

KEAMBIGUAN DALAM BAHASA ALAMI

Muhammad Erwin Ashari Haryono

Laboratorium Pemrograman dan Informatika Teori,
Jurusan Teknik Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia, Yogyakarta
E-mail: meah@fti.uui.ac.id

ABSTRAKSI

Penelitian ini akan menganalisis macam-macam keambiguan yang terjadi dalam pengolahan bahasa alami, terutama pada bahasa Indonesia. Manusia sangat sulit untuk membedakan makna yang terdapat ambiguitas di dalam suatu bahasa, terlebih suatu mesin yang hanya dapat mengenal logika komputasi saja. Proses bagaimana suatu bahasa alami dapat diterapkan pada komputer dan dimengerti menjadi suatu perintah kerja yang baik menjadi fokus pada penelitian kami. Untuk mengerti dan menerapkan perintah kerja manusia dalam mesin komputer, tidak saja harus mengerti makna leksikal, gramatikal, dan sintaktik saja, tetapi makna semantik lah yang harus lebih ditekankan dalam proses *understanding* di dalamnya. "All About Nature Technology" mencoba menerapkan kemampuan bahasa alami untuk interaksi manusia dan komputer ke depannya, termasuk adaptasi kolaborasi antara dua disiplin ilmu yang berbeda, yaitu ilmu komputer (*computer science*) dan ilmu kebahasaan (*Linguistic*), sehingga penggabungan kedua bidang ini dapat meningkatkan kemampuan bahasa alami dalam komputer, terutama pada analisis ambiguitas.

Kata kunci: ambigu, bahasa alami, komputer, linguistik, semantik

1. PENDAHULUAN

Pada era informasi saat ini dimana kemajuan teknologi terutama teknologi ICT semakin baik dan terus berkembang, maka penulis mencoba untuk mengembangkan suatu teknologi yang mendukung pada proses alami (*All About Nature Technology*), yaitu kemampuan teknologi terutama dalam kaitannya dengan interaksi pada manusia yang alami. Konsep bahasa alami (*natural language*) telah banyak dikembangkan, tetapi kesulitan dalam bahasa alami adalah bagaimana membuat suatu bahasa yang multi bahasa, dan dapat mengerti makna-makna yang terkandung didalamnya, tanpa terjadi perbedaan makna karena adanya bentuk tulis, maksud, pembacaan, dan pemotongan kalimat yang bisa terjadi kapan saja. Bahasa alami, secara potensial memang banyak mengandung unsur-unsur yang ambigu, padahal di dalam sistem computer tidak dimungkinkan representasi unsur-unsur yang ambigu itu terjadi. Oleh karena itu, informasi mengenai unsur-unsur mana di dalam bahasa alami yang ambigu itu perlu dijabarkan secara rinci, untuk memudahkan analisis berikutnya, di dalam proses pengkomputerannya.

Keambiguan di dalam bahasa alami itu dapat dibedakan –paling tidak- atas tiga macam, yaitu (i) keambiguan kategori, (ii) keambiguan makna, dan (iii) keambiguan struktural. Keambiguan kategori adalah keambiguan menyangkut "kelas kata" (*parts of speech*), seperti yang terdapat pada kata *water* di dalam bahasa Inggris; periksa contoh [1] dan [2] berikut.

The water is boiling [nomina][1]

They water the plant twice a day [verba][2]

Keambiguan makna di dalam bahasa Indonesia, misalnya, terdapat pada kata *genting*. Makna kata itu pada frasa [3] berbeda dengan makna kata itu pada frasa [4].

Peristiwa 1 Maret sangat **genting**[3]

Atap **genting** di rumahku bocor[4]

Contoh keambiguan struktural dapat dilihat pada kedua kalimat berikut (dikutip dari Hutchins 1982:28). Pada [5] frasa berpreposisi *on the table* itu mengait pada verba *put*, sedangkan pada [6] frasa berpreposisi itu mengait pada *the book*. Pada contoh kalimat bahasa Indonesia, baru [7] dapat ditafsirkan sebagai kata – berupa adjektiva – yang mengait pada anggota "pendaftaran anggota baru, akan kami lakukan minggu depan". dan dapat pula ditafsirkan sebagai kata – berupa adverbial – yang mengait pada minggu depan "pendaftaran anggota, baru akan kami lakukan minggu depan". Biasanya makna ambigu pada struktural terjadi pada pemotongan pada kalimat tersebut.

