

Prospek Penggunaan Korpus untuk Studi Kebahasaan dan Proses Pembelajaran Bahasa Kedua

Priyono

Abstract: This article will argue that intuition-based description by native speakers to some extent may not be accurate in depicting what people really say and think about the language. Since the problem of learning a second language is essentially lexical, the introduction of machine-readable text corpora has been seen as an alternative for describing lexical behaviour of a natural language and media which second language teachers and learners can benefit from (see Fox, 1988, Willis, 1990, Francis, 1993). The corpus programs such as Concordancing, Sentence, and Collocation are capable of providing data of lexical properties, and thus, overcoming the failings of intuition. In addition, it also helps learners and teachers understand the characteristics of lexicon, various meanings of the word and acquire its syntactic properties.

Kata-kata kunci: korpus, studi kebahasaan, pembelajaran bahasa, bahasa kedua.

Kebanyakan guru bahasa Inggris di sekolah lanjutan (SL) dan sekolah menengah (SM) beranggapan bahwa problem utama dalam menguasai bahasa asing adalah penguasaan gramatika. Hal ini dapat disimpulkan dari hasil survai bahasa Inggris di SL dan SM seluruh Indonesia yang dilaksanakan oleh Depdikbud pada 1990. Dilaporkan bahwa 54,1% guru bahasa

Priyono adalah dosen Pendidikan Bahasa Inggris FKIP Universitas Mataram. Artikel ini telah ditelaah oleh Penyunting Ahli Tamu, Nuril Huda.

Inggris di Indonesia lebih menekankan pengajaran tata bahasa, hanya 7,3% menekankan kosa kata (*vocabulary*), dan sisanya menekankan aspek bahasa lainnya. Kecenderungan ini bertentangan dengan berbagai hasil studi tentang pemerolehan bahasa kedua/asing yang memberi indikasi bahwa kesulitan belajar bahasa yang sering dikaitkan dengan penguasaan struktur sebenarnya adalah masalah leksikal. Dengan kata lain, kendala yang dihadapi dalam mempelajari struktur disebabkan oleh penguasaan kata yang tidak sempurna. Penguasaan kata yang sempurna meliputi aspek fonologis, morfologis, semantis, sintaktis, dan pragmatis.

Dalam kajian ini yang dibicarakan adalah aspek semantis dan sintaktis. Permasalahannya adalah: jika seorang siswa telah "mengetahui" arti suatu kata, sejauh manakah pengetahuannya telah mencakup kekayaan sintaktis dari kata itu? Misalnya, untuk kata *sell* dalam bahasa Inggris, dipastikan hampir semua siswa SL mengetahui artinya dalam bahasa Indonesia, yaitu "menjual". Pengetahuan semantis ini jika diterjemahkan ke dalam realisasi sintaktisnya dapat berwujud dalam kalimat *I sold my car* atau *I will sell my car* dan sejenisnya. Tingkat penguasaan seperti ini ternyata hanya sebagian kecil saja dari kekayaan sintaktis seluruhnya dari kata *sell*. Padahal kata tersebut dapat digunakan dalam berbagai struktur kalimat seperti yang diilustrasikan di bawah ini.

1. *The gallery sells highly quality Australian art and craft.*
2. *The car normally sells for \$36,000 on the road.*
3. *"I have to say now, no album sells itself," the singer said.*
4. *The bookmaker sells me a little piece of pasteboard....*
5. *He sells the cane to the mill...*

Kelima struktur kalimat di atas memiliki struktur kalimat yang berbeda, tetapi merupakan struktur yang tertinggi pemakaiannya (Priyono, 1998). Pada kalimat pertama dan kelima, arti kata *sell* sama dengan yang dicontohkan sebelumnya, *I will sell my car*. Pada kalimat nomor 2, 3, dan 4 kata *sell* tidak lagi sepadan dengan "menjual". Ini baru sebagian persoalan semantis. Dari segi sintaktis, kalimat nomor 1 adalah *simple transitive*, kalimat 2 intransitif, kalimat 3 refleksif, kalimat 4 datif, dan kalimat 5 benefaktif. Di antara kelima struktur kalimat itu, nomor 1 dan 5 mungkin tidak asing atau tidak "aneh" bagi siswa sekolah lanjutan, tetapi kalimat intransitif, refleksif, dan datif seperti pada nomor 2, 3, dan 4 tidak mudah diterima oleh siswa.

