

PENGURAIAN DENGAN METODE STATISTIK UNTUK BAHASA INDONESIA

Rosa Ariani Sukamto

23507024

Pendahuluan

- Latar Belakang
- Rumusan Masalah
- Tujuan
- Ruang Lingkup
- Batasan Masalah
- Metode Penelitian

Latar Belakang

- Pemrosesan bahasa alami berawal dari keinginan manusia untuk berkomunikasi dengan komputer menggunakan bahasa manusia.
- Untuk memahami makna bahasa manusia dengan benar maka komputer perlu mengetahui tata pola bahasa manusia.
- Pengurai (*parser*) akan membentuk pohon pola tata bahasa sehingga dapat disimpulkan dimana inti dari sebuah kalimat berbahasa manusia.
- Sebelumnya telah dilakukan beberapa riset mengenai topik tesis ini untuk berbagai bahasa, seperti
 - bahasa Inggris (Michael Collins, Daniel Jurasky dan James H. Martin),
 - bahasa Melayu Malaysia (Mohd Juzaidin Ab Aziz dkk)
 - bahasa Indonesia dengan perangkat PC-PATR (Ria H. Gusmita , Ruli Manurung
- Belum ada paper atau riset resmi yang dipublikasikan menggunakan pengurai Collins untuk bahasa Indonesia

Rumusan Masalah

- Rumusan masalah dalam tesis ini adalah sebagai berikut:
 - Bagaimana melakukan penyesuaian kumpulan *file* masukan dari pengurai Collins dengan bahasa Indonesia,
 - Bagaimana melakukan penentuan jenis kata (*POS tagging*) untuk bahasa Indonesia guna mendukung pemrosesan awal.
 - Bagaimana kinerja pengurai Collins setelah menerima masukan kumpulan *file* hasil pemrosesan awal (apakah sudah cukup mampu merepresentasikan pola tata bahasa Indonesia).

Tujuan

- Tujuan dari tesis ini adalah sebagai berikut:
 - Mengimplementasikan pemrosesan awal kumpulan file masukan dari pengurai Collins menggunakan contoh-contoh teks yang berisi kumpulan kalimat berbahasa Indonesia; modul pemrosesan awal pengurai Collins akan diimplementasikan dalam sebuah program,
 - Mengimplementasikan POS *tagging* untuk bahasa Indonesia guna mendukung pemrosesan awal.
 - Melakukan evaluasi kinerja model pola tata bahasa yang dihasilkan; hasil model pola tata bahasa akan dievaluasi kinerjanya, apakah sudah sesuai dengan pola tata bahasa Indonesia baku.

Ruang Lingkup

- Pekerjaan yang dilakukan dalam tesis ini yang akan dilaksanakan oleh penulis adalah:
 - mengimplementasikan pemrosesan awal (*preprocessing*) untuk mempersiapkan kumpulan *file* masukan dari pengurai Collins dari *file* teks yang berisi kumpulan kalimat berbahasa Indonesia,
 - melakukan beberapa penyesuaian terhadap pengurai Collins agar dapat dijalankan untuk bahasa Indonesia.

Batasan Masalah

- Proses-proses yang dilakukan untuk pemrosesan awal (*preprocessing*) dan penentuan jenis kata (*POS tagging*) menggunakan jenis kalimat berbahasa Indonesia tanpa melibatkan ahli bahasa,
- Tidak semua aturan *grammar* (tata bahasa) bahasa Indonesia dipakai dalam tesis ini.
- Permasalahan ambiguitas penguraian kalimat tidak diujikan pada tesis ini.

Metode Penelitian

- Studi literatur yang berkaitan dengan *parsing* probabilistik, tata tulis bahasa Indonesia, riset-riset terkait pemrosesan temu balik informasi bahasa Indonesia
- Analisis masalah
- Implementasi metode dan algoritma untuk memodelkan pola tata bahasa Indonesia
- Pengujian hasil implementasi metode dan algoritma untuk memodelkan pola tata bahasa Indonesia
- Evaluasi dan penarikan kesimpulan

Tinjauan Pustaka

- Model-model Pola Tata Bahasa
- Pengurai (*parser*)
- Penguraian dengan Metode Probabilistik

Model-Model Pola Tata Bahasa [1]

- *Grammar* secara teknis merupakan sebuah alat untuk merepresentasikan sebuah bahasa
- *Grammar* yang berbasis struktur frasa (*phrase structure*) antara lain seperti *context-free grammar* (CFG)
- CFG juga terdiri dari empat parameter (*4-tuple*) yaitu kumpulan simbol non-terminal, kumpulan simbol terminal, kumpulan aturan produksi, dan kumpulan simbol awal.

Model-Model Pola Tata Bahasa [2] : PCFG

- PCFG terdiri dari lima buah *tuple* yaitu $G = (N, \Sigma, P, S, D)$ (simbol non-terminal, kumpulan simbol terminal, kumpulan aturan produksi, dan kumpulan simbol awal, kumpulan probabilitas aturan produksi)
- $P(T, S) = p(r(n))$,
- Jika aturan produksi $\alpha \rightarrow \beta$ maka

$$- P(\beta | \alpha) = \frac{\text{jumlah}(\alpha \rightarrow \beta)}{\text{jumlah}(\alpha)}$$

Model-Model Pola Tata Bahasa [3]

- *Stochastic Lexicalized Context-Free Grammar* (SLCFG) oleh Yves Schabes dan Richard C. Waters (1993) (model 11 tuple).
- Glen Carroll (1995) melakukan sebuah penelitian mengenai pembelajaran tata bahasa probabilistik untuk permodelan bahasa (mulai menggunakan PCFG).
- Mark Johnson (1998) melakukan penelitian mengenai model PCFG (*Probabilistic Context-Free Grammar*) untuk representasi pohon (membandingkan dengan model-model lain)

Parser (Pengurai) [1]

- *Parser* (pengurai) dalam tesis ini merupakan pengurai kalimat yang digunakan dalam pemrosesan bahasa alami.
- Fungsi *parser* (pengurai) pada tesis ini adalah sebagai pengurai kalimat untuk membuat pohon pola tata bahasanya dari teks masukan yang berisi kumpulan kalimat (*corpus*) berbahasa Indonesia

Parser (Pengurai) [2]

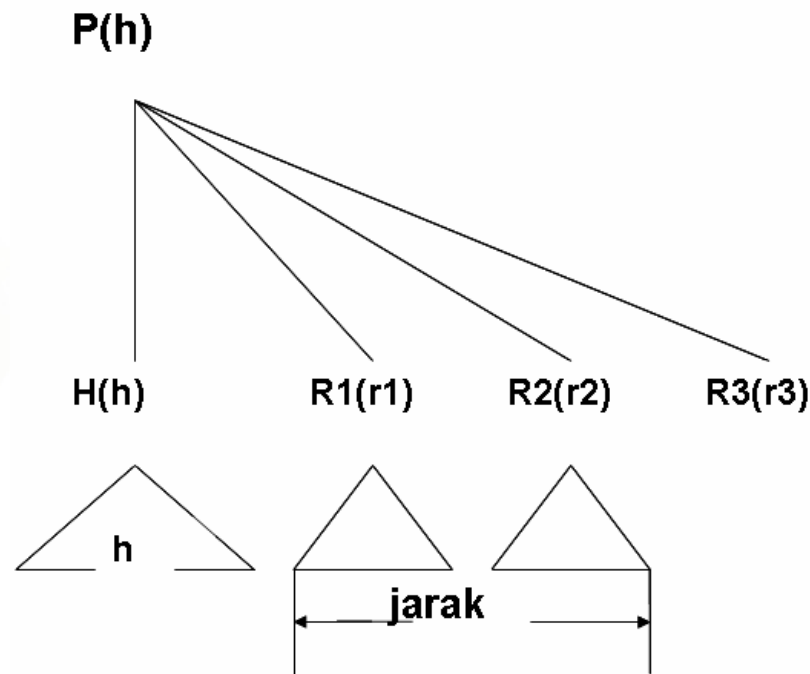
- *Parser* (pengurai) yang dibangun oleh Charniak (1997) adalah *parser* (pengurai) untuk bahasa Inggris dan menggunakan *treebank* untuk membangun sistem *parser* (pengurai) (*bottom up*)
- Charniak melakukan penelitian mengenai *parser* (Pengurai) dengan Menggunakan Entropi Maksimum (2000) (*top down*)

Parser (Pengurai) [3]

