

# *Bahasa Alami 1*

Peranan Pengetahuan dalam Bahasa  
Masalah Bahasa Alami  
Proses Sintaksis  
Grammar dan Parser  
Automated Transition Network

Referensi

Luger & Stubblefield : bab-3

Rich & Knight : bab 15

# Peranan Pengetahuan dalam Bahasa

- Komunikasi dengan bahasa alami (teks/ucapan) bergantung pada pengetahuan dari domain pembicaraan
- Pemahaman bahasa tidak hanya ditransmisi dari kata2; tapi butuh inferensi tentang tujuan dan asumsi-asumsi dari pembicara dan tentang konteks interaksi
- Implementasi dari program pemahaman bahasa alami membutuhkan representasi sejumlah besar pengetahuan dan alasan2.

# Masalah Bahasa Alami

- Bahasa [1] merupakan fenomena yg kompleks, melibatkan berbagai proses, mis: pengenalan suara atau huruf tercetak, parsing sintak, inferensi semantik tingkat tinggi & komunikasi isi emosional melalui ritme & inflection
- Masalah pemrosesan bahasa ini secara garis besar di bagi dua: [2]
  1. Pemrosesan teks yg ditulis, menggunakan leksikal, sintaksis, & pengetahuan semantik dari bahasa
  2. Pemrosesan bahasa bicara, menggunakan semua informasi di atas, ditambah informasi mengenai *phonology*

# Masalah Bahasa Alami (lanjutan)

- Untuk mengatasi kompleksitas dari bahasa ini didefinisikan tingkat analisis bahasa alami : [1]
  1. *Prosody*, mengerjakan ritme & intonasi bahasa
  2. *Phonology* menguji suara yg dikombinasikan untuk membentuk bahasa
  3. *Morphology* mengenai komponen2( morfem2) yg membentuk kata2  
Termasuk aturan2 pengembangan formasi kata2 seperti efek dari prefiks (un-, non-, anti-, dll), & sufiks (-ing, -ly, dll) yg memodifikasi arti dari akar kata

# Masalah Bahasa Alami (lanjutan)

4. *Syntax* mempelajari aturan2 kombinasi kata2 ke dalam perkataan dan kalimat serta penggunaan aturan2 tsb untuk dianalisis & menghasilkan kalimat
5. *Semantics* berhubungan dengan arti dari kata, perkataan & kalimat serta cara dimana artinya terdapat pada bahasa alami
6. *Pragmatics* merupakan studi pada cara di mana bahasa digunakan & efeknya pada si pendengar
7. *World knowledge* termasuk pengetahuan dunia nyata secara fisik, dunia interaksi sosial & manusia, serta peranan, tujuan & maksud dari komunikasi

# Pemrosesan Bahasa Alami

- Menurut Rich [2], pemrosesan bahasa alami dapat dibagi atas beberapa proses:
  1. *Morphological analysis*, dimana kata secara individu dianalisis ke dalam komponennya dan yang bukan kata (sep. tanda baca) dipisahkan dari kata
  2. *Syntactic analysis*, urutan linier dari kata ditransformasikan ke dalam struktur yang menunjukkan bagaimana kata saling berhubungan. Urutan kata yang melanggar aturan bahasa dalam pengkombinasian kata, akan ditolak
  3. *Semantic analysis*, dibuat mapping antara struktur sintaksis & objek dalam domain kerja. Struktur dimana tidak ada mapping yang memungkinkan, akan ditolak

# Pemrosesan Bahasa Alami (lanjutan)

4. *Discourse integration*, dimana arti dari kalimat secara individu bergantung pada kalimat2 yg mendahului & mungkin dapat mempengaruhi arti dari kalimat2 yang mengikutinya
5. *Pragmatic analysis*, dimana struktur yang merepresentasikan apa yang diucapkan, diinterpretasikan kembali untuk menentukan apa arti sebenarnya.

