



Translating Definiteness and Indefiniteness: A Corpus-based Study of Malay Translational Equivalences of English Articles

Imran Ho-Abdullah*, Azhar Jaludin and Intan Safinaz Zainudin

School of Language Studies and Linguistics, Faculty of Social Sciences and Humanities, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

ABSTRACT

The use of parallel corpora and the methodology of corpus linguistics have yielded innovative insights in translation studies and cross linguistics studies. The tools and methodology of corpus linguistics have made it possible to systematically explore and capture enormous quantities of text and their translation effectively and efficiently, hence providing researchers with more accurate data for analysis. This paper builds on this tradition and investigates the translation equivalents of English articles *a(n)* and *the* into Malay utilising a 250,000 words unidirectional parallel corpus of English–Malay. In particular, the distributions of the translational equivalents of English articles *a(n)* and *the* are examined in relation to the functions of the articles in terms of the definiteness – indefiniteness referencing system of article usage in English. A total of 1355 instances of the definite article (*the*) and 407 instances of the indefinite article (*a/an*) were detected and their translational equivalents in Malay identified using ParaConc. The results of the analysis of the Malay translation reveal that due to cross linguistics differences in the article system, 71% of the definite article and 56% of the indefinite articles are translated as Ø. In addition, the corpus analysis also reveals that there are shared translational equivalents that are used to indicate both definite and indefiniteness as well as unique translational equivalents that are used only for definiteness or indefiniteness.

Keywords: Corpus linguistics, definite-indefinite articles, English-Malay translation, parallel corpus, translational equivalents

ARTICLE INFO

Article history:

Received: 13 September 2012

Accepted: 15 March 2013

E-mail addresses:

imranho@ukm.my (Imran Ho-Abdullah), azharj@ukm.my (Azhar Jaludin), intansz@ukm.my (Intan Safinaz Zainudin)

* Corresponding author

INTRODUCTION

Over the last two decades, the use of parallel corpora and the methodology of corpus linguistics have developed dramatically in translation and cross-linguistic studies.

Baker (1993, p. 243) asserts that the availability of large corpora of both original and translated texts, along with the development of a corpus-based approach, would help translation scholars unearth the nature of translated text. Since then, scholars in the field of translation studies, as well as corpus linguistics, have started to recognize the corpus-based approach as a vital and linguistically effective way within which translation and translating can be systematized (Laviosa, 1998, p. 474). In addition, Baker (1999), Olohan (2004) and Frankenberg-Garcia (2009), among others, have argued that corpus linguistics in terms of both the tools and methodology has given an opportunity for a systematic exploration and compilation of enormous quantities of text and their translation effectively and efficiently, hence providing researchers with more accurate data for analysis. In addition to this evolution of practice, Laviosa (2003) remarks that not only has the study of corpora become fully integrated into Translation Studies since the early 90s, but most importantly, corpus research has grown so rapidly and has influenced so significantly the way in which we conceptualize, study and teach translation. Recently, corpus-based translation studies emerged as a key paradigm that has transformed analysis within the discipline of translation studies. It can be defined as the use of corpus linguistic technologies to demonstrate and exhibit the translation process, something that is accessible in the advent of new millennium and the on-going advances in computer technology. Kruger (2002, p. 70) shows how translation studies can benefit when corpus

linguistics methodology that involves the analysis of large corpora of authentic running texts is integrated by the means of computer software.

The present paper utilizes a computerized parallel corpus to investigate the translation of the English articles into Malay, that is, it proposes a corpus-based study of translational equivalents of English articles *a/ (an)* and *the* into Malay utilising a 250,000 words unidirectional parallel corpus of English – Malay. More specifically, the main purpose of this paper is to show how the translational equivalents are examined in relation to the functions of the articles in terms of the definite–indefinite referencing system of article usage in English. By using the parallel corpus in the current paper, we shall be able to generate and uncover the distribution of Malay translation equivalents for the English articles.

REVIEW OF RELATED LITERATURE

Articles are considered as a subclass of determiners in English. Determiners are words which specify the range of reference of a noun in various ways, such as by making it definite (*the* book), indefinite (*a* book) or by indicating quantity (*many* books). In English, determiners always precede the noun they determine. Articles come under the category called central determiners, which are the most important in the determiner category. The articles can be clearly shown in the following diagram (Fig. 1, as cited in Selvan, 2010, p. 3).

