

COMPUTATIONAL LINGUISTICS
(Model Baru Kajian Linguistik dalam Perspektif Komputer)

Oleh: Musthofa

Fakultas Adab UIN Sunan Kalijaga Yogyakarta
Jl. Marsda Adisucipto Yogyakarta 55281
e-mail: top_bsa@yahoo.com

Abstract

This paper describes a new discipline in applied linguistics studies, computational linguistics. It's a new model of applied linguistics which is influenced by computer technology. Computational linguistics is a discipline straddling applied linguistics and computer science that is concerned with the computer processing of natural languages on all levels of linguistic description. Traditionally, computational linguistics was usually performed by computer scientists who had specialized in the application of computers to the processing of a natural language. Computational linguists often work as members of interdisciplinary teams, including linguists (specifically trained in linguistics), language experts (persons with some level of ability in the languages relevant to a given project), and computer scientists. The several areas of computational linguistics study encompasses such practical applications as speech recognition systems, speech synthesis, automated voice response systems, web search engines, text editors, grammar checking, text to speech, corpus linguistics, machine translation, text data mining, and others. This paper presents the definition of computational linguistics, relation between language and computer, and area of computational linguistics studies.

Tulisan ini menjelaskan sebuah disiplin baru dalam kajian linguistik terapan yang dinamai "linguistik komputer" atau "linguistik komputasi". Kajian ini merupakan model baru dalam kajian linguistik terapan yang dipengaruhi oleh

teknologi komputer. Linguistik komputasi adalah disiplin kajian interdisipliner antara linguistik terapan dan ilmu komputer yang menitikberatkan kajiannya pada pemrosesan bahasa natural oleh komputer pada semua tingkatan deskripsi linguistik. Linguistik komputasi memiliki beberapa sumber dan bidang kajian yang masing-masing memiliki tujuan, metode, dan latar belakang sejarah sendiri. Untuk itu, tulisan ini menyajikan beberapa hal berkaitan dengan definisi linguistik komputasi, hubungan antara bahasa dan komputer, dan bidang kajian linguistik komputer.

Kata kunci: komputer; linguistik komputer; bahasa; pemrosesan bahasa natural.

A. PENDAHULUAN

Kehadiran teknologi komputer dan perkembangannya yang begitu pesat dewasa ini telah berpengaruh besar terhadap berbagai aspek kehidupan manusia, termasuk bahasa. Disadari atau tidak, ketika seseorang mengoperasikan sebuah komputer, ia akan senantiasa bersentuhan dengan bahasa. Mulai dari saat seseorang menghidupkan komputer hingga ia menjalankan aplikasi-aplikasi sesuai pilihannya di dalam komputer, ia akan selalu bersentuhan dengan bahasa yang ditampilkan di layar monitor.

Ketika seseorang mulai menekan tombol “*power*” atau “*on*” hingga proses *booting* komputer selesai misalnya, ia akan disugahi informasi-informasi di layar monitor dalam bentuk bahasa yang menginformasikan kepadanya mengenai proses yang sedang berjalan di dalam komputer hingga komputer selesai *booting* dan siap digunakan. Pada saat itu, tampak di layar monitor sebuah program aplikasi seperti windows, mac, linux, atau yang lainnya. Kemudian setelah program aplikasi tersebut siap dan tampak di layar monitor, maka seseorang akan disugahi berbagai menu, juga dalam bentuk bahasa yang menginformasikan berbagai bentuk program aplikasi yang disediakan di dalamnya. Ada program untuk memproses teks, gambar, musik, video, dan lain sebagainya. Bahasa-bahasa ini

akan selalu ada pada setiap program aplikasi di dalam komputer, sehingga setiap kali seseorang berinteraksi dengan komputer, dia akan selalu bersentuhan dengan bahasa.

Pada saat seseorang berinteraksi dengan komputer ini, maka seseorang seakan-akan berdialog dengan komputer, dan komputer seakan-akan mampu berbicara dan memberi tahu pengguna tentang apa yang dibutuhkannya melalui bahasa yang bisa dipahami oleh manusia. Padahal kita tahu, bahwa komputer hanyalah sebuah mesin atau teknologi, tetapi senyatanya, komputer mampu menampilkan bahasa manusia sehingga bisa dipahami oleh manusia. Dari sini menjadi jelas, bahwa komputer tidak bisa dilepaskan dari bahasa dan bahasa merupakan bagian penting bagi komputer.

Lazimnya, yang biasa merangkai, menggunakan, memproses, menampilkan bahasa, dan bahkan mengucapkannya adalah manusia. Namun, mesin teknologi yang dikenal sebagai "komputer" ini senyatanya saat ini mampu melakukan kegiatan memproses bahasa sebagaimana yang dilakukan oleh manusia sebagai pemilik bahasa. Dalam hal ini, komputer mampu memproses bahasa-bahasa yang dimiliki manusia, yang dalam istilah linguistik dikenal sebagai "bahasa natural" (*natural language*) dalam kerangka aturan-aturan linguistik yang juga telah ditetapkan oleh manusia. Komputer, sebagaimana kita ketahui, mampu merangkai bahasa, menggunakan, memproses, menampilkan, mengoreksi kesalahan penulisan kata, memaknai kata, menerjemahkan kalimat, dan bahkan mengucapkannya sebagaimana yang dilakukan oleh manusia. Bahkan apa yang dilakukan oleh mesin komputer seringkali lebih akurat ketimbang yang dilakukan oleh manusia. Kemampuan komputer di dalam memproses bahasa natural (*natural language*)¹ sebagaimana yang dilakukan oleh manusia ini kemudian dikenal dengan *natural*

¹ Bahasa natural (*natural language*) adalah bahasa yang dikembangkan oleh manusia dan dijadikan sebagai sarana untuk berkomunikasi antar sesama (Walter, 2008)

language processing 'pemrosesan terhadap bahasa natural' yang biasa disingkat menjadi NLP.²

Tentu saja, komputer tidak bisa melakukan semua proses terhadap bahasa natural itu sendiri tanpa adanya campur tangan manusia. Bagaimanapun juga, komputer hanyalah sebuah mesin yang tidak bisa berpikir sebagaimana manusia berpikir dan komputer juga tidak bisa bekerja memproses bahasa natural itu sendiri tanpa adanya campur tangan manusia. Artinya, sebuah mesin komputer membutuhkan manusia untuk memprogramnya agar dapat berpikir dan bekerja di dalam memproses bahasa natural sebagaimana yang dilakukan oleh manusia. Upaya inilah yang kemudian memunculkan sebuah disiplin ilmu baru dalam bidang linguistik yang disebut "linguistik komputer" (*computational linguistics*). Apa dan bagaimana gambaran mengenai disiplin baru dalam kajian linguistik ini dan apa saja ruang lingkup dan bidang kajiannya, hal inilah yang akan coba diuraikan dalam tulisan ini.