- Penelitian mengenai *parser* juga dilakukan oleh Michael Collins (1996). Penelitian ini mengenai *parser* (Pengurai) berbasis statistik pada ketergantungan bigram leksikal
- Collins juga melakukan penelitian mengenai *parsing* (penguraian) bahasa alami dengan model statistik berbasis *head-driven* (1999)
- Daniel M. Bikel (2004) dengan *Framework* untuk *parser* (pengurai)
- Collins melakukan penelitian mengenai pengurai Collins untuk bahasa Czech

Parser (Pengurai) [4]: Perhitungan Probabilistik Model 1

- $P(h) \rightarrow Ln(ln)...L1(l1)H(h)R1(r1)...Rm(rm)$



Parser (Pengurai) [5]: Perhitungan Probabilistik Model 2

- Adanya pembedaan pelengkap/keterangan dan pengkategorian sub kalimat yang menjadi pelengkap/keterangan
- menambahkan simbol –C pada simbol non-terminal sebagai penanda sub kalimat pelengkap/keterangan

Parser (Pengurai) [6]: Perhitungan Probabilistik Model 3

- Kalimat dengan kata hubung (majemuk) dan kalimat dengan kata tanya (kalimat bersarang)
- menambahkan simbol +gap untuk melakukan penelusuran terhadap kalimat, dan menghentikan penelusuran saat bertemu simbol TRACE

Parsing Probabilistik [1]

- Penelitian mengenai teknik pembangkitan pola tata bahasa untuk ekstraksi relasi pada bahasa Malaysia dilakukan oleh Mohd Juzaidin Ab Aziz dkk (2006)
- Penelitian mengenai *parsing* (penguraian) probabilistik dilakukan oleh Daniel Jurafsky dan James H. Martin (2000)

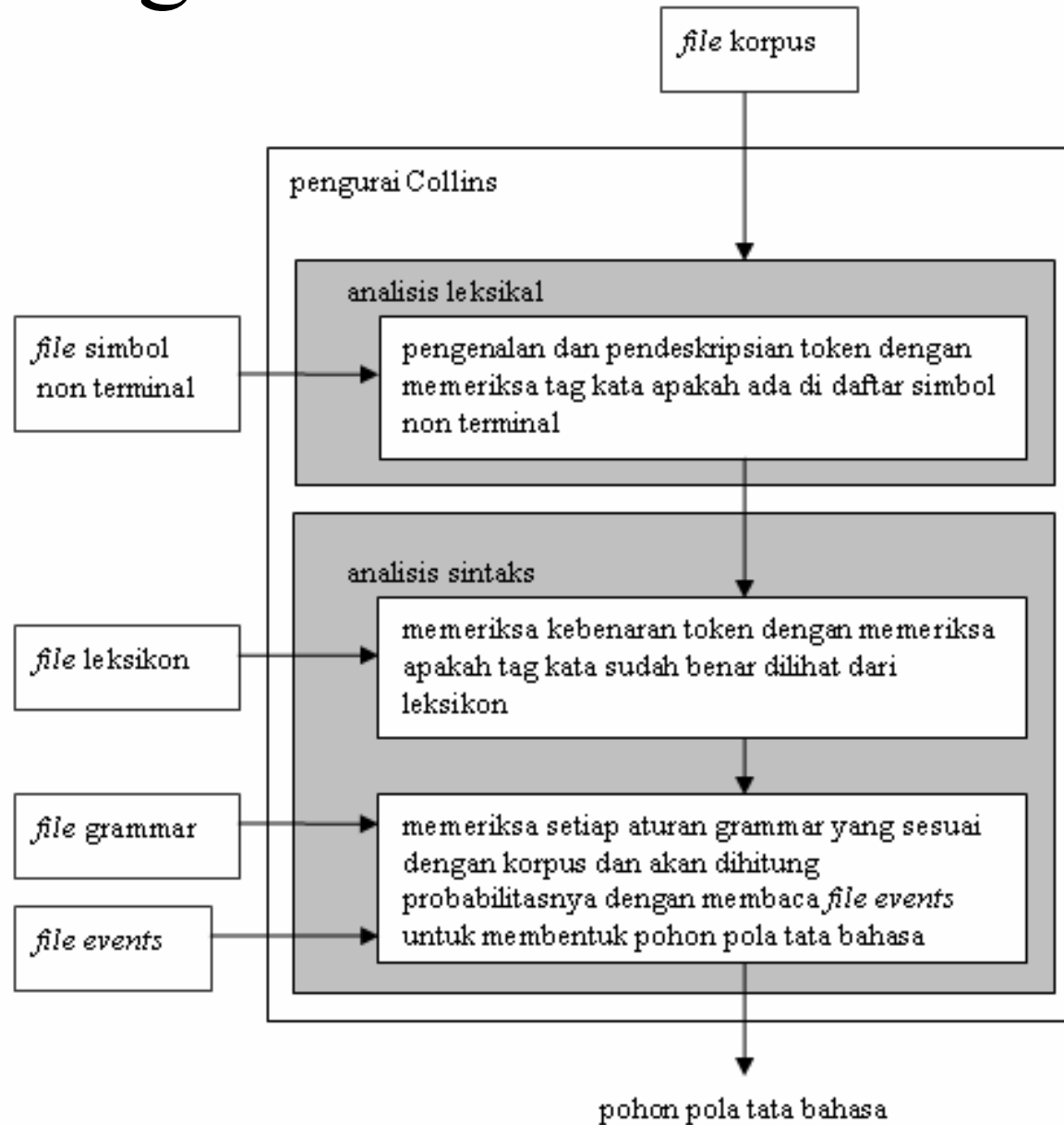
Parsing Probabilistik [2]

- Penelitian yang dilakukan Ramon Lefuel dan Brian J. Ross (2004) menggabungkan *parsing* probabilistik dengan algoritma genetik (kromosom = *parse-tree*)
- Penelitian yang dilakukan Ria Hari Gusmita dan Ruli Manurung menggunakan perangkat PC-PATR