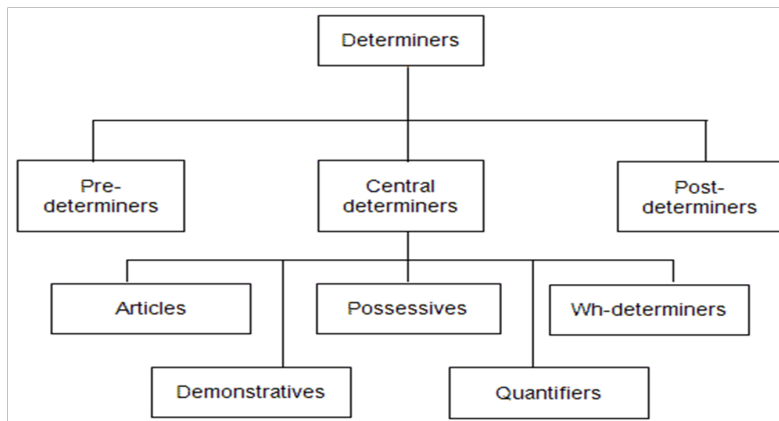


Fig.1: The English Determiners

In the English article system, there are two articles, namely, the definite and the indefinite. Broadly speaking, the indefinite article ‘a’ is used before a singular countable noun beginning with a consonant sound whereas the other indefinite ‘an’ is used before a singular countable beginning with a vowel sound. The indefinite article cannot be used before uncountable nouns as well as plural nouns. The second article of the English system of articles is the definite article ‘the’, which is used before the nouns that are definite and specific that refers to shared knowledge of the situation or contexts and nouns that refer to unique objects. It is also used before unique adjectives, superlative additives and nationality nouns. It is also used to refer to systems of communication, mass media and means of transportation when ‘by’ is excluded. Furthermore, two different systems of article use have been set up, depending on the type of reference in English, i.e., definite and indefinite. While definite reference points out to a particular object, indefinite reference designates an

unknown and unspecified object being referred to. According to Heim (1991), ‘the’ is specified [+definite], and ‘a’ is underspecified for definiteness. This means that ‘the’ can be used only when the conditions on definiteness have been satisfied, while ‘a’ has no such requirements. As a result, whenever the conditions on definiteness have been met, ‘the’ is used. If these conditions have not been fulfilled, ‘the’ cannot be used, and ‘a’ or *zero* in the case of plurals and mass nouns is used instead. An illustration of the use of ‘the’ and ‘a’ is shown as in the following sentence:

1. I saw a dog. I gave the dog some meat.

Celce-Murcia and Larsen-Freeman (1998, pp. 197-180) observe that the use of articles in English to convey generic meaning is a matter that is not very well treated in reference grammars remarking that Quirk and Greenbaum (1973), for instance, provide the following examples and assert that all these patterns reveal generic meaning:

2. The German or A German is a good musician.
3. The German or Germans are good musician.

The above examples deal with countable nouns being used generically. However, mass nouns can be used to express generic meaning and when they are, no article is being used, as shown in the following example: water is essential for life (p. 181).

In contrast, Malay does not have a system of articles but instead rely on various demonstrative determiners. According to Marsden (as cited in Chan, 1996), the Malay word *itu* can be used as a transitional equivalent to the definite article 'the' and the Malay pronouns are either demonstrative or definitive. However, he does not explain what he meant by those two terms. Likewise, in Winstedt's Malay Grammar (as cited in Chan, 1996), *itu* is seen as being equivalent to "the, that, those" with the sense of 'the distant, remote in space and time, the aforesaid.' Although Winstedt claims that there are no 'articles' in Malay, he asserts that there are a few exceptions in *itu* and *yang*. These are demonstrative pronouns that seem to have the quality of the article and when they are required to indicate particular objects, use is made of these demonstrative pronouns. However, in a translated text of Gonda by Hussein and Siahaan (as cited in Chan, 1996), the strong statement by Gonda about the non-existence of an article in Malay is found. For him, the morpheme *itu* below is not an article.

4. *Perempuan itu lemah badannya.*
'That woman is physically weak.'

The subject *perempuan itu* in the Malay sentence above refers to a definite person. The definite perspective is indicated through the use of *itu*, which is akin to the demonstrative pronoun *that*.

Khazryati *et al.* (2006) note that, at times, the demonstrative adjectives *itu*, *ini* are equivalents to the definite article that are in the sense the thing that the noun refers is familiar. As pronouns, these markers precede the noun. According to Nik Safiah Karim *et al.* (2004, pp. 349–361), there are two types of determiners in the Malay language, pre-determiners and post-determiners. A pre-determiner, on the one hand, occurs in front of nouns, and it can consist of numerals followed by classifiers, as shown in (5) and (6) below (as cited in Wong & Quek, 2007, pp. 214–215).

