

Bab 3

Pengolahan Bahasa Alami

POKOK BAHASAN:

- ✓ Aplikasi Pengolahan Bahasa Alami
- ✓ Grammatika
- ✓ Parsing
- ✓ Semantik

TUJUAN BELAJAR:

Setelah mempelajari materi dalam bab ini, mahasiswa diharapkan :

- ✓ Memahami Definisi dari Pengolahan Bahasa Alami
- ✓ Mengetahui Aplikasi dalam Pengolahan Bahasa Alami
- ✓ Mengerti bidang pengetahuan dari Pengolahan Bahasa Alami yaitu Grammatika, Parsing dan Semantik.

3.1. PENDAHULUAN

Bahasa sebagai bagian yang penting dari kehidupan manusia, dalam bentuk tulis dapat merupakan catatan dari pengetahuan yang didapat oleh umat manusia dari satu generasi ke generasi berikutnya sedangkan dalam bentuk lisan merupakan sarana komunikasi antar individu dalam suatu masyarakat. Tujuan dalam bidang *Natural Language* ini adalah melakukan proses pembuatan model komputasi dari bahasa, sehingga dapat terjadi suatu interaksi antara manusia dengan komputer dengan perantaraan bahasa alami. Model komputasi ini dapat berguna untuk keperluan ilmiah misalnya meneliti sifat - sifat dari suatu bentuk bahasa alami maupun untuk keperluan sehari - hari dalam hal ini memudahkan komunikasi antara manusia dengan komputer.

Sebuah *Natural Language System* harus memperhatikan pengetahuan terhadap bahasa itu sendiri, baik dari segi kata yang digunakan, bagaimana kata - kata tersebut digabung untuk menghasilkan suatu kalimat, apa arti sebuah kata, apa fungsi sebuah kata dalam sebuah kalimat dan sebagainya. Akan tetapi kita juga harus mempertimbangkan ada satu hal lagi yang sangat berperan dalam bahasa, yaitu kemampuan manusia untuk mengerti dan kemampuan untuk itu didapat dari pengetahuan yang didapat secara terus menerus sewaktu hidup. Sebagai contoh dalam suatu percakapan, seseorang mungkin dapat menjawab suatu pertanyaan atau ikut dalam suatu percakapan dengan tidak hanya berdasar pada kemampuan berbahasa tapi juga harus tahu misalnya kata istilah yang umum digunakan dalam kelompok percakapan itu atau bahkan harus tahu konteks dari percakapan itu sendiri.

3.2. BIDANG PENGETAHUAN DALAM NATURAL LANGUAGE

Secara singkat pengolahan bahasa alami (*natural language processing*) mengenal beberapa tingkat pengolahan yaitu :

Dibawah ini dijelaskan bidang-bidang pengetahuan yang berhubungan dengan *Natural Language Processing* :

1. **Fonetik dan fonologi** : berhubungan dengan suara yang menghasilkan kata yang dapat dikenali. Bidang ini menjadi penting dalam proses aplikasi yang memakai metoda *speech based system*.
2. **Morfologi** : yaitu pengetahuan tentang kata dan bentuknya dimanfaatkan untuk membedakan satu kata dengan lainnya. Pada tingkat ini juga dapat dipisahkan antara kata dan elemen lain seperti tanda baca. Sebagai contoh kata *going* :
going (word)
go (root)
ing (suffix)
kata *understand* :
under(prefix)
stand(root)
3. **Sintaksis** : yaitu pemahaman tentang urutan kata dalam pembentukan kalimat dan hubungan antar kata tersebut dalam proses perubahan bentuk dari kalimat menjadi bentuk yang sistematis. Meliputi proses pengaturan tata letak suatu kata dalam

kalimat akan membentuk kalimat yang dapat dikenali. Selain itu dapat pula dikenali bagian - bagian kalimat dalam suatu kalimat yang lebih besar. Sebagai contoh kalimat S dibentuk dari noun phrase (NP) dan verb phrase (VP)

