

KIM VS. KIM: KAJIAN LEKSIS BERDASARKAN ANALISIS TEKS SELARI

Rusli Abdul Ghani

(rusli@dbp.gov.my)

Norhafizah Mohamed Husin

(hafizah@dbp.gov.my)

Bahagian Penyelidikan Bahasa

Dewan Bahasa dan Pustaka

(<http://www.dbp.gov.my>)

(Dibentangkan dalam Seminar Sehari Jabatan Bahasa Melayu dan Terjemahan,
Universiti Kebangsaan Malaysia pada 7 Februari 2001)

Abstrak

Penterjemahan merupakan suatu proses pengkaryaan yang menuntut pelbagai keterampilan berdwibahasa, khususnya kemahiran untuk mentafsirkan struktur peristiwa, struktur wacana dan leksis yang terdapat dalam teks sesuatu bahasa dan kemudian mengekodkan perkara yang sepadan dalam teks bahasa yang lain. Kertas kerja ini merupakan hasil penelitian terhadap aspek leksis yang terdapat dalam dua versi 'Kim' (Rudyard Kipling), yang satu dalam bahasa asalnya Inggeris dan yang satu lagi terjemahannya dalam bahasa Melayu. Kedua-dua teks ini dianalisis dengan menggunakan perisian teks analisis dan perisian konkordans untuk mendapatkan maklumat tentang kekerapan kata dan juga kolokasi. Kajian ini merupakan satu contoh pemanfaatan teks atau korpus selari dalam usaha untuk mendapatkan wawasan linguistik (linguistic insight) tentang struktur dan fungsi bahasa, khususnya hubungan bentuk-makna dalam pelbagai bahasa.

Penghargaan

*Dengan takzimnya kami merakamkan setinggi-tinggi penghargaan kepada kumpulan penyelenggara **Project Gutenberg** kerana menyediakan teks digital untuk penyelidikan dan kepada **Michael Barlow** kerana menyebarkan, secara percuma setakat ini, versi beta ParaConc untuk tujuan pengujian dan penyelidikan.*

1.0 Pendahuluan

Penelitian teks selari¹ boleh dilaksanakan pada aras pasangan teks atau pada aras kumpulan teks (korpus) yang lebih besar dan pelbagai, dengan melibatkan dua bahasa (COMPARA²; ESPC³; INTERSECT⁴; PELCRA⁵) atau lebih (Resnik et al.⁶; Danielsson dan Ridings 1996).

Kajian teks/korpus selari ini telah menyerlahkan wawasan linguistik yang bermanfaat dalam penyelidikan dan pemprosesan bahasa tabii, terutamanya dalam penterjemahan berkomputer (sebagai contoh, Brown 1996; Brown et al. 1990; Karlgren et al. 1994; Klavans dan Tzoukermann 1995; Picchi dan Peters⁷), pembinaan leksikon (Marinai et al. 1992; Kumano dan Hirakawa 1994) dan dalam leksikografi komputeran (Hartmann 1994; Resnik dan Melamed 1997; Warwick dan Russell 1990).

Justeru itu, melihat pada spektrum kajian seluas dan serencam yang kami lakarkan di atas, kebanyakan penyelidikan teks selari mutakhir lebih tertumpu pada pemprosesan bahasa tabii yang ternyata jauh lebih teknikal. Namun, teks selari juga boleh berperanan sebagai sumber rujukan langsung kepada pengguna manusiawinya.

2.0 Kerangka Kajian

Dengan perisian yang sesuai, teks selari boleh digunakan secara langsung oleh penterjemah, penyusun kamus berbilang bahasa dan juga pengajar dan pelajar bahasa asing, khususnya untuk mendapatkan contoh-contoh terjemahan atau padanan yang autentik⁸ pada peringkat unit leksikal, frasa dan ayat, mengingat bahawa,

In considering the theoretical and practical aspects of translation, it seems wise to look to those linguistic approaches to form and meaning that are tied to actual language usage and are capable of accounting for the data coming from corpora rather than from intuition (Barlow⁹).

Penelitian ini secara umumnya tertumpu pada;

- i. pengenalan alat dan teknik pemprosesan teks selari; dan
- ii. aspek leksis¹⁰ yang berpotensi untuk pengkajian dengan teks selari.

¹ Teks selari, menurut takrifan Baker (1995) dan Sinclair (1996), terdiri daripada teks yang asalnya ditulis dalam bahasa A bersama-sama terjemahannya dalam bahasa B (bahasa C ... dan bahasa n).

² Korpus terjemahan selari bahasa Portugis-Inggeris. 24 Disember 2000 <<http://www.portugues.mct.pt/COMPARA/About.html>>

³ Korpus selari bahasa Inggeris-Swedish (Altenberg 1999).

⁴ Projek korpus selari bahasa Inggeris-Jerman (Salkie 1995).

