

## BAB 3

### DATA KORPUS

#### 3.1 Pengenalan

Dalam bab ini, perkara yang akan dibincangkan ialah sumber bukti leksikografi. Sinclair (1991: 37) telah menggariskan sumber utama bagi bukti leksikografi mengikut urutan kepopularan kepada tiga, iaitu:

- (i) Kamus-kamus lain.

Ini bererti bahawa maklumat atau sumber leksikografi telah sedia tersusun. Oleh itu, penyusun meneliti kamus-kamus yang telah diterbitkan bagi mendapatkan bukti leksikografi. Kamus yang sedia ada ini membantu penyusun untuk menyemak liputan. Ini adalah kerana berdasarkan kamus ini makna dan konteks kata ditunjukkan.

- (ii) Idea pengguna terhadap bahasa yang mereka gunakan.

Melalui kaedah ini, ahli leksikografi boleh menjalankan ujian sesuatu kata terhadap informan-informannya. Tetapi dalam hal-hal tertentu, penyusun kamus sering kali menggunakan intuisi mereka sendiri untuk menghuraikan sesuatu kata.

- (iii) Pemerhatian terhadap bahasa yang digunakan.

Kaedah ini dapat dibahagikan kepada dua, iaitu melalui kutipan daripada teks-teks tertentu yang biasanya ditulis di atas kad-kad bahan dan pemerhatian terhadap baris konkordans yang dikeluarkan daripada data korpus.

Walaupun kajian ini cuba mengetengahkan data korpus berkomputer, namun untuk mengkomputerkan sesuatu aktiviti, sistem kerja secara manual perlu diteliti terlebih dahulu. Setelah itu, barulah maklumat tersebut dipindahkan ke dalam komputer dengan membina perisian yang sesuai. Oleh itu, perbincangan seterusnya dalam

bab ini adalah berkaitan pengumpulan data perbahaman secara manual dan secara komputer. Perkara yang diberi penekanan ialah sumber leksikografi ketiga, iaitu kutipan yang ditulis di atas kad bahan dan baris-baris konkordans yang dijana daripada data korpus.

### **3.2 Pengumpulan Data Secara Manual**

Bahagian ini akan membincangkan pengumpulan data perbahaman dengan lebih mendalam berbanding bahagian lain kerana perbahaman merupakan prasyarat asas dalam penyusunan sesebuah kamus, khususnya kamus ekabahasa. Tumpuan pula akan diberikan terhadap cara-cara data perbahaman dikumpulkan oleh Dewan Bahasa dan Pustaka apabila badan ini merangka untuk menghasilkan kamus ekabahasa. Dalam proses penyusunan kamus, terdapat lima tahap untuk menyusun kamus secara manual, iaitu:

- i. Pengumpulan data perbahaman.
- ii. Pendrafan.

iii. Pengeditan.

iv. Pembacaan Pruf.

v. Percetakan dan Penerbitan.

(Noresah Baharom, 1993: 63)

Bagi **Kamus Dewan** 1970 yang diketuai oleh Teuku Iskandar, terdapat lima orang yang ditugaskan untuk mengumpulkan bahan (Teeuw, A., 1964: 393-394). Peringkat awal pengumpulan bahan, penelitian dan program pembacaan dilakukan terhadap pelbagai sumber bacaan seperti buku, majalah, akhbar, laporan dan kertas kerja. Sumber bacaan seperti ini dipilih kerana daripada sumber inilah definisi sesuatu kata boleh dihasilkan. Tambahan pula, contoh-contoh penggunaan kata boleh didapati daripada bahan-bahan yang benar-benar wujud. Ini dapat mengelakkan penyusun daripada mereka-reka sesuatu definisi atau contoh ayat.

Bahan-bahan bacaan yang dipilih ini mestilah menggambarkan penggunaan bahasa yang terdiri daripada pelbagai laras atau bidang. Ini adalah kerana pemilihan seperti ini boleh

menggambarkan penggunaan bahasa secara umum oleh penutur bahasa tersebut. Hasil daripada bacaan tersebut, bentuk-bentuk kata yang telah dikenal pasti lazimnya akan dicatat di atas kad-kad kecil yang dikenali sebagai kad bahan. Di samping bentuk kata, dituliskan juga contoh penggunaan kata dan juga sumber bahan tersebut. Ini dapat dijelaskan melalui contoh yang berikut.