S -> NP,VP

Dan berikutnya :

NP -> DET,N

VP -> V,NP

NP -> N

4. **Semantik** : yaitu pemetaan bentuk struktur sintaksis dengan memanfaatkan tiap kata ke dalam bentuk yang lebih mendasar dan tidak tergantung struktur kalimat . Semantik mempelajari arti suatu kata dan bagaimana dari arti kata - arti kata tersebut membentuk suatu arti dari kalimat yang utuh. Dalam tingkatan ini belum tercakup konteks dari kalimat tersebut.
5. **Pragmatik** : pengetahuan pada tingkatan ini berkaitan dengan masing - masing konteks yang berbeda tergantung pada situasi dan tujuan pembuatan sistem.
6. **Discourse Knowledge** : melakukan pengenalan apakah suatu kalimat yang sudah dibaca dan dikenali sebelumnya akan mempengaruhi arti dari kalimat selanjutnya. Informasi ini penting diketahui untuk melakukan pengolahan arti terhadap kata ganti orang dan untuk mengartikan aspek sementara dari informasi.
7. **World Knowledge** : mencakup arti sebuah kata secara umum dan apakah ada arti khusus bagi suatu kata dalam suatu percakapan dengan konteks tertentu.

Definisi ini tidaklah bersifat kaku, dan untuk setiap bentuk bahasa alami yang ada biasanya ada pendefinisian lagi yang lebih spesifik sesuai dengan karakter bahasa tersebut. Pada beberapa masalah mungkin hanya mengambil beberapa dari pendekatan tersebut bahkan mungkin ada yang melakukan tambahan proses sesuai dengan karakter dari bahasa yang digunakan dan sistem yang dibentuk.

Selain yang sudah disebutkan di atas masih ada lagi satu masalah yang cukup menantang dalam *Natural Language* yaitu ambiguitas atau makna ganda dari suatu kata atau kalimat. Dari satu masukan yang sama dapat menjadi beberapa arti yang berbeda dan masing - masing dapat bernilai benar tergantung pada keperluan pemakai. Hal ini dapat terjadi pada hampir semua tingkatan pendekatan di atas.

3.3 APLIKASI DALAM BIDANG NATURAL LANGUAGE

Jenis aplikasi yang bisa dibuat pada bidang *Natural Language* adalah: *text - based application* dan *dialogue - based applications*.

Text - based application mencakup segala macam aplikasi yang melakukan proses terhadap text tertulis seperti misalnya buku, berita di surat kabar, e-mail dan lain sebagainya. Contoh penggunaan dari *text - based application* ini adalah :

- a. mencari topik tertentu dari buku yang ada pada perpustakaan.
- b. mencari isi dari surat atau *e-mail*
- c. menterjemahkan dokumen dari satu bahasa ke bahasa yang lain.

Akan tetapi tidak semua system yang dapat melakukan hal - hal seperti di atas menggunakan pendekatan *Natural Language*, karena seperti misalnya contoh pencarian topik dari suatu buku di perpustakaan dapat didekati dengan sistem *database* yang cukup lengkap. Tetapi kalau dihadapkan pada pertanyaan yang cukup kompleks dengan bahasa alami yang ada maka akan dirasakan bahwa pendekatan dengan *Natural Language* lebih efisien. Salah satu bentuk yang cukup menarik adalah apabila sistem diminta untuk mencari isi dari suatu berita atau artikel, untuk hal ini pendekatan yang dilakukan hampir serupa dengan pendekatan yang dilakukan manusia apabila menghadapi suatu tes *reading and comprehension*.

Bentuk berikutnya adalah bentuk *dialogue - based application*. Idealnya pendekatan ini melibatkan bahasa lisan atau pengenalan suara, akan tetapi bidang ini juga memasukkan interaksi dengan cara memasukkan teks pertanyaan melalui *keyboard*. Aplikasi yang sering ditemui untuk bidang ini adalah :

- a. sistem tanya jawab, dimana *natural language* digunakan dalam mendapatkan informasi dari suatu *database*.
- b. sistem otomatis pelayanan melalui telepon
- c. control suara pada peralatan elektronik
- d. sistem *problem - solving* yang membantu untuk melakukan penyelesaian masalah yang umum dihadapi dalam suatu pekerjaan.