5. *seorang askar* 'A/one soldier'
6. *sepuluh buah bangunan*
'Ten buildings'

Although not an equivalent, the use of *seorang* in (5) could be described as functioning almost like the indefinite article *a*. The post determiners, on the other hand, are elements that occur after nouns, and consist of *itu* 'that' and *ini* 'this', as shown in (7) and (8).

7. *orang itu* 'That person'
8. *rumah ini* 'This house'

The following example shows the sequence of the pre-determiner with *itu*

and *ini* at the very end of the post-modifier structure, as in (9):

9. *baju tebal orang itu* “The person’s thick clothing”

The use of *itu* ‘that’ and *ini* ‘this’ in the above sentences demonstrates a notion of definiteness in the Malay NPs, although they are not equivalent to the English article *the*. Similarly, Cumming (1991, p. 23) remarks that in Malay, “there are no morphemes which are specialized for marking definiteness, specificity, or identifiability; but as in many languages, other resources may be co-opted for these purposes.” Thus, the numeral *satu/se* ‘one’ is used to indicate indefiniteness and *ini* ‘this’ and *itu* ‘that,’ the deictic pronouns/ demonstratives may be used to mark definiteness.

RESEARCH QUESTIONS

Bearing in mind the fact that Malay does not have a sub-class of determiners akin to the articles in English, the following research questions have been formulated in this paper:

1. What are the translational equivalents for the articles (*a*), (*an*) and (*the*) in Malay?
2. How are various translational equivalents distributed?

MATERIALS AND METHODS

This work is based on a unidirectional parallel corpus comprising of original English texts and their Malay translations. A ‘parallel corpus’ can be defined as a

‘collection of texts, each of which is translated into one or more other languages than the original’ (Sinclair, 1995, p. 32). Martínez Magaz (2006, p. 55) argues that ‘a parallel corpus is formed by original texts in language A and their translations (and pseudo-translations) in language B’. Hence, parallel corpora may be bilingual or multilingual, depending on how many target languages of the same source language text are involved. Others, such as Uzar (2002), have used the term ‘translation corpus’ to refer to such collection of texts. Meanwhile, Biel (2010) simplifies the relationship between these two terms by arguing that ‘a parallel corpus is a translation corpus in the strictest sense’. In addition, Frankenberg-Garcia (2009) draws a distinction between parallel corpora that are ‘unidirectional’ and ‘bidirectional’. The distinction captures the difference between a parallel corpus that contains source texts in one language (L1) and their translation into another language (L2). The single L1 > L2 alignment is regarded as a unidirectional parallel corpus. In contrast, a parallel corpus that contains source texts in both L1 and L2 as well as their translations in L2 and L1 respectively would be a bidirectional parallel corpus. In the case of the bidirectional parallel corpus, there will be two sets of alignments: L1 > L2 and L2 > L1. In other words, the bidirectional parallel corpus is a corpus containing ‘source texts in language A and the target texts in language B’, as well as the ‘source texts in language B and their translation into language A’ (Olohan, 2004, p. 24).

The utilization of the parallel corpus as a basis for translation studies has several advantages. The opportunity offered by a parallel corpus for the alignment and hence matching and mapping of the original texts and their translations are the key features of using a parallel corpus in translation studies (cf. Kenny, 2007). In turn, the alignment can allow for the processing and manipulation of the texts in a systematic, quantifiable and methodical way (through the use of software such as ParaConc) towards a more effective description of translation data (cf. Baker, 1995).

The unidirectional English-Malay parallel corpus is compiled by the 2011 Corpus Linguistics MA class at Universiti Kebangsaan Malaysia. The source texts are in English comprising of samples of the first five pages of five university Linguistics and Applied Linguistics textbooks. The target texts were identified from the translations of these textbooks published by Dewan

Bahasa and Pustaka. In total, a selection of 10 texts from five different textbooks was compiled for the corpus (5 English original and 5 Malay translations). The basic statistics of the parallel corpus is presented in Table 1. The number of words in the source text is 19,182. The translation in the target text is slightly smaller with 18,088 words. There are also more types in the source texts (3,151) as compared to the target texts (2,837). The type-token ratio is 16.5 in the source text but 15.84 in the target text. However, the number of sentences in the target text exceeded the number in the source text. The corpus was aligned manually at the sentence level. The accuracy at the sentence alignment level exceeded 99%. The 702 sentences in the source texts were translated into 705 sentences in the Malay target texts. This is a result of some long and complex sentences in English being translated into two or more sentences.