⁵ Korpus selari bahasa Inggeris-Polish. PELCRA Corpora. Kemas kini mutakhir 1 Oktober 2000. Lodz University. 23 Disember 2000 <<http://www.uni.lodz.pl/pelcra/index.htm>>

⁶ Creating a Parallel Corpus from the "Book of 2000 Tongues." 22 Disember 2000 <<http://www.stg.brown.edu/conferences/tei10/tei10.papers/resnik.html>>

⁷ "Bilingual Reference Corpora for Translator and Translation Studies." Picchi, Eugenio dan Carol Peters. 22 Disember 2000 <<http://www.ilc.pi.cnr.it/pisystem/procedure/procedure/comparable/articolo/corpo.html>>

⁸ Dari segi pemakaiannya dalam terjemahan sebenar.

⁹ "Investigating Form and Meaning." Michael Barlow. 21 Disember 2000 <<http://www.ruf.rice.edu/~barlow/>>

¹⁰ Kami menggunakan istilah 'leksis' sebagai bermaksud "kosa kata sesuatu bahasa atau sub-bahasa, dan terdiri daripada unit/butir leksikal atau leksem" (McArthur 1992: q.v. LEXIS)

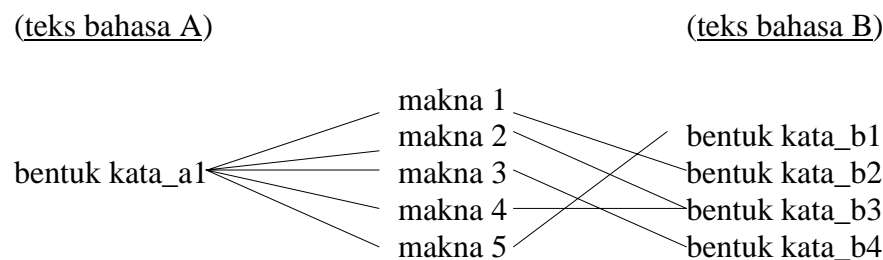
Kajian ringkas ini diharap dapat memperlihatkan manfaat perisian dan data teks selari dalam menampilkan wawasan linguistik tentang proses penterjemahan dan penyusunan kamus berbilang bahasa (tentunya dengan mengambil kira keterbatasan data¹¹ dan cakupan yang dihuraikan di bawah).

2.1 Matlamat

Salah satu unsur yang menyumbang kepada pemahaman fenomena penterjemahan ialah kajian leksikal di antara teks sumber dan teks sasaran. Kajian leksikal boleh bersifat meluas (dengan mengambil kira sebahagian besar leksem yang terdapat dalam teks) atau mendalam (dengan mengangkat beberapa butir leksikal sebagai contoh dan menggarap butiran tersebut selengkap mungkin). Kekangan kertas ini menjuruskan kami kepada kaedah yang kedua.

Matlamat khusus kami adalah untuk mengkaji hubungan bentuk–makna (atau dalam kes teks selari, lebih tepat lagi dikatakan rangkaian bentuk–makna–bentuk), dengan tumpuan terhadap butir-butir leksikal yang berpolisemi, kerana bentuk-bentuk beginilah yang sering kali menyulitkan penterjemah, penyusun kamus dan pengajar atau pelajar bahasa asing.

Secara umum, hubungan ini dapat digambarkan seperti dalam Rajah 1.0 yang berikut;



Rajah 1.0 Gambar Rajah Hubungan Bentuk–Makna

Penelitian kolokasi dan konteks penggunaan bentuk kata_a1 dalam teks bahasa A mungkin membayangkan padanannya yang paling sesuai dalam bahasa B.

Dengan menggunakan perisian ParaConc, semua kehadiran sesuatu bentuk kata dalam teks bahasa tertentu dapat dikumpulkan bersama-sama perenggan yang sepadan dalam teks terjemahannya. Dengan cara ini, penelitian dapat dilakukan dengan tertib dan tepat kerana untuk mencari semua kehadiran sesuatu butir leksikal yang terdapat dalam sesebuah teks (yang kadang-kadang ratusan banyaknya) secara manual dan kemudian merunut pula padanannya dalam perenggan-perenggan terjemahan teks tersebut bukanlah suatu kerja yang mudah.

2.2 Data

Pemilihan data untuk penelitian ini terikat sepenuhnya pada kepraktisan. Kami tidak dari awal menentukan jenis mahupun tajuk teks yang diperlukan untuk kajian kerana kami tidak berhasrat untuk bermula dengan penyediaan data digital (menaip atau

¹¹ Untuk memperoleh dapatan yang benar-benar bererti dalam penyelidikan, sebuah pangkalan korpus selari dengan teks yang banyak, pelbagai dan berpenanda (*tagged* atau *annotated*) perlu dibina terlebih dahulu.

mengimbas dan menyemak teks). Dalam hal ini kami bersikap oportunistik (berlandaskan falsafah praktis “mana ada, guna”¹²) dengan mencari data (teks bahasa Melayu dan teks bahasa Inggeris) yang sedia terakam dalam bentuk elektronik.

