Coherence Analysis Issues on English-Translated Quran

Arash Amini Tabrizi, Rohana Mahmud
Department of Artificial Intelligence
Faculty of Computer Science and Information Technology
University of Malaya
Kuala Lumpur, Malaysia
E-mail: aminitabrizi@siswa.um.edu.my, rmana@um.edu.my

Abstract—This study shows issues of comparing English-translations of Holy Quran and its Arabic text from discourse structure point of view. There are several different translations of Quran, which differ in structure and word domain. In these translations, the order of sentences, phrases, and words is different, not only with each other but also with the Arabic holy Quran, which affects computational text analyzing. This research is a preliminary stage of study the issues, constructing a platform and defining of some preliminary rules for comparing and evaluating discourse structure of translations.

Keywords-Quran English translation; coherence; lexical cohesion; translation evaluation; discourse processing issues

I. INTRODUCTION

Holy Quran is the religious text of more than 1.5 billion Muslims around the world which are speaking in different languages. Although Muslims use the Arabic text of Holy Quran for reading, it can be helpful to have its translation and well-formed interpretation in every nation’s mother tongue or an internationally studied language like English for creating a better understanding of Quran. In addition, non-Muslims can have the ability of understanding the holy Quran using its exact translations which provides them to have a fair view of Islam. It is not to forgotten the obligation of resistance against serious attempts for constructing an incorrect view of Muslims’ believes. Translation of Quran to English which is a part of widely translation efforts on religious texts is not a new domain. First English translation of Quran has been offered 1649 by Alexander Ross from French, and the first direct translation from Arabic to English was produced at 1734 by George Sale reflecting his missionary stance. Nowadays we can read more than 80 translations of Quran which their accuracy should be discussed from various aspects. Because of the nature of Quran as a divine text, even having a professional translator, its translation to English encounters with several hardships. Ambiguities of words, lack of word equivalence between English and Arabic, and different structures of word, sentence, and discourse in these two languages are from the numerous issues in translation.

These translations are a significant part of Muslims’ documents, so their accuracy and equivalence with the original Arabic Quran are arguable and need to be studied in several aspects. Coherence and cohesion is two significant factors which can help researchers to have a discourse structure of text and make them able to compare translations. Discourse processing methods can be useful for opening a discussion on this field. In recently years, some of these translations have been converted to computerized texts and even implemented in data-base structures for using in computational analysis.

Based on this corpus, it is a new idea to study translations of Quran form entity coherence and lexical cohesion point of view, as an evaluation factor of existing translations and even using that for machine translation in the future. In this paper, we discuss the preliminary problems of this analysis.

II. PREVIOUS WORKS

In this part, we have a general review of previous studies and implementations on discourse level knowledge of Quran. Although the Quran is one of the prominent religious books of the world, it seems to be a lack of computational discourse analysis studies on Quran, particularly on its translations. Contributions to this field include the analysis of Quran text using entity-based coherence evaluation methods for developing a structural model of every translation and comparing it with the Arabic original text structure.

In one side, from a linguistic point of view, there are numerous studies on Quran discourse, especially on coherence. There are some books written on these subjects [11,18], in addition to some attempts of researchers to clarify issues and facts of discourse analysis on Quran [14,4]. Although these texts have discussed regarding discourse analysis of Arabic text or English translation of Quran, almost all of them have not used computational methods or definitions for their analysis. As a result, they cannot be used as an evaluation and comparing method on Arabic and translated texts. Consequently, they have not faced with problems of processing methods on Quran.

On the other side, even inside the domain of computational processing of Quran, the focus of this new study is on discourse, so the researches which are useful in the other levels, cannot be referenced as background works. One of the most integrated corpus-based studies on Quran is an ongoing interest of the Language Research Group in the School of Computing in University of Leeds. In their project, they produced a complete annotation of Quran [10], for text mining, information extraction, text categorization, concept linkage, and discovery of association [27]. This project has a
word-by-word translation of Quran. In the other attempt, they have collected six translations, which provide readers with the ability of comparing between several meanings of every verse. In addition, they have a tree bank of grammatical dependency for every verse [10]. This project is in parallel with an implementation of a Discourse Tree Bank for Modern Arabic Language which can be widely used as supportive system in Arabic studies. Although we will use results of this system in the next stage of this research as a robust platform, except their pronunciation resolution system for Arabic Quran, this project and its related attempts have not introduced any direct idea on discourse processing level of translated Quran, especially in analyzing of coherence.

