Proceedings of the New Trends in Information Technology (NTIT-2017).
The University of Jordan, Amman, Jordan. 25-27 April 2017.

Construction of an English-Arabic Political Parallel Corpus

Alia Al-Sayed Ahmad
Linguistics Department, University of Jordan, Amman, Jordan
Bassam Hammo
Computer Information Systems Department, King Abdullah II School for Information Technology, University of Jordan, Amman, Jordan
Sane Yagi
Linguistics Department, University of Jordan, Amman, Jordan
E-mail: aliaahmadsh@gmail.com, b.hammo@ju.edu.jo, saneyagi@gmail.com

Abstract

This study reports on the construction of an English Arabic Political Parallel Corpus (EAPPC)، which will be a useful resource for research in translation, language learning and teaching, bilingual lexicography, contrastive studies, political science studies, and cross-language information processing. It describes the phases of corpus compilation, and explores the corpus, by way of illustration, to discover the translation strategies used in rendering the word *takfiri*. A total of 189 speeches, 80 interviews, and 68 letters, which were translated by the Royal Hashemite Court, were selected and culled from King Abdullah II's official website, in addition to the textual material of the English and Arabic versions of His Majesty's book, *Our Last Best Chance: The Pursuit of Peace in a Time of Peril (2011)*. The texts were meta-annotated, segmented, tokenized, English-Arabic aligned, stemmed, and POS-tagged. The challenges encountered in corpus compilation were found to be the scarcity of freely available machine-readable Arabic-English translated texts, and the deficiency of software that process Arabic texts.

Keywords: parallel corpus, political, English-Arabic translation, corpus compilation, challenges

1. INTRODUCTION

A parallel corpus is a collection of original (i.e. source language) texts and their translations (i.e. target language). These texts can be aligned at different levels, such as paragraph, sentence, phrase, or word level. Bilingual concordances are used to display all the occurrences of a search term in the source language (SL) together with their equivalents in the target language (TL). This type of corpora plays a crucial role in research that involves two or more languages, such as machine translation, cross-language information processing, contrastive studies, language research, language learning and teaching, and bilingual lexicography [1-6].

1. 1 Parallel and Comparable Corpora

Parallel and comparable corpora are two types of multilingual corpora. Both parallel and comparable corpora consist of texts in two or more languages, but the first type requires that there be a source language and a translation version of the same texts. The latter, on the other hand, makes no such requirement. The texts it contains are merely of the same sampling frame (e.g., the same text size from the same genres, and published in the same period of time). Parallel and comparable corpora are invaluable resources for translation and contrastive studies in particular[7].

1. 2 Challenges of Compiling Parallel Corpora

Developing a parallel corpus is not a straightforward task. This is due to technical and linguistic challenges encountered at most stages of construction, such as text selection, conversion, segmentation, stemming, alignment, and annotation.[8-11].

In the text selection process, it is challenging to find a large number of translated texts that are available in the desired language combination. These texts should be machine-readable, accessible, and representative samples of the specific language combination [8, 9, 12].

Moreover, the inefficient tools for text conversion (e.g., OCR), segmentation, tokenization, and part of speech tagging cause a series of challenges, especially when Arabic is involved.

As far as automatic text alignment is concerned, Anderman and Rogers (2008) emphasized that it is not an easy task for minority languages or languages of technologically underdeveloped countries. Similarly, Veronis (2000) stressed the issue of the high cost of aligning a large amount of texts. That is

why it is difficult to find parallel texts of any significant size that are aligned at word or phrase level [13]. Arabic is not a minority language since it has more than 300 million speakers around the world, yet it lacks reliable text-conversion, text-alignment, and text-annotation tools.

2. LITERATURE REVIEW

2. 1 English Parallel Corpora

Parallel corpora began to appear in 1988, when Bell Communications Research and the IBM T J. Watson Research Center compiled the Hansard corpus, the first parallel corpus of French and English [13]. It included 50 million words collected from transcriptions of the Canadian Parliament debates between 1975 and 1988 [13]. Since then, many parallel corpora projects began to be compiled. The European Corpus Initiative (ECI) contains about 19 million words from French, English, and Spanish texts; the English-Norwegian Parallel Corpus (ENPC)joined the German-Norwegian, French-Norwegian, and Russian and Norwegian parallel corpora to form the Oslo Multilingual Corpus, etc.[14]. The Open Parallel Corpus (OPUS) consists of nearly 352 million tokens in many European and Asian languages including Arabic.