# Proses Sintaksis

- Proses sintaksis [2] atau disebut juga *parsing*, merupakan langkah dimana kalimat masukan secara flat diubah ke dalam struktur hirarkis yang berhubungan dengan unit2 arti dalam kalimat
- Proses sintaksis sangat berperan dalam pemahaman bahasa alami dikarenakan
  - Proses semantik harus beroperasi pada pilihan2 kalimat. Jika tidak ada tahap parsing sintaksis maka sistem semantik harus memutuskan atas pilihannya sendiri. Jika parsing dilakukan, hal ini akan membatasi pilihan2 dari semantik.
  - Tidak selalu dimungkinkan untuk mengekstrak arti kalimat tanpa menggunakan fakta gramatikal



# Proses Sintaksis (lanjutan)

- Semua sistem menggunakan dua komponen :
  1. Representasi deklaratif, disebut *grammar*, fakta sintaksis tentang bahasa
  2. Prosedur, disebut *parser*, dimana membandingkan *grammar* dengan kalimat2 masukan untuk membentuk struktur yang dianalisis

# Grammar & Parser

- Grammar direpresentasikan sebagai sekumpulan aturan-aturan produksi, seperti pada gambar 1.
- Aturan pertama, 'Sebuah kalimat terdiri atas kata benda (NP) diikuti kata kerja (VP)'.
- $I$  menyatakan 'atau/or',
- $\varepsilon$  menyatakan string kosong
- Simbol2 yg diperluas oleh aturan-aturan disebut *nonterminal symbols*, sedangkan simbol2 yg berhubungan langsung dgn string yg ditemukan pada kalimat masukan disebut *terminal symbols*

# Gambar 1. Grammar Untuk Bahasa Inggris

S → NP VP  
NP → the NP1  
NP → PRO  
NP → PN  
NP → NP1  
NP1 → ADJS N  
ADJS → ε | ADJ ADJS  
VP → V  
VP → V NP  
N → file | printer  
PN → Bill  
PRO → I  
ADJ → short | long | fast  
V → printed | created | want

# Top-Down vs Bottom-Up Parsing

- Untuk menganalisis atau mengupas sebuah kalimat, ada 2 cara :

