DISCUSSION PUBLISHERS

Research Center · Open Access Edition

CORPUS LINGUISTICS: METHOD, THEORY, AND PR

Dana Waskita

Discussion Research ID	DPR-2026-1313
Year	2017
Field	Arts and Humanities
DOI	10.5614/sostek.itbj.2017.16.1.12
Author ORCID	0000-0001-7610-6408
Licence	CC-BY-SA
Original source	Jurnal Sosioteknologi
Permanent link	https://discussionpublishers.com/research/a88d0c20-a2aa-4320-a7f7-fb02f68bd2a8
Publisher	Discussion Publishers discussionpublishers.com

Abstract

Buku ini merupakan versi pertama yang memaparkan bagaimana linguistik korpus berkembang dan digunakan sebagai metodologi, teori utama, dan penggunaan korpus dalam bidang linguistik dan lintas disiplin. Buku ini terdiri dari 9 Bab, yang diawali dengan Bab I, What is Corpus Linguistics? Penulis mendefinisikan linguistik korpus sebagai sebuah bidang yang memfokuskan pada prosedur, atau metode mempelajari atau meneliti bahasa. McEnery dan Hardie juga menyinggung mengenai pendekatan yang digunakan dalam linguistik korpus yang juga dikemukakan oleh Tognini-Bonelli (2001) yang menyatakan adanya dua pendekatan linguistik korpus, yaitu: berbasis korpus (corpus-based) dan yang dikendalikan korpus (corpus-driven). Keduanya memiliki perbedaan dalam melihat korpus sebagai bukti yang mendukung teori. Yang pertama menggunakan pendekatan yang bersifat deduktif. Sementara itu corpus-driven approach menganggap korpus sebagai bukti yang harus menjadi acuan teori, sehingga bersifat induktif.

Republished by Discussion Publishers under the original CC-BY-SA licence. Copyright remains with the original authors; please cite the original work.

Cite as: Dana Waskita (2017). CORPUS LINGUISTICS: METHOD, THEORY, AND PRACTICE. Discussion Publishers, DPR-2026-1313.

RESENSI BUKU

Judul	: *Corpus Linguistics: Method, Theory and Practice*
Penulis	: Tony McEnery dan Andrew Hardie
Tebal Buku	: 294 halaman
Edisi/ISBN	: 2012/ 978-0-521-83851-1
Penerbit	: Cambridge

Oleh Dana Waskita

Buku ini merupakan versi pertama yang memaparkan bagaimana linguistik korpus berkembang dan digunakan sebagai metodologi, teori utama, dan penggunaan korpus dalam bidang linguistik dan lintas disiplin. Buku ini terdiri atas sembilan bab, yang diawali dengan bab kesatu, What is Corpus Linguistics? Penulis mendefinisikan linguistik korpus sebagai sebuah bidang yang memfokuskan pada prosedur, atau metode mempelajari, atau meneliti bahasa. McEnery dan Hardie juga menyinggung mengenai pendekatan yang digunakan dalam linguistik korpus yang juga dikemukakan oleh Tognini-Bonelli (2001). Tognini menyatakan adanya dua pendekatan linguistik korpus, yaitu berbasis korpus (*corpus-based*) dan yang dikendalikan korpus (c*orpus-driven*). Keduanya memiliki perbedaan dalam melihat korpus sebagai bukti yang mendukung teori. Yang pertama menggunakan pendekatan yang bersifat deduktif. Sementara itu *corpus-driven approach* menganggap korpus sebagai bukti yang harus menjadi acuan teori, sehingga bersifat induktif.

Bab kedua buku ini membahas cara menilai dan menganalisis data korpus. Cara menilai dan menganalisis data korpus dapat dimulai dari metadata, penanda teks (*textual markup*), dan anotasi. Metadata memberi informasi siapa yang memproduksi teks tersebut, kapan, dan dalam bahasa apa teks tersebut dibuat. Penanda teks meliputi tanda yang digunakan dalam penulisan teks itu sendiri seperti tipologi, misalnya cetak miring. Anotasi memberikan informasi mengenai fitur linguistik data tersebut, bisa berupa kelas kata seperti verba, nomina, ajektiva, atau adverbial.