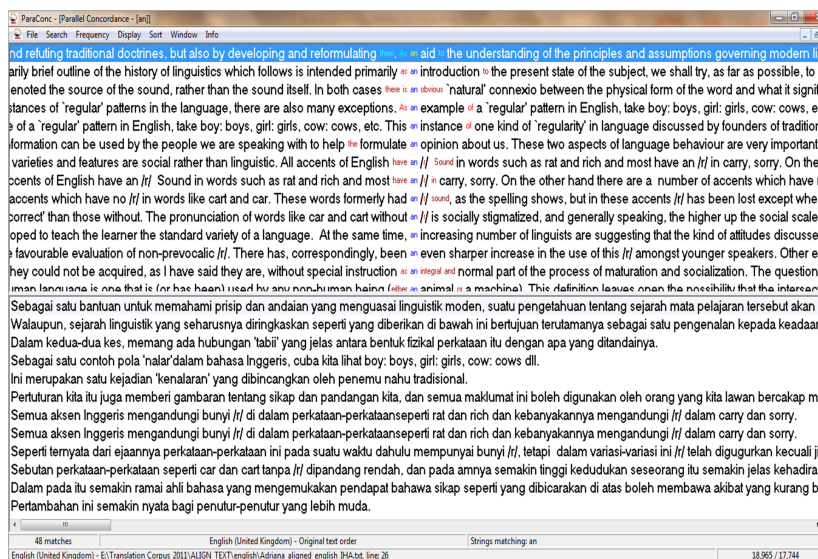


Fig.2: Search Results Using ParaConc (© 2001, 2003, Michael Barlow)

TABLE 1
Basic Statistics of the Unidirectional English-Malay Parallel Corpus

	Source (English)	Target (Malay)
Tokens	19,182	18,088
Types	3,151	2,937
Type token ratio	16.58	15.84
Sentences	702	705

Using the search function in ParaConc, the definite article ‘the’ and the indefinite articles ‘a/an’ and their translational equivalents in Malay were identified.

RESULTS

The results of the analysis of the translational equivalents of the English definite article ‘the’ are summarized in Table 2.

The most common Malay translational equivalents for the English definite article are *Ø*, *itu*, *-nya*, *ini* and *tersebut*. These five forms account for 88.7% of the data. Out of the 1355 occurrences of the definite article ‘the’ in English text, 966 (71.29%) were translated as *zero* in Malay. In all these instances, the NPs in the source texts in English are preceded by the definite article ‘the’ but no translational equivalents for the definite article (or any form of determiner) were found with the corresponding NPs in the Malay target texts. Meanwhile, only 9.0% (n = 122) of the English definite articles ‘the’ were translated into ‘*itu*’. An additional 3.6% (n = 49) of the total number of definite articles in the English text were translated as *-nya*; 3.4% (n = 46) as *ini* (*inilah*) in Malay, and 1.7% (n = 23) as *tersebut*. The occurrence of the numeral

‘one’ (*satu*), as a translational equivalent for the definite article, is also surprising since the use of the numeral ‘one’ as a NP pre-determiner normally indicates non-identifiability and hence indefiniteness. However, the frequency of ‘*satu*’ as the translational equivalent of ‘the’ is rather small, i.e. accounting for less than 1% of the total (n = 9).

TABLE 2
The Malay Translational Equivalents for the Definite Article ‘the’

Translational Equivalent	n	Percentage
<i>Ø</i>	966	71.29%
<i>itu</i>	122	9.00%
<i>-nya</i>	49	3.62%
<i>se-demikian (1) buah (2) orang (1)</i>	4	0.30%
<i>sese-orang (3)</i>	3	0.21%
<i>sesuatu</i>	1	0.07%
<i>satu</i>	9	0.66%
<i>ini (inilah)</i>	46	3.39%
<i>tersebut</i>	23	1.70%
<i>specific NP repeated</i>	6	0.44%
<i>tadi</i>	4	0.30%
<i>mereka</i>	1	0.07%
<i>tertentu</i>	1	0.07%
<i>yang</i>	1	0.07%
<i>berkenaan</i>	4	0.30%
<i>di sini</i>	1	0.07%
<i>NP not translated</i>	40	2.95%
<i>phrasal</i>	74	5.46%
Total	1355	100.00%