The list of parallel corpora that include English is too long to cover here. This is not the case for Arabic however.

2. 2 Arabic Parallel Corpora

Arabic parallel corpora began to come into existence only in the late 1990's when the English-Arabic Parallel Egypt Corpus was developed at John Hopkins University in 1999for the purpose of facilitating machine translation [15]. It consisted of the Qur'an in English and Arabic. Then in 2004, the English-Arabic Parallel Corpus was compiled by Al-Ajmi. It contained three million words that were collected from the Kuwaiti World of Knowledge book series of translated books about a variety of topics in history, economics, arts, science, and literature. There is also the Open Parallel Corpus OPUS (2004), which is a large-scale multilingual parallel corpus that contains translated documents in 60 languages including Arabic [16]. Furthermore, Samy, et al.(2004) built the parallel Spanish-Arabic corpus from the annual reports of different United Nations institutions for the purpose of experimenting with alignment at sentence level. Samy, et al. reused available tools for the Spanish language on Arabic texts. One year later, Samy, et al. (2006) added English texts from the United Nations documents to the previous corpus and developed the Arabic-Spanish-English multilingual corpus. Dukes and Habash (2010) compiled The Quranic Arabic Corpus[1], which received prominence in "Morphological Annotation of Quranic Arabic". Salhi (2013) compiled the English-Arabic Parallel Corpus of United Nations Texts (EAPCOUNT)[18]. Al Kahtani (2015) constructed what he labeled as 'free high quality' parallel Arabic-English corpus for the purpose of improving the performance of machine translation (MT). Finally, LDC[2] developed several English Arabic parallel corpora from broadcast conversation (e.g., talk shows), broadcast news, and news wires, which amount to around 40 million words.

Although these may appear like numerous parallel corpora, they are either proprietary experimental, or restricted to specific text types. They are also aligned at either paragraph or sentence level but not at phrase and word levels. Most of them are not POS-tagged. There is a clear need for annotated Arabic resources that translators, learners, educators, researchers, and language engineers can use free of charge. To fill this gap, the present study developed a freely available English Arabic political parallel corpus (EAPPC) that contains around one million words culled from His Majesty King Abudillah II's writings and speeches. To the best of our knowledge, this is the first work that aligns English-Arabic parallel texts at multiple levels (i.e., sentence, clause, phrase, and word levels). To make this resource even more valuable, the corpus texts have not only been stemmed and POS-tagged automatically but also manually verified.

3. BUILDING THE CORPUS

3. 1 Data Selection

A preliminary survey of Arabic texts on the World Wide Web (WWW) was conducted to identify the kinds of existing Arabic texts that were translated into English and English texts that were translated into Arabic. The survey results showed that there are different types of these texts such as UN documents, news (e.g., Petra News Agency and the British Broadcasting Corporation BBC's news texts), novels (e.g., Najib Mahfouth, Ghassan Kanafani, and Agatha Cristy's novels), books (e.g., historical and scientific books). However, most of these texts were not freely available. Furthermore, some of them were available in a small data size and others were available in a Portable Document Format (PDF) that cannot be turned into machine-readable text.

In order to obtain high quality textual material for the present corpus, we considered the following selection criteria:

1. Arabic data should be in Modern Standard Arabic (MSA) and must have an English translation.
2. English data should be in Standard English regardless of geographic origin and must have a translation in Modern Standard Arabic.
3. The translation must be of good quality.

[1] http://corpus.quran.com/

[2] https://catalog.ldc.upenn.edu/

4. The data should be available in machine-readable textual format.
5. The data should be representative of MSA in general or of a particular MSA genre.
6. Texts had not been used in previous parallel corpora.
7. The copyright must permit corpus compilation.