There are some efforts in morphological and syntactic layer which show that processing on Quran in these levels has grown during recent years [2,9].

In conclusion, working on coherence computational analysis of Quran translations, as a tool for their evaluation, is a new field in Quranic studies.

III. QURAN TEXT STRUCTURE

Quran consists of 114 suras (as chapters), which have obvious boundaries in text. In general, the longer suras appear earlier in the Quran, while the shorter ones appear later. Each sura is formed from several ayats or verses. Neither the number of verses in suras, nor the word counting of verses is the same. The Arabic Quran corpus consists of 77784 word tokens and 19287 word types [1].

Arabic language of Quran differs from Modern Standard Arabic (MSA) which is used in writing and formal speaking. The main difference is in Syntax and lexicicon. Based on this difference, it would be arguable that even the methods and tools for analyzing Arabic modern text may need to change for Quranic discourse analysis.

IV. ENGLISH TRANSLATIONS OF QURAN

Translation of Quran can be done in a range of methods, varies from very literal to free translation.

In one side, we have form-oriented or literal method, which supports formal equivalence. Every word and structure in its position is translated to the target language. This method produces the nearest translation to the source text from a structural and grammatical point of view. In the dark side, because of syntactic and semantic differences between Arabic and English languages, the final text is not communicative and the reader should have background knowledge of the source language to understand the target text. Translators use parenthetical texts and footnotes to clarify the meaning of the source text to solve this problem.

In the other side, in Paraphrastic method, which can be a free translation, translator tries to show the aim of every phrase and part of source text in the target language, based on his personal interpretation of the text. Translations based on this method are understandable for even unfamiliar readers, but they are inevitably combined by interpretation. As a result, they can be questionable in accuracy level, because there is no clear boundary between the source text and interpretation of translator.

In the middle way, content-oriented methods, which cover dynamic-equivalence [22], translate words to their natural equivalent, using local idiomatic expressions, common vernacular, and contemporary speech in English.

Nowadays we can read more than 80 translations of Quran. Most translations of the Quran are source language oriented [4].

Word-by-word translation of Leeds project 1 is an instance of formal translation. Table I shows nine of the prominent Quran translations. These translations, which we use as the source data, have been converted and computerized in digital formats like SQL and XML2.

V. ENTITY COHERENCE

Entity coherence results from the way in which, every element of discourse is related to subsequent ones by reference of some entities (e.g. noun phrases). Methods of entity-based coherence analysis investigate pronoun resolution and other types of entity relationships in discourse [17,19] Centering theory [12,15], theory theme and rhyme [13], and Entity Grid [16] are the prominent methods for modeling entity coherence in computational linguistics.

Miltatskaki and Kukich [20] defined the automatic evaluation of student essays using entity coherence. Cherry studied the effects of coherence on machine translation systems [7].

Entity coherence is related to the lexical and referential cohesions. The term cohesion is often used synonymously with coherence, as well as being used to describe a kind of coherence [29].

VI. RESULTS OF ISSUES STUDY

In this section, we demonstrate two categories of issues on study of coherence in English translations of Quran. If we want to analyze coherence in Arabic and English texts and compare them, it is compulsory to have a parallel study in Arabic and English corpus and highlight the issues which are realized during this study.

A. General Issues

First group of problems is summarized from the previous sections as a general view of Quran English translations.

<table>
<thead>
<tr>
<th>Table I. Samples of Quran English Translations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Translator</strong></td>
</tr>
<tr>
<td>M. Pickthall [23]</td>
</tr>
<tr>
<td>M. H. Shakir [26]</td>
</tr>
<tr>
<td>S. A. Qura'</td>
</tr>
</tbody>
</table>

1 http://corpus.quran.com/wordbyword.jsp
2 http://qurandatabase.org/Database.aspx
1) Synonymy in translations: Synonymy is a factor that shows the grammatical and lexical similarity between texts. The style of translators to choose different structure for sentences affects synonymy in grammar, whereas word domain of translation affects the lexical synonymy between several translations of a same text [28].