The results of the translational equivalents of the English indefinite article ‘a / an’ are summarized in Table 3. The total number of the indefinite articles ‘a

/ an' is 407 in the entire corpus. More than half of the occurrences (228 out of 407) of the indefinite article have the translational equivalents in Malay as *zero article*. Sixty-two out of the total number of these indefinite articles were translated using the numeral *satu* in Malay (15.23%), whereas twenty-six of them were translated using the numeral *se-* (6.39%) followed by a classifier. Twenty out of the total number of the indefinite articles of English involved were translated as *sesuatu*. Six out of the whole have their transitional equivalent in Malay as *satu-satu* (1.91%). Others are varied as shown in the table below with the percentage and the number from the actual results of the translation.

TABLE 3
The Malay Translational Equivalents for the Indefinite Articles 'a/an'

Translational Equivalents	n	Percentage
Ø	228	56.02%
<i>satu</i>	62	15.23%
<i>satu-satu</i>	6	1.47%
<i>sesuatu</i>	20	4.91%
<i>suatu</i>	9	2.21%
<i>se-</i> <i>sebahagian</i> (2) <i>sebatang</i> (1) <i>sebaya</i> (1) <i>sebilangan</i> (1) <i>sebuah</i> (3) <i>seekor</i> (1) <i>sejenis</i> (2) <i>seorang</i> (9) <i>sepanjang</i> (1) <i>seperangkat</i> (1) <i>sepintas</i> (1) <i>seratus</i> (1) <i>setahun</i> (1) <i>setiap</i> (1)	26	6.39%
<i>sese-buah</i> (1) <i>orang</i> (3)	4	0.98%
<i>mana-mana</i>	1	0.25%
<i>itu</i>	2	0.49%
<i>NP not translated</i>	15	3.69%
<i>phrasal</i>	34	8.35%
Total	407	100%

Our analysis revealed that the translation of the source texts (English) into the target texts has some common translational equivalents that are used for both the definite and indefinite articles of English. This is an interesting finding since the two articles in English (definite and indefinite) have different referencing functions. In more specific, the shared translational equivalents between the definite and indefinite articles of English are Ø, *itu*, *satu*, *se-*, *sese-* and *sesuatu*. The frequency of these shared translational equivalents is presented in Table 4.

TABLE 4
The Shared Translational Equivalents between the Definite and Indefinite Articles

Malay Translational Equivalents	Definiteness	Indefiniteness
Ø (<i>zero article</i>)	71.29%	56.02%
<i>satu</i>	0.66%	15.23%
<i>itu</i>	9%	0.49%
<i>se-</i>	0.30%	6.36%
<i>sese-</i>	0.21%	0.98%
<i>sesuatu</i>	0.7%	4.91%

Two translational equivalents, namely, the zero article and the demonstrative *itu*, are used more often to translate the definite than the indefinite. The zero article (Ø) is used as the translational equivalent for the definite article around 71.29% of the time as compared to only 56.02% for the indefinite articles. *Itu* is used as the translational equivalent for the definite article in 9% of the occurrence in contrast to 0.49% for the indefinite article. The other four translational equivalents, namely, *satu*, *se-*, *sese-* and *sesuatu* are used more frequently to translate the indefinite than the definite. The numeral '*satu*' (one) is used

as the translational equivalent 15.23% of the occurrences of the indefinite article compared to 0.66% for the definite article. Likewise, the numeral *se-*, followed by a classifier, is used 6.36% as the translational equivalent for the indefinite articles compared to 0.30% for the definite articles. Like *se-*, the Malay *sesuatu* and *sese-*, which are used as singular numeral adjectives, are used more often as the translational equivalents to indicate the indefinite (4.91% for *sesuatu* and 0.98% for *sese-*) in contrast to 0.70% and 0.21% respectively for definiteness.

There are also translational equivalents that are unique for the definite article ‘the’. In total, nine items (namely, *ini*, *tersebut*, *tadi*, *mereka*, *tertentu*, *yang*, *berkenaan*, *di sini*) and also the suffix *-nya* are used to translate the definite article. The suffix *-nya* functions in Malay as the third person possessive pronoun and hence can play the role of identifying and making definite the NP through co-referencing. In total, 49 (3.62%) of the definite articles in the source text were translated using the equivalent *-nya*. The demonstrative pronoun ‘*ini*’ is akin to the English ‘this’ and also serves to provide definiteness to the NP being translated. Meanwhile, 46 occurrences (3.39%) of the article ‘the’ were translated using ‘*ini*’. The items *tersebut* (above-mentioned or aforesaid); *tertentu* (specified or specific) and *berkenaan* (particular) are used as the translational equivalents of the definite article. These items function as adjectives in the NP construction to denote definiteness. However, these adjectives account for only 2% of the equivalents

used to translate the definite article. The distribution of the unique translational equivalents for the definite article ‘the’ is summarized in Table 5. Another translation strategy used to indicate definiteness is the repetition of the NP, which accounts for 0.44% of the data.

