

# Yang Selari dan Yang Setanding: Peranan Korpus dalam Penterjemahan\*

*Rusli Abdul Ghani*  
([rusli@dbp.gov.my](mailto:rusli@dbp.gov.my))

*Norhafizah Mohamed Husin*  
([hafizah@dbp.gov.my](mailto:hafizah@dbp.gov.my))

Bahagian Penyelidikan Bahasa  
Dewan Bahasa dan Pustaka  
(<http://www.dbp.gov.my>)

\*Dibentangkan dalam Persidangan Penterjemahan Antarabangsa Ke-8 pada 3–5 September 2001 di Langkawi, Kedah.

## **Abstrak**

Kertas kerja ini menghuraikan peranan korpus selari dan korpus setanding dalam konteks penyediaan prasarana linguistik untuk penterjemahan.

Korpus selari terbina daripada teks-teks dalam suatu bahasa sumber serta terjemahan teks-teks tersebut dalam bahasa(-bahasa) sasaran. Korpus setanding pula merupakan kumpulan teks terpilih dalam bahasa sumber serta teks-teks yang serupa dalam varieti bahasa tertentu (misalnya kumpulan teks bahasa Melayu, bahasa Melayu Brunei dan bahasa Indonesia) atau dalam bahasa(-bahasa) yang lain.

Matlamat utama kajian ini ialah pembinaan sebuah pangkalan data frasa. Pangkalan data ini akan memuatkan frasa lazim dan ungkapan tetap dalam bahasa sumber (bahasa Inggeris dalam kes kajian ini) dengan padanannya dalam bahasa sasaran (bahasa Melayu). Frasa dan ungkapan tetap serta padanan masing-masing diperoleh daripada korpus selari dan juga korpus setanding.

## **1.0 PENDAHULUAN**

*“... in the early 1960s, an apocryphal tale went around about a computer that the CIA had built to translate between English and Russian: to test the machine, the programmers decided to have it translate a phrase into Russian and then translate the result back into English, to see if they'd get the same words they started with. The director of the CIA was invited to do the honors; the programmers all gathered expectantly around the console to watch as the director typed in the test words: "Out of sight, out of mind." The computer silently ground through its calculations. Hours passed. Then, suddenly, magnetic tapes whirred, lights blinked, and a printer clattered out the result: -- Invisible insanity.”*  
(Budiansky 1998)

Dalam ranah pengkomputeran automatik anekdot seperti di atas terlalu lazim sehinggakan sudah *cliché* dan *passé* (untuk variasi cerita di atas lihat sebagai contoh Hutchins 16:1986). Namun, hakikatnya, bukan penterjemah digital sahaja yang menghadapi masalah dengan ungkapan tetap dan idiom. Penterjemah berdarah berdaging pun sukar mencari padanan yang sesuai untuk idiom asing.

Lantaran itu, yang bermasalah bukannya teknologi tetapi aspek-aspek bahasa yang menyangkut makna, “...*understanding, and social and cultural differences of human communication*” (Hutchins dan Somers 9:1992).

Sebahagian masalah penterjemahan dapat ditangani dengan menyediakan senarai istilah dan kamus-kamus dwibahasa. Akan tetapi, bahan rujukan sedemikian sering kali mengabaikan butir-butir linguistik seperti metafora dan idiom yang sering kali melibatkan perbezaan sosiobudaya.

Kajian kami ini merupakan suatu usaha awal untuk mewujudkan pangkalan data yang memuatkan frasa, sama ada idiomatik atau tidak, berdasarkan penggunaan sebenar padanan frasa tersebut dalam teks terjemahan.

## **2.0 KERANGKA KAJIAN**

Dalam kajian ini, kami hanya menumpukan perhatian terhadap aspek data dan pemprosesan data untuk mendapatkan frasa dan padanan frasa. Aspek pangkalan itu sendiri, baik seni binanya mahupun cara pencapaian dan pencarian tidak kami teliti kerana perisian untuk ini sudah banyak di pasaran (misalnya Trados<sup>1</sup> dan Wordfast<sup>2</sup>). Perisian seperti Babylon<sup>3</sup>, melalui Babylon builder, juga boleh digunakan untuk membangunkan pangkalan data frasa.

