KESETANDINGAN MAKNA TERJEMAHAN BERKOMPUTER ARAB-MELAYU

Oleh

TAJ RIJAL BIN MUHAMAD ROMLI

HAK CIPTA

05-4506832 pustaka.upsi.edu.my Perpustakaan Tuanku Bainun Kampus Sultan Abdul Jalil Shah PustakaTBainun ptbupsi

05-4506832 pustaka.upsi.edu.my Perpustakaan Tuanku Bainun Kampus Sultan Abdul Jalil Shah PustakaTBainun ptbupsi

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk Ijazah Doktor Falsafah

KESETANDINGAN MAKNA TERJEMAHAN BERKOMPUTER ARAB-MELAYU

Oleh

TAJ RIJAL BIN MUHAMAD ROMLI

April 2017

Pengerusi : Abd Rauf Hassan, PhD
Fakulti : Bahasa Moden dan Komunikasi

Kajian ini merupakan satu kajian perbandingan teks antara bahasa Arab dengan bahasa Melayu yang bertujuan mencari makna setanding untuk dijadikan model terjemahan. Kajian ini timbul berikutan masalah kualiti terjemahan yang ditawarkan oleh mesin terjemahan *Google* ditambah pula dengan faktor ketiadaan mesin terjemahan yang efektif yang dapat membantu usaha penterjemahan buku Arab ke bahasa Melayu. Kajian ini mengfokuskan kepada mengenal pasti data-data korpus Arab-Melayu setanding yang wujud dalam talian terbuka. Data-data ini dikumpulkan, diklasifikasikan dan dinilai mengikut tahap makna kesetandingan. Kajian terjemahan berkomputer ini akan memanfaatkan data-data korpus terbuka yang ditawarkan oleh perkhidmatan atas talian dan laman sesawang seperti perisian korpus *Google*, *Webcorp*, *arabiCorpus* dan Pangkalan Data Bahasa Melayu DBP. Metode pencarian data setanding akan menggunakan kaedah pencarian korpus setanding (*comparable corpora*) cerita sama dan cerita berkaitan yang diperkenalkan oleh Aker dan rakan-rakan (2012). Dapatan data-data setanding kemudiannya dinilai berdasarkan lima aras kesetandingan (*equivalence*) Baker (1992) iaitu aras perkataan, gabungan perkataan, aras tatabahasa, aras tekstual dan aras pragmatik. Mutu setiap tahap lima kesetandingan ini seterusnya dinilai terdasarkan tiga level Guidere (2002) iaitu Kesetandingan Kuat (KK), Kesetandingan Sederhana (KS) dan Kesetandingan Lemah (KL). Di peringkat akhir, data-data yang berada dalam tiga level Guidere akan dirumus untuk melihat nilai terjemahannya berdasarkan kaedah sama entiti dan sama peristiwa seperti yang diperkenalkan oleh Richard Sproat, Tao Tao dan ChengXiang Zhai (2006). Hasil daripada proses penapisan ini, banyak data yang setanding boleh dikategorikan mempunyai nilai kesetandingan makna terjemahan Arab-Melayu, bahkan boleh dijadikan model terjemahan yang bermutu kepada pelajar atau penterjemah amatur yang mengambil kursus terjemahan atau bahasa Arab di peringkat pengajian tinggi. Kajian ini diharap akan dimanfaatkan sebagai data leksikal untuk leksikografi Arab-Melayu dan pembinaan korpus setanding Arab-Melayu yang lebih besar pada masa akan datang.

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the Degree of Doctor of Philosophy

EQUIVALENT MEANING OF ARABIC-MALAY COMPUTER TRANSLATION

By

TAJ RIJAL BIN MUHAMAD ROMLI

April 2017

Chairman : Abd Rauf Hassan, PhD
Faculty : Modern Languages and Communication

This study is a comparative text study between Arabic and Malay languages, which aims to find equivalent meaning to be a translation model. This study arose following problems regarding translation quality offered by *Google* translation engines. Furthermore, another factor is the absence of effective machine translation which can assist the effort of translating Arabic books to Malay language. It focuses on identifying comparable Arabic-Malay data corpus which existed in open online. These data are collected, classified and evaluated according to the level of comparable meaning. This computer translation study will make use of data from open corpus offered by online services and websites such as *Google* corpus software, *Webcorp*, *arabiCorpus* and DBP Malay Database. Comparable corpora search method of similar stories and related stories introduced by Aker et.al. (2012) will be applied as method of searching comparable data. The finding from comparable data are then evaluated based on Baker's five levels of equivalence (1992); the level of words, above words level, grammatical level, textual level, and pragmatic level. The quality of the five levels is then valued based on Guidere's (2002) three levels of equivalence; Strong Equivalence (SE/KK), Medium (ME/KS) and Weak (WE/KL). In the final stage, the data within Guidere's three levels will be formulated to evaluate the translation based on the method of same entity and same event as introduced by Richard Sproat, Tao Tao and Chengxiang Zhai (2006). As a result of this filtering process, many data can be categorized as having equivalent meaning for Arabic-Malay translation, and even can be used as a model for quality translation to the students or amateur translators who take translation courses or Arabic language in higher education. This study is expected to be utilized as lexical data for Arab-Malay lexicography, thus the formation of a larger Arabic-Malay comparable corpus in the future.

PENGHARGAAN

الحمد لله حمدًا كثيرًا طيّبًا مباركًا فيه كما يحبّ ربّنا ويرضى، وأشهد أن لا إلهَ إلا الله وحدَه لا شريكَ له العليّ الأعلى، وأشهد أنّ سيدنا ونبينا محمّدًا عبده ورسوله نبيّ الهدى، صلى الله وسلم وبارك عليه وعلى آله وأصحابِه الأتقياء، أما بعد...

Alhamdulillah segala puji dipanjatkan kepada Allah S.W.T. kerana dengan izin dan limpah kurniaNya, saya dapat menyiapkan tesis ini bagi memenuhi syarat penganugerahan ijazah Doktor Falsafah Bahasa Arab,Universiti Putra Malaysia.

Dalam kesempatan yang ada ini saya ingin merakamkan penghargaan dan terima kasih yang tidak terhingga kepada barisan Jawatankuasa Penyeliaan tesis ini iaitu Dr. Abd. Rauf bin Hassan dan Dr. Hasnah Mohamad yang telah banyak membantu, membimbing, memberi tunjuk ajar dan menyelia kerja penyelidikan sehingga tesis ini dapat disiapkan dengan jayanya. Seribu doa dipanjatkan kepada Almarhum PM Dr. Muhammad Fauzi Jumingan yang telah menyelia dan membimbing tesis ini daripada peringkat awal hingga ia hampir siap sehinggalah beliau ditimpa kesakitan yang membawa ajal. Moga Allah S.W.T. mencucuri rahmat dan keampunanNya ke atas ruh beliau.

Terima kasih juga saya ucapkan kepada semua kakitangan dan pensyarah Jabatan Bahasa Asing, Fakulti Bahasa Moden dan Komunikasi yang banyak memberi kerjasama dan membantu sepanjang pengajian saya di Universiti Putra Malaysia. Terima kasih juga kepada rakan sejawatan di Universiti Pendidikan Sultan Idris serta staf yang banyak memacu dan menghulurkan pelbagai kemudahan untuk menyiapkan tesis ini. Seterusnya jutaan penghargaan kepada Bahagian Tajaan Pendidikan, Kementerian Pelajaran Malaysia yang telah memberikan saya peluang untuk menyambung pelajaran selama tiga tahun di Universiti Putra Malaysia.

Juga tidak dilupakan penghargaan ini kepada ayahanda Hj. Muhamad Romli bin Abdullah dan bonda Hajah Khatijah Mk. Othman yang sentiasa mendoakan kejayaan saya. Juga buat isteri dan anak-anak tersayang, serta semua anggota keluarga kerana amat memahami, memberi dorongan dan sokongan sepanjang penyelidikan dan penulisan tesis.

Akhir kata, harapan saya agar usaha ini dapat memberi manfaat kepada bidang bahasa Arab dan terjemahan terutamanya para pelajar untuk diaplikasikan dalam pengajian mereka. Moga usaha ini diberi ganjaran sebagai sebahagian daripada amal kebajikan oleh Allah S.W.T. Sesungguhnya segala kekurangan dan kelemahan datangnya daripada saya kerana sesuatu yang lengkap dan sempurna itu milikNya sahaja. Terima kasih.

TAJ RIJAL BIN MUHAMAD ROMLI
Tg. Malim, Perak
18 Januari 2017

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the Degree of Doctor of Philosophy

EQUIVALENT MEANING OF ARABIC-MALAY COMPUTER TRANSLATION

By

TAJ RIJAL BIN MUHAMAD ROMLI

April 2017

Chairman : Abd Rauf Hassan, PhD
Faculty : Modern Languages and Communication

This study is a comparative text study between Arabic and Malay languages, which aims to find equivalent meaning to be a translation model. This study arose following problems regarding translation quality offered by *Google* translation engines. Furthermore, another factor is the absence of effective machine translation which can assist the effort of translating Arabic books to Malay language. It focuses on identifying comparable Arabic-Malay data corpus which existed in open online. These data are collected, classified and evaluated according to the level of comparable meaning. This computer translation study will make use of data from open corpus offered by online services and websites such as *Google* corpus software, *Webcorp*, *arabiCorpus* and DBP Malay Database. Comparable corpora search method of similar stories and related stories introduced by Aker et.al. (2012) will be applied as method of searching comparable data. The finding from comparable data are then evaluated based on Baker's five levels of equivalence (1992); the level of words, above words level, grammatical level, textual level, and pragmatic level. The quality of the five levels is then valued based on Guidere's (2002) three levels of equivalence; Strong Equivalence (SE/KK), Medium (ME/KS) and Weak (WE/KL). In the final stage, the data within Guidere's three levels will be formulated to evaluate the translation based on the method of same entity and same event as introduced by Richard Sproat, Tao Tao and Chengxiang Zhai (2006). As a result of this filtering process, many data can be categorized as having equivalent meaning for Arabic-Malay translation, and even can be used as a model for quality translation to the students or amateur translators who take translation courses or Arabic language in higher education. This study is expected to be utilized as lexical data for Arab-Malay lexicography, thus the formation of a larger Arabic-Malay comparable corpus in the future.

Saya mengesahkan bahawa satu Jawatankuasa Peperiksaan Tesis telah berjumpa pada 20 April 2017 untuk menjalankan peperiksaan akhir bagi Taj Rijal bin Muhamad Romli bagi menilai tesis beliau yang bertajuk "Kesetandingan Makna Terjemahan Berkomputer Arab-Melayu" mengikut Akta Universiti dan Kolej Universiti 1971 dan Perlembagaan Universiti Putra Malaysia [P.U.(A) 106] 15 Mac 1998. Jawatankuasa tersebut telah memperakukan bahawa calon ini layak dianugerahi ijazah Doktor Falsafah.

