

**WAJARAN PENGGUNAAN DATA KORPUS DALAM PENULISAN ILMIAH:
DIMENSI BAHARU SUKATAN PELAJARAN BAHASA MELAYU
SIJIL TINGGI PELAJARAN MALAYSIA (STPM)**

*(The Use of Data Corpus in Academic Writing: A New Dimension in STPM
Malay Language Syllabus)*

HISHAMUDIN ISAM
Universiti Utara Malaysia
din@uum.edu.my

FAIZAH AHMAD
Universiti Utara Malaysia
a.faizah@uum.edu.my

MASHETOH ABD. MUTALIB
Universiti Utara Malaysia
mashetoh@uum.edu.my

Dihantar pada:
20 September 2014

Diterima pada:
27 Oktober 2014

Koresponden:
din@uum.edu.my

Abstrak: Penggubalan sukatan pelajaran bagi mata pelajaran Bahasa Melayu yang dibuat mulai tahun 2012 oleh Kementerian Pelajaran Malaysia bertujuan menggantikan sukatan pelajaran tersebut yang telah digunakan semenjak peperiksaan Sijil Tinggi Persekolahan Malaysia pada tahun 2000. Sebagai satu strategi serampang dua mata, penggubalan semula sukatan pelajaran ini diharapkan dapat melahirkan murid-murid yang seimbang daripada aspek kemahiran kognitif, manipulatif, dan kemahiran insaniah (*soft skills*). Penelusuran sukatan pelajaran baharu mata pelajaran Bahasa Melayu menemukan pengkaji dengan kepelbagaian kelebihan yang boleh diteroka oleh pelbagai pihak, khususnya yang melibatkan guru dan murid dalam menjayakannya. Justeru, kertas kerja konsep ini bertujuan untuk meneliti kandungan baharu sukatan pelajaran tersebut, dan menelusur skop dan ruang yang boleh dimanipulasi oleh guru-guru Bahasa Melayu Tingkatan Enam, terhadap hasil daripada kemajuan dan kejayaan kajian-kajian dalam bidang bahasa dan linguistik, untuk diselarikan dengan keperluan sukatan pelajaran baharu. Berdasarkan komponen-komponen baharu yang diperkenalkan dalam sukatan pelajaran seperti pengenalan komponen penulisan ilmiah dalam sistem penggal, maka dapat diandaikan hasrat untuk menjadikan kelayakan Tingkatan Enam setanding dengan program prauniversiti yang ditawarkan oleh institusi pendidikan swasta, dan usaha memperbaiki dan menaik taraf sistem peperiksaan STPM, sesuai dengan keperluan orientasi pengajaran dan pembelajaran di peringkat kolej dan universiti, dapat direalisasikan.

Kata kunci: Sukatan pelajaran, Bahasa Melayu, penulisan ilmiah, data korpus, analisis korpus

Abstract: The initiation of syllabus drafting for the Malay Language subject in the year 2012 by the Ministry of Education aims at replacing its previous account which has been used since the Malaysian Higher School Certificate in 2000. The drafting, which also acts as a double-barrelled strategy aspires to produce students with equal knowledge of cognitive, manipulative and soft skills. The search for the new Malay language syllabus has also presented the researchers with various advantages that can be exploited by various parties, particularly the teachers and students. Thus this conceptual paperwork seeks to analyse the reformulated content of the syllabus whilst examines possible scopes and areas to be exercised and manipulated by the Form Six Malay Language teachers with respect to the proliferations and accomplishments of studies in the field of language and linguistics to be associated with the needs of the new syllabus. Based on the new components introduced in the syllabus, such as the introduction of academic writing component on a termly basis, it is assumed that the positive prospect of having Form Six's qualification comparable to pre-university program offered by private institutions, and the elevation of STPM examination status according to the needs of teaching and learning orientations at colleges and universities are possible to be realised.

Keywords: Syllabus, Malay Language, academic writing, corpus data, corpus analysis

PENGENALAN

Bermula pada tahun 2012, Kementerian Pendidikan Malaysia (KPM) telah menstrukturkan semula Tingkatan 6 dan Sijil Tinggi Persekolahan Malaysia (STPM) / Sijil Tinggi Agama Malaysia (STAM) bagi menggalakkan lebih ramai murid mengikuti pilihan kelayakan tersebut. Usaha ini akan menjadikan kelayakan Tingkatan 6 setanding dengan program prauniversiti yang ditawarkan oleh institusi pendidikan swasta. Sekolah yang menawarkan Tingkatan 6 akan diberi kuasa membuat keputusan bagi memperkenalkan pelbagai inisiatif untuk mengupayakan murid menjadi lebih bersedia bagi kemasukan ke universiti (Bahagian Pendidikan Guru, 2014). Satu daripada aktiviti penstrukturkan semula tersebut ialah penggubalan sukatan pelajaran untuk semua mata pelajaran.

Tujuan utama aktiviti penggubalan sukatan pelajaran ini adalah selari dengan matlamat Majlis Peperiksaan Malaysia (MPM) yang sedang berusaha memperbaiki dan menaik taraf sistem peperiksaan STPM, sesuai dengan keperluan orientasi pengajaran dan pembelajaran di peringkat kolej dan universiti (MPM, 2013). Selain itu, keperluan ini juga dirancang untuk memastikan murid-murid yang mengikuti saluran ini turut sama dapat menyumbang ke arah usaha merealisasikan Transformasi Pendidikan Negara, selari dengan Anjakan 1 Pelan Pembangunan Pendidikan Malaysia (PPPM) iaitu inisiatif menyediakan kesamarataan akses kepada pendidikan berkualiti bertaraf antarabangsa (KPM, 2012).