Satu contoh pangkalan data yang memuatkan idiom ialah Idioms für Windows.<sup>4</sup> Pangkalan data ini memuatkan idiom bahasa Inggeris dengan padanannya dalam bahasa Jerman. Pangkalan data yang memuatkan peribahasa dan ungkapan lain pula dinamakan Sayings für Windows<sup>5</sup>. Gambar rajah yang berikut menunjukkan contoh paparan Idioms für Windows.

---

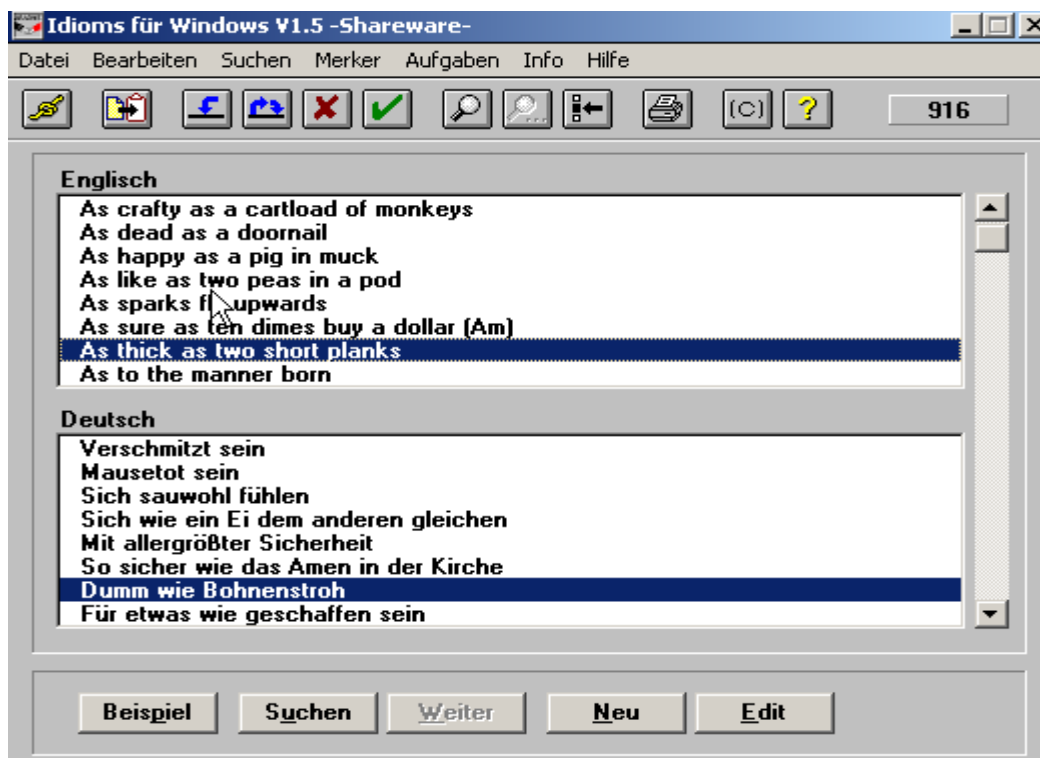
<sup>1</sup> <http://www.trados.com/>

<sup>2</sup> <http://www.champollion.net/>

<sup>3</sup> <http://www.babylon.com/>

<sup>4</sup> <http://www.netcologne.de/~nc-raylero/idi.html>

<sup>5</sup> <http://www.netcologne.de/~nc-raylero/say.html>



## 2.1 Matlamat

Matlamat kajian ini adalah untuk meneliti kemungkinan sebuah pangkalan data frasa dibina berdasarkan hasil penterjemahan sebenar atau penggunaan sebenar sesuatu frasa, klausa atau ayat. Pangkalan data ini kemudiannya boleh digunakan sebagai alat bantu penterjemahan.

## 2.2 Data

Ada dua bentuk korpus yang digunakan dalam kajian ini: korpus selari dan korpus setanding.

Korpus selari terbina daripada teks-teks dalam suatu bahasa sumber serta terjemahan teks-teks tersebut dalam bahasa(-bahasa) sasaran. Korpus setanding pula merupakan kumpulan teks terpilih dalam bahasa sumber serta teks-teks yang serupa (dari segi domain) dalam varieti bahasa tertentu (misalnya kumpulan teks bahasa Melayu, bahasa Melayu Brunei dan bahasa Indonesia) atau dalam bahasa(-bahasa) yang lain (untuk perbincangan yang lebih lengkap tentang korpus selari dan korpus setanding lihat antara lain Laviosa 1997; Laviosa-Braithwaite 1997; Sinclair 1996; Teubert 1996).