2.2.1 Mengapa Kim?

Pangkalan Data Korpus Bahasa Melayu yang dibangunkan oleh DBP kini mengandungi pelbagai jenis teks dengan jumlah perkataan sekitar 70 juta. Teks digital ini disimpan dalam sub-korpus tertentu berdasarkan bentuk terbitan (seperti buku, majalah dan akhbar) atau jenis karya (seperti terjemahan, puisi, drama dan sastera tradisional).

Untuk mendapatkan teks selari, kami meneliti teks yang menjuzuki sub-korpus terjemahan dan mencari teks asal bahasa Inggerisnya di Internet/WWW. Daripada senarai calon teks yang kami teliti hanya *Kim* karya Kipling sahaja tersedia sebagai data digital dalam bahasa Melayu dan juga bahasa Inggeris.

A. *Kim* versi bahasa Melayu

Kim versi bahasa Melayu (Kipling 1989) merupakan edisi terjemahan daripada teks Inggeris (Kipling 1987) dan tersedia dalam fail teks ASCII¹³ (saiznya 703 KB).

Walaupun tataletak halaman, saiz fon dan muka taip dalam fail teks ini berbeza daripada naskhah cetaknya, namun teks ini dapat dimanfaatkan kerana maklumat sedemikian tidak relevan dalam kajian leksikal.

B. *Kim* versi bahasa Inggeris

Data *Kim* versi bahasa Inggeris (fail “kimrk10.txt” bersaiz 613.63 KB) diperoleh dari laman Web Projek Gutenberg¹⁴ (Kipling, Rudyard, 1865-1936. - 2000. - Kim - Urbana, Illinois (USA): Project Gutenberg. Etext #2226. - First Release: Jun 2000 - ID:2639).

Setiap teks dalam Projek Gutenberg mempunyai kepala teks (*header*)¹⁵ yang antara lain menyentuh hal-hal hak cipta, sumbangan dana, dan juga maklumat ringkas tentang teks itu sendiri. Seperti kebanyakan teks lain dalam simpanan Projek Gutenberg, teks *Kim* juga mungkin disediakan berdasarkan “...multiple editions, all of which are in the Public Domain ...” (Kepala teks fail kimrk10.txt).

Oleh sebab teks ini mungkin berasaskan pelbagai edisi, kami terlebih dahulu membandingkannya dengan edisi *Kim* yang digunakan sebagai teks sumber untuk penterjemahan edisi bahasa Melayu. Ternyata kedua-dua teks ini pada hakikatnya serupa dan sedikit kelainan yang ada (tataletak halaman, fon dan tanda baca) tidak memudaratkan kerangka analisis kami.

¹² Tentunya dengan mengambil kira hak cipta penerbit dan pengarang masing-masing.

¹³ Singkatan bagi ‘American Standard Code for Information Interchange.’

¹⁴ Project Gutenberg. Kemas kini mutakhir 20 Disember 2000. PROMO.NET. 22 Disember 2000 <<http://www.promo.net/pg/>>. Fail dimuat turun dari <<ftp://ibiblio.org/pub/docs/books/gutenberg/etext00/kimrk10.txt>> (Versi mampatnya, kimrk10.zip – 244.12 KB, terdapat di <<ftp://ibiblio.org/pub/docs/books/gutenberg/etext00/kimrk10.zip>>).

¹⁵ Bersesuaian dengan piawaian TEI (Text Encoding Initiatives) <<http://www.uic.edu:80/orgs/tei/>> atau <<http://www.tei-c.org/>>.

2.3 Perisian

Kajian teks selari memerlukan kedua-dua teks itu diproses dan dipaparkan sepadan, sama ada ayat dengan ayat atau perenggan dengan perenggan.¹⁶ Untuk itu, kami memerlukan sebuah perisian.

Sekali lagi, kami mengandalkan Internet untuk perisian yang sesuai. Dua perisian yang berkemungkinan ialah Multiconcord¹⁷ dan ParaConc.¹⁸

2.3.1 Mengapa ParaConc?

Kami memilih ParaConc atas alasan gratis semata-mata. Buat sementara waktu, perisian ParaConc versi β disebarikan secara percuma untuk tujuan pengujian. Bagaimanapun, versi lengkapnya¹⁹ dikenakan bayaran.

2.4 Pemprosesan

Data *Kim* melalui tiga peringkat pemprosesan, iaitu pensejajaran teks, penyenaian kata dan penselarian teks, sebelum analisis leksikal sebenar dapat dilakukan.

2.4.1 Pensejajaran Teks

Pensejajaran teks merupakan peringkat yang paling lambat (dan leceh) kerana kedua-dua teks perlu dibandingkan perenggan demi perenggan. Walaupun kajian pensejajaran automatik telah banyak dilakukan (lihat sebagai contoh Gale dan Church 1993; Johansson et al. 1994; Ahrenberg et al. 1998) namun untuk pemprosesan dengan ParaConc, setiap perenggan dalam teks terjemahan perlu disejajarkan dengan perenggan teks asal secara manual.

2.4.2 Penyenaian Kata

Selain ParaConc, kami juga menggunakan perisian MATA²⁰ (*Malay Text Analysis*) untuk mendapatkan senarai kata daripada teks *Kim* asal dan terjemahannya. Senarai kata ini juga akan menunjukkan kekerapan butir-butir leksikal itu di dalam teks (Lihat Jadual 1.0).