Sehubungan itu, kertas kerja konsep ini bertujuan untuk meneliti kandungan baharu sukatan pelajaran, khusus untuk mata pelajaran Bahasa Melayu STPM. Selain meneliti kandungan baharu, kajian ini juga bertujuan untuk menelusur skop dan ruang yang boleh dimanipulasikan oleh guru-guru Bahasa Melayu Tingkatan 6, terhadap hasil daripada kemajuan dan kejayaan kajian-kajian dalam bidang bahasa dan linguistik, khususnya kajian berteraskan data korpus dalam meningkatkan prestasi murid, selari dengan usaha kementerian meningkatkan imej Tingkatan 6. Sebagai usaha rintis, perbincangan akan dimulakan dengan meneliti sukatan pelajaran baharu bagi mata pelajaran Bahasa Melayu dan perancangan mengendalikan sukatan ini dalam aliran sistem penggal.

SUKATAN PELAJARAN BAHARU BAHASA MELAYU STPM DAN PENGENDALIANNYA DALAM SISTEM PENGGAL

Bagi mata pelajaran Bahasa Melayu STPM, penggubalan semula ini dibuat untuk menggantikan sukatan pelajaran tersebut yang telah digunakan semenjak peperiksaan STPM pada tahun 2000. Sesuai dengan perkembangan semasa, sukatan pelajaran baharu ini memberi penekanan terhadap aspek meningkatkan kemahiran menulis, mencari dan menganalisis maklumat, berfikir secara kritis dan kreatif, serta berkomunikasi secara berkesan dalam kalangan calon. Khusus untuk agenda kemahiran menulis, mencari dan menganalisis maklumat, sukatan baharu ini mewajibkan calon menjalankan kerja kursus. Aspek ini dianggap sangat penting untuk membolehkan calon menerapkan kemahiran insaniah ke dalam diri mereka melalui pengalaman yang ditempuhi semasa menyiapkan kerja kursus tersebut. Hasilnya, calon yang lahir daripada sukatan pelajaran baharu ini diharap dapat berfikiran matang, berpengetahuan luas, dan berupaya untuk menyampaikan idea secara berkesan dengan menggunakan pelbagai bentuk komunikasi.

Berkenaan dengan sistem penggal, pengajian di Tingkatan 6 akan dibahagikan kepada tiga penggal dan calon akan menduduki peperiksaan pada setiap akhir penggal. Pada penggal pertama, murid-murid berpeluang untuk mempelajari topik Sejarah dan perkembangan bahasa Melayu (termasuk di dalamnya subtopik berkaitan asal usul bahasa Melayu, bahasa Melayu kuno, bahasa Melayu klasik, bahasa Melayu moden, dan unsur asing dalam bahasa Melayu, topik fonologi (subtopik fonetik dan fonemik, penggolongan dan penghasilan bunyi bahasa, struktur suku kata bahasa Melayu, sistem ejaan bahasa Melayu, dan aspek tanda baca), topik pembacaan kritis dan analitis, serta penulisan esei tidak berformat.

Dalam penggal kedua pula, murid-murid akan didedahkan dengan topik morfologi (termasuk di dalamnya subtopik bentuk kata, golongan kata, dan proses pembentukan kata), topik sintaksis (jenis dan binaan frasa, jenis dan binaan klausa, pola ayat dasar, susunan konstituen ayat, ragam ayat, jenis ayat, bentuk ayat, dan proses penerbitan ayat), topik analisis kesalahan bahasa, dan topik penulisan esei berformat. Sementara itu, pada

penggal ketiga, murid-murid akan mempelajari topik wacana, retorik, variasi bahasa, laras bahasa, semantik, dan pragmatik (Majlis Peperiksaan Malaysia, 2012).

Selain topik-topik tersebut, pada penggal ketiga juga, murid-murid dikehendaki menulis sebuah penulisan ilmiah, iaitu penulisan berdasarkan isu-isu bahasa Melayu yang boleh dibincangkan berdasarkan topik-topik yang telah dipelajari sejak daripada penggal pertama lagi. Di sinilah menariknya sukatan pelajaran baharu bagi mata pelajaran Bahasa Melayu, kerana murid-murid akan didedahkan dengan pembelajaran ke arah andragogi, iaitu menyediakan penulisan ilmiah, hasil daripada penyelidikan yang dijalankan oleh murid-murid itu sendiri. Keadaan ini sekali gus memberi peluang kepada guru-guru Bahasa Melayu untuk bertindak kreatif mencari skop dan ruang baharu dalam membantu membimbing murid masing-masing semasa menjalankan penyelidikan untuk menghasilkan sesebuah penulisan ilmiah. Keadaan ini juga sebenarnya membolehkan guru-guru ini memanfaatkan peluang yang ada untuk meningkatkan kepakaran dan kemahiran dalam bidang bahasa dan penyelidikan bahasa.

KOMPONEN PENULISAN ILMIAH

Topik penulisan ilmiah yang telah ditetapkan untuk dikendalikan dalam penggal ketiga bertujuan untuk melatih murid-murid menulis esei ilmiah mengikut format yang betul, dan mengembangkan isi secara tersusun, padat, dan meyakinkan supaya terhasil sebuah wacana yang utuh. Topik ini juga diketengahkan untuk membolehkan murid-murid menghasilkan esei yang isinya menepati tajuk dengan menggunakan bahasa yang lancar dan berkesan. Akhirnya, diharapkan murid-murid yang telah berjaya menghasilkan penulisan ilmiah, dapat membentangkan hasil penulisan secara berkumpulan dengan menggunakan intonasi dan bahasa badan yang sesuai (MPM, 2012).

Sebagai strategi serampang dua mata, pengendalian topik ini diharapkan dapat melahirkan murid-murid yang seimbang daripada aspek kemahiran kognitif, manipulatif, dan kemahiran insaniah (*soft skills*). Untuk kemahiran kognitif, murid-murid akan dipimpin untuk memahami dan mengaplikasi pengetahuan dan

sistem Bahasa Melayu, mengaplikasi teori dan kaedah yang diajar dalam bilik darjah dengan tugas yang diberikan, mentafsir, mensintesis, serta merumus data dan maklumat berkaitan dengan isu yang dibentangkan.