TABLE 5: Unique Malay Translational Equivalents for the Definite Article ‘the’

Translational Equivalent	n	Percentage
<i>-nya</i>	49	3.62%
<i>ini (inilah)</i>	46	3.39%
<i>tersebut</i>	23	1.70%
<i>specific NP repeated</i>	6	0.44%
<i>tadi</i>	4	0.30%
<i>mereka</i>	1	0.07%
<i>tertentu</i>	1	0.07%
<i>yang</i>	1	0.07%
<i>berkenaan</i>	4	0.30%
<i>di sini</i>	1	0.07%
Total	1355	100.00%

As with unique translational equivalents for the definite articles, there are translational equivalents that are unique to the indefinite article ‘a / an’. In the case of the indefinite articles, three unique translational equivalents for the English article of indefiniteness with different frequency of occurrence in the target text are found in the corpus – *satu-satu*, *suatu* and *mana-mana*. Their frequency of occurrence is shown in Table 6.

TABLE 6
Unique Malay Translational Equivalents for the
Indefinite Articles 'a/an

Translational Equivalent	n	Percentage
<i>satu-satu</i>	6	1.47%
<i>suatu</i>	9	2.21%
<i>mana-mana</i>	1	0.25%
Total	16	3.93%

DISCUSSION

The use of the unidirectional parallel corpus in this study has allowed us to examine the translational equivalents in the target language (Malay) for the definite and indefinite articles of the source language. The results reveal that the most frequent translational equivalent in the target language (Malay) for the articles in the source language (English) is the zero article. It can be used as a translational equivalent for the article of definiteness as well as for the articles of indefiniteness. The high frequencies of the use of (\emptyset) *zero* as the translational equivalents for both the definite (71.29%) and indefinite articles (56.02%) are due to the fact that the target language has no articles akin to the English articles. Apart from the use of (\emptyset) *zero*, the denotation of definiteness in the target language relies on the demonstrative pronoun *itu* (that), which can equally function to point out a particular person or thing or to indicate persons or things already mentioned. Surprisingly, the item is also used as an indefinite translational equivalent for 'a/an'. However, as expected, *itu* is used more as the translational equivalent for the

definite article (9%) than for the indefinite article (0.49%). In contrast, indefiniteness is manifested in the target texts through the use of the numeral *satu* (one) as translational equivalent of 'a / an'. It ranks second after (\emptyset) *zero* and accounts for (15.23%) of the translational equivalents for the indefinite article. Although *satu* is also found as a translational equivalent for the definite article 'the', the frequency is very low (0.22%). The cognates of *satu* – namely *se-*, *sese-* and *sesuatu* are also used more often as the translational equivalents for indefiniteness in the source texts than for the article of definiteness. Hence, despite the corpus results which reveal shared translational equivalents for the articles of definiteness and indefiniteness, the distributions of these equivalents indicate that they are primarily used either to indicate definiteness or indefiniteness. While *zero* is used for both definite and indefinite and is not specific to either, the translational equivalent using the demonstrative pronoun *itu* is more definite than indefinite while the use of the numerals and *se* + classifier) is primarily indefinite. However, because these items also appear as the translational equivalent for both definite and indefinite, the distinction between definiteness/ indefiniteness in English may not necessarily be as clear cut and unambiguous as in Malay. Hence, when translating the definite articles from English into Malay, the choice of the translational equivalent includes items that indicate definiteness and also the items that indicate indefiniteness, and vice-versa. Our findings support Khazryati

et al. (2011, p. 148) who found that the presence of the Malay numeral classifiers in modern Malay texts may be used to indicate the notion of indefiniteness of a noun. In addition, there is no one-to-one relationship between the articles of definiteness and indefiniteness in the source language and the translational equivalents used for them in the target language, and this is mainly due to the linguistics cross differences between the article systems of the source language (English) and the target language (Malay). Another finding of this study is the occurrence of two groups of unique translational equivalents. The unique translational equivalents of definiteness, which include the third person possessive pronoun *-nya*, the demonstrative pronoun *ini* (*inilah*), and the adjectives *tertentu*, *tersebut* and *berkenaan*. The unique translational

equivalents of indefiniteness are *satu-satu*, *suatu* and *mana-mana*.