2.4.3 Penselarian Teks

Butir leksikal yang diipilih kemudian dikutip daripada teks asal dan teks terjemahan secara serentak dengan menggunakan ParaConc.²¹

¹⁶ Dengan andaian bahawa pola biasa terjemahan ialah satu ayat diterjemahkan dengan satu ayat, ayat pendek diterjemahkan dengan ayat pendek dan ayat panjang diterjemahkan dengan ayat panjang (Church dan Gale 1991).

¹⁷ "Multiconcord: the Lingua Multilingual Parallel Concordancer for Windows. 23 Disember 2000. <http://web.bham.c.uk/johnstf/l_text.htm> (Lihat juga Bonhomme dan Romary 1995; Romary et al. 1995).

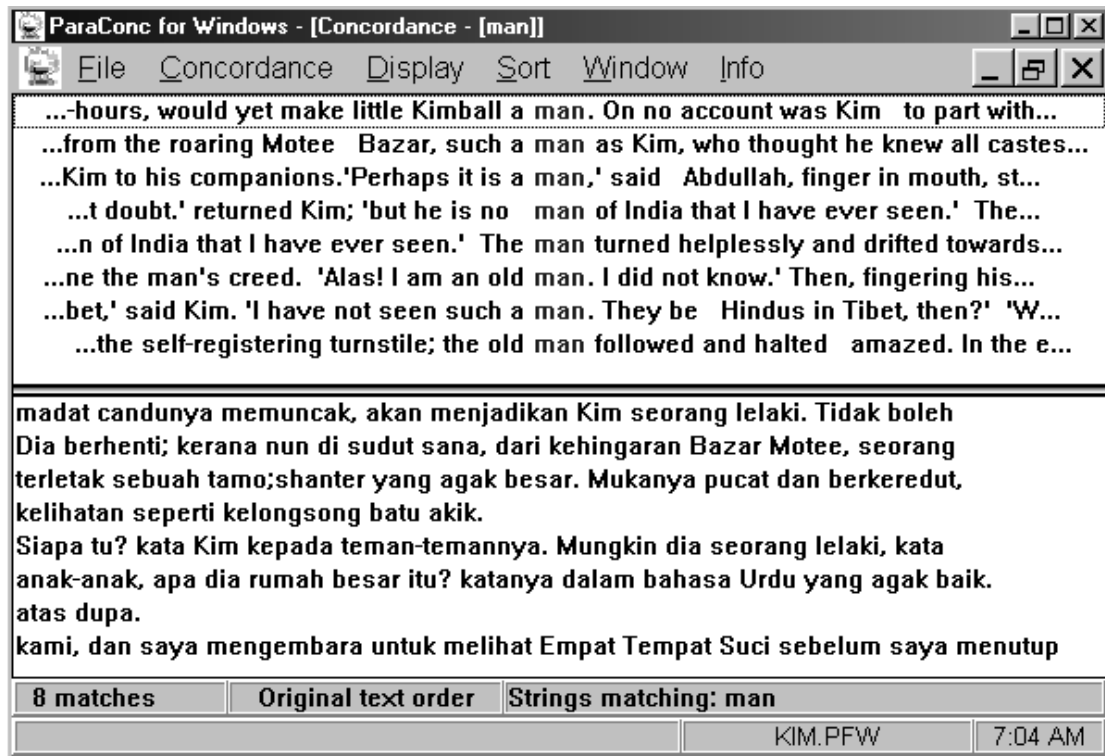
¹⁸ ParaConc for Windows Version 0.7 β © 1996 Michael Barlow. Rice University. 22 Disember 2000 <<http://www.ruf.rice.edu/~barlow/Pfw07.html>>.

¹⁹ Versi komersial untuk sistem operasi Windows terdapat di <<http://www.athel.com/>>.

²⁰ Perisian ini dibangunkan bersama-sama Sistem Korpus oleh Unit Terjemahan Melalui Komputer, USM untuk pembinaan Pangkalan Data Korpus Bahasa Melayu DBP.

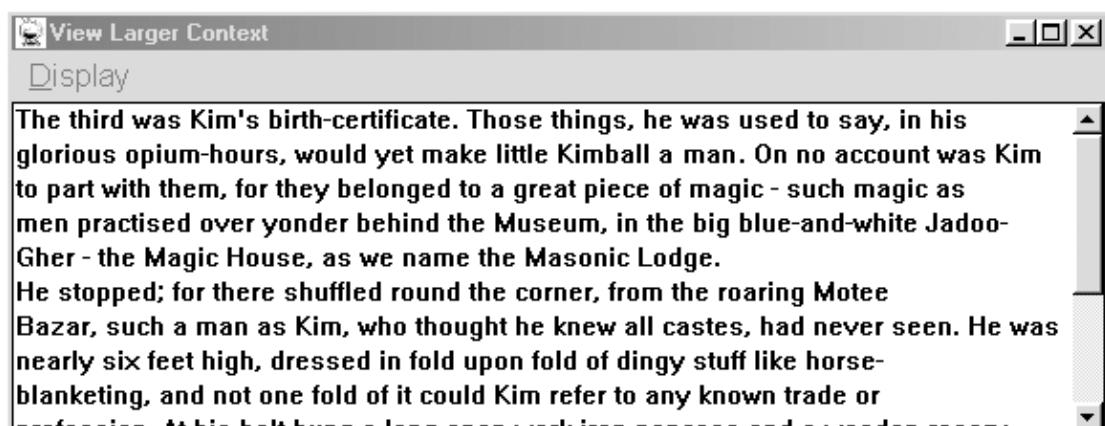
²¹ Untuk panduan penggunaan ParaConc sila lihat Barlow 1995a; Barlow 1995b; Barlow 1996.