B. BAHASA SEBAGAI BAGIAN PENTING DARI KOMPUTER

Jika berbicara mengenai bahasa, mungkin jarang terlintas di benak pemikiran banyak orang, bahwa bahasa memiliki keterkaitan erat dengan teknologi komputer karena pada umumnya bahasa hanyalah dianggap sebagai sarana komunikasi bagi manusia dan tidak memiliki kaitan dengan komputer. Namun, jika diperhatikan dengan seksama, maka sebenarnya bahasa memiliki hubungan keterkaitan yang sangat erat dengan komputer. Bahkan bisa dikatakan, bahwa komputer tidak bisa dilepaskan dari bahasa karena bahasa merupakan aspek penting bagi komputer.

² *Natural Language Processing* (pemrosesan terhadap bahasa natural) adalah sebuah analisis terhadap bahasa manusia (bahasa natural) yang dilakukan oleh komputer. Sebagai contoh adalah analisis otomatis terhadap sebuah teks untuk menentukan struktur gramatikal yang digunakannya, atau sebuah proses analisis akustik terhadap input bahasa ucap (Richards, 2002: 352).

Semenjak diciptakan dan dikembangkannya teknologi komputer, bahasa kemudian menjadi salah satu komponen penting dalam rancang bangun sebuah sistem yang ada di dalam komputer. Sebuah komputer tidak bisa bekerja tanpa adanya bantuan bahasa di dalam sistemnya, karena sebuah komputer membutuhkan sistem kerja yang dirancang menggunakan bahasa dalam bentuk sederetan program-program aplikasi. Hal inilah, sebenarnya, yang menjadikan komputer tidak bisa dilepaskan dari bahasa.

Ditinjau dari sisi keterkaitannya, hubungan antara bahasa dan komputer bisa diklasifikasikan menjadi 4 (empat) macam, yaitu:

1. Bahasa nama produk atau komponen (*product name language*). Bahasa ini berhubungan dengan setiap nama produk ataupun nama bagi setiap komponen yang ada di dalam komputer. Adanya produk atau komponen-komponen dalam komputer ini kemudian berpengaruh terhadap munculnya istilah-istilah baru dalam bahasa dalam bidang komputer, seperti *keyboard*, *mouse*, *windows*, *motherboard*, dan *harddisk*. Tanpa adanya bahasa ini, sulit bagi seseorang untuk mengenali dan memahami komputer. Dalam konteks ini, maka komputer butuh terhadap bahasa.
2. Bahasa mesin (*machine language*), yaitu bahasa yang berkaitan dengan mesin dan hanya dapat dipahami oleh mesin, sehingga mesin bisa mengerjakan apa saja yang dikehendaki oleh pengguna komputer. Bahasa ini hanya terdiri dari kode-kode angka yang hanya terdiri dari gabungan angka 1 dan 0 yang biasa disebut *binary digit* disingkat menjadi *bit*. Bahasa mesin inilah yang dijadikan sebagai basis konversi bagi bahasa natural (*natural language*) sehingga bahasa natural bisa dibaca dan dipahami oleh mesin dan ditampilkan kembali di layar monitor untuk selanjutnya bisa dipahami oleh pengguna komputer. Sejak komputer membutuhkan bahasa untuk membangun sistem kerjanya, para ahli komputer kemudian menciptakan sistem

konversi baku atas semua huruf atau abjad dan angka dalam bahasa natural ke dalam kode-kode angka biner. Di antara kode konversi standar yang banyak dikenal adalah ASCII dan UNICODE. Pada awalnya, pengkodean abjad dan angka yang banyak digunakan adalah ASCII. Namun, karena adanya tuntutan agar komputer mampu mengakomodasi berbagai bahasa dunia yang berbeda-beda, maka para ahli komputer kemudian mengembangkan sistem pengkodean abjad dan angka yang kemudian dikenal dengan UNICODE. Melalui sistem pengkodean standar UNICODE ini, semua bahasa di dunia ini bisa dikenali, dibaca, dan dipahami oleh komputer. Contoh untuk sistem pengkodean standar ASCII dan UNICODE adalah sebagai berikut.

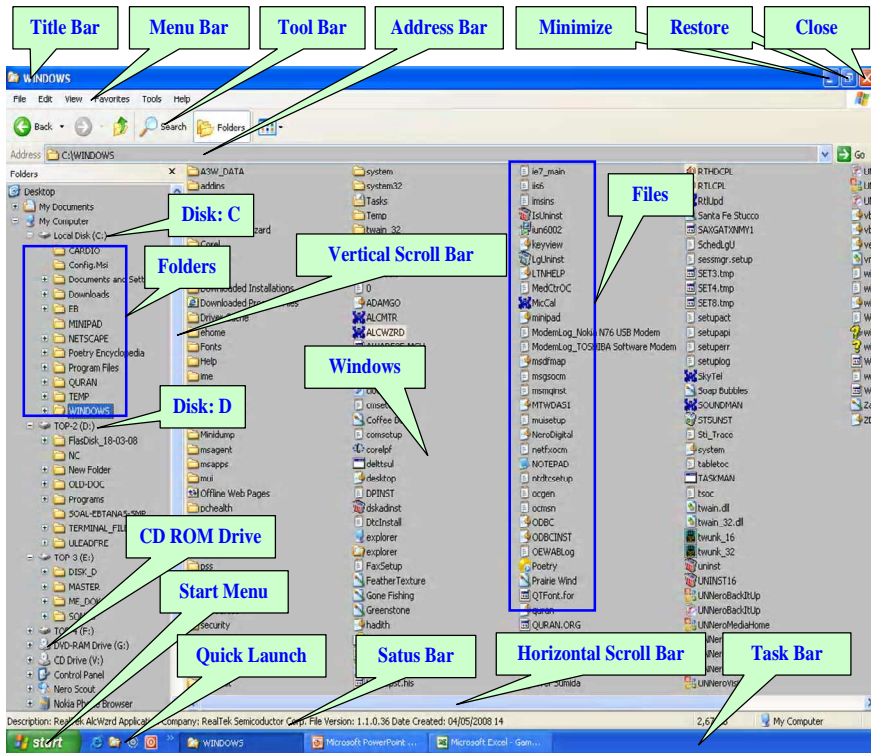
No Abjad	Karakter	Keterangan	Kode Biner ASCII	UNICODE
65	A	Abjad "A" besar	1000001	U+0041
97	a	Abjad "A" kecil	1100001	U+0061
66	B	Abjad "B" besar	1000010	U+0042
70	F	Abjad "F" besar	1000110	U+0046

Melalui sistem pengkodean huruf, abjad atau karakter ini, komputer bisa memproses dan memahami bahasa natural.