Sebelumnya perlu diberikan batasan bahwa untuk sistem yang dapat melakukan interaksi melalui bahasa lisan ada bagian *speech recognition* yang merupakan bagian terpisah dari *Natural Language*.

3.4. GRAMATIKA

Grammar suatu bahasa dapat dilihat sebagai suatu aturan yang menentukan apakah suatu kumpulan kata dapat diterima sebagai kalimat oleh bahasa tersebut. Grammar dari Chomsky Hierarchy yaitu Context Free Grammar memiliki sifat lebih mudah dipahami perilakunya dan pengolahannya serta masih dapat diolah dalam bentuk program yang terstruktur.

Sebuah bahasa L dapat dijelaskan sebagai set dari *string*, dimana *string* dibentuk dari bagian terkecil yang disebut *symbol*. Kelompok tertentu v dari *symbol* biasa dikenal sebagai alfabet atau perbendaharaan kata. Sebuah kalimat yang dapat dikenali dibentuk dengan berdasarkan aturan - aturan yang ada yang biasa disebut *grammar*.

Sebuah *grammar* G dapat dibentuk dari 4 *tuple* yaitu : simbol non terminal, simbol terminal, simbol awal dan aturan penulisan atau (*rules*). Definisinya adalah :

$$\blacksquare G = (V_n, V_t, s, P)$$

Sebagai contoh dapat kita lihat dari *grammar* G sederhana berikut ini :

DictJenis = {Kata_Benda, Kata_Kerja, Frasa_Benda, Frasa_Kerja, Keterangan}

DictKata = {Orang, Makan, Telur, Ayam, Terbang, Tinggi}

dengan aturan :

$s \rightarrow \text{Frasa_Benda Frasa_Kerja}$

$\text{Frasa_Benda} \rightarrow \text{Kata_Benda Kata_Benda}$

$\text{Frasa_Kerja} \rightarrow \text{Kata_Kerja Keterangan}$

$\text{Kata_Benda} \rightarrow \{\text{Orang, Telur, Ayam}\}$

$\text{Kata_Kerja} \rightarrow \{\text{Makan, Terbang}\}$

$\text{Keterangan} \rightarrow \{\text{Tinggi}\}$

Dari grammar G dapat dibentuk kalimat :

Orang Makan Ayam

Ayam Terbang Tinggi

Orang Terbang Tinggi

Ayam Makan Orang

semua kalimat tersebut apabila dicari pembentukannya melalui *grammar* G dapat dikatakan benar akan tetapi harus diingat bahwa kalimat dengan *grammar* yang benar hanya berarti benar secara struktural bukan berarti selalu benar dalam makna. Seperti kalimat ketiga yang hanya benar apabila berada dalam konteks 'orang memakai alat'

misalnya pesawat terbang. Sedangkan kalimat keempat malah sama sekali tidak mungkin dapat dimengerti maknanya, selain hanya akan menimbulkan tanda tanya bagi orang yang membaca. Dari *grammar* kita dapat mempelajari bahasa dari segi struktur dan bukan dari segi makna bahasa itu sendiri.

3.5. CHOMSKY HIERARCHY OF GENERATIVE GRAMMAR

Noam Chomsky menyusun *grammar* dalam urutan yang dia sebut tipe 0, 1, 2 dan 3. Tipe 0 adalah bentuk yang paling bebas dan paling sulit dikenali, biasa disebut *recursively enumerable set*, untuk mengenali bentuk ini biasa dipakai *Turing Machine*. Berikutnya adalah tipe 1 yang disebut *context sensitive grammar*. Tipe 2 dari *grammar* yaitu *context free grammar* dinyatakan dengan aturan umum yaitu : $\langle \text{symbol1} \rangle \rightarrow \langle \text{symbol1} \rangle \dots \langle \text{symbolk} \rangle$ dengan $k \geq 1$ dan bagian kiri dari rule adalah *single non terminal symbol*. *Grammar* tipe 3 bernama *finite state* atau *regular grammar*, tipe ini paling sederhana dan mudah dipahami sifatnya.