Untuk membina korpus selari kami bersikap oportunistik dengan mengambil data teks bahasa Melayu dan teks bahasa Inggeris yang sudah sedia terakam dalam bentuk elektronik. Korpus selari kami terdiri daripada dua teks asal serta teks terjemahannya. Yang pertama ‘*Kim*’ karya Kipling (1987;1902)<sup>6</sup> dan terjemahannya (Kipling 1989) dan yang kedua ialah ‘*The Prince and The Pauper*’<sup>7</sup> karya Mark Twain (1904) dan terjemahannya ‘*Anak Raja dengan Anak Papa*’ (Za’ba 2001).

Untuk korpus setanding kami mengandalkan WWW lalu mengumpulkan teks yang berkaitan dengan teknologi maklumat dan komunikasi (TMK) yang tersiar dalam akhbar-akhbar harian tempatan. Teks Inggeris diperoleh daripada *Computimes* NSTP<sup>8</sup> dan ruangan *In Tech* The Star<sup>9</sup>, manakala teks bahasa Melayu diperoleh daripada ruangan *Megabait* Utusan Malaysia<sup>10</sup> dan ruangan *Komputer* Berita Harian.<sup>11</sup> Semua teks yang berkaitan dengan TMK dikumpulkan untuk membentuk korpus setanding.



<sup>6</sup> Dimuat turun daripada Project Gutenberg <<http://promo.net/pg>>.

<sup>7</sup> Dimuat turun daripada Project Gutenberg <<http://promo.net/pg>>.

<sup>8</sup> <http://www.nstpi.com.my/TECH/Computimes/>

<sup>9</sup> <http://star-techcentral.com/intech/>

<sup>10</sup> <http://www.utusan.com.my/utusan/>

<sup>11</sup> [http://www.nstpi.com.my/Current\\_News/BH/Section/khas\\_html?mysec=Komputer](http://www.nstpi.com.my/Current_News/BH/Section/khas_html?mysec=Komputer)



Dua contoh makalah di atas membentuk teks setanding kerana kedua-duanya dalam domain TMK dan memperkatakan topik yang sama. Kedua-dua makalah ini diteliti untuk mencari frasa yang bersepadanan.

### 2.3 Perisian

Korpus selari diproses dengan perisian ParaConc<sup>12</sup> untuk memaparkan ayat dengan ayat atau perenggan dengan perenggan.<sup>13</sup> Perisian ParaConc versi  $\beta$  ini disebarluaskan secara percuma untuk tujuan pengujian. Bagaimanapun, versi lengkapnya<sup>14</sup> dikenakan bayaran.

Korpus setanding pula dibandingkan secara manual kerana teks ini tidak dapat disejajarkan secara automatik.

<sup>12</sup> ParaConc for Windows Version 0.7 $\beta$  © 1996 Michael Barlow. Rice University. 22 Disember 2000  
<<http://www.ruf.rice.edu/~barlow/Pfw07.html>>.

<sup>13</sup> Dengan andaian pola biasa terjemahan ialah satu ayat diterjemahkan dengan satu ayat (Church dan Gale 1991  
Gale dan Church 1993).

<sup>14</sup> Versi komersial untuk sistem operasi Windows terdapat di <<http://www.athel.com/>>.

### 3.0 DAPATAN

Bagi korpus selari, teks sumber diteliti secara manual untuk mendapatkan frasa, ungkapan tetap atau idiom. Kemudian padanan frasa, ungkapan tetap atau idiom dicari dengan menggunakan ParaConc.

Bagi korpus setanding, setiap makalah sumber dipadankan dengan teks sasaran mengikut keserasian topik. Frasa dan padanannya diperoleh secara manual.

Setiap frasa yang berkemungkinan dijadikan entri pangkalan data frasa akan disemak dengan menggunakan kamus dwibahasa (KIMD 1992) dan kamus ekabahasa (NSOED 1997 dan KD3 1994).

### 3.1 Korpus Selari ‘Kim’

Daripada korpus selari ‘Kim’ kami memperoleh beberapa frasa dan padanannya yang boleh dijadikan entri bagi pangkalan data frasa.