Pengenalan kata dengan struktur semantis dan sintaktis yang cukup kompleks melalui pengajaran di kelas tidak mudah dilakukan. Eksposisi terhadap bacaan saja tidak cukup memadai dan menjamin penguasaan kata yang sempurna, karena eksposisi semacam itu tidak terfokus pada kata yang sedang dipelajari. Dalam wacana inilah, korpus dapat dipakai sebagai sumber data khususnya untuk memahami kekayaan leksikal dari kata-kata yang ingin dipelajari, karena guru yang bukan penutur asli memiliki keterbatasan dalam mengidentifikasi struktur sintaktis yang perlu mendapat penekanan.

Artikel ini mencoba membuka perspektif baru penggunaan korpus dalam memahami sifat dan karakteristik kosakata bahasa Inggris yang sekaligus dapat membantu proses belajar dan mengajar. Artikel ini dikembangkan berdasarkan studi tentang properti leksikal kata kerja dalam bahasa Inggris dan implikasinya terhadap proses pembelajaran bahasa Inggris sebagai bahasa kedua/asling dengan menggunakan korpus sebagai sumber data. Kalimat-kalimat yang dicontohkan bukan buatan penulis, tetapi diambil dari korpus yang berasal dari teks bahasa Inggris otentik.

PERSPEKTIF HISTORIS PERKEMBANGAN KORPUS

Sejalan dengan pesatnya perkembangan teknologi komunikasi dan informasi, komputer telah menjadi peralatan yang sangat andal. Dengan perangkat komputer, suatu jaringan informasi internasional dapat diakses dalam waktu relatif singkat, misalnya melalui internet. Adanya *World Wide Web* (WWW) membuat informasi di hampir segala bidang dapat dicari dan ditemukan jauh lebih mudah dan cepat.

Dalam perkembangannya, perangkat komputer telah dimanfaatkan secara luas dalam berbagai sendi kehidupan. Di bidang perdagangan, pendidikan, perbankan, keamanan, perhubungan, pariwisata, dan lain-lain, komputer telah menjadi bagian peralatan yang tak terpisahkan. Rasanya sulit saat ini dibayangkan pekerjaan di salah satu bidang tersebut dilaksanakan tanpa komputer. Keunggulan utama penggunaan komputer terletak pada pengelolaan data, informasi, dan dokumentasi, suatu kelebihan yang juga telah banyak dimanfaatkan untuk keperluan studi bahasa.

Penggunaan komputer dalam bidang linguistik dan linguistik terapan bermula dengan diciptakannya *machine-readable text corpora*, suatu perangkat lunak komputer *database* yang digunakan untuk analisis dan deskripsi bahasa dan leksikografi. Sebelum komputerisasi diperkenalkan dalam

bidang kebahasaan, analisis dan deskripsi bahasa banyak bergantung pada intuisi peneliti yang sering dipengaruhi oleh subjektivitas penafsirannya. Gagasan ini jauh sebelumnya merupakan impian para pakar bahasa. Sinclair (1991), misalnya, menyatakan bahwa sekitar tahun 1960-an dirasakan tidak mungkin mengolah teks dengan ukuran jutaan kata dalam waktu singkat. Sepuluh tahun kemudian mulai terlihat kemungkinannya meskipun secara marginal. Pemikiran tersebut pada waktu itu masih dianggap luar biasa. Baru pada awal 1980 pemikiran itu mulai diterima meskipun belum meluas. Kini penggunaan komputer *database* untuk keperluan pengajaran dan pembelajaran bahasa Inggris sebagai bahasa kedua serta studi kebahasaan telah menjadi umum di negara-negara yang berpenutur bahasa Inggris seperti Inggris, Amerika, dan Australia, tetapi belum banyak disosialisasikan di negara-negara yang menempatkan bahasa Inggris sebagai bahasa kedua atau asing, seperti di Indonesia.

Kemajuan teknologi komputer telah menghasilkan *machine-readable text corpora* atau korpus, yaitu kumpulan teks baik lisan maupun tertulis yang diambil dari berbagai sumber yang otentik dalam skala relatif besar sebagai representasi dari penggunaan bahasa yang nyata (*natural language*). Melalui korpus penggunaan bahasa dapat diamati dengan lebih cermat. Sejak kemunculannya, penggunaan korpus dalam studi kebahasaan telah dirasakan semakin penting dan sangat diperlukan (Levin, 1991; Sinclair, 1993). Beberapa korpus yang telah diciptakan dan dikenal dalam bidang kebahasaan, antara lain adalah *The British Lancaster Oslo/Bergen (LOB)*, *The American Brown Corpus*, *The London-Lund Corpus of Spoken English*, dan yang terbesar yang dikenal selama ini adalah korpus yang dikembangkan oleh proyek kerjasama antara penerbit Collins dan Birmingham University di Inggris yang kini dikenal dengan sebutan *The Bank of English*. Korpus ini telah mencapai ratusan juta kata (Kjellmer, 1991).