Sebagai contoh, penggelintaran dengan kata carian 'man' menghasilkan output seperti yang ditunjukkan dalam Rajah 2.0 di bawah (dengan bahagian atas memaparkan baris-baris yang mempunyai kata carian daripada satu bahasa dan bahagian bawah memperagakan baris yang sepadan dalam bahasa yang satu lagi);

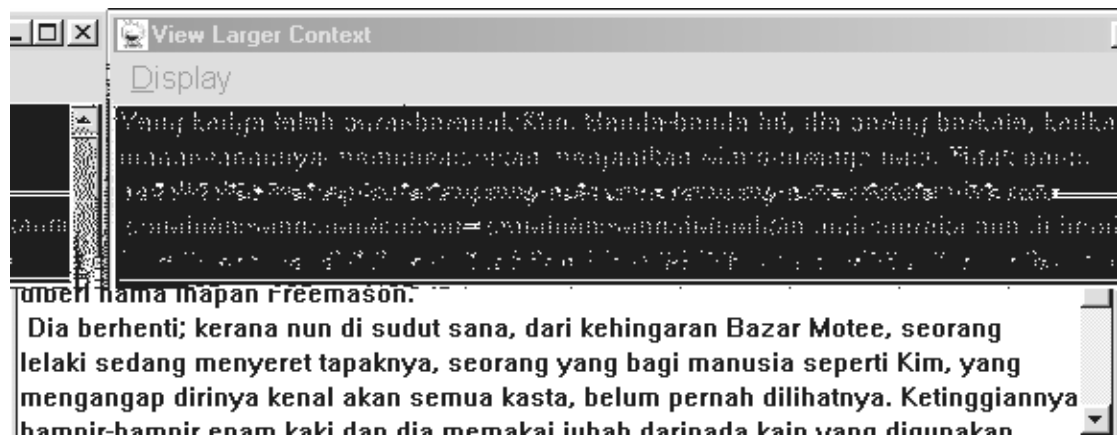


Rajah 2.0 Contoh Hasil Pemprosesan dengan Kata Carian 'man'

Setiap baris dalam paparan tersebut boleh dikembangkan untuk menunjukkan konteks penggunaan yang lebih luas bagi sesuatu butir leksikal yang diteliti. Demikian juga dengan baris yang sepadan dengannya. Sebagai contoh, baris pertama dalam Rajah 2.0 dikembangkan menjadi;



manakala baris yang sepadan dengannya dikembangkan menjadi;



Daripada kedua-dua konteks yang diperluas ini, dapatlah kajian terhadap leksikal 'man' dan padanannya dilakukan dengan lebih cepat dan berkesan.

3.0 Analisis dan Perbincangan

Dapatan daripada pemprosesan dianalisis secara umum untuk mengenal pasti butir kata yang akan dikaji selanjutnya.

3.1 Analisis Umum

Maklumat yang diperoleh daripada pemprosesan MATA merupakan maklumat statistik tentang leksis dalam kedua-dua teks *Kim* (Jadual 1.0).

Maklumat statistik tentang saiz teks, dengan jumlah keseluruhan perkataan (token) bagi *Kim* Inggeris sekitar 98,500 dan *Kim* Melayu sekitar 120,000 perkataan, tidak mempunyai implikasi yang bererti dalam kajian ini tetapi mungkin berguna untuk kajian ciri teks terjemahan (misalnya untuk menguji hipotesis bahawa teks terjemahan sering kali lebih panjang daripada teks sumber atau sebaliknya).

Bagaimanapun, butir-butir kata yang paling kerap terdapat dalam kedua-dua teks, selayang pandang pun, menimbulkan pelbagai persoalan.

Sebagai contoh, 'Kim' dalam teks asalnya muncul sebanyak 818 kali manakala dalam terjemahannya terdapat sebanyak 949 'Kim.' Ini bermakna ada pertambahan sebanyak 131 butir tersebut dalam teks terjemahan.

Hakikat ini dapat dikaji dengan menggunakan ParaConc untuk mendapatkan semua 'Kim' dalam teks bahasa Melayu dan meneliti perenggan yang sepadan dalam bahasa Inggeris.