Ahli Jawatankuasa Peperiksaan Tesis adalah seperti berikut:

Pabiyah Hajimaming @ Pabiyah Toklubok, PhD
Pensyarah Kanan
Fakulti Bahasa Moden dan Komunikasi
Universiti Putra Malaysia
(Pengerusi)

Mohd Azidan bin Abdul Jabar, PhD
Profesor Madya
Fakulti Bahasa Moden dan Komunikasi
Universiti Putra Malaysia
(Pemeriksa Dalam)

Mohd. Zaki bin Abd. Rahman, PhD
Pensyarah Kanan
Universiti Malaya
Malaysia
(Pemeriksa Luar)

Rosni bin Samah, PhD
Profesor Madya
Universiti Sains Islam Malaysia
Malaysia
(Pemeriksa Luar)

NOR AINI AB. SHUKOR, PhD
Profesor dan Timbalan Dekan
Sekolah Pengajian Siswazah
Universiti Putra Malaysia

Tarikh: 6 July 2017

Tesis ini telah dikemukakan kepada Senat Universiti Putra Malaysia dan telah diterima sebagai memenuhi syarat keperluan untuk ijazah Doktor Falsafah. Ahli Jawatankuasa Penyeliaan adalah seperti berikut:

Abd Rauf Hassan, PhD
Profesor Madya
Fakulti Bahasa Moden dan Komunikasi
Universiti Putra Malaysia
(Pengerusi)

Hasnah Mohamad, PhD
Pensyarah Kanan
Fakulti Bahasa Moden dan Komunikasi
Universiti Putra Malaysia
(Ahli)

ROBIAH BINTI YUNUS, PhD
Profesor dan Dekan
Sekolah Pengajian Siswazah
Universiti Putra Malaysia

Tarikh:

Perakuan pelajar siswazah

Saya memperakui bahawa:

- tesis ini adalah hasil kerja saya yang asli;
- setiap petikan, kutipan, dan ilustri telah dinyatakan sumbernya dengan jelas;
- tesis ini tidak pernah dimajukan sebelum ini, dan tidak dimajukan serentak dengan
- ini, untuk ijazah lain sama ada di Universiti Putra Malaysia atau di institusi lain;
- hak milik intelek dan hak cipta tesis ini adalah hak milik mutlak Universiti Putra Malaysia, mengikut Kaedah-Kaedah Universiti Putra Malaysia (Penyelidikan) 2012;
- kebenaran bertulis daripada penyelia dan Pejabat Timbalan Naib Canselor (Penyelidikan dan Inovasi) hendaklah diperoleh sebelum tesis ini diterbitkan (dalam bentuk bertulis, cetakan atau elektronik) termasuk buku, jurnal, modul, prosiding, tulisan popular, kertas seminar, manuskrip, poster, laporan, nota kuliah, modul pembelajaran atau material lain seperti yang dinyatakan dalam Kaedah-Kaedah Universiti Putra Malaysia (Penyelidikan) 2012;
- tiada plagiat atau pemalsuan/fabrikasi data dalam tesis ini, dan integriti ilmiah telah dipatuhi mengikut Kaedah-Kaedah Universiti Putra Malaysia (Pengajian Siswazah) 2003 (Semakan 2012-2013) dan Kaedah-Kaedah Universiti Putra Malaysia (Penyelidikan) 2012. Tesis telah diimbaskan dengan perisian pengesanan plagiat.

Tandatangan: _____ Tarikh: 1/5/2017

Nama dan No. Matrik: Taj Rijal Bin Muhamad Romli,GS27562

Perakuan Ahli Jawatankuasa Penyeliaan:

Dengan ini, diperakukan bahawa:

- penyelidikan dan penulisan tesis ini adalah di bawah seliaan kami;
- tanggungjawab penyeliaan sebagaimana yang dinyatakan dalam Universiti Putra Malaysia (Pengajian Siswazah) 2003 (Semakan 2012-2013) telah dipatuhi.

Tandatangan: _____
Nama Pengerusi Jawatankuasa Penyeliaan: Abd Rauf Hassan, PhD

Tandatangan: _____
Nama Ahli Jawatankuasa Penyeliaan: Hasnah Mohamad, PhD

JADUAL KANDUNGAN

SENARAI JADUAL

SENARAI RAJAH

SENARAI SINGKATAN

BCET	Birmingham Collection of English Texts
CATT	Computer Aided Translation Technology
CCA	Corpus of Contemporary Arabic
CLARA	Corpus Linguae Arabicae
CLIR	Cross-Language Information Retrieval
Cobuild	Collins Birmingham University International Language Database
COCA	Corpus of Contemporary American English
COHA	Corpus of Historical American English
DS	Data Setanding
KK	Kesetandingan Kuat
KL	Kesetandingan Lemah
KS	Kesetandingan Sederhana
MT	Mesin Terjemahan
PDK	Pangkalan Data Korpus
PdP	Pengajaran dan Pembelajaran
SDL KbTS	SDL Knowledge-based Translation Sistem
TS	Teks Sumber
TT	Teks Terjemahan

PANDUAN TRANSLITERASI EJAAN ARAB KE RUMI
(Dewan Bahasa Dan Pustaka, 2008)

Huruf Arab	Huruf Rumi
Konsonan	
ب	b
ت	t
ث	th
ج	j
ح	ḥ/Ḥ
خ	kh
د	d
ذ	dh
ر	r
ز	z
س	s
ش	sy
ص	ṣ/Ṣ
ض	ḍ/Ḍ
ط	ṭ/Ṭ
ظ	ẓ/Ẓ
ع	'
غ	gh
ف	f
ق	q
ك	k

ل	l
م	m
ن	n
و	w
ه	h
ء	'
ي	y
ة	h
Vokal pendek	
ـَ	a
ـِ	i
ـُ	u
Vokal panjang	
ـَا	ā/Ā
ـِي	ī/Ī
ـُو	ū/Ū
Diftong	
ـَيْ	ay
ـَوْ	aw

BAB 1

PENDAHULUAN

1.1 Pengenalan

Statistik kajian *Common Sense Advisory* bulan Jun (2016) melaporkan bahawa nilai pasaran global industri bahasa akan mencecah ASD (Dolar Amerika Syarikat) 45 bilion pada tahun 2020. Purata peningkatan pasaran adalah lebih lima peratus setiap tahun. Menurut laporan *IBISWorld* (2015), perkhidmatan penterjemahan dan interpretasi sahaja dijangka akan terus berkembang kepada nilai ASD 37 bilion pada tahun 2018. Ini menunjukkan bahawa terjemahan merupakan satu bidang yang penting yang boleh menjana ekonomi dunia dan ini semestinya memberi kesan kepada pasaran terjemahan di Asia khususnya di Malaysia.

Menurut Abdullah Hassan (2011), beliau menyarankan Malaysia mengikuti jejak langkah Taiwan yang telah berjaya menerbitkan lapan ribu judul buku terjemahan setahun iaitu sebanyak 20 peratus daripada jumlah buku yang diterbitkan. Untuk menerbitkan buku terjemahan sejumlah itu, Malaysia memerlukan penambahbaikan terhadap perisian teknologi dan menaik taraf alat bantu terjemahan supaya dapat membantu mempercepatkan proses penterjemahan terutamanya yang berkaitan dengan penterjemahan Arab-Melayu.

Dengan kemajuan dan penggunaan teknologi yang berkembang pesat, penterjemahan sudah menjadi nadi yang menyumbang kelangsungan tamadun manusia. Terjemahan dan teknologi ibarat aur dengan tebing. Masa dan kecepatan proses menjadi pemacu kepada keberkesanan produk dan pemindahan maklumat. Kelewatan dalam memproses menyebabkan sesuatu negara itu akan ketinggalan. Dahulu, masyarakat bergantung kepada kepakaran dan kebolehan individu penterjemah. Tiada alat atau bahan seperti kamus yang dapat membantu mereka memahami sesuatu bahasa. Sekiranya tiada dalam kalangan mereka yang mahir dwi bahasa maka mereka akan menunggu lama sehinggalah muncul mereka yang berkemampuan. Kelewatan ini tidak memberi kesan yang ketara pada masa itu kepada masyarakat kerana keadaan kehidupan mereka yang tidak global.

Selepas kamus yang lebih tersusun dicipta untuk kegunaan masyarakat secara meluas pada kurun ke lapan Masihi (al-'Ilwani, 2007; Yaakub, 1985), maklumat dan ilmu pengetahuan mampu disaring dan dipindahkan secara tersusun, sama ada oleh penterjemah amatur ataupun individu yang kurang berkemahiran dalam dwi bahasa. Proses pencarian makna perkataan dalam kamus dilakukan secara manual iaitu dengan membuka satu persatu helaian kamus. Sekiranya kamus itu tebal maka pencarian semakin lambat kecuali mereka yang sudah berpengalaman. Proses ini sudah dikira konvensional kerana teknologi sudah mengambil alih peranan kamus. Proses teknologi terkini menawarkan kepada pengguna dengan hanya menulis atau

menyebut sesuatu perkataan melalui input mesin, kemudian mesin akan menterjemahkanya secara spontan dan hasilnya boleh didapati terus melalui bunyi suara atau tulisan di paparan skrin.

Pada awal milenium, perkakas yang membantu penterjemahan semakin berkembang, terutamanya penggunaan bantuan komputer yang dinamakan *Computer Aided Translation* (CAT) (lihat, Wan Rose Eliza, 2013). Kemudian kamus disimpan dalam perisian komputer, bukan sekadar dalam bentuk kosa kata, tetapi dalam bentuk frasa dan ayat. Perisian ini dikembang dan dibangunkan sehingga boleh menterjemah frasa dan ayat yang tertentu. Ia yang dinamakan sebagai mesin terjemahan (MT). Data daripada kosa kata, frasa dan ayat yang disimpan ini dinamakan korpus. Data ini boleh dicapai di mana sahaja dengan perkembangan pengunaan internet.

Dalam dunia penterjemahan, semakin banyak data-data bahasa disimpan dalam perisian komputer semakin banyak ia memberi kesan kepada kelancaran dan kecepatan proses terjemahan, dalam masa yang sama membantu penterjemah menjadi lebih produktif (Hutchins, 2001). Kecanggihan dalam memproses data-data besar menjadi pemacu kepada penguasan bidang komunikasi dan maklumat. Dengan teknologi inilah yang menyebabkan bahasa Eropah menjadi bahasa yang menonjol. Teknologi membantu empayar Eropah mengembangkan bahasa dan budaya mereka, termasuklah bahasa Inggeris dan Perancis. Amerika Syarikat sebagai satu-satunya negara kuasa besar menjadikan Inggeris sebagai 'lingua franca' dalam dunia moden. Kekuatan negara bukan penyebab perkembangan sesuatu bahasa itu, tetapi dengan perkembangan teknologi mesin perterjemahan yang digunakan (Yang, 2009).

Maklumat yang terjemahkan dengan pantas memberi input besar kepada negara. Negara yang mampu menyaingi kemajuan ini, merekalah yang akan menguasai dunia. Negara Barat dan negara maju seperti Amerika Syarikat, Eropah, China dan Jepun sentiasa berlumba-lumba membina dan mengembangkan penciptaan mesin terjemahan ini kerana melihat kepentingan dan kesannya dalam komunikasi dunia dan pembangunan ilmu sejagat.

1.2 Latar Belakang Kajian

Bahagian ini akan membicarakan sejarah pengunaan mesin dalam penterjemahan, sistem terbaru Hibrid, perkidmatan penterjemahan yang ditawarkan, mesin terjemahan (MT), kajian teks berkomputer dan kajian korpus Arab-Melayu.

1.2.1 Menterjemah Menggunakan Mesin

Kepesatan dunia teknologi menuntut manusia menghasil terjemahan secepat dan selari dengan produk baharu yang dikeluarkan. Tenaga manusia tidak mampu untuk mengejar kecepatan teknologi melainkan terjemahan itu sendiri dibantu dengan

teknologi. Cadangan untuk menggabungkan tenaga mesin dengan tenaga manusia adalah satu cadangan yang sangat praktikal. Mesin dicipta untuk menterjemah teks dan manusia sebagai editornya. Peratus untuk mendapat ketepatan tinggi dan kualiti dalam penterjemahan adalah lebih banyak.

