

PENGURAIANKATA PADA KALIMAT BAHASA KOMERING RASUAN BERDASARKAN KAIDAH BAHASA INDONESIA MENGGUNAKAN TEORI AUTOMATA

Alvi Syahrini Utami", Julianisya Tri Parasta
alvisyahrini22@yahoo.com

ABSTRACT

Untuk mengerti suatu bahasa, bahasa tersebut dapat dipelajari pada level-level pembelajaran yang berbeda, salah satunya adalah pada level sintaksis. Level sintaksis dalam analisa linguistik difokuskan dengan bagaimana kumpulan kata-kata saling menyusun membentuk suatu konstruksi. Teori automata yang digunakan dalam penelitian ini adalah Shift-Reduce Parsing dan pengaplikasian pohon penurunan sebagai hasil parsing dari suatu kalimat. Jenis kalimat yang diolah berupa kalimat yang hanya mengacu pada sintaksis secara gramatikal dan hanya sebatas satu kalimat positif formal sederhana dan keluarannya berupa bentuk penurunan kalimat menggunakan pohon penunman. Jenis bahasa yang diproses dalam penelitian ini adalah bahasa daerah Komerling Rasuan. Jenis bahasa daerah yang digunakan dapat beragam selama masih memiliki sintaksis yang sama dengan kaidah Bahasa Indonesia, meskipun demikian, masih diperlukan beberapa pengembangan lebih lanjut untuk mengolah bentuk kalimat yang lebih kompleks.

Keywords: bahasa Komerling Rasuan, bahasa Indonesia, Shift-reduce parsing, pohon parsing, automata

I. PENDAHULUAN

Bahasa Komerling adalah bahasa keseharian yang dipakai oleh masyarakat yang berdomisili di sepanjang sungai Komerling, Sumatera Selatan. Tiap desa dalam mempergunakan bahasanya memang sebagian memiliki sedikit perbedaan katanya, termasuk juga di desa Rasuan, tetapi tetap pada satu bahasa induk yaitu bahasa Komerling [4]. Di Sumatera Selatan sendiri terdapat banyak sekali bahasa daerah yang berlainan kosakatanya.

Kurangnya pembelajaran mengenai bahasa daerah Komerling maupun bahasa daerah lainnya akan mengaburkan kebudayaan asli Indonesia yang semakin lama semakin digeser oleh kebudayaan barat yang modern. Karena itu ada perlunya mempelajari bahasa daerah guna pelestarian kebudayaan. Selain itu, mutasi para pekerja baik dari departemen-pemerintahan maupun tenaga kerja lain yang ditempatkan di desa Komerling Rasuan tentunya membutuhkan pembelajaran khusus mengenai bahasa daerah ini guna dapat membangun komunikasi dengan masyarakat asli Komerling Rasuan.

Untuk mengerti suatu bahasa, bahasa tersebut dapat dipelajari pada level-level pembelajaran yang berbeda, salah satunya adalah pada level

sintaksis. Level sintaksis dalam analisa linguistik difokuskan dengan bagaimana kumpulan kata-kata saling menyusun membentuk suatu konstruksi [1]. Setiap bahasa didasarkan pada sebuah kosa kata. Elemen-elemennya biasanya disebut dengan kata; dalam bahasa formal, mereka disebut dengan simbol (dasar).

Berdasarkan pembelajaran berdasarkan sintaksisnya, sejumlah deret kata akan disebut benar jika kalimat yang terbentuk tersusun dengan baik. Suatu kalimat dapat dikatakan kalimat jika memiliki unsur minimal subjek dan predikat. Adanya susunan subjek, predikat, objek dan lainnya akan membentuk suatu standar struktur kalimat itu sendiri. Klen O dan Manning menggunakan Tree-bank Universitas Pennsylvania sebagai tag standar simbolisasi sintaks bahasa natural yang mereka teliti [7]. Bahasa komering Rasuan merupakan bagian dari bahasa Indonesia, maka bentuk kalimat formal dari bahasa komering itu sendiri berdasarkan pada struktur bahasa Indonesia, sedangkan kalimat dalam bahasa Indonesia sedikitnya harus memiliki unsur subjek dan predikat.