Secara umum dikatakan bahwa pemakaian *context free grammar* secara murni (tanpa tambahan metoda tertentu) adalah tidak cukup untuk pengolahan bahasa alami. Akan tetapi karena bentuk *context free* dan *regular grammar* tersebut yang paling dipahami perilaku dan pengolahannya, maka beberapa cara telah dikembangkan untuk dapat melakukan pengolahan bahasa alami dengan bentuk *grammar* tersebut.

3.6. PARSING

Parsing adalah suatu proses menganalisa suatu kumpulan kata dengan memisahkan kata tersebut dan menentukan struktur sintaksis dari tiap kata tersebut.. Gramatika yang dipakai juga sangat berkaitan dengan proses parsing apa yang digunakan. Pada Bottom-Up Parsing gramatika yang dipakai akan lebih banyak bercabang ke arah simbol non-terminal. Hal lain yang juga berkaitan erat dengan proses *parsing* adalah kamus atau leksikon yang digunakan. Dalam leksikon disimpan daftar kata yang dapat dikenali sebagai simbol terminal dalam *grammar* dan informasi yang diperlukan untuk tiap kata tersebut untuk proses parsing yang bersangkutan.

Dari pendekatan dalam mengenali struktur suatu kalimat, proses *parsing* dapat dibagi menjadi dua bagian besar yaitu *Top Down parsing* dan *Bottom Up parsing*. *Top*

Down parser memulai pemeriksaan dari simbol awal *s* dan mencoba untuk mencari bentuk simbol terminal berikutnya yang sesuai dengan jenis kata dari kalimat masukan. Cara sebaliknya diterapkan untuk *Bottom Up parser* yaitu mencari dari simbol - simbol terminal menuju ke arah pembentukan simbol awal *s*.

3.7. SEMANTIK

Semantic analyzer mempunyai himpunan rule dalam basis pengetahuan untuk menginterpretasikan sebuah kalimat.

Rule 1 :

IF determiner adalah bagian pertama dalam kalimat dan diikuti oleh noun THEN noun tersebut dianggap sebagai subyek.

Rule 2 :

IF verb diikuti subyek THEN verb menjelaskan tentang apa yang dikerjakan oleh subyek.

Rule 3 :

IF noun diikuti subyek dan verb THEN noun tersebut dianggap sebagai obyek

Rule 4 :

IF kalimat mempunyai bentuk subyek, ver, obyek THEN subyek mengerjakan (verb) yang ada hubungannya dengan obyek.

Kalimat 'A plane flew home' mentrigger rule 1 yang mengidentifikasi plane sebagai subyek, lalu rule 2 menjelaskan bahwa plan flew. Rule 3 dan 4 mengidentifikasikan home sebagai obyek.

Natural language processing dapat dipakai sebagai front (bagian depan) pada sistem AI, dimana data dilewatkan secara verbal.

Pendekatan yang dipakai adalah pendekatan Semantic Grammar yang dipadukan dengan Dictionary tambahan dan Template Grammar.

Semantic Grammar sebagai grammar utama dipilih dengan alasan dalam grammar ini sudah terkandung unsur semantic yang dapat membantu pembentukan semantic dari kalimat. Selain itu karena format dari kalimat sudah dibatasi pada bentuk tertentu (representasi data tabel) maka grammar ini dapat diandalkan terutama untuk bentuk - bentuk tanya dan perintah.