Daripada aspek kemahiran manipulatif pula, murid-murid akan digembleng untuk mencari dan mengumpul pelbagai jenis data dan maklumat daripada pelbagai sumber, memperoleh sumber melalui aktiviti memerhati, merekod, menemu bual, dan mengkaji selidik, mempersembahkan data dan maklumat yang diperoleh dalam bentuk grafik, jadual, rajah, carta, dan tulisan, mengaplikasi pembelajaran dengan kaedah, teknik, dan strategi yang ilmiah, seperti dengan memperoleh ilmu melalui sumber kajian, pengumpulan maklumat daripada pelbagai sumber, dan menyampaikan dapatan dan maklumat dalam bentuk komunikasi yang teratur, jelas, dan sah.

Sementara itu, usaha menerapkan topik penulisan ilmiah yang bermatlamatkan penghasilan murid-murid berkemahiran insaniah (*soft skills*), diharap dapat mengupayakan murid-murid untuk berkomunikasi secara lisan dan tulisan, bekerjasama dalam kumpulan, mengorganisasikan tugas, dan bekerja sebagai satu pasukan, mengawal aktiviti kerja kursus yang disediakan secara berkesan, serta mengurus masa mengikut jadual kerja yang ditetapkan. Murid-murid juga akan didedahkan dengan keperluan untuk berusaha menyelesaikan masalah, menghasilkan tugas secara kreatif dan inovatif, mampu menjalankan kajian kuantitatif dan kualitatif secara saintifik, dan berkebolehan untuk menjalankan kajian secara jujur, amanah, dan beretika.

Kesemua kemahiran insaniah yang dipupuk ini akhirnya diharap dapat melahirkan murid-murid yang berupaya meningkatkan keyakinan diri dan bermotivasi. Gabungan ketiga-tiga komponen ini yang disatukan dalam satu topik yang diadun dalam sesuatu mata pelajaran, dan diajarkan dengan kandungan yang utuh, kaedah dan pendekatan yang terbukti berkesan, serta dilengkapi pula bahan pengajaran yang mampan, maka sudah pasti mampu melahirkan murid yang bukan sahaja berilmu bahkan berketerampilan pada tahap yang maksimum (Mahzan, 2012).

ANALISIS KORPUS SEBAGAI SUATU KAJIAN ALTERNATIF

Pada setiap tahun, MPM akan memperuntukkan dua tema sebagai panduan kepada guru dan murid-murid untuk menghasilkan penulisan ilmiah. Pada penggal ketiga tahun 2013 misalnya, tema yang digariskan ialah berkaitan variasi bahasa dan morfologi, manakala pada penggal ketiga tahun 2014, temanya pula ialah fonetik dan fonologi, dan laras bahasa. Berdasarkan kedua-dua tema ini, murid-murid diminta memilih satu daripadanya, dan kajian yang dihasilkan mestilah hasil daripada penyelidikan yang dijalankan dengan kepanjangan antara 2000 hingga 2500 patah perkataan.

Jika dirujuk kepada tema-tema tersebut, didapati bahawa terdapat ruang dan peluang yang boleh digunakan oleh para guru dalam membimbing murid-murid untuk menghasilkan penulisan ilmiah yang menarik. Kelebihan ini boleh dirujuk dengan mengaplikasikan penggunaan data korpus, iaitu data yang menjadi bahan autentik dalam kerja-kerja penyelidikan bahasa dan linguistik dewasa ini.

Data Korpus

Umumnya, data korpus bermaksud himpunan data dari pelbagai genre yang digunakan untuk meneliti pelbagai aspek kebahasaan dalam menghurai fenomena sebenar penggunaan sesuatu bahasa. Konsep ‘himpunan’ membawa maksud himpunan teks digital yang dikumpulkan berdasarkan kriteria tertentu.

Dalam linguistik moden, korpus mempunyai pengertian tambahan, yang diterjemahkan sebagai bahan yang ‘terbacakan dan terolahkan komputer’ (Dewan Bahasa dan Pustaka, 2009). Prinsip ‘terbacakan dan terolahkan komputer’ bermakna setiap data dari kumpulan teks digital tersebut boleh diproses dengan teknik dan kaedah linguistik perkomputeran untuk menampilkan pola dan hubung kait sesuatu lexis dengan suatu lexis yang lain. Prinsip ini mempunyai persamaan dengan pandangan yang diberikan oleh Renouf (1987), yang merujuk korpus sebagai himpunan teks, sama ada dalam bentuk tulisan atau lisan, yang disimpan dan diproses di dalam komputer untuk tujuan penyelidikan linguistik.

Menurut Sinclair (1991:171), data korpus ialah “*a collection of naturally occurring language*

text, chosen to characterize a state or variety of a language”. Hal ini bermaksud, data korpus dianggap sebagai koleksi teks bahasa yang terhasil secara semula jadi (bukan data yang direka-reka untuk tujuan tertentu) yang dipilih untuk menandakan kepelbagaian bahasa. Crystal (1994) mengatakan data korpus ialah koleksi data linguistik, sama ada dalam bentuk teks bertulis atau transkripsi ucapan yang dirakam, yang boleh digunakan sebagai titik permulaan pemerian linguistik atau sebagai satu kaedah untuk mengesahkan hipotesis tentang sesuatu maklumat dalam bahasa. Hal yang hampir serupa turut dinyatakan oleh Kennedy (1998:91) dengan mendefinisikan data korpus sebagai “*a collection of text in an electronic database can be many questions for there are many different kinds of corpora*”.

Sementara itu, Cook (2003:102) mendefinisikan korpus sebagai “*the systematic analysis and description of extensive databanks of language which has actually occurred in use*”, manakala Hasmidar (2006) menjelaskan data korpus sebagai koleksi atau himpunan teks yang dikumpulkan mengikut prinsip yang tertentu untuk tujuan yang tertentu. Korpus ini amat bernilai kerana komponennya membolehkan suatu kenyataan yang menyeluruh dibuat terhadap sesuatu bahasa. Kenyataan pengkategorian juga boleh dilakukan kerana korpus ini disusun dengan teratur mengikut kategori yang tertentu. Terkini, Rusli, Nurhafizah dan Chin Lee Yim (2008) mendefinisikan korpus yang bersinonim dengan korpus komputer sebagai himpunan teks digital yang dikumpulkan berdasarkan kriteria tertentu, dan digunakan untuk meneliti pelbagai aspek kebahasaan.