Hal lain yang paling penting dalam korpus adalah tersedianya alat yang memungkinkan untuk mencari sebuah korpus dalam rangkaian unit tertentu seperti kata atau frasa yang ditampilkan dalam format baris secara vertikal, atau disebut konkordansi. Alat berupa *software* komputer yang digunakan untuk melakukan konkordansi berkembang dari generasi ke generasi. Yang pertama menggunakan komputer besar yang ditempatkan di satu lokasi dan hanya menampilkan KWIC, serta perlu *software* lain untuk mengolah yang lainnya. Generasi kedua menggunakan *Personal Computer* (PC) sehingga setiap orang dapat menggunakannya dan lebih banyak lagi data yang dapat dimasukkan, program bukan hanya menggunakan konkordansi KWIC, tetapi juga bisa mengurut secara alfabetis konteks sebelah kiri dan kanan kata yang dicari, serta dapat melakukan

Discussion Publishers edition DPR-2026-1313 https://discussionpublishers.com/research/a88d0c20-a2aa-4320-a7f7-fb02f68bd2a8

perhitungan statistik deskriptif dasar, hanya tidak bisa mengidentifikasi kata yang ditandai. Generasi ketiga dapat mengolah data lebih besar melalui PC, sedangkan generasi keempat berbasis *web*. McEnery dan Hardie (2003) juga membahas penggunaan statistik dalam linguistik korpus. Perhitungan statistik yang dapat digunakan untuk mengolah data linguistik korpus meliputi statistik deskriptif dan tes signifikansi untuk menguji seberapa mungkin hasil tertentu merupakan sebuah kebetulan, misalnya dalam menghitung kata kunci dan kolokasi.

Bab ketiga buku ini membahas permasalahan yang dihadapi dalam membuat korpus yang salah satunya adalah masalah hukum, khususnya hak cipta (*copyright*). Untuk memasukan dokumen sebagai data korpus perlu izin untuk dapat menampilkannya pada mesin pengolah korpus, apalagi bila disajikan pada *web*. Untuk mengantisipasi hal itu ada beberapa cara, pertama menghubungi pemegang hak cipta tersebut untuk meminta izin menggunakan dokumen untuk mesin pengolah korpus. Kedua adalah mengumpulkan data tanpa harus meminta izin, dengan konsekuensi hasilnya tidak didistribusikan kepada yang lain.

Pada bab keempat, penulis memaparkan "Linguistik Korpus Bahasa Inggris" yang sudah dirintis oleh beberapa universitas, antara lain University College London, yang membuat korpus terbesar untuk studi komparatif variasi bahasa Inggris yaitu International Corpus of English (ICE). Lancaster University, dipelopori oleh Leech dan Garside (1971), membangun CLAWS (Constituent Likelihood Automatic Tagging System) yaitu sistem anotasi kelas kata yang otomatis. Di University of Birmingham dipelopori oleh Sinclair membuat *corpus-based dictionary* (Collins COBUILD dictionary). Universite Catholique de Louvain oleh Sylviane Granger membuat International Corpus of Lerner English (ICLE). University of Nothingham memelopori korpus bentuk lisan tata bahasa lisan dalam bahasa Inggris. Pada Northern Arizona University dan Amerika pengembangan korpus dipelopori oleh Biber yang membuat korpus seperti ARCHER (A Representative Corpus of Historical English Registers), AAACL (American Association of Applied Corpus Linguistics). Selain itu, LDC (Linguistics Data Consortium), MICASE (Michigan Corpus of Academic Spoken English), BYU (Brigham Young) dan merupakan pusat linguistik korpus di Amerika.

Pada bab kelima, penulis memaparkan penelitian berbasis korpus variasi sinkronik dan diakronik. Banyak sekali karya berbasis korpus mengenai variasi register, antarwaktu, dan variabel internal lain. Sebagian karya variasi diakronik telah dilakukan dengan menggunakan metodologi umum perbandingan korpus antarwaktu (diakronik). Sebaliknya, penelitian variasi sinkronik berbeda bergantung pada variabel sosial atau registernya. Akan tetapi, keduanya dapat menggunakan metode kuantitatif ataupun kualitatif.

