

Perancangan Penganalisis Struktur Kalimat Bahasa Indonesia dengan Menggunakan Constraint-Based Formalism

I Made Dwi Jendra Sulastra
Politeknik Negeri Bali
e-mail: dwijendrasulastra@yahoo.com

Abstrak

Tujuan utama penelitian ini adalah menganalisa struktur kalimat bahasa Indonesia dengan menggunakan metode *constraint-based formalisms*. *Constraint based-formalisms* adalah suatu metode yang memungkinkan aturan-aturan sintaks untuk diberikan suatu batasan-batasan sehingga dapat meminimalkan keambiguan dari aturan sintaks yang dibuat. Struktur sintaks yang dibangun pada penelitian ini mengacu pada aturan-aturan sintaks yang terdapat di dalam buku tata bahasa baku bahasa Indonesia edisi ketiga. Pada penelitian ini, penulis memfokuskan pada pembuatan aturan sintaks untuk frasa nominal. Pembuatan pengurai sintaks kalimat bahasa Indonesia ini menggunakan alat bantu PC-PATR. PC-PATR ini merupakan implementasi dari *constraint-based formalisms* pada personal computer. Domain kalimat-kalimat yang akan digunakan pada penelitian ini adalah kalimat-kalimat yang terdapat pada media massa elektronik. Hasil uji coba yang dilakukan terhadap kumpulan kalimat pada media massa elektronik cukup baik. Sebanyak 93,68 % dari 380 kalimat masukan dapat diuraikan dengan benar oleh pengurai, dan berhasil menolak 3,16 % kalimat input yang salah struktur kalimatnya; sedangkan 2,11 % dari 380 kalimat masukan ini tidak berhasil diuraikan walaupun struktur kalimat masukan sesuai dengan tata bahasa baku dan juga terdapat sebanyak 1,05 % kalimat masukan yang hasil penguraiannya salah.

Kata Kunci: struktur kalimat bahasa indonesia, *constraint-based formalisms*, frasa nominal, *pc-patr*, media massa elektronik

Abstract

The main purpose of this study is to analyze the structure of Indonesian sentences by using *constraint-based formalisms*. *Constraint - based formalisms* is a method that allows the syntax rules for a given boundaries so as to minimize ambiguity of syntax rules are made. Syntax structure built in this study refers to the syntax rules contained in the book of Raw Indonesian grammar third edition. In this study, the authors focused on the manufacture of syntax rules for noun phrases. Making Indonesian sentence syntax parser uses PC - PATR tools. PC - PATR is an implementation of a *constraint-based formalisms* in the personal computer. Domain sentences that will be used in this study are the sentences contained in the electronic mass media. Results of experiments performed on a collection of sentences on electronic mass media quite well. A total of 93.68 % of the 380 input sentence can be parsed correctly by the parser, and managed to resist 3.16% input the wrong sentence grammar; while 2.11% of the 380 sentences described the input does not work even though the structure of the input sentence in accordance with the procedures standard language and also contained 1.05 % as a result of breakdown in the input sentence wrong.

Keywords: the structure of Indonesian sentences, *constraint-based formalisms*, noun phrases, *pc-patr*, electronic mass media

1. Pendahuluan

Pemrosesan bahasa alami merupakan suatu topik menarik yang banyak diperbincangkan oleh para ahli dari berbagai bidang ilmu seperti misalnya: ahli komputer, guru, psikolog, ahli bahasa,

dan masih banyak lagi ahli-ahli lainnya. Sebagian besar dari ahli ini memiliki suatu pertanyaan yang hampir sama, yaitu seberapa jauh teknologi komputer yang ada sekarang ini mampu menerjemahkan atau menginterpretasikan bahasa manusia atau bahasa alami ini ke dalam bahasa komputer [1].

Ketertarikan mereka pada pemrosesan bahasa alami ini tidak lain disebabkan karena pemrosesan bahasa alami ini mempunyai banyak *fitur* yang dapat dikembangkan untuk mempermudah aktivitas kerja sehari-hari mereka [1]. Banyak tugas penting yang dapat dilakukan jika komputer dapat diprogram untuk membaca isi dari suatu pesan yang dikirim dan disimpan dalam bentuk bahasa alami. Sebagai contoh, seseorang dapat mengetik dengan cepat jika mereka dapat menggunakan bahasa alami untuk memberikan perintah atau instruksi kepada komputer. Dengan kata lain, pemrosesan bahasa alami ini dapat merubah cara penggunaan komputer dan cara seseorang dalam berinteraksi dengan komputer [2] [3] [4] [5] [6] [11].

Akan tetapi semua hal di atas, tidak lah mudah untuk dicapai. Pemrosesan bahasa alami merupakan suatu hal yang sangat kompleks. Sangat sulit untuk merancang dan membangun suatu aplikasi komputer yang dapat menganalisa dan memahami bahasa yang biasa dipakai oleh manusia dalam kehidupan sehari-hari. Pemrosesan bahasa alami memerlukan pengetahuan tentang kamus kata dan aturan tata bahasa [6] [7] [8] [10]. Kamus kata diperlukan untuk memeriksa apakah kata-kata yang dimasukkan merupakan kata-kata yang *valid*. Aturan Tata bahasa diperlukan untuk memeriksa apakah kalimat yang dimasukkan sesuai dengan kaidah tata bahasa dari bahasa yang digunakan. Ironisnya, aturan tata bahasa yang ada sampai saat ini, masih jauh dari sempurna.