Daripada kajian mudah ini, dapat dilihat bahawa pertambahan 'Kim' dalam teks bahasa Melayu adalah disebabkan penterjemahan *possessive pronoun* seperti yang berlaku dengan frasa 'his mother/feet/head' menjadi ibu Kim, kaki Kim dan kepala Kim. Persoalan yang lebih sukar dan menarik adalah untuk mengenal pasti bila masa 'his mother' menjadi 'ibu Kim' dan bila pula ia menjadi 'ibunya.'²²

²² Mungkin juga keadaan ini tidak ditentukan oleh leksis tetapi oleh stilistika.

Jadual 1.0 Sebahagian daripada Maklumat Leksikal dalam Teks *Kim*

Teks	Kim Versi Bahasa Inggeris		Kim Versi Bahasa Melayu	
Bilangan Perkataan (token)	98,503		120,141	
Bilangan Jenis Perkataan (taip)	10,191		9633	
Senarai Perkataan dan Kekerapan (mengikut susunan kekerapan menurun dan dengan mengabaikan kata tugas)	Kata	Kekerapan	Kata	Kekerapan
	kim	818	kim	949
	said	662	kata	599
	lama	370	lama	499
	man	356	dalai	412
	go	287	pergi	360
	old	261	sahib	298
	little	240	tahu	295
	mahbub	229	mahbub	267
	sahib	210	lelaki	259
	come	195	kuda	227
	men	191	anak	223
good	177	tua	211	

Demikian juga dengan ‘lama’ yang jumlahnya lebih banyak dalam bahasa Melayu. Penelitian kolokasi menunjukkan bahawa peristiwa ini adalah disebabkan fenomena homonimi di antara “lama_sami” dan “lama_masa.” Penelitian kolokasi selanjutnya menunjukkan bahawa padanan untuk ‘lama’ ialah ‘dalai lama.’

Aspek lexis begini perlu dikaji secara mendalam untuk mendapatkan wawasan tentang ciri-ciri penterjemahan, termasuk sifat teks terjemahan dan gaya penterjemah individu (lihat sebagai contoh Baker 1995; Baker 1999; Baker 2000).

3.2 Analisis Khusus

Daripada sekian banyak butir leksikal dalam senarai kata *Kim*, kami memilih ‘man’ untuk analisis selanjutnya kerana, daripada 25 bentuk kata yang paling kerap (setelah diketepikan kata tugas), butir ini merupakan kata yang sangat berpolisemi²³ dan boleh dijadikan contoh untuk menunjukkan hubungan bentuk–makna.

Sifat polisemi ini diimbaskan oleh padanannya dalam bahasa Melayu. Daripada sejumlah 547 ‘man/men,’ yang menjadi ‘lelaki’ hanyalah 206. Ke mana menghilangnya yang lain? Hakikat ini menunjukkan bahawa tidak semua ‘man’ itu ‘lelaki.’

Untuk meneliti fenomena ini dengan lebih lanjut, semua ‘man’ dan ‘men’ disarikan daripada teks Inggeris dan padanannya daripada teks Melayu dengan menggunakan ParaConc.

Butiran ‘man’ muncul sebanyak 356 kali dalam *Kim*, manakala ‘men’ sebanyak 191. Berdasarkan penelitian teks selari, padanan bagi lema²⁴ ‘man’ adalah seperti yang ditunjukkan dalam Jadual 2.0.

²³ Entri ‘man’ dalam *Kamus Inggeris Melayu Dewan* (DBP 1992) mempunyai 13 makna dan dua sub-makna, manakala dalam *Collins English Dictionary* (Hanks 1979) ‘man’ mempunyai 15 makna dan empat sub-makna.

²⁴ Bentuk berkanun yang dijadikan kata masukan kamus; lema ‘man’ merangkumi juga ‘men.’

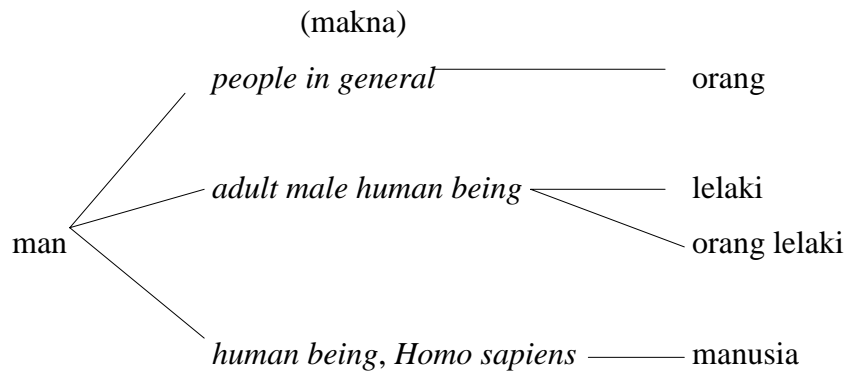
Jadual 2.0 Butir Leksikal ‘Man/Men’ dan Padanan

Butir Leksikal	Padanan	Kekerapan
man (356) men (191)	orang	364
	lelaki	129
	orang lelaki	29
	manusia	23
	gembala	1
	anak jantan	1

Hubungan ini dapat dipetakan seperti yang berikut (dengan huraian makna dipetik daripada *Kamus Inggeris Melayu Dewan* ibid.);

(bahasa Inggeris)

(bahasa Melayu)



Butiran ‘gembala’ dan ‘anak jantan’ tidak dipetakan kerana kekerapan masing-masing hanyalah satu.

