Equivalent Malay-Arabic Data Corpus Collection

Taj Rijal Muhamad Romli

Department of Modern Languages, Faculty of Languages and Communication, UPSI, 35900 Tg. Malim, Perak

Department of Foreign Languages, Faculty of Modern Languages of Communications, UPM, 43400 Serdang, Selangor

taj.rijal@fbk.upsi.edu.my

Abd Rauf Hassan

Department of Foreign Languages, Faculty of Modern Languages of Communications, UPM, 43400 Serdang, Selangor

raufh@upm.edu.my

Hasnah Mohamad

Department of Malay Language, Faculty of Modern Languages of Communications, UPM, 43400 Serdang, Selangor

hasnah_m@upm.edu.my

Abstract

This paper aims to introduce a search strategy and collecting comparable sentences of Arab-Malay corpus data. This method was introduced for the use of students, researchers and amateur translators to search and compare the structure of sentences in Arabic and Malay. The first stage is to collect data corpus with high impact titles from the press and must be able to enlarge the scope of study as stated by Maia (2003). The second stage is to search using the specified key words based on selected high-impact titles such as the Football World Cup year 2010 and 2014. Data search is by using Webcorp engine http://www.webcorp.org.uk/live/ corpus and also open database Google https://www.google.com. The third stage is to filter the data by using Aker et.al (2012) and Braschler's (1998) method based on similar story, related story and similar aspects. At the fourth stage every category is measured by Guidere's (2002) equivalence strength which is strong comparability (SC), medium (MC) and weak (WC). At the last stage comparable sentences between the two languages are compiled in parallel according to Mona Baker's (1992) level of grouping which are sentence level, combination of words, grammatical, pragmatic and textual level. The result from data analysis based on Mona Baker and Vinay & Darbelnet's (1995) comparable theory proved the existence of some sentences in large quantities are on the same level of comparability from the point of information delivery. This can be used as the basis of additional evidence concerning the validity of 'universal theory.' in the science of translation.

Keywords: software, comparable, parallel

Introduction

The effort to develop Arabic-Malay data corpus in Malaysia is still at the beginning. The search strategy and the collection of comparable data by using data corpus of open source is proposed in order to save time and effort as compared to the old method usually done by students, amateur translators and teachers. This strategy is also expected to be used to develop specific software which is online sentence dictionary that functions as a dictionary of sentences featuring on-screen display of comparable sentences between two languages or more. This can be said as an effort to assist and improve the use of dictionaries in the schools and universities that typically offer translation word by giving examples of usage only.

BACKGROUND

According to Taj Rijal (2015), Arabic corpus is limited which has been pioneered by researchers from the West, such as Leeds Arabic Internet study (http://corpus.leeds.ac.uk/internet.html) developed by the University of Leeds contains 170 million entries (Atwell, 2011), Tunisian Arabic corpus (http://tunisiya.org) contains 405,590 entries (Karen McNeil and Miled Faiza, 2011), the Arabic-Czech corpus (CLARA - corpus Linguae Arabicae) by Charles University (Zamanek, 2001) and arabiCorpus (http: // arabicorpus.byu.edu) contains 30 million entries run by Dilworth Parkinson's, who is a professor of Arabic at Brigham Young University.

Among other corpus projects are Corpus of Contemporary Arabic (CCA) by Latifa Al-Sulaiti & Eric Atwell (2006) as well as Arabic-Dutch corpus (Vertaalwoordenboek Arabisch-Nederlands) by Mark Van Mol. According to Mark (2002) corpus developed by him since 1996, consists of two types, one based on word, containing 26,000 entries and another based on sentences, containing 4,000,000 entries which have been tagged, taken from over 1,200,000 various sources of discourses. By developing such corpuses has helped them to improve tools and methods of teaching Arabic in their country .

The need for Arab-Malay corpus is very important, especially for the development of the Arabic language in Malaysia. This study will contribute to the efforts of building and developing the Arabic data corpus that has a parallel translation in Malay. By applying the theory of lexicography and computer corpus analysis has systematized the compilation of dictionary with latest methods. In addition, it also can create an effective machine translation that can help translators expedite their business.