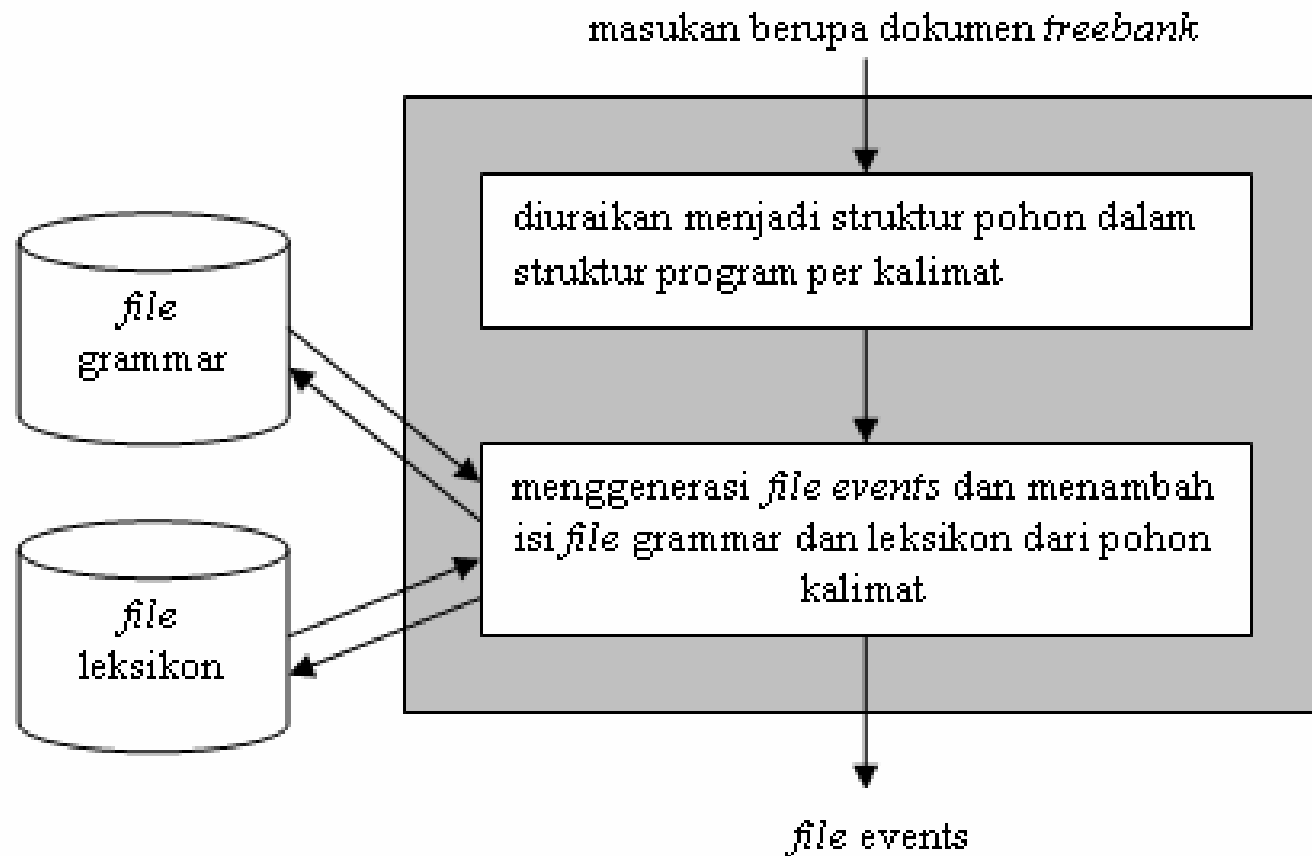
Analisis Proses

- Pemrosesan Awal (*preprocessing*)
- Penentuan Jenis Kata (*POS tagging*)

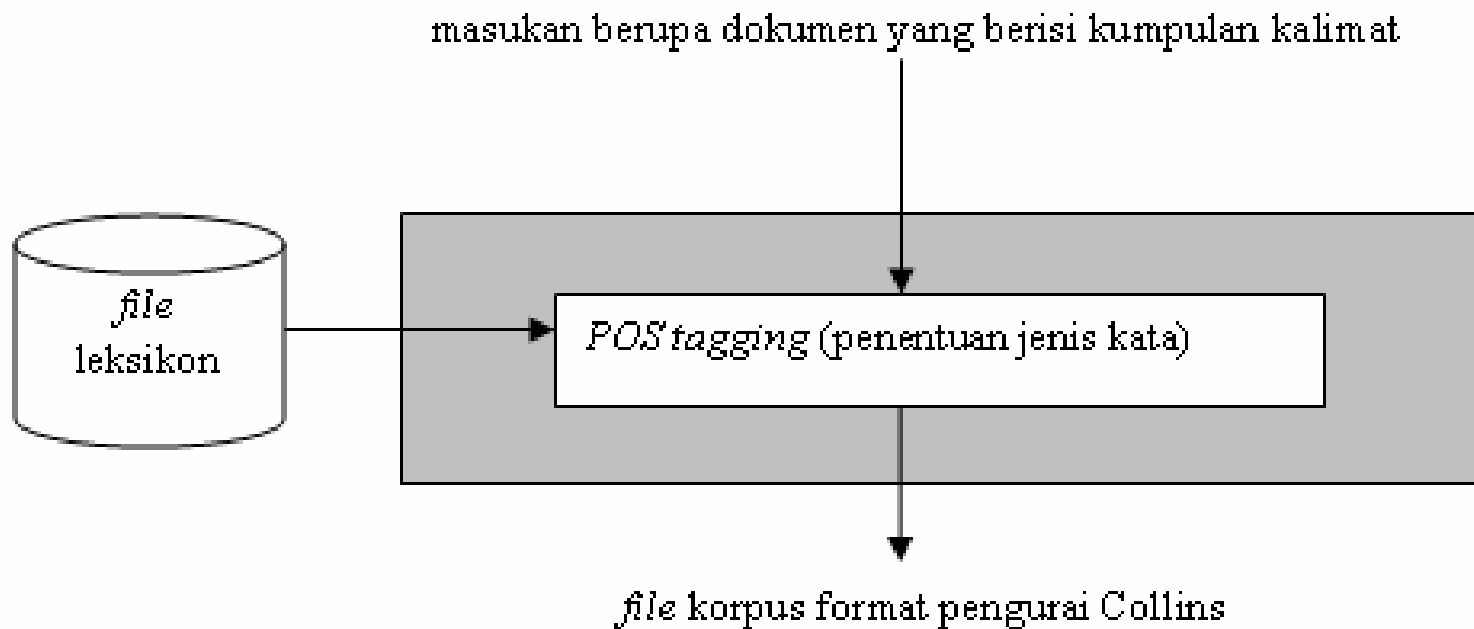
Pengurai Collins



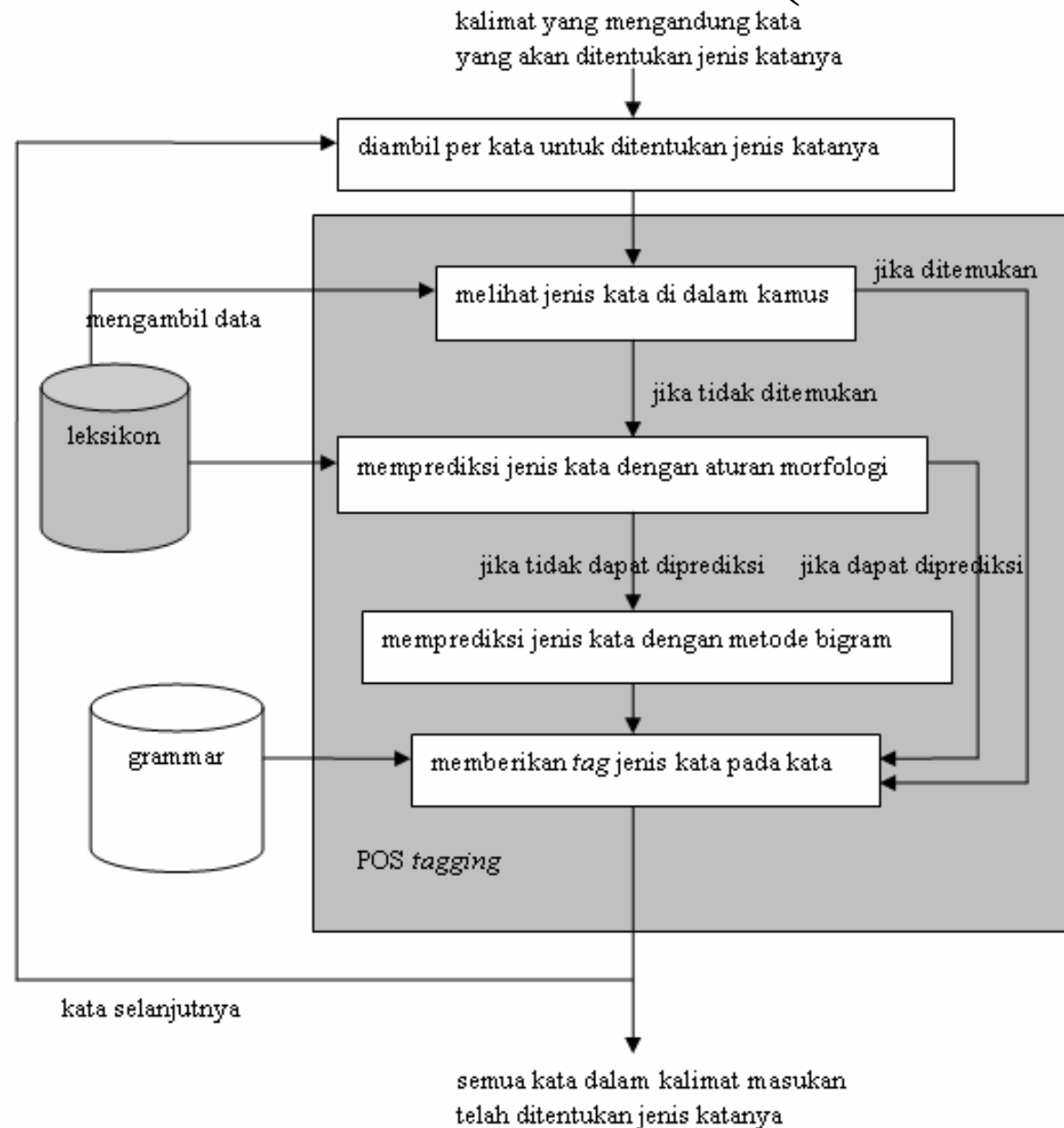
Pemrosesan Awal [1]