It is worth noting that these two groups of equivalents are used in a few cases and the main explanation lies in the resources of the Malay linguistics system of indicating definiteness/ indefiniteness. The advantage of using a unidirectional parallel corpus is also obvious as it allows us to align the translation and generate a large quantity of instances of the source items and the translation equivalent in a very short time. It also allows us to generate and calculate the frequency of each translational equivalent accurately. The following line chart provides a complete overview of the translational equivalents in the target language (Malay) for the corresponding article of the source language (English) in the corpus.

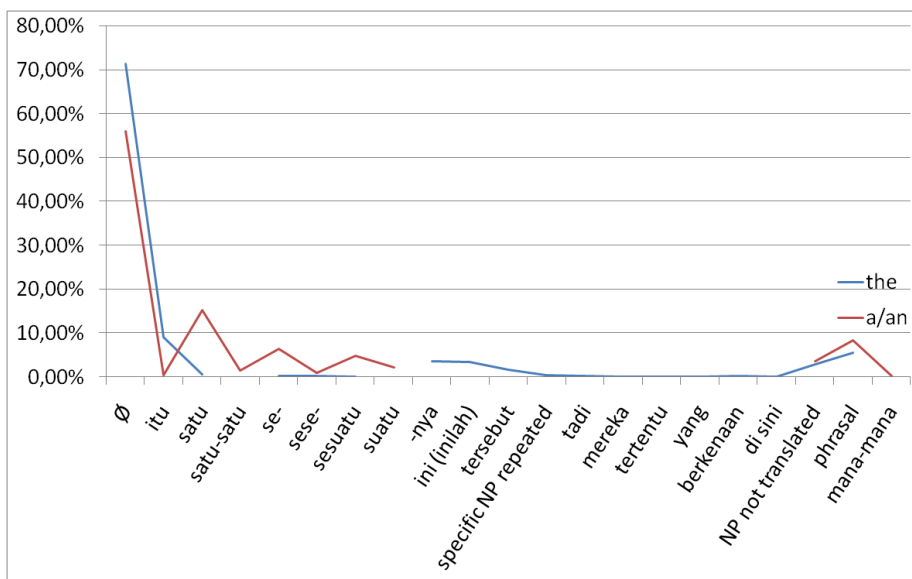


Fig.3: The Overall Distribution of the Translational Equivalents

CONCLUSION

In this paper, we have attempted to utilize a unidirectional parallel corpus to investigate how a particular grammatical feature of English (not present in the target language) is translated. By using the parallel corpus in the current study, the patterns for the translation equivalents of both the definite article 'the' and the indefinite article 'a/an' were observed and explained. Four distinctive patterns of the translational equivalents of the articles of both definiteness and indefiniteness in English do exist. Firstly, the shared transitional equivalent pattern contains six Malay translational equivalents - *ø*, *itu*, *satu*, *se-*, *sese-* and *sesuatu*. The translational equivalents of this particular pattern are varied in their degree in definiteness and indefiniteness. Secondly, the unique translational equivalents for definite article 'the' pattern include ten Malay translational equivalents. Similarly, there is a set of unique translational equivalents for the indefinite article of English 'a/an', namely, *satu-satu*, *suatu* and *mana-mana*.

In order to understand the occurrences, the translational equivalents were examined in relation to the functions of the articles in terms of the definite-indefinite referencing system of article usage in English. A total of 1,355 instances of the definite article (*the*) in English and 407 instances of the indefinite article (*a/an*) were detected and the translational equivalents in Malay were identified using the ParaConc. The results of the analysis of the Malay translational equivalents have mainly revealed that due

to cross linguistics differences in the article systems of the two languages involved, the English articles are most often translated as *ø*; numerals; and demonstrative pronouns. By conducting the current corpus-based study of the translational equivalents of English articles into Malay, we are able to uncover the distribution of the translation equivalents for the system of the English articles into Malay. Undeniably, the use of the parallel corpus in this paper and the findings observed are encouraging. However, a bidirectional parallel corpus is needed to serve as a control for the findings observed along its counterpart perspective.