Disamping masalah ketidakbakuan aturan tata bahasa di atas, pemrosesan bahasa alami juga menemui banyak hambatan dalam masalah keambiguan kelas kata atau kamus kata. Sebagai contoh, dalam kalimat "Tolong sapu kamar saya." kata sapu di sini masuk dalam kategori kelas kata Verba, sedangkan dalam kalimat "Sapu itu sudah rusak." kata sapu di sini masuk dalam kategori kelas kata Nomina. Hal-hal seperti inilah yang menyebabkan komputer mengalami kesulitan dalam menginterpretasikan/menerjemahkan bahasa alami ke dalam bahasa komputer.

Proses penguraian kalimat dalam bahasa manusia mirip dengan proses penguraian tata bahasa pemrograman dalam dunia komputer. Perbedaan yang mendasar pada keduanya adalah tata bahasa dalam dunia komputer merupakan tata bahasa yang bebas konteks (*context free grammar*), sedangkan tata bahasa pada manusia adalah tata bahasa yang peka terhadap konteks (*context sensitive*). Pendefinisian tata bahasa yang peka terhadap konteks untuk diproses oleh komputer merupakan hal yang kompleks. Oleh karena itu, salah satu alternatif solusi supaya komputer dapat menginterpretasikan bahasa manusia ke dalam bahasa komputer adalah dengan memisahkan analisa sintaks dari suatu kalimat dengan analisa semantiknya [2].

Penelitian dalam bidang pemrosesan bahasa alami ini sudah banyak dilakukan di luar negeri. Bukti konkritnya dapat dilihat di internet [7] [8] [9] [10] [12]. Sementara itu, penelitian pemrosesan bahasa alami untuk bahasa Indonesia sendiri masih jarang dilakukan. Oleh karena itu, penulis mencoba melakukan penelitian pemrosesan bahasa alami untuk bahasa Indonesia. Penelitian ini difokuskan pada pembuatan aturan sintaks kalimat bahasa Indonesia.

Dikarenakan luasnya cakupan penelitian tentang struktur kalimat bahasa Indonesia maka dalam penelitian ini diperlukan adanya suatu pembatasan masalah sehingga penelitian ini bisa fokus untuk memecahkan suatu jenis permasalahan. Adapun batasan-batasan masalah dalam penelitian ini adalah:

1. Struktur kalimat bahasa Indonesia yang dibuat dalam penelitian ini adalah struktur kalimat yang sesuai dengan aturan sintaks tata bahasa baku bahasa Indonesia.
2. Pengurai struktur kalimat ini dibuat dengan menggunakan metode aturan bebas konteks (*context free grammar*). Artinya, setiap kalimat diuraikan tanpa melakukan analisa semantis (analisa arti dari kalimat).

3. Pengurai struktur kalimat ini dibangun dengan tujuan untuk menguraikan struktur kalimat pada artikel media massa elektronik, dalam hal ini yaitu kolom berita utama, kolom berita ekonomi, dan kolom berita olahraga dari Kompas dan Media Indonesia. Artikel untuk pengujian ini terbatas pada artikel terbitan bulan Mei, Juni, dan Juli.
4. Jenis kalimat yang dapat diuraikan oleh pengurai struktur kalimat di atas:
 - a. Berdasarkan jumlah klausanya: kalimat tunggal, dan kalimat majemuk.
 - b. Berdasarkan kategori predikatnya: kalimat berpredikat verbal, kalimat berpredikat adjektival, kalimat berpredikat nominal (termasuk pronominal), dan kalimat berpredikat numeral.
 - c. Berdasarkan kemungkinan kehadiran nomina atau frasa nominal objeknya: kalimat intransitif, kalimat ekatransitif, kalimat semitransitif dan kalimat dwitransitif.
 - d. Berdasarkan peran subjeknya: kalimat aktif (jika subjek berperan sebagai pelaku) dan kalimat pasif (jika subjek berperan sebagai sasaran).
 - e. Berdasarkan kemajemukannya: kalimat majemuk setara dan kalimat majemuk bertingkat.
 - f. Berdasarkan bentuk atau kategori sintaksisnya: kalimat deklaratif atau kalimat berita.
 - g. Berdasarkan struktur subjek dan predikatnya: kalimat biasa (subjek mendahului predikat).
5. Kalimat yang mengalami peniadaan unsur-unsurnya tidak diuraikan. Hasil keluaran dari pengurai struktur kalimat ini adalah pohon pengurai (*parse tree*) jika kalimat masukan sesuai dengan aturan tata bahasa baku bahasa Indonesia. Sebaliknya jika kalimat tidak sesuai dengan aturan tata bahasa baku bahasa Indonesia, maka struktur pohon pengurainya tidak akan terbentuk.
6. Aturan pengurai sintaks yang dibangun dalam penelitian ini memfokuskan pada pembangunan aturan sintaks untuk frasa nominal. Aturan sintaks frasa nominal ini kemudian digunakan untuk membangun aturan-aturan konstituen pembentuk kalimat, yaitu subjek, predikat, objek, pelengkap, dan keterangan. Selanjutnya, semua aturan yang telah dibangun untuk frasa nominal dan konstituen-konstituen tersebut digunakan untuk membentuk aturan kalimat deklaratif.