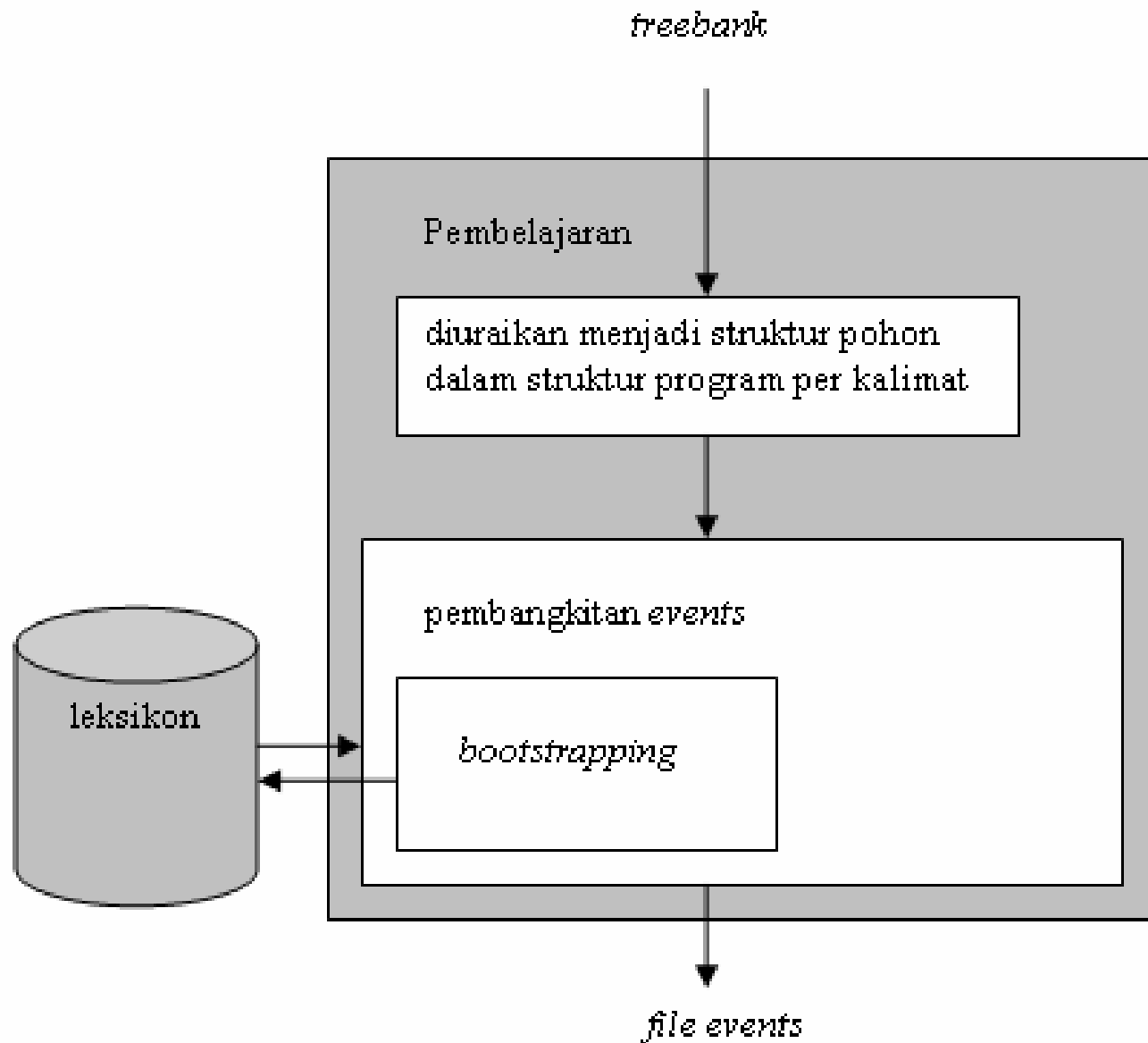
Pemrosesan Awal [2]



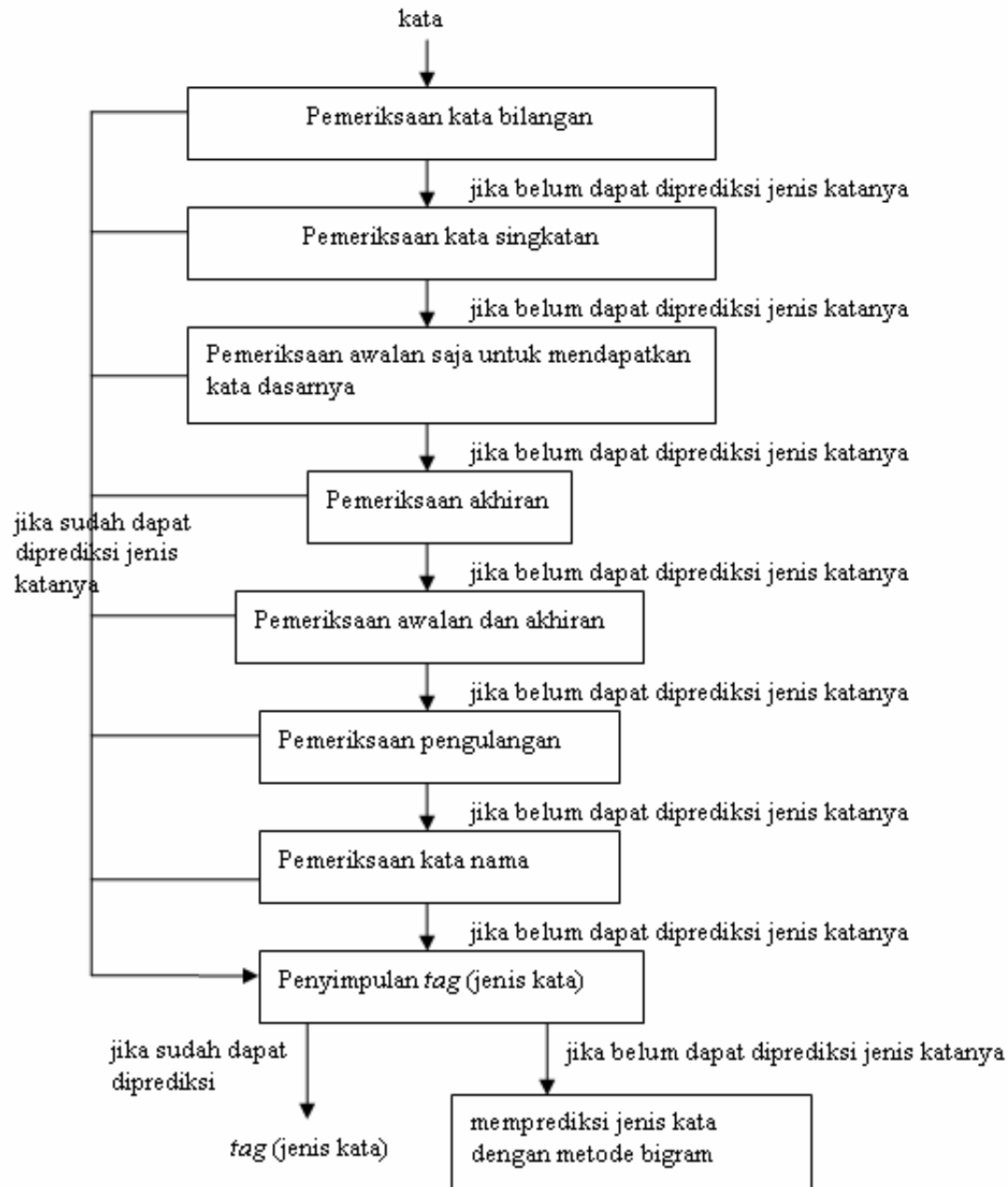
Penentuan Jenis Kata (POS Tagging)



Memperkaya Kamus



Proses Pemeriksaan Morfologi



Prediksi Jenis Kata dengan Bigram

kalimat yang mengandung kata yang tidak bisa diprediksi kelas katanya

periksa kata di depan dan di belakang kata yang tidak dapat diprediksi kelas katanya

cari pola tata bahasa/aturan *grammar* yang mengandung simbol jenis kata di depan kata yang akan diprediksi jenis katanya

cari pola tata bahasa/aturan *grammar* yang mengandung simbol jenis kata di belakang kata yang akan diprediksi jenis katanya

memberikan *tag* pada kata yang tidak diketahui jenis katanya berdasarkan aturan *grammar* yang memiliki probabilitas terbesar

kelas kata/*tag*

Tujuan Pengujian

- Pengujian dilakukan untuk mengetahui kinerja pengurai Collins jika digunakan untuk bahasa Indonesia dengan menggunakan kumpulan *file* masukan hasil pemrosesan awal (*preprocessing*) dari tesis ini.