Related to several factors as the time period of translation, background knowledge, and translator’s literary style, the structure and word domain of translations are different.

As an instance, we compare four translations of a short verse.

3 (sura 1: verse 7)

Yusuf Ali “The way of those on whom Thou hast bestowed Thy Grace, those whose (portion) is not wrath, and who go not astray.”

Shakir “The path of those upon whom Thou hast bestowed favors. Not (the path) of those upon whom Thy wrath is brought down, nor of those who go astray.”

Pickthall “The path of those whom Thou hast favoured; Not the (path) of those who earn Thine anger nor of those who go astray.”

Sarwar “The path of those to whom You have granted blessings, those who are neither subject to Your anger nor have gone astray.”

The highlighted parts of translations realize that the word domain of translators and their method to find words and phrases can be different with each other. In addition, tense of verbs is different in translations. Finally, their style of translation is different with each other.

From coherence analysis point of view, algorithms and methods for processing these texts, either should work on a small group of similar translations, or must be a comprehensive solution with a low degree of attention on effects of difference.

2) Difference in Translation Method: Method of translation can affect the possibility of language analysis. As the first Group, formal translations of Quran are accepted for syntactic, morphological, and ontological analysis, while their fragmentation in form and structure, frequently use of parenthetical words and phrases, and semantic problems in the target language should be discussed in discourse studies.

In the other side, the overall same order of words in translated and Arabic texts can be extremely suitable for entity-based coherence evaluation, which uses relationships of noun phrases. As an instance, the translation of the first part of verse (1:7), with annotated information of pronouns and verbs is:

|
| مسارط أَنْعَمْتَ عَلَيْهِمْ غير المخضوب عليهَمْ ولا أُصِرَٰطَ (the) path of |
| ﴿وَٱلَّذِينَ (those)﴾ |
| انَّمَ + ت [perfect verb] you [2nd person masculine singular subject pronoun → Allah] + have bestowed Favors |
| علىَ [on] |
| ﴿وَٱلَّذِينَ (Them [3rd person masculine plural])﴾ |

For the remainder of this verse, it is difficult to translate word-by-word Arabic to English without ambiguity in the understanding and lose of the structure.

The second group, Paraphrastic translations of Quran, which has interpreted and paraphrased the source text to the target language, can be studied easily in some domains of NLP as text summarization, but they suffer from a high degree of structure and word lose. As an instance, two translations of verse (1:7) can be compared with previous ones:

Qara’i “the path of those whom You have blessed such as have not incurred Your wrath, nor are astray.”

Daryabadi “The path of those whom Thou hast favoured. Not of those on whom is indignation brought down, nor of the astray.”

The third group, content-oriented, as the largest group of Quran translations, can be tested in coherence evaluation, although behavior of these translations during processing should be studied and compared, based on their use of best equivalence methods in every part of the translation. We pointed a sample of this type in the issue 1.

Table II shows a comparing between six English translations by counting parenthetical cues. The first four translations have used third Translation method, and their parenthetic cues are between 59 to 174 in the Al-Emran sura by 200 verses and about 6000 word tokens. As a character of his method, Daryabadi has not used parenthesis

<table>
<thead>
<tr>
<th>Translation</th>
<th>Count of () and []</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ahmad Ali</td>
<td>59</td>
</tr>
<tr>
<td>Mohammad Ali</td>
<td>70</td>
</tr>
<tr>
<td>Sarwar</td>
<td>80</td>
</tr>
<tr>
<td>M. Pickthall</td>
<td>174</td>
</tr>
<tr>
<td>A. Daryabadi</td>
<td>1</td>
</tr>
<tr>
<td>Formal Translation (Leeds project)</td>
<td>799</td>
</tr>
</tbody>
</table>

3 All translations are from free service of Online Quran Project (http://community.al-quran.info/)

4 http://corpus.quran.com/wordbyword.jsp
in translation (three in the whole of the Quran and one in this sura). In the opposite side, Leeds project, as an instance of word-by-word translation, has used numerous parenthetical cues in this sura, which has been counted by the human annotator.

