

JELAJAH LINGUISTIK

Rubrik ini membuka peluang untuk saling berbagi di antara kita tentang beberapa kemungkinan topik ini:

- a. penancangan metode penelitian linguistik yang belum lazim digunakan
- b. daur-ulang metodologi penelitian linguistik
- c. persoalan data yang – meskipun barangkali belum ditemukan pemecahannya – penelusurannya berpeluang membuka sesuatu yang baru yang belum pernah menjadi perhatian peneliti terdahulu
- d. penerapan teori linguistik tertentu untuk menjelaskan data bahasa seperti bahasa Indonesia yang membuat peneliti mempersoalkan teori yang bersangkutan

BAGAIMANA TEKNOLOGI DAPAT MEMBANTU METODOLOGI

Faizah Sari

Universitas Katolik Indonesia Atma Jaya

faizah.sari@atmajaya.ac.id

Selama hampir delapan tahun menggeluti kajian partikel pragmatik (*pragmatic particles*), saya memperoleh satu pelajaran penting mengenai metodologi analisis fungsi partikel pragmatik dalam penggunaannya dalam interaksi lisan yang kerap tertuang pada dua fokus analisis (Östman, ed. 2011): (1) pengelolaan wacana (*discourse*) dalam percakapan, dan (2) strategi pengungkap sudut interpersonal penutur pada saat menggunakan partikel. Fokus pertama analisis fungsi partikel pragmatik sebagai pengelola wacana merupakan tanda memulai atau memperkenalkan topik dalam percakapan, mengecek pengetahuan kawan bicara tentang topik yang sedang dibicarakan, dan memberi tekanan (*emphasis*) pada bagian kata tertentu yang ingin disampaikan. Fokus yang kedua, partikel pragmatik merupakan strategi penutur untuk menjaga kesetaraan (*solidarity*) kawan bicara dan menjaga citra penutur (*face-saving*).

Penelitian partikel pragmatik menjadi semakin menarik karena jumlah partikel pragmatik yang cukup banyak dalam bahasa-bahasa di Indonesia dan ada sebagian yang berfungsi mirip. Pentingnya fungsi partikel pragmatik sebagai pengelola wacana dan pengungkap sudut interpersonal penutur bergantung erat pada sejauh mana penyebaran partikel pragmatik pada percakapan (Sari, 2008). Hal ini menjadi krusial karena penelitian di bidang ini memasuki tahap di mana pendekatan tradisional seperti merekam dan mentranskripsi satu bahasa saja tidak mampu menguak spektrum fungsi partikel pragmatik. Pada titik inilah intervensi teknologi diperlukan – atau paling tidak kolaborasi sejumlah metodologi yang dapat membantu peneliti mengolah data dengan efektif sehingga memberikan hasil yang komprehensif.

Pada penelitian bahasa-bahasa di dunia yang sudah populer, misalnya Bahasa Inggris, salah satu dukungan teknologi yang membantu penelitian partikel pragmatik adalah penggunaan korpus. Sayangnya, hingga tahun 2013 ini korpus bahasa-bahasa di Indonesia belum dikelola dengan skala besar dan belum ada satu jaringan data yang sudah bebas akses seperti layaknya bahasa-bahasa besar lain di dunia. Cukup dimaklumi penelitian bahasa-bahasa di Indonesia masih berhadapan dengan infrastruktur akademik, jumlah peneliti, pilihan bahasa yang diteliti, dan aspek yang diteliti. Di samping itu, pemeliharaan korpus adalah pertemuan dua disiplin ilmu besar, yaitu *Information Technology* dan *Linguistics*.

Sebenarnya, kolaborasi dua disiplin ilmu itu dapat menghasilkan, mengumpulkan, dan mengatur data linguistik untuk tujuan penelitian dan akademis lainnya, termasuk pedagogi. Penelitian linguistik Indonesia telah menarik kolaborasi seluruh dunia yang memperluas batas-batas ilmiah dan menemukan kembali makna baru lintas-budaya (Sneddon, 2003; Kaswanti Purwo, 1989; 2003). Pusat penelitian terkemuka seperti Max Planck Institute Jakarta Field Station dan SIL International telah menjadi dukungan yang konsisten untuk penelitian linguistik

dan sosial budaya tentang dan di Indonesia. Kedua pusat internasional telah mengembangkan kerja sama dengan Pusat Kajian Bahasa dan Budaya di Universitas Katolik Indonesia Atma Jaya Indonesia, Jakarta selama dua dekade terakhir. Kolaborasi lain yang telah dijalin antara lain dengan The Center of International Corporation untuk Komputerisasi (CICC, Jepang) dan Pusat Studi Teknologi dan Aplikasi (Badan Pengkajian dan Penerapan Teknologi), dalam penelitian tentang terjemahan mesin (Kaswanti Purwo, 1989). Sayangnya, kembali semua upaya tersebut terwujud agak sporadis (lih. Moeliono, 1993; Ewing, 2000). Saat inilah waktu yang tepat untuk membangun kembali jaringan penelitian dan menawarkan solusi praktis untuk pengembangan korpus.