Kedudukan korpus dalam studi kebahasaan menjadi begitu kuat sehingga tidak perlu lagi diragukan pentingnya sebagai sumber informasi tentang *grammar* dari suatu bahasa. Pernyataan ini memperoleh banyak dukungan luas dari mereka yang memanfaatkan korpus dalam studi kebahasaan, misalnya dalam upaya menyusun deskripsi suatu bahasa atau pengembangan materi pelajaran bahasa asing atau bahasa kedua (literatur studi kebahasaan berdasarkan korpus ini dapat dilihat pada Willis, 1990; Aijmer dan Altenberg, 1991; Fox, 1991; Levin, 1991; Sinclair, 1991; Francis, 1993).

INFORMASI DALAM KORPUS

Korpus dapat dikembangkan dari berbagai teks sumber seperti cerita pendek, novel, majalah, surat kabar, buku teks, bahkan dari rekaman bahasa lisan. Pengembangan korpus untuk studi kebahasaan dilakukan dengan pertimbangan akan kemampuannya dalam memberikan data tentang penggunaan bahasa yang sebenarnya oleh penutur asli. Data tersebut dapat berupa informasi bahasa, antara lain perilaku kata, kolokasi, kelaziman penggunaan bahasa, frekuensi penggunaan kata, dan variasi struktur kalimat.

Perilaku Kata

Yang dimaksud dengan perilaku kata dalam hal ini adalah struktur sintaktis yang membuat sebuah kata dapat berfungsi dalam membentuk kalimat secara gramatikal. Sebagai contoh, dilihat dari struktur sintaktisnya, kata kerja dalam bahasa Inggris dapat digolongkan sebagai transitif, intransitif, atau kedua-duanya seperti kata *sell* pada kalimat 1—5. Kata kerja transitif memiliki struktur sintaktis yang mengharuskan adanya objek dalam kalimat. Apabila objek itu dihilangkan, kalimat bentukannya menjadi tidak gramatikal. Misalnya, kata *hit* tidak pernah muncul dalam kalimat tanpa objek, seperti yang dicontohkan berikut ini.

6a. *He put the ball on the sofa.*

6b. *He put on the sofa.*

Kalimat 6a dan 6b menunjukkan bahwa kata *put* selalu memerlukan objek dalam membentuk kalimat yang gramatikal. Kata itu adalah contoh transitif murni (*strictly transitive*). Sebaliknya, kata *happen* tidak pernah memerlukan objek. Kehadiran objek justru membuat struktur kalimatnya secara gramatikal tidak berterima.

7a. *The explosion happened in the morning.*

7b. *The explosion happened something...*

Kalimat 7a dan 7b membuktikan bahwa kata *happen* tidak dapat diikuti oleh objek. Kata yang berperilaku seperti itu adalah intransitif murni (*strictly intransitive*). Berbeda dengan kedua kata tersebut di atas, kata kerja lainnya seperti *eat*, *read*, *write* tidak mengikuti aturan yang ketat. Kata kerja pada kategori itu dapat muncul dalam kalimat dengan atau tanpa objek.

8a. *I will eat dinner in a restaurant.*

8b. *I will eat in a restaurant.*

Dalam contoh 8a, kata *eat* diikuti dengan objek *dinner*, sedang dalam kalimat kedua (8b) tanpa objek. Sekalipun dalam makna kontekstual me-

ngandung sesuatu objek (sesuatu yang dimakan), kata tersebut menampakkan perilaku yang berbeda dengan kedua kata pada kalimat 6 dan 7.

Di samping itu, ada kata kerja yang dalam alternatifnya mengalami perubahan posisi dan fungsi subjek dan objeknya. Misalnya, objek dalam kalimat aktifnya berubah posisi dan fungsi sebagai objek dalam kalimat pasif. Kata kerja ini dalam bahasa Inggris disebut kata kerja ergatif. Contoh perilaku kata kerja ini dapat dilihat pada kalimat berikut.