Perancangan Pengujian [1]

- membagi dua kelompok
 - kelompok pertama
 - *treebank* dengan 42 pohon kalimat sederhana
 - korpus dengan 7 kalimat sederhana yang mirip dengan aturan *grammar* pada *treebank*
 - kelompok kedua
 - *treebank* dengan 190 pohon kalimat sederhana dan kompleks
 - korpus dengan 15 kalimat kompleks

Perancangan Pengujian [2]

- pada kedua kelompok dilakukan hal-hal berikut:
 - memasukkan *file treebank* yang berisi pohon kalimat untuk menguji proses generasi *file events*, penambahan grammar, dan leksikon,
 - memasukkan *file* kalimat berbahasa Indonesia untuk mempersiapkan *file* korpus,
 - diujikan pada pengurai Collins

Hasil Pengujian [1]

- Kelompok pertama
 - dari 7 kalimat, dapat diuraikan semua kalimat
 - 6 kalimat berhasil diuraikan dengan benar
 - 1 kalimat memiliki bagian yang kurang tepat diuraikan
 - Kurang tepat diuraikan karena aturan *grammar* yang tepat tidak terdapat pada *treebank*

Hasil Pengujian [2]

- Kelompok kedua
 - Dari 15 kalimat, yang dapat diuraikan 8 kalimat
 - Kalimat yang tidak dapat diuraikan hasil probabilitas kalimatnya adalah 0.
 - Justifikasi nilai 0 diambil oleh Collins dalam pengurainya karena kalimat tidak berhasil diuraikan keseluruhan (terputus di tengah), karena ada bagian pohon yang tidak cocok dengan satupun *grammar* di *file grammar*

Hasil Pengujian [3]

- Kalimat yang berhasil diuraikan tidak ada yang benar hasil penguraiannya secara kesatuan kalimat.
- Beberapa hasil penguraian bagian kalimat ada yang benar, tapi juga ada yang kurang tepat
- Bobot terbesar didapat dari hasil penguraian kalimat ke 15 dengan bobot 173.707.
- Bobot terkecil dari kalimat yang berhasil diuraikan didapat dari kalimat ke 13 dengan bobot 21.83.
- Semakin besar bobot pohon yang dihasilkan tidak mencerminkan semakin benar penguraian kalimat secara kesatuan kalimat

Kesimpulan

- Pengurai Collins dapat digunakan untuk bahasa Indonesia dengan melakukan adaptasi kumpulan file masukan pengurai Collins.
- Kinerja pengurai Collins untuk bahasa Indonesia dengan terbatasnya *treebank*:
 - dapat menguraikan seluruh kalimat pada kelompok pengujian pertama
 - dapat menguraikan 8 kalimat dari 15 kalimat dari kelompok pengujian kedua.
 - Hasil penguraian juga telah mampu merepresentasikan pohon kalimat berbahasa Indonesia walau tidak semua bagian kalimat benar diuraikan.

Saran [1]

- Permasalahan paling mendasar dari tesis ini
→ keterbatasan *treebank* berbahasa Indonesia
 - Dibuat sebuah perkumpulan atau *consortium* untuk bersama-sama berkontribusi membuat *treebank* berbahasa Indonesia
 - Melakukan translasi Penn WSJ Treebank menjadi bahasa Indonesia, tapi tetap perlu ada pembenahan pola tata bahasa karena beberapa aturan dalam bahasa Inggris tidak berlaku pada bahasa Indonesia
 - Membuat sebuah permainan → “Are You Really Indonesian Citizen”

Saran [2]

- Perlu dibuat kamus/leksikon yang valid
- Bagaimana membuat format yang tepat untuk kamus/leksikon (satu kata mempunyai dua tag)
- Perlu bimbingan ahli bahasa untuk membuat *treebank*
- Pemilihan kalimat yang tepat (tidak acak)
- Tesis ini tidak melakukan pengujian mengenai penanganan ambiguitas pohon pola tata bahasa untuk setiap kalimat
- Model 2 dan Model 3 perlu dibuat *treebank*-nya agar dapat diujicobakan untuk bahasa Indonesia.

Daftar Pustaka [1]

- [1] _____, 2008, *Wikipedia, The Free Encyclopedia : N-gram*, [html], (<http://en.wikipedia.org/wiki/N-gram>, diakses tanggal 6 November 2008)
- [2] Adriani, Mirna dkk. (2007) : Stemming Indonesian: A Confix-Stripping Approach, *ACM Transaction on Asian Language Information Processing*, **6**, 13:1 – 13:33.
- [3] Asian, Jelita (2007) : Effective Techniques for Indonesian Text Retrieval, Tesis Program Master of Science, School of Computer Science and Information Technology, 9-204.
- [4] Aziz, Mohd Juzaidin Ab dkk. (2006) : Pola Grammar Technique for Grammatical Relation Extraction Malay Language, *Malaysian Journal of Computer Science*, **19**, 59-72.
- [5] Bikel, Daniel M. (2004) : *On The Parameter Space of Generative Lexicalized Statistical Parsing Models*, Disertasi program Doctor of Philosophy, University of Pennsylvania, [halaman].
- [6] Blum, Avrim dan Tom Mitchell (1998) : Combining Labeled and Unlabeled Data With Co-training, *Proceedings of the 11th Annual Conference on Computational Learning Theory*, 92-100.
- [7] Charniak, Eugene. (1993) : *Statistical Language Learning*, Massachusetts Institute of Technology.
- [8] Charniak, Eugene. (1997) : Statistical Parsing with a Context-free Grammar and Word Statistics, *American Association for Artificial Intelligence: AAAI Press*.
- [9] Charniak, Eugene. (2000) : A Maximum-Entropy-Inspired Parser, *Proceedings of NAACL-2000*.
- [10] Clark, Stephen (2003) : Bootstrapping POS Taggers Using Unlabelled Data, *Proceedings of the Seventh CoNLL Conference*.

Daftar Pustaka [2]

- [11] Carroll, Glen. (1995) : *Learning Probabilistic Grammars for Language Modeling*, Tesis program Master of Science, Brown University.
- [12] Collins, Michael. (1996) : A New Statistical Parser Based on Bigram Lexical Dependencies, *In Proceedings of the 34th Annual Meeting of ACL*, 184-191.
- [13] Collins, Michael. (1999) : *Head-Driven Statistical Models for Natural Language Parsing*, Disertasi program Doctor of Philosophy, University of Pennsylvania.
- [14] Collins, Michael, Jan Hajic, Lance Ramshaw, Cristoph Tillmann (1999) : A Statistical Parser for Czech, *Proceedings of the 37th Annual Meeting of the ACL*.
- [15] Gusmita, Ria Hari & Ruli Manurung (2008) Some initial experiments with Indonesian probabilistic parsing. Second MALINDO Workshop. 1-5.
- [16] Iskak Hendrawan, (1999) : *Pengurai Sintaks Kalimat untuk Bahasa Indonesia dengan Metode Linguistic String Analysis*; Fakultas Ilmu Komputer Universitas Indonesia, Depok 1999
- [17] Johnson, Mark (1998) : PCFG Models of Linguistic Tree Representations, *Association for Computational Linguistics*, **24**, 613-632.
- [18] Jurafsky, Daniel dan Martin, James H. (2000) : *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall, New Jersey, 443-471.
- [19] Lefuel, Ramon dan Brian J. Ross (2004) : Parsing Probabilistic Context Free Languages with Multiple-Objective Genetic Algorithms. *Technical Report*. Brock University.
- [20] Linz, Peter. (2001) : *An Introduction to Formal Languages and Automata*, Jones and Bartlett Publisher, Inc, Massachusetts, 126-148.