Fortunately, the required data were found in His Majesty's official website (http://www.kingabdullah.jo/). His speeches, interviews, and letters met the selection criteria. However, we decided to add His Majesty's book, *Our Last Best Chance: The Pursuit of Peace in a Time of Peril (2011)* despite the fact that it was not available in electronic format. This was done for two reasons: first, to enlarge the corpus size; second, to shed light on the real problems that are often encountered by corpus compilers when they deal with text non-availability. The nature of the data defined the type of parallel corpus to compile. Thus, this corpus turned to be of political nature.

3. 2 Data Description

The present corpus consists of 351 documents that fall into four categories: speeches, interviews, letters, and books. Tables 1, 2, and 3 below describe these categories:

Table 1: The data extracted from His Majesty's speeches, interviews, and letters

Text	Speeches	Interviews	letters
Time range	1999-2015	1999-2015	1999-2015
Total number	269	206	145
Translated	189	80	68
Trans. from Eng.	131	45	0
Trans. from Ar.	58	35	68
Untranslated	80	126	77
Translators	RoyalCourt	RoyalCourt	RoyalCourt
Length (words)	250-2350	308-6403	200-3050
English tokens	176,907	178,748	45,385
Arabic tokens	159,850	138,357	39,730

Table 2: TheKing's book (English version)

English title	*Our Last Best Chance: The Pursuit of Peace in a Time of Peril*
Publisher	Viking Press
Date of publication	2011
Place of publication	New York
No. of chapters	27
Chapter length	2300-8194
English words	109491

Table3: TheKing's book (Arabic version)

English title	فرصتنا الأخيرة: السعي نحو السلام في وقت الخطر
Publisher	Daralsaqi
Date of publication	2011
Place of publication	Beirut
Translator	Shukri Rahim
No. of chapters	27
Chapter length	2300-8194
Arabic words	107634

3. 3 Data Extraction

After obtaining permission from the Royal Hashemite Court to use the King's speeches and writings for non-commercial purposes, Arabic and English data were extracted from the official website of His Majesty King Abdullah II and from His book.

The texts on the website were saved in Microsoft Word document format. The English version of the book was scanned and submitted to an optical character recognition (OCR) process to convert picture to electronic text (e-text). This process is relatively fast, yet its product is not without inaccuracies. Therefore, the e-text had to be manually checked and edited. OCR was only possible for the English version of the book but not for the Arabic one. Available OCR tools were not capable of satisfactorily converting the Arabic version of the King's book into e-texts. Thus, the Arabic version of the book was manually retyped, checked, and saved in an e-text format to render it ready for machine processing.

3. 4 Data Processing

Data processing involved six stages: metadata annotation, text segmentation, tokenization, alignment, stemming, and POS tagging.

3. 4. 1 Meta Data Annotation

Metadata include text title, author, year, era, category, occasion, region or place, source language, and text translator. All texts in the corpus were annotated with these nine metatags.

3. 4. 2 Text Segmentation

In this corpus, sentence boundaries in ST and TT represent a major challenge for the segmentation process. This is because the boundaries do not always correspond. The relations between text segments and their translations are not always in one to one correspondence. One sentence in one language might be translated into one sentence or two sentences; or two sentences might be translated into one. Besides, there are many examples where a sentence corresponds to a clause or to a phrase. Therefore, text segmentation is manually carried out at sentence, clause, and phrase levels. This is done in order to obtain the best matching between ST and TT segments. Each segment is thereafter referred to as a corpus line.

3. 4. 3 Tokenization

Tokenization is carried out manually. Word boundary identification during the tokenization process is also a challenge. In many cases, two or more words are kept together as one phrase. This is because a single word in Arabic might correspond to a phrase in English (e.g.,سَتُوَفِّرُ corresponds to will save). On the other hand, a single word in English might correspond to two or more words in Arabic (e.g., cousins corresponds to أبناء العمومة).

This stage also involves manual omitting of all punctuation marks and involves separating the

conjunction particle و (and) from the words it is conjoined to.

3. 4. 4 Alignment

Accurate alignment is crucial for extracting information out of a parallel corpus. It enables users to easily and swiftly find equivalents of search terms or phrases. For example, when the user types a search term in one language, the concordance displays all occurrences of this word in that language. It also displays all the aligned equivalents in the target language. The results can be then extracted and analyzed [20]

To automatically align the corpus texts, SDL Trados WinAlign 2011 was used as an alignment tool. However, the output was unsatisfactory. Furthermore, Win Align altered the predesignated text segmentation.