Menggunakan pohon analisis seperti pada gambar 2

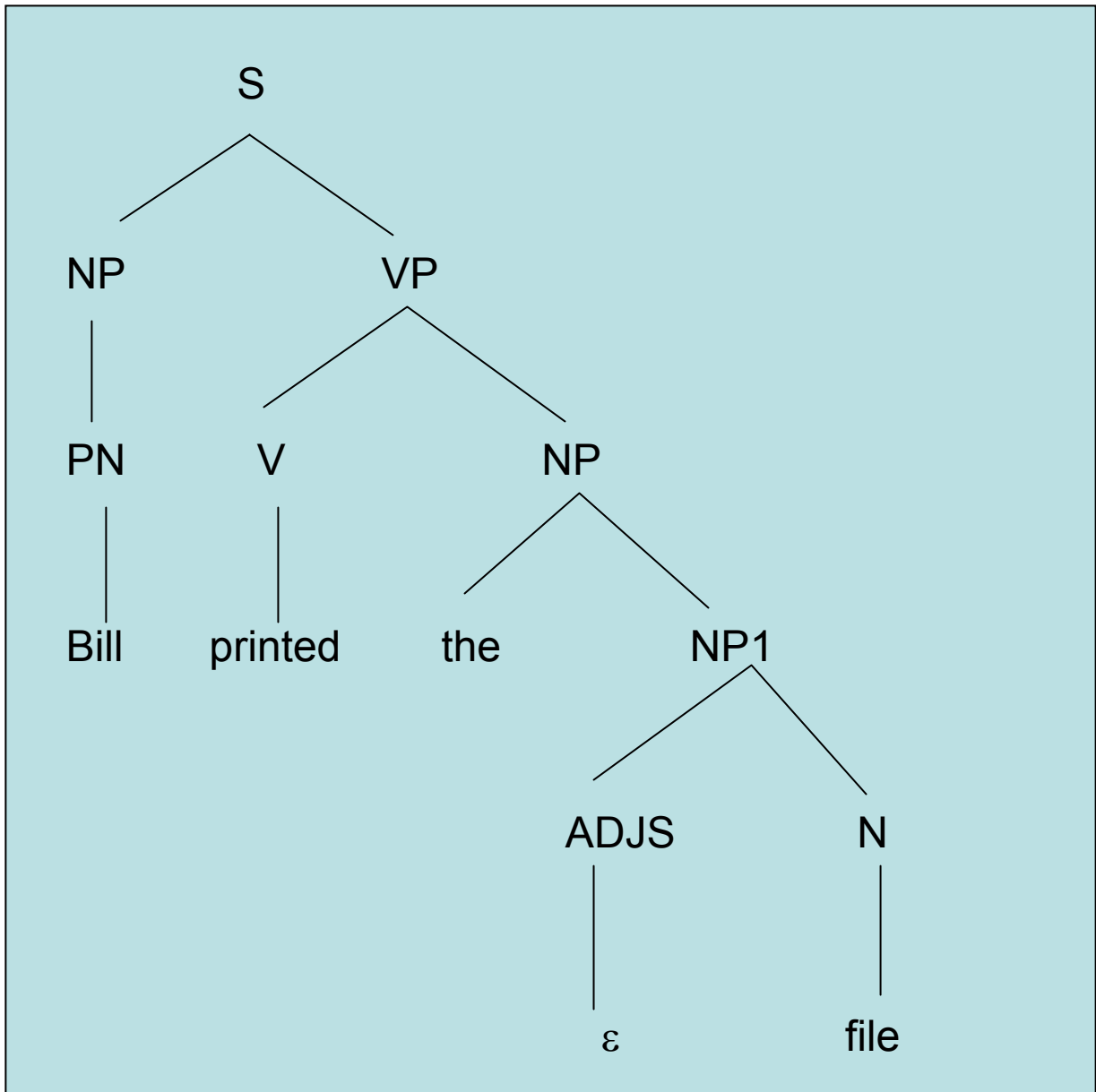
## 1. *Top-Down Parsing*

Dimulai dari simbol 'mulai' dan diaplikasikan aturan grammar sampai simbol pada terminal pohon berhubungan dengan komponen2 kalimat yang dianalisis

## 2. *Bottom-Up Parsing*

Dimulai dari kalimat yang dianalisis dan diaplikasikan aturan grammar secara backward sampai pohon tunggal dimana terminalnya adalah kata-kata dari kalimat dan titik paling atas terbentuk simbol mulai

## Gambar 2. Pohon Analisis Untuk Sebuah Kalimat



# Automated Transition Networks (ATN)

- *Automated Transition Network (ATN)* merupakan prosedur top-down parsing dimana bermacam-macam pengetahuan digabungkan ke sistem analisis sehingga dapat beroperasi lebih efisien.
- ATN (seperti pada gambar 3) menyerupai *finite-state machine* dimana klas label digabungkan pada garis berarah yang mendefinisikan transisi antara suatu keadaan.
- Garis berarah ini dapat dilabeli (dalam bahasa Inggris) dengan

## ATN (lanjutan)

- Kata2 spesifik, seperti 'in'
- Kategori2 kata, seperti 'noun'
- Didorong ke jaringan lain untuk mengenali komponen lain pada suatu kalimat, seperti jaringan untuk mengenali prepositional phrase (PP) maka pada garis berarah ditanyakan noun phrase (NP)
- Prosedur yang menjalankan tes pada input saat itu dan pada komponen kalimat yang telah diidentifikasi
- Prosedur2 yang membangun struktur yang akan membentuk bagian akhir dari analisis