O'Brien (2012, ms.101) menyatakan sehingga kini MT telah berjaya menghasilkan terjemahan yang baik dalam beberapa tugasan yang tertentu tetapi masih belum boleh menggantikan manusia untuk menterjemah teks sastera, ada makna yang tersirat tidak akan dapat difahami oleh komputer. Mesin hanya mampu menterjemahkan maksud perbendaharaan kata dan merekodkan perubahan tatabahasa, tetapi manusia mampu menterjemahkan makna dan memahami struktur dokumen itu. Secara umumnya mesin masih belum boleh menandingi akal secanggih manusia (Vitek, 2011). Namun Krikke (2006, ms.4) menolak dakwaan Vitek dengan menyatakan bahawa bahawa sistem hibrid yang diperkenalkan di Amerika Syarikat telah membuktikan bahawa mesin boleh menterjemah. Sistem hibrid dianggap jalan penyelesaian kepada MT yang sistemnya tidak mampu memahami bahasa manusia. Ia satu kaedah mengabungkan mesin dengan data-data statistik terjemahan manusia. MT menyimpan data-data terjemahan dalam multi bahasa dan memprosesnya berdasarkan logik komputer. Walaupun begitu manusia sebagai penterjemah yang mahir dan berkelayakan tetap diperlukan untuk mengesahkan kesahihan terjemahan MT hibrid ini. Mesin ini tidak bertujuan menggantikan manusia tetapi sebenarnya ia dapat meningkatkan keupayaan manusia, penjimatan masa dan produktiviti (Vitek, 2011, ms.7).

1.2.2 Perkembangan Sistem Hibrid

Sistem hibrid ini sudah berkembang semenjak 1994 dimulakan oleh *IBM Translation Manager* untuk menangani masalah kualiti mesin terjemahan. Kemudian sistem ini berkembang dengan pelbagai pendekatan dan kaedah supaya mesin terjemahan tidak hanya berpandu kepada logik algoritma sahaja tetapi digabungkan dengan pelbagai kaedah antaranya gabungan data memori terjemahan, terjemahan komputer berasaskan kes yang dinamakan Terjemahan Mesin Strategi Interaktif Hibrid (IHSMT), terjemahan komputer berasaskan konteks, terjemahan komputer berasaskan korpus dan terjemahan komputer berasaskan ingatan. (Wai-Chan, 2012, ms34).

Malaysia masih ketinggalan dalam bidang terjemahan berkomputer berbanding Amerika Syarikat dan Jepun. Ini kerana bahasa Melayu antara bahasa-bahasa yang tidak dimasukkan dalam kajian dan data-data MT yang dibina (Hutchins, 2009). Oleh yang demikian, apa yang Malaysia boleh lakukan ialah memanfaatkan peluang dengan memberi kerjasama melalui kajian antara negara-negara tersebut. Contohnya kajian kerjasama antara universiti-universiti dari Jepun, China, Korea dan Malaysia pada tahun 2002 yang diketuai oleh Profesor Toru Ishida dari Universiti Kyoto (Ishida, 2012).

Kaedah yang diperkenalkan ialah menggunakan bahasa Inggeris sebagai bahasa perantaraan. Sekiranya maklumat dalam bahasa Melayu hendak dihantar ke Jepun, maklumat tersebut akan diterjemahkan ke bahasa Inggeris terlebih dahulu. Kemudian diterjemahkan ke dalam bahasa Jepun berdasarkan teks terjemahan bahasa Inggeris tersebut dan begitulah sebaliknya, (Krikke, 2006). Kaedah ini dikenali sebagai terjemahan pihak ketiga, terjemahan ini tidak menjamin ketepatan terjemahan kerana berpandukan struktur bahasa Inggeris sedangkan struktur bahasa Melayu adalah berbeza.

Pembinaan korpus ini terlalu penting bagi sesebuah negara, sebagaimana yang disarankan oleh Normaziah (2008), membina sistem teknologi bahasa dan mengembangkan pengetahuan teknologi kepada pengguna bahasa Melayu akan mengangkat maruah bangsa dan mengekalkan jati diri bahasa.

1.2.3 Khidmat Penterjemahan Mesin

Penciptaan mesin terjemahan (MT), bukan sebenarnya untuk menggantikan kepakaran manusia dalam menterjemah. Kepakaran manusia tetap diperlukan untuk mengesahkan mutu terjemahan (Tate, 2008). MT dicipta untuk membantu manusia sebagaimana yang dinyatakan oleh Tim Cavalier (2001) bahawa MT boleh mengurangkan bebanan kerja penterjemahan. Penterjemah hanya perlu memperbaiki draf yang telah diterjemahkan oleh MT. Draf pertama yang diproses biasanya dapat mempercepatkan *post-editing* untuk fasa kedua. Kos menjadi yang lebih rendah dan terjemahan dapat disiapkan tepat pada masanya.

Persaingan antara kuasa-kuasa besar antara penyebab manusia berlumba-lumba dan berusaha mencipta MT yang terbaik. Semakin baik mutu terjemahan MT, maka semakin mudah kerja-kerja penterjemahan dapat dilakukan dan semakin banyak maklumat dapat dipindahkan. Negara-negara besar seperti Amerika Syarikat, Eropah, China dan Jepun adalah negara-negara yang terkehadapan dalam pembinaan mesin terjemahan. Antara MT yang afektif penggunaannya ialah *SRI International's IraqComm Speech Translator* (MT IraqComm) yang dicipta oleh Institut Antarabangsa SRI yang berpengkalan di Amerika Syarikat (W. Frandsen dan rakan-rakan, 2010).

MT *IraqComm* ini, direka untuk menterjemah perbualan antara penutur berbahasa Arab dalam dialek Iraq dan penutur Inggeris. Komponen pertuturan ialah sebuah mikrofon dengan pembesar suara yang jelas bunyinya. Penutur akan bercakap melalui mikrofon yang disambung kepada MT *IraqComm*, MT ini kemudiannya memproses suara yang diterima dan menterjemahkan kepada bahasa yang dikehendaki. Hasil terjemahan diproses keluar melalui pembesar suara. Sistem korpus pada mesin ini menyimpan berpuluh ribu kosa kata dialek Iraq dan bahasa Inggeris yang berkaitan dengan peperangan, perkhidmatan majlis perbandaran, asas-asas perubatan dan pengambilan baharu kakitangan awam (Michael W. Frandsen dan

rakan-rakan, 2010). Ia mula diguna pakai pada tahun 2006 oleh tentera Amerika Syarikat yang beroperasi di Iraq.

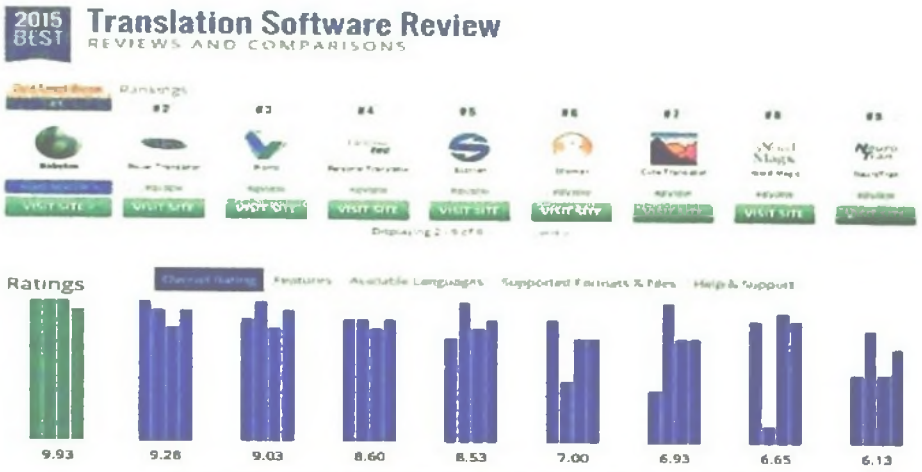
MT *IraqComm* ini terus berkembang dan diperbaik mutu terjemahannya. Kajian Kristin Precoda dan rakan-rakan (2007) terhadap ketepatan MT *IraqComm* dalam menterjemah, mendapati kesalahan bagi mengesan suara penutur Inggeris ialah sebanyak 6-15%, manakala kesalahan mengesan penutur dialek Iraq sebanyak 18-30%. Setelah dilakukan beberapa pembaikan, MT *IraqComm* telah menunjukkan perkembangan yang positif. Menurut Murat Akbacak dan rakan-rakan (2009), penilaian dari sudut ketepatan dan kelajuan menterjemah mesin ini telah meningkat sebanyak 6% hingga 8% pada tahun 2008.

MT *IraqComm* ialah mesin terjemahan bentuk baharu untuk generasi yang baharu. Ia sudah terkehadapan dalam mengembangkan teknologi alat bantu penterjemahan. Dengan bahasa yang mudah, ia adalah kamus bertutur dan boleh menggantikan sebahagian kerja-kerja jurubahasa. Adalah dijangkakan dalam beberapa tahun akan datang peranan jurubahasa akan berkurangan dengan sebab penciptaan MT *IraqComm* ini. Fakta ini sendiri telah dibuktikan oleh beberapa kajian yang dilakukan oleh beberapa syarikat terhadap kemajuan MT dalam menterjemah. Steven Klein yang juga pengarah syarikat Context-Based Machine Translation (CBMT) di New York, mengumumkan bahawa syarikatnya telah berjaya menjalankan uji kaji terhadap algoritma baharu membuktikan bahawa MT boleh mengantikan penterjemah manusia (Jan Krikke dan Benjamin Alfonsi, 2006). Juga tiada siapa dapat menafikan bahawa mesin boleh menterjemah dengan memuaskan dari segi mutu dan jumlah dengan syarat sumber telah disesuaikan dengan algoritma sistem dan boleh dikemas kini dengan mudah (Gouadec, 2015).

1.2.3.1 Mesin Terjemahan Dari Sumber Tertutup

Terdapat dua jenis perisian MT yang ditawarkan kepada umum. Perisian tertutup yang dikhaskan untuk dijual dengan harga yang tertentu dan perisian terbuka yang ditawarkan penggunaannya secara percuma. Perisian tertutup kebanyakannya dibeli oleh syarikat-syarikat besar yang memerlukan penjimatan masa yang cekap dan pantas supaya produk-produk dapat dipasarkan ke serata dunia dalam masa yang telah ditetapkan. Segala istilah dan olahan ayat telah diprogram dalam perisian secara tersusun dalam pelbagai bahasa dan pihak syarikat hanya perlu menterjemahkan tek-teks produk menggunakan perisian ini.

Perisian ini juga dibeli oleh badan-badan, instisusi-institusi atau individu yang sifat kerja mereka adalah global seperti Institut Terjemahan dan Buku Malaysia (ITBM). Berdasarkan laporan *Translation Software Review* pada tahun 2015, sembilan perisian berikut sebagaimana dalam Rajah 1.1 menduduki carta terbaik berdasarkan mutu terjemahan dan ciri-ciri yang ditawarkan.



(http://translation-software-review.toptenreviews.com, 2015)

Rajah 1.1: Carta perisian terjemahan 2015

Turutan perisian terjemahan tertutup berdasarkan penilaian 2015 ialah:

1. Babylon Translation
2. Power Translator
3. Promt Standard
4. Personal Translator
5. Systran
6. Idomax
7. Cute Translator
8. Word Magic
9. NeuroTran

Adapun Perisian yang menawarkan terjemahan daripada bahasa Inggeris ke bahasa Arab ialah Babylon Translation, Promt Standard, Systran, Cute Translator dan NeuroTran, tetapi tiada yang menawarkan penterjemahan Arab-Melayu atau sebaliknya.