Given the particular importance of alignment to the parallel corpus, it is regarded vital to be carried out manually. The text alignment is carried out at sentence, clause, phrase, and word levels. However, it is not a straightforward process either. In fact, translation strategies, such as omission and lengthy explanation affect the alignment process. For example, one of the main challenges is the numerous cases of omission, where an ST line has no equivalent in the TT. Another challenge emerges when a TT line contains information that explains a SL term, which is usually unknown to the target reader, and has no equivalent line in the ST. In such cases, corpus lines are either deleted or attached to a neighboring line.

Alignment at word level was more complex. This is due to word order differences between Arabic and English. Words with no equivalents were left unaligned.

3. 4. 5 Stemming

The Arabic light stems and roots are automatically extracted using Khoja's stemmer. Its accuracy rate was reported to be 96% [15]. English words were stemmed using the Porter stemmer.

3. 4. 6 Part of Speech Tagging

POS is of paramount concern to the linguist who wants to know how language works, how it is structured, and how it is used. Hence, corpora are often annotated with POS.

In the present corpus, Arabic words were automatically POS tagged using the Khoja POS tagger. However, the outcome was unsatisfactory as its accuracy rate is reported to be 86% [21]. It skipped words, wrongly tagged a noun as a verb and a verb as a noun. It also offered only two tags viz. verb and noun, so the particles in the corpus were tagged either as verbs or as nouns. For example, the noun أُسُسْ (foundations) was tagged as a verb and the verb أَسْمَعُ (hear) was tagged as a noun. Furthermore, the particles من (from) and في (in) were tagged as a noun and a verb respectively.

Arabic POS tagging is challenging primarily because most texts on the net are not fully vowelized. Thus, many words can be read in different ways and can be assigned different POS tags. For example, the word تطور in تطور المجتمع المدني can be read in two ways. It can be read as a verb تَطَوَّرَ 'developed' or as a noun تَطَوُّرُ'development'.

Due to the critical importance of POS annotation, it was deemed necessary that it be carried out manually as well. The POS tags deemed essential were: Noun, Verb, and Particle. We had to go through close to half a million words deciding their part of speech. It was necessary that the POS tags reflect the category of a word in its context of use. We had to read all the King's speeches, letters, interviews, and book and label each word with its appropriate POS tags. This is a valuable feature of our corpus as its POS annotation is context-based. The example in the previous paragraph, for instance, was tagged as noun because the context portrayed it so.

4. CORPUS EXPERIMENTATION

Parallel corpora can be used to study different aspects of language such as the features of source and target languages, the influence of SL on TL, the translation strategies used, and the ideology and style of individual translators [22-25]. Furthermore, translators may learn strategies and use them in their translation tasks [26]. In order to demonstrate the EAPPC utility in translation, the corpus was used to investigate how different Royal Court translators rendered the term تكفيري (takfiri) in the speeches, interviews, letters and book of His Majesty King Abdullah II.

Translating political texts has been labeled a complex activity [27-29]. Translators of such texts attempt to maintain the ideological and cultural aspects of the ST during the translation process [27, 29]. Hence, translators need to use the translation strategies and techniques that would enable them to preserve the ideology of the SL text and to cope with translation problems that surface during the translation process. [29, 30]. One of the problematic issues that often encounter translators is the translation of culture-specific concepts. This emerges when a given concept is either unknown or known but is unlexicalized in TL. Baker (1992:26-42) listed eight strategies used by translators to deal with such challenges:

1. Translating the unlexicalized concept by using a hypernym.
2. By using a more neutral/ less expressive term.
3. By cultural substitution.
4. By using a loan word or a loan word plus explanation.
5. By paraphrase using a related term.
6. By illustration and exemplification.
7. By omission of this unlexicalized concept.