B. Discourse Study Issues

3) Pronoun Resolution: Pronoun resolution, anaphoric or deictic, helps us to have more detailed understanding from the aim of the text. In discourse processing field, pronoun resolution is an fundamental ability to find discourse relations and evaluate coherence and cohesion, especially from entity-based point of view. The role of pronouns seems even more critical than common circumstances, when it is connected to the other complexities in discourse. In Quran, direction of discussion can be changed inside every part of the text. The role of pronouns in these changes can be studied in detail.

There are some implementations for finding reference of pronouns in Arabic Quran. Sharaf [27], in his discussion on his implementation, have expressed the issues of pronoun resolution in Quran. Number agreement, multiple pronouns in a single Arabic word, reference to a verb, action or entity (like a phrase), ambiguity of particular or general referencing, and having several reference for a pronoun are some of the issues in this field [27]. This implementation has categorized pronoun of Quran to three groups and covered personal pronouns group.

For implementation of the same system in English to use in discourse processing, we should study that from two overall points of view. In one side, we should have an evaluation of Arabic text, which can be human annotated for entity-base coherence studies. On the other side, it needs an attempt to evaluate existing translations on their referential accuracy in comparison with the main text.

In both situations, we should face with several issues. The first group of issues which has been discussed by Sharaf is propagated from Arabic language to English, and every attempt for using pronoun resolution in discourse level should face with their difficulties and ambiguities. The next group of problems is created in English translation, which we point only one of them.

Every pronoun potentially has a referent which should be in agreement with in some aspects. In Arabic language, we have a complete category of pronouns which support every gender and number.

As a small group, the types of pronouns for “you” in Arabic are:

- أي: 2nd person masculine single
- ائت: 2nd person feminine single
- ائضا: 2nd person masculine or feminine for two people
- أتم: 2nd person masculine or for a group of both, more than two
- أئن: 2nd person feminine plural, more than two

This structure is similar for every type of pronouns.

As an instance of pronoun structure, we point a complicated situation.

(1) You: explicit pronoun, single, masculine, implicit reference is Mohammad
(2) Say (you): implicit pronoun, single, masculine, implicit reference is Mohammad
(3) Ask (you): implicit pronoun, single, masculine, implicit reference is Mohammad

Pronoun shift

(4) You: explicit pronoun, more than two, masculine, explicit reference is the People of the Book and the illiterate ones

Pronoun shift

(5) Your: explicit pronoun, single, masculine, implicit reference is Mohammad

This verse has several pronoun shifts, but only two of them are highlighted. In this translation, the pronouns you and your and their dependent verbs (say and ask) refer to Prophet Mohammad, but in the middle of the verse, we see pronoun you which refers to the People of the Book and the illiterate ones. In Arabic, the number of these pronouns is different and can be recognized easily, but in modern English, we face ambiguity of pronoun resolution process.

On the other side, pronouns should have an agreement with their reference in number and gender which may be a complicated situation during translation.

(8) Pickthall

“...And the believers, men and women, are protecting friends of one another; they enjoin the right and forbid the wrong, and they establish worship and they pay the poor due, and they obey Allah and His messenger. As for these, Allah will have mercy on them. Lo! Allah is Mighty, Wise.”
poor-rate, and obey Allah and His Apostle; (as for) these, Allah will show mercy to them; surely Allah is Mighty, Wise.”

Shakir

These referential relations are a tool for analyzing coherence in Arabic text. In English translation of Quran, if these factors are not supported in pronoun and verb structure, we will lose a wide ability in evaluating coherence. For analyzing this effect, we implemented a simple pronoun counter for Quran English translations which can count every pronoun word in a part or all of the Quran. Table III shows a part of this system results as the comparison of the second person pronouns usage in five translations of Quran.

