent Text Processing*, Springer Berlin Heidelberg, 2002, pp. 168-186.

[3] R. Mitkov et al., "Introduction to the special issue on computational anaphora resolution". *Computational Linguistics,* Vol. 7, No. 4, 2001, pp. 473-477.

[4] A. Refoufi, "Pronominal Anaphora Resolution Using XML Tagged Documents" *J Comput Eng Inf Technol,* Vol. 3 No. 1, 2014.

[5] J. D'Souza and V. Ng, “Anaphora resolution in biomedical literature: a hybrid approach”, in *Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine*. ACM, 2012, pp. 113-122.

[6] S. Broscheit, et al. "BART: A multilingual anaphora resolution system”, in *Proceedings of the 5th International Workshop on Semantic Evaluation*. Association for Computational Linguistics, 2010, pp. 104-107.

[7] S. Lappin and H. Leass, “An algorithm for pronominal anaphora resolution” *Computational Linguistics*, Vol. 20 No. 4, 1994, pp. 535-561.

[8] B. Baldwin, “CogNIAC: high precision coreference with limited knowledge and linguistic resources”, in *Proceedings of a Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts*, Association for Computational Linguistics, 1997, pp. 38-45.

[9] C. Kennedy and B. Boguraev, “Anaphora for everyone: pronominal anaphora resolution without a parser”, in: *Proceedings of the 16th International Conference on Computational Linguistics (COLING'96)*, Copenhagen, Denmark, Vol. 1, 1996, pp. 113-118

[10] R. Mitkov, “Robust pronoun resolution with limited knowledge”, in *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics,* Vol. 2, Association for Computational Linguistics, 1998, pp. 869-875.

[11] R. Mitkov and L. Belguith, “Multilingual robust anaphora resolution”, In *Proceedings of the 3rd Conference on Empirical Methods in Natural Language Processing*, Granada, Spain, 1998, pp. 7-16

[12] R. Mitkov et al., “Anaphora resolution: To what extent does it help NLP applications?”, *Anaphora: Analysis, Algorithms and Applications*. Springer Berlin Heidelberg, 2007. pp. 179-190.

[13] C. Hardmeier and F. Marcello, “Modelling pronominal anaphora in statistical machine translation”, in *Proceedings of the International Workshop on Spoken Language Translation*, 2010, pp. 283-289.

[14] J. Steinberger et al., “Two uses of anaphora resolution in summarization”, Information Processing & Management Vol. 43, No. 6, 2007, pp. 1663-1680.

[15] C. Orasan, “The Influence of Pronominal Anaphora Resolution on Term-based Summarisation”, *Recent Advances in Natural Language Processing V: Selected Papers from RANLP 2007*, John Benjamins, 2009, pp. 291-300

[16] O. B. Abiola et al., “Review of the Various Approaches to Text to Text Machine Translations”, *International Journal of Computer Applications*, Vol. 120, No. 18, 2015, pp. 7-12.

[17] M. Baker, *In other words: A Coursebook on translation,* Routledge, 2011.

[18] T. Liang and D. S. Wu, “Automatic Pronominal Anaphora Resolution in English Texts” in *Computational Linguistics and Chinese Language Processing* , Vol. 9, No.1, 2004, pp. 21-40

[19] C. Barbu and R. Mitkov, “Evaluation tool for rule-based anaphora resolution methods”, in *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, Association for Computational Linguistics, 2001, pp. 34-41.

[20] R. Mitkov, “Outstanding issues in anaphora resolution” in *Computational Linguistics and Intelligent Text Processing*, Springer Berlin Heidelberg, 2001, pp. 110-125.

[21] S. Hammami et al., “Arabic Anaphora Resolution: Corpora Annotation with Coreferential Links”, *The International Arab Journal of Information Technology*, Vol. 6, 2009, pp. 480-488.