Berdasarkan pelbagai pandangan yang diberikan, dapat dirumuskan bahawa korpus ialah himpunan data daripada teks lisan atau tulisan (pelbagai genre) yang memperlihatkan ciri-ciri bahasa semua jadi (bahasa yang digunakan secara nyata), tersimpan secara sistematik dalam bentuk pengkomputeran linguistik dan boleh diproses untuk mendapatkan pelbagai jawapan berkenaan maklumat linguistik dan kebahasaan (berkaitan struktur, sistem, tatabahasa) mengenai sesuatu bahasa. Oleh kerana data ini melibatkan data daripada suatu sistem korpus yang berformat, maka hasil analisis dari kajian yang menggunakan

data korpus akan menzahirkan dapatan yang boleh dipertanggungjawabkan.

Kewajaran Pemilihan Data Korpus sebagai Data Kajian

Pemilihan data korpus sebagai data kajian dianggap sangat bertepatan dengan kehendak dan keperluan kajian, termasuk juga kajian penulisan ilmiah oleh murid-murid Tingkatan 6. Terdapat lima faktor penting berkaitan data korpus yang membolehkannya dianggap sebagai data yang berautentik untuk digunakan sebagai data kajian bahasa dan linguistik.

Data yang Sebenar

Data korpus adalah data sebenar dan realistik, tidak dipantau dan menggambarkan konteks sosial yang sebenar. Ia juga boleh dianggap sebagai data sahih dan bukan data yang direka-reka untuk memenuhi keperluan kajian. Menurut Azhar (1993), keadaan ini bermaksud data yang digunakan memaparkan bahasa tabii (*natural language*) yang logik dan tidak ditokok tambah kerana ia mengkaji ayat dalam konteks penggunaan. Oleh itu sesuatu kajian bahasa yang menggunakan data korpus dianggap data yang lebih adil dan tidak bias mengikut acuan objektif yang diharapkan. Hal ini kerana, korpus penting sejajar dengan kepentingan data yang empiris. Ia membolehkan seseorang pengkaji bahasa membuat kenyataan yang objektif atau berdasarkan persepsi kognitif dalaman seseorang individu terhadap bahasa dengan dibuktikan oleh data yang empirikal (Leech, 1997).

Dalam membincangkan hal ini, McEnery dan Wilson (2001) mengatakan, penjelasan berdasarkan persepsi kognitif melalui data korpus dapat dimanfaatkan dalam menyelesaikan permasalahan linguistik seperti semantik, pragmatik dan sintaksis. Oleh itu, murid-murid Tingkatan 6 yang menggunakan data ini berpeluang untuk membandingkan rumus dalam tatabahasa pedagogi (rumus yang dipelajari daripada buku tatabahasa untuk tujuan pengajaran) dengan rumus tatabahasa praktis (rumus daripada bahasa yang digunakan oleh pengguna bahasa sehari-hari). Kepekaan meneliti perbandingan ini akan membolehkan murid-murid dapat memahami kepelbagaian jenis tatabahasa yang ada, dan

menghasilkan penulisan ilmiah yang sangat menarik untuk dibincangkan.

Mudah untuk Dicapai dan Digunakan

Kebanyakan korpus sudah tersedia oleh pihak yang bertanggungjawab terhadap pengembangan bahasa, sama ada badan kerajaan (contohnya Dewan Bahasa dan Pustaka) ataupun pihak universiti (contohnya Universiti Kebangsaan Malaysia, dengan kerjasama Dewan Bahasa dan Pustaka), dan disediakan sama ada pada harga yang rendah atau percuma. Data ini boleh diakses secara atas talian dan mudah dicanai dengan menggunakan pelbagai program yang bersesuaian dengan kehendak kajian. Kemudahan capaian dan penggunaan data korpus banyak dibantu oleh perkembangan teknologi perkomputeran.

Berhubung hal ini, menurut Asmah (2001) data korpus yang diperkenalkan dalam linguistik korpus telah membawa cara baru dalam kajian penelitian data bahasa, dan kaedah ini diperkuatkan lagi dengan perkembangan teknologi perkomputeran. Rundell (1996) mengatakan data korpus telah menjadi sumber kajian bahasa yang utama kerana dua faktor iaitu ia membawa kepada perubahan ideologi, iaitu perubahan besar ke arah kajian bahasa secara empiris dan statistik. Selain itu, data korpus juga telah membawa kepada perubahan dalam teknologi kerana kemampuannya yang besar dalam menyimpan data yang banyak pada kos yang rendah, dan dapat membantu memproses data secara terperinci.

Pewajaran ini sepatutnya menjadi suatu kelebihan kepada guru-guru, kerana murid-murid sekarang secara amnya memang sangat meminati dan mengikuti perkembangan teknologi semasa, khususnya penggunaan komputer. Oleh itu, sekiranya sebelum ini pengajaran dan pembelajaran bahasa hanya melibatkan kaedah hafalan rumus semata-mata, melalui penggunaan data korpus, murid-murid dapat didedahkan dengan pengajaran dan pembelajaran bahasa berasaskan komputer dan aplikasinya. Keadaan ini menjadikan proses penyelidikan yang akan dijalankan lebih menarik dan tidak membosankan.

Percontohan dan Perbilangan

Korpus boleh dijadikan sampel penyelidikan kerana dapat mewakili satu populasi dan sebarang dapatan yang diperoleh daripada korpus boleh

digeneralisasikan pada populasi yang lebih besar. Misalnya, sehingga 25 November 2008 (Dewan Bahasa dan Pustaka, 2009), Pangkalan Data Korpus Dewan Bahasa dan Pustaka telah mengandungi lebih kurang 135 juta perkataan yang dikelompokkan dalam sepuluh subkorpus seperti buku, majalah, akhbar, dan sebagainya. Percontohan dan perbilangan yang pelbagai dari sepuluh subkorpus ini boleh dianggap sebagai mewakili satu populasi yang cukup untuk menghasilkan dapatan kajian yang utuh.