3. Bahasa pemrograman (*programming language*), yaitu bahasa yang berisi sederetan perintah yang menjembatani antara bahasa mesin dan bahasa pengguna atau bahasa natural. Sederet perintah dalam bahasa pemrograman ini berupa sintaksis-sintaksis tertentu yang dirancang secara khusus untuk membuat dan menjalankan program-program tertentu pula. Ada banyak bahasa program dalam dunia komputer, yang di antaranya adalah bahasa Dos, bahasa C,

bahasa Visual Basic, Foxpro, Delphi, Prolog, PHP, dan Java yang masing-masing memiliki spesifikasinya yang berbeda-beda. Namun, setiap bahasa program tersebut sudah disusun secara khusus dan menjadi semacam rumus khusus yang berisi sintaksis-sintaksis dari bahasa natural. Sederet perintah yang berisi sintaksis-sintaksis tertentu ini kemudian dikompilasikan menggunakan program *compiler* sehingga membentuk serangkaian program atau sistem yang saling berkaitan satu dengan yang lainnya secara otomatis. Melalui perantara bahasa pemrograman inilah, komputer bisa dirancang seakan menjadi seperti manusia yang bisa memahami bahasa natural, memprosesnya, dan menampilkannya kembali di layar monitor sebagaimana bahasa natural yang biasa digunakan dan dipahami oleh manusia. Dengan demikian, bahasa natural apapun yang ada di dunia ini bisa diproses dan dipahami oleh komputer.

4. Bahasa perantara antara komputer dan penggunanya (*user interface language*), yaitu bahasa tampilan yang ada di layar monitor komputer yang dapat dipahami oleh pengguna komputer. Biasanya bahasa ini berupa bahasa natural yang sudah diprogram secara otomatis dan ditampilkan di layar monitor saat komputer hidup. Bahasa ini berbentuk menu-menu, ikon, nama-nama *file* atau folder ataupun bahasa bantuan (*help*) yang berfungsi untuk memandu pengguna komputer di dalam proses mengoperasikan komputer. Pada umumnya, bahasa perantara antara komputer dan penggunanya (*user interface language*) ini merupakan bagian dari sebuah sistem operasi (*operating system*) seperti Dos, Windows, MAC, UNIX, Linux, bagian dari program aplikasi perkantoran (*office application programs*) seperti Microsoft Office, Mac Office, StarOffice, dan KDE, bagian dari program aplikasi yang dirancang untuk membuat program seperti Visual Basic, Foxpro, Delphi atau program-program pendukung yang lain. Contoh untuk bahasa antarmuka pengguna tersebut dapat dilihat pada gambar berikut.



Gambar 1. Tampilan *Windows Explorer* berbahasa Inggris

Pada gambar di atas, tampak jelas bahwa setiap bagian pada tampilan *windows explorer* yang ada di layar monitor komputer berkaitan dengan bahasa. Artinya, bahasa dibutuhkan untuk menandai atau menamai setiap bagian yang tampak pada tampilan *windows explorer* sebagaimana tampak pada gambar, baik yang berupa menu-menu, ikon, nama-nama file, folder, atau bagian-bagian lain yang tampak pada gambar di atas. Bahkan, setiap ikon yang tampak pada layar monitor tersebut jika ditunjuk menggunakan *mouse* maka pada ikon tersebut akan ditampilkan bahasa dalam bentuk kata atau rangkain kata-kata sesuai dengan fungsi dari setiap ikon tersebut. Bahasa tersebut tidak ada begitu saja, tetapi bisa ada setelah melalui proses pemrograman.

Dengan demikian, tampak jelas bahwa komputer tidak bisa dilepaskan dari bahasa. Komputer membutuhkan bahasa untuk memandu pengguna di dalam proses mengoperasikan komputer. Dalam hal ini, bahasa memiliki peran dan fungsi yang sangat penting di dalam komputer. Bahasa merupakan sarana yang sangat penting untuk mengeksplorasi dan menjelaskan sistem dan cara kerja komputer. Bahasa merupakan sarana di dalam membuat program-program aplikasinya dan juga berfungsi sebagai bahasa perantara antara komputer dan penggunanya (*user interface language*) yang ada di layar monitor sehingga dapat dipahami oleh pengguna komputer. Di sisi lain, bahasa juga membutuhkan program-program komputer untuk mengelola dan memproses data-data bahasa.

Melalui komputer, bahasa dapat dikelola, digunakan, dan diperlakukan sebagai apa saja sebagaimana manusia mengelola, menggunakan, dan memperlakukan bahasa. Dalam konteks komputer, bahasa bisa diperlakukan sebagai media komunikasi yang tertampilkan di layar monitor, bisa diucapkan secara *audio* sebagaimana manusia mengucapkannya, bisa dibuat sebagai kamus elektronik, dan bahasa bisa dijadikan sebagai *command* 'perintah' yang mampu menggerakkan sebuah sistem di dalam komputer, sehingga dengan bahasa ini komputer mampu memprogram dan menjalankan apa saja yang dikehendaki oleh manusia. Perpaduan antara komputer dan bahasa inilah yang kemudian memunculkan sebuah disiplin baru linguistik yang disebut *Natural Language Programming* (pemrograman bahasa natural) disingkat NLP, atau dalam istilah lain dikenal dengan *Computational Linguistics* (linguistik komputer).

C. LINGUISTIK KOMPUTER (COMPUTATIONAL LINGUISTICS)

Di masa lampau, sebelum ditemukannya teknologi komputer, bahasa natural (*natural language*) atau bahasa yang dimiliki oleh manusia hanya dikelola dan diproses secara manual oleh manusia. Namun, sejak ditemukannya komputer dan seiring pula

dengan kemampuan komputer yang mampu mengelola dan memproses berbagai bentuk data, maka manusia bukan lagi menjadi satu-satunya subjek di dalam mengelola dan memproses bahasa natural karena komputer dengan berbagai kecanggihan dan kemampuannya juga mampu mengelola dan memproses bahasa natural sebagaimana yang dilakukan oleh manusia.