Mendefinisikan *semantic* dan arti sebenarnya adalah proses yang sulit karena hal ini tergantung pada maksud dalam kalimat dan juga adanya kemungkinan arti lain dalam kalimat. Seperti misalnya ‘makan hati ‘dapat diartikan makan dengan lauk hati atau perasaan sedih yang ada pada hati seseorang, tergantung pada letaknya dalam kalimat. Apabila terdapat pada kalimat; karena baru mendapat rejeki anak itu makan hati di restoran; maka artinya adalah yang pertama tetapi jika pada kalimat; dia makan hati karena ditinggal pergi pacarnya; berarti yang kedua. Pada bagian ini akan diterangkan beberapa pendekatan *semantic* yang biasa dilakukan pada suatu *Natural Language system*.

3.8. SEMANTIC GRAMMAR

Pada Semantik Grammar, dipakai sekumpulan *rule* yang bukan hanya bersifat sintaksis tapi juga bersifat semantis dan pragmatis. Hasil dari proses *parsing* dengan menggunakan *grammar* tersebut adalah langsung berupa representasi semantis dari kalimat yang diolah. Dapat dilihat pada contoh dengan *domain* sistem jadwal penerbangan pesawat udara :

the flight to Chicago

the 8 o'clock flight

flight 457 to Chicago

Grammar untuk sistem ini pada umumnya dikenali sebagai :

NP → DET CNP (the flight)

CNP → N (flight)

CNP → CNP PP (flight to Chicago)

CNP → PRE-MOD CNP (8 o'clock flight)

NP → N NUMB (flight 457)

Tetapi perlu diingat bahwa *grammar* tersebut masih bersifat umum, masih memungkinkan terjadi kesalahan karena luasnya jangkauan *grammar* seperti :

the city to Chicago

the 8 o'clock city

Untuk itu maka dilakukan pembatasan dengan jalan memberikan kategori leksikal baru pada suatu kata yang berdasar pada keperluan semantis. Dapat kita gambarkan bahwa pada :

NP → DET CNP (the flight) diubah menjadi

FLIGHT-NP → DET FLIGHT-CNP

CNP → N (flight)

diubah menjadi

FLIGHT-CNP → FLIGHT-N

dengan demikian kita sudah melakukan pembatasan bahwa yang mungkin dibentuk untuk kata benda dari sistem adalah ‘the flight’ dan ‘flight’ tidak untuk yang lain. Perlu diingat perubahan ini juga akan merubah kata yang lain misal ‘Chicago’ dari NP menjadi misalnya CITY-NAME. Secara lengkap maka *grammar* di atas tadi berubah menjadi :

FLIGHT-NP → DET FLIGHT-CNP

FLIGHT-CNP → FLIGHT-N

FLIGHT-CNP → FLIGHT-CNP FLIGHT-DEST

FLIGHT-CNP → FLIGHT-CNP FLIGHT-SOURCE

FLIGHT-CNP → FLIGHT-N FLIGHT-PART

FLIGHT-CNP → FLIGHT-PRE-MOD FLIGHT-CNP

FLIGHT-NP → FLIGHT-N NUMB

CITY-NP → CITY-NAME

CITY-NP → DET CITY-CNP

CITY-CNP → CITY-N

CITY-CNP → CITY-MOD CITY-CNP CITY-MOD-ARG

Dari *grammar* dasar di atas dapat dibentuk *grammar* tambahan misalnya untuk pertanyaan

TIME-QUERY → When does FLIGHT-CNP (When does flight to Chicago)

Dengan pendekatan ini, interpretasi dari *rule* yang bersangkutan menjadi lebih mudah karena sebagian besar dari informasi *semantic* yang diperlukan dapat dilihat dari *rule* yang digunakan. Kekurangan dari pendekatan ini adalah *domainsistem* yang tidak begitu besar, dimana *domain* yang baru akan memerlukan aturan yang baru yang sesuai. Selain itu terjadi pembengkakan jumlah *rule* yang diperlukan, hal ini karena dengan langsung mengacu ke semantik maka banyak generalisasi linguistik yang harus diperinci lebih jauh. Seperti pada contoh di atas kita harus memisahkan antara NP untuk ‘flight’ dan ‘Chicago’ menjadi FLIGHT-NP dan CITY-NAME.