Pangkalan data UKM-DBP pula, walau hanya mengandungi sekitar lima juta perkataan, boleh dianggap mewakili satu bentuk populasi data yang boleh dijadikan sampel kajian, sehingga berjaya menghasilkan sebuah buku berkaitan nahu praktis yang berwibawa kerana menggunakan data yang autentik dan kontemporari sifatnya (Zaharani, 2008). Sekiranya dibimbing dengan berkesan, faktor percontohan dan perbilangan yang besar ini boleh membantu menghasilkan penulisan ilmiah yang dijalankan oleh murid-murid menunjukkan suatu pola dapatan yang menarik untuk dipertimbangkan sebagai suatu rumus bahasa. Hal ini kerana, murid-murid berupaya memperoleh contoh yang banyak yang dapat menggambarkan suatu fenomena bahasa yang jelas dan relevan.

Data yang Diperluas

Banyak korpus yang telah diperluas dengan maklumat linguistik tambahan seperti ulasan bahagian pertuturan, pecahan-pecahan ayat, dan transkripsi prosodi. Hal ini menjadikan data yang telah diberikan maklumat linguistik tambahan lebih mudah dan spesifik. Misalnya, melalui maklumat linguistik tambahan, pengetahuan skema berdasarkan korpus dapat dikembangkan.

Menurut Aston (1997), kajian bahasa yang berasaskan korpus membolehkan pengkaji menyingkap pelbagai aturan tertentu dalam wacana dan ruang lingkup yang boleh dikaji dengan gabungan yang tetap dan separa tetap. Sebagai contoh, selain maklumat pola bahasa berdasarkan baris konkordans, murid-murid juga boleh mendapatkan maklumat pola bahasa berdasarkan gugusan (gugusan rangkaian kata), senarai kata dan *collocates* (kolokasi kata). Kesemua maklumat yang diperoleh berdasarkan kepelbagaian fungsi daripada perisian yang digunakan untuk menganalisis data korpus ini

dapat membantu murid-murid menghasilkan dapatan yang sangat menarik untuk dikupas.

Data sebagai Bukti Kajian Sainifik

Kajian bahasa yang bersifat saintifik seharusnya menggunakan data yang boleh dianggap sebagai bukti terbaik yang dapat membezakan kebenaran daripada pembohongan dan pendayaan. Dalam hal ini, Idris (2008) mengatakan, kemajuan hebat dalam teknologi perkomputeran telah sama-sama membantu menyingkirkan unsur 'prejudisme' dalam kajian bahasa dan linguistik berasaskan himpunan jutaan data (data korpus) yang boleh dijadikan sumber utama sesebuah kajian.

Menurut Milroy dan Milroy (1999), sebagai pengkaji, kita mesti mencuba mendasarkan disiplin ilmu ini pada pengamatan fakta sepada mungkin dan tidak pada set prejudis disebabkan oleh penggunaan data yang direka-reka, data buatan dan data secara introspeksi. Kenyataan ini menjadi dalil kukuh untuk mengatakan bahawa data korpus merupakan bukti sesebuah kajian (khususnya kajian linguistik) boleh dikategorikan sebagai kajian bersifat saintifik. Faktor ini merupakan faktor pengukuh mengapa murid-murid Tingkatan 6 yang diminta untuk menghasilkan penulisan ilmiah, perlu menggunakan data korpus sebagai satu daripada data kajian yang boleh digunakan. Hal ini kerana, seperti mana yang telah diperkatakan sebelum ini, matlamat utama keperluan komponen penulisan ilmiah ialah melahirkan murid-murid yang seimbang daripada aspek kemahiran kognitif, manipulatif, dan kemahiran insaniah (*softskills*).

Sekiranya dibimbing dengan berkesan, ketiga-tiga aspek kemahiran ini mampu diraih berdasarkan penggunaan data korpus sebagai data kajian kerana secara tidak langsung, penggunaan data korpus membolehkan murid-murid memahami, mentafsir, mensintesis, serta merumus data dan mengaplikasikan pengetahuan sistem bahasa Melayu (kemahiran kognitif), mencari dan mengumpul pelbagai jenis data dan maklumat daripada pelbagai sumber (kemahiran manipulatif), serta mengawal aktiviti kerja kursus yang disediakan secara berkesan, serta mengurus masa mengikut jadual kerja yang ditetapkan (kemahiran insaniah).

Kemampuan Data Korpus sebagai Data Autentik dalam Kertas Projek Penulisan Ilmiah

Suatu persoalan yang mungkin bermain di fikiran guru-guru ialah, apakah kemampuan data korpus sebagai data autentik yang boleh dimanfaatkan dalam kertas projek penulisan ilmiah murid-murid Tingkatan 6? Secara umumnya, dalam lapangan dunia bahasa dan linguistik di Malaysia, ramai penyelidik telah menggunakan data korpus dalam kajian kebahasaan untuk meneliti beberapa persoalan yang selama ini belum terungkai. Diperturunkan di sini beberapa contoh kajian empirikal berbantu data korpus oleh penyelidik, untuk dijadikan contoh kajian yang boleh dijalankan.

Contoh 1

Pengkaji : Zaharani Ahmad (2005)
 Bidang tumpuan : Fonologi dan Morfologi
 Tajuk : Isu morfofonologi bahasa Melayu: Menangani kepelbagaian varian dalam bahasa Melayu
 Data/jumlah data : DBP/UKM / 5 000 000
 Persoalan : Isu morfofonologi sering menghadapi masalah apabila berhadapan dengan fenomena kepelbagaian varian dalam kata terbitan bahasa Melayu
 Hasil kajian : Kesemua varian boleh dijana oleh unsur-unsur fonologi yang asas yang mendasari nahu bahasa Melayu, dengan berbantu data korpus sebagai data kajian