Daftar Pustaka [3]

- [21] Marcus, Mitchell P. dkk (1992) : Building a Large Annotated Corpus of English: The Penn Treebank. Departmet of Computer and Information Science University of Pennsylvania.
- [22] Mulyono, Anton M. (1991) : Tata Bahasa Baku Bahasa Indonesia, Balai Pustaka, Jakarta
- [23] Pisceldo, Femphy dkk. (2008) : A Two-Level Morphological Analyser for Indonesian Language, *Australian Language Technology Association (ALTA) Workshop*.
- [24] Schabes, Yves dan Waters, Richard C. (1993) : Stochastic Lexicalized Context-Free Grammar, *International Workshop on Parsing Technology*.
- [25] Shavitri, Shelly. (1999) : *Analisa Struktur Kalimat Bahasa Indonesia dengan Menggunakan Pengurai Kalimat Berbasis Linguistic String Analysis*. Tugas Akhir untuk Sarjana Ilmu Komputer, Universitas Indonesia.
- [26] Sutarmo, Sarwiji Suwandi (2008) : Bahasa Indonesia: Bahasa Kebanggaanku : untuk SMP dan MTs kelas VII. Pusat Perbukuan Departemen Pendidikan Nasional.
- [27] Tala, Fadillah Z. (2003) : *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*, Tesis program Master of Logic, Institute for Logic, Language, and Computation Netherland.



Sekian dan Terima Kasih

Mohon kritik dan sarannya.....

<http://www.gangsir.com>

Simbol Non-terminal Jenis Kata 1

Simbol	Jenis Kata	Keterangan	Contoh
JJ	Adjektiva	Kata sifat; kata yang memberi penjelasan tentang suatu benda	cantik, baik, buruk
RB	Adverbia	Kata keterangan	nanti , sekarang
AR	Artikula	Kata sandang	si, sang
CC	Konjungtor Koordinatif	Kata hubung yang menghubungkan klausa pada kalimat majemuk setara.	dan, lalu
CS	Konjungtor Subordinatif	Kata hubung pada kalimat majemuk bertingkat	ketika, walaupun
MD	Modal	Kata Keterangan Modalitas	boleh
PR	Pronomina	Kata ganti; kata yang dipakai untuk menggantikan kata atau yang dibendakan	saya, itu
WH	Kata Tanya	Kata yang digunakan untuk menanyakan sesuatu	siapakah, bagaimanakah
NN	Nomina	Kata benda; kata yang menyebut benda atau yang dibendakan	buku, meja, orang
CD	Numeralia	Kata bilangan; kata yang menyatakan jumlah benda atau jumlah kumpulan atau jurusan dari nama-nama benda	seribu

Simbol Non-terminal Jenis Kata 2

Simbol	Jenis Kata	Keterangan	Contoh
IN	Preposisi	Kata depan; kata yang merangkaikan kata-kata atau bagian-bagian kalimat	di, ke, dari
UH	Interjeksi	Kata seru	ai, ah, ceile
RP	Partikel	Kata tugas partikel	pun, per
VB	Verba	Kata kerja; kata yang bermakna melakukan aktivitas atau kegiatan, atau lebih jelas kalau dikatakan melakukan pekerjaan	mencoba, lempar, menari
AUX	Kata bantu	Kata bantu	akan, dapat
FW	Kata asing	Kata asing	<i>download, notebook</i>
PU	Tanda baca	Tanda baca	., , , ; (,) , “ , ‘ , ” , ’
SYM	Simbol matematika	Simbol matematika	+, #, \$,
X	<i>unknown</i>	Kata yang tidak dapat diprediksi jenis katanya	

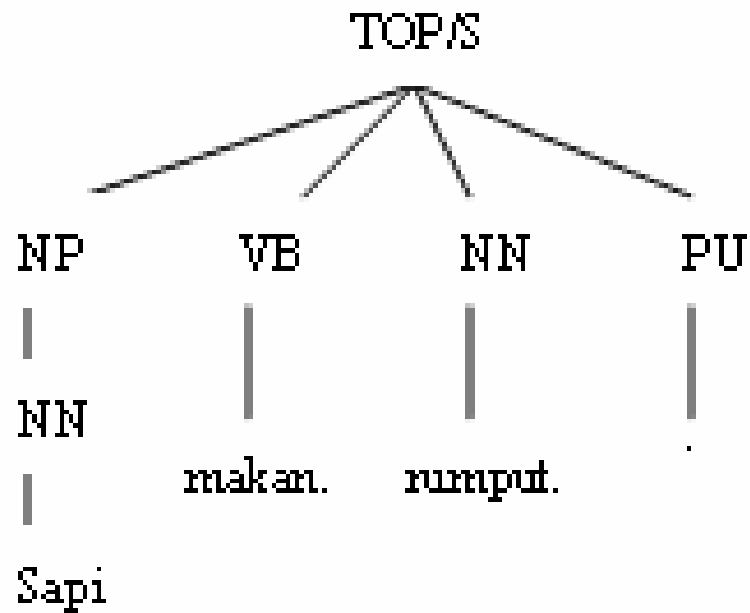
Simbol Non-Terminal Level Frase

Simbol	Keterangan
S	Kalimat
ADJP	frase yang menyatakan sifat (frase adjektiva)
ADVP	frase yang menyatakan keterangan (frase adverbial)
NP	frase yang menyatakan benda (frase nomina)
SBAR	sub kalimat majemuk
SBARQ	sub kalimat setelah kata tanya
VP	frase yang menyatakan kerja (frase verba)

Contoh pohon kalimat

```
PROB 34 -134.275 0
TOP NP -134.275 NN 0 Sapi
  VB 0 makan
  NN 0 rumput
  PU 0 .
(TOP~(NP~rumput~4~3 Sapi/NN makan/VB rumput/NN ./PU ) )
TIME 0
```