These results realize that, translations of Ahmad Ali and Sarwar have used two pronoun, you and your in the majority of positions which makes an ambiguous situation between second singular and second plural reference in pronoun resolution process. However, Mohammad Ali, Pickthall, and Daryabadi translations use such words as thou, thee, thy, and thine, exclusively (single second pronoun) and ye (plural second pronoun), which can be very useful in comparing between Arabic and English pronoun resolutions of Quran.

As an instance, the Pickthall’s translation of verse 20 in sura 3 is more unambiguous than the Shakir’s translations in pronoun resolution, which previously has been shown.

(sura 3: verse 20):
“...And if they argue with thee, (O Muhammad), say: I have surrendered my purpose to Allah and (so have) those who follow me. And say unto those who have received the Scripture and those who read not: Have ye (too) surrendered? If they surrender, then truly they are rightly guided, and if they turn away, then it is thy duty only to convey the message (unto them). Allah is Seer of (His) bondmen.”

Pickthall

Another way to solve the problem is referring the main text and trying to find similar references in translated text. It will be a human annotation and can be implemented in word-by-word translations. In Quran project of Leeds, it has been done in Arabic corpus and has been related to formal translation of Quran. Unfortunately, this method cannot be used for other translations of Quran as a general solution, because there is not an annotated implementation for all of them, so it will be an attempt to find a general method to solve pronoun resolution in a wide range of translations.

2) Word and Phrase Ordering: Like every translation process, the order and structure of words in a text may be changed in the target language, for having a smooth and understandable result. These changes can be affect methods of discourse processing, which depend to word numbering and word ordering, especially for comparing translations with each other and the main text. The other aspect of the issue is happened when these changes are in the phrase or sentence level. Rerordering, merging or splitting of discourse elements in translation, due to specific semantic and idiomatic reasons, cause difficulties in analyzing Quran.

A simple instance of these issues is shown for a small part of a verse by three words and five word tokens:

(Yusuf Ali “On them will God pour His mercy”)
(3 words or 5 word tokens)

Yusuf Ali “On them will God pour His mercy”

Pickthall “It is they to whom Allah will soon grant His mercy”

Shakir “(As for) these, Allah will show mercy to them”

Sarwar “God will have mercy on them”

The third aspect is paraphrasing. In complications of Quran translation, some words change to a phrase or vice versa. As an instance, word المغضوب عليهم has no equivalent in translations and should be translated to a phrase.

(Yusuf Ali “those whose (portion) is not wrath”)
(Pickthall “Not the (path) of those who are neither subject to Your anger”)
(Shakir “Not the (path) of those upon whom Thy wrath is brought down”)

Table IV shows the word counting of sura Al-Emran (third sura of Quran by 200 verses) in four translations. Based on Morphological analysis of Arabic Quran, Produced by Tanzil website, word counting for this sura is 5752

(sura 7: verse 48: part 1)

— Translated by Tanzil website

5 http://tanzil.info

<table>
<thead>
<tr>
<th>Table III. Second person pronouns counting for translations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>you</td>
</tr>
<tr>
<td>yourself</td>
</tr>
<tr>
<td>yours</td>
</tr>
<tr>
<td>your</td>
</tr>
<tr>
<td>thou</td>
</tr>
<tr>
<td>thee</td>
</tr>
<tr>
<td>thyself</td>
</tr>
<tr>
<td>thine</td>
</tr>
<tr>
<td>yourselves</td>
</tr>
<tr>
<td>ye</td>
</tr>
<tr>
<td>thine</td>
</tr>
</tbody>
</table>
This increase of word counting can depend on each of the three described structural changes during the translation process.

Coherence and cohesion analysis of translations has more difficulties in situations which process methods depend on word and phrase order and word counting in text. Its first and foremost effect is in segmenting of a text to its elements for coherence analysis.

VII. CONCLUSION

In this research, we discussed some issues which entity-coherence and lexical cohesion analysis on English translation of Quran should face them. In one side, differences in word domain, structure selection, and method of translations as the common issues and in the other side, pronoun resolution and word and phrase ordering as structural issues were discussed.

REFERENCES