Contoh 2

Pengkaji : Imran Ho Abdullah (2005)
 Bidang tumpuan : Sintaksis
 Tajuk : Pendekatan linguistik kognitif dalam pemerian 'di antara' / 'antara'
 Data/jumlah data : DBP/UKM / 1 020 575
 Persoalan : Konflik penggunaan kata sendi nama *di antara* / *antara* dikatakan tidak menurut rumus / peraturan yang ditetapkan oleh buku *Tatabahasa Dewan*
 Hasil kajian : Kajian korpus mendapati bahawa pengguna lebih kerap menggunakan struktur yang

dikatakan 'salah' dari segi rumus / peraturan yang ditetapkan oleh buku *Tatabahasa Dewan*

Contoh 3

Pengkaji : Nor Hashimah Jalaluddin (2005)
 Bidang tumpuan : Semantik dan Pragmatik
 Tajuk : Gandingan kata bilangan dan penggandaan bahasa Melayu: Sama tak serupa
 Data/jumlah data : DBP/UKM / 5 000 000
 Persoalan : Kekaburan makna dalam penggunaan kata bilangan dan penggandaan bahasa Melayu
 Hasil kajian : Bukti korpus dapat menyelesaikan masalah kekaburan makna dengan menunjukkan pola tertentu, untuk mengelakkan sifat bias dalam membentuk peraturan nahu

Contoh 4

Pengkaji : Norsimah Mat Awal (2003)
 Bidang tumpuan : Terjemahan
 Tajuk : Analisis struktur dalaman dan peluasan makna kata adjektif dari perspektif semantik kognitif: Terjemahan karya Shanon Ahmad
 Data/jumlah data : DBP/UKM / 28, 547 perkataan
 Persoalan : Kajian terjemahan masih tertumpu pada analisis hasil, hingga mengabaikan aspek paling penting dalam kajian terjemahan iaitu proses mendapatkan hasil
 Hasil kajian : Panduan pencarian makna (proses) penting untuk menentukan ketepatan hasil terjemahan. Kaedah meneliti struktur dalaman dan peluasan makna sesuatu kata dapat menyelesaikan masalah ini, dengan berbantuan data korpus. Hasil kajian menunjukkan, hampir kesemua kata adjektif dan terjemahannya berkongsi ciri-ciri dalam kategori makna teras.

Keempat-empat kajian kebahasaan yang dipaparkan di atas, telah menggunakan data korpus sebagai data kajian bahasa untuk merungkaikan pelbagai permasalahan yang sering menimbulkan kekeliruan, kekaburan dan salah faham terhadap pengguna bahasa. Dapatan daripada hasil kajian telah membuktikan bahawa data korpus sememangnya berupaya mengemukakan bukti dapatan yang bersifat empirikal dan autentik.

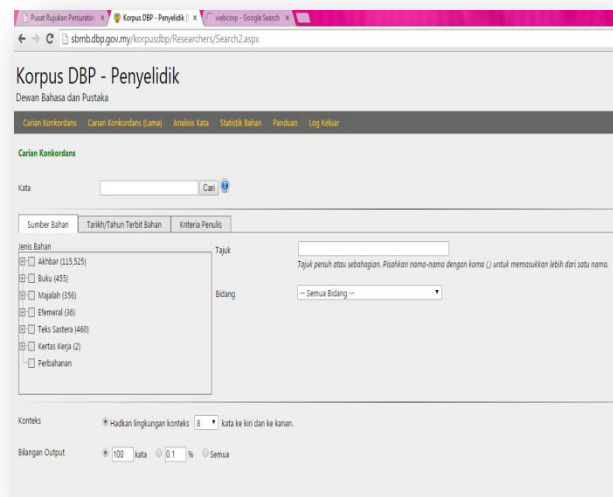
Berdasarkan contoh-contoh kajian yang telah dijalankan tersebut, murid-murid Tingkatan 6 (dengan bimbingan guru), yang menjalankan penyelidikan untuk menghasilkan penulisan ilmiah, boleh turut menggunakan data korpus sebagai data kajian. Walaupun penyelidikan tersebut berskala kecil (hanya sekitar 2000 hingga 2500 patah perkataan sahaja), pelbagai kajian rintis boleh dijalankan, khususnya untuk objektif memahami, mengenal pasti, mengklasifikasikan, menghuraikan, dan menjelaskan sesuatu fenomena kebahasaan. Antaranya ialah seperti:

1. Meneliti rumus tatabahasa (misalnya bidang morfologi dan sintaksis) yang terdapat dalam buku tatabahasa berdasarkan data korpus
2. Meneliti perbandingan rumus tatabahasa daripada buku tatabahasa dengan rumus baharu (kelainan-kelainan berdasarkan contoh) yang dapat dibuktikan berdasarkan data korpus
3. Memeriksa semula rumus tatabahasa yang boleh diaplikasikan oleh pengguna bahasa
4. Membincangkan permasalahan makna (semantik) berdasarkan konteks (pragmatik)
5. Menghurai fenomena kebahasaan daripada perspektif sosial (variasi bahasa) berdasarkan kepelbagaian sumber data subkorpus (majalah, buku cerita, akhbar dan teks klasik)

Guru-guru boleh memilih untuk menggunakan sumber pangkalan data korpus yang bersesuaian dengan jenis penyelidikan yang bakal dijalankan oleh murid-murid. Umumnya terdapat tiga sumber pangkalan data yang sangat mudah dicapai, diaplikasikan dan digunakan sebagai sumber pangkalan data, dan ketiga-tiganya adalah percuma (boleh dilayari secara terus daripada web www.google.com.my).