ISI  
perisian  
perisian berborak

Visual IRC merupakan perisian berborak dalam bentuk sidang video yang tidak memerlukan anda membaca baris demi baris perkataan atau mesej seperti yang terdapat pada perisian berborak konvensional.

KOSMIK/MEI 99/46/ZYA

Daripada contoh kad bahan di atas, di bahagian kiri akan dicatatkan kata dasar sesuatu bentuk kata. Ini diikuti oleh bentuk kata yang akan dijelaskan dalam kad tersebut. Setelah itu, dimasukkan contoh penggunaan kata dengan kata yang hendak ditakrif digariskan. Di bahagian bawah sebelah kanan dimasukkan maklumat sumber bahan dan

tarikh bahan tersebut diterbitkan serta singkatan nama pengumpul bahan tersebut.

Kad-kad bahan yang telah dicatatkan kemudiannya disusun mengikut jenis kata, iaitu bermula dengan kata tunggal, kata terbitan, kata ganda dan kata majmuk dan juga mengikut urutan abjad. Sebagai contoh, bagi kata dasar *buru*, susunannya akan bermula dengan kata *buru*, *berburu*, *berburu-buru*, *memburu*, *terburu-buru*, *buruan* dan *perburuan*. Akhirnya, kata tersebut akan ditakrif mengikut amalan mentakrif sesebuah kamus.

Bagi pengumpulan bahan untuk **Kamus Dewan** 1970, dalam jangka masa satu minggu, hampir 1500 kad bahan dikumpulkan (Teeuw, A., 1964: 394). Oleh itu, dalam masa satu tahun hampir 75,000 kad berjaya dikumpulkan. Walaupun begitu, bagi perkataan yang biasa, banyak contoh kata yang boleh didapati daripada sumber bacaan, tetapi bagi perkataan yang jarang digunakan, hanya sedikit sahaja contoh yang berjaya dikutip dan kadang-kadang hampir tiada contoh. Justeru bahan-bahan

yang dikumpulkan dalam jangka waktu satu tahun masih tidak mencukupi (Teeuw, A., 1964: 394). Oleh itu, dari semasa ke semasa, pengumpulan data perbahanan terus dilakukan.

Walaupun melalui kaedah secara manual ini Dewan Bahasa dan Pustaka telah berjaya menerbitkan kamus yang berwibawa seperti **Kamus Dewan**, namun disedari bahawa kaedah ini banyak kelemahannya. Antaranya:

- i. Memakan masa yang lama untuk menyediakan data yang lengkap.
- ii. Sukar untuk mendapatkan contoh penggunaan yang banyak, sedangkan apabila mentakrif sesuatu kata, penyusun memerlukan contoh yang banyak bagi menghasilkan takrifan yang tepat. Malahan, contoh yang sedikit menyukarkan penyusun untuk menggambarkan penggunaan sesuatu kata dalam pelbagai konteks dan sukar mendapatkan contoh-contoh yang sesuai.

- iii. Sukar mendapatkan maklumat statistik bagi sesuatu kata.
- iv. Mengambil masa yang lama untuk pengeditan dan pembacaan pruf.
- v. Sukar mengesan kewujudan kata atau makna baharu secara cepat untuk dimuatkan di dalam kamus.

(Noresah Baharom, 1993: 63)

Menyedari kelemahan penyusunan kamus secara manual ini, dan pada waktu yang sama terdapat kecanggihan dalam bidang teknologi maklumat, maka penyusunan kamus telah menggunakan komputer sepenuhnya, begitu juga dengan pengumpulan data perbahanan yang kini menggunakan komputer.

### **3.3 Pengumpulan Data Secara Berkomputer**

Peningkatan dan kamajuan dalam bidang teknologi maklumat telah menjadikan komputer lebih mudah untuk dikendalikan. Komputer kini mampu

menyimpan data dalam jumlah yang banyak dan mempunyai rangkaian yang lebih besar. Kecanggihan dalam bidang teknologi maklumat ini telah menyebabkan perkamusuan tidak terkecuali menggunakan komputer dalam segala aktiviti penyusunan kamus. Misalnya, data perbahanan yang berjuta-juta jumlahnya lebih mudah disimpan di dalam komputer berbanding kad bahan. Malahan, kerja-kerja perbahanan tidak lagi bermula dengan pembacaan teks-teks terpilih, tetapi dengan pemerolehan teks untuk dijadikan data korpus.