Oalam ilmu komputer dikenal suatu ilmu yang disebut dengan otomata atau automata. Teori automata dapat digunakan sebagai salah satu solusi dalam masalah pengolahan bahasa natural,

yaitu dengan menggunakan pohon penurunan untuk menghasilkan parsing dari suatu kalimat.

Banyak penelitian Bahasa natural dan bergagai macam buku linguistik yang menggunakan pohon parsing. Ratnapharki [10], Utami [13], dan beberapa penelitian bahasa natural lainnya kerap kali menggunakan pohon parsing sebagai pemodelan struktur kalimat bahasa natural. Sedangkan sebelum dapat membentuk pohon penurunan, Shift-Reduce Parser digunakan oleh Ratnapharki [10], Nivre [9] dan Sagae K, Livre A [13] sebagai pengolah leksikal input suatu kelompok kata.

II. STRUKTUR KALIMAT BAHASA INDONESIA

A. Sintaksis

Sintaksis mencakup dua hal, yaitu studi tentang bagaimana kata-kata membentuk kalimat dan pokok-pokok aturan yang mengatur pembentukan kalimat. Sintaksis mencakup hubungan antar kata, frase, ataupun klausa dalam kalimat serta aturan-aturan yang terlibat [3]. Sintaksis membicarakan penataan dan pengaturan kumpulan kata kedalam satuan yang lebih besar, yang disebut satuan-satuan sintaksis yaitu: kata, frase, klausa, kalimat, dan wacana. [2][3]. Aturan-aturan dalam sintaksis suatu bahasa dapat digunakan ke dalam bentuk algoritma parsing [1].

B. Kelas Kata dalam Bahasa Indonesia

Banyak penelitian linguistik mengenai kelas kata dalam bahasa Melayu dan Indonesia, sehingga penulis menggunakan salah satu hasil penelitian kelas kata bahasa Melayu dan Indonesia yang menyimpulkan sebagian besar hasil penelitian linguistik sebelumnya mengenai kelas kata. Berikut ini adalah kelas kata dalam bahasa Indonesia [6];

1. Verba (V)

Sebuah kata dapat dikatakan suatu Verba dari perilakunya dalam suatu frase, yaitu kemungkinan untuk didampingi partikel tidak dalam konstruksinya dan tidak dapat didampingi dengan partikel di, ke, dari, sangat, lebih, atau agak.

2. Ajektiva (A)

Ajektiva adalah kategori yang ditandai oleh beberapa kemungkinan:

- a. Bergabung dengan partikel tidak,
- b. Mendampingi nomina,
- c. Didampingi partikel seperti lebih, sangat, agak,
- d. Mempunyai ciri-ciri morfologis tertentu
- e. Dapat dibentuk menjadi nomina dengan konfiks ke-an.

3. Nomina (N)

Nomina adalah kategori yang secara sintaksis tidak mempunyai potensi untuk bergabung dengan partikel tidak dan mempunyai potensi untuk didahului oleh partikel dari.

4. Pronomina (Pron)

Pronomina adalah kategori yang berfungsi untuk menggantikan nomina. Pronomina tidak bisa memiliki afiks, tetapi dalam beberapa di antaranya bisa direduplikasikan. Kata pronomial dapat dijadikan frase pronomial misalnya aku ini, kamu sekalian, mereka semua, dan lain-lain.

5. Numeralia (Num)

Dalam konstruksi sintaksis, numeralia dapat mendampingi nomina. Numeralia mempunyai potensi untuk mendampingi numeralia lain dan tidak dapat bergabung dengan tidak atau sangat.

6. Adverbia (Adv)

Adverbia adalah kategori yang mendampingi ajektiva, numeralia, proposisi dalam konstruksi sintaksis.

7. Interogativa (Intg)

Interogativa adalah kategori dalam interogatif yang berfungsi menggantikan sesuatu yang ingin diketahui oleh pembicara atau mengukuhkan apa yang telah diketahui pembicara.

8. Demonstrativa (Oem)

Demonstrativa adalah kategori yang berfungsi untuk menunjukkan sesuatu di dalam maupun di luar wacana.