3.8. TATA BAHASA INDONESIA

Sub bab ini membahas secara singkat beberapa pengertian dasar tentang tata bahasa bahasa Indonesia yang memiliki kaitan erat dengan pengolahannya dengan komputer.

Kalimat dapat dirumuskan sebagai konstruksi sintaksis terbesar yang terdiri dari dua kata atau lebih. Hubungan struktural antara kata dan kata, atau kelompok kata dan kelompok kata lain, berbeda – beda. Antara ‘kalimat’ dan ‘kata’ terdapat dua satuan sintaksis antara, yaitu ‘klausa’ dan ‘frasa’. Klausa merupakan satuan sintaksis yang terdiri dari dua kata atau lebih, yang mengandung unsur predikasi. Unsur predikasi adalah subyek, predikat, obyek, pelengkap, atau keterangan. Frasa adalah satuan sintaksis yang terdiri dari dua kata atau lebih yang tidak mengandung unsur predikasi (Krulee : 1991) dalam Wiryana(2001)

Kalimat umumnya berwujud rentetan kata yang disusun sesuai dengan kaidah yang berlaku. Tiap kata dalam kalimat mempunyai tiga klasifikasi, yaitu berdasarkan (1) [bentuk dan kategori kata / frasa](#), (2) [fungsi sintaksis](#), dan (3) peran semantisnya. Contoh penguraian kalimat berdasarkan klasifikasi di atas terlihat pada Tabel 1.

Tabel 1 Contoh Hubungan Bentuk, Kategori, Fungsi dan Peran Unsur-Unsur Kalimat

Bentuk	<i>Ibu</i>	<i>saya</i>	<i>tidak</i>	<i>mem-</i>	<i>baru</i>	<i>untuk</i>	<i>kami</i>
uk	<i>u</i>	<i>a</i>	<i>k</i>	<i>beli</i>	<i>ju</i>	<i>u</i>	<i>uk</i>
Kata	N	Pro n	Ad v	V	N	Adj	Pre p
Fras a	FN		FV		FN		FPrep
Fung si	Subjek		Predik at		Objek		Pelengkap
Pera n	Pelaku		Perbuatan		Sasaran		Peruntung

[Macam ragam kalimat](#) / jenis kalimat dapat ditinjau dari beberapa hal antara lain berdasarkan jumlah klausanya. **Klausa** adalah satuan sintaksis yang terdiri atas dua kata atau lebih yang mengandung unsur predikasi. Kalimat dapat terdiri dari satu klausa atau

lebih. Dengan demikian, dalam kalimat yang mempunyai lebih dari satu klausa, terdapat suatu [hubungan antar klausa](#) yang menentukan struktur kalimat tersebut.

Bentuk dan kategori Kata/ Frasa

Kata dapat dibedakan berdasarkan kategori sintaksisnya. Artinya, kata yang mempunyai bentuk serta perilaku yang sama dimasukkan dalam satu kelompok. **Kategori sintaksis** sering juga disebut **kategori** atau **kelas kata**. Kelas kata yang ada dalam bahasa Indonesia adalah:

1. Verba (kata kerja)
2. Nomina (kata benda)
3. Adjektiva (kata sifat)
4. Adverbial (kata keterangan)
5. Kata tugas :
 - a. Preposisi (kata depan)
 - b. Konjungtor (kata sambung)
 - c. Partikel penegas
 - d. Artikula
 - e. Interjeksi (kata seru)

Nomina, verba, maupun adjektiva dapat dikembangkan dengan tambahan pembatas (pewatas) tertentu menjadi sebuah frasa. **Frasa** adalah satuan sintaksis yang terdiri atas dua kata atau lebih yang tidak mengandung unsur predikasi. Jika dua kata atau lebih itu mengandung unsur predikasi maka satuan sintaksis tersebut disebut **klausa**. **Kategori frasa** dibedakan berdasarkan unsur utamanya, yaitu:

1. Frasa Nominal (FN), contoh: *gedung* → *gedung sekolah, gedung bagus*
2. Frasa Verbal (FV), contoh: *makan* → *makan pelan-pelan, pergi* → *sudah pergi*
3. Frasa Adjektival (FAdj), contoh: *manis* → *sangat manis*
4. Frasa Preposisional (FPrep), contoh: *kepada* → *kepada saya*

3.8.1. Verba dan Frasa Verbal

Verba memiliki fungsi utama sebagai predikat atau sebagai inti predikat dalam kalimat. Verba mengandung makna perbuatan (aksi), proses, dan keadaan yang bukan sifat/kualitas. Pada dasarnya verba terdiri atas verba transitif dan verba taktransitif. Ketransitifan verba ditentukan oleh dua faktor: (1) adanya nomina yang berdiri di belakang verba yang berfungsi sebagai objek dalam kalimat aktif dan (2) kemungkinan objek itu berfungsi sebagai subjek dalam kalimat pasif.

3.8.2. Nomina, Pronomina dan Numeralia

3.8.2.1. Nomina dan Frasa Nominal

Nomina adalah kata yang mengacu pada manusia, binatang, benda, konsep, dan pengertian. Nomina menduduki fungsi subjek, objek, pelengkap, dan keterangan. Nomina diingkarkan dengan kata *bukan*.

Pada **frasa nominal**, nomina berfungsi sebagai inti frasa. Pewatas depannya dapat berupa numeralia (*lima lembar*) atau kata tugas (*di atas*). Numeralia sebagai pewatas depan dapat diikuti oleh nomina penggolong (*dua buah buku*). Nomina bisa juga berada di depannya jika tidak ada pewatas lain (*buku dua buah*).

Nomina umumnya diikuti dengan adjektiva dan dapat didahului kata *yang*, contoh: *anak (yang) baik*. Pewatas belakang nomina bisa juga berupa nomina lain (*atap rumah*), verba (*jam kerja*), pronomina (*buku saya*), atau frasa preposisional (*uang untuk pondokan*) dan ditutup dengan pronomina orang (seperti: *saya, kamu*) dan pronomina tunjuk (*ini, itu*). Pewatas belakang nomina dapat juga berupa klausa yang diawali dengan kata *yang*, contoh: *paman yang tinggal di Bandung*.

3.8.2.2. Pronomina dan Frasa Pronomina

Pronomina adalah kata yang dipakai untuk mengacu kepada nomina lain, contoh: kata *dia* mengacu kepada *perawat*. Pronomina menduduki fungsi subjek, objek, maupun predikat. Pronomina terbagi atas:

1. Pronomina persona, yaitu pronomina yang dipakai untuk mengacu pada orang. Contoh: *saya, aku, kami, kita, engkau, kamu, kalian, dia, beliau, mereka*.
2. Pronomina penunjuk, yaitu pronomina yang dipakai untuk mengacu pada benda, tempat atau ihwal. Contoh: *ini, itu, sini, sana, begini, begitu, demikian*.

Selain itu ada juga *yakni* dan *yaitu* yang mengacu ke depan maupun ke belakang. Bagian di belakangnya merupakan penjas/uraian. Contoh: Camat wilayah itu, *yakni* Pak Sitepu, masih berkerabat dengan saya.

3. Pronomina penanya, yaitu pronomina yang dipakai sebagai pemarah pertanyaan. Contoh: *siapa, apa, mana, mengapa, kenapa, kapan, bila(mana), di/ke/dari mana, bagaimana, berapa*.

Frasa pronominal adalah frasa yang intinya berupa pronomina. Pronomina dapat diikuti oleh numeralia kolektif (*kamu sekalian*), kata penunjuk (*saya ini*), kata *sendiri* (*mereka sendiri*), ataupun klausa yang diawali dengan kata *yang* (*mereka yang tidak hadir akan ditegur*).