1.2.3.2 Mesin Terjemahan Dari Sumber Terbuka

Masyarakat awam khususnya pelajar dan pengkaji bahasa hanya menumpukan kepada penggunaan perisian terbuka yang ditawarkan atas talian secara percuma disebabkan kewangan yang terhad. Antara perisian terbuka yang mudah dilayari dan dimuat-turun ialah seperti berikut sebagaimana yang dinyatakan oleh Rusli (2009) dan Aiken et. Al (2011):

1. Google translate
2. Bing

3. Freetranslation
4. ImTranslator
5. WorldLingo
6. Systran
7. Altavista Babelfish
8. Latin Dictionary

Walaupun perisian terbuka ini mudah dan cepat. Ia tetap memberi masalah kepada anda apabila menterjemah teks yang panjang. *Google Translate* adalah perisian yang terbaik setakat ini, masih lagi memiliki kelemahan yang ketara yang berkait dengan perubahan strukur bahasa dan pemilihan kata terjemahan yang tepat khususnya teks Arab-Melayu atau Melayu-Arab. Kajian-kajian yang lepas banyak membuktikan mutu terjemahan Google masih perlu ditingkatkan dari semasa ke semasa. Kajian Farah Hanan & Muhammad Fauzi Jumingan (2012), membuktikan terdapat kesalahan dalam terjemahan kolokasi. Kajian Khadijah Sjahrony dan Maheram Ahmad (2012) mendapati kesalahan dari sudut frasa idiomatik. Begitu juga kajian Fauzi Azmi dan Saiful Johari (2011) mendapati kesalahan terjemahan Arab-Melayu dari sudut sintaksis.

1.2.4 Mesin Terjemahan Khusus Arab-Melayu

Sehingga kini belum ada usaha secara kolektif atau usaha peribadi untuk membina mesin terjemahan Arab-Melayu yang khusus. Usaha yang wujud adalah tertumpu kepada pembinaan MT Inggeris-Melayu. Antaranya yang telah dicipta oleh Institut Antarabangsa SRI yang diberi nama MalayComm. Ia adalah hasil daripada kajian yang dilakukan terhadap MT IraqComm itu sendiri (W. Frandsen dan rakan-rakan, 2010). Kemudian terdapat usaha-usaha dari universiti-universiti tempatan dalam menilai dan membina korpus Inggeris-Melayu. Korpus adalah asas dalam membina sistem mesin penterjemahan, sistem korpus yang baik akan memberikan natijah terjemahan mesin yang baik. Antaranya kajian Tengku Sepora Tengku Mahadi dan rakan-rakan (2010) dari USM dan kajian Rogayah A. Razak dan rakan-rakan (2008) dari Universiti Kebangsaan Malaysia (UKM), Universiti Teknologi Malaysia (UTM) dan Universiti Malaya (UM).

Dalam kajian ini, pengkaji tidak akan menumpukan kepada bagaimana sistem mesin terjemahan itu berfungsi berdasarkan algoritma tertentu. Tetapi kajian ini akan menumpukan kepada sistem korpus Arab-Melayu dari segi bagaimana ia dapat digunakan dalam menghasilkan sesuatu terjemahan yang terbaik. Ia adalah sebagai tapak mula kepada membina mesin terjemahan Arab-Melayu yang berasaskan komputer.

1.2.5 Kajian Teks Berkomputer

Teknologi komputer telah memberi sumbangan besar dalam pengajaran dan pembelajaran di Malaysia, begitu juga kepada membina sistem penyusunan kamus dan mesin terjemahan supaya menjadi lebih cepat dan mudah. Menurut Zaharin Yusoff (2007), "mesin terjemahan boleh mempercepatkan kerja penterjemahan dan menyahkan terjemahan-terjemahan yang berulang atau sekurang-kurang dapat menjadikan terjemahan itu lebih konsisten" (ms.66). Dengan perkembangan teknologi terkini, kamus tidak lagi diterbitkan sekadar dalam bentuk buku, tetapi telah muncul dalam bentuk data-data yang disusun dalam perisian komputer. Perisian komputer dapat menganalisis data dan memproses maklumat secara pantas bagi membantu ahli bahasa, pelajar dan penterjemah dalam bidang kajian mereka.

Sehubungan itu, komputer boleh dimanfaatkan dalam konteks menyusun prasa dan ayat yang lebih lengkap dan bermutu. Proses menterjemah sukar dan memakan masa yang lama jika tidak menggunakan komputer kerana kerja-kerja mencatat maklumat di atas kertas, menyemak kesalahan dan mengatur huruf terpaksa dibuat secara manual. Begitulah juga kaitannya, dengan membina data-data dalam bahasa Arab untuk digunakan dalam proses pembelajaran dan penterjemahan di Malaysia. Mengikut Rahimi Md. Saad dan Zawawi Ismail (2005), iaitu pada sepuluh tahun yang lepas, "konsep e-pembelajaran bahasa Arab di Malaysia masih berada pada tahap awal. Menjayakannya adalah suatu cabaran, ...banyak kekurangan yang perlu diperbaiki"(ms.12). Kini antara usaha yang telah berjaya dilakukan ialah membangunkan e-pembelajaran *EZ-Arabic* khusus untuk pelajar sekolah rendah oleh Muhammad Sabri Sahrir, Mohd Firdaus Yahaya & Mohd Shahrizal Nasir (2013, 2012).

Sebelum wujudnya komputer, ahli bahasa terpaksa menganalisis teks halaman demi halaman untuk mendapatkan contoh-contoh penggunaan bahasa yang membantu dalam proses pemerian makna. Kerjanya rumit, juga sukar untuk mengesan apa jua perubahan bahasa, pengulangan sesuatu kata atau ayat dan penggunaan yang sama secara serentak dalam kuantiti yang banyak.

Satu inisiatif untuk membantu mengembangkan penyelidikan bahasa, telah diusahakan oleh Dewan Bahasa dan Pustaka (DBP). Mengikut Wan Norasikin (2012), DBP mengadakan kerjasama dengan Universiti Sains Malaysia membina satu pangkalan data yang diberi nama Korpus Bahasa Melayu menggunakan perisian komputer. Tujuannya ialah untuk memberi kemudahan dalam penyelidikan bahasa, perkamusan dan peristilahan. Dengan terbinanya korpus ini, kerja-kerja penyelidikan menjadi rancak dan hasil kajian-kajiannya adalah sangat mengujakan. Banyak kos, tenaga dan masa dapat dijimatkan.

Menurut Nor Hashimah dan rakan-rakan (2009), bahawa kajian bahasa melalui korpus dapat dianalisis dengan lebih cepat serta terjamin mutu keasliannya. Menurutnya lagi, bahawa penyelidik dari Universiti Kebangsaan Malaysia (UKM)

yang mendapat dana IRPA untuk mengendalikan kajian 'Nahu Prakstis Bahasa Melayu: Data Korpus Berkomputer untuk tahun 2003-2005 telah berjaya menghasil kajian berkaitan semantik dan korpus. Semoga dengan usaha yang dimulakan oleh DBP ini menjadi pencetus dan penggerak kepada pembinaan pangkalan data korpus Arab-Melayu di Malaysia.

1.2.5.1 Kajian Korpus Bahasa

Kajian lingusitik telah berkembang dengan pesat setelah diperkenalkan komputer. Komputer telah banyak menolong penyelidik bahasa untuk memproses data-data yang dikaji. Data-data ini disimpan di dalam sistem yang dinamakan korpus. Apakah itu korpus? Berdasarkan pandangan Sinclair (2005): "Korpus bermaksud himpunan jalinan bahasa atau teks lengkap dalam format elektronik. Ia yang dipilih dan disusun mengikut kriteria linguistik yang eksplisit dan digunakan sebagai sampel sesuatu bahasa" (ms.12). Menurut Gries (2009) korpus ialah frekuensi pengulangan unsur-unsur bahasa sebagai contoh berapa banyak kekerapan unsur morfologi, perkataan atau pola nahu berlaku dalam data korpus tersebut. Biasanya dapatan ini diberi nama senarai frekuensi.

Korpus ini diberi penanda, kod dan difonmatkan secara piawai serta dapat dicapai dan diproses dengan komputer. Kualiti pembinaan korpus boleh memberi impak kepada keputusan sesuatu kajian. Ini telah dinyatakan sendiri oleh E. Morin, B. Daille, K. Takeuchi dan K. Kageura (2007) dan Sinclair (1991) bahawa kajian yang dijalankan adalah bergantung kepada kualiti korpus itu sendiri. Begitu juga, korpus yang menyimpan data yang besar turut memberi kesan kepada keputusan sesuatu kajian. Baker (1993) melihat bahawa korpus yang besar mampu menyediakan peluang yang tiada taranya kepada pengkaji-pengkaji teori dalam melakukan pemerhatian dan penelitian yang mendalam terhadap bahan seperti kajian terhadap bahasa secara umum , khusus atau apa-apa interaksi sosial yang lain.

Projek pembinaan sistem korpus berkomputer atau pangkalan data bahasa ini dimulakan oleh Universiti Brown Francis dan Kučera pada tahun 1964 (Rusli dan rakan-rakan, 2004). Kemudian ianya berkembang hinggalah terbinanya Projek Cobuild (Collins Birmingham University International Language Database) pada awal tahun 1980an dan telah mempelopori penyusunan kamus yang berasaskan data korpus. Sehingga kini korpus ini terus diperkemas dan dimuatkan dengan data-data terkini.

Menurut fakta yang dinyatakan pada tahun 2004 oleh Rusli dan rakan-rakan (2004):

> "...saiz korpus lain seperti *Birmingham Collection of English Texts* (BCET) membesar daripada 7.3 juta kata kepada 20 juta menjelang tahun 1985 diikuti dengan pangkalan korpus lain yang jauh lebih besar, seperti *British National Corpus*

(http://www.natcorp.ox.ac.uk/), dengan teks tulisan dan lisan sebesar 100 juta kata." (ms.2).

Terkini, sistem *Google Books Corpus* mengandungi 155 bilion perkataan dalam bahasa Inggeris Amerika yang direkod bermula tahun 1810 hinggalah ke tahun 2009. Selain itu, sistem *Corpus of Historical American English* (COHA) mengandungi 400 juta perkataan dan sistem *Corpus of Contemporary American English* (COCA) mengandungi lebih 425 juta perkataan. Ketiga-tiga sistem korpus ini telah dibina oleh Mark Davies (2011) dan boleh dicapai di laman web (http://corpus.byu.edu).

Di Malaysia, usaha awal pembangunan pangkalan data ialah pada tahun 1983 di bawah Projek Analisis Teks Secara Komputer yang mensasarkan data teks sebesar dua juta kata (Rusli dan rakan-rakan, 2004, ms. 2). Kemudian pada tahun 1994, DBP mula membina korpus baharu menggantikan sistem yang lama melalui projek usahasama dengan Universiti Sains Malaysia. Tujuan utama pembinaan data korpus ini untuk menyediakan prasarana penelitian yang objektif dan autentik sifatnya. Sehingga kini kiraan mutakhir data yang terkumpul melebihi 135 juta kata (DBP, 2008).

1.2.5.2 Korpus Bahasa Arab

Kajian dan usaha pembinaan data korpus bahasa Arab di Malaysia masih pada tahap permulaaan. Usaha dan inisiatif yang dilakukan oleh Shariman dan rakan-rakannya (2015) dalam membina korpus Arab di Malaysia adalah dengan usaha sendiri tanpa mendapat perbelanjaan daripada mana-mana pihak. Maklumat mengenai pembinaan korpus ini boleh dicapai di laman sesawang http://prokamus.wordpress.com yang kini mencecah 30 juta kata. Adapun pembinaan korpus Arab-Melayu di Malaysia khususnya di atas talian masih belum wujud, untuk membinanya pada peringkat awal perlu kepada sokongan tenaga pakar dan bantuan kewangan yang cukup. Pangkalan korpus luar negara seperti *British National Corpus* dan Pangkalan Data DBP dalam negeri boleh dijadikan model dalam pembinaan ini.