Istilah korpus digunakan untuk merujuk kepada himpunan teks, sama ada dalam bentuk tulisan atau lisan, yang disimpan dan diproses di dalam komputer untuk tujuan penyelidikan linguistik (Renouf, A., 1988: 1). Sementara itu, menurut Leech dan Flingstone (1992), korpus berkomputer ialah sampel teks yang disimpan, yang boleh dibaca oleh mesin dan dapat dicapai secara elektronik (Mohd Zulkifli Bahari, 1993: 30). Sinclair pula menganggap korpus sebagai dokumentasi bahasa (1991: 3), sementara W. Nelson

Francis (1982) menyifatkan korpus sebagai koleksi bahan teks yang dicirikan pemberatan tertentu, yang mewakili sebilangan besar teks sesuatu bahasa, dialek atau subset sesuatu bahasa yang digunakan untuk analisis linguistik (Ibrahim Ahmad, 1994: 84). Justeru dapat disimpulkan bahawa korpus berkomputer merupakan himpunan teks yang disimpan di dalam komputer dan digunakan untuk tujuan penyelidikan dalam bidang linguistik.

Ditinjau dari segi sejarah, penggunaan komputer dalam bidang linguistik, iaitu penyelidikan terhadap bahasa Inggeris, telah bermula pada tahun 1961. Pada waktu tersebut, W. Nelson Francis dan Henry Kucera telah mula menggunakan komputer untuk menyusun Brown Corpus di Universiti Brown, Amerika Syarikat (Mohd. Zulkifli Bahari, 1993: 30). Projek Korpus Brown inilah merupakan projek korpus berkomputer yang pertama. Sejak itu, data korpus berkomputer telah berkembang pesat di seluruh dunia terutamanya sejak tiga dekad yang lalu. Jadual di bawah ini

menunjukkan sebilangan korpus komputer yang terdapat di dunia.

Jadual : 6

**Projek Korpus Berkomputer di Seluruh Dunia**

Jenis Korpus	Saiz Korpus
Korpus BBC	120 000 000
Korpus Birmingham	20 000 000
Korpus Brown	1 000 000
Korpus Dutch	60 000 000
Korpus French	190 000 000
Korpus Jerman	27 500 000

(Mohd. Zulkifli Bahari, 1993: 30)

Di Birmingham, penghasilan dan kajian korpus berkomputer telah menjadi satu kebiasaan sejak tahun 1970-an lagi. Pada tahun 1980 pula, menerusi Projek COBUILD, Universiti Birmingham dengan Penerbit Collins telah merancang dan memulakan kerjasama pembinaan korpus. Ini disebabkan kedua-dua pihak ini mempunyai minat yang sama untuk membina sesuatu yang baharu berkaitan dengan kajian terhadap bahasa Inggeris

dan merekodkan pemerhatian mereka terhadap perlakuan linguistik bagi memperlihatkan penggunaan semula jadi sesuatu teks (Renouf, A. 1988: 1). Dalam bidang leksikografi, korpus berkomputer merupakan prasyarat dalam penghasilan kamus, khususnya kamus ekabahasa. Dalam bahagian yang berikutnya, perbincangan akan menumpukan perhatian kepada pembinaan data korpus, konkordans dan kolokasi.

### **3.3.1 Pembinaan Data Korpus**

Data korpus berkomputer amat penting dalam bidang leksikografi kerana data korpus seperti ini menjadikan kerja-kerja penyusunan dan pengemaskinian kamus lebih mudah dan cepat. Dalam menghasilkan sesuatu data korpus, reka bentuknya amat bergantung pada tujuan penghasilannya. Jika tujuan utamanya adalah untuk menghasilkan kamus bagi bahasa kontemporari dan untuk pengguna umum bahasa tersebut, maka penghasil data korpus perlulah mendapatkan sumber-sumber yang berkaitan.