9. Artikula (Art)

Artikula dalam bahasa Indonesia adalah kategori yang mendampingi nomina dasar, pronomina. Artikula berupa partikel, sehingga tidak dapat berafiksasi.

IO. Preposisi (Prep)

Preposisi adalah kategori yang terletak depan kategori lain (terutama nomina) sehingga terbentuk frase.

II. Konjungsi (Konj)

Konjungsi adalah kategori yang berfungsi untuk menghubungkan satuan yang lain dalam konstruksi hipotaktis, dan selalu menghubungkan dua satuan lain atau lebih dalam konstruksinya.

12. Interjeksi (Intj)

Interjeksi adalah kategori yang bertugas mengungkapkan perasaan pembicara, dan secara sintaktis tidak berhubungan dengan kata-kata lain dalam ujaran. Interjeksi selalu mendahului ujaran sebagai teriakan yang lepas atau berdiri sendiri,

C. Pohon Penurunan (Parse Tree)

Suatu pohon (tree) adalah suatu graph terhubung tidak sirkuler yang memiliki satu simpul (node)/vertex disebut akar (root) dan dari situ memiliki lintasan ke setiap simpul [5]. Pohon penurunan merupakan pohon yang menunjukkan esensi suatu asal mula [5]. Pohon penurunan (derivation tree/parse tree) berguna untuk memperoleh suatu string (untai) dengan cara menurunkan simbol-simbol variabel menjadi simbol-simbol terminal. Setiap simbol variabel akan diturunkan menjadi terminal sampai tidak ada yang belum tergantikan.

Sebagai contoh, kita perhatikan kalimat "cat sleeps". Kata "cat" adalah subyek dan "sleeps" adalah predikat. Kalimat ini kepunyaan suatu bahasa yang mungkin di definisikan dengan sintaksis berikut ini.

< kalimat > ::= < subyek > < predikat >

< subyek > ::= cat | dog

< predikat > ::= sleeps | eat

Arti dari tiga baris di atas adalah sebagai berikut:

1. Kalimat tersusun atas subyek di ikuti dengan predikat.
2. Subyek dapat berupa kata "cat" atau kata "dog".
3. Predikat dapat berupa kata "sleeps" atau "eat".

Dengan demikian sebuah kalimat dapat diturunkan dari simbol awal (starting symbol) <kalimat> dengan menerapkan aturan penggantian secara berulang. Bentuk utama <kalimat>, <subyek> dan <predikat> disebut dengan simbol non-terminal, kata cat, dog, sleeps dan eat disebut dengan simbol terminal dan aturannya disebut aturan produksi. Simbol ::= dan

disebut dengan simbol mata. Jika, untuk ringkasnya, kita menggunakan sebuah huruf kapital untuk menandai simbol non terminal dan huruf lower case untuk menandai simbol terminal, maka contoh di atas dapat di tulis sebagai Contoh I.

S ::= AB

A ::= xly

B ::= zlw

Bahasa yang di definisikan dengan sintaksis di atas terdiri atas empat kalimat: xz, yz, xw, yw. Aplikasinya pada kalimat bahasa Inggris yang pertama harus diketahui adalah makna dari suatu kata dalam bahasa Inggris, misalkan kata "cat", apakah termasuk kata benda, kata kerja atau kata keterangan? Untuk itu disusunlah deskripsi kata ataupun frasa seperti yang digunakan oleh Marcus (1993) dan Klein (2001) yang disebut dengan Pennsylvania Tree Bank [7][8]. Dicontohkan suatu kalimat: "cat sleeps on the floor", maka parsing yang didapat berdasarkan tabel PenTreeBank menjadi:

S ::= <subject><predicate>

<subject phrase> ::= NP

<predicate phrase> ::= VBZ PP

NP ::= NN | DTNN

PP ::= INNP

NN ::= cat

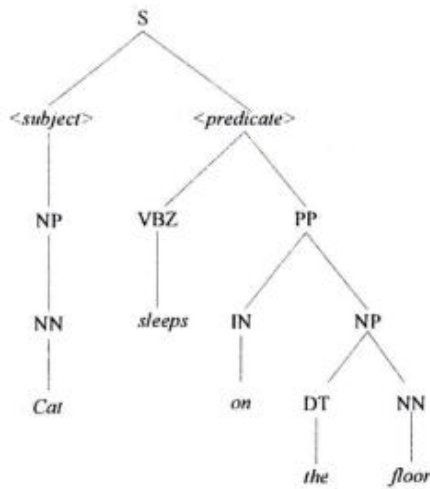
VBZ ::= sleeps

IN ::= on DT ::=

the

NN ::= floor

Lalu penggambaran pohon parsing yang terbentuk dapat dilihat pada gambar I. Jadi dari hasil penurunan menggunakan parse tree didapatkan S(<subject>(NP(NN cat)<predicate><VBZ sleeps>(PP<IN on>(NP(DT = the)(NN = floor))))). Parsing dapat dilakukan dengan berbagai macam cara: Bottom-up Parsing, Top-down Parsing atau pendekatan dinamik seperti chart Parsing atau algoritma CYK.



Gambar I. Parsing Tree untuk Kalimat "cat sleeps on the floor"

D. Shift Reduce Parsing

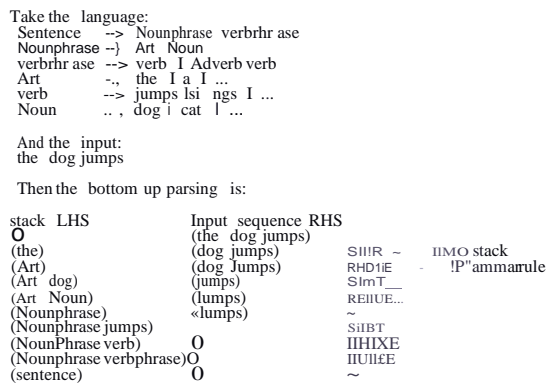
Shift-Reduce Parsing (SR Parsing) merupakan teknik parsing yang termasuk kategori bottom-up parsing yang paling umum dipakai dan paling unggul. SR parsing digunakan sebagai peruntun token dan membentuk barisan produksi untuk membangun pohon parse (parse tree). SR Parsing menggunakan tumpukan (stack) guna menjaga urutan masing-masing token.

Tiap langkah dalam SR parsing terdapat dua langkah dasar [11][12]: operasi shift, yang merupakan operasi penambahan kata dari kalimat masukan pada elemen teratas stack yang sering disebut sebagai top. Dan operasi reduce yang merupakan operasi pemindahan elemen top pada stack dan menggantinya dengan elemen baru yang berupa grammar rule sesuai informasi elemen yang digantikan tersebut. Langkah dasar dari Shift Reduce Parser dapat dijabarkan sebagai berikut:

1. Diawali dengan kalimat yang akan diurai per token kedalam bentuk stack
2. Hingga stack kosong, lakukan:
 - a. I. Scan melalui input sampai dikenali sesuatu yang sesuai dengan RHS dari salah satu aturan produksi (ini disebut handle)
 - b. 2. Terapkan aturan produksi secara terbalik, yaitu, menggantikan RHS dari aturan yang muncul dalam bentuk sentensial dengan

LHS aturan (tindakan yang dikenal sebagai reduce)

Pada langkah 2.1 "shift" simbol input ke satu sisi (LHS); maka parser yang beroperasi dengan berulang kali dengan menerapkan langkah 2,1 dan 2,2 dikenal sebagai parser shift-reduce. Operasi SR parsing ini menggunakan dua stack, RHS (Right Handle Stack) yang menyimpan masukan kalimat yang jika dipecah menjadi urutan token dan disimpan dalam bentuk stack, dan LHS (Left Handle Stack) yang menampung hasil operasi (shift dan reduce) dari token pada RHS.