4.1. Methodology

This research uses EAPPC, classical and modern Arabic dictionaries, *such as, Al-a'ayn (786* CE*), Mu'jammaqāyīs al-lughah*(1004 CE), *al-mufradātfīgharīb al-Qur'ān*(1109 CE), *Lisān al-'Arab*(1311 CE) and*Mu'jam al-Lughah al-'Arabīyah al-Mu'aasira* (2003 CE)[3], the Corpus of Contemporary American (COCA)[4], the British National Corpus (BNC)[5], and the News on the Web Corpus (NOW)[6].

First the Arabic dictionaries were consulted in order to determine the meaning of the word تكفيري (takfiri). Next, the EAPPC corpus was explored using the Arabic root (كفر)asa query term. The parallel corpus concordance displayed, in the KWIC style, all occurrences of the term along with their translations. These data were then exported to an Excel spreadsheet and analyzed. After that, COCA, BNC, and NOW were explored for instances of the wordsتكفير (takfir) andتكفيري (takfiri) in English texts. Finally, the study used Baker's taxonomy of translation strategies to analyze the data and to discover the adopted translation strategies.

4.2. Findings and Discussion

The term تكفير (takfir) is an abstract noun derived from the verb of intensification, kaffara, and takfiri تكفيري is a relative noun derived from تكفير (takfir). This word has multiple senses as summarized in Table 5below:

Table 5: Senses of the termتكفير (takfir)

Sense	Meaning in English	Dictionary
إيْمَاءُ الذِّمِّيّ برأسِه	nodding	Al-a'ayn
تَتْوِيْجُ المَلِكِ بتاجٍ	enthronement	Al-a'ayn
الذُّلّ والخضوع	abasement, submissiveness	Lisān al-'Arab
الانحناءُ الشَّدِيد	bowing	Lisān al-'Arab
سترالشيء وتغطيته	to shield, cover	Al-mufradātfīgharīb al-Qur'ān
أن يَتَكَفَّرَ المُحارِبُ في سلاحه	cover the body with weapons	Lisān al-'Arab
تُكَفِّرَ الخطيئة أي تمحوها	expiation	Lisān al-'Arab
التَّكْفير: أَن يَخْضَعَ الإِنسانُ لغَيْرِه ويَخْنَي ويُطَأْطِئَ رَأْسَه قَرِيبا من الرُّكوع،	to offer obeisance	Lisān al-'Arab
جماعة **تكفيرية**: جماعة متشدّدة تنسب العصاة والمذنبين إلى الكُفر، أو عدم الإيمان بالله، أو الزّندقة	takfiri group: extremists who call people apostates.	Mu'jam al-Lughah al-'Arabīyah al-Mu'aasirah

The majority of these senses have changed over time and only one sense has survived. Table 5shows تكفير (takfir) to have had the meanings of abasement, submissiveness, obeisance, wearing armor, expiation of sins, nodding, and enthronement. Only expiation, however, has survived the ravages of time. Mo3jam Al-Lugha Al-3arabiya Al-Mo3asira by Ahmad Mukhtar Omar reflects our modern conception of تكفيري (takfiri) as the "attribute of ascribing apostasy to others".

Using EAPPC, seventy-seven instances of derivatives of the root (كفر) have been used by His Majesty. Their distribution in the corpus is shown in Figure 1 below.

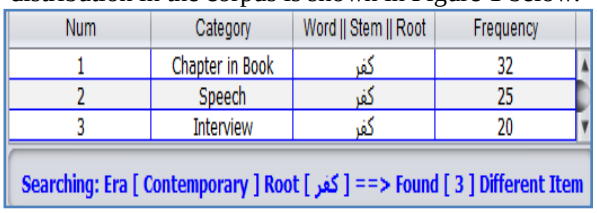

Figure 1: Search Results of the Root كفر (kfr) in the Parallel Corpus

Furthermore, Table 6 shows the distribution of the derivatives of the root كفر (kfr) in the corpus.