Sumber yang dipilih perlulah menggambarkan penggunaan bahasa tersebut dalam pelbagai bidang atau laras dan sumber tersebut dapat mewakili penggunaan bahasa sebenar oleh penutur jatinya. Bagi **COBUILD**, salah satu strategi untuk menghasilkan korpus induk atau korpus utama adalah dengan menjadikan perpustakaan sebagai mikrokosma bagi bahasa tulisan. Strategi ini telah diamalkan oleh penyusun seperti Yang pada tahun 1985 di Shanghai. Selain itu, **COBUILD** juga bergantung pada sumber bibliografi yang telah mempunyai kedudukan yang komprehensif seperti yang dilakukan dalam penghasilan korpus LOB. Dalam memilih sumber ini, **COBUILD** telah memutuskan beberapa perkara, iaitu korpus utama perlulah mengandungi:

- i. Bahasa tulisan dan bahasa lisan.
- ii. Bentuk umum berbanding bentuk-bentuk teknikal.
- iii. Bahasa kontemporari, iaitu bentuk bahasa yang dikumpulkan bermula pada tahun 1960-an hingga kini.
- iv. Teks yang berlaku secara semula jadi dan bukan dalam bentuk drama.

- v. Memasukkan prosa, termasuklah karya-karya fiksyen tetapi tidak memasukkan puisi.
- vi. Bahasa yang digunakan oleh orang dewasa, iaitu 16 tahun ke atas.
- vii. Bahasa Inggeris baku dan tidak memasukkan dialek.
- viii. Kebanyakkan sumber merupakan bahasa Inggeris Britain dan hanya sedikit sahaja bahasa Inggeris Amerika dan bahasa-bahasa Inggeris lain.

Dari segi bentuk pula **COBUILD** telah memutuskan bahawa 75 peratus daripada sumber terdiri daripada bahasa tulisan, manakala selebihnya, iaitu 25 peratus dalam bentuk lisan. Penulis pula terdiri daripada penulis lelaki dan wanita dengan perkadarana 75 peratus lelaki dan 25 peratus wanita (Sinclair, 1988: 2-3). Bahan-bahan tulisan ini tidak dipilih secara rawak, tetapi berdasarkan beberapa kriteria:

- i. Mempertimbangkan buku-buku yang paling banyak dibaca di Britain dan di luar

negeri. Untuk tujuan ini, mereka menyelidiki senarai bacaan sekolah dan senarai buku yang paling laris jualannya dan hasil-hasil tulisan yang telah mantap, sama ada dalam bentuk fiksyen atau tidak.

- ii. Mengenal pasti tajuk atau judul buku yang paling popular dibaca di luar negara. Ini dilakukan dengan memohon bantuan daripada Perpustakaan British Council di seluruh dunia. Selain itu, buku terlaris yang disiarkan dalam akhbar-akhbar utama dan penerbitan berkala turut diteliti.
- iii. Mendapatkan katalog-katalog seperti katalog perumahan.

Setelah mempertimbangkan kriteria-kriteria ini dan meneliti segala maklumat, maka digariskan senarai awal buku, akhbar, jurnal, dan bahan *ephemeral*<sup>1</sup> untuk kegunaan data korpus.

---

<sup>1</sup> Bahan *ephemeral* ialah bahan-bahan yang tidak kekal seperti katalog, teks ucapan, iklan dan brosur (Nor Azizah Abu Bakar, 1999).

Bagi teks lisan pula, strategi pemilihan sumber yang berbeza-beza telah digunakan. Oleh itu, untuk mendapatkan data lisan yang sesuai, sama ada dalam bentuk pita atau transkripsi, pelbagai badan di Britain telah dihubungi. Misalnya, beberapa buah universiti telah dihubungi bagi mendapatkan teks-teks perbincangan di antara pensyarah dengan pelajar dalam bentuk transkripsi. Selain itu, data juga dikumpulkan daripada individu-individu yang secara suka rela menyumbang data lisan. Temu duga radio di University of Sussex juga diambil sebagai data korpus. Malahan, pihak BBC juga dihubungi untuk mendapatkan data-data dalam bentuk transkripsi yang terdiri daripada laporan, temu duga dan perbincangan melalui siaran radio. Kesemua data yang dikumpulkan ini merupakan bahasa kontemporari yang digunakan oleh penutur dewasa. Melalui cara ini, bahan yang dikumpulkan meliputi bidang yang pelbagai, antaranya seperti bidang kebudayaan, sastera, kesenian dan kewangan. Walau bagaimanapun, bahan tersebut digunakan oleh penutur yang bukan pakar dalam bidang tersebut.