9a. *Bill broke the mirror.*

9b. *The mirror broke.*

Yang unik dari contoh kalimat 9a dan 9b di atas adalah bahwa objek kalimat 9a, *the mirror*, yang merupakan benda mati, dapat muncul sebagai subjek kalimat seolah-olah dia pelaku dari predikat *broke*. Dalam alternatif ini bentuk kata kerja dari kalimat aktif ke kalimat pasif tidak mengalami perubahan dan tidak menampakkan tanda-tanda morfologis seperti halnya dalam bahasa Indonesia. Contoh kalimat-kalimat di atas (6—9) memperlihatkan bahwa kata kerja memiliki perilaku sintaktis yang khas dan membedakan satu sama lain. Perilaku sintaktis yang telah diilustrasikan tersebut akan dengan mudah diamati melalui korpus.

Kolokasi

Informasi bahasa lain yang dapat diberikan oleh korpus adalah kolokasi kata. Ada kecenderungan bahwa suatu kata "memilih" kata-kata tertentu lainnya untuk berdampingan dalam kalimat atau wacana yang lebih luas. Kemunculan kata tertentu untuk berkolokasi dengan suatu kata sering tidak dapat diprediksi. Ini disebabkan karena pemilihan kata (*lexical selection*) sangat dipengaruhi oleh sifat idiosinkratis dari kata yang bersangkutan. Ada dua macam kolokasi dalam kalimat, yaitu kolokasi gramatikal dan kolokasi leksikal. Yang pertama terjadi apabila suatu kata bermakna (*content word*) berkolokasi dengan kata fungsi (*function word*). Contoh dalam bahasa Inggris adalah munculnya *phrasal verb* yang terdiri atas kata kerja dan preposisi, misalnya *wait for*, *look over* dan *listen to*. Yang kedua terjadi apabila kolokasi tersebut terdiri atas dua atau lebih kata bermakna, seperti *student diary*, *press conference*, dan *price tag*. Informasi tentang kolokasi kata ini dapat diamati melalui korpus dengan ukuran besar (misalnya 10 sampai 20 juta kata).

Kelaziman Penggunaan Bahasa

Dengan ukuran besar, korpus dapat memberikan contoh nyata cara penutur menggunakan bahasa secara wajar. Kinerja korpus ini dimungkinkan karena bahasa yang dipakai tidak artifisial, melainkan terjadi dalam wacana yang alami. Contoh-contoh yang diambil dari korpus adalah yang lazim dipakai masyarakat penutur bahasa. Dengan demikian otentisitasnya terjamin. Misalnya, kata *cook*, dilihat dari kekayaan sintaktisnya, dapat dipakai dalam struktur sintaktis *simple transitive*, *benefactive*, dan *dative* seperti berikut.

10. *I have to cook Chinese food.*

11. *She went home to cook a meal for him.*

12. *We will cook you some Italian food.*

Akan tetapi, selain ketiga struktur di atas, kata *cook* juga lazim dipakai dalam kalimat intransitif (lihat Priyono, 1998:124-25).

13. *...the fire wouldn't go and the breakfast wouldn't cook...*

Kalimat 13 ini, meskipun sering dipakai oleh penutur asli, memiliki struktur yang tampak tidak berterima secara gramatika (cf. Ard dan Gass, 1987; Yip, 1994). Dalam perspektif orang Indonesia yang sedang belajar bahasa Inggris, kalimat tersebut sulit diterima sebagai ungkapan yang benar. Permasalahannya, yang menjadi subjek kalimat adalah *nonhuman*, sedangkan kata kerja *cook* dianggap sebagai kata kerja agentif yang subjeknya berfungsi sebagai pelaku dalam peristiwa yang diisyaratkan oleh predikat atau kata kerjanya. Kata tersebut menampakkan seolah-olah subjek (*breakfast*) menjadi pelaku dari peristiwa yang dinyatakan oleh kata *cook*. Faktanya tidaklah demikian. Kata *breakfast* dalam kalimat tersebut hanya berfungsi sebagai subjek, tetapi tidak berfungsi sebagai pelaku (agen).