Yang digunakan sebagai sumber data untuk bahan penelitian ini adalah media massa elektronik. Pemilihan media massa elektronik sebagai domain pengujian dilakukan karena media massa elektronik tersebut memiliki karakteristik seperti di bawah ini:

1. Jumlah media massa elektronik ini banyak, dan pembacanya meliputi sebagian besar masyarakat.
2. Penggunaan bahasanya dapat dianggap baku. Asumsi ini diambil karena sebagian besar para jurnalis yang menjadi penulis berita pada media massa elektronik ini merupakan orang-orang yang memiliki latar belakang pendidikan bahasa Indonesia yang cukup baik.
3. Golongan pembacanya pada umumnya berasal dari lapisan masyarakat golongan menengah ke atas. Golongan pembaca ini dapat dikategorikan pada golongan masyarakat "intelektual".

Berdasarkan alasan di atas, harian yang dipilih sebagai sumber data adalah media massa elektronik Kompas dan Media Indonesia.

Adapun kolom berita yang dipilih sebagai bahan penelitian adalah kolom berita utama, kolom berita ekonomi, dan kolom berita olahraga. Kolom berita utama ini dipilih karena kolom ini dianggap mempunyai tingkat kebakuan yang tinggi dalam penggunaan tata bahasa. Sementara itu, kolom berita ekonomi dan kolom berita olahraga dipilih karena kedua kolom berita ini mempunyai jumlah pembaca yang paling banyak.

Tujuan dari penelitian ini adalah:

1. Membangun aturan-aturan sintaks untuk pengurai sintaks kalimat bahasa Indonesia dengan menggunakan metode *constraint-based formalisms*.
2. Menganalisa hasil penguraian kalimat dengan aturan sintaks yang telah didefinisikan.

2. Metode Penelitian

Penelitian dimulai dengan mempelajari buku tata bahasa baku bahasa Indonesia terbitan Balai Pustaka [14]. Setelah buku tersebut dipahami dengan baik, penulis kemudian melakukan analisa terhadap contoh-contoh kalimat yang ada pada buku tersebut.

Dari hasil analisa tersebut, penulis kemudian mencari suatu metode yang tepat untuk menguraikan struktur kalimat dari kalimat-kalimat contoh yang telah dipilih tersebut. Strategi penguraian yang digunakan dalam penelitian ini adalah penguraian yang bebas konteks (*context free*). Strategi bebas konteks ini digunakan karena strategi ini sesuai atau mirip dengan tata bahasa pemrograman yang sering digunakan dalam bidang komputer. Disamping itu, strategi ini juga sudah sering digunakan dalam pemrosesan bahasa alami [1] [10].

Setelah menetapkan strategi penelitian, penulis kemudian mulai mempelajari *tools* yang akan digunakan untuk membuat aturan-aturan sintaks untuk menguraikan struktur kalimat tersebut. *Tools* yang dipergunakan untuk membuat pengurai sintaks ini adalah PC-PATR. PC-PATR adalah suatu pengurai sintaks yang menggunakan metode *constraint-based formalisms*. *Constraint-based formalisms* adalah suatu metode yang memungkinkan aturan-aturan sintaks untuk diberikan suatu batasan-batasan (*constraints*) sehingga dapat meminimalkan keambiguan dari aturan sintaks yang dibuat. PC-PATR adalah pengurai sintaks yang memungkinkan pengidentifikasian terhadap ambiguitas yang terjadi pada aturan-aturan sintaks yang dibangun. Akibatnya, hasil keluaran dari pengurai struktur kalimat ini dapat berupa satu atau lebih dari satu buah pohon struktur [16].

Setelah metode penguraian PC-PATR dipahami dengan baik, penulis kemudian mulai membangun aturan-aturan sintaks untuk kalimat-kalimat sederhana yang ada di buku tata bahasa baku bahasa Indonesia. Setelah aturan ini selesai dibuat, penulis kemudian mulai melakukan pengujian terhadap kalimat-kalimat sederhana ini. Pengujian ini merupakan pengujian tahap pertama. Setelah memperoleh hasil yang cukup baik, kemudian sedikit demi sedikit aturan-aturan tersebut dimodifikasi untuk dapat menerima kalimat-kalimat yang lebih kompleks (kalimat-kalimat ini masih diambil dari buku tata bahasa baku bahasa Indonesia). Setelah pengujian tahap pertama ini selesai, pengujian kemudian dilanjutkan ke tahap yang kedua.