[22] K. M. Seddik et al., “Arabic Anaphora Resolution: Corpus of the Holy Quran Annotated with Anaphoric Information”, *International Journal of Computer Applications*, Vol. 124, No. 15, 2015, pp. 35-43

[23] A. M. Sharaf and E. Atwell, “QurAna: Corpus of the Quran annotated with Pronominal Anaphora”, in *LREC*, 2012, pp. 130-137.

[24] A. A. Tabrizi and R. Mahmud, “Issues of coherence analysis on English translations of Quran” in *1st International Conference on Communications, Signal Processing, and their Applications (ICCSPA),* IEEE, 2013, pp. 1-6.

[25] M. P. Marcus et al., “Treebank III”, 1999.

[26] M. P. Marcus et al., “Building a large annotated corpus of English: The Penn Treebank”, *Computational linguistics,* Vol. 19, No. 2, 1993, pp. 313-330.

[27] W. N. Francis and H. Kucera, “Brown corpus manual”, *Brown University Department of Linguistics,* 1979.

[28] A. Abu-Mahfouz, “Some issues in Translating Nouns in Abdullah Yusuf Ali’s Translation of the Meanings of the Holy Quran”, *Jordan Journal of Modern Languages and Literature*, Vol. 3, No. 1, 1975, pp. 65-83.

[29] H. Abdul-Raof, *Qur'an translation: Discourse, texture and exegesis*. Routledge, 2013.

[30] S. M. El-Awa, *Textual relations in the Quran: relevance, coherence and structure*, London, RoutledgeCurzon, 2006.

[31] J. Al-Qinai et al., “Convergence and Divergence in the Interpretation of Quranic Polysemy and Lexical Recurrence”, *Studies About Languages*, No. 19, 2011, pp. 27-38.

[32] J. Dror et al., “Morphological Analysis of the Qur’an”, *Literary and linguistic Computing*, Vol. 19, No. 4, 2004, pp. 431–452.

[33] K. Dukes et al., “Syntactic Annotation Guidelines for the Quranic Arabic Dependency Treebank”, in *Proc LREC’2010: Language Resources and Evaluation Conference*, 2010, pp. 1822-1827.

[34] K. Dukes and N. Habash, “Morphological Annotation of Quranic Arabic”, in *Proc LREC’2010: Language Resources and Evaluation Conference*, 2010.

[35] A. M. Sharaf, “The Qur’an Annotation for Text Mining”, *Review Literature And Arts Of The Americas*, December 2009.

[36] E. D. Sharifabad et al., “Linguistic Ambiguity in the Holy Quran and Its English Translations”, *International Journal of Linguistics*, Vol. 4, No. 3, 2012, pp. 352.

[37] M. Novák and Z. Zabokrtský, “Cross-lingual Coreference Resolution of Pronouns,” In *COLING*, 2014, pp. 14-24.

[38] M. Ali, *The Holy Qur'an,* 1917.

[39] M. Pickthall, *The Meaning of the Glorious Qur’an.* Hyderabad, Hyderabad Government Press, 1930.

[40] A. J. Arberry, *The Koran Interpreted: A Translation,* New York, George Allen & Unwin Ltd., 1955.

[41] M. H. Shakir, *Holy Qur'an*, New York, Tahrike Tarsile Qur'an Inc., 1987.

[42] M. T. Sarwar, *The Holy Qur'an*, Islamic Seminary Inc., 2001.

[43] A. M. Daryabadi, *Holy Qur'an*, Taj Company Ltd., Karachi, Pakistan, 1970.

[44] A. Ali, *The Qur'an: A Contemporary Translation*, Karachi, Akrash Printing, 1984.

[45] A. Nooruddin et al., *The Holy Qur'an: Arabic Text-English Translation*. Holy Quran, 1997.

[46] A. Ünal, *The Qur'an with Annotated Interpretation in Modern English*. Tughra Books, 2006.

[47] S. F. Haque, *The Holy Quran, from Kanzul Iman,* Delhi, 1988.

[48] H. S. Aziz, *The Meaning of The Holy Quran, with Explanatory Note*, online at http://www.qurandatabase.org

[49] E. Loper and S. Bird. "Natural Language Toolkit", 2005.