Gambar 2. Contoh ilustrasi Shift parsing pada kalimat berbahasa Inggris

III. SHIFT-REDUCE PARSING SEBAGAI PEMERIKSA SINTAKS

Pemeriksaan sintaks pada pola kalimat berbentuk dari sekumpulan kata memegang peranan penting pada perangkat lunak ini. Jika sekumpulan kata yang dimasukkan dapat diterima sebagai suatu kalimat secara sintaksnya, maka hasil akhir dari perangkat lunak yang akan dibangun akan didapat, yaitu berupa visualisasi pohon penurunan yang menggambarkan struktur kalimat berdasarkan sintaksnya. Sebaliknya, jika sekumpulan kata tersebut menurut sintaksnya tidak diterima sebagai suatu kalimat, maka hasil akhir berupa visualisasi pohon penurunan tidak dapat ditampilkan.

Secara umum, dalam algoritma Shift-Reduce Parsing memiliki 4 aksi sebagai berikut:

- a. Shift - menambah satu elemen (token yang didapat dari masukan) pada stack. Aksi Shift hanya berupa pemindahan (shifting) item pertama (bagian teratas dalam tumpukan kata, dalam hal ini per-item> berupa satu buah kata) dari RHS (Right Handle Stack) ke LHS (Left Handle Stack)
- b. Reduce - menghapus elemen teratas pada LHS dan menggantinya dengan menambah satu elemen nonterminal yang sesuai.
- c. Accept - mengenali kalimat jika hanya terdapat simbol root dan masukan kosong
- d. Error - terjadi jika tiga poin di atas tidak mungkin dilakukan lagi, yang mengartikan bahwa masukan bU kan berupa kalimat

Jika LHS kosong, maka hanya aksi Shift yang dapat dilakukan. Jika RHS kosong, hanya aksi Reduce yang dilakukan. Jika RHS dan LHS tidak kosong, maka terdapat kemungkinan aksi yang

terjadi adalah keduanya, dan pemroses harus memberikan satu kondisi untuk menentukan aksi yang dilakukan. Jika Aksi yang dilakukan adalah Reduce, maka ditentukan suatu non-terminal (dalam hal ini rule) apa yang harus ditambah ke dalam LHS menggantikan item teratas dari LHS itu sendiri. Jika aksi yang dilakukan Shift, maka akan terbentuk suatu node terminal baru sebagai leaf dari pohon parsing dan akan terbentuk subtree baru. Dalam perancangan perangkat lunak ini, masukan yang berupa kumpulan kata dalam bahasa Komerling Rasuan akan diartikan terlebih dahulu ke dalam bahasa Indonesia dan kemudian aturan sintaks yang digunakan adalah sintaks bahasa Indonesia.

Pengaplikasian Algoritma Shift-Reduce Parser Ditentukan rule sintaks sebagai berikut:

- E->S PIS POIS P KISPOK
- S->FNIN
- P->FYIYIFAdj IAdjIFNum
- O->FNIN
- K->FPrep
- FN->NIFN NIFN DemlFN YIFN AdjAdv FNJFN Num
- FY->YIAdv FYJFY AdvlFY NIFY Adj
- FAdj-> AdjlFAdj AdjlFAdj NIFAdj AdvlAdv FadjlFadj V
- FPrep->Prep FNIPrep NIFPrep FN
- FNum->Num NINum Nurnll-Num NIFNum
- Num

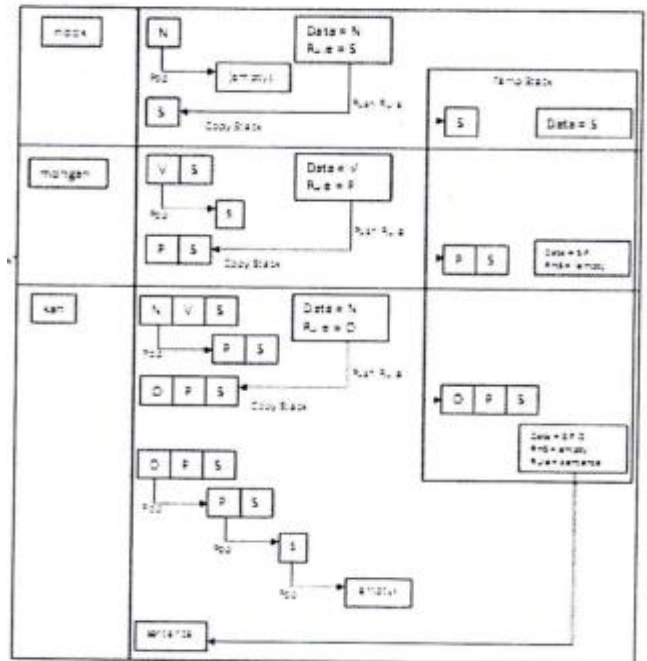
Dimisalkan masukan beberapa kumpulan kata: *indok mongan kan* (ibu makan nasi), Dengan berdasarkan pada data kamus, didapat hasil sebagai berikut:
 indok;ibu;N
 mongan;makan; Y
 kan;nasi;N