Pada pengujian tahap kedua ini, penulis mengujicobakan aturan-aturan sintaks yang telah dibuat dalam pengujian tahap pertama pada kalimat-kalimat yang lebih spesifik, yaitu kalimat-kalimat yang terdapat pada kolom berita utama, kolom berita ekonomi dan kolom berita olahraga (Kompas dan Media Indonesia).

3. Struktur Kalimat Bahasa Indonesia

3.1 Bagian-Bagian Kalimat

Dilihat dari segi bentuknya, kalimat dapat dirumuskan sebagai konstruksi sintaksis terbesar yang terdiri atas dua kata atau lebih. Antara "kalimat" dan "kata" terdapat dua satuan sintaksis antara, yaitu "klausa" dan "frasa". Klausa merupakan satuan sintaksis yang terdiri atas dua kata, atau lebih, yang mengandung unsur predikasi. Sedangkan frasa adalah satuan sintaksis yang terdiri atas dua kata atau lebih yang tidak mengandung unsur predikasi.

3.2 Struktur Kalimat Dasar

Yang dimaksud dengan kalimat dasar adalah kalimat yang (i) terdiri atas satu klausa, (ii) unsur-unsurnya lengkap, (iii) susunan unsur-unsurnya menurut urutan yang paling umum, dan (iv) tidak mengandung pertanyaan atau pengingkaran [14]. Dengan kata lain, kalimat dasar di sini identik dengan kalimat tunggal deklaratif afirmatif yang urutan unsur-unsurnya paling lazim. Berdasarkan ciri-ciri yang dimilikinya, kalimat dasar dapat dibedakan ke dalam enam tipe [14].

Tabel 1. Pola Kalimat Dasar

Fungsi Tipe	S	P	O	Pel	Ket
1. S-P	Saya	dosen			
2. S-P-O	Rani	mendapat	uang		
3. S-P-Pel	Pancasila	adalah		dasar negara kita	
4.S-P-Ket	Itu	terjadi			tadi pagi
5. S-P-O-Pel	Dia	mengirim	kami	uang	
6. S-P-O-Ket	Beliau	memberi	kami		rasa aman

Pada struktur kalimat dasar, di antara kalimat dan kata, biasanya ada satuan-antara yang berupa kelompok kata. Kelompok kata yang menjadi salah satu unsur pembentuk kalimat ini dikenal dengan sebutan frasa. Dalam bahasa Indonesia dikenal lima buah frasa yang lazim menjadi unsur pembentuk kalimat. Kelima buah frasa ini, antara lain yaitu : *frasa verbal*, *frasa nominal*, *frasa pronominal*, *frasa adjektival*, dan *frasa numeralia*.

3.3 Fungsi Sintaksis Unsur-Unsur Kalimat

Terdapat lima fungsi sintaksis yang digunakan untuk membangun sebuah kalimat, yaitu subjek, predikat, objek, pelengkap dan keterangan. Suatu pernyataan merupakan suatu kalimat jika di dalam pernyataan tersebut sekurang-kurangnya terdapat subjek dan predikat, baik disertai objek, pelengkap, atau keterangan maupun tidak.

4. *Constraint-Based Formalisms* dan PC-PATR

4.1 *Constraint-Based Formalisms*

Constraint-based formalisms merupakan suatu metode pereduksian *gramatical theories* dari suatu struktur gramatikal yang lebih besar dan kompleks menjadi struktur gramatikal yang lebih kecil dengan menggunakan batasan-batasan atau *constraints* [15].

Alasan penggunaan metode *constraint-based formalisms* pada penelitian tugas akhir ini adalah untuk mengatasi permasalahan-permasalahan yang terjadi pada pengembangan pengurai sintaks bahasa Indonesia yang telah dilakukan oleh Shelly Salvitri [13] dengan menggunakan metode *Backus-Naur-Form* (BNF). Penggunaan metode *Backus-Naur-Form* (BNF) ini mengharuskan kita untuk membuat suatu aturan baru jika terjadi suatu kasus khusus. Akibatnya, semakin banyak kasus khusus yang muncul, maka akan semakin banyak aturan sintaks yang harus dibuat. Dengan penggunaan metode *constraint-based formalisms* ini diharapkan kasus di atas akan dapat diatasi.

Implementasi dari *constraint-based formalisms* yang digunakan dalam penelitian ini adalah *constraint-based formalisms* yang berdasarkan *feature structures* dan *unification*.

4.2 PC-PATR

PC-PATR merupakan implementasi dari metode *constraint-based formalisms* pada *personal computer* (PC). PC-PATR menggunakan *left corner chart parser* dengan karakteristik seperti berikut [16]:

1. *Bottom-up parse* dengan *top-down filtering* berdasarkan kategori/kelas kata.
2. *Left-to-right order*—setelah tiap-tiap kata dipetakan/diformulasikan, semua kemungkinan hasil *parsing* akan dimunculkan.