Padanan ini ternyata serasi sebahagiannya dengan apa yang disajikan dalam *Kamus Inggeris Melayu Dewan Edisi Ringkas*²⁵ (Shahrezad 1996: q.v. **man**) seperti yang berikut;

man *n* **1.** *adult male human being*, (orang) lelaki: ... **2.** *human being, Homo sapiens*, manusia: ... **3.** *people in general*, orang: ... **4.** *male worker*, pekerja (lelaki); ... **5.** (*colloq*) *male adult closely associated with female person*, **a.** (husband) suami: ... **b.** (*lover, sweetheart*) kekasih, pak we; **6.** (*colloq*), (*form of address used when speaker is angry, surprised, etc*) kawan: ...

Bentuk kata berkekerapan tinggi mendukung makna-makna teras, dan hakikat ini boleh dijadikan panduan untuk pengajaran bahasa asing dan juga penyusunan kamus pelajar (sama ada dalam pemilihan kata masukan, pengutipan ayat contoh atau penyusunan makna). Dan, jika korpus selari itu cukup besar, maka bentuk-bentuk kata dengan makna pinggir juga akan timbul dalam teks dan ini dapat membantu

²⁵ Sebuah kamus penterjemah.

penterjemah untuk mendapatkan padanan bagi bentuk-bentuk yang sukar atau yang jarang-jarang digunakan.

Selain itu, penelitian kolokasi sesuatu padanan mungkin juga dapat memberikan petunjuk atau isyarat tentang makna yang didukung oleh sesuatu bentuk kata. Sebagai contoh, daripada pola kolokasi 'orang' dapat dilihat bahawa 'man' dipadankan dengan 'orang' apabila kata itu merupakan unsur kata majmuk. Misalnya, 'orang putih' merupakan padanan bagi 'white man' (tidak terdapat kata majmuk 'lelaki/manusia putih') manakala 'orang bukit' merupakan padanan bagi 'hillfolk/ hill people' (tidak terdapat 'lelaki/manusia bukit').

Butir leksikal 'man' dipadankan dengan 'lelaki' atau 'orang lelaki' apabila makna jantina mahu diketengahkan; misalnya, 'lelaki Inggeris' lebih spesifik daripada 'orang Inggeris.' Namun demikian kolokasi tidak menunjukkan sebarang pola yang ketara. Lagi pun, 'lelaki' sering kali digunakan secara bertukar ganti dengan 'orang,' misalnya dalam frasa 'lelaki/orang tua' atau 'lelaki/orang alim.'

Padanan 'manusia' untuk 'man' sering kali berkait dengan bentuk yang sejagat dan lebih luas seperti 'alam,' 'ragam,' 'sifat' dan 'dunia.' Namun pola kolokasi yang lebih jelas tidak kelihatan melainkan data teks selari ini diperbesar saiz dan diperluas skopnya.

4.0 Kesimpulan

Walaupun relatif kecil, teks selari *Kim* telah dapat menampilkan pelbagai kemungkinan kajian dan memperlihatkan beberapa fenomena linguistik.

Korpus teks selari yang besar tentu sekali lebih bermanfaat sebagai sumber rujukan dalam penyelidikan linguistik, pengajaran bahasa, perkamusan dan juga penterjemahan, baik manusiawi mahupun mesin.

Data selari yang pelbagai dan representatif merupakan tunjang empirisme dan asas objektiviti dalam sebarang kajian linguistik bandingan atau kontrasif, dan wawasan linguistik yang terserlah daripada penelitian data ini sering kali berbeza daripada apa yang diperolehi daripada korpus ekabahasa atau intuisi.

Rujukan

- Ahrenberg, L., Andersson, M., dan Merkel, M. 1998. "A Simple Hybrid Aligner for Generating Lexical Correspondences in Parallel Texts." *Proceedings of 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, Montréal, Canada, 10-14 Ogos 1998, 29-35.
- Altenberg, Bengt, Karin Aijmer dan Mikael Svensson 1999. "Manual: The English-Swedish Parallel Corpus (ESPC)." 29 Disember 2000
<<http://www.englund.lu.se/research/corpus/corpus/esp.html>>
- Baker, M. 1995. "Corpora in Translation Studies: An Overview and Some Suggestions for Future Research." *Target* 7(2): 223-243.
- 1999. 'The Role of Corpora in Investigating the Linguistic Behaviour of Professional Translators.' *International Journal of Corpus Linguistics* 4(2): 281-298.
- 2000. "Linguistic Perspectives on Translation.", dalam Peter France (ed.) *The Oxford Guide to Literature in English Translation*, Oxford: Oxford University Press, 20-26.