Translation studies of Arabic-Malay and vise-versa in Malaysia using corpus data has been simplified and mobilized effectively as the basic of corpus usage and its result had been proven successful by European researchers. By using this method, all of the latest data of the language usage are stored. The saved corpus data can be reached through concordance system that allows researchers to see the usage of language and its estimated latest meaning. As specified by St. John (2001), concordance is a display of lines of words or combination of words in a context that is removed from corpus text. Researchers must use the keyword search for certain words, so the desired data search can be generated.

OBJECTIVE

This paper aims to explain the steps needed to be done by students and amateur translators in finding comparable data on Arabic-Malay using the search engine of open source corpus. Every step taken is based on the methods and theories that have been introduced by researchers and experts in the field of translation corpus.

RESEARCH SIGNIFICANCE

The study is expected to introduce a strategy of searching comparable sentences using search engine of open source corpus. This strategy is hoped to be used as a foundation in developing online Arabic-Malay sentence dictionary using text comparison by comparable method between Arabic and Malay languages known as bilingual comparable corpora. This strategy is also expected to become essential in proving the validity of 'universal' theory in translation that still need evidences from the findings of studies in language, corpus and translation. At the same time, it can be used as a methodology of teaching and learning in the classroom and home studies for students and amateur translators learning practical knowledge in translation. This study was also designed to help school and university students to solve problems in structuring sentences in Arabic, especially in the writing and translation courses.

LITERATURE REVIEW

According to Zanetten(1998), Rusli & Norhafizah (2001) and Kruger (2004), there are two types of corpus that can be used as items of study to replace the dictionary. The first, referred to as parallel corpora which compares the original with the translated text. The second, known as comparable bilingual corpus (comparable corpora) which compares the text in two different languages sharing the same topic. For example, some topics from world press reporting the news about important events in multiple languages (Li Shao and Hwee Tou Ng, 2004).

According to Rusli & Nurhafizah (2001), the DBP had made an attempt to build a database of Malay idiomatic and unidiomatic phrases based on actual usage of the language in the translated text. This database contains common phrases

and regular expressions of source language (English) with its equivalent in the target language (Malay). Phrases and regular expressions with their equivalences are derived from parallel and comparable corpora.

In Europe, comparable corpus studies had started since 1990s. Comparable corpus is bilingual texts that are not parallel but interrelated and deliver redundant infomations derived from various webs such as news issued by news agencies such as CNN and BBC. Among the studies that utilize comparable corpus are such as made by Munteano and Marcu (2005) and Munteano (2006).

Various techniques have been introduced by those researchers such as Rapp (1995), made an assumption that comparable word that can be translated may appear in the same context though from unrelated texts. Rapp took 100 words with their translations representing the context as a vector representing the same event (co-occurrence vector). The result is, a matrix of the same event becomes more similar when the composition of the words in the matrix is similar in both languages.

Aker et. al (2012) in collaboration with Google has established a simple technique to collect comparable corpus from the web. This is because the techniques introduced by former researchers such as Rapp (1999), Monteanu and Marcu (2002), Resnik (1999), Huang et.al. (2010), Talvensaari (2008), and others are time-consuming and requires a lot of resources. The objective of his research is to reduce the amount of time and resources. Previously, researchers have to go through three steps to collect and build a of comparable corpus, namely:

First: by downloading the document from the list of titles of the two languages. The process of downloading the document takes a long time and has to go through many obstacles.

Second: the process of matching with comparable data,and thus the data extraction. Headlines beneficial to the study from various categories of the selected language are taken together with the time and date of the newscast.

Third is to divide the title into several entities in the source language and named after people, places or organizations. The next phase is the process of making document alignment to compare the titles of articles from collected corpora. If it is comparable to the actual article it is then downloaded to obtain the matching corpus.

BAKER'S EQUIVALENCE LEVEL

Mona Baker (1992) introduces five levels of equivalence. The first is word level. This level exists in almost all languages of the world. One word represents one unit in searching equivalent meaning.

The second is combined words level (above word level). The equivalence is by combining words to give a meaning such as collocation.