Frekuensi Penggunaan Kata dan Variasi Kalimat

Korpus juga menampilkan data frekuensi penggunaan kata dalam teks. Daftar kata berdasarkan frekuensi dapat diperoleh seperti yang telah dikembangkan oleh *Cobuild Project (CCEED)* yang memuat 2.000 kata yang paling sering digunakan. Daftar kata semacam ini dapat dimanfaatkan untuk kepentingan pengembangan materi belajar kosakata. Di samping itu variasi kalimat atau struktur sintaktis yang dapat dibentuk oleh kata dapat pula diamati frekuensi kejadiannya. Bagi pengajaran dan proses belajar, informasi mengenai frekuensi variasi kalimat dapat dipakai untuk menentukan

skala prioritas materi kebahasaan yang perlu dikuasai oleh pelajar bahasa. Misalnya, kata kerja *pour* dapat dipakai dalam berbagai struktur sintaktis seperti *simple transitive, benefactive, dative, intransitive*. Akan tetapi struktur sintaktis ini tidak seluruhnya lazim dipakai. Dalam studi berdasarkan korpus (Priyono, 1998), ditemukan bahwa struktur sintaktis yang paling sering dipakai oleh penutur asli bahasa Inggris adalah (*pour + NP + Prep + Location*) seperti pada kalimat *I could pour the diesel into the filters*. Informasi semacam ini dengan mudah dapat diperoleh melalui sitasi korpus, dan penemuan ini sangat bermanfaat untuk memberikan aksentuasi pengajaran bahasa Inggris pada properti sintaktis yang perlu didahulukan.

Singkatnya, penggunaan korpus membantu peneliti untuk membuat deskripsi bahasa dengan lebih cermat. Kehadiran korpus memberikan peluang lebih besar bagi pengamatan bahasa terutama yang menyangkut perilaku kata dalam kalimat.

RASIONAL PENGGUNAAN KORPUS DALAM PENELITIAN BAHASA

Beberapa studi dan penerbitan yang berkaitan dengan perilaku kata (Selkirk, 1982; Ard dan Gass, 1987; Atkins, 1988; Fisher, 1994; Little, 1994; Pinker, 1996) telah memberikan indikasi bahwa kesulitan bahasa yang sering dikaitkan dengan penguasaan struktur (*grammatical rules*) sebenarnya adalah masalah leksikal. Dengan kata lain, kendala yang dihadapi dalam mempelajari struktur disebabkan oleh penguasaan kata (*acquisition of lexicon*) yang tidak komprehensif.

Penguasaan kata yang komprehensif meliputi kekayaan fonologis, morfologis, semantis, dan sintaktis. Secara rinci kekayaan leksikal ini dapat dijabarkan sebagai berikut: lafal (bagaimana kata harus diucapkan dan ditulis); struktur kata (kata dasar dan afiksasi); makna (leksikal dan pragmatis); pola kalimat yang dapat dibentuk; kolokasi; sinonimi, antonimi, dan hiponimi.

Penguasaan kata yang komprehensif akan mencakup aspek gramatika. Di sinilah letak titik singgung yang menunjukkan bahwa ranah leksikal tampak sebagai persoalan gramatika. Contoh kasus ini dapat diambil dari hasil penelitian yang dilakukan oleh Ard dan Gass (1987). Dalam penelitian itu responden yang terdiri atas mahasiswa asing di *University of Michigan* yang penguasaan bahasa Inggrisnya masuk kategori *advanced* dihadapkan pada, antara lain, bentuk kalimat *inchoative* seperti pada kalimat 4b. Dalam

penilaian responden, ternyata bentuk kalimat tersebut dianggap *ungrammatical*. Menurut mereka, bentuk yang benar adalah *The mirror was broken*. Di sini nampak persoalan yang sebenarnya adalah bahwa struktur sintaktis (*inchoative* atau *causative*) merupakan konsekuensi dari *lexical property* dari kata *break*. Oleh karena itu permasalahannya adalah leksikal, bukan masalah gramatika. Kesulitan yang dihadapi responden adalah penguasaan kata (*break*) yang tidak mencakup semua *syntactic property*. Penggunaan kata *break* dalam kalimat *The mirror was broken* dan *inchoative* (kalimat 4b) adalah lazim dan benar menurut tata bahasa Inggris. Keduanya merupakan *syntactic variant* dari kata *break*. Oleh karena permasalahan belajar bahasa kedua secara substansial adalah penguasaan kata, maka korpus menjadi media alternatif yang dapat diandalkan karena korpus mampu menghadirkan data otentik penggunaan bahasa.

KEUNGGULAN KORPUS

Publikasi tentang *corpus linguistics* (Aarts, 1991; Halliday, 1991; Kjellmer, 1991; Levin, 1991; Francis, 1993; Willis, 1990) telah membuktikan reliabilitas korpus sebagai sumber informasi tentang perilaku kata dalam kalimat. Keunggulan korpus ini disebabkan karena korpus dapat diciptakan dari contoh penggunaan bahasa yang nyata dan alami serta otentik, dan sumbernya dapat diambil dari berbagai publikasi, bahkan dari rekaman lisan. Di samping itu, korpus juga dapat dikembangkan untuk tujuan khusus, misalnya untuk membuat deskripsi dialek tertentu, seperti *The Corpus of Spoken American English* (CSAE) atau untuk tujuan pengajaran.