Table 6: Derivatives of the root كفر (kfr) in the parallel corpus

Derivative	Speeches	Interviews	Letters	Book	Frequency
تكفير	21	6	0	3	30
تكفيريين	0	3	0	14	17
تكفيريون	0	0	0	10	10
تكفيري	1	8	0	0	9
تكفيرية	0	0	0	4	4
يُكَفِّر	2	0	0	0	2
يُكَفِّرون	0	2	0	0	2
كافِر	0	1	0	0	1
كَفْر	1	0	0	0	1
كُفّار	0	0	0	1	1
Total	25	20	0	32	77

Table 6 indicates that the word تكفير (takfir) mainly occurs in His Majesty's speeches. The word تكفيري (takfiri) mostly occurs in his interviews, while its feminine form تكفيرية, plural form in the nominative case تكفيريون, and plural form in the accusative case تكفيريين most frequently occur in the King's book. All the corpus instances reflect our modern usage of this term.

In order to examine how each instance of the terms تكفير (takfir) and تكفيري (takfiri) were rendered in Arabic and English, it is important to discover the translation strategies that were used by the King's translators.

The terms تكفير (takfir) and تكفيري (takfiri) were both used in the English ST despite the fact that they are culture-specific terms (see Table 7).

Table 7: Examples of the term تكفير (takfir) and تكفيري (takfiri) in His Majesty's English STs

Source	English Text	Arabic Text
Interview 25	and basically, takfir ideology if you don't agree with me, I have the right to kill you,	وأيديولوجية التكفير، بشكل أساسي، مفادها أنه إذا لم تتفق معي فلي الحق في أن أقتلك،
Interview 25	In my discussions with the Muslim Brotherhood here, I don't believe that the majority of them are takfiris,	وفي مناقشاتي مع الإخوان المسلمين هنا، لا أعتقد أن غالبيتهم تنادي بالفكر التكفيري،

The terms takfir and takfiri are used in English source texts as loan words. To prove that, we searched COCA, BNC, and NOW corpora for instances of these terms and found 26 instances per million of the word takfir in COCA, three instances per million in BNC, and 141

3http://lisaan.net
4http://corpus.byu.edu/coca/
5http://corpus.byu.edu/bnc/
6http://corpus.byu.edu/now/

instances per million in NOW. There are for the word takfiri, on the other hand, 11 instances in COCA and 357 instances in NOW, but no instances in BNC. The contexts for most of these instances are the news and political and academic texts. Despite the very small rate of occurrence of takfir and takfiri in COCA, BNC, and NOW, there is support for the conclusion that these terms are used in English source texts as loan words.

The term تكفير (takfir) in the Arabic SL component of the EAPPC occurs three times in one interview and 14 times in five speeches. In the English SL component of this corpus, the term occurs three times in one interview, six times in six speeches, and three times in one chapter in the book.

EAPPC evidence shows that translators adopted these strategies when rendering تكفير (takfir) from Arabic into English: The use of loan words, loan words plus explanation, English equivalents, and English equivalents with the TL terms between brackets.

In many instances, translators would introduce the loan word takfir with an explanation (e.g., calling others apostates) and then use it without explanation in subsequent occurrences, as shown in the following example from an interview given by His Majesty on22 April 2006 to *Al Sabah Al Jadid Newspaper:*

ST (Arabic):

"كما حظي بتوافق إجماعي يدين ممارسات **التكفير** التي يلجأ إليها المتطرفون لتبرير العنف. ولأننا نقف ضد التطرف **والتكفير**، فقد أصبحنا مستهدفين من الجماعات الإرهابية في العراق".

TT (English)

"This declaration condemned [the practice of] **takfir [calling others apostates]** that extremists use to justify violence. And because we stand against extremism and **takfir**, we have become targets of terrorist groups in Iraq,"

Another strategy for rendering تكفير (takfir) from Arabic into English is translation by TL equivalents, as illustrated in the following examples from His Majesty's speech at *the opening session of the International Islamic Conference*on4 July 2005:

ST (Arabic):

"وعدم جواز **تكفير** أي مسلم من أتباعها."

TT (English):

"and that **declaring any one of them an apostate** is unacceptable.

Another strategy is using a TL equivalent with the loan word between brackets. For example, translators paraphrased تكفير (takfir) as *apostasy* and used the loan word *takfir* between brackets, as demonstrated in the following example from His Majesty's speech at the opening of the third extraordinary session of the *Islamic Summit*on7 December 2005:

ST (Arabic):

"لأن عدم الاتفاق على هاتين المسألتين هو سبب الفرقة والاختلاف وتبادل تهم **التكفير**والاقتتال بين أبناء الدين الواحد"

TT (English):

"The absence of consensus on these two issues has led to divisions and differences, accusations of **apostasy (takfir)** and internecine fighting.".