Setelah bahan lisan dan tulisan dikumpulkan, bahan tersebut diubah menjadi bentuk elektronik dengan cara memuatkannya ke dalam komputer. Pada masa ini, terdapat tiga cara untuk mengubah bahan-bahan tersebut ke dalam bentuk elektronik (Sinclair, 1991: 14):

- i. Adaptasi bahan-bahan yang telah sedia ada dalam bentuk elektronik. Melalui kaedah ini, bahan-bahan yang terdapat dalam bentuk disket, yang merupakan storan kecil akan dikonversikan semula ke dalam komputer untuk dijadikan data korpus.
- ii. Penggunaan skan optik, iaitu mesin pembaca optik. Melalui kaedah ini, bahan-bahan bercetak seperti buku, akhbar atau teks-teks bertulis yang lain tidak perlu ditaip semula, tetapi hanya perlu diskan oleh mesin pembaca. Misalnya, **COBUILD** telah menggunakan Kurzweil Data Entry Machine (KDEM). KDEM ini mampu mengenal pasti apa jua bentuk huruf atau tulisan. Konsep mesin ini

sama seperti mesin fotokopi. Buku atau teks akan dihadapkan di atas plat kaca mesin tersebut dan mesin tersebut akan membaca keseluruhan teks berkenaan (skan). Semua huruf akan disalin semula dan dapat dilihat di skrin komputer. Cara ini amat berkesan untuk memasukkan data tulisan kerana secara purata, kaedah ini dua kali lebih pantas daripada konversi melalui paparan kekunci.

- iii. Kaedah melalui paparan kekunci. Kaedah ini memerlukan ketangkasan menaip sesuatu bahan. Biasanya, bahan-bahan tulisan tangan dan transkripsi bahasa lisan akan menggunakan kaedah ini untuk dijadikan data korpus. Walaupun kaedah ini mengambil masa yang agak lama bagi sesuatu bahan untuk disiapkan, namun kesalahan penaipan amat sedikit berbanding penggunaan skan optik.

Melalui sistem korpus yang dibina ini, konkordans akan dijana untuk menganalisis sesuatu kata. Oleh itu, komputer memaparkan maklumat kehadiran dalam bentuk konkordans.

### **3.3.2 Konkordans**

Konkordans ialah baris petikan penggunaan sesuatu kata daripada sekumpulan data korpus, dengan kata yang hendak dikaji diletakkan di tengah-tengah baris (Noresah Baharom, 1994: 9). Menurut Sinclair (1991: 32), konkordans ialah himpunan kehadiran sesuatu bentuk kata dengan setiap bentuk kata mempunyai lingkungan teksnya yang tersendiri. Dengan kata yang lebih mudah, konkordans ialah indeks, iaitu konkordans akan membentuk indeks pengabjadan untuk semua perkataan dalam korpus yang menunjukkan konteks kewujudan sesuatu kata.

Pada masa ini, format KWIC (*Key Word In Context*) digunakan secara menyeluruh dalam

pemprosesan data. Format ini membantu penyelidik meneliti setiap kehadiran dengan lebih mudah dan cepat. Melalui format ini, bentuk kata yang dikaji hadir di tengah-tengah setiap baris konkordans. Bagi membezakan dan menunjukkan kata yang dikaji dengan kata lain di sekitarnya, maka diwujudkan ruang di kedua-dua belah sisi kata tersebut. Format KWIC ini ditunjukkan melalui contoh ayat yang berikut.

JBA memotong bekalan air.

Melalui ayat tersebut, satu bentuk kata dicetak di salah satu bahagian kata yang hendak dikaji, iaitu rentang<sup>2</sup> ± 1. Oleh itu, baris yang ditunjukkan oleh konkordans adalah seperti yang berikut:

bekalan	air	
memotong	bekalan	air
	JBA	memotong
JBA	memotong	bekalan

Berdasarkan contoh di atas, kata yang hendak dikaji diletakkan di tengah-tengah dan mempunyai ruang di kedua-dua sisinya untuk membezakannya

---

<sup>2</sup> Rentang atau *span* ialah ukuran perkataan sesuatu konteks bagi kata kunci atau kata yang hendak dikaji. Contohnya rentang -4,+4 bermaksud 4 kata di sebelah kanan dan 4 kata di sebelah kiri kata kunci atau kata yang hendak dikaji.

dengan perkataan lain. Malahan, penyusun kamus juga sering kali menebalkan cetak huruf bagi kata yang hendak dikaji. Umumnya kepanjangan setiap baris konkordans diputuskan melalui paparannya pada skrin komputer. Daripada contoh di atas juga, jenis huruf, sama ada huruf kecil atau besar tidak diambil kira. Oleh itu, perkataan 'JBA', 'Jba' atau 'jba' akan dikekalkan bentuknya. Justeru dalam pendefinisian kata, perbezaan antara huruf kecil dengan besar tidak dihiraukan dan perkataan tersebut dianggap perkataan yang sama.