Penutur asli suatu bahasa mungkin dapat mengandalkan intuisi mereka dalam memahami dan mendeskripsikan bahasanya. Akan tetapi dalam beberapa hal intuisi sering kurang akurat menggambarkan cara sebenarnya penutur asli menggunakan bahasanya. Hal ini semakin terasa apabila peneliti bahasa menghadapi kata-kata dengan frekuensi tinggi dan struktur semantis yang kompleks. Kata-kata kategori ini memerlukan pengamatan yang lebih cermat melalui data bahasa dalam skala besar. Namun tidak berarti intuisi tidak penting bagi seorang pakar bahasa seperti yang dikemukakan oleh Francis sebagai berikut:

"... intuition maybe useful to linguists in a number of ways, but for the purposes of saying exactly how language is used, it is notoriously unreliable. Even when people are describing their own language prac-

tices, there is often a difference between what they think they say and what they actually do say ..." (1993:139).

Masalahnya, dalam pemakaian bahasa, apa yang terpikir oleh penutur sering tidak sejalan dengan apa yang mereka katakan. Perbedaan ini terkait dengan subjektivitas penutur bahasa. Kesimpulannya, intuisi penutur bahasa tidak sepenuhnya dapat diandalkan untuk mendeskripsikan suatu bahasa. Ilustrasi dari kasus ini dapat diambil dari Willis (1990) yang menemukan bahwa kata *some* dalam tata bahasa Inggris sebagai bahasa asing yang sering dianggap sebagai properti kalimat afirmatif sedang kata *any* cenderung diasosiasikan dengan kalimat negatif dan interogatif, ternyata tidak seluruhnya benar. Melalui *corpus-based analysis*, Willis mendapatkan 23 dari 38 pemakaian kata *any* dalam kalimat afirmatif, 11 dalam kalimat negatif dan hanya 4 dalam kalimat interogatif. Bukti serupa juga dikemukakan oleh Fox (1991) dengan temuan yang tak terduga tentang penggunaan kata *torrent*. Data yang diperoleh dari korpus adalah sebagai berikut.

For the extended sense of the word torrent, intuitive knowledge of the language seems to tell us that a torrent of abuse would be the most typical use; yet a quick glance at the corpus reveals far more lines for a torrent of outrage, a torrent of confession and explanation, and the torrent of investive than there are lines for a torrent of abuse (Fox, 1991:139).

Dari contoh-contoh di atas jelas bahwa penggunaan korpus dapat mengatasi keterbatasan intuisi peneliti bahasa. Studi berdasarkan korpus sebagai sumber data telah mengingatkan kelemahan yang akan dijumpai apabila peneliti terlalu mengandalkan intuisi. Hal ini sekaligus memperkuat keandalan korpus sebagai sumber informasi bahasa. Ketidakandalan intuisi penutur asli yang telah diisyaratkan itu juga memberikan peringatan kepada penutur nonasli (*non native speakers*) yang berupaya untuk menyusun deskripsi bahasa dari *judgment* mereka sendiri (cf. Ard dan Gass, 1987). Dalam kerangka ini, *Text Corpora* menjadi alternatif terbaik sebagai *provider* informasi bahasa.

SUMBER TEKS DAN BESAR KORPUS

Telah disinggung sebelumnya bahwa korpus dapat diciptakan dari berbagai sumber dan secara terus menerus dapat diperluas dan diperbaharui

sesuai dengan tujuan. Sumber teks dapat berasal dari novel, surat kabar, cerita pendek, fiksi/nonfiksi, puisi, majalah, biografi, dan lain-lain. Untuk tujuan khusus korpus dapat dikembangkan dari bahasa lisan atau kumpulan dari dialek tertentu. Dengan demikian korpus dapat diharapkan menjadi representasi dari pemakaian bahasa yang nyata dari suatu komunitas bahasa.