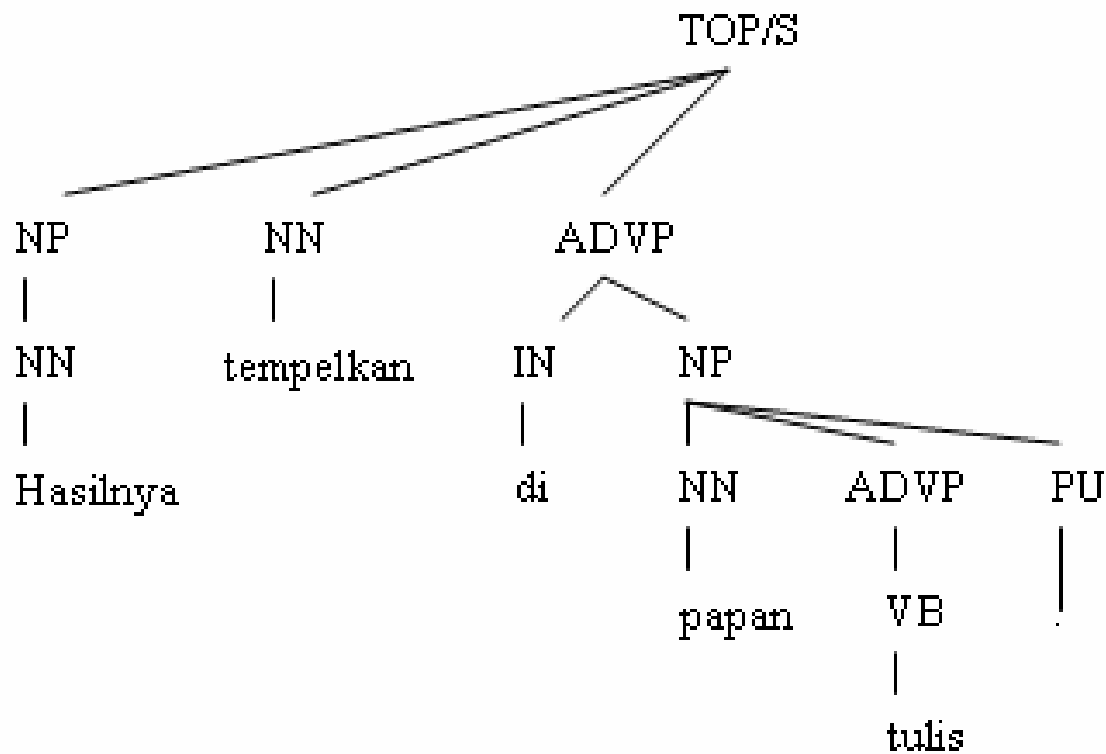
Contoh pohon kalimat



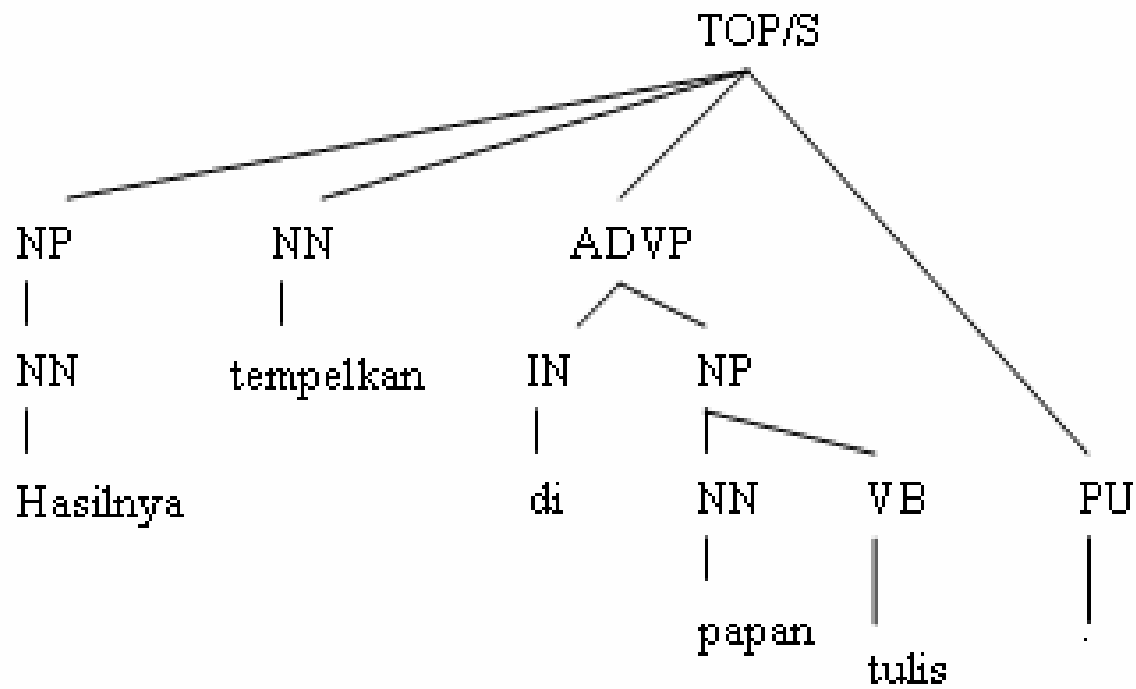
Contoh Pohon: Kalimat 13 (Hasil)

```
PROB 89 -21.83 0
TOP NP -21.83 NN 0 Hasilnya
  NN 0 tempelkan
  ADVP -16.0802 IN 0 di
    NP -13.6505 NN 0 papan
      ADVP -4.89085 VB 0 tulis
        PU 0 .
(TOP~(NP~Hasilnya~3~1 Hasilnya/NN tempelkan/NN (ADVP~di~2~1 di/IN
(NP~papan~3~1 papan/NN (ADVP~tulis~1~1 tulis/VB ) ./PU ) ) ) )
TIME 0
```

Contoh Pohon: Kalimat 13 (Hasil)



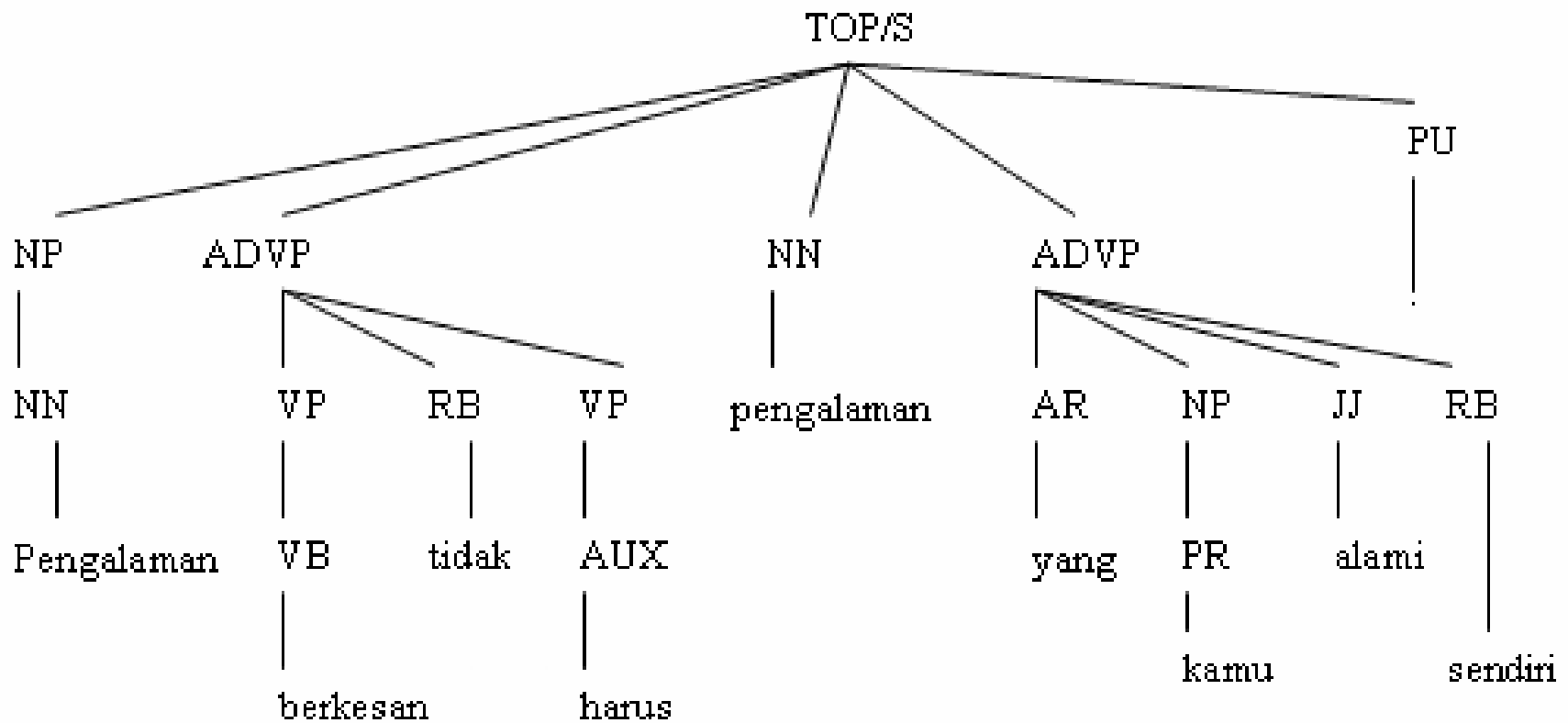
Contoh Pohon: Kalimat 13 (Seharusnya)