Bagi memudahkan penyusun kamus menganalisis konkordans, konkordans boleh disusun mengikut urutan. Urutan yang paling biasa ialah urutan abjad, iaitu kata disusun atau diisih mengikut urutan abjad di sebelah kanan atau kiri kata kuncinya (Renouf, A., 1988:19).

Contoh isih kanan:

182 untuk ukiran tebuk, untuk memotong bahagian-bahagian yang hend 602 dah pandai menjahit, pandai memotong baju untuk abang dan anak2 483 h terbongkok-bongkok hendak memotong batang cengal sebesar dua 562 dan kerana Khalijah telah memotong beberapa ekor ayam. Tok k 495 tengah, bersungguh-sungguh memotong beberapa helai daun cenerai 30 u risau..." tempelak Suisei memotong bicara Fid. Hawa tersenyum 368 hulu Pabitha dengan segera memotong cakap, 'Kita masih sayang 482 ebas, menjak, mencangkul, memotong dahan dan ranting- ranting 112 air, setiap hakim hendaklah memotong dari 1 - 3 mata dari mata 187 au membuat peraturan untuk memotong gajinya untuk tujuan pelabu 569 beri oleh che' gu- nya, dan memotong gambar pengetahuan bagi di- 476 engambil upah menuai padi, memotong getah dan mencucuk atap jik B00088

Contoh isih kiri:

78 kepada pokok rumput akibat dipotong atau diragut menjejaskan p 144 kan menyebabkan markah anda dipotong. Ini mungkin jelas, tetapi 29 ut mengeluh. Cakap bininya dipotong pantas. Cakap Tepawa belum 84 ntonnya rumput Guinea boleh dipotong 4-5 minggu sekali atau 9 k 329 i orang Islam, kemaluannya dipotong ujung kulitnya. "Buka, ri 330 pi jalan. Setiap kenderaan dipotong Hamid. Hati Zahid rusuh. 11 nnya pula pinang yang telah dipotong kecil-kecil serta secebis g 6 itaburkan bawang yang telah dipotong nipis dan akhirnya dipicik K00137 99 tapi kaki kirinya terpaksa dipotong. Saya berasa senang dan leg 39 takdirkan kaki aku terpaksa dipotong? Masih sudikah kau...." J70001 26 ai Icut. Cakapnya terputus dipotong oleh orang tua itu. "Tak so 140 ng. Apabila tandan tersebut dipotong ia akan terus jatuh ke dala B28007 K00057 K00119 K00063 B00037 K00088

Daripada contoh baris-baris konkordans tersebut, penyusun kamus dengan cepat dapat meneliti ko-teks sesuatu kata, sama ada kata di sebelah kiri atau di sebelah kanan kata kunci. Ini dilakukan dengan melihat struktur sintaksis sesuatu kata. Struktur sintaksis sesuatu kata kunci perlu dilihat kerana daripada struktur ini, peranan kata kunci dapat dilihat dengan jelas, sama ada sebagai subjek ayat, predikat, objek,

pelengkap, keterangan atau fungsi-fungsinya yang lain.

Melalui konkordans yang dipaparkan pada skrin komputer, ahli leksikografi juga boleh mendapatkan maklumat tentang sesuatu kutipan. Misalnya konkordans yang dihasilkan oleh **COBUILD**, sumber sesuatu kutipan boleh diketahui melalui kod yang hadir di sebelah kiri setiap baris konkordans. Antara maklumat yang boleh didapati ialah:

i. Sumber sesebuah teks.

Antara kod yang digunakan ialah:-

- G merujuk kepada teks umum
- T merujuk kepada teks teknikal
- W merujuk kepada bahan tulisan
- S merujuk kepada bahan lisan.

ii. Identiti budaya penulis.

Antara kod yang digunakan ialah:

- BR merujuk kepada Britain
- AM merujuk kepada Amerika
- AU merujuk kepada Australia
- AA merujuk kepada Anglo-Amerika
- AG merujuk kepada Anglo-Jerman
- OT merujuk kepada lain-lain negara

iii. Negara penerbit.

Antara kod yang digunakan ialah:

BR merujuk kepada Britain  
AM merujuk kepada Amerika  
AU merujuk kepada Australia

iv. Muka surat.