Perspektif teoretis yang menginformasikan titik persimpangan korpus bahasa dan akses untuk penggunaannya melalui teknologi dapat secara luas digambarkan sebagai "pembangunan korpus" yang memerlukan sarana sistematis analisis untuk menjelaskan fungsi linguistik, termasuk analisis teks tertulis dan wacana lisan, dengan menangkap fitur yang paling dasar dari bagian leksema. Korpus adalah "sampel yang representatif dari bahasa, dikompilasi untuk tujuan analisis linguistik" (Crystal, 1997:474). Suatu analisis linguistik kontemporer pasti melibatkan berbagai tahapan yang direncanakan dalam penyelidikan ilmiah dan empiris bahasa, misalnya, pengumpulan data melalui seleksi penutur, rekaman, dan menyalin, dan menyimpan. Sebuah korpus melalui aplikasi yang berbeda dalam sistem yang dapat menampilkan kata-kata tertentu, frasa, suara tertentu, terjemahan, konstruksi, atau frekuensi kata, menyimpan data linguistik untuk penggunaan peneliti. Korpus kemudian merupakan bagian integral dari gambar yang lebih besar studi dalam linguistik.

Pengembangan korpus jelas mengambil keuntungan dari teknologi dalam menyediakan sebuah protokol sistem yang merevolusi organisasi data, kecepatan, dan ketepatan untuk membuat korpus untuk beroperasi secara efektif. Saat ini ada korpora yang terus berkembang dan tampil lebih baik dalam memberikan data linguistik dari perspektif yang berbeda. Misalnya, beberapa menyediakan data linguistik sebagai bahasa keseluruhan, sedangkan korpora jenis lain menyediakan data teks tertentu. Contoh-contoh korpus yang sudah berkembang dan dapat diakses dari mana saja termasuk korpus Bahasa Inggris Amerika (*Corpus of Contemporary American English*) yang terdiri dari sekitar 400 juta kata. Korpus Nasional Inggris (*British National Corpus*) berisi sekitar 100 juta kata. Di sisi lain, beberapa korpora sangat spesifik, misalnya Korpus Bahasa Inggris Akademik Michigan (MICASE) yang menyediakan lebih dari 1,8 juta kata data dari bahasa lisan akademik.

Korpus bahasa-bahasa Indonesia sebaiknya berdampingan dengan korpora bahasa kontemporer kawasan Asia Tenggara yang serupa, seperti Bahasa Melayu (*Malay Concordance Project*), sehingga selanjutnya korpus Indonesia bermaksud untuk secara aktif memperluas akan tersedia untuk tujuan akademik dengan menyediakan akses *online* bagi para peneliti dan mahasiswa Indonesia untuk belajar lebih banyak tentang penggunaan bahasa. Pada akhirnya, korpus bisa menjadi peta gagasan plurikultural bahasa-bahasa di Indonesia dan menjadi sumber dokumentasi bahasa dan pangkalan data (*database*) serta penelitian bahasa di Indonesia.

Singkatnya, pengolahan teks dan penciptaan korpora meliputi pencarian teks, pengindeksan, dan data mengkategorikan kata-kata dan frasa Indonesia. Data yang sudah ada, ditambah rencana konsisten untuk meningkatkan jumlah kata, akan menjadi titik balik perkembangan korpus yang beragam dan komprehensif. Lebih penting lagi, korpus dapat membantu menghasilkan, mengumpulkan, dan mengatur data linguistik untuk tujuan penelitian dan akademis lainnya, termasuk pedagogi. Pada titik pandang penelitian, protokol penelitian interdisipliner terlihat berbeda dan dapat 'bercabang' tidak hanya membantu aspek khusus seperti penelitian tentang partikel pragmatik, namun penelitian tentang korpora sendiri.

Upaya untuk mengembangkan dan memperluas korpus Indonesia yang meliputi pangkalan data yang dapat diandalkan berisi kata-kata, frasa, dan kalimat dari berbagai sumber merupakan upaya kolaborasi yang berasal dari lebih dari satu disiplin penelitian. Saat ini wilayah yang paling relevan dalam platform korpus adalah linguistik dan teknologi informasi.

Di satu sisi, studi linguistik menyediakan data bahasa penting melalui serangkaian pengujian empiris yang sering melibatkan analisis percakapan, studi etnografi, sosiolinguistik interaksional, dan metode lain dalam menganalisis wacana. Di sisi lain, teknologi informasi berusaha untuk berimprovisasi alat dan mengelola sejumlah besar data dengan merancang desain sistem yang akan mampu mengakomodasi semua proses penelitian.