Dalam pengembangannya, ukuran korpus sering menjadi masalah besar. Ukuran pasti (dalam jumlah kata) suatu korpus bahasa sulit ditentukan. Idealnya, semakin besar korpus, semakin representatif fungsinya (cf. Sinclair, 1991). Ukuran korpus sebenarnya tidak terkait langsung dengan tingkat representasi bahasa. Realitanya, standar korpus terus berubah untuk alasan pragmatis. Pada tahun 1960-an satu korpus dengan ukuran satu juta kata dianggap cukup besar dan representatif. Seiring dengan perkembangan teknologi di bidang komputer, korpus dengan ukuran 20 juta kata tidak lagi dianggap terlalu besar. Tingkat representasi juga bergantung pada keberagaman teks sumber. Suatu korpus dengan ukuran puluhan juta kata yang terdiri atas satu macam sumber, misalnya, tidak cukup representatif kecuali jika korpus itu dirancang untuk tujuan khusus. Jadi besar kecilnya korpus tergantung pada pertimbangan pragmatis, misalnya tujuan pembuatan korpus dan kemampuan teknologi penunjangnya.

PROGRAM-PROGRAM KORPUS

Dengan komputer *database*, korpus dapat menawarkan beberapa program, antara lain *concordance*, kalimat, dan kolokasi. Program-program ini dirancang dengan tujuan khusus sesuai dengan macam informasi yang dicari.

Konkordansi

Sinclair (1991:170) mendefinisikan konkordansi (*concordance*) sebagai indeks kata dalam teks yang memberikan akses terhadap contoh pola kalimat tempat kata yang bersangkutan dimunculkan. Sitasi memperlihatkan kata kunci yang berada di lingkungan kata dalam teks. Dalam bahasa korpus, Sinclair menyebutnya sebagai *Key Word in Context* (KWIC). Ketika kata kunci dimunculkan, tampak berbagai macam penggunaan kata tersebut dalam konteks kebahasaan. Misalnya, jika kata itu adalah kata kerja, sitasi akan memperlihatkan tipe objek atau subjek, struktur kalimat pada umumnya, serta kolokasi. Hanya saja, program *concordance* ini mem-

berikan konteks penggunaan kata secara "terbatas". Biasanya produk *concordance* menempatkan kata kunci di tengah satu baris teks. Di satu segi, sitasi seperti ini sangat berguna untuk pengamatan sepintas, di sisi lain tidak memberikan akses terhadap adanya informasi kontekstual yang menarik serta bermanfaat bagi pemahaman yang lebih akurat dan lengkap. Bagian penting tersebut kadang-kadang dapat berada sekitar 80 karakter jauhnya dari kata kunci sehingga hilang dari monitor.

Kalimat

Program lain adalah tayangan kalimat (*sentence*) secara utuh, yang dimulai dari kata awal dengan huruf pertama kapital dan diakhiri dengan tanda titik. Inilah yang disebut dengan program *sentence*. Program ini menyajikan lebih banyak informasi bahasa daripada yang diberikan oleh program *concordance*. Kata kunci pada tayangan kalimat berada dalam konteks kebahasaan yang lebih luas, sehingga makna yang terkandung dapat dipahami dengan lebih mudah dan akurat. Namun program *sentence* lebih banyak membutuhkan waktu dalam pengamatan daripada *concordance*.

Kolokasi

Program kolokasi khusus memberikan informasi tentang kata-kata yang berkolokasi dengan kata kunci serta frekuensi kejadiannya. Kolokasi terjadi setelah atau sebelum kata kunci. Kegunaan program ini adalah memberikan tipikalitas kata-kata yang lazim dipakai berdampingan secara semantis dengan kata kunci. Hanya saja program kolokasi tidak membedakan kolokasi gramatikal dan kolokasi leksikal. Hasil yang diperoleh mungkin dalam bentuk daftar kata dari frekuensi tertinggi sampai terendah. Kelemahan ini semakin terasa karena makna kolokasi berkaitan dengan hubungan semantis dan sintaktis antara kata kunci dan mitra katanya (*the node and the collocates*).

Masalah lain program kolokasi adalah jarak antara *node* dan *collocate*. Hal ini disebabkan karena *node* dan *collocate* sering terpisah oleh beberapa kata yang tidak memiliki kaitan semantis secara langsung. Padahal kolokasi sering diasosiasikan dengan kedekatan pada struktur permukaan dan paling mudah ditandai. Ilustrasi di bawah ini mengisyaratkan bahwa menandai kolokasi semata-mata dari kedekatan saja dapat memberikan gambaran yang salah.

The distinctions we draw between better and...

I found that I had moved, without realizing it, through the gateway....

Kedua kalimat di atas memberikan contoh kolokasi yang tidak bersambung (*discontinuous collocations*). Kata *distinctions* berkolokasi dengan *between*, dan *moved* dengan *through*. Keduanya terpisah oleh dua/tiga kata yang tidak berkaitan langsung baik secara semantis maupun sintaktis. Kesulitan yang akan dialami oleh alat komputasi adalah untuk menghasilkan data kolokasi. Program komputer dalam hal ini terlalu mekanis untuk menemukan nuansa perbedaan yang hanya dapat dirasakan oleh manusia.