Kodnya nnn

- (Renouf, A. 1988:7)

Berdasarkan kod ini, buku **The Five Next Time** yang ditulis oleh James Balwin, dikodkan sebagai GW 008 AM BR (027). Tempat atau negeri sesebuah buku diterbitkan penting kerana kod ini dapat memberikan gambaran awal kemungkinan berlakunya campuran dalam proses suntingan bagi buku yang dijual di Britain. Kod seperti ini penting kepada ahli leksikografi kerana daripada kod ini mereka mengetahui ragam bahasa, laras dan kawasan geografi bagi setiap baris konkordans (Krishnamurthy, R. 1987: 64). Bagaimanapun, sistem kod seperti ini sering kali dilakukan pengubahsuaian. Misalnya, pada tahun 1981, halaman rujukan telah digugurkan untuk membolehkan lebih

banyak konteks hadir dalam setiap baris. Pada tahun berikutnya pula, halaman rujukan telah dimasukkan semula tetapi maklumat penulis dan penerbit telah dipindahkan. Walaupun maklumat ini penting kepada penyelidik, namun bagi memaksimumkan jumlah konteks setiap baris, maka maklumat ini telah disimpan di bahagian lain. Bagaimanapun, di DBP kod yang digunakan hanyalah untuk merujuk kepada sumber bahan tersebut sama ada daripada buku, akhbar, majalah atau buku klasik. Kod diletakkan di bahagian kanan pada setiap baris. Angka di sebelah kiri setiap baris konkordans pula merujuk kepada bilangan barisan konkordans yang terhasil daripada sejumlah data korpus yang hendak dikaji.

Konkordans dapat menyediakan maklumat penting tentang makna dan konteks penggunaan kata. Dari pada konkordans yang dibina, ahli linguistik boleh menganalisis sesuatu bentuk kata dan memperoleh pelbagai maklumat, antaranya:

- Maklumat semantik, sintaksis, pragmatik, gaya huraiann makna dan ayat contoh.

- Maklumat statistik, iaitu frekuensi kewujudan sesuatu kata dalam keseluruhan korpus atau sebahagian korpus.
- Penggunaan sesuatu kata dalam konteks yang lebih luas, iaitu dua atau tiga ayat sebelum atau selepas sesuatu ayat yang mengandungi sesuatu kata yang dikaji.

(Noresah Baharom, 1993: 64)

Maklumat penting yang disediakan oleh konkordans ialah maklumat tentang konteks sesuatu kata. Konteks yang disediakan ini dapat memberikan makna dengan jelas bagi kebanyakan kata kunci atau kata yang hendak dikaji. Ini dilakukan dengan melihat kolokasi sesuatu kata melalui struktur sintaksisnya.

### 3.3.3 Kolokasi

Melalui Teori Medan, Trier memperlihatkan hubungan paradigmatis, iaitu hubungan menegak antara bentuk yang menduduki tempat yang sama dalam sesuatu struktur. Tetapi pada tahun 1934, Pozig menyatakan bahawa selain hubungan paradigmatis, hubungan sintagmatik juga amat penting. Beliau memberikan contoh perkataan *gigit* dan *gigi*; *menyalak* dan *anjing*. Sementara itu, Firth pada tahun 1951 menghujahkan bahawa sesuatu perkataan itu boleh diketahui dengan melihat perkataan yang wujud bersama-sama. Contohnya perkataan *ass* (*keldai*) dalam *You silly ass* dan *Don't be such an ass* digunakan dalam kumpulan kata adjektif yang terhad seperti *silly*, *obstinate*, *stupid*, dan *awful*. Menurut Firth, perkataan yang digunakan secara bersama-sama ini dikenali sebagai kolokasi (dalam Plamer, 1992: 82-83).

Svensen, Bo (1993: 98) mentakrifkan kolokasi sebagai sekumpulan perkataan yang

mempunyai makna tertentu apabila perkataan-perkataan tersebut wujud bersama-sama. Sinclair (1991: 170) pula menyatakan bahawa kolokasi ialah "the occurrence of two or more word within a short space of each other in a text". Dengan kata yang lebih mudah, kolokasi ialah "lexical co-occur of word". Moon, R. (1988: 92) menganggap kolokasi sebagai realisasi leksikal bagi konteks situasi. Moon memberikan beberapa contoh kolokasi yang menunjukkan kata yang wujud dalam konteks situasi yang berbeza mempunyai makna yang berbeza. Contohnya *gap* mempunyai empat makna utama apabila berkolokasi dengan konteks situasi yang berbeza.