A related term is *takfiri* which is the adjectival form of *takfir*. The corpus offers 40 such instances. Twenty eight of them occur in His Majesty's book, and one in an English interview. In the subcorpus of Arabic STs, on the other hand, *takfiri* occurs 10 times in four interviews and once in a speech.

The word *takfiri* in English SL texts was rendered as تكفيري (takfiri), تكفيرية (takfiriyah), تكفيريين (takfiriyeen), or (takfiriyoon) تكفيريون in Arabic in accordance with the requirements of syntactic inflection.

In some instances, the word *takfiri* is not found in the SL text but the translator understood that it was intended. In such a case, the translator made it explicit by using *takfiri* in the TL text, as shown in the following example from His Majesty's book, *Our Last Best Chance (2011):*

- **ST (English):**

"and we helped the Americans understand **what** to look for"

- **TT (Arabic):**

"وقد ساعدنا الأمريكيين في التعرف على **التكفيريين**."

'and we helped the Americans recognize **takfiris'**

In other cases, the author referred to the word *takfiri* by using an anaphoric pronoun. In this case, the translators explicitly used the equivalent word تكفيري (takfiri), as illustrated in the following example from His Majesty's book, *Our Last Best Chance(2011):*

ST (English):

"Islam celebrates life; **they** seek to destroy it."

TT (Arabic):

"فإذ يحترم الإسلام الحياة الإنسانية ويصونها، لا يتردد **التكفيريون** في تدميرها والقضاء عليها."

'Even though Islam respects and protects human life, **takfiris** do not hesitate to destroy it and quell it'

Although the term تكفير (takfir) has been translated into English using multiple strategies, تكفيري (takfiri) has only been rendered using the loan word strategy, as shown in His Majesty's interview on 22 April 2006 given to *Al Sabah Al Jadid Newspaper*:

ST (Arabic):

"وجد الفكر **التكفيري** ما يغذي أهدافه البعيدة كل البعد عن قيم الإسلام الحقيقية"

TT (English):

"takfiri thought found feeding ground for its aims that are alien to true Islamic ethics and values."

To sum up, translators tended to render the terms takfir and takfiri from English into Arabic by using the same terms as they are Arabic in the first place. On the other hand, when they translated into English they employed several strategies: translation using loan words, loan words plus explanation, translation by TL equivalence, and translation by equivalents with the loan word between brackets.

5. CONCLUSION

This study has described the construction of EAPPC. The ultimate aim of EAPPC is to provide translators, learners, educators, researchers, and language engineers with a freely available tagged parallel corpus whose annotation has been manually verified.

To illustrate its utility, we have carried out an experiment that examined the translation strategies

used in rendering a culture-specific term. The results demonstrated the ease with which knowledge about translation strategies can be gained from this parallel corpus.

6. REFERENCES

[1] J. Sinclair, *Corpus, concordance, collocation*: Oxford University Press, 1991.

[2] M. Baker, "Corpora in translation studies: An overview and some suggestions for future research," *Target*, vol. 7, pp. 223-243, 1995.

[3] D. Biber, S. Conrad, and R. Reppen, *Corpus linguistics: Investigating language structure and use*: Cambridge University Press, 1998.

[4] L. Bowker, "Towards a methodology for a corpus-based approach to translation evaluation," *Meta: Journal des traducteursMeta:/Translators' Journal*, vol. 46, pp. 345-364, 2001.

[5] S. Hunston, *Corpora in applied linguistics*. Cambridge: Cambridge University Press, 2002.

[6] F. Zanettin, S. Bernardini, and D. Stewart, *Corpora in translator education*. London, NY: Routledge, 2014.

[7] K. Aijmer, B. Altenberg, and M. Johansson, *Languages in contrast: papers from a symposium on text-based cross-linguistic studies, Lund 4-5 March 1994* vol. 88: Lund studies in English, 1996.