- Barlow, M. 1995a. *A Guide to ParaConc*. Athelstan. Lihat juga
<<http://www.ruf.rice.edu/~barlow/para.html>>.
- 1995b. "ParaConc: A Concordancer for Parallel Texts." *Computers and Texts* 10. (CTI Textual Studies).
- 1996. "Analysing Parallel Texts with ParaConc." ALLC/ACH '96. University of Bergen, Norway.
- Bonhomme, P. dan L. Romary JUNE 1995. "The Lingua Parallel Concordancing Project: Managing Multilingual Texts for Educational Purpose," dalam *Proceedings of Language Engineering*.
- Brown, P., Della Pietra, S., Della Pietra, V. dan Mercer, R. 1993. "The Mathematics of Machine Translation." *Computational Linguistics* 19(2): 263-311.
- Brown, Ralf D. 1996. "Example-Based Machine Translation in the Pangloss System." *Proceedings of the Sixteenth International Conference on Computational Linguistics* August 5-9, 1996. Copenhagen, Denmark, 169-174. 24 Desember 2000 <<http://www.cs.cmu.edu/~ralf/papers.html>>
- Church, K.W. dan Gale, W. 1991. "Concordances for Parallel Text" dalam *Using Corpora*, Proc. 7th Annual Conference of the UW Centre for the New OED and Text Research. Oxford: OUP, 40-62.
- COLING 94. *Proceedings of the 15th International Conference on Computational Linguistics* Vol. 1 & 2, August 5-9 1994, Kyoto, Japan.
- Danielsson, Pernilla dan Daniel Ridings 1996. "PEDANT: Parallel Texts in Göteborg." Göteborgs Universitet. 24 Desember 2000
<<http://spraakbanken.gu.se/lb/pedant/parabank/node1.html>>
- DBP 1992. *Kamus Inggris Melayu Dewan*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Gale, W. A. dan Kenneth W. Church 1993. "A Program for Aligning Sentences in Bilingual Corpora." *Computational Linguistics* 19(3): 75-102.
- Hank, P. 1979. *Collins English Dictionary*. London & Glasgow: Collins.
- Hartmann, R.R.K. 1994. "The Use of Parallel Text Corpora in the Generation of Translation Equivalents for Bilingual Lexicography." Amsterdam: Euralex 1994 Proceedings. 291-297.
- Johansson, S., J. Ebeling dan K. Hofland 1994. "Coding and Aligning the English-Norwegian Parallel Corpus." 25 Desember 2000
<<http://www.hd.uib.no/enpreports.html>>.
- Karlgren, Hans et al. 1994. "Dilemma – An Instant Lexicographer." *COLING 94* Vol. 1: 82-84.
- Kipling, Rudyard 1987 (1901). *Kim*, World's classics Edition. Edited with an introduction by Alan Sandison. Oxford:OUP.
- 1989. *Kim*. Edisi Terjemahan(Berdasarkan Edisi OUP 1987). Zainal Abidin Bakar. Kuala Lumpur: DBP.
- Klavans, J. dan E. Tzoukermann 1995. "Combining Corpus and Machine-Readable Dictionary Data for Building Bilingual Lexicons." *Machine Translation* 10 (3), 185-218.
- Kumano, Akira dan Hideki Hirakawa 1994. "Building an MT Dictionary from Parallel Texts based on Linguistics and Statistical Information." *COLING 94* Vol. 1: 76-81.
- Marinai, E., Peters, C. dan Picchi, E. 1991. "Bilingual Reference Corpora: A System for Parallel Text Retrieval," dalam *Using Corpora*, Proc. of 7th Annual Conference of the UW Centre for the New OED and Text Research. Oxford: OUP, 63-70.

- McArthur, T. 1992. *The Oxford Companion to the English Language*. London: OUP.
- Resnik, Philip dan I. Dan Melamed 1997. "Semi-Automatic Acquisition of Domain-Specific Translation Lexicons," dalam *Fifth Conference on Applied Natural Language Processing*, Washington, D.C.
- Romary, L., Mehl, N. dan D. Woolls. 1995. "The Lingua Parallel Concordancing Project: Managing Multilingual Texts for Educational Purposes." *Text Technology* 5/3.
- Salkie, R. 1995. "INTERSECT: A Parallel Corpus Project at Brighton University." *Computers & Texts* 9:4-5. (Published by CTI Textual Studies. 25 Desember 2000 <<http://users.ox.ac.uk/~ctitext2/index.html>>)
- Shahrezad Ibrahim 1995. *Kamus Inggeris Melayu Dewan Edisi Ringkas*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Sinclair, J. 1996. "Preliminary Recommendations on Corpus Typology." EAG-TCWG-CTYP/P. Mei 1996. EAGLES. 23 Desember 2000 <<http://www.ilc.pi.cnr.it/EAGLES96/corpustyp/corpustyp.html>>
- Warwick, S. dan G. Russell 1990. "Bilingual concordancing and bilingual lexicography," dalam *Proceedings of the Fourth International EURALEX Conference*, 1-4. Malaga.