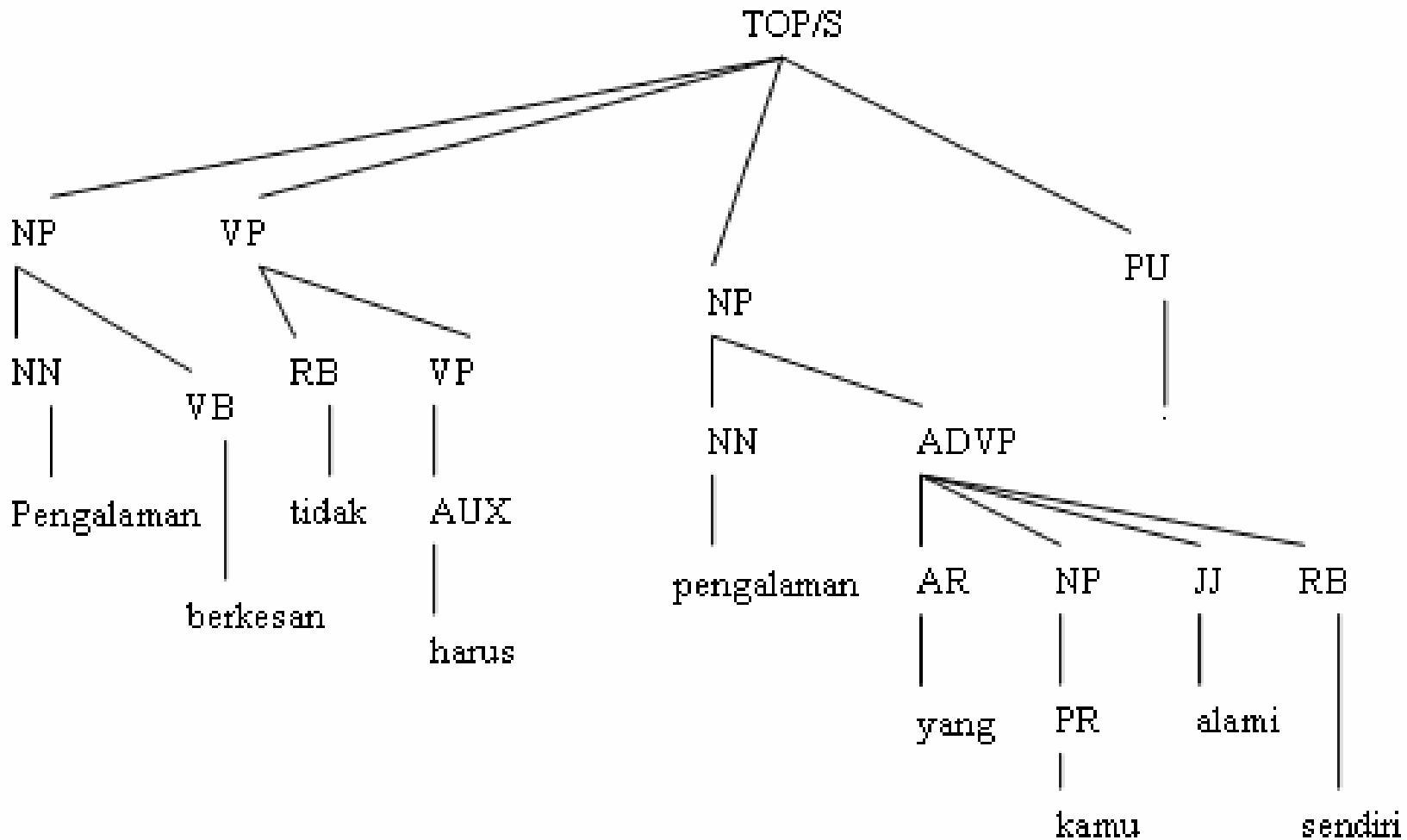
Contoh Pohon: Kalimat 15 (Hasil)

```
PROB 195 -173.707 0
TOP NP -173.707 NN 0 Pengalaman
  ADVP -61.349 VP -53.0975 VB 0 berkesan
    RB 0 tidak
      VP -2.86092 AUX 0 harus
        NN 0 pengalaman
          ADVP -95.8387 AR 0 yang
            NP -0.0122658 PR 0 kamu
              JJ 0 alami
                RB 0 sendiri
                  PU 0 .
                    (TOP~(NP~Pengalaman~5~1 Pengalaman/NN (ADVP~berkesan~1~1
                    (VP~berkesan~3~1 berkesan/VB tidak/RB (VP~harus~1~1 harus/AUX ) )
                    ) pengalaman/NN (ADVP~alami~4~3 yang/AR (NP~kamu~1~1 kamu/PR )
                    alami/JJ sendiri/RB ) ./PU ) )
TIME 0
```

Contoh Pohon: Kalimat 15 (Hasil)



Contoh Pohon: Kalimat 15 (Seharusnya)



File Korpus

21 Benar JJ , PU Malin NN Kundang NN berisi VB cerita NN tentang CC anak NN yang AR durhaka JJ kepada IN ibunya NN sehingga CS ia PR dikutuk VB oleh IN Tuhan NN dan CC menjadi VB batu NN . PU

18 Tema NN cerita NN Malin NN Kundang NN dari IN Sumatra NN Barat NN ini PR ternyata VB juga JJ bisa AUX ditemui VB di IN daerah NN lain JJ di IN Indonesia NN . PU

8 Apakah WH daerahmu NN juga JJ ada VB cerita NN semacam VB ini PR ? PU

9 Kalau CS ada VB , PU tulislah VB dengan IN singkat JJ cerita NN itu PR ! PU

12 Setelah RB kamu PR dengarkan VB dongeng NN tadi RB , PU buatlah VB pertanyaan-pertanyaan NN tentang CC isi NN dongeng NN . PU

.....

File Events

```
.....  
6 7 Kenapa WH kamu PR berpikir VB John NN akan IN pergi VB . PU  
3 Kenapa WH S PR 00000 00000  
2 #STOP# #STOP# Kenapa WH #STOP# S PR 000000 110 0 0  
2 berpikir VB Kenapa WH SBARQ S PR 000000 010 0 0  
2 . PU Kenapa WH PU S PR 000000 000 0 0  
2 #STOP# #STOP# Kenapa WH #STOP# S PR 000000 000 0 0  
3 berpikir VB SBARQ VB 00000 00000  
2 kamu PR berpikir VB NP SBARQ VB 000000 110 0 0  
2 #STOP# #STOP# berpikir VB #STOP# SBARQ VB 000000 100 0 0  
2 John NN berpikir VB NN SBARQ VB 000000 010 0 0  
2 pergi VB berpikir VB ADVP SBARQ VB 000000 000 0 0  
2 #STOP# #STOP# berpikir VB #STOP# SBARQ VB 000000 000 0 0  
3 kamu PR NP PR 00000 00000  
2 #STOP# #STOP# kamu PR #STOP# NP PR 000000 110 0 0  
2 #STOP# #STOP# kamu PR #STOP# NP PR 000000 010 0 0  
3 pergi VB ADVP VB 00000 00000  
2 akan IN pergi VB AUX ADVP VB 000000 110 0 0  
2 #STOP# #STOP# pergi VB #STOP# ADVP VB 000000 100 0 0  
2 #STOP# #STOP# pergi VB #STOP# ADVP VB 000000 010 0 0  
.....
```

File Treebank

.....

(S (NP (PR saya)) (>VB mencoba) (ADVP (IN untuk) (>VB pergi)) (PU .))

(S (NP (PR saya)) (>VP (VB ingin) (>VB menjadi)) (JJ pandai) (PU .))

(S (NP (PR saya)) (>VP (JJ percaya)) (ADVP (IN akan) (>VB menjadi) (JJ pandai)) (PU .))

(S (NN John) (>VB dipaksa) (ADVP (IN untuk) (>VB pergi)) (PU .))

(S (NN Yohanes) (>VB dipercaya) (ADVP (IN untuk) (>VB menjadi) (JJ pandai)) (PU .))

(S (NN John) (>VB diinginkan) (ADVP (IN untuk) (>VB pergi)) (PU .))

(S (NN John) (ADJP (JJ lebih) (JJ suka)) (>VB parkir) (ADVP (IN di) (RB sini)) (PU .))

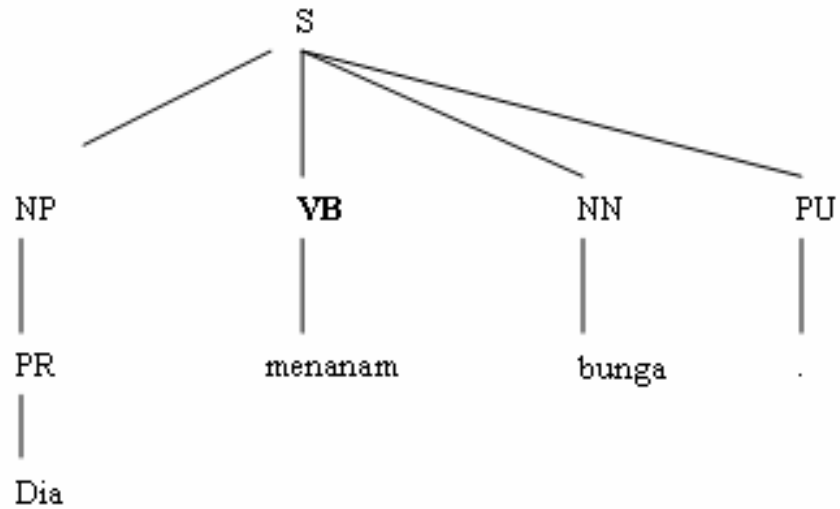
.....

File Leksikon

.....
, PU 1
. PU 1
: PU 0
; PU 0
= SYM 0
? PU 0
@ SYM 0
\ PU 0
aba-aba NN 0
abad NN 0
abadi JJ 0
abah NN 0
abah-abah NN 0
abakus NN 0
abang NN 0
abangan NN 0
abar NN 0
abatoar NN 0
abdas NN 0
.....

<http://www.gangsir.com>

File Grammar



L S NP VB
R S VB NN
R S NN PU
U NP PR
X NP PR 00000
Y NP PR 00000

Chart Parsing

Dia	menanam	bunga
PR	VB	NN
NP		

edge:

$S \rightarrow PR$

$S \rightarrow VB$

$S \rightarrow NN$

$S \rightarrow NP$

$NP \rightarrow PR$

<http://www.gangsir.com>