[8] D. Samy, A. Moreno Sandoval, and J. M. Guirao, "An Alignment Experiment of a Spanish-Arabic Parallel Corpus," in *Proceedings of the International Conference on Arabic Language Resources and Tools (NEMLAR 2004)*, 2004, pp. 85-89.

[9] D. Samy, A. M. Sandoval, J. M. Guirao, and E. Alfonseca, "Building a Parallel Multilingual Corpus (Arabic-Spanish-English)," in *Proceedings of the 5th Intl. Conf. on Language Resources and Evaluations, LREC*, 2006.

[10] M. Tadić, "Building the Croatian-English parallel corpus," in *The 5th International Conference on Language Resources and Evaluation (LREC'2006), pp.523-530, Athens, Greece.*, 2000.

[11] S. Singh, T. McEnery, and P. Baker, "Building a parallel corpus of English/Panjabi," in *Parallel Text Processing*, ed: Springer, 2000, pp. 335-346.

[12] L. Rura, W. Vandeweghe, and M. Montero Perez, "Designing a parallel corpus as a multifunctional translator's aid," in *XVIII FIT World Congress= XVIIIe Congrès mondial de la FIT*, 2008.

[13] J. Véronis, "From the Rosetta stone to the information society," in *Parallel Text Processing*, ed: Springer, 2000, pp. 1-24.

[14] H. Hasselgård, "Contrastive analysis / contrastive linguistics," in *The Routledge Linguistics Encyclopedia*, K. Malmkjær, Ed., Third Edition ed London, NY: Routledge, 2010, pp. 98-101.

[15] M. S. S. Sawalha, "Open-source resources and standards for Arabic word structure analysis: Fine grained morphological analysis of Arabic text corpora," PhD Thesis, University of Leeds, 2011.

[16] J. Tiedemann and L. Nygaard, "The OPUS Corpus-Parallel & Free," in *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04).* Lisbon, Portugal, 2004.

[17] K. Dukes and N. Habash, "Morphological Annotation of Quranic Arabic," in *LREC'10*, Valletta, Malta, 2010.

[18] H. Salhi, "Investigating the Complementary Polysemy and the Arabic Translations of the Noun Destruction in EAPCOUNT," *Meta: Journal des traducteursMeta:/Translators' Journal*, vol. 58, pp. 227-246, 2013.

[19] A. Hussein, *Our Last Best Chance. The Pursuit of Peace in a Time of Peril*. London, UK: Penguin, 2011.

[20] M. Ghadessy, A. Henry, and R. L. Roseberry, *Small corpus studies and ELT: theory and practice* vol. 5. Philadelphia, USA: John Benjamins Publishing Co., 2001.

[21] L. Al-Sulaiti and E. Atwell, "Designing and developing a corpus of contemporary Arabic," MA Thesis, School of Computing, University of Leeds, UK, 2004.

[22] M. Baker, "Corpus linguistics and translation studies: Implications and applications," in *Text and technology: In honour of John Sinclair*, M. Baker, G. Francis, and E. Tognini Bonelli, Eds., ed Philadelphia, USA: John Benjamins Publishing Co., 1993.

[23] S. Laviosa, *Corpus-based translation studies: theory, findings, applications*. Amsterdam – New York, NY: Rodopi B.V, 2002.

[24] M. Olohan, *Introducing corpora in translation studies*. London, NY: Routledge, 2004.

[25] C. Fantinuoli and F. Zanettin, *New directions in corpus-based translation studies*. Berlin: Language Science Press, 2015.

[26] G. Shen, "Corpus-based Approaches to Translation Studies," *Cross-Cultural Communication*, vol. 6, pp. 181-187, 2011.

[27] A. Shunnaq, "Arabic-English translation of political speeches," *Perspectives: Studies in Translatology*, vol. 8, pp. 207-228, 2000.

[28] G. Quentel, "Translating a Crucial Political Speech," ed: Retrieved October, 2006.

[29] K. Sárosi-Márdirosz, "Problems related to the translation of political texts," *Acta Universitatis Sapientiae Philologica*, pp. 159-180, 2014.

[30] M. Baker, *In Other Words. A coursebook on translation*. London, NY: Routledge, 1992.