Struktur PC-PATR ini terdiri dari sekumpulan *rule* (aturan) dan *lexicon* (kamus kata). Setiap *rule* terdiri dari *context-free phrase structure rule* dan sekumpulan *feature constraints*. *Feature constraints* ini adalah unifikasi pada *feature structures* yang disesuaikan dengan konstituen kalimat pada *rule*.

1. *Phrase structure rules*

Context-free phrase structure rule tersebut dapat direpresentasikan seperti berikut:

$$\text{LHS} \rightarrow \text{RHS}_1 \text{ RHS}_2 \quad (1)$$

'LHS' (*left hand symbol*) merupakan simbol nonterminal untuk *phrase* 'RHS₁ RHS₂'. 'RHS₁ RHS₂' adalah urutan konstituen dari *phrase* yang ingin dibentuk. Konstituen-konstituen ini bisa berupa simbol nonterminal ataupun simbol terminal. Simbol nonterminal berarti simbol tersebut akan muncul lagi pada sisi kiri sebagai 'LHS' untuk mendefinisikan *phrase* yang ingin dijelaskan. Sementara itu, simbol terminal berarti simbol yang sudah bisa digantikan dengan penggantinya yang ada di *lexicon*.

2. *Feature constraints*

Feature constraints pada *phrase structure rule* terdiri dari sekumpulan ekspresi unifikasi (*unification constraints*). Setiap ekspresi unifikasi mempunyai 3 bagian, dengan urutan sebagai berikut:

- a. *Feature path*: elemen pertama dari suatu *phrases structure rule*.
- b. Tanda sama dengan (=)
- c. Suatu nilai sederhana ataupun *feature path* yang lain yang juga dimulai dengan suatu simbol dari *phrase structure rule*.

Untuk lebih jelasnya, penggunaan *feature constraints* ini dapat dilihat pada contoh di bawah ini:

$$\begin{array}{ll} \text{Rule S} & \rightarrow \text{NP VP (Subcl)} \\ <\text{NP head agr}> & = <\text{VP head agr}> \\ <\text{NP head case}> & = \text{NOM} \\ <\text{S subj}> & = <\text{NP}> \\ <\text{S head}> & = <\text{VP head}> \end{array} \quad (2)$$

Aturan ini mempunyai dua *feature constraints* yang membatasi kemunculan kembali NP dan VP. Dengan kata lain aturan ini mempunyai dua fitur yang dapat mencegah munculnya ambiguitas dari VP dan NP. Disamping itu, pada aturan di atas juga terdapat dua *feature constraints* yang membentuk suatu *feature structures* dari aturan produksi S.

5. Analisa dan Perancangan

5.1 Penentuan Kelas Kata

Kalimat umumnya berwujud rentetan kata yang disusun sesuai dengan kaidah yang berlaku. Dengan kata lain, kata adalah unsur terkecil dari struktur sintaks suatu bahasa, sedangkan kalimat adalah unsur terbesarnya. Tiap kata dalam kalimat mempunyai tiga klasifikasi, yaitu berdasarkan (1) kategori sintaksis, (2) fungsi sintaksis, dan (3) peran semantis [14]. Dalam penguraian struktur sintaks suatu kalimat, kelas kata merupakan unsur yang paling sederhana yang tidak dapat diuraikan lagi menjadi unsur lain yang lebih sederhana. Sehingga dalam mendefinisikan aturan-aturan sintaks, kelas kata ini akan menjadi terminal atau *token* [10]. Dalam proses penguraian struktur kalimat, jenis-jenis kelas kata ini akan diperiksa pada saat proses analisa leksikal atau pada saat penyusunan kamus kata. Jenis-jenis kelas kata yang digunakan dalam penelitian ini dapat dilihat pada tabel II di bawah ini:

Tabel 2. Kelas Kata yang Digunakan dalam Penelitian

Kelas Kata	Simbol	Contoh
Verba transitif	Vtran	membeli, membelikan
Verba intransitif	Vintr	adalah, dibelikan, bekerja
Adjektiva	Adj	cantik, lembut, sopan
Nomina	N	baju, kebangsaan, negara
Adverbia	Adv	sangat, selalu, hamper
Pronomina	Pron	dia, saya

Numeralia	Num	satu, pertama, sejuta
Preposisi	Prep	di, ke, dari
Konjungtor koordinatif	Kkord	dan, atau, maupun
Konjungtor subordinatif	Ksub	sedangkan, ketika, selama
Konjungtor kalimat	Kkal	ironisnya, sementara itu
Interjeksi	Intr	ayo, mari, brengsek
Artikula	Art	si, sang, para
Partikel	Part	lah, kah, tah, pun
Penggolong	Gol	orang, buah, ekor

Seperti yang sudah dijelaskan pada bab I, PC-PATR yang akan digunakan sebagai *parser* dalam penelitian ini adalah suatu pengurai yang memungkinkan terjadinya keambiguan dalam pendefinisian aturan sintaksnya. Akibatnya, jika kita langsung menggunakan kelas kata yang terdapat pada tabel II, maka akan timbul banyak keambiguan dari aturan sintaks yang akan dibuat nantinya. Oleh karena itu, dengan menggunakan *feature structures* yang dimiliki oleh pengurai PC-PATR, maka setiap kelas kata yang terdapat pada tabel II, kemudian diberikan sub kategori untuk lebih menspesifikasikan jenis kelas kata yang diinginkan dalam pembuatan kamus kata. Berikut akan dijelaskan mengenai kelas kata berikut sub kategorinya yang digunakan dalam penelitian ini.