Bab keenam buku ini menjabarkan Linguistik Korpus Neo Firthian yang terdiri dari kolokasi dan wacana (*discourse*). Kolokasi menunjukkan bahwa aspek penting makna sebuah kata bukan hanya terkandung didalamnya tetapi juga bergantung pada kata lain yang menyertainya. Caranya yaitu dengan meneliti pola yang muncul bersamaan dalam korpus. Wacana

Discussion Publishers edition DPR-2026-1313 https://discussionpublishers.com/research/a88d0c20-a2aa-4320-a7f7-fb02f68bd2a8

(*discourse*) mengacu pada unit bahasa di atas tingkatan kalimat. Analisis Wanaca Kritis (AWK) memberi makna lebih mengenai *discourse* bahwa *discourse* merupakan praktik, sebuah karakteristik bahasa yang ditemukan dalam sebuah kelompok teks atau lebih luas lagi dalam bahasa komunitas. Dalam bab ini penulis juga menyatakan bahwa prosodi semantik yang disebut oleh Stubb merupakan sebuah konsep yang berasal dari konkordansi berdasarkan analisis neo-Firthian tentang kolokasi. Hal lain yang menjadi perhatian neo-Fithtian adalah leksis dan tata bahasa yang meliputi prinsip idiom, pola tata bahasa, dan *priming* leksikal. Dikatakan pula bahwa korpus masih jadi perdebatan apakah dianggap data atau dapat dianggap teori.

Pada bab ketujuh, penulis memaparkan klasifikasi linguistik teoretis menjadi dua, yaitu fungsionalisme dan metodologis. Fungsionalisme menginvestigasi bentuk bahasa, tetapi menerangkannya menurut fungsi bahasa itu digunakan. Bahasa tidak dianggap sebagai sistem yang terisolasi namun sesuatu yang digunakan untuk menyampaikan makna. Kelompok fungsionalisme selalu mempertimbangkan konteks dalam meneliti objek kajiannya.

Pada bab kedelapan McEnery dan Hardie (2003) menjelaskan hubungan antara linguistik korpus dan psikolinguistik dan linguistik fungsionalis. Hubungan korpus dengan psikolinguistik adalah data korpus dapat digunakan untuk menguji kealamiahan bahasa yang digunakan dalam soal ujian (tes), juga dalam penguasaan bahasa kedua. McEnery dan Hardie (2003) juga menjelaskan perbedaan antara linguistik korpus dan linguistik komputasional (*computational linguistics*). Mereka mengatakan linguistik korpus bertujuan untuk mencari sifat dan penggunaan bahasa, sedangkan linguistik komputasional bertujuan memodelkan sifat bahasa dengan menggunakan komputer. Selain itu, memfokuskan dirinya pada pemecahan masalah teknis yang melibatkan bahasa.

Secara umum kehadiran buku ini bisa menginspirasi penelitian interdisipliner antara ilmu kebahasaan dengan menggunakan teknologi komputer. Kemajuan teknologi informasi dan komputer dapat bersinergi dengan ilmu kebahasaan atau linguistik sehingga dapat membantu penelitian linguistik lebih ekstensif. Perkembangan linguistik korpus juga dapat diterapkan lebih luas dalam membangun korpus bahasa Indonesia.

DAFTAR PUSTAKA

Adolphs, S. (2006). I*ntroducing Electronic Text Analysis: A practical guide for language and literary studies*. New York: Routledge.

Baker, P. (2010). *Sociolinguistics and Corpus Linguistics.* Edinburgh: Edinburgh University Press.

Hardie, A. and McEnery, T. (2003). "The were-subjunctive in British rural dialects: Marrying corpus and questionnaire data." *Computers and the Humanities*, 37 (2), 205–208.

Leech, G. (1971). *Meaning and the English Verb*. London: Longman.

Stubbs, M. (2001). *Words and Phrases: Corpus Studies of Lexical Seman.* Oxford: Blackwell Publisher.

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam: John Benjamins.

Discussion Publishers edition DPR-2026-1313 https://discussionpublishers.com/research/a88d0c20-a2aa-4320-a7f7-fb02f68bd2a8