# Gambar 3. Contoh ATN

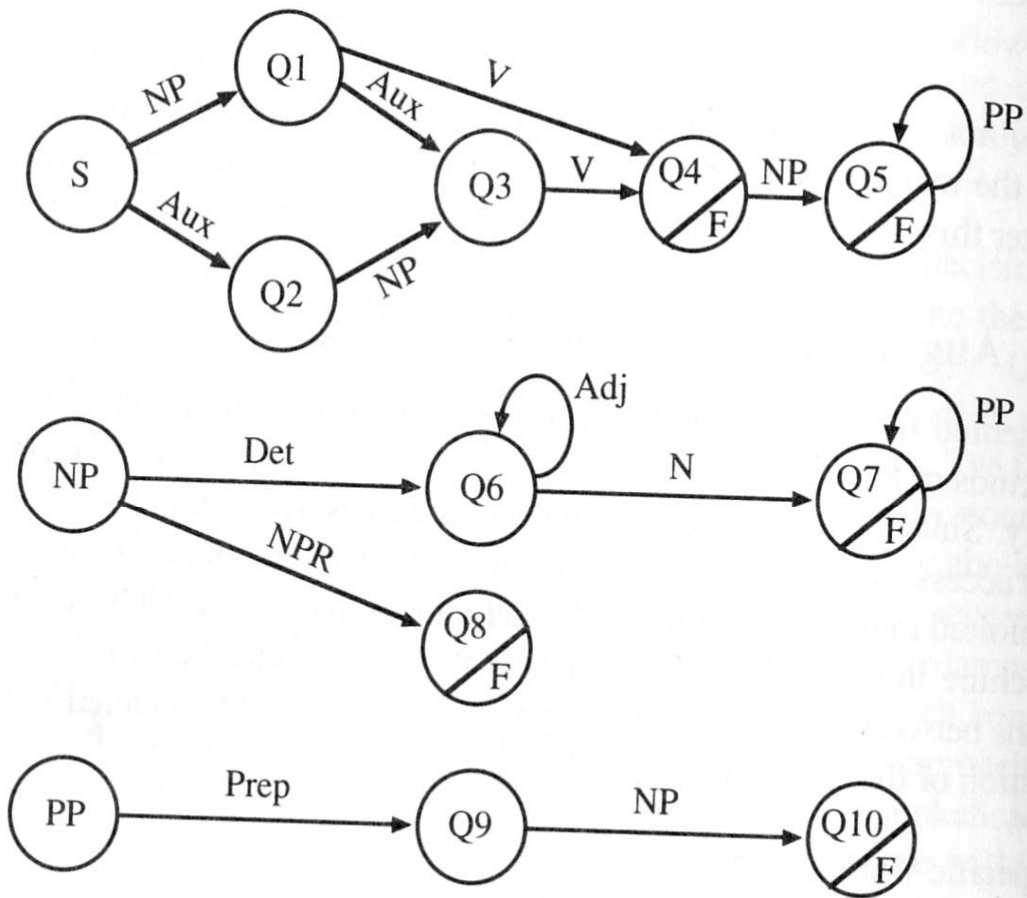


Figure 15.8: An ATN Network for a Fragment of English



## ATN (lanjutan)

- Cara kerja ATN pada kalimat:

‘The long file has printed’

1. Mulai pada keadaan S
2. Masuk ke NP
3. Kerjakan tes kategori untuk melihat jika ‘the’ merupakan *determiner*
4. Tes sukses, maka register DETERMINER diset ke DEFINITE lalu ke keadaan Q6
5. Kerjakan tes kategori untuk melihat jika ‘long’ merupakan *adjective*
6. Tes sukses, maka register ADJS ditambahkan ‘long’. Tetap tinggal di Q6
7. Kerjakan tes kategori untuk melihat jika ‘file’ merupakan *adjective*. Tes gagal
8. Kerjakan tes kategori untuk melihat jika ‘file’ merupakan *noun*. Tes sukses, register NOUN diset ke ‘file dan lalu ke keadaan Q7

## ATN (lanjutan)

9. Masuk ke PP
10. Kerjakan tes kategori untuk melihat 'has' merupakan *preposition*. Tes gagal, sehingga keluar
11. Tidak ada lagi yang dikerjakan di Q7, lalu keluar dan kembali ke struktur (NP(FILE(LONG)DEFINITE)), sehingga kembali ke keadaan Q1, dengan register SUBJ diset ke struktur yang dikembalikan dan register TYPE diset ke DCL
12. Kerjakan tes kategori untuk melihat jika 'has' merupakan verb. Tes sukses, sehingga register AUX diset ke NIL dan register V diset ke 'has'. Lanjut ke Q4
13. Masuk ke keadaan NP, dikrnkan kata 'printed' bukan *determiner* ataupun *noun* murni, maka NP dikeluarkan
14. Masih tersisa input, berarti analisis belum selesai. Lalu diperlukan *backtracking*.

## ATN (lanjutan)

15. Pilihan terakhir terdapat pada Q1, sehingga kembali ke Q1. Register AUX dan V tidak diset
16. Kerjakan tes kategori untuk melihat jika 'has' merupakan *auxiliary*. Tes sukses sehingga 'has' diset ke register AUX dan lalu ke Q3
17. Kerjakan tes kategori untuk melihat jika 'printed' merupakan *verb*. Tes sukses, maka register V diset ke 'printed'. Lanjut ke Q4.
18. Karena input telah habis, maka Q4 merupakan keadaan terakhir. Lalu keluar dan kembali ke struktur  
(S DCL (NP (FILE(LONG) DEFINITE))  
HAS  
(VP PRINTED))

Struktur ini merupakan output dari analisis.

Lanjut ke bahasa alami 2