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 survei 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 dikaikan 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 pastebord...
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 diconohkan 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 khusunya 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 peren-kanan.

Artikel ini mencoba membuka perspektif baru penggunaan korpus dalam memahami sifat dan karakteristik kosakata bahasa Inggris yang sekali-gus 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 se-bagai bahasa kedua/asing dengan menggunakan korpus sebagai sumber data. Kalimat-kalimat yang diikhtohok bukan buatan penulis, tetapi diambil dari korpus yang berasal dari teks bahasa Inggris otentik.

PERSPEKTIF HISTORIS PERKEMBANGAN KORPUSS

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.


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


INFORMASI DALAM KORPUSS

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 diconohkan 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-

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 alternasi 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**

 Kelaziman Penggunaan Bahasa


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.


13. ...the fire wouldn’t go and the breakfast wouldn’t cook...


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 PENGUNGAAN KORPUSS 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 dikaikan 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 (lexikal 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 masih 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 KORPUSS**


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 intuasi tidak penting bagi seorang pakar bahasa seperti yang dikemukakan oleh Francis sebagai berikut:

"... intuition may be 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).


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 invective than there are lines for a torrent of abuse (Fox, 1991:139).


**SUMBER TEKS DAN BESAR KORPUSS**

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.


PROGRAM-PROGRAM KORPUSS

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 tayangkan kalimat (sentence) secara utuh, yang dimulai dari kata awal dengan huruf pertama kapital dan diakhiri dengan tanda titik. Nilai 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


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....

Kedu 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


DAFTAR RUJUKAN