Korpus bahasa Arab yang ada adalah terhad. Ia banyak dipelopori oleh pengkaji-pengkaji di barat, seperti kajian *Leeds Arabic Internet* (http://corpus.leeds.ac.uk/internet.html) yang dibina oleh University of Leeds mengandungi 170 juta entri (Atwell, 2011), *Tunisian Arabic Corpus* (http://tunisiya.org) mengandungi 405,590 entri (Karen McNeil dan Miled Faiza, 2011), korpus bahasa Arab-Czech (CLARA – Corpus Linguae Arabicae) oleh Charles University (Zamanek, 2001) dan *arabiCorpus* (http://arabicorpus.byu.edu) mengandungi 30 juta entri yang diusahakan oleh Dilworth Parkinson, beliau merupakan profesor bahasa Arab di Universiti Brigham Young.

Antara projek korpus yang lain ialah, *Corpus of Contemporary Arabic (CCA)* oleh Latifa Al-Sulaiti dan Eric Atwell (2006). Korpus Arab-Belanda (*Vertaalwoordenboek Arabisch-Nederlands*) oleh Mark Van Mol. Menurut Mark (2002) Korpus yang dibina oleh beliau semenjak 1996, terdiri daripada dua jenis korpus, iaitu korpus berdasarkan perkataan mengandungi 26,000 entri dan korpus berdasarkan ayat mengandungi 4,000,000 entri yang telah ditagkan yang diambil dari lebih 1,200,000 sumber lisan yang pelbagai. Pembinaan korpus-korpus ini telah membantu mereka mengembangkan alat dan kaedah pengajaran bahasa Arab di negara mereka. Berdasarkan korpus ini Mark (2000) telah berjaya menyusun sebuah kamus Arab bernama *Arabic-Dutch/Dutch-Arabic Learner's Dictionary.*

Keperluan kepada korpus Arab-Melayu adalah sangat penting, khususnya bagi perkembangan bahasa Arab di Malaysia. Kajian yang dilakukan dapat menyumbang usaha kepada membina dan mengembangkan data korpus Arab yang mempunyai terjemahan yang selari dalam bahasa Melayu. Penerapan teori leksikografi dan analisis korpus berkomputer akan menjadikan corak penyusunan kamus mengikuti kaedah yang sistematik dan terkini. Di samping itu, ia juga dapat mewujudkan satu mesin terjemahan yang berkesan yang boleh membantu penterjemah dalam mempercepatkan usahanya.

Kajian terjemahan Arab-Melayu dan Melayu-Arab dengan menggunakan data korpus adalah satu kajian baharu, kaedah ini dipilih kerana korpus dapat menawarkan data-data khususnya penggunaan bahasa yang terkini. Kemudahan pencapaian kepada data-data korpus melalui sistem konkordans membolehkan pengkaji melihat kepenggunaan bahasa dan nilai makna yang terkini dalam sesebuah masyarakat supaya proses pemindahan makna yang berlaku dalam terjemahan adalah tepat dan relevan.

Menurut St. John (2001), konkordans ialah satu paparan baris-baris perkataan atau kombinasi perkataan dalam bentuk konteks yang dikeluarkan dari korpus teks. Pengkaji perlu mengunakan kata kunci carian kata-kata tertentu, supaya carian data yang diingini dapat dihasilkan. Dalam kata lain konkordans ialah senarai contoh item yang telah dicari, biasanya didatangkan sekali dengan konteks ayat yang lengkap untuk memudahkan kajian secara mendalam (Leech dan Fligelstone, 1992; Roberts, Al-Sulaiti & Atwell, 2006).

Berikut ialah contoh paparan pencarian perkataan yang menggunakan kata kuncinya "تسونامي". Pencarian dilakukan di pangkalan data *arabicorpus*, hasilnya dipaparkan dalam bentuk konkordans, seperti dalam jadual 1.1 berikut.

Jadual 1.1: Bentuk konkordans

sort word	10 words after	word	10 words before	subsection
1	الزواج في بريطانيا!!..	تسونامي		Thawra

Thawra	صباح يوم 26 كانون الأول خبر العاصفة والمد البحري	تسونامي	اللذي خرب المِنطَقةِ كلهَا وأحدث خرابا مدمرا وضحايا بالالف.	2
Thawra	بوش الأب وكلينتون في أندونيسيا لدعم متضرري	تسونامي	...إجتماع دولي في نيروبي لبحث الكوارث الطبيعية	3
Thawra	أسبوعان على	تسونامي	والأرض مازالت تتأرجح ...عاصفة هوجاء تضرب أوروبا الشمالية وتوقع	4
Thawra	التعاطف الدولي مع ضحايا كارثة	تسونامي	يتصاعد.. 350 مليون دولار مساعدات أميركية ومليار دولار من	5

Daripada paparan dalam jadual 1.1 di atas, jelas diperlihatkan kata kunci "تسونامي" dipaparkan di permulaan dan di tengah-tengah baris konkordans. Sistem ini memperlihatkan bagaimana sesuatu kata itu digunakan dalam ayat-ayat tertentu yang dipetik daripada teks-teks akhbar dan kemudiannya dipaparkan secara baris demi baris. Beberapa kesalahan bahasa dalam paparan di atas sengaja dikekalkan untuk menzahirkan mutu bahasa teks tersebut. Bentuk paparan jadual ini boleh dikembangkan lagi seperti mengklasifikasikan ayat-ayat tersebut berdasarkan masa atau melakukan bancian jumlah kekerapan kata kunci tersebut dalam tempoh tertentu atau dalam teks yang ditentukan sendiri oleh pengkaji.

1.2.6 Data Setanding

Pilihan kata merupakan asas yang penting dalam penterjemahan. Kesilapan memilih kata yang tepat menatijahkan kesalahan perpindahan makna yang hendak disampaikan kepada pembaca. Kemahiran dalam memilih perkataan yang setanding antara dua bahasa mesti datang dari penterjemah yang berpengalaman dan mahir terutamanya dalam menterjemah peribahasa atau *amthal* dari bahasa Arab ke bahasa Melayu (Mohd. Zaki Bin Abdul Rahman, Ahmad Arifin Bin Sapar, Maheram Ahmad, 2013). Penterjemah yang berpengalaman akan mengetahui jenis-jenis kata dan bagaimana susunan ayat dirangka dalam bahasa yang hendak diterjemahkan. Kemahiran penterjemah sangat diperlukan khususnya dalam situasi penterjemahan perkataan yang tiada setandingnya.

Istilah 'kesetandingan' dipilih oleh pengkaji sebagai istilah kajian ini yang membawa maksud *equivalence* dalam bahasa Inggeris. Ia adalah istilah yang dipakai oleh pakar-pakar penterjemah dalam memberi maksud terjemahan yang terbaik. Antara mereka yang menafsirkan terjemahan sebagai pencarian kesetandingan (equivalence) antara dua bahasa ialah Baker (2011, 1992), Catford (1996), Vinay dan Darbelnet (1995), Bell (1991), Nida dan Taber (2003), dan Newmark (1981). Pandangan mereka ini boleh dirumuskan dalam pandangan Baker bahawa kesetandingan dalam terjemahan bermaksud, teks yang diterjemah boleh menyamai teks bahasa sumber

dari semua sudut sama ada dari sudut tatabahasa, konteks ayat, makna pragmatik dan lain-lain.

Namun beberapa ahli teori seperti Jakobson (1959) dan Hervey & Higgins (1992) memustahilkan pandangan ini. Walau bagaimanapun, pengkaji tidak membahaskan perselisihan teori ini, tetapi lebih cenderung kepada pandangan Nida dan Taber (2003) yang berusaha mencari kesetandingan makna terjemahan secara natural. Pengkaji akan memfokus kepada mengaplikasi teori kesetandingan menggunakan data-data yang dianalisis melalui komputer, berdasarkan kaedah-kaedah yang akan dinyatakan dalam bab tiga.

1.2.7 Kajian Antara Dua Korpus

Kajian mencari kesetandingan makna mengguna dua data korpus bahasa Arab dan Melayu di Malaysia agak baharu dan masih dalam idea perlaksanaan. Sebenarnya, kaedah meletakkan data dwibahasa dalam keadaan selari dan serentak dengan menggunakan program komputer yang direka khas banyak memberi natijah positif dalam menyelesaikan masalah kajian linguitik dan terjemahan. Mengikut Olohan (2004), "...data korpus akan mendedahkan lebih banyak penggunaan konteks dari sudut makna dan kaedah penggunaannya. Ia membantu penterjemah memilih makna yang setanding dan yang paling hampir dengan teks asal" (ms.177). Kaedah ini merupakan kaedah yang lebih ketara kesannya jika dibandingkan dengan penggunaan kamus secara manual.

Wilkinson (2005) menjelaskan bahawa teknologi korpus atau *Computer Aided Translation Technology* (CATT) ini, boleh digambarkan sebagai kumpulan teks yang besar dalam format elektronik. Korpus elektronik ini diperkaya dengan sistem 'tag' yang diprogramkan untuk mengenal pasti teks-teks berdasarkan jenisnya, dan ini sangat bermanfaat dalam rangka membolehkan para penyelidik melaksanakan kajian linguistik yang moden. Wilkinson (2005) menambah lagi, kalau sistem itu tidak menggunakan sistem 'tag', ia tetap juga boleh membantu ahli bahasa dan penterjemah dengan menjadikannya alat yang berguna untuk meningkatkan prestasi dalam menterjemah, misalnya ia dapat membantu dalam mengesahkan atau menolak keputusan mengenai penggunaan bahasa atau mendapatkan maklumat tentang kolokasi, menguatkan pengetahuan tentang pola bahasa dan belajar bagaimana menggunakan olahan ayat baharu.

Berdasarkan Zanettin (1998), Rusli dan Norhafizah (2001) dan Kruger (2004), terdapat dua jenis korpus yang boleh dijadikan bahan kajian menggantikan kamus. Pertama, yang disebut sebagai korpus selari (*parrallel corpora*), iaitu korpus yang membandingkan teks asal dengan teks terjemahannya. Yang kedua dikenali sebagai korpus setanding dwibahasa (*comparable corpora*), iaitu korpus yang membandingkan teks dalam dua bahasa yang berlainan tetapi berkongsi topik yang sama, seperti beberapa topik akhbar dunia melaporkan berita utama berkenaan peristiwa penting dalam pelbagai bahasa (Li Shao dan Hwee Tou Ng, 2004).

Oleh itu, boleh dikatakan bahawa kajian terhadap korpus setanding adalah kajian yang mencari kesamaan makna dalam penyampaian teks antara dua bahasa. Harus diingatkan di sini bahawa ia bukanlah terjemahan, tetapi dua teks yang berlainan bahasa yang kedua-duanya berkongsi kesamaan dari segi maklumat yang hendak disampaikan. Setiap teks ditulis mengunakan olahan dan gaya bahasa ibunda penulis sendiri dan tidak dipengaruhi oleh kesan bahasa-bahasa asing. Begitulah sebaliknya pada teks bandingan yang kedua.

1.2.7.1 Contoh korpus selari:

Teks 1 ialah teks asal daripada buku Seribu Satu Malam dalam bahasa Arab yang dipetik daripada pangkalan data *arabicorpus*, manakala teks kedua ialah yang dipetik daripada teks Kisah Seribu Satu Malam terjemahan oleh Rahmani Astusi, terbitan Mizan pada tahun 2004.