#### A. King-of-the-castle game

Permainan ini ditakrifkan sebagai “a children’s game involving dislodging a player from a mound” (NSOED). Daripada ayat sumber dan ayat sasaran dapat dilihat bahawa padanan yang digunakan untuk permainan ini ialah ‘raja-mahligai.’ (Mungkin padanan alternatif ialah ‘raja bukit’)

As he drummed his heels against Zam-Zammah he turned now and again from his king-of-the-castle game with little Chota Lal and Abdullah the sweetmeat-seller’s son, to make a rude remark to the native policeman on guard over rows of shoes at the Museum door.

Ketika dia menghentak-hentak tumitnya pada Zam-Zammah, dia menoleh berulang kali daripada permainan raja-mahligainya dengan Chota Lal dan Abdullah, anak si penjual halwa, untuk memerli polis peribumi yang sedang mengawal deretan kasut di muka pintu Muzium.

#### B. Those who beg in silence starve in silence

‘Those who beg in silence starve in silence,’ said Kim, quoting a native proverb.

Sesiapa meminta sedekah secara senyap-senyap akan kebulur sepi, kata Kim, menyebut serangkap peribahasa peribumi.

#### C. If the worst came to the worst

Dalam NSOED ada frasa **if it comes to the worst, if the worst comes to the worst** dengan maksud “if things turn out as badly as possible or conceivable”

He called for a new light-ball to his hookah, and considered the case. If the worst came to the worst, and the boy came to harm, the paper would incriminate nobody.

Dia meminta sebuah bola api yang baru untuk ogarnya, sambil menimbang-nimbang keadaan itu. Sekiranya yang buruk berlaku, dan budak itu mendapat bencana, kertas itu tidak memburukkan sesiapa.

#### D. A heart of gold

Dalam KD3 ada frasa *berhati berlian (emas)* dengan maksud “sangat baik hatinya.”

Dalam KIMD **have a ~ of gold**, berhati emas; manakala dalam NSOED ada frasa **heart of gold** dengan maksud “a kind or generous nature.”

'O holy man! A woman with a heart of gold gives us lodging for the night.

Wahai orang alim! Seorang wanita berhati emas menawarkan penginapan kepada kita buat malam ini.

#### E. In the seventh heaven

KIMD: **in seventh ~**, (*colloq*) seronok betul: *when he's in his studio surrounded by his brushes and paints, he's in seventh ~*, dia seronok betul apabila berada di studionya dikelilingi oleh berus dan cat;

NSOED: *in the seventh heaven*, “in a state of extreme delight or exaltation.”

But Kim was in the seventh heaven of joy.

Tetapi Kim berada di syurga ketujuh dengan kegembiraan.

#### F. stolen a march on

NSOED: **steal a march on** “get ahead of by stealth; gain an (unnoticed) advantage over.”

If your Anglican Chaplain thinks you've stolen a march on him -' Bad luck to Bennett!

Jika sahabat Anglo anda berfikir yang anda telah mencuri darinya, Celaka menimpa Bennett!

Mungkin dalam kes ini, pemilihan padanan tidak menampakkan makna idiomatik.

### G. a weight off my mind

NSOED: **a load off a person's mind** “a source of anxiety removed”; *take a load off a person's mind*, bring a person relief from anxiety.

It's a weight off my mind, but - this thing here?' - he waved the note of hand

Lapang pikiran saya, tapi benda ini? dia mengibas-ngibaskan nota jaminan itu.

### H. to the letter

NSOED: **to the letter** “with adherence to every detail; in accordance with a strict literal interpretation (*of*).”

This prophecy was fulfilled to the letter a month later.

Telahan itu telah terbukti dengan tepatnya sebulan kemudian.

### I. to boot

KIMD: **boot**<sub>2</sub> *n*; **to** ~, dan juga, serta: *he was offered an excellent salary and living quarters to ~*, dia ditawarkan gaji yg lumayan dan juga tempat kediaman.

NSOED: *to boot*, **besides, as well, additionally.**

I was very young, and a fool to boot, when I took Mahbub's message to Umballa.

Saya amat muda, dan amat bodoh sekali, ketika saya membawa pesanan Mahbub ke Umballa.

### J. evil eye

KIMD: **evil eye** *n* mata syaitan.

NSOED: **evil eye** “a malicious look; such a look superstitiously believed to do material harm; the ability to cast such looks.”