The girl put the book **on the table**[5]

The girl saw the book **on the table**[6]

Pendaftaran anggota **baru** akan kami lakukan minggu depan[7]

Akan tetapi, keambiguan seperti yang terdapat pada contoh [3] dan [4], dan juga pada contoh-contoh berikut, dapat diatasi oleh komputer karena ada ciri pembeda yang dapat disimpulkan. Kata *genting* [3] adalah adjektiva, sedangkan kata *genting* [4] adalah nomina. Kata *memerah* [8]

adalah kata transitif, sedangkan kata **memerah** [9] adalah verba tak transitif. Kata mengandung[10] dirangkaikan dengan subjek insan, sedangkan kata mengandung [11] berkaitan dengan subjek bukan insan (atau benda).

Pekerjaannya **memerah** susu sapi [8]

Wajahnya **memerah** ketika ia dimarahi..... [9]

Ibu sedang **mengandung** enam bulan.....[10]

Garam **mengandung** potassium dan sodium[11]

Bentuk [8],[9],[10], dan [11] sebenarnya juga dalam macam keambiguan makna seperti bentuk [3] dan [4]. Namun, keambiguan yang berikut ini hingga kini masih merupakan persoalan bagi komputer, karena komputer baru dapat menjangkau analisis kalimat, belum sampai pada analisis konteks. Keambiguan seperti pada contoh [8-11] itu sudah dapat dipecahkan hanya dengan mengandalkan analisis kalimat. Adapun keambiguan pada contoh berikut ini, sebelum komputer dapat melakukan analisis konteks, masih merupakan persoalan bagi komputer.

Batu sekarang mahal[12]

Bisa ular kobra sangat beracun[13]

Adik berumur 5 tahun sudah **bisa** membaca.....[14]

Mereka melihat kera **dengan kata mata**[15]

Dia ingin menikah **karena** mengejar status[16]

Batu [12] dapat diartikan sebagai 'batu untuk membangun rumah' atau 'batu untuk menghidupkan transistor' (yaitu batu baterai). Bisa [13] dapat diartikan sebagai 'racun' (sehingga jawabannya bisa ular) dan dapat pula diartikan sebagai 'dapat' (sehingga jawabannya adalah bisa menyanyi, misalnya). Kalimat [15] dapat ditafsirkan sebagai 'melihat dengan kaca mata' atau 'kera yang memakai kaca mata'. Karena [16] dapat ditafsirkan sebagai preposisi (dan dapat digantikan dengan kata sekedar), dan dapat pula ditafsirkan sebagai konjungsi; pada tafsiran yang pertama belum tentu bahwa "si dia" itu tidak menikah, tetapi pada tafsiran yang kedua sudah pasti bahwa "si dia" itu tidak menikah. Untuk memecahkan keambiguan tersebut tidak dapat ditempuh analisis kalimat; harus digunakan analisis konteks.

Selain macam keambiguan tersebut Newmark (1987:218-220) mengidentifikasi lima macam keambiguan lain yang disebutnya:

- (1) keambiguan pragmatic
- (2) keambiguan kebudayaan (*cultural*)
- (3) keambiguan idiolektual

(4) keambiguan rujukan (referensial), dan

(5) keambiguan metaforik

Lima macam keambiguan di atas nampaknya termasuk kategori lain dari ketiga macam keambiguan di atas, yaitu kategori, makna, dan struktural. Tiga macam keambiguan NewMark itu termasuk 'sumber' keambiguan yaitu keambiguan pragmatic, kebudayaan, dan idiolektual; sedangkan macam yang keempat adalah fungsi unsure bahasa (=merujuk). Rasanya, macam kelima, yaitu metaforik, tidak perlu disebut; penjelasannya pun tidak ada. Lagipula, dapat kita tambahkan dua macam keambiguan lagi berdasarkan sumbernya, yaitu keambiguan fonologis, dan keambiguan ejaan.

Sebagai contoh kedua macam keambiguan terakhir di atas ialah (dari bahasa inggris) (1) dan (2).

[1] /rayt/ = *write, rite, right*

[2] *lead* = /led/ 'timah' dan /liy/ 'memimpin'

Dalam contoh pertama, ucapan ketiga kata itu sama, tetapi tulisannya berbeda dan maknanya pun berbeda, yaitu masing-masing 'menulis, upacara, kanan'. Keambiguan oleh persamaan ucapan itu disebut "homofoni" (*homophony*). Dalam contoh kedua, ejaan atau tulisan ini disebut "homografi" (*homography*). Dalam masalah pengolahan bahasa alami, hanya tipe kedua inilah, homografi yang menimbulkan masalah keambiguan.

2. PEMECAHAN KEAMBIGUAN

Keambiguan merupakan permasalahan bagi pendengar, penerjemah atau penganalisis suatu teks, sebab ia bisa mengaburkan atau menyesatkan pengertian. Memang hal ini kurang diperhatikan oleh para linguist, sebaliknya, keambiguan juga merupakan permasalahan bagi pemerhati bahasa: "bagaimana memerikan suatu bentuk bahasa (= kata atau kalimat) agar tidak berambigu?".