Teks 1

" حُكِيَ وَاللّهُ أَعْلَمُ أَنَّهُ كَانَ فِيمَا مَضَى مِنْ قَدِيمِ الزَّمَانِ وَسَالِفِ العَصْرِ وَالأَوَانِ مَلِكٌ مِنْ مُلُوكِ سَاسَانَ بِجَزَائِرِ الهِنْدِ وَالصِّينِ صَاحِبُ جُنْدٍ وَأَعْوَانٍ وَخَدَمٌ وَحَشَمٌ لَهُ وَلَدَانِ أَحَدُهُمَا كَبِيرٌ وَالآخَرُ صَغِيرٌ وَكَانَا بَطَلَيْنِ وَكَانَ الكَبِيرُ أَفْرَسُ مِن الصَّغِيرِ وَقَدْ مَلَكَ البِلَادَ وَحَكَمَ بِالعَدْلِ بَيْنَ العِبَادِ وَأَحِبَّهُ أَهْلِ بِلَادِهِ وَمَمْلَكَتِهِ وَكَانَ اسْمُهُ المَلِكُ شَهْرَيَار وَكَانَ أَخُوهُ الصَّغِيرُ اسْمُهُ المَلِكُ شَاه زَمَان وَكَانَ مَلِكُ سَمَرْقَنْدَ العَجَمِ وَلَمْ يَزَل الأَمْرُ مُسْتَقِيمًا فِي بِلَادِهِمَا وَكُلُّ وَاحِدٍ مِنْهُمَا فِي مَمْلَكَتِهِ حَاكِمٌ عَادِلٌ فِي رَعِيَّتِهِ مُدَّةَ عِشْرِينَ سَنَةٍ وَهُمْ فِي غَايَةِ البَسْطِ وَالانْشِرَاحِ."

Teks 2

"Diceritakan – tetapi Tuhanlah yang paling mengetahui dan melihat apa yang tersembunyi dalam sejarah masyarakat dan masa yang telah lewat – bahwa dahulu kala, pada masa pemerintahan wangsa Sasaniah, di Jazirah India dan Indocina, hiduplah dua orang raja bersaudara. Yang lebih tua bernama Syahrayar dan yang paling muda bernama Syahzaman. Syahrayar adalah seorang kesatria bermartabat tinggi dan seorang pahlawan yang berani, tak terkalahkan, penuh semangat, dan keras kepala. Kekuasaannya mencapai seluruh negeri itu dan penduduknya, sehingga negeri itu setia kepadanya, dan rakyatnya mematuhinya. Syahrayar sendiri hidup dan memerintah di India dan Indocina, sementara adiknya diberi tanah Samarkand untuk dikuasainya."

Hasil perbandingan dua teks asal (Teks 1) dengan teks terjemahan ini (Teks 2), akan dapat mencetuskan banyak analisis dan kajian teks yang dilihat dari sudut linguistik dan khususnya dari sudut ilmu penterjemahan.

1.2.7.2 Contoh korpus setanding:

Teks setanding yang dibandingkan mempunyai perkongsian topik yang dapat memberikan input yang banyak dari segi penggunaan bahasa dan terjemahan, seperti dalam Teks 3 dalam bahasa Arab dan Teks 4 dalam bahasa Melayu di bawah. Teks-teks ini di petik daripada laman sesawang Akhbar Arab *Al-Arabia* dan disetandingkan dengan akhbar tempatan yang dipetik daripada laman sesawang *Utusan Melayu*. Topiknya ialah "Pindaan perlembangaan di Mesir".

Teks 3

٧٧٪ مِن ١٨ مِلْيُونِ مِصْرِيّ أَيَّدُوا تَعْدِيلَاتِ الدُّسْتُورِ فِي الاسْتِفْتَاءِ

أَيَّدَ ٧٧,٢٪ تَعْدِيلَاتِ المَوَادِ التِّسْعِ فِي الدُّسْتُورِ المِصْرِيِّ الَّتِي جَرَى الاسْتِفْتَاءُ عَلَيْهَا السَّبْتَ ١٩ مَارِسِ / آذَارَ الجَارِي فِيمَا رَفَضَهَا ٢٢,٨٪. أَعْلَنَ ذَلِكَ المُسْتَشَارُ مُحَمَّد أَحْمَد عَطِيَّة، رَئِيسُ اللَّجْنَةِ القَضَائِيَّةِ المُشْرِفَةِ عَلَى الاسْتِفْتَاءِ، فِي مُؤْتَمَرٍ صَحَافِيّ بِالقَاهِرَةِ مَسَاءَ الأَحَدِ 20-3-2011.

وَقَالَ إِنَّ إِجْمَالِي الَّذِينَ يَحِقُّ لَهُمُ التَّصْوِيتُ ٤٥ مِلْيُونِ شَخْصٍ، حَضَرَ مِنْهُمْ ١٨ مِلْيُونًا وَ٥٣٧ أَلْفًا وَ٩٥٤ شَخْصًا بِنِسْبَةِ ٤١,٩٪. وَأَضَافَ أَنَّ الَّذِينَ أَيَّدُوا التَّعْدِيلَاتِ بَلَغُوا ١٤ مِلْيُونًا وَ١٩٢ أَلْفًا وَ٥٧٧ صَوْتًا، فِيمَا رَفَضَهَا ٤ مَلَايِينَ وَ٧٤ أَلْفًا وَ١٨٧ صَوْتًا."

Teks 4

"14 juta rakyat Mesir undi pinda Perlembagaan

KAHERAH 21 Mac - ....Lebih 14 juta rakyat Mesir atau 77.2 peratus daripada mereka yang mengundi dalam pungutan suara, menyokong pindaan perlembagaan yang bertujuan untuk dijadikan panduan oleh negara ini dalam menghadapi pilihan raya Presiden dan Parlimen dalam tempoh enam bulan, kata pengerusi suruhanjaya penganjur, Mohammed Attiya. Seramai empat juta lagi

atau 22.8 peratus menolak pindaan itu, katanya pada sidang media semalam.

Sejumlah 18.5 juta orang atau 41 peratus daripada kira-kira 45 juta pengundi berdaftar keluar kelmarin untuk kali pertama menunaikan hak mereka di bawah sistem demokrasi sebenar, selepas demonstrasi 18 hari menamatkan pemerintahan autoritarian Mubarak selama 30 tahun.... AFP."

Kaedah perbandingan korpus setanding ini mampu untuk menganalisis bagaimana sesuatu situasi itu diolah dalam bahasa kedua, dan dalam masa yang sama melihat kesamaannya dalam bahasa pertama. Ia merupakan satu cara membantu penterjemah memilih kosa kata, frasa dan olahan ayat. Kemungkinan besar juga, penterjemah boleh menjadikannya sebagai bahan yang sudah diterjemah, padahal tidak berlaku pun kerja-kerja penterjemahan tersebut. Cuma, penterjemah perlu memastikan bahawa teks itu berkongsi maklumat yang sama. Kaedah akan menjadi lebih mudah, dan boleh dilakukan dengan lebih kerap sekiranya bahan-bahan ini disimpan dalam pangkalan data korpus. Antara sumber-sumber yang dijangka boleh memberi sumbangan besar dalam bidang kesetandingan ini ialah bahan-bahan yang diambil daripada akhbar-akhbar harian, novel, cerpen dan majalah-majalah.

1.2.8 Mengaplikasikan Makna Setanding dalam Terjemahan

Maksud kesetandingan dalam terjemahan secara deskriptif ialah kesetaraan antara teks sumber (TS) dengan teks terjemahan (TT) yang secara langsung berkaitan dengan satu sama lain (Hervey and Haggins, 2002). Teori ini hangat diperbincangkan dalam kalangan ahli bahasa dan penterjemah sejak lima dekad dahulu. Pelbagai pandangan pro dan kontra dikemukakan terhadap teori ini.

Teori ini menjadi masyhur apabila diperkembangkan oleh Nida dan Taber pada tahun 1969 dengan teori formal dan kesetandingan yang dinamik (dynamic equivalence). Catford (1996) mentafsirkan penterjemahan sebagai proses pertukaran teks antara satu bahasa dengan bahasa lain dengan menitik berat kesetandingan (equivalence). Kemudian dikembangkan oleh House dan Baker. Baker (1992) menengahkan takrif *equivalence* dalam bentuk yang berbeza. Di mana beliau mensyaratkan kepada kesetandingan tatabahasa, konteks ayat, makna pragmatik dan lain-lain. Vinay and Darbelnet (1995) menganggap terjemahan sebagai *equivalence-oriented study.* Mereka menyatakan bahawa *equivalence* adalah metode yang ideal dan praktikal dalam mana-mana amalan penterjemahan.

Pada akhir-akhir ini, teori *equivalence* mula dikritik dan diperdebatkan semula dalam kalangan ahli bahasa dan penterjemah, berikutan ketidaksefahaman mereka dalam maksud *equivalence.* Bagi pengkritik, adalah mustahil setiap bahasa boleh disamakan. Kerana setiap bahasa mempunyai tatabahasa, kolokasi, makna pragmatik dan budaya yang berbeza. Jakobson (1959, ms.234) dalam teori terjemahannya

melihat bahawa dalam tahap terjemahan antara bahasa, tiada di sana kesetaraan yang sempurna boleh dilakukan walaupun dengan cara terjemahan yang paling mudah. Maklumat yang disampaikan hanya dikira mencukupi dari segi makna untuk disampaikan.

Ini bermakna terjemahan setanding berlaku tidak sepenuhnya. Begitu juga Hervey & Higgins (1992) melihat tidak mungkin sesuatu terjemahan itu boleh disetarakan kerana setiap makna bahasa mempunyai makna yang lebih atau berkurang walaupun dari segi zahirnya nampak sama antara kedua bahasa. Namun pengkaji lebih bersetuju kepada pandangan bahawa kesetandingan boleh dilakukan berdasarkan pandangan Nida (2003) yang berusaha mencari kesetandingan makna terjemahan yang paling hampir. Dalam masalah ini, kajian terjemahan ini akan menggunakan teori *equivalance* sebagai dasar kajian dan mahu mengesahkan perdebatan ini melalui terjemahan Arab-Melayu berkomputer. Melalui kaedah terjemahan korpus setanding Arab-Melayu ini, pengkaji mengharapkan dapat memberi beberapa keputusan yang positif dalam mengukuhkan teori setanding dalam ilmu penterjemahan Arab-Melayu.

Dalam konteks pencarian makna setanding melalui data-data setanding dalam korpus terbuka atas talian, secara tidak langsung kajian ini akan berkait rapat dengan teori bahasa natural iaitu kewujudan hubungan makna secara semula jadi dalam setiap bahasa. Dalam maksud lain, setiap bahasa memiliki cara tersendiri dalam mengolah sesuatu situasi yang berlaku atau maklumat yang hendak disampaikan. Sebagai penterjemah yang mementingkan makna dan pengolahan ayat dalam bahasa ibunda maka ciri-ciri makna dan penstrukturan ayat terjemahan secara natural mestilah menjadi dasar dalam kajian ini.

Penterjemahan secara natural ini adalah ciri-ciri terjemahan yang mementingkan kesetandingan makna. Sebagaimana yang Nida (2003, ms.159) takrifkan terjemahan dinamis itu ialah terjemahan yang bertujuan menghasilkan ungkapan secara natural dan menyesuaikan ungkapan tersebut dengan budaya setempat. Begitu juga dengan pandangan Vinay dan Darbelnet (1995), mereka menggunakan kaedah terjemahan secara bersahaja (natural) apabila menterjemah dengan mengekalkan fungsi bahasa tetapi berlainan istilah yang dikenali sebagai mengadaptasi budaya dalam pengguna bahasa sasaran. Terjemahan itu seseolah merupakan proses menyalin semula situasi yang sama seperti teks yang asal, walaupun diolah dalam bahasa yang berbeza. Prosedur inilah yang perlu dimahiri oleh penterjemah agar dapat mengekalkan kesan gaya teks bahasa sumber terhadap bahasa teks sasaran (Leonardi, 2000).