Kemampuan komputer di dalam mengelola dan memproses bahasa natural ini tidak bisa ada begitu saja, tetapi harus melalui rekayasa pemrograman secara sistematis yang dilakukan oleh manusia terhadap komputer. Dengan rekayasa pemrograman tersebut, komputer mampu mengelola dan memproses berbagai hal yang berkaitan dengan bahasa sebagaimana yang biasa dilakukan oleh manusia. Keinginan dan kebutuhan manusia untuk menjadikan komputer mampu mengelola dan memproses berbagai hal yang berkaitan dengan bahasa natural inilah yang melatarbelakangi munculnya disiplin baru dalam linguistik yang disebut *Computational Linguistics* (Linguistik Komputer).

Berkaitan dengan hal tersebut, ada beberapa definisi yang dikemukakan oleh para ahli bahasa dan komputer mengenai Linguistik Komputer (*Computational Linguistics*). Igor A. Bolshakov and Alexander Gelbukh menyatakan,

“Computational linguistics might be considered as a synonym of automatic processing of natural language, since the main task of computational linguistics is just the construction of computer programs to process words and texts in natural language (Linguistik komputer dianggap sebagai padanan bagi istilah pemrosesan secara otomatis terhadap bahasa natural, sebab tugas utama linguistik komputer hanya mengkonstruksikan program-program komputer agar dapat memproses kata-kata dan teks pada bahasa natural)” (Bolshakov, 2004: 25).

Sementara itu, Ruslan Mitkov menyatakan, *“Computational linguistics is an interdisciplinary field concerned with the processing of language by computer”* ‘Linguistik Komputer adalah sebuah bidang kajian interdisipliner berkenaan dengan pemrosesan bahasa natural oleh komputer’ (Mitkov, 2004: ix). Dalam kamus *Routledge*

Dictionary of Language and Linguistics disebutkan bahwa linguistik komputer didefinisikan sebagai:

Discipline straddling linguistics and (applied) computer science that is concerned with the computer processing of natural languages on all levels of linguistic description. 'Sebuah disiplin ilmu yang menggabungkan antara ilmu linguistik dengan aplikasi ilmu komputer yang berkenaan dengan pemrosesan bahasa natural oleh komputer dalam berbagai level deskripsi linguistik' (Bussmann, 1996: 224).

Definisi lain mengenai linguistik komputer juga dikemukakan oleh Jack C. Richards and Richard Schmidt. Mereka mengatakan,

Computational Linguistics is the scientific study of language from a computational perspective. Computational linguists are interested in providing computational models of natural language processing (both production and comprehension) and various kinds of linguistic phenomena. The work of computational linguists is incorporated into such practical applications as speech recognition systems, speech synthesis, automated voice response systems, web search engines, text editors, and language instruction materials. 'Linguistik komputer merupakan studi ilmiah tentang bahasa dari perspektif komputer. Para ahli linguistik komputer tertarik untuk menyediakan berbagai model komputer untuk pemrosesan bahasa natural, baik dalam bentuk produksi maupun pemahaman, dan berbagai hal yang berkaitan dengan fenomena linguistik. Kerja para ahli linguistik komputer dipadukan ke dalam aplikasi-aplikasi praktis seperti sistem pengenalan terhadap ujaran, pemaduan ujaran, sistem perespon suara secara otomatis, mesin pencari di web, pengedit teks, dan berbagai materi intruksi bahasa' (Richards, 2002: 100).

Adapun Carl Vogel menyatakan,

Linguistics is the science of human language, whether spoken, written or signed; computational linguistics is the science of language with particular attention given to the processing complexity constraints dictated by the human cognitive architecture. Like most sciences, computational linguistics also has engineering applications. 'Linguistik adalah ilmu pengetahuan tentang bahasa manusia, baik bahasa ucap, bahasa tulis maupun bahasa dalam bentuk tanda, sedangkan linguistik komputer adalah ilmu pengetahuan tentang bahasa yang menitikberatkan perhatiannya secara khusus pada pemrosesan bahasa natural dengan berbagai kompleksitasnya,

yang dipandu oleh rekayasa program yang secara kognitif dibuat oleh manusia' (Vogel, 2006: 6).

Berdasarkan berbagai pernyataan di atas, maka dapat dipahami, bahwa Linguistik Komputer (*Computational Linguistics*) memiliki beberapa pengertian sebagai berikut.

1. Sebuah studi ilmiah mengenai bahasa dari sudut pandang komputer.
2. Sebuah kajian interdisipliner bahasa yang melibatkan antara bahasa natural dan komputer.
3. Sebuah studi yang menitikberatkan pada pemrosesan bahasa natural dengan berbagai fenomenanya secara otomatis oleh komputer
4. Studi tentang rekayasa program aplikasi komputer untuk pemrosesan bahasa natural dengan berbagai kompleksitasnya.

Dengan demikian, linguistik komputer dapat dipahami sebagai sebuah studi ilmiah interdisipliner mengenai bahasa dari sudut pandang komputer yang menitikberatkan pada rekayasa program aplikasi komputer, sehingga komputer mampu melakukan pemrosesan dan pengelolaan terhadap bahasa natural sebagaimana yang biasa dilakukan oleh manusia. Untuk ini, maka komputer perlu direkayasa sedemikian rupa sehingga secara formal komputer mampu melakukan pemrosesan dan pengelolaan terhadap bahasa natural sesuai dengan aturan-aturan yang ada pada linguistik. Berkaitan formalisasi bahasa dari sudut pandang komputer ini Stefano Crespi Reghizzi menyatakan,

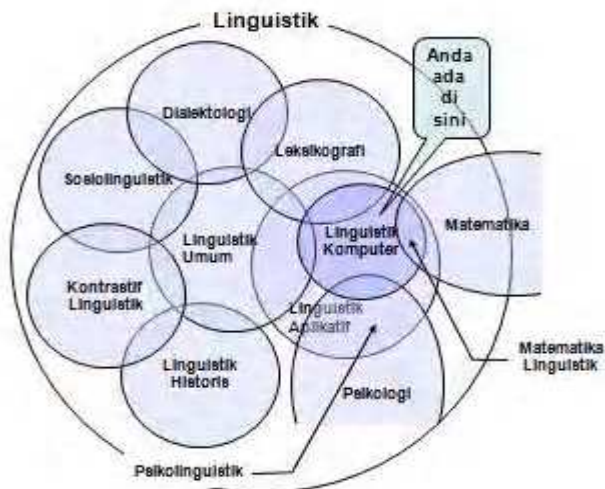
"For a language to be formalized (or formal), the form of sentences (or syntax) and their meaning (or semantics) must be precisely and algorithmically defined. In other words, it should be possible for a computer to check that sentences are grammatically correct, and to determine their meaning" 'bagi bahasa yang akan diformalisasi, bentuk kalimat dan maknanya harus secara tepat bisa didefinisikan atau dipahami secara logaritma. Dengan kata lain, definisi logaritmatik terhadap kalimat tersebut akan memungkinkan komputer bisa mengecek bahwa kalimat tersebut

secara gramatikal benar dan untuk menentukan maknanya' (Reghizzi, 2009: 5).