Satu daripadanya ialah pangkalan data korpus Dewan Bahasa dan Pustaka yang boleh dilayari secara terus untuk mendapatkan data melalui laman web www.dbp.gov.my. Melalui pangkalan

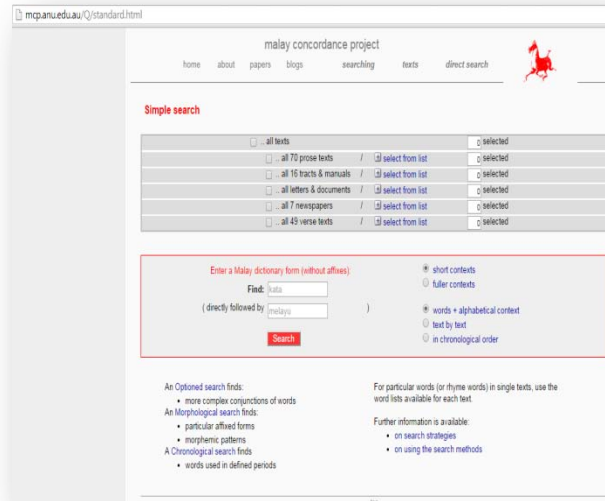
data ini, pelbagai subkorpus data boleh diperoleh seperti data daripada akhbar, buku, majalah, efemeral, teks sastera, kertas kerja dan perbahanan. Kesemua bahan ini pula boleh dipilih berdasarkan tarikh atau tahun diterbitkan dan kriteria penulis (contoh paparan muka hadapan seperti dalam Rajah 1. Oleh sebab data ini adalah data daripada dokumen yang telah diterbitkan, maka data daripada pangkalan ini merupakan data yang telah diedit (oleh editor bertauliah) dan boleh dirujuk sebagai data yang sahih untuk dijadikan bahan kajian. diedit (oleh editor bertauliah) dan boleh dirujuk sebagai data yang sahih untuk dijadikan bahan kajian.



Rajah 1: Muka hadapan pangkalan data korpus Dewan Bahasa dan Pustaka

Selain pangkalan data korpus Dewan Bahasa dan Pustaka, guru-guru juga boleh mencuba untuk membimbing murid-murid yang ingin menjalankan penyelidikan berdasarkan data teks klasik, menggunakan sumber data daripada pangkalan data korpus klasik secara khusus iaitu pangkalan data korpus *Malay Concordance Project* (seperti dalam Rajah 2). Pangkalan yang boleh ditelusuri daripada laman web www.mcp.anu.edu.au ini mengandungi keseluruhan teks klasik yang berupa teks prosa, manual, surat dan dokumen pengurusan, surat khabar dan teks verse. Penggunaan data daripada pangkalan ini membolehkan kajian bersifat diakronik (perbandingan melibatkan tempoh masa) dijalankan. Misalnya kajian yang membandingkan

penggunaan kata penanda wacana klasik seperti *alkisah, hatta, dan arakian, dengan kata penanda wacana moden seperti sehubungan itu, walau bagaimanapun, dan justeru.*

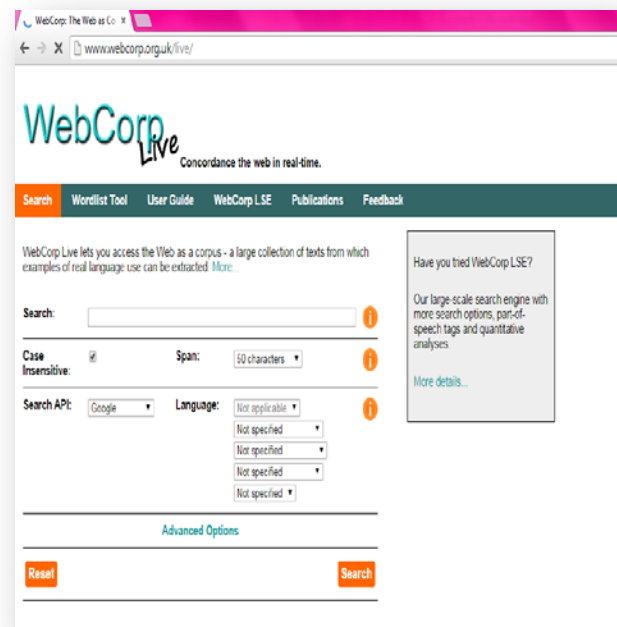


Rajah 2: Muka hadapan pangkalan data *Malay Concordance Project*

Jika pangkalan data Dewan Bahasa dan Pustaka adalah data yang telah diedit, manakala data dari pangkalan *Malay Concordance Project* adalah data teks klasik, pangkalan data *webcorp* (*web corpus*) yang boleh ditelusuri daripada laman web www.webcorp.org.uk pula adalah data yang tidak diedit dan merupakan data moden. Pangkalan data *webcorp* ialah pangkalan data terkini yang menghimpunkan keseluruhan teks yang terdapat dalam laman web. Menariknya data tersebut adalah data segar (asli) hasil garapan penulis seperti tulisan daripada blog persendirian, laman sosial (seperti *facebook dan twiter*) dan laman web rasmi sesuatu persatuan atau pertubuhan. Penggunaan data daripada *webcorp* membolehkan kajian terhadap fenomena penggunaan struktur bahasa Melayu versi moden dijalankan. Muka hadapan pangkalan data ini adalah seperti dalam Rajah 3.

Penggunaan satu daripada pangkalan data ini diyakini membolehkan penulisan ilmiah yang dihasilkan oleh murid-murid Tingkatan 6 mampu memberi impak positif, bukan sahaja kepada perkembangan pendidikan mereka, tetapi dapat

menyumbang kepada perkembangan pengkajian bidang bahasa dan linguistik Melayu di Malaysia.



Rajah 3: Muka hadapan pangkalan data *webcorp*

KESIMPULAN

Hasil penelusuran terhadap sukatan pelajaran baharu mata pelajaran Bahasa Melayu Tingkatan 6 yang telah dijalankan mendapati, terdapat banyak kelebihan yang boleh diteroka oleh pelbagai pihak, khususnya yang melibatkan guru dan murid dalam menjayakannya. Berdasarkan komponen-komponen baharu yang diperkenalkan dalam sukatan pelajaran seperti pengenalan komponen penulisan ilmiah dalam sistem penggal, maka dapat diandaikan hasrat untuk menjadikan kelayakan Tingkatan 6 setanding dengan program prauniversiti yang ditawarkan oleh institusi pendidikan swasta, dan usaha memperbaiki dan menaik taraf sistem peperiksaan STPM, sesuai dengan keperluan orientasi pengajaran dan pembelajaran di peringkat kolej dan universiti, dapat direalisasikan.