Fokus utama dalam kajian ini ialah usaha untuk mengangkat mutu terjemahan yang mempunyai nilai kesetandingan komunikatif kerana ia sangat memberi kesan kepada pembaca sebagaimana pembaca bahasa ibunda terkesan (Newmark, 1981). Sesuatu terjemahan yang mempunyai nilai komunikatif itu ialah kerana ia bersifat lancar, mudah dan jelas. Walaupun terjemahan ini adalah terjemahan bebas yang membawa maksud tidak terikat dengan teks sumber supaya terjemahan itu nampak hidup dan harmoni kepada pembaca, hakikatnya bebas itu ada batasannya. Batasan-batasan

itulah yang diletakkan oleh Popovic (1976) dan Baker (1992) agar setiap makna dan maklumat itu dipindah berdasarkan prosedur dan ada dasar-dasar yang perlu diikuti. Setiap data-data setanding itu apabila dipindahkan tidak akan berlaku lebihan atau kekurangan maklumat.

Berdasarkan kertas-kertas kerja dan penulisan terjemahan terkini, penggunaan komputer dalam kajian terjemahan masih hangat dan sedang berjalan dengan rancak di serata dunia. Gelombang penyelidikan dalam terjemahan mesin ini sebenarnya, secara kebetulan datang bersama kerancakkan kajian teori *equivalence*. Oleh yang demikian, kajian ini dikira mengambil peluang yang ada dalam rangka meneruskan usaha-usaha terjemahan berkomputer, khususnya pembinaan pengkalan data korpus Arab-Melayu di Malaysia.

1.2.9 Keperluan kepada Pembinaan Korpus Arab-Melayu

Menurut Wan Hashim Wan Teh (2008, ms.3) kadar penghasilan buku terjemahan dianggarkan secara purata masih di bawah 1% daripada jumlah tahunan penerbitan buku. Anggaran ini meliputi semua judul dalam pelbagai bahasa. Sekiranya anggaran ini dinisbahkan kepada judul bahasa Arab nescaya puratanya adalah tersangat kecil. Berdasarkan Abdullah Hassan (2009) juga, kekurangan buku-buku terjemahan ke bahasa Melayu, adalah disebabkan kekurangan kepakaran, dana dan tenaga pakar. Kenyataan beliau ini, meliputi fakta daripada semua bahasa termasuklah bahasa Arab.

Rata-rata buku-buku terjemahan Arab-Melayu yang dipasarkan di Malaysia kebanyakkannya adalah terjemahan Indonesia. Adapun yang diterjemah ke dalam bahasa Melayu tidak banyak. Kita tidak boleh meletakkan usaha terjemahan ini kepada DBP dan ITNM sahaja, hasil terjemahan institusi ini adalah terbaik, namun ia terhad bilangannya. Malaysia perlu mencapai sasaran seperti Taiwan yang menerbitkan buku 38,000 judul setahun, 8,000 judul daripadanya adalah buku terjemahan (lihat Abdullah Hassan, 2009:4). Bagaimana misi terjemahan ini dapat disetandingkan dengan Taiwan, sedangkan ITNM hanya mensasarkan 300 judul terjemahan sahaja untuk tahun 2014 (Mohd. Khair Ngadiron, 2009:26). Bagaimana pula dengan institusi-institusi, syarikat-syarikat dan individu-individu yang lain? Berapakah sasaran mereka?

Oleh kerana itu, kerja-kerja membina korpus penterjemahan Arab-Melayu di Malaysia pada masa kini sudah menjadi tuntutan utama. Kelemahan kita dalam menceburi bidang korpus ini, akan menyebabkan kita ketinggalan dalam bidang penyelidikan bahasa Arab yang menggunakan teknologi komputer, sedangkan negara seperti Amerika Syarikat (W. Frandsen dan rakan-rakan, 2010), Belanda (Mark, 2002), Britain (Atwell, 2011) dan Czechoslovakia (Zamanek, 2001) telah pun berjaya menghasilkan penerbitan kamus dan mencipta mesin terjemahan yang sangat berkesan dalam bahasa Arab ke bahasa ibunda mereka.

Faktor yang menyebabkan usaha ini perlu disegerakan ialah kerana ia adalah mesin yang menjadi alat bantuan yang efektif dalam kerja-kerja penterjemahan Arab-Melayu dan Melayu-Arab di Malaysia. Mesin ini menyimpan segala kosa kata, istilah, prasa, ayat, makalah, cerita, laporan dan sebagainya yang selari antara dua bahasa. Mesin-mesin yang ada ditawarkan di laman sesawang adalah sangat tidak membantu dan perlu kepada penambah baikan yang banyak. Gunilla dan Margaret (2008) menjelaskan bahawa, kajian bahasa mengunakan korpus pada dekad 80-an telah menampakkan potensi dalam teks terjemahan, khususnya, teks terjemahan novel sastera. Kerja berdasarkan mesin lebih mudah dilakukan berbanding kerja secara manual. Kajian dalam mencari keselarian makna antara dua teks asal dan terjemahan dengan mengggunakan korpus merupakan satu keadah yang perlu dikembangkan dalam kajian terjemahan Arab-Melayu. Usaha individu yang dimulakan oleh Shariman dalam dan rakan-rakan dalam http://prokamus.wordpress.com perlu disokong sepenuhnya dan dikembangkan supaya korpus Arab-Melayu akan menjadi kenyataan dalam masa terdekat.

1.3 Penyataan Masalah

Terdapat dua permasalahan yang perlu diberi keutamaan dalam bidang penterjemahan Arab-Melayu, pertamanya yang berkaitan dengan penggunaan teknologi alat bantu menterjemah terutamanya *google translation* dan kedua penerbitan buku terjemahan daripada bahasa Arab ke bahasa Melayu. Isu yang timbul dalam penggunaan mesin penterjemahan ini ialah, *google translation* masih tidak mampu untuk memberi khidmat yang terbaik dalam menghasilkan mutu terjemahan yang dikehendaki. Terjemahan yang dihasilkan menggunakan *google translation* meragukan sehingga menyebabkan ramai pelajar universiti terperangkap sekali dalam kesalahan tersebut.

Kajian mengenai masalah kebolehterjemahan mesin ini telah banyak dilakukan oleh ahli akademik dan pengkaji bahasa. Nur Faezah Mohd Ayob dan Hasnah Mohamad (2015, ms.232) menyatakan bahawa mesin *google translation* hanya berjaya menterjemah 31.6 peratus daripada keseluruhan teks teknikal. Khadijah Sjahrony dan Maheram Ahmad (2012, ms. 107) menilai bahawa penterjemahan manusia lebih baik manakala mesin terjemahan *Google* terbatas dengan korpus yang terhad dan bergantung kepada bank perkataan yang ada. Mohamad Nor Amin & Naimah (2011) merumuskan bahwa mesin terjemahan *Google* tidak mampu menterjemah perkataan bahasa Arab yang mempunyai pelbagai makna ke bahasa Melayu. Manakala Radiah Yusoff dan Wan Rose Eliza Abdul Rahman (2008) bersetuju bahawa komputer tidak mampu memahami konsep makna dan budaya kerana pengaruh bahasa, pengalaman dan pengetahuan penutur.

Penggunaan kamus yang bercetak tidak cukup membantu untuk menyampaikan maksud terjemahan, kerana kamus terhad dalam menterjemah perkataan dan frasa ayat, manakala contoh yang diberikan pula tidak banyak. Penggunaan kamus yang khusus perlu kepada latihan, kekurangan latihan menyebabkan penterjemah terjebak dengan masalah pemilihan perkataan yang sesuai seperti mana yang dinyatakan oleh

Law (2009), "*... they were not familiar with their tools, and found some difficulties in choosing the right words from the dictionary, and in putting the words in the translation context*" (ms.199). Anida dan Siti Nuur Roihan (2014) pula mendapati kamus dalam kajiannya tidak menawarkan contoh penggunaan kata untuk pengguna. Manakalah Intan, Nor Hashimah dan Imran (2014) menyatakan bahawa "*kebanyakan kamus yang dihasilkan didapati hanya memberi padanan secara intuisi ataupun berdasarkan maklumat yang diterjemahkan secara harfiah daripada kamus lain*" (ms.7).

Faktor yang kedua ialah, kekurangan alat bantu penterjemahan Arab-Melayu yang menjadi salah satu punca kelembapan percetakan buku-buku terjemahan agama dan ilmiah di Malaysia, sebagaimana yang telah dinyatakan oleh Wan Hashim Wan Teh (2008:3). Buku terjemahan Indonesia pula menjadi laris kerana buku terjemahan Melayu tidak mendominasi pasaran (Muhammad Bukhari, 2004; Ainon, 2009; Azman, Ahmad Fakrulazizi dan Azarudin, 2016). Oleh yang demikian, kerja-kerja penterjemahan Arab-Melayu di Malaysia pada masa kini adalah sangat dituntut terutama dalam membina mesin terjemahan khusus Arab-Melayu supaya proses penterjemahan itu cepat dan dapat melahirkan penterjemah profesional Arab-Melayu.

Nithya dan Joseph (2013, ms.15) menyatakan bahawa dengan membina mesin terjemahan yang baik dapat mempercepatkan proses penterjemahan dengan menyimpan data-data yang telah diterjemah, "*...This integration speeds up the translation process as the recent translations are cached.*" Dalam masa yang sama dapat menjimatkan kos dan masa pencarian istilah (Bundgaard, Christensen dan Schjoldager, 2016; Jaworski, 2013; Seljan, 2006).

Salah satu cara penyelesaian yang dicadangkan pengkaji ialah dengan membina data korpus Arab-Melayu setanding yang boleh menyimpan segala kosa kata, istilah, frasa, ayat, makalah, cerita, laporan dan sebagainya yang selari antara dua bahasa untuk digunakan sebagai alat bantu terjemahan. Walaupun ia memerlukan kepada tenaga ramai dan masa yang panjang, namun usaha dengan langkah pertama membolehkan ianya diperkembangkan dari semasa ke semasa. Kelemahan dalam menceburi bidang korpus ini, akan menyebabkan bidang penterjemahan ketinggalan jauh, sedangkan Eropah telah lebih terkehadapan dalam lapangan teknologi ini.

1.4 Objektif Kajian

Setelah melihat kepada penyataan masalah, maka pengkaji telah meletakkan sasaran kajian berdasarkan objektif berikut:

1. Mengumpul dan mengenal pasti data korpus bahasa Melayu dan bahasa Arab dari sumber terbuka untuk dijadikan bahan kajian kesetandingan terjemahan.

2. Menganalisis dan menghuraikan dapatan melalui kaedah perbandingan teori makna setanding antara bahasa Arab dengan bahasa Melayu.

3. Menyusun dapatan analisis data setanding bahasa Arab dengan bahasa Melayu secara selari berdasarkan susunan korpus setanding.

1.5 Persoalan Kajian

Soalan-soalan kajian yang perlu diutarakan berdasarkan objektif kajian ialah:

1. Adakah terdapat korpus bahasa Melayu dan korpus bahasa Arab yang boleh dijadikan bahan dalam kajian makna setanding.

2. Apakah metodologi yang dipakai dalam kajian untuk mengumpul dan mengklasifikasikan teks yang setanding antara bahasa Arab dan bahasa Melayu.