Tabel 3. Sub Kategori Nomina dan Pronomina

Kelas Kata	Sub Kategori	Contoh
Nomina	Waktu	kemarin, sekarang
	Tempat	kota, rumah, lapangan
	Benda	meja, kursi, gunting
	Orang	orangtua, pria, istri
	Profesi	tentara, petani, presiden
	Istilah	abolisi, acara, adat, administrasi
	Milik	bukunya, pernyataannya
	Majemuk	daerah istimewa
Pronomina	Orang	kakak, saya, kamu, mereka
	Tunjuk	ini, itu, sana, sini

5.2 Rancangan Struktur Kalimat

Dengan menggunakan pola kalimat dasar, disertai dengan informasi yang terdapat pada tabel kelas kata dan tabel sub kategori, maka dirancang aturan-aturan sintaks untuk kalimat bahasa Indonesia (kalimat deklaratif) seperti dibawah ini :

- Kalimat berpredikat verba transitif dengan subkategori "ekatransitif"**
KALIMAT = S + P + O + (K)
- Kalimat berpredikat verba transitif dengan subkategori "dwitransitif"**
KALIMAT = S + P + (O) + PEL + (K)
- Kalimat berpredikat verba transitif dengan subkategori "semitransitif"**
KALIMAT = S + P + (O) + (K)
- Kalimat berpredikat verba intransitif dengan subkategori "berpelengkap"**
KALIMAT = S + P + PEL + (K)
- Kalimat berpredikat verba intransitif dengan subkategori "tberpelengkap"**
KALIMAT = S + P + (K)
- Kalimat berpredikat verba intransitif dengan subkategori "pasif"**
KALIMAT = S + P + (PEL) + (K)
- Kalimat berpredikat numeralia**
KALIMAT = S + P + (K)
- Kalimat berpredikat nomina**
KALIMAT = S + P + (K)
- Kalimat berpredikat adjektiva dengan subkategori "warna"**
KALIMAT = S + P + (K)
- Kalimat berpredikat adjektiva dengan subkategori "bandingan"**
KALIMAT = S + P + PEL + (K)

11. Kalimat berpredikat adjektiva dengan subkategori “biasa”

KALIMAT = S + P + (PEL) + (K)

6. Implementasi**6.1 Implementasi Aturan Sintaks untuk Frasa Nominal****1. Aturan Sintaks Frasa nominal untuk Non Subjek**

Rule FN -> {(WDEPANFN)

```

{
    {{NMINORG_4 / N} {WBLKGFN_1 / WBLKGFN_4}}
    /
    {{NMINORG_1 / N_1} WBLKGFN_2}
    /
    {NMINORG_2 WBLKGFN_3}
    /
    {{NMINORG}
    {WBLKGFN_6 / WBLKGFN_7 / WBLKGFN_8 / WBLKGFN_9 / WBLKGFN_10}}
    /
    {NMINORG_5 WBLKGFN_5}
    /
    NMINORG_3
    /
    {N_2 / N_3 / {N_4 N_5}}
}

```

(WBELAKANG) ((KKORD_1) PENJELAS)} (KKORD_2 FN_1)

<N subcat>	= orang
<N_1 subcat>	= orang
<N_2 subcat>	= orang
<N_3 subcat>	= milik
<N_4 subcat>	= milik
<N_5 subcat>	= orang
<WBLKGFN_1 subcat>	= FPRON
<WBLKGFN_2 subcat>	= FADJ
<WBLKGFN_3 subcat>	= FNUM
<WBLKGFN_4 subcat>	= waktu
<WBLKGFN_5 subcat>	= tempat
<WBLKGFN_6 subcat>	= orang
<WBLKGFN_7 subcat>	= benda
<WBLKGFN_8 subcat>	= profesi
<WBLKGFN_9 subcat>	= istilah
<WBLKGFN_10 subcat>	= majemuk
<FN subcat>	= <NMINORG_4 subcat>
<FN subcat>	= <N subcat>
<FN subcat>	= <NMINORG_1 subcat>
<FN subcat>	= <N_1 subcat>
<FN subcat>	= <NMINORG_2 subcat>
<FN subcat>	= <WBLKGFN_6 subcat>
<FN subcat>	= <WBLKGFN_7 subcat>
<FN subcat>	= <WBLKGFN_8 subcat>
<FN subcat>	= <WBLKGFN_9 subcat>
<FN subcat>	= <WBLKGFN_10 subcat>
<FN subcat>	= <NMINORG_5 subcat>
<FN subcat>	= <NMINORG_3 subcat>
<FN subcat>	= <N_2 subcat>
<FN subcat>	= <N_3 subcat>
<FN subcat>	= <N_5 subcat>
<FN_1 subcat>	= <FN subcat>