3.8.2.3. Numeralia dan Frasa Numeralia

Numeralia adalah kata yang dipakai untuk menghitung banyaknya nomina. Pada dasarnya, dalam bahasa Indonesia, ada dua macam numeralia yaitu numeralia pokok dan numeralia tingkat.

Numeralia pokok adalah bilangan dasar yang menjadi sumber dari bilangan-bilangan yang lain. Numeralia pokok terbagi menjadi numeralia pokok tentu (*0-nol* sampai *9-semilan*), pokok kolektif (*kedua gedung, lima sekawan, kami berdua, berpuluh-puluh, puluhan*), pokok distributif (*masing-masing, tiap-tiap, setiap*), pokok taktentu (*banyak, berbagai, beberapa, semua, seluruh, segala, segenap*), numeralia klitika (*eka-, dwi-, tri-, catur-, panca-, sapta-, dasa-*), numeralia ukuran (*lusin, kodi, meter, liter, gram*).

Numeralia tingkat diletakkan dibelakang nomina untuk menyatakan tingkatan nomina tersebut. Contoh: *anak pertama, gedung kedua*.

Frasa numeralia dibentuk dengan menambahkan kata penggolong. Contoh: *tiga buah (buku)*. Penggolong secara umum terbagi atas penggolong orang yaitu *orang*, penggolong binatang yaitu *ekor*, dan penggolong untuk nomina yang bukan manusia dan bukan binatang yaitu *buah*.

3.8.3. Adjektiva dan Frasa Adjektival

Adjektiva adalah kata yang memberikan keterangan (bersifat atributif) terhadap nomina dalam kalimat. Nominanya menjadi subjek, objek atau pelengkap dalam kalimat. Letak adjektiva adalah di sebelah kanan nomina. Contoh: baju *baru*, baju *putih* yang *baru* dan *bersih* Adjektiva mengungkapkan kualitas (disebut adjektiva bertaraf) atau keanggotaan dalam suatu golongan (disebut adjektiva tak bertaraf) dari acuan nomina yang diatasinya.

Frasa adjektival adalah frasa yang inti frasanya adalah adjektiva. Adjektiva ini dapat diatasi oleh pemarkah aspektualitas dan pemarkah modalitas di sebelah kirinya. Contoh: *tidak bodoh, sudah akan sembuh*. Adjektiva juga dapat diikuti pewatas yang berposisi di sebelah kanannya. Contoh: *sakit lagi, kaya juga*.

3.8.4. Adverbia

Dalam tataran klausa, adverbia mewatasi/menjelaskan fungsi-fungsi sintaksis, khususnya predikat. Dalam tataran frasa, adverbia adalah kata yang menjelaskan verba, adjektiva atau adverbia lain. Adverbia juga menjelaskan nomina, frasa preposisional, pronomina, dan numeralia. Ada empat macam posisi adverbia, yaitu:

1. mendahului kata yang diterangkan, contoh: *lebih tinggi, hanya menulis*
2. mengikuti kata yang diterangkan, contoh: *merah sekali, duduk-duduk saja*
3. mendahului atau mengikuti kata yang diterangkan, contoh: *amat mahal, mahal amat*
4. mendahului dan mengikuti kata yang diterangkan, contoh: *sangat manis sekali*

3.8.5. Kata Tugas

Kata tugas adalah kata yang hanya mempunyai arti gramatikal dan tidak memiliki makna leksikal. Artinya, kata tugas baru akan mempunyai arti apabila dirangkaikan dengan kata lain, contoh: *ke pasar*. Ciri lainnya adalah kata tugas tidak dapat menjadi dasar untuk membentuk kata lain, dan kata tugas merupakan kelas kata tertutup (tidak ada padanannya). Berdasarkan peranannya dalam frasa atau kalimat, kata tugas dibagi menjadi