The third is grammatical level where each language has different grammar. These differences pose problems in finding equivalent meaning directly in the translation. It also causes significant changes in how the message or information is transferred. These changes will cause an increase or decrease of information in a language.

The fourth is pragmatic level where the equivalence is from the aspect of coherence and interpretation process such as speech acts. Text should be evaluated based on the intention of the author in the target culture so that readers in target language can understand it.

The fifth is the textual level where the equivalence is on thematic structure and data (information and messages) and the use of cohesive tools such reference, connectors, replacement, ellipsis and lexical cohesion.

VINAY & DARBELNET EQUIVALENCE

Equivalence introduced by Vinay & Darbelnet (1995) can be said as different from the meaning of dynamic equivalence introduced by Nida & Taber. In Leonardi (2000), Vinay & Darbelnet concluded that equivalence is a process of reproducing the same situation as found in the original text, although phrased in a different language. If the procedure can be applied in the process of translating, it can mantain the effect of source language text style to the language of target text. This method is more ideal to find equivalent translation for proverbs, idioms, cliches, adjective phrases, and onomatopoeia of animal sounds.

Vinay & Darbelnet (in Pym, 2010) had used natural method of translation in translating to maintain the same function of language with different terminology. This is referred to as culture adaptation of target language. Vinay & Darbelnet prioritize style effects compared to Nida who gave priority to the effect of message to the target user. Although they both state that the translation is a procedure that is based on equivalence or *equivalence-oriente* which is a process of replicating a similar situation with the original text with different words differ (Leonardi, 2000); but they argued that semantic meaning in the dictionaries is not enough to help produce a successful translation.

Thus the theory is seen as fit to be applied as a method in this study because it maintains the effect of culture and style of the text. Therefore, students will be able to understand the true form of language structure.

STRATEGY

The study will gather topics that have a probability of sharing the same information both in Malay and Arabic. The search for comparable meaning will be assisted by an open corpus online using corpus search engines of Webcorp http://www.webcorp.org.uk/live/ besides Google database search engine; https://www.google.com.

Ratings are based on heuristic Aker's techniques (2012), the TS (title similarity), HS (time difference), and TLD (title length difference) when used in a combination of TS-HS-TLD and Braschler &Schäuble (1998) for the same story, related story and similar aspects. Each category is then measured the highest level of comparability of three levels as recommended by the Guidere (2002) as strong equivalence, medium equivalence and weak equivalence.

Figure 1 shows a general overview of these steps:

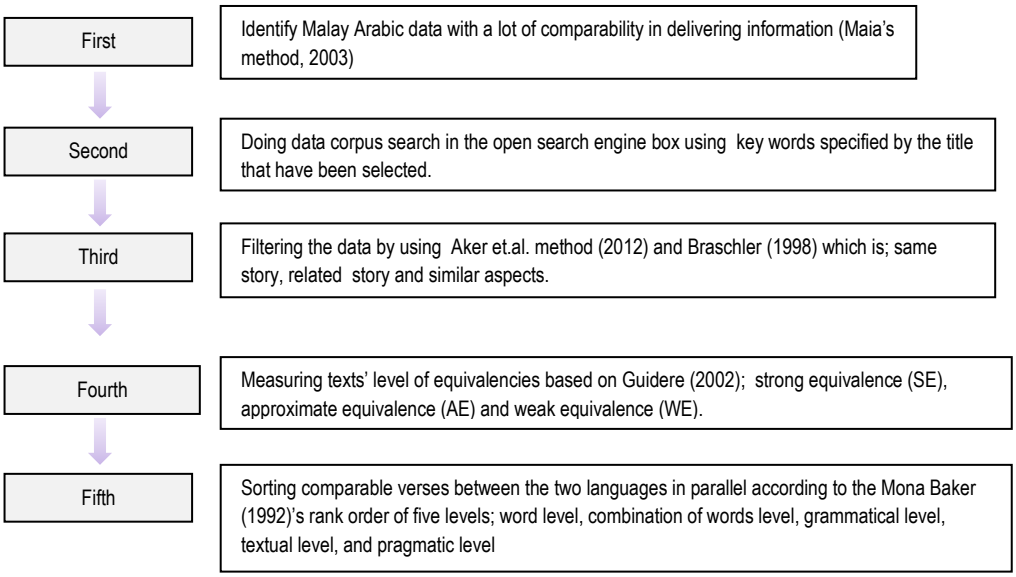

Figure 1

STEPS

In particular, the search and compilation steps can be described as follows:

a) Identifying data:

Source texts are taken from any sources that publish the article, document or text in both Arabic and English. The selected title is a topic that dominated the world news in a particular time whether in politics, economy, war and general genres. The

study will have a problem to find an equivalent texts if the searched topic do not dominate the news in general as the equivalence score is not enough to be relied as equivalent due to the lack of data collected. On the other hand, the topics that dominated the discussion in every major newspaper world will open up a wider debate, as defined by Maia (2003), thus triggering the levels of new language usage and many terms related to this topic will appear. This study chooses football from sports genre as the topic particularly the World Cup in 2010 and 2014 because of the wider impact and being the headlines around the world.

Through general inspection of several titles based on their importance and impact in the headline of daily newspapers. Among the titles preferred are World Cup Final Match 2010, match between top teams and World Cup Final Match 2014.

b) Finding equivalent article:

This step comes after determining the topic in the genre of football. The task of searching articles is by utilizing the Google and Webcorp search engine. Based on the big topic of the final match of the 2010 World Cup, the researcher can choose a few key words such as 'World Cup 2010 tournament,' 'final match', 'Spain champion', 'Spain beat the Netherlands' and 'winning goal'. In the same time making the timing of the match between July 10, 2010 to July 13, 2010. Also setting the place, which is 'Malaysia' to limit the scope of the language in Malay only. Among the titles that were achieved by the search engine in Arabic are:

"الغضب الأحمر" بطلا للعالم – الإمارات اليوم

المونديال- صحيفة الشعب اليومية. نهائي في الثالثة للمرة تسقط وهولندا الأول لقبها تحرز أسبانيا

النيلين – بهدف هولندا على الفوز بعد 2010 العالم كأس بطل إسبانيا

وصدقت توقعات بول.. انييستا يضع اسبانيا على عرش العالم- مجموعة البورصة المصرية

للعالم- فرانس24 بطلة وتتوج هولندا تفوز1-0 على أسبانيا

نهائي كأس العالم لكرة القدم 2010 - ويكيبيديا

لكرة القدم- بي. بي. سي. العربية 2010 العالم كأس وتحرز هولندا على تفوز أسبانيا

القدم- صحيفة مباشرالعربية لكرة 2010 العالم بطل أسبانيا

القدم- وكالة فاكت نيوز الأنباء كرة في 2010 العالم كأس وتحرز هولندا على تفوز أسبانيا

العالمية- جريدة الخبر الكرة عرش على تتربع إسبانيا

While the headlines listed in the Malay language are:

Sepanyol juara Piala Dunia

Andres Iniesta wira julang Sepanyol juara Piala Dunia 2010

Perlawanan Akhir Piala Dunia FIFA 20103

Final: Belanda 0 Sepanyol 1

Piala Dunia: Sepanyol juara, sotong gurita juga cipta sejarah

Andres Iniesta wira Sepanyol

Sepanyol sekat hajat Marwijk kubur "Total Football" buat selama-lamanya.

c) Filtering data:

The selected data in Arabic and Malay languages in this study are classified according to two of five categories: as introduced by Aker et.al. (2012). The first category is based on the same story,thus two articles with the same title and the same information. The second category is based on related story, thus two articles covering news on two related events.

d) Measuring the equivalence:

Once the texts are divided into two categories as above they are then sorted based on the level of the strongest equivalent. The equivalence as recommended by the Guidere (2002) is to find equivalence from the aspect of language structure and sentence. For texts with strong equivalence (SE), the evaluation is based on the similarity of number of words, its sequence and meaning. Approximate equivalence (AE), is assessed by the number of words and same meaning, but not its sequence. Weak equivalence (WE) is the sequence and the number of words are different, but the lexical meaning is the same.

e) Organizing the findings:

Organizing comparable sentences between the two languages in parallel according to Mona Baker (1992)'s rank order of five levels, word level, combination of words level, grammatical level, textual level, and pragmatic level.