Sukatan pelajaran baharu ini juga diharap dapat memainkan peranan sebagai batu loncatan untuk melahirkan murid-murid yang seimbang daripada aspek kemahiran kognitif, manipulatif, dan kemahiran insaniah, seterusnya dapat

menyumbang ke arah usaha merealisasikan Transformasi Pendidikan Negara, selari dengan Anjakan 1 Pelan Pembangunan Pendidikan Malaysia (PPPM), iaitu inisiatif menyediakan kesamarataan akses kepada pendidikan berkualiti bertaraf antarabangsa.

RUJUKAN

- Asmah Omar. (2001). Dasar dan setia bahasa dalam masyarakat pelbagai bahasa. Syarahan umum bahasa Melayu. Anjuran Dewan Bahasa dan Pustaka Wilayah Utara, Perbadanan Perpustakaan Awam Perlis, dan Maktab Perguruan Perlis, 17 Julai 2001, Dewan Besar, Maktab Perguruan Perlis.
- Aston, G. (1997). Enriching the learning environment: Corpora in ELT. Dlm. A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (Eds.), *Teaching and language corpora*, 51-64. London: Longman.
- Azhar M. Simin. (1993). *Sintaksis wacana "Yang" dalam bahasa Melayu*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Bahagian Pendidikan Guru, Bahagian Pengurusan Sekolah Harian. (2014). *Modul Kursus Pemantapan Pedagogi PdP dalam Sistem Pentaksiran STPM: Guru Tingkatan Enam Tahun 2014*. Kuala Lumpur: Kementerian Pelajaran Malaysia.
- Cook, G. (2003). *Applied linguistics: Oxford introductions to language study*. Oxford: Oxford University Press.
- Crystal, D. (1994). *An encyclopedic dictionary of language and languages*. London: Penguin Books.
- Dewan Bahasa dan Pustaka. (2009). *Pangkalan data korpus*. www.dbp.gov.my/laman [16 Mac 2009]
- Hasmidar Hassan. (2006). Kata hubung DAN dan TETAPI: satu analisis perbandingan nahu dan pragmatik. Tesis Doktor Falsafah, Pusat Pengajian Bahasa dan Linguistik, Universiti Kebangsaan Malaysia.
- Idris Aman. (2008). Kata bantu AKAN. Dlm. Zaharani Ahmad (Pnyt.). *Aspek nahu praktis bahasa Melayu*, 195-204. Bangi: Penerbit Universiti Kebangsaan Malaysia.
- Imran Ho Abdullah. (2005). Pendekatan linguistik kognitif dalam pemerian 'di antara' / 'antara'. *Jurnal Dewan*, 5 (3): 1-29.
- Kementerian Pelajaran Malaysia. (2012). *Pelan Pembangunan Pendidikan Malaysia 2013-2025: Laporan awal, ringkasan eksekutif*. Putrajaya: Kementerian Pelajaran Malaysia.
- Kennedy, G. (1998). *An introduction to corpus linguistics*. London: Longman.
- Leech, G. (1997). Introducing corpus annotation. Dlm. R. Garside, G. Leech & T. McEnery, *Corpus annotation: Linguistics information from computer text corpora*, 1-18. London: Longman.
- McEnery, T. & Wilson, A. (2001). *Corpus linguistics: an introduction*. 2nd Edition. Edinburgh: Edinburgh University Press.
- Mahzan Arshad. 2012. Anjakan paradigm dalam memartabatkan kurikulum bahasa Melayu bagi mendepani abad ke-21. Dlm. Mahzan Arshad, IdrisMohd. Radzi&Naffi Mat (Eds), *Memartabatkan Bahasa Melayu: Pengajaran dan Pembelajaran*.Tanjong Malim: Penerbit UPSI.
- Majlis Peperiksaan Malaysia. (2012). *Peperiksaan Sijil Tinggi Persekolahan Malaysia (STPM): Bahasa Melayu, sukatan pelajaran dan kertas soalan contoh*. Kuala Lumpur: Majlis Peperiksaan Malaysia.
- Majlis Peperiksaan Malaysia. (2013). *Sistem pentaksiran baharu STPM*. Selangor: Majlis Peperiksaan Malaysia.
- Milroy, J. & Milroy, L. (1999). *Authority in language: Investigating standard English*. London: Routledge.
- Nor Hashimah Jalaluddin. (2005). Gandingan kata bilangan dan penggandaan bahasa Melayu: Sama tak serupa. *Jurnal Dewan*, 5 (3): 30-52.
- Norsimah Mat Awal. (2003). *Analisis struktur dalaman dan peluasan makna kata adjektif dari perspektif semantik kognitif: Terjemahan karya Shahnnon Ahmad*. Tesis untuk keperluan Ijazah Doktor Falsafah. Pusat Pengajian Bahasa dan Linguistik, Fakulti Sains Sosial dan Kemanusiaan. Universiti Kebangsaan Malaysia, Bangi.
- Renouf, A. (1987). Corpus development. Dlm. J. M. Sinclair (Ed.), *Looking up: an account of the COBUILD project of lexical computing*, hlm.1-40. London: Collins ELT.
- Rundell, M. (1996). Computer corpora and their impact on lexicography and language teaching. Dlm. C.Mullings, M. Deegan, S. Ross, & S. Kensa (Eds.), *New technologies*

- for the humanities*, 198-216. London: Bowker-Sauer.
- Rusli Abdul Ghani, Norhafizah Mohamed Husin & Chin Lee Yim. (2008). Pangkalan data korpus DBP: Perancangan, pembinaan dan pemanfaatan. Dlm. Zaharani Ahmad (Pnyt.), *Aspek nahu praktis bahasa Melayu*, 21-36. Bangi: Penerbit Universiti Kebangsaan Malaysia.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Zaharani Ahmad. (2005). Isu morfofonologi bahasa Melayu: Menangani kepelbagaian varian dalam korpus. *Jurnal Dewan*, 5 (3): 53-77.
- Zaharani Ahmad (Pnyt.). (2008). *Aspek nahu praktif bahasa Melayu*. Bangi: Penerbit Universiti Kebangsaan Malaysia.