Pertanyaan kritis dalam pembangunan suatu korpus adalah: bagaimana hasil penelitian sebelumnya diintegrasikan ke dalam korpus bahasa Indonesia dapat dibuat dan akhirnya akan tersedia untuk digunakan bagi siapa saja yang mungkin memerlukan data linguistik? Pengaruh eksponensial teknologi informasi terus menjadi indikasi titik balik dalam penelitian empiris melalui program yang dapat merevolusi analisis data, kecepatan, dan akurasi. Teknologi informasi dapat membantu merancang sistem yang diperlukan untuk membangun platform yang berjalan efektif untuk menjaring titik persimpangan dari akademisi dan budaya melalui pemodelan sebuah teknologi *open source* yang idealnya dapat menjadi akses publik.

Pembangunan korpus memerlukan aplikasi sistem *concordancing* untuk mengatur dan menyimpan data linguistik efektif, yang meliputi perbandingan kata atau sinonim, analisis kata kunci, frekuensi kata, indeks, atau banyak unsur linguistik lainnya. Beberapa menggunakan aplikasi perangkat lunak *open source corpora* (ApSIC XBench, WordSmith, MonoConc, GlossaNet/Unitex), yang lain mengembangkan perangkat lunak bebas yang digunakan dalam akademisi, seperti AdTAT (The University of Adelaide, Australia Selatan), CorpusEye (Syddansk Universitet, Denmark), dan *Malay Concordance Project* (Australia National University, Australia). Aplikasi *concordancing* membantu memilah data ke memori tertentu untuk digunakan sebagai masukan untuk setiap jenis analisis yang dibantu komputer, termasuk terjemahan bahasa dan rekonstruksi.

Sarana umum yang terlibat di balik cara kerja sistem *concordancing* yang baik adalah alat pengolahan teks, dikenal sebagai *tagging* dan *parsing* (Crystal, 1997; Altenberg & Granger, 2002; Fromkin et al., 2003). *Tagging* merupakan pemberian label pada sebuah kata untuk menunjukkan kelas kata (nomina, verba tindakan, atau yang lain) atau konstruksi besar (frasa, klausa, atau informasi gramatikal lainnya). Misalnya, *London-Lund Corpus of Spoken English* adalah salah satu *corpora online* yang memperkenalkan *tag* diatur melalui pemetaan otomatis menggunakan penjelasan leksiko-gramatikal (lih. *London-Lund Corpus of Spoken English*, Crystal, 1997:415). *Parsing*, di sisi lain, merupakan prosedur untuk menentukan urutan gramatikal kata dalam korpus sesuai dengan aturan struktur sintaksis untuk mendapatkan tingkat yang lebih luas makna. Misalnya, SYSTRAN adalah salah satu perangkat lunak mesin terjemahan terkemuka yang menggunakan *parsing* untuk menghasilkan struktur semantik kalimat (lih. SYSTRAN, Crystal, 1997:352). Kedua prosedur dalam mengembangkan korpus elektronik ini akhirnya mampu membantu mempercepat pencarian (*search*) di dalam pangkalan data.

DAFTAR PUSTAKA

- Altenberg, Bengt dan Granger, Sylviane. 2002. *Lexis in Contrast. Corpus-based Approaches*. Amsterdam: John Benjamins.
- Crystal, David. 1997. *The Cambridge Encyclopedia of Language* (Edisi ke-2). Cambridge: Cambridge University Press.
- Ewing, Michael. 2000. Conversations of Indonesian language students on computer-mediated projects: Linguistic responsibility and control. *Computer Assisted Language Learning*, 13(4-5), 333-356.
- Fromkin, Victoria, et al. 2003. *An Introduction to Language*. Edisi ke-7. Boston: Heinle.
- Kaswanti Purwo, Bambang. (ed.) 1989. *Linguistik dan Teknologi Komputer: Pemrosesan Bahasa Alami*. Bandung: ITB Press.

- Moeliono, Anton. 1993. The first effort to promote and develop Indonesian. Dalam J. A. Fishman (Ed.), *The Earliest Stage of Language Planning: The First Congress Phenomenon* (hlm. 129-142). Berlin: Mouton de Gruyter.
- Östman, Jan-Ola, Zienkowski, Jan, & Verschueren, Jef (ed.). 2011. *Discursive Pragmatics*. Amsterdam: John Benjamins.
- Sari, Faizah. 2008. Teaching Pragmatic Particles in the LCTL Classroom. *The Journal of the National Council of Less Commonly Taught Languages*, 5, 129-159.
- Sneddon, James. 2003. *The Indonesian Language: Its History and Role in Modern Society*. Sydney: University of New South Wales Press.