Skema dasar *shift-reduce* pada kalimat '*indok mongan kan*' dapat dilihat pada gambar 3.

LHS (J)	RHS(Input)	Aksi
[indokJ	[indok mongan kan]	shift kat a ke LHS
[N]	[mongan kan]	Reduce dengan rule yang ses
[N mongan]	[mongan kan]	shift .
[N Vj	[kan]	Reduce .
[s V	[kan]	Reduce .
[s v kan]	[kan]	shift .
[s V NJ		Reduce .
[s P NJ		Reduce .
[s P oj		Reduce .
[sentenceJ		Accept

Gambar 3. Skema dasar Shift-Reduce Parsing pada kalimat "indok mongan kan "

Berikut digambarkan skema aksi *reduce* pada implementasi Shift Reduce Parser dengan penggunaan *stack* temporer serta modifikasi pada algoritmanya.



Gambar 4. Proses Reduce (posisi LHS) dalam Proses parsing kalimat 'indok:mongan kan '

IV. KESIMPULAN

Kesimpulan yang dapat diperoleh dari penelitian ini adalah sebagai berikut :

1. Dari sampel yang ada, persentase hasil pemeriksaan sintaks menggunakan Shift-Reduce Parsing pada perangkat lunak yang dibangun sebesar 84%. Dan perangkat lunak dapat menampilkan hasil visualisasi penggambaran pohon penurunan untuk 84% sampel yang struktur kalimatnya diterima.
2. Perancangan dan implementasi aplikasi penguraian kata pada suatu kalimat dalam bahasa Komerling Rasuan telah berhasil dilakukan, Kata-kata dalam bahasa Komerling Rasuan dapat ditranslasikan berdasarkan kamus data yang ada..
3. Data atau kamus bahasa masih terbatas pada kata-kata yang terdapat dalam kamus referensi. Input kata ganti nama orang masih belum dapat dikenali sebagai suatu bentuk nomina.
4. Jika input terdapat suatu kesalahan seperti kesalahan eja atau kata tidak terdapat dalam kamus, maka perangkat lunak akan berhenti melakukan pemrosesan kalimat. Begitu pula halnya jika input secara sintaks-nya tidak dapat diterima sebagai suatu bentuk kalimat.

- [7]. Klein D, Manning CD. 2001. Parsing with Treebank Grammars: Empirical Bounds, Theoretical Models, and the Structure of the Penn Treebank. Proceedings of the 39th Annual Meeting on Association for Computational Linguistics; Toulouse, 6 Jul 2001. Morristown: Association for Computational Linguistics. Hlm 338-345.

DAFTAR PUSTAKA

- [1]. Akrekar R, Joshi M. 2008. Natural Language Interface Using Shallow Parsing. International Journal of Computer Science and Applications; 5(3):70-90.
- [2]. Carnie, Andrew. 2002. Syntax: A Generative Introduction, parts I and 2. Oxford: Blackwell Publishers.
- [3]. Chaer A. 2009. Sintaksis Bahasa Indonesia (Pendekatan Proses). Jakarta: Rineka Cipta.
- [4]. Hanie A. 2008. Kamus Bahasa Daerah Komerling Rasuan: OKU Timur Sumatera Selatan. Jakarta: Cikoro Tri Rasuandar.
- [5]. Hopcroft JE, Motwani R, Ullman JD. 2001. Introduction to Automata Theory, Languages, and Computation. New York: Addison Wesley.
- [6]. Kardilaksana H. 2007. Kelas Kata dalam Bahasa Indonesia. Jakarta: Gramedia Pustaka Utama.