3. Bagaimana hasil kajian ini dapat membantu membina korpus dwi bahasa Arab-Melayu sebagai alat bantu penterjemahan.

1.6 Kaedah Kajian

Kaedah kajian ini ialah kaedah kualitatif iaitu penyelidikan yang perlu kepada pengamatan terhadap data-data kepustakaan secara terperinci yang terdiri daripada data-data yang ada disimpan dalam korpus elektronik secara terbuka di atas talian. Dokumen atau teks yang dikumpulkan dianalisa secara deskriptif interpretatif mengikut objektif iaitu berdasarkan prosedur yang dibina. Prosedur atau langkah-langkah menganalisis data-data korpus secara ringkasnya dapat diterangkan seperti berikut.

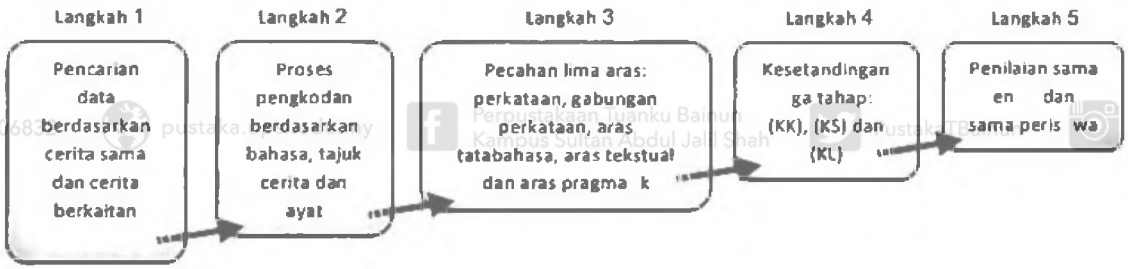
Pada permulaan dalam mencari data-data setanding Arab-Melayu dari sumber terbuka, kajian akan membuat pencarian bahan-bahan yang telah dikelompokkan dalam dua kumpulan iaitu **cerita sama** dan **cerita berkaitan** sebagaimana yang diperkenalkan oleh Aker dan rakan-rakan (2012). Data-data yang telah dikumpulkan ini, diletakkan dalam satu kumpulan data berdasarkan tajuk cerita sama dan cerita berkaitan. Data-data kajian terjemahan berkomputer ini diambil dengan menggunakan sistem enjin pencarian korpus terbuka yang ditawarkan oleh perkhidmatan atas talian dan laman sesawang seperti perisian korpus Google, Webcorp, arabiCorpus dan Pangkalan Data Bahasa Melayu DBP.

Data-data yang berjaya dikumpulkan berdasarkan cerita sama dan cerita berkaitan tersebut diletakkan selari untuk fasa pengecaman pertama dan ditagkan dengan kod-

kod berdasarkan jenis bahasa, jenis cerita dan ayat. Data-data yang tidak dikod dan yang tidak berkaitan akan dipisahkan. Ayat-ayat yang telah dikodkan ini kemudiannya dipecahkan lagi mengikut tahap kesetandingan lima aras kesetandingan (equivalence) Mona Baker (1992) iaitu aras perkataan, gabungan perkataan, aras tatabahasa, aras tekstual dan aras pragmatik. Namun kajian ini akan mengutamakan dua aras tertinggi dahulu sekiranya tiada pada tahap tersebut maka aras atau pecahan yang lebih rendah akan dikira.

Seterusnya data-data ini akan dinilai berdasarkan tiga tahap kesetandingan Guidere (2002) iaitu Kesetandingan Kuat (KK), Kesetandingan Sederhana (KS) dan Kesetandingan Lemah (KL). Di peringkat akhir, data-data yang berada dalam tiga tahap Guidere akan dirumus untuk dilihat nilai terjemahannya berdasarkan kaedah **sama entiti** dan **sama peristiwa** seperti yang diperkenalkan oleh Richard Sproat, Tao Tao dan ChengXiang Zhai (2006).

Secara ringkasnya data-data kajian ini akan melalui lima langkah penapisan seperti dalam gambar Rajah 1.2 berikut:



Rajah 1.2: Langkah-langkah penapisan data setanding Arab-Melayu

1.7 Kepentingan Kajian

Menurut Anderman & Rogers (2010) bahawa bahasa Inggeris telah berkembang sebagai bahasa dunia, di samping bahasa Rusia, Cina dan Arab. Ia memerlukan satu usaha penterjemahan antara bahasa dan dari semua bahasa ini dan ke bahasa lain supaya bahasa-bahasa tamadun besar in dapat disebarkan ke seluruh dunia.

Anderman & Rogers (2010) menjelaskan lagi bahawa, kajian bahasa mengunakan korpus pada dekad 80-an telah menampakkan potensi dalam teks terjemahan, khususnya, teks terjemahan novel sastera. Oleh yang demikian, kerja berdasarkan mesin lebih mudah dilakukan berbanding kerja secara manual. Jiang (2008) yang melakukan kajian antara dua korpus setanding Inggeris-Cina menyatakan bahawa kajian korpus selari ini menawarkan pendekatan baharu dalam pembelajaran terjemahan, bahkan ia telah menjadi satu sumber dan alat latihan yang sangat bermanfaat.

Kajian dalam mencari keselarian makna antara dua teks asal dan terjemahan dengan mengggunakan korpus merupakan satu kaedah yang perlu dikembangkan dalam kajian terjemahan Arab-Melayu kerana ia dapat menilai mutu terjemahan kata demi kata, frasa demi frasa dan ayat demi ayat secara sistematik. Kajian yang berdasarkan teori dan konteks terhadap kata dalam bahasa Arab ini diharap dapat memberikan sumbangan yang besar kepada bidang penterjemahan Arab-Melayu dan Melayu-Arab khususnya kepada:

1. Penterjemah

Kajian ini dapat membantu para penterjemah Arab-Melayu dan Melayu-Arab mengenal pasti makna yang terkini dan relevan yang diterima oleh masyarakat dalam konteks pemahaman bahasa mereka yang sebenar. Penggunaan data korpus adalah satu alternatif yang sungguh berkesan dalam mencari perluasan penggunaan bahasa yang terkini dan relevan. Ini memudah para penterjemah melakukan terjemahan yang efektif dan bermutu. Contohnya dalam kajian Abdullah Yusof (2006), beliau mendapati bahawa perkataan *letih* membawa pelbagai makna berdasarkan konteks situasi yang berbeza, hasil kajian beliau merumuskan terdapat 36 makna baharu yang diperoleh daripada analisi ayat dalam data korpus.

2. Pengguna bahasa Arab dan bahasa Melayu

Secara umumnya, dengan adanya kajian yang mendalam terhadap makna ayat yang terdapat dalam bahasa Arab dan Melayu, mampu menzahirkan kepada pelajar dan penterjemah jenis-jenis penyampaian makna yang wujud dalam sesuatu ayat. Selama ini pelajar dan penterjemah hanya memahami tafsiran makna kata melalui kamus yang terhad. Kaedah penyampaian kamus moden sekarang tidak mampu menghuraikan perkataan secara terperinci bahkan agak sedikit contoh-contoh ayat yang diberikan.

3. Ahli-Ahli bahasa dan akademik

Manfaat hasil daripada kajian ini ialah kepada penyelidik dari kalangan ahli bahasa dan akademik. Diharap usaha ini diteruskan dan dikembangkan melalui kajian yang baharu dan dengan data yang lebih besar berdasarkan keperluan dan tuntutan semasa serta kepentingan teori dalam mendasari sesuatu kajian. Hal ini akan menyuburkan displin kajian teks terjemahan yang terdapat dalam bahasa Arab dan bahasa Melayu. Seterusnya akan mewujudkan pusat atau pasukan khas bertanggungjawab setiap masa menyelidik dan mengembangkan korpus Arab-Melayu-Arab di Malaysia.

1.8 Batasan Kajian

Kajian ini adalah kajian kesetandingan makna antara bahasa Arab dengan bahasa Melayu dari sudut penterjemahan. Bahan kajian hanya akan memfokuskan kepada pencarian makna setanding antara dua bahasa ini. Kedua-dua merupakan bahasa sumber. Pencarian teks adalah berdasarkan teks umum yang terdapat dalam akhbar utama dunia. Topik utama peristiwa dunia yang menjadi berita utama di dada-dada akhbar dan majalah harian atau mingguan akan dipilih antara tahun 2010 sehingga 2014.

Topik yang terpilih dibataskan kepada tiga topik utama sahaja, setiap topik yang mengandungi artikel berita utama dalam akhbar dihadkan kepada 10 artikel bagi setiap bahasa. Bermakna 10 artikel dalam bahasa Melayu dan 10 artikel dalam bahasa Arab yang berkongsi topik yang sama. Dari data inilah kajian akan menyenaraikan perkataan, frasa, klausa dan ayat yang mempunyai kesetandingan makna antara kedua bahasa. Kemudian membuat konklusi mengenai terjemahan berdasarkan perbahasan teori dan analisis.

Sumber teks-teks ini akan diperoleh daripada sumber-sumber yang terbuka dan mudah dicapai berdasarkan kepantasan dan ciri-ciri yang ditawarkan. Pangkalan data terbuka *Google* beralamat di https://www.google.com. Pangkalan data *arabiCorpus* yang beralamat di http://arabicorpus.byu.edu/search.php, mempunyai lebih 30 juta kata dalam bahasa Arab yang diambil daripada pelbagai media cetak dari pelbagai negara Arab. Pangkalan data Webcorp di alamat http://www.webcorp.org.uk/live/ yang dibina oleh Universiti Liverpool dengan kemampuan pencarian data dari pelbagai bahasa (Olohan, 2004). Pangkalan data *Leeds Arabic Internet* beralamat di (http://corpus.leeds.ac.uk/internet.html) mengandungi 170 juta kata.

Selain daripada Google dan Webcorp, teks-teks daripada bahasa Melayu diambil daripada Pangkalan Data Korpus (PDK) yang dibangunkan oleh Dewan Bahasa dan Pustaka bersama Universiti Sains Malaysia (PDK DBP-USM) pada tahun 1993. PDK DBP-USM yang beralamat di http://sbmb.dbp.gov.my/korpusdbp dianggap pangkalan data korpus bahasa Melayu yang terulung kerana menyimpan berjuta data korpus bahasa Melayu dalam bentuk digital di Malaysia ini.

Analisis kajian ini akan berkisar dalam teori kesetandingan (equivalence) dan menggunakan metode kajian kesetandingan korpus (comparable corpora). Kajian ini juga dijangka akan memaparkan dapatan baharu yang berada diluar lingkungan atau batas linguistik dalam memerikan makna bahasa Melayu dan bahasa Arab.

1.9 Kesimpulan

Bahasa Inggeris telah berkembang sebagai bahasa dunia, di samping bahasa Rusia, Cina dan Arab. Ia memerlukan satu usaha penterjemahan dari bahasa ke bahasa lain supaya bahasa-bahasa tamadun besar ini dapat disebarkan ke seluruh dunia (J.R. Firth, 1968; Gunilla dan Margaret, 2008). Oleh itu, tidak menjadi keberatan untuk dikatakan di sini bahawa, salah satu cara merealisasikan saranan J.R. Firth, Gunilla dan Margaret ini ialah, dengan mengembeling tenaga dan usaha yang ada untuk mewujudkan satu data korpus dwi bahasa, iaitu korpus Arab-Melayu. Dibina dalam bentuk perisian dan sistem yang direka khas, manakala data-data yang dikumpulkan adalah tepat dan mudah dicapai, agar kelak, kajian-kajian bahasa yang dilakukan ini dapat dihasilkan dengan lebih efektif. Hasil daripada kajian inilah, diharap dapat membantu membuka ruang baharu kepada perkembangan teknik dan kaedah penterjemahan Arab-Melayu.