Dari uraian di atas, tampak bahwa linguistik komputer merupakan sebuah kajian linguistik interdisipliner, sebagaimana yang dikemukakan oleh Ruslan Mitkov, yang melibatkan teori linguistik umum, teori komputer, dan bahkan teori matematika dan logaritma. Hal ini sesuai pendapat Roland Hausser yang menyatakan,

"Computational Linguistics combines the methods of traditional grammar and theoretical linguistics with the method of effectively verifying explicit hypotheses by implementing formal grammars as efficient computer programs and testing them automatically on realistic"
'Linguistik komputer menggabungkan metode-metode tatabahasa tradisional dan linguistik teoretis dengan metode yang seefektif mungkin mampu memverifikasi berbagai hipotesis dengan jelas melalui cara pengimplementasian tatabahasa formal dalam bentuk program komputer yang efisien dan mampu mengujinya secara otomatis dan realistis' (Hausser, 2001: 18).

Jika demikian, dimanakah letak dan posisi linguistik komputer (*computational linguistics*) di antara kajian-kajian linguistik yang lain? Jawaban untuk hal ini dapat dilihat pada gambar berikut.



Gambar 2. Posisi Linguistik Komputer

Pada gambar di atas (Bolshakov, 2004: 25), tampak jelas bahwa disiplin linguistik komputer merupakan bagian dari bidang kajian linguistik aplikatif (*applied linguistics*). Di samping itu, linguistik komputer, sebagaimana tampak pada gambar di atas, juga melibatkan berbagai disiplin ilmu lain seperti linguistik umum, matematika, leksikografi, dan psikologi. Urgensi berbagai disiplin ilmu tersebut di dalam linguistik komputer dapat diuraikan sebagai berikut.

1. Linguistik umum penting bagi linguistik komputer karena tema-tema dalam bidang kajian linguistik umum seperti fonologi, morfologi, sintaksis, semantik, dan pragmatik juga menjadi tema kajian linguistik komputer. Hanya saja, di dalam linguistik komputer semua tema dalam bidang kajian linguistik umum tersebut dikemas dalam kerangka program aplikasi pemrosesan bahasa natural oleh komputer (*natural language processing*). Aturan-aturan yang digunakan oleh komputer di dalam pemrosesan bahasa natural juga didasarkan pada aturan-aturan yang ada di dalam linguistik umum.
2. Teknologi komputer dirancang dan bekerja atas dasar logika-logika matematika, mulai dari aljabar, aritmatika, logaritma maupun yang lainnya. Aturan-aturan yang ada dalam logika matematika inilah yang juga dijadikan sebagai dasar dan kerangka di dalam mengkonversi bahasa natural, sehingga komputer mampu mengenali dan memproses bahasa natural sebagaimana yang biasa dilakukan oleh manusia.
3. Dalam linguistik umum, leksikografi berkaitan dengan prinsip-prinsip pengkodean dan penyusunan entri-entri leksikal (*lexical entries*). Prinsip-prinsip ini sangat penting di dalam linguistik komputer untuk menyusun indeks data-data leksikal elektronik atau digital sesuai dengan aturan dan bentuk-bentuk morfologisnya sehingga data-data leksikal elektronik atau digital ini dapat dikenali secara otomatis oleh mesin komputer.

4. Psikologi adalah ilmu pengetahuan yang mempelajari perilaku manusia dalam hubungan dengan lingkungannya. Salah satu bidang kajian yang menghubungkan antara psikologi dan bahasa adalah psikolinguistik. Psikolinguistik adalah ilmu yang mempelajari faktor-faktor psikologis dan neurobiologis yang memungkinkan manusia mendapatkan, menggunakan, dan memahami bahasa. Oleh karenanya, psikolinguistik sangat erat kaitannya dengan psikologi kognitif, yang dalam penelitian modern hal ini menggunakan biologi, neurologi, ilmu kognitif, dan teori informasi untuk mempelajari cara otak memproses bahasa. Berbagai aspek psikologis yang berkaitan dengan perilaku manusia di dalam mendapatkan, menggunakan, dan memahami bahasa dan bagaimana cara otak memproses bahasa ini menjadi pertimbangan penting di dalam merancang program aplikasi komputer yang berkaitan dengan pemrosesan bahasa natural (*natural language processing*) oleh komputer. Dengan kata lain, melalui pertimbangan aspek psikologis ini komputer ingin dimodelkan seperti manusia yang mampu mengenali, merespon, mengucapkan, dan memproses bahasa sebagaimana manusia.

Berbagai aturan, prinsip, dan cara kerja yang ada di berbagai disiplin ilmu tersebut menjadi dasar pemodelan bagi linguistik komputer dalam memproses bahasa natural. Oleh karenanya, linguistik komputer (*computational linguistics*) merupakan model baru kajian linguistik dari sudut pandang komputer dan masuk dalam bidang kajian linguistik aplikatif (*applied linguistics*).

Disiplin kajian linguistik komputer bisa dikatakan sebagai model baru dalam kajian linguistik karena disiplin ilmu ini ada dan muncul setelah ada dan berkembangnya teknologi komputer. Di sisi lain, kajian linguistik komputer ini masih relatif baru masuk di negeri ini, belum begitu dikenal, dan belum begitu mendapatkan tempat di kalangan ahli linguistik. Linguistik

komputer masih menjadi bagian kecil dari kajian bidang komputer yang dilakukan oleh para ahli komputer negeri ini. Padahal mestinya para ahli linguistiklah yang paling punya kepentingan terhadap linguistik komputer karena sebenarnya, merekalah yang telah banyak bergelut dengan persoalan bahasa dengan berbagai aturan dan kompleksitasnya.

Hanya saja, tantangan yang dihadapi oleh para ahli linguistik untuk masuk dan bergelut dengan disiplin linguistik komputer ini tidaklah ringan, yakni tuntutan bagi mereka untuk bisa memahami teori dan cara kerja teknologi komputer yang memang cukup rumit, dan kemudian harus dikaitkan dengan teori linguistik. Oleh karenanya, disiplin kajian linguistik komputer ini masih menjadi sesuatu yang baru di kalangan para ahli linguistik negeri ini, apalagi secara historis munculnya kajian linguistik komputer ini juga relatif belum lama.

Secara historis, istilah *computational linguistics* 'linguistik komputer' baru muncul pada tahun 1974 sebagai padanan bagi istilah *machine translation* yang dimuat di sebuah jurnal *American Journal of Computational Linguistics*. Kemudian sejak tahun 1980 istilah *computational linguistics* mulai dikenal dan eksis sampai sekarang (Mitkov, 2004: xvii). Meski demikian, embrio bagi munculnya disiplin *computational linguistics* 'linguistik komputer' telah dimulai sejak tahun 1949 ketika Warren Weaver membuat pernyataan yang sangat terkenal dan mengejutkan banyak kalangan terkenal, bahwa *translation by machine might be possible* 'penerjemahan bahasa oleh mesin sangat dimungkinkan' (Mitkov, 2004: xvii). Kemudian, pada tahun 1955, setelah melalui berbagai penelitian yang dilakukannya, ia menegaskan pernyataannya tersebut dengan menyatakan bahwa *computers might be useful for the solution of world-wide translation problems* 'teknologi komputer sangat mungkin bisa digunakan untuk memecahkan problem-problem penerjemahan yang terjadi di dunia secara luas' (Barr, 1981: 226). Dengan demikian, istilah *computational linguistics* 'linguistik komputer' baru muncul setelah munculnya istilah

machine translation atau *mechanical translation* 'penerjemahan oleh mesin atau mekanik'.

Istilah '*computational linguistics*' (linguistik komputer) seringkali dipadankan dengan istilah '*natural language processing*' (pemrosesan terhadap bahasa natural) (Barr, 1981: 226) yang biasa disingkat menjadi NLP karena pada prinsipnya kedua istilah tersebut digunakan untuk menunjuk pada sebuah disiplin ilmu yang mengkaji persoalan pemrosesan bahasa natural oleh komputer (Richards, 2002: 352). Meski demikian, ada pendapat yang menyatakan, bahwa kedua disiplin tersebut adalah berbeda, sebagaimana yang dikemukakan oleh Martin Kay. Ia menyatakan,

"...he remarked on the distinction between Computational Linguistics and Natural Language Processing (NLP). Computational Linguistics is about using computers to investigate linguistic theory, while the NLP field concerns the engineering of text processing applications to solve particular tasks for practical reasons. Computational Linguistics is then a science, while NLP is the set of all its technological implications. Computational Linguistics is a branch of general linguistics, while NLP is more properly an engineering problem" 'Ia membuat pernyataan tentang perbedaan antara linguistik komputer dan pemrosesan bahasa natural, NLP. Linguistik komputer berbicara tentang penggunaan komputer untuk menginvestigasi teori linguistik, sedangkan bidang kajian NLP berkenaan dengan persoalan bagaimana merekayasa aplikasi-aplikasi pemrosesan teks untuk memecahkan tugas-tugas tertentu untuk alasan-alasan praktis. Dengan demikian, linguistik komputer adalah ilmu pengetahuan, sedangkan NLP adalah perangkat untuk segala hal yang muncul sebagai implikasi dari adanya teknologi. Linguistik komputer adalah cabang dari linguistik umum, sedangkan NLP lebih tepat disebut sebagai persoalan yang berkaitan dengan problem *engineering*' (Gliozzo, 2009: 1).

Demikian halnya dengan pendapat Daniel Jurafsky and James H. Martin yang menyatakan bahwa

computational linguistics (is) in linguistics, natural language processing in computer science, speech recognition in electrical engineering, computational psycholinguistics in psychology. 'Disiplin linguistik komputer masuk dalam bidang kajian linguistik, NLP ada pada ilmu komputer, pengenalan ujaran

ada pada permesinan elektronik, dan psikolinguistik komputer ada pada bidang kajian psikologi' (Jurafsky, 2000: 10).

Dengan demikian, dapat dipahami bahwa linguistik komputer (*computational linguistics*) adalah sebuah disiplin ilmu yang masuk dalam bidang kajian linguistik umum (*general linguistics*) dan secara khusus masuk dalam kajian linguistik aplikatif (*applied linguistics*) yang membahas persoalan bahasa dengan berbagai kompleksitasnya dari sudut pandang komputer. Tujuan dari disiplin ini adalah menjadikan bahasa bisa diucapkan, ditulis, disusun secara gramatikal, dimaknai, diterjemahkan, dan dipahami oleh komputer. Dengan kata lain, tujuan dari linguistik komputer adalah upaya untuk menjadikan komputer bisa mengelola dan memproses bahasa, baik dalam bentuk bahasa ucap maupun tulis, sebagaimana yang biasa dilakukan oleh manusia. Tugas utama dan orientasi ke depan dari linguistik komputer, menurut Roland Hausser, adalah mengembangkan mesin komputer kognitif, di tempat manusia bisa berbincang bebas dengannya, dengan tetap memperhatikan aturan-aturan yang ada pada bahasa natural. Jika demikian, maka linguistik komputer akan bisa mengembangkan teori fungsional tentang bahasa, menjadi sarana verifikasi yang objektif, dan memiliki bidang aplikasi yang luas (Hausser, 2001: vii).

D. RUANG LINGKUP KAJIAN LINGUISTIK KOMPUTER

Sebagaimana telah disebutkan, linguistik komputer (*computational linguistics*) adalah sebuah disiplin ilmu yang masuk dalam bidang kajian linguistik umum (*general linguistics*) dan secara khusus masuk dalam kajian linguistik aplikatif (*applied linguistics*). Sebagai sebuah disiplin ilmu, linguistik komputer juga memiliki ruang lingkup kajian yang bisa mencakup ke berbagai area ilmu pengetahuan. Menurut Andy Dong, bidang kajian linguistik komputer begitu luas dan utamanya adalah meneliti dan menciptakan desain model komputasi terhadap berbagai

fenomena yang ada di dalam linguistik. Bidang kajian linguistik komputer ini masih akan terus terbuka bagi model-model atau desain baru studi bahasa dengan mempertimbangkan interpretasi komputasional terhadap konsep-konsep teori linguistik secara kritis (Dong, 2009: 48). Sementara itu, Roland Hausser menyatakan bahwa

computational linguistics has its sources in several areas of science, each with its own goals, methods, and historical background. Thereby, it has remained unclear which components fit together and which do not. 'linguistik komputer memiliki banyak sumber di sejumlah wilayah ilmu pengetahuan yang masing-masing memiliki tujuan, metode, dan latar belakang sejarahnya. Dengan demikian, masih belum begitu jelas mana bidang yang masuk dan tidak masuk dalam kajian linguistik komputer' (Hausser, 2001: v).

Pendapat Roland Hausser tersebut mengindikasikan adanya perbedaan di kalangan para ahli linguistik komputer di dalam menetapkan mana bidang yang masuk dan tidak masuk dalam kajian linguistik komputer. Namun, jika kita mengacu pada namanya, yakni linguistik komputer yang menggabungkan antara linguistik dan komputer, maka paling tidak ruang lingkup kajian linguistik komputer bisa merambah pada dua area disiplin ilmu tersebut, yaitu disiplin linguistik dan komputer. Atas dasar ini, maka Richard Sproat menyatakan bahwa pada dasarnya ruang lingkup dan bidang kajian linguistik komputer adalah sama dengan ruang lingkup dan bidang kajian yang ada pada linguistik itu sendiri (Sproat, 2003: 608).

Hanya saja, tema-tema dalam bidang kajian linguistik tersebut dikaji dan dibahas dalam kerangka logika mesin komputer. Dengan demikian, ruang lingkup dan bidang kajian linguistik komputer memang cukup luas dan hal inilah yang barangkali menyebabkan munculnya perbedaan di antara para ahli linguistik komputer di dalam menetapkan pembagian secara rinci mengenai ruang lingkup atau bidang kajian linguistik komputer. Hal ini tampak pada berbagai buku linguistik komputer yang menyajikan berbagai kajian yang berbeda, baik menyangkut basis teori, metode, dan model aplikasinya.

Roland Hausser, di dalam bukunya *Foundations of Computational Linguistics: Human-Computer Communication in Natural Language* misalnya, menyebutkan bahwa bidang kajian linguistik komputer mencakup: 1) Fonologi (*sound/phonology*); 2) Morfologi (*word form/morphology*); 3) Leksikon (*lexicon*); 4) Sintaksis (*sentence/syntax*); 5) Semantik (*literal meaning/semantics*); dan 6) Pragmatik (*language use in communication/pragmatics*) (Hausser, 2001: 21). Pendapat yang hampir sama telah dikemukakan oleh Igor A. Bolshakov and Alexander Gelbukh. Keduanya menyatakan, bahwa bidang kajian linguistik komputer mencakup: 1) Perangkat matematika dan stuktur linguistik komputer (*matemathical tools and structures of computational linguistics*); 2) Fonologi (*phonology*); 3) Morfologi (*morphology*); 4) Sintaksis, baik pada level permukaan maupun dalam (*syntax of both surface and deep levels*); dan 5) Semantik (*semantics*) (Bolshakov, 2004: 8).

Apa yang dikemukakan oleh Roland Hausser, Igor A. Bolshakov, dan Alexander Gelbukh ini merupakan bagian dari bidang kajian fundamental dalam linguistik komputer yang dikaji dari sudut pandang logika matematika dan komputer sehingga menghasilkan kerangka dasar teoritis bagi linguistik komputer. Setiap bidang kajian tersebut, di dalam aplikasinya, bisa menghasilkan berbagai model aplikasi dalam konteks linguistik komputer seperti: 1). pengolahan teks atau pengeditan teks (*text preparation or text editing*) yang mencakup: penggabungan kata secara otomatis (*automatic hyphenation*), pengecekan ejaan (*spell checking*), pengecekan tatabahasa (*grammar checking*), pengecekan gaya bahasa (*style checking*) dan perenferensian terhadap kata tertentu (*referencing specific words*), penggabungan kata (*word combinations*) dan hubungan antar makna kata (*link between words*); 2) Pemerolehan informasi (*information retrieval*); 3) Penerjemahan otomatis antar bahasa (*automatic translation*); 4) Antar-muka bahasa natural (*natural language interfaces*); 5) Ekstraksi data faktual (*extraction of factual data*); 6) Pemahaman terhadap bahasa natural (*natural language understanding*); 7) Pengenalan karakter

huruf oleh lensa optik (*optical character recognition*); dan 8) Pengenalan ujaran (*speech recognition*) (Bolshakov, 2004: 53–54). Semua ini merupakan bidang kajian dalam linguistik komputer.

Di samping itu, gambaran mengenai ruang lingkup dan bidang kajian linguistik komputer yang lebih komprehensif dipaparkan oleh Ruslan Mitkov dalam bukunya yang berjudul *The Oxford Handbook of Computational Linguistics*. Di dalam bukunya tersebut ia mengorganisasikan bidang-bidang kajian linguistik komputer menjadi 3 (tiga) bagian, mulai dari teori dasar, proses, metode hingga bentuk-bentuk aplikasinya (Mitkov, 2004: v-x.). Tiga kelompok bidang kajian linguistik komputer tersebut adalah sebagai berikut.

1. Kelompok kajian fundamental dalam linguistik komputer, yang mencakup area kajian seperti: fonologi (*phonology*), morfologi (*morphology*), leksikografi (*lexicography*), sintaksis (*syntax*), semantik (*semantics*), wacana (*discourse*), pragmatik (*pragmatics*), dan dialog (*dialogue*). Semua hal tersebut dikaji dalam perspektif komputer, di samping juga dikaitkan dengan isu-isu mutakhir berkaitan dengan matematika linguistik (*mathematical linguistics*), tatabahasa formal (*formal grammar*), dan berbagai kompleksitasnya.
2. Kelompok kajian yang berkaitan dengan proses, metode, dan sumber dalam linguistik komputer, yang mencakup dasar-dasar proses dan tingkat kajian dalam linguistik komputer, tugas dan fungsi, metode, dan sumber kajian yang harus disiapkan dalam rangka pemrosesan bahasa natural secara otomatis oleh komputer. Model kajian linguistik komputer dalam kelompok ini adalah seperti: pembagian teks (*text segmentation*), pelabelan bagian-bagian ujaran (*part-of speech tagging*), pemenggalan kalimat (*parsing*), pengenalan ujaran (*speech recognition*), pemaduan pengucapan teks (*text to speech synthesis*), teknologi stata hingga (*finite-state technology*), resolusi anapora (*anaphora resolution*), pembelajaran melalui mesin (*machine learning*), dan bank data bahasa (*corpus linguistics*). Semua ini

merupakan kajian mengenai proses, metode, dan penyediaan sumber mengenai bahasa dalam perspektif komputer.

3. Kelompok kajian yang berkaitan dengan bentuk-bentuk aplikasi dalam linguistik komputer. Kelompok bidang kajian ini mencakup: penerjemahan oleh mesin (*machine translation*), pemerolehan informasi (*information retrieval*), penjawaban atas pertanyaan (*question answering*), peringkasan teks (*text summarization*), penggalian data teks (*text data mining*), penggunaan bahasa natural dalam sistem multimodal dan multimedia (*natural language in multimodal and multimedia systems*), dan pemrosesan bahasa natural multi-bahasa secara online (*multilingual online natural language processing*).

Dengan demikian, dapat dipahami, bahwa ruang lingkup dan bidang kajian linguistik komputer cukup luas, mulai dari teori linguistik fundamental dalam perspektif komputer, metode pemrosesan bahasa natural oleh komputer, sampai pada bentuk-bentuk atau model-model aplikasi otomatis dalam pemrosesan bahasa natural oleh komputer. Ruang lingkup dan bidang kajian linguistik komputer ini masih sangat mungkin berkembang sesuai dengan tuntutan perkembangan teknologi komunikasi dan komputer.

E. PENUTUP

Linguistik komputer dapat dipahami sebagai sebuah studi ilmiah interdisipliner mengenai bahasa dari sudut pandang komputer, yang menitikberatkan pada rekayasa program aplikasi komputer sehingga komputer mampu melakukan pemrosesan dan pengelolaan terhadap bahasa natural sebagaimana yang biasa dilakukan oleh manusia. Di samping itu, linguistik komputer merupakan sebuah kajian linguistik interdisipliner yang melibatkan teori linguistik umum, teori komputer, dan bahkan teori matematika, logaritma, dan psikologi.

Tujuan dari linguistik komputer adalah menjadikan komputer bisa mengelola dan memproses bahasa, baik dalam bentuk bahasa ucap maupun tulis, sebagaimana yang biasa dilakukan oleh manusia. Dengan kata lain, komputer ingin dimodelkan seperti manusia yang mampu mengenali, merespons, mengucapkan, dan memproses bahasa sebagaimana manusia. Ini adalah sesuatu yang baru. Pada dasarnya, sebelum adanya komputer, yang mengelola dan memproses bahasa, baik dalam bentuk bahasa lisan maupun tulis adalah manusia. Oleh karenanya, disiplin kajian linguistik komputer ini bisa dikatakan sebagai model baru dalam kajian linguistik aplikatif karena disiplin ilmu ini ada dan muncul setelah ada dan berkembangnya teknologi komputer.

Ruang lingkup kajian linguistik komputer cukup luas. Kajian linguistik komputer ini adalah mencakup bidang linguistik fundamental seperti: fonologi (*phonology*), morfologi (*morphology*), leksikografi (*lexicography*), sintaksis (*syntax*), semantik (*semantics*), wacana (*discourse*), pragmatik (*pragmatics*), dan dialog (*dialogue*), yang kesemuanya dikaji dalam pespektif komputer. Kemudian yang berkaitan dengan proses dan metode kajiannya adalah mencakup: pemenggalan kalimat (*parsing*), pengenalan ujaran (*speech recognition*), pemaduan pengucapan teks (*text to speech synthesis*), teknologi stata hingga (*finite-state technology*), resolusi anapora (*anaphora resolution*), pembelajaran melalui mesin (*machine learning*), dan bank data bahasa (*corpus linguistics*). Semua ini merupakan kajian mengenai proses, metode, dan penyediaan sumber mengenai bahasa dalam perspektif komputer.

Ada kelompok kajian yang berkaitan dengan bentuk-bentuk aplikasi dalam linguistik komputer, yang mencakup: penerjemahan oleh mesin (*machine translation*), pemerolehan informasi (*information retrieval*), penjawaban atas pertanyaan (*question answering*), peringkasan teks (*text summarization*), penggalian data teks (*text data mining*), penggunaan bahasa natural dalam sistem multimodal dan multimedia (*natural*

language in multimodal and multimedia systems), dan pemrosesan bahasa natural multi-bahasa secara online (*multilingual online natural language processing*).

DAFTAR PUSTAKA

- Barr, Avron and Edward A. Feigenbaum. 1981. *Handbook of Artificial Intelligence Volume 1*, Stanford, Los Astos: Heuristech Press, William Kaufmann.
- Bolshakov, Igor A. and Alexander Gelbukh. 2004. *Computational Linguistics: Models, Resources, Applications*. Mexico: Ciencia De La Computación.
- Bussmann, Hadumod. 1996. *Routledge Dictionary of Language and Linguistics*. New York: Routledge.
- Dong, Anda. 2009. *The Language of Design, Theory and Computation*. London: Springer-Verlag London Limited.
- Giozzo, Alfio and Carlo Strapparava. 2009. *Semantic Domains in Computational Linguistics*. Dordrecht; Heidelberg; London; New York: Springer.
- Hausser, Roland. 2001. *Foundations of Computational Linguistics: Human-Computer Communication in Natural Language*. Berlin; Heidelberg; New York; Barcelona; Hong Kong: Springer.
- Jurafsky, Daniel and James H. Martin. 2000. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. New Jersey: Prentice Hall, Englewood Cliffs.
- Mitkov, Ruslan. 2004. *The Oxford Handbook of Computational Linguistics*. Oxfor: Oxford University Press.
- Reghizzi, Stefano Crespi. 2009. *Formal Languages and Compilation*. London: Springer.

- Richards, Jack C. and Richard Schmidt. 2002. *Longman Dictionary of Language Teaching and Applied Linguistics*. Edinburgh: Pearson.
- Sproat, Richard, Christer Samuelsson, Jennifer Chu-Carroll, and Bob Carpenter. "Computational Linguistics" dalam Mark Aronoff and Janie Rees-Miller. 2003. *Handbooks of Linguistics*, USA: Blackwell Publishing.
- Vogel, Carl (Dir.). 2006. *Computer Science, Linguistics and a Language: Course Handbook 2006–2007*. Dublin: University of Dublin.
- Walter, Elizabeth. 2008. *Cambridge Advanced Learner's Dictionary*. Cambridge: Cambridge University Press, versi CD, edisi ke-3.