- i. *Gap* bermaksud *a physical space* apabila berkolokasi dengan *mountain, teeth, in* dan *between*.
- ii. *Gap* bermaksud *the interval of time* apabila berkolokasi dengan *year* dan *of*.
- iii. *Gap* bermaksud *deficiency* apabila berkolokasi dengan *fill* dan *record*.
- iv. *Gap* bermaksud *discrepancy* apabila berkolokasi dengan *close, poor, rich, widen, trade* dan *generation*.

Daripada contoh di atas, jelas kelihatan bahawa perbezaan persekitaran linguistik akan memperlihatkan perbezaan dari segi makna.

Svensen, Bo (1993: 99), membahagikan kolokasi kepada tiga, iaitu:

i. Kolokasi

Apabila sesuatu kata mempunyai kombinasi yang tetap dengan sesuatu perkataan lain, kombinasi tersebut dikatakan sebagai kolokasi. Contohnya, makna *have a feeling* hanya boleh bergabung dengan perkataan seperti *comfort, courage, offence, pride, pity*. Kombinasi ini dikatakan lebih terbatas cakupan atau lingkungannya berbanding kombinasi bebas. Oleh sebab itu, kolokasi perlu dibezakan daripada kombinasi bebas kerana komponen atau perkataan dalam kombinasi bebas biasanya bebas untuk bergabung dengan apa juar kata. Contohnya perkataan *take* yang membawa maksud *bring, carry* dan *move* bebas bergabung dengan apa sahaja objek tepat. Misalnya *take a bag (to*

*the station), take a letter (to the post office), take a child (to the school) dan sebagainya. Tetapi kolokasi mempunyai batasan kata yang tertentu.*

ii. Kolokasi Nahu

Kolokasi nahu melibatkan kombinasi perkataan yang 'dominan' seperti kata nama, kata kerja dan kata adjektif dengan kata sendi. Misalnya *adhere to, amazement at* dan *angry with*.

iii. Kolokasi Semantik dan Leksikal

Menurut Benson, M (1985: 61), kolokasi semantik dan leksikal biasanya mengandungi status yang hampir sama. Set kolokasi sesuatu kata boleh sahaja banyak atau sedikit bagi perkataan yang hampir maksud atau maknanya. Contohnya kata kerja *run* membawa maksud *in-charge, organize* apabila berkolokasi dengan objek seperti *a business, a company, an institution*, juga apabila berkolokasi dengan *a gym, a pawnshop, a pizza parlour*. Walau bagaimanapun, makna yang hampir sama dengan

*run* seperti *manage* atau *direct* hanya boleh berkolokasi dengan *a business*, *a company*, *an institution* tetapi tidak dengan *a gym*, *a pawnshop*, *a pizza parlour*. Ini menunjukkan bahawa set kolokasi bagi perkataan yang hampir sama maksudnya mempunyai cakupan atau persekitaran linguistik yang tidak seratus peratus sama. Terdapat kata yang luas cakupan kolokasinya dan ada kata yang sedikit cakupan kolokasinya.

Maklumat seperti ini amat penting dalam penyusunan sesebuah kamus terutamanya kamus ekabahasa kerana kolokasi dapat menunjukkan perbezaan sesuatu makna. Malahan, kamus seperti **COBUILD** berjaya menunjukkan bentuk tipikal sesuatu kata berdasarkan kolokasinya. Pada tahap-tahap tertentu, kolokasi yang dipaparkan di dalam kamus dapat membimbing pengguna memilih bentuk kata yang sesuai digunakan bagi sesuatu perkataan. Melalui kolokasi juga pengguna dapat melihat dengan jelas bahawa setiap kolokasi mempunyai fungsi yang berbeza-beza.

Kemajuan dalam bidang teknologi maklumat telah dimanfaatkan oleh pelbagai bidang ilmu dan bidang leksikografi tidak terkecuali. Komputer telah digunakan sepenuhnya oleh ahli leksikografi bermula daripada pengumpulan data korpus sehingga terhasilnya sesebuah kamus. Kecanggih pengkomputeran kini telah mengukuhkan lagi kualiti bukti leksikografi. Malahan, kamus lebih cepat dihasilkan dengan berbantukan komputer.