2. Aturan Sintaks Frasa nominal Subjek

Rule FNSUBJ -> {(WDEPANFN)

```
{
    {{NMINORG_4 / N} WBLKGFN_1}
    /
    {{NMINORG_1 / N_1} WBLKGFN_2}
    /
    {NMINORG_2 WBLKGFN_3}
    /
    {{NMINORG}
    {WBLKGFN_4 / WBLKGFN_5 / WBLKGFN_6 / WBLKGFN_7 / WBLKGFN_8 /
    WBLKGFN_9 / WBLKGFN_10}
}
```

```
/
NMINORG_3
/
{N_2 / N_3 / {N_4 N_5}}
```

```
}
(WBELAKANG)
((KKORD_1) PENJELAS))
(KKORD_2 FNSUBJ_1)
```

```
<N subcat> = orang
<N_1 subcat> = orang
<N_2 subcat> = orang
<N_3 subcat> = milik
<N_4 subcat> = milik
<N_5 subcat> = orang
<WBLKGFN_1 subcat> = FPRON
<WBLKGFN_2 subcat> = FADJ
<WBLKGFN_3 subcat> = FNUM
<WBLKGFN_4 subcat> = waktu
<WBLKGFN_5 subcat> = tempat
<WBLKGFN_6 subcat> = orang
<WBLKGFN_7 subcat> = benda
<WBLKGFN_8 subcat> = profesi
<WBLKGFN_9 subcat> = istilah
<WBLKGFN_10 subcat> = majemuk
<FNSUBJ subcat> = <WBLKGFN_1 subcat>
<FNSUBJ subcat> = <NMINORG_1 subcat>
<FNSUBJ subcat> = <N_1 subcat>
<FNSUBJ subcat> = <NMINORG_2 subcat>
<FNSUBJ subcat> = <WBLKGFN_4 subcat>
<FNSUBJ subcat> = <WBLKGFN_5 subcat>
<FNSUBJ subcat> = <WBLKGFN_6 subcat>
<FNSUBJ subcat> = <WBLKGFN_7 subcat>
<FNSUBJ subcat> = <WBLKGFN_8 subcat>
<FNSUBJ subcat> = <WBLKGFN_9 subcat>
<FNSUBJ subcat> = <WBLKGFN_10 subcat>
<FNSUBJ subcat> = <NMINORG_3 subcat>
<FNSUBJ subcat> = <N_2 subcat>
<FNSUBJ subcat> = <N_3 subcat>
<FNSUBJ subcat> = <N_5 subcat>
<FNSUBJ subcat> = <FNSUBJ_1 subcat>
```

7. Pengujian

7.1 Pengujian Tahap Pertama

Pengujian tahap pertama ini menggunakan masukan contoh-contoh kalimat yang terdapat pada buku tata bahasa baku bahasa Indonesia [14]. Pengujian tahap pertama ini difokuskan pada kalimat-kalimat tunggal.

Tujuan dari uji coba tahap pertama ini adalah untuk memeriksa dan mengalisa kebenaran dari aturan-aturan sintaks yang telah dirancang dan didefinisikan pada penelitian ini. Oleh karena itu, jenis masukan yang digunakan sebagai sampel pada pengujian tahap pertama ini adalah kalimat-kalimat yang strukturnya sudah didefinisikan dalam aturan-aturan sintaks pada penelitian ini.

7.2 Pengujian Tahap Kedua

Pengujian tahap kedua ini menggunakan kalimat-kalimat masukan yang berasal dari media massa elektronik Kompas (KP) dan Media Indonesia (MI). Hasil penguraian kalimat dari pengujian tahap kedua ini dapat dilihat pada tabel di bawah ini :

Tabel 4. Hasil Penguraian Kalimat

	Domain Pengujian	Jenis-Jenis Hasil Penguraian								
		0	1	2	3	4	5	6	7	8
MI	Utama	1	32	6	-	1	-	-	-	-
	Ekonomi	5	45	14	1	-	-	-	-	-
	Olahraga	2	69	13	1	-	-	-	-	-
KP	Utama	1	25	8	2	4	-	-	-	-
	Ekonomi	4	45	15	-	-	-	1	-	-
	Olahraga	7	59	16	-	1	-	-	1	1
Jumlah Total Penguraian		20	275	72	4	6	0	1	1	1
Persentase Hasil Penguraian		5,26	72,37	18,95	1,05	1,58	0	0,26	0,26	0,26
Persentase kalimat terurai terhadap yang tidak terurai adalah 94,7 %										

Tabel 5. Kalimat Tunggal dan Kalimat Majemuk

	Domain Pengujian	Kalimat Tunggal		Kalimat Majemuk Bertingkat		
		Aktif	Pasif	Aktif	Pasif	Setara
Media Indonesia	Utama	17	8	11	3	0
	Ekonomi	28	7	18	6	1
	Olahraga	41	16	21	4	1
Kompas	Utama	13	6	18	1	1
	Ekonomi	29	5	23	2	2
	Olahraga	43	9	21	3	2
Jumlah Kalimat		171	51	112	19	7
Persentase dari 360 kalimat yang dapat terurai		47,5 %	14,2 %	31,1 %	5,3 %	1,94 %
Persentase Kalimat Tunggal yang dapat terurai = 61,67 %						
Persentase Kalimat Majemuk yang dapat terurai = 38,33 %						