Examples of comparable data

Table 1 below is an example of the analysis and the formulation of a number of comparable text taken from the titles of 2010 World Cup final match between Germany and Argentina, as measured by Braschler and Schäuble (1998)'s method which is; same story, related story and similar aspects. After removing the difference of extra level of phrases and words in both text, these sentences are formulated as equivalent.

Arabic	Equivalence	Malay
(AC1-1 **ونجح منتخب ألمانيا في إحراز لقبه المونديالي الأول منذ كأس العالم 1990 بإيطاليا وكأس العالم الرابع في تاريخه**) (AC1-2 **ليقود ألمانيا لإحراز اللقب العالمي للمرة الرابعة في تاريخها.**) (AC1-3 **ليقود ألمانيا لإحراز لقب كأس العالم لكرة القدم**) (AC1-4 **فازت المانشافت مساء الأحد على ملعب ماراكانا في ريو دي جانيرو بكأس العالم للمرة الرابعة في تاريخها**) (AC1-6 **فاز المنتخب الألماني بكأس العالم 2014**) (AC1-7 **وأحرزت كأس العالم لكرة القدم للمرة الرابعة في تاريخها**) (AC1-8 **وتحرز بذلك لقب بطولة كأس العالم**)	Approximate Equivalence (AE) Weak Equivalenve (WE)	(MC1-1) membawa Jerman menjulang Piala Dunia 2014 (MC1-2) JERMAN ialah juara Piala Dunia 2014 (MC1-3) Jerman meraih Piala Dunia pertama mereka selepas penyatuan antara Jerman Barat dan Timur (MC1-4) Jerman muncul juara Piala Dunia 2014 tewaskan Argentina 1-0 (MC1-5) Jerman menjulang Piala Dunia dengan kemenangan 1-0 (MC1-6) Jerman Menafikan Misi Argentina memburu kejuaraan Piala Dunia Kali ke-3

Table 1

APPROXIMATE EQUIVALENCE (AE)

Text level in comparable data AC1-3 is approximate compared with MC1-1 data. The meaning and the sequence is the same but number of words are different. Arabic data is (in the form of re-translation):

AC1-3: *membawa Jerman meraih gelaran Piala Dunia Bola Sepak* / ليقود ألمانيا لإحراز لقب كأس العالم لكرة القدم

and the Malay text is:

MC1-1: membawa Jerman menjulang Piala Dunia 2014

There was an extra word 'gelaran' in AC1-3 data and '2014' in MC1-1 data.

WEAK EQUIVALENCE (WE)

Text level in comparable data AC1-1, AC1-2, AC1-4, AC1-6, AC1-7 and AC1-8 is ranked as weak (WE) when compared with data MC1-3, MC1-4, MC1-5 and MC1-6. The meaning is the same but number of words and sequence are different words. These data are listed in the following Arab (in the form of re-translation):

AC1-1: Pasukan Jerman berjaya merangkul gelaran Dunia terulung / ونجح منتخب ألمانيا في إحراز لقبه المونديالي الأول

AC1-2: Membawa Jerman merangkul gelaran Dunia untuk kali ke-4 / ... ليقود ألمانيا لإحراز اللقب العالمي للمرة الرابعة

AC1-4: Pasukan Jerman berjaya menjulang Piala Dunia untuk kali ke-4/ ... فازت المانشافت.... بكأس العالم للمرة الرابعة

AC1-6: Jerman memenangi Piala Dunia 2014 / 2014 فاز المنتخب الألماني بكأس العالم

AC1-7: Merangkul Piala Dunia Bola Sepak untuk kali ke-4/ ... وأحرزت كأس العالم لكرة القدم للمرة الرابعة في

AC1-8: Dengan demikian merangkul gelaran Juara Piala Dunia /وتحرز بذلك لقب بطولة كأس العالم

and from Malay data are:

MC1-2: JERMAN ialah juara Piala Dunia 2014

MC1-3: Jerman meraih Piala Dunia pertama mereka selepas penyatuan antara Jerman Barat dan Timur

MC1-4: Jerman muncul juara Piala Dunia 2014 tewaskan Argentina 1-0

MC1-5: Jerman menjulang Piala Dunia dengan kemenangan 1-0

MC1-6: Jerman Menafikan Misi Argentina memburu kejuaraan Piala Dunia Kali ke-3

Equivalence level in the above data are measured at sentence level as follows:

AC1-1: ونجح منتخب ألمانيا في إحراز لقبه

AC1-2: ليقود ألمانيا لإحراز اللقب العالمي

AC1-4: فازت المانشافت.... بكأس العالم

AC1-6: فاز المنتخب الألماني بكأس العالم

AC1-7: وأحرزت كأس العالم لكرة القدم

AC1-8: وتحرز بذلك لقب بطولة كأس العالم

MC1-2: JERMAN ialah Juara Piala Dunia

MC1-3: Jerman meraih Piala Dunia pertama

MC1-4: Jerman muncul juara Piala Dunia

MC1-5: Jerman menjulang Piala Dunia

MC1-6: Jerman Menafikan Misi Argentina memburu kejuaraan Piala Dunia

In overall, the texts above explain the meaning of German's victory in the final match with Argentina.

The phrases and clauses such as: وأحرزت كأس العالم / فازت.... بكأس العالم/ يقود ... لإحراز اللقب / ونجح منتخب.... في إحراز / وتحرز بذلك لقب بطولة كأس العالم

Are equivalent to the following phrases: World Cup Champion, won the World Cup / appear as world Cup winners / winning the World Cup.

CONCLUSION

This methodology is expected to be used as an effective and easy method as a guide to find comparable data between Arabic and Malay languages. At the same time, it becomes as a platform for developing software of Arabic-Malay dictionary online. This method is also suitable to be used by lecturers, teachers and students in teaching and learning Arabic and Malay translation aided with corpus.

REFERENCE

Aker, A. Kanoulas, E. and Gaizauskas, R. (May, 2012). A light way to collect comparable corpora from the Web. *Proceedings of LREC*, 21-27.

ArabiCorpus. http://arabicorpus.byu.edu. Brigham Young University.

Al-Sulaiti L., Atwell Eric. (2006). The design of a corpus of contemporary Arabic. *International Journal of Corpus Linguistics,* 11, 135-171.

Atwell, ES; Brierley, C; Dukes, K; Sawalha, M; Sharaf. (2011). *An Artificial Intelligence Approach to Arabic and Islamic Content on the Internet. Proceedings of NITS 3rd National Information Technology Symposium.*

Baker, Mona. (1992). *A Coursebook on Translation*. London: Routledge.

Braschler M., Schäuble P. .(1998). Multilingual information retrieval based on document alignment techniques. *Proceedings of the 2nd European Conference on Research and Advanced Technology for Digital Libraries,* 183-197.

Davies Mark. (2011). *Online Corpora*. http://corpus.byu.edu/corpora.asp .

E. Morin, B. Daille, K. Takeuchi, K. Kageura. (2007). *Bilingual terminology mining - using brain, not brawn comparable corpora.* Proc. ACL.

Guidere, Mathieu. (January, 2002). Toward corpus based machine translation for standard arabic. *Translation Journal*, 6(1).

Jan Krikke, Benjamin Alfonsi. (Mar./Apr., 2006). *IEEE Intelligent Systems*. The News, 21(2), 4-7.

K. Precoda, J. Zheng, D. Vergyri, H. Franco, C. Richey, A. Kathol, S. Kajarekar. (2007). In Iraqcomm: a next generation translation system. *Interspeech*, 2841-2844.

Karen McNeil and Miled Faiza. (February, 2011). *The Jil Jadid conference at University of Texas, Austin*. Retrieved from http://tunisiya.org/media/TunisianArabicCorpusSummary.pdf.

Kruger. (2004). Corpus-Based Translation Research: Its Development And Implications For General. *Literary And Bible Translation*. Retrieved from http://www.ajol.info/index.php/actat/article/viewFile/5455/29593.