Who has died in thy house?' asked Kim in the vernacular. 'None. But she may have the Evil Eye - that sorceress,' the Babu replied.

Siapa yang mati di rumah anda? tanya Kim dalam bahasa tempatan. Tak ada siapa. Tapi dia mungkin mempunyai mata hantu nenek kebayan tu, jawab babu itu.



### K. by Jove

KIMD: **Jove** *n*; **by ~**, (*old-fashioned & colloq*), [*no specif translation*]: *by ~, he's clever!*, wah, pandainya dia!; *he's tall, by ~!*, masya-Allah, tingginya dia!; *I'm going to do it, by ~!*, saya akan membuatnya, tengoklah!

NSOED: **by Jove** *colloq*. “*expr. surprise or approval.*”

Well, there I hear what you have done so well, so quickly, upon the instantaneous spur of the moment. I tell our mutual you take the bally bun, by Jove! It was splendid. I come to tell you so.'

Ya, di situlah saya dengar yang anda telah melakukan sesuatu dengan bagus sekali, pada waktu yang serta-merta dan spontan. Saya memberitahu sahabat bersama kita yang anda mengambil kesempatan, amboi! Anda hebat sekali. Saya datang ni nak memberitahu anda.

My native sepoi-guard, they laughed at me! By Jove! I was such a fearful man. Askar sepoi peribumi yang mengawal saya, mentertawakan saya! Wah, saya seorang yang dalam ketakutan.

Padanan yang pelbagai dalam pemakaiannya mengesahkan huraian yang terdapat dalam KIMD, iaitu [*no specif translation*] untuk frasa ini. Padanan bergantung pada konteks ayat. Dengan adanya banyak contoh padanan sedemikian, penterjemah lebih mudah membuat pilihan padanan yang paling sesuai dengan keperluan teks asal.

## 3.2 Korpus Selari ‘The Prince and The Pauper’

### A. by and by

NSOED: **by and by**

- †1 **One by one, successively.** Only in ME.
- †2 **Continuously.** ME–E17.
- 3 **Before long; presently; soon; shortly.** E16.
- †4 **Therefore, as a consequence.** M16–E17.

The youth led the King a crooked course through Southwark, and by-and-by struck into the high road beyond.

Orang muda tadi berjalan berleenggok-leenggok dan berbelit-belit membawa anak raja itu iaitu melalui Southwark. Setelah menempuh beberapa puluh simpang dan lorong lalu ia masuk ke jalan raya di belakang kampung Southwark itu.

By-and-by he saw the officer coming again, in a great hurry; so he plucked his courage together, purposing to meet his trouble as became a man.

Tiada berapa lama dilihatnya pegawai tadi keluar balik dengan tersangat gopoh rupanya. Jadi ia pun memberanikan hati dengan seboleh-boleah, dan tetaplah hendak menerima hukuman yang akan menimpa itu sebagai laki-laki.

By-and-by he found himself at Temple Bar, the farthest from home he had ever travelled in that direction.

Lama-kelamaan dilihatnya dirinya telah sampai ke Lorong Temple Bar, iaitulah penghabisan jauh yang ia telah ada berjalan dari rumahnya arah ke situ, belum pernah lebih dari itu lagi.

Frasa ini juga mempunyai padanan yang pelbagai. Semua bentuk padanan ini perlu disenaraikan dalam pangkalan data frasa untuk tujuan rujukan penterjemah.

### B. in the flesh

KIMD: **in the ~**, orangnya: *I've never met him in the ~*, saya belum pernah berjumpa dgn orangnya;

NSOED: **in the flesh** "in bodily form, in life."

And still his desire to look just once upon a real prince, in the flesh, grew upon him, day by day, and week by week, until at last it absorbed all other desires, and became the one passion of his life.

Dengan hal yang demikian akhir-akhirnya terbitlah satu cita-cita dalam hatinya yang tak dapat sekali-kali dihilangkannya siang dan malam, iaitu ia ingin hendak melihat anak raja yang sebenarnya dengan matanya sendiri.

### **3.3 Korpus Setanding**

Korpus setanding yang kami teliti merupakan korpus yang terdiri daripada teks dalam bidang teknologi maklumat dan komunikasi. Dalam korpus ini tidak banyak ungkapan tetap atau idiom yang digunakan walaupun kami pada mulanya menjangkakan

bahawa akan ada banyak bentuk ungkapan yang digunakan kerana bidang ini cenderung untuk membentuk ungkapan tidak formal. Apa yang kami peroleh adalah sekadar bentuk-bentuk istilah dan padanannya. Sebagai contoh, *worm* – cecacing dalam konteks jenis virus komputer, *compression software* – perisian pemampatan, *e-commerce* – e-dagang.