8. Kesimpulan

Kesimpulan yang dapat diambil berdasarkan hasil pengujian yang dilakukan dalam penelitian ini memiliki hasil implementasi pengurai yang memuat struktur kalimat tunggal dan kalimat majemuk bahasa Indonesia cukup baik. Hasil analisa struktur kalimat bahasa Indonesia terhadap domain kalimat sampel dengan menggunakan aturan-aturan sintaks yang telah didefinisikan pada penelitian ini dalam kolom berita utama pada Media Indonesia dan Kompas

memiliki persentase yang sama, sedangkan pada kolom berita olahraga memiliki nilai persentase yang berbeda. Terlihat bahwa hasil penguraian persentase keteruraian kalimat-kalimat pada media massa elektronik Media Indonesia sedikit lebih baik dibandingkan dengan kalimat-kalimat pada media massa elektronik Kompas. Jika dihubungkan dengan frasa nominal, ketidak teruraian suatu kalimat sebagian besar disebabkan oleh pembatasan-pembatasan yang ada pada aturan sintaks frasa nominal.

Daftar Pustaka

- [1] Sager, N., "Natural Language Information Processing : A Computer Grammar of English and its Applications", Addison-Wesley Publishing Company, Massachusetts 1981.
- [2] Allen, J., "Natural Language Understanding", The Benjamin/Cumming Publishing Company, Inc., Redwood City, CA 1994.
- [3] Kamlesh Sharma, S.V.A.V. Prasad, T. V. Prasad, "A Hindi speech actuated computerinterface for web search," International Journal of Advanced Computer Science and Applications, Vol. 3, No. 10, 2012.
- [4] Md. Musfique Anwar, Nasrin Sultana Shume and Md. Al-Amin Bhuiyan, "Structural analysis of Bangla sentences of different tenses for automatic Bangla Machine Translator," International Journal of Computer Science and Information Security, Vol. 8, No. 9, December 2010
- [5] Matthias Scheutz, Rehj Cantrell, Paul Schermerhorn, "Toward humanlike task-based dialogue processing for human robot interaction," Association for the Advancement of Artificial Intelligence. ISSN 0738-4602.
- [6] Gauri Rao, Chanchal Agarwal, Snehal Chaudhry, Nikita Kulkarni, S.H. Patil, "Natural language query processing using semantic grammar," International Journal on Computer Science and Engineering Vol. 02, No. 02, 2010, 219-223.
- [7] K. M. Azharul Hasan, Al-Mahmud, Amit Mondal, Amit Saha, "Recognizing Bangla grammar using predictive parser," International Journal of Computer Science & Information Technology (IJCSIT) Vol 3, No 6, Dec 2011.
- [8] Mohammad Bahrani, Hossein Sameti, Mehdi Hafezi Manshadi, "A computational grammar for Persian based on GPSG," Lang Resources & Evaluation (2011) 45:387–408 Springer Science+Business Media B.V. 2011.
- [9] Nitin Hambir, Ambrish Srivastav, "Hindi parser-based on CKY algorithm," International Journal Computer Technology & Applications, Vol 3 (2), 851-853.
- [10] P.G. Magdum, D. V. Kodavade, "Grammar for checking the syntactic validity of English sentence," International Journal of Advanced Research in Computer Science and Electronics Engineering Volume 2, Issue 1, January 2013.
- [11] DU Jia-li, YU Ping-fang, "A computational linguistic approach to natural language processing with applications to garden path sentences analysis", International Journal of Advanced Computer Science and Applications, Vol. 3, No. 9, 2012.
- [12] Yang Huang, J Henry Lowe, Dan Klein, J. Russel Cucina, "Improved identification of noun phrases in clinical radiology reports using a high-performance statistical natural language parser augmented with the UMLS specialist lexicon," Journal of the American Medical Informatics Association, May/Jun 2005, 12, 3, ProQuest pg. 275.
- [13] Shelly Salvitri, "Pengembangan Lanjut Pengurai Sintaks Bahasa Indonesia dengan Metode Linguistic String Analysis," Skripsi Sarjana Ilmu Komputer, 1999.
- [14] Alwi, H., Soenjono Dardjowidjojo, Hans Lapoliwa, Anton M. Moeliono, "Tata Bahasa Baku Bahasa Indonesia", Departemen Pendidikan dan Kebudayaan Republik Indonesia, Jakarta, 1998.
- [15] Daniel Jurafsky & James H. Martin, "Speech and Language Processing – An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition", Prentice Hall, Inc, 2000.
- [16] Stephen McConnel, "PC-PATR Reference Manual – a unification based syntactic parser - version 1.0", Oktober 1998.