University of Leeds. *Leeds Arabic Internet*. http://corpus.leeds.ac.uk/internet.html.

Leonardi, Vanessa. (2000). Equivalence in Translation: Between Myth and Reality. *Translation Journal*, 4(4). Retrieved from http://translationjournal.net/journal/14equiv.htm,.

Li Shao, Hwee Tou Ng. (2004). Mining New Word Translations from Comparable Corpora. *COLING '04 Proceedings of the 20th international conference on Computational Linguistics*. Retrieved from http://www.mt-archive.info/Coling-2004-Shao.pdf.

Maia, Belinda. (2003). What are comparable corpora?. *Proceedings of the workshop on multilingual corpora: linguistic requirements and technical perspectives, at the corpus linguistics.*, 27–34. UK: Lancaster.

Muhammad Fauzi Jumingan. (2007). *Status Bahasa Arab di Malaysia: Bahasa Asing atau Bahasa Kedua* - Research Grant Scheme (RUGS). Serdang: UPM.

Munteanu, D.S., Marcu, Daniel. (2005). Improving Machine Translation Performance by Exploiting Non-Parallel Corpora. *Computational Linguistics*, 31(4), 477-504.

Rusli Abdul Ghani dan Norhafizah Mohamed Husin. (2001, Sept., 3). Yang Selari dan Yang Setanding: Peranan Korpus dalam Penterjemahan. *Kertas Kerja Persidangan Penterjemahan Antarabangsa Ke-8*, Langkawi, Kedah.

Rusli Abdul Ghani et al. (2004, Mar. 30). Pangkalan Data Korpus DBP: Perancangan, Pembinaan dan Pemanfaatan. *Kertas Kerja Seminar Sehari Linguistik*. Pusat Pengajian Bahasa dan Linguistik, UKM, Bangi.

Sinclair, J. (2005). Corpus and Text. *Basic Principles in Developing Linguistic Corpora: a Guide to Good Practice*, ed. M. Wynne. Oxford: Oxbow Books: 1-16. Retrieved from http://www.ahds.ac.uk/creating/guides/linguistic-corpora/chapter1.htm.

St. John, Elke. (2001). A Case for Using a Parallel Corpus and Concordancer for Beginners of a Foreign Language. *Language Learning & Technology*. 5(3), 185-203. Online journal retrieved from http://llt.msu.edu/vol5num3/stjohn.

Taj Rijal Muhamad Romli, Muhamad Fauzi Jumingan. (Sep-Dec, 2015). Open Source Corpus as a Tool for Translation Training. *European Journal of Language and Literature Studies*. 3(1), 61-68.

Tengku Sepora Tengku Mahadi, Helia Vaezian, Mahmoud Akbari. (2010). Design and development procedure of an English-Malay parallel corpus. *Universal Corpora for Comparative and Translation Studies (UCCT)*. UK: Edge Hill University.

Van Mol, Mark. (2002). The Semi-automatic tagging of Arabic Corpora. *Workshop Proceedings. Arabic Language Resources and Evaluation, Status and Prospects, LREC*, 40-44.

Vinay, Jean-Paul and Jean Darbelnet. (1995). A methodology for translation. *Juan C. Sager and M.-J. Hamel (eds. and trans.) Comparative Stylistics of French and English: A Methodology for Translation*, 31-42, reprinted in Venuti 2000:84-93 (Also in the 2004 edition).

Zaharin Yusoff. (2007). Machines and Translation. *Membina kepustakaan dalam bahasa Melayu*. Kuala Lumpur: PTS,.

Zanettin, Federico. (1998). Bilingual comparable corpora and the training of translators. *Meta: Translators' Journal*, 43(4), 616-630. Retrieved from http://www.erudit.org/erudit/meta/v43n04/zanettin/zanettin.html

Zemánek, Petr. (2001). CLARA (Corpus Linguae Arabicae): An Overview. *Proceedings of ACL/EACL Workshop on Arabic Language Processing: Status and Prospects*. Retrieved from http://www.elsnet.org/arabic2001/zemanek.pdf.