REFERENCES

- Baker, M. (1993). Corpus-Linguistics and Translation Studies: Implication and Applications. In Baker, M., Francis, G., & Tognini-Bonelli, E. (Eds.), *Text and technology: in honour of John Sinclair* (pp. 233-250). Amsterdam & Philadelphia: John Benjamins.
- Baker, M. (1995). Corpora in translation studies: An overview and some suggestions for future research. *Target*, 7(2), 223-243.
- Baker, M. (1999). The Role of Corpora in Investigating the Linguistic Behaviour of Professional Translators. *International Journal of Corpus Linguistics*, 4(2), 281-298.
- Biel, Ł. (2010). Corpus-Based Studies of Legal Language for Translation Purposes: Methodological and Practical Potential. In Heine, C. & Engberg, J. (Eds.), *Reconceptualizing LSP. Online proceedings of the XVII European LSP Symposium 2009*. Aarhus 2010. Retrieved on May 15, 2012 from <http://www.asb.dk/fileadmin/www.asb.dk/isek/biel.pdf>.

- Celce-Murcia, M., & Larsen-Freeman, D. (1998). *The Grammar Book: An ESL/EFL Teacher's Course*. Boston: Heinle ELT.
- Chan, K. K. (1996). *'Itu' and the notion of definiteness in Malay* (Master thesis dissertation). Universiti Malaya, Kuala Lumpur, Malaysia.
- Cumming, S. (1991). *Functional change: the case of Malay constituent order*. Berlin: Mouton de Gruyter.
- Frankenberg-Garcia, A. (2009). Compiling and using a parallel corpus for research in translation. *International Journal of Translation*, 21(1), 57-71.
- Hawkins, J. A. (1978). *Definiteness and indefiniteness*. London: Croom Helm.
- Heim, I. (1991). *Articles and definiteness*. Published in German as 'Artikel und Definitheit'. In Stechow, A. V., & Wunderlich, D. (Eds.), *Semantics: An international handbook of contemporary research*. (pp. 487-535). Berlin: Mouton de Gruyter. Retrieved on May 15, 2012 from http://isites.harvard.edu/fs/docs/icb.topic980565.files/Sept%2021/heim_articles.pdf.
- Khazriyati Salehuddin, Tan, K. H., & Marlyna Maros. (2006). Definiteness and indefiniteness: A contrastive analysis of the use of determiners between the Malay language and English. *GEMA Online™ Journal of Language Studies*, 6(1), 21-30.
- Khazriyati Salehuddin, Winskel, S., & Marlyna Maros. (2011). The pragmatic functions of numeral classifiers in a modern Malay written corpus. *GEMA Online™ Journal of Language Studies*, 11(2), 137-153.
- Kruger, A. (2002). Corpus-based translation research: Its development and implications. *Acta Theological Supplementum*, 22, 70-106.
- Laviosa, S. (1998). The corpus-based approach: A new paradigm in translation studies. *Meta Translation Studies*, 43, 474-479.
- Martinez-Magaz, J. (2006). TRADI IMT (xx-xxi): Recent proposals for the alignment of a diachronic parallel corpus. *ICAME Journal*, 30, 55-65.
- Master, P. (1997). The English article system: Acquisition, function and pedagogy. *System*, 25, 215-232.
- Moropa, K. (2007). Analysing the English-Xhosa parallel corpus of technical texts with Paraconc: a case study of term formation processes. *Southern African Linguistics and Applied Language Studies*, 25(2), 183-205.
- Nik Safiah Karim, Farid M Onn, Hashim Haji Musa, & Abdul Hamid Mahmood. (2004). *Tatabahasa Dewan*. Kuala Lumpur: Dewan Bahasa & Pustaka.
- Olohan, M. (2004). *Introducing Corpora in Translation Studies*. London: Routledge.
- Selvan, A. R. (2010). *Problem of Equivalence in Translating English Articles into Tamil*. Retrieved on May 15, 2012 from <http://www.translationdirectory.com/articles/article22019>.
- Sinclair, J. (1995). Corpus typology: A framework for classification. In Melchers, G., & Warren, B. (Eds.), *Studies in Anglistics (Stockholm studies in English 85)* (pp. 17-33). Stockholm: Almqvist and Wiksell International.
- Uzar, R. S. (2002). A Corpus Methodology for Analysing Translation. In Agnin, S. E .O. T. (Ed.), *Cadernos de Tradução: Corpora e Tradução*. (pp. 235-263). Florianópolis: NUT.
- Wong, B. E., & Quek, S. T. (2007). Acquisition of the English definite article by Chinese and Malay ESL learners. *Electronic Journal of Foreign Language Teaching*, 4, 210-234.