#### **4.0 KESIMPULAN**

Apa yang diperlihatkan dalam kajian awal ini ialah kemungkinan korpus selari dan korpus setanding dimanfaatkan untuk membina sebuah pangkalan data frasa. Kekangan utama dan nyata tentu sekali saiz korpus yang agak kecil. Namun demikian, beberapa contoh frasa dan ungkapan telah dapat dipetik daripada korpus selari, manakala korpus setanding tidak produktif dari segi frasa idiomatik tetapi kaya dengan istilah.

Andainya korpus selari dan korpus setanding itu cukup besar, maka akan timbul suatu pola yang jelas tentang sesuatu frasa. Frasa dengan padanan yang sama daripada teks yang berlainan (dan penterjemah yang berlainan) memberikan petunjuk bahawa sudah ada konsensus yang meyakinkan bahawa padanan itulah yang paling cocok. Dalam keadaan terdapatnya satu frasa sumber, berbilang padanan, penterjemah boleh membuat pilihan sendiri berdasarkan kesesuaian. Dalam hal ini, pangkalan data frasa sekadar bertindak sebagai pencetus atau penggalak ingatan.

#### **RUJUKAN**

- Barlow, M. 1995a. *A Guide to ParaConc*. Athelstan. Lihat juga
- Barlow, M. 1995b. "ParaConc: A Concordancer for Parallel Texts." *Computers and Texts* 10. (CTI Textual Studies).
- Barlow, M. 1996. "Analysing Parallel Texts with ParaConc." ALLC/ACH '96. University of Bergen, Norway.
- Budiansky, S. 1998. "Lost in Translation." Dalam *The Atlantic Monthly*: Digital Edition, Disember. <<http://www.theatlantic.com>>
- 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.

- Gale, W. A. dan Kenneth W. Church 1993. "A Program for Aligning Sentences in Bilingual Corpora." *Computational Linguistics* 19(3): 75-102.
- Hutchins, W.J. 1986. *Machine Translation: Past, Present, Future*. Chichester: Ellis Horwood Ltd.
- Hutchins, W.J. dan Somers, H.L. 1992. *An Introduction to Machine Translation*. London: Academic Press Ltd.
- Johansson, S., Ebeling, J. dan Hofland, K. 1994. "Coding and Aligning the English-Norwegian Parallel Corpus." 25 Desember 2000
- KD3 1994. *Kamus Dewan* Edisi Ketiga. DBP: Kuala Lumpur.
- KIMD 1992. *Kamus Inggeris Melayu Dewan*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Kipling, R. 1987 (1901). *Kim*, World's classics Edition. Edited with an introduction by Alan Sandison. Oxford:OUP. Etext daripada Project Gutenberg <<http://promo.net/pg>>
- Kipling, R. 1989. *Kim*. Edisi Terjemahan(Berdasarkan Edisi OUP 1987). Zainal Abidin Bakar. Kuala Lumpur: DBP.
- Laviosa, S. 1997. "How Comparable can 'Comparable Corpora' Be?", *Target*, 9 (2), hlm. 289-319.
- Laviosa-Braithwaite, S. 1997. *The English Comparable Corpus (ECC): A resource and a methodology for the empirical study of translation*. Manchester: UMIST.
- NSOED 1997. *New Shorter Oxford English Dictionary*. Versi CD-ROM. Oxford: OUP.
- 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/corpusyp/corpusyp.html>>
- Teubert, W. 1996. "Comparable or parallel corpora?". Dalam Sinclair, J., Payne J. dan Pérez Hernandez, C. (Ed.) *Corpus to corpus: a study of translation equivalence. International Journal of Lexicography*, 9, 3 (special issue). 238-264.
- Twain, M. 1904. *The Prince and the Pauper*. Chatto & Windus edition. Etext daripada Project Gutenberg <<http://promo.net/pg>>
- Za'ba 2001(1958). *Anak Raja dengan Anak Papa*. Edisi terjemahan *The Prince and the Pauper* oleh Mark Twain. Kuala Lumpur: DBP.