PENUTUP

Penggunaan korpus untuk studi bahasa dan pembelajaran bahasa memiliki prospek yang menjanjikan. Dengan sajian bahasa yang realistis dan otentik, korpus dapat menjadi sumber data yang andal. Sitasi korpus memberikan informasi tentang perilaku sintaktis dari kata yang diamati. Sebagai sumber data dan refleksi penggunaan bahasa faktual (*natural language*), kemampuan korpus akan menutupi kekurangan yang ada dalam penelitian bahasa yang semata-mata mengandalkan intuisi. Prospek pengembangan *text corpora* kini melihat peluang untuk bahasa Indonesia. Mengingat bahasa Indonesia memiliki keunikan morfosintak dan kekayaan makna, suatu korpus bahasa Indonesia akan memberikan objek pengamatan yang sangat bermanfaat bagi studi kebahasaan dan proses pembelajaran.

DAFTAR RUJUKAN

- Aarts, J. 1991. Intuition-based and Observation-based Grammars. Dalam Aijmer, K. dan Altenberg, B. (Eds.). *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. New York: Longman.
- Aijmer, K. dan Altenberg, B. 1991. *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. New York: Longman.
- Ard, J. dan Gass, S. 1987. Lexical Constraints on Syntactic Acquisition. *Studies in Second Language Acquisition*, Vol. 9 No. 2, hlm. 233—52.
- Atkins, B.T., Kegl, J., dan Levin, B. 1988. Anatomy of a Verb Entry: From Linguistic Theory to Lexicographic Practice. *International Journal of Lexicography*, Vol. 1, No. 3.
- Fisher, C. 1994. Structure and Meaning in the Verb Lexicon: Input for a Syntax-added Verb Learning Procedure. *Language and Cognitive Processes*, Vol. 9, 1994, No. 4, hlm. 473—517.

- Fox, G. 1988. The Case for Examples. Dalam Sinclair, J. *Looking Up: An Account of the Cobuild Project in Lexical Computing*. London and Glasgow: Collins Publishers.
- Francis, G. 1993. A Corpus-Driven Approach to Grammar-Principles, Methods and Examples. Dalam Baker, M., Francis, G. dan Tognini-Bonelli, E. (Eds.). *Text and Technology: In Honour of John Sinclair 1993*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Halliday, M.A.K. 1991. Corpus Studies and Probabilistic Grammar. Dalam Aijmer, K. dan Altenberg, B. (Eds.). *English Corpus Linguistics: Studies in Honour of Jan Svartvik 1991*. New York: Longman.
- Kjellmer, G. 1991. A Mint of Phrases. Dalam Aijmer, K. and Altenberg, B. (Eds.). *English Corpus Linguistics: Studies in Honour of Jan Svartvik 1991*. New York: Longman.
- Levin, B. 1991. Building a Lexicon: The Contribution of Linguistics. *International Journal of Lexicography*, Vol. 4, No. 3.
- Little, D. 1994. Words and Their Properties: Arguments for a Lexical Approach to Pedagogical Grammar. Dalam Odlin, T. (Ed.). *Perspectives on Pedagogical Grammar*. New York: Cambridge University Press.
- Pinker, S. 1996. *Language Learnability and Language Development*. Cambridge, Massachusetts: Harvard University Press.
- Priyono. 1998. *The Study on the Verb Lexicon and Its Implication for Learning English as a Second Language*. Disertasi tidak diterbitkan. Sydney: Macquarie University.
- Selkirk, E. 1982. *The Syntax of Words*. Cambridge, Massachusetts: MIT Press.
- Sinclair, J. 1991. *Corpus Concordance Collocation*. Oxford: Oxford University Press.
- Sinclair, J. 1993. Text Corpora: Lexicographers' Needs. *Zeitschrift fuer Anglistik und Amerikanistik*, Vol. 41, No. 1, hlm. 5—14 dan Vol. 41, No. 2, hlm. 165—167.
- Willis, D. 1990. *The Lexical Syllabus: A New Approach to Language Teaching*. London: Colloing E.L.T.
- Yip, V. 1994. Grammatical Consciousness-raising and Learnability. Dalam Odlin, T. (Ed.). *Perspective on Pedagogical Grammar*. New York: Cambridge University Press.