Dari sudut pandangan pengertian, pemecahan keambiguan secara umum ialah dengan memperhitungkan konteks kebahasaan dan situasi pragmatik suatu unsur bahasa. Sebagai contoh, perkataan kali dalam bahasa Indonesia mempunyai tiga makna : (1) sungai; (2) frekuensi; dan (3) suatu perlakuan matematik. Kalau kita melihat perkataan kali dalam kalimat bersama kata dalam, lebar, dsb., kita akan mengerti bahwa kali adalah 'sungai'. Jadi lingkungan kebahasaan perkataan kali itu menghindari atau memecahkan keambiguan itu (*disambiguate*). Demikian juga, kalau perkataan kali terlihat dengan bilangan-bilangan, frasa sama dengan, dsb., kita akan tahu bahwa perkataan kali bermakna 'suatu perlakuan matematik'.

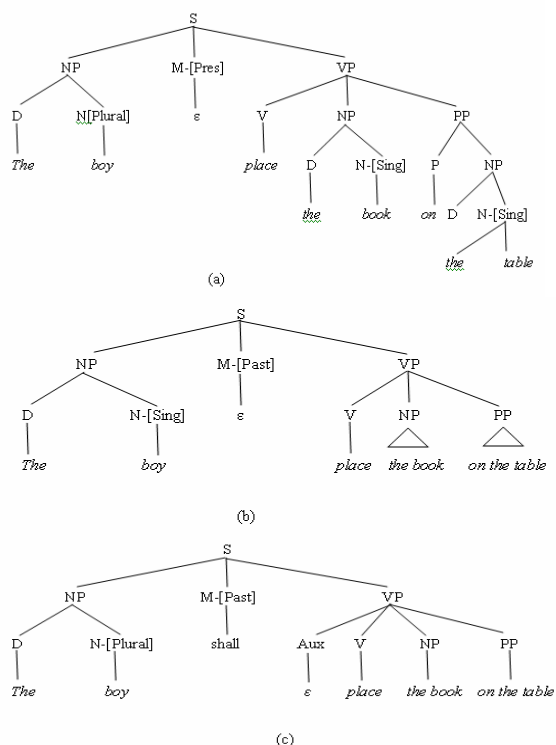
3. HIPOTESIS MASALAH

Dalam pemrosesan pengolahan bahasa Alami keambiguan secara structural dapat dijelaskan dan ditemukan dengan mengekstrak atau menguraikan kalimat tersebut menjadi lebih dari dua struktur pohon (sesuai dengan konsep hirarki Chomsky). Misalkan bila suatu kalimat pada contoh 17 dibawah ini kita uraikan menjadi diagram pohon (proses parsing) dapat diidentifikasi bahwa kalimat tersebut mengandung arti yang mendua (ambigu) gambar 2.

Dari susunan gambar yang diuraikan dari kalimat *The boy place the book on the table* dapat diturunkan menjadi tiga struktur pohon. Setiap struktur pohon yang diuraikan mengandung makna dan arti yang berbeda. Masalah ini barulah menguraikan arti keambiguan secara struktural.

Keambiguan yang lain seperti yang telah diuraikan di atas (sub bab 2) pada prinsipnya bila diturunkan untuk mengetahui makna semantiknya sama dengan penurunan pada gambar 2 tersebut.

The boy place the book on the table..... (17)



Gambar 2. Diagram pohon penurunan ambigu dari susunan pemecahan kalimat.

DAFTAR PUSTAKA

- AHO, A.V. Indexed Grammars: An Extension of Context-Free Grammars. *Journal of the ACM*. Vol. 15 4, pp. 647-671. 1968.
- AHO, A.V. and J.D. Ullman. *The Theory of Parsing, Translation, and Compiling*, Vol 1: Parsing, Prentice-Hall, Engelwood Cliffs, N.J. 1972.
- Barr, A.,P. R. Kohen, and E. A. Feigenbaum eds. *The Handbook of Artificial Intelligence*, Vol IV, Addison-Wesley, Reading, Mass (1989).
- Krulle, Gilbert. *Computer Processing of Natural Language*. Prentice-Hall International Editions. 1991.
- Kaswanti Purwo, Bambang. *Linguistik dan Teknologi Komputer, Pemrosesan Bahasa Alami*. Penerbit ITB Bandung. 1989.
- Munir, Rinaldi. *Matematika Informatika. Diktat kuliah teknik Informatika Institut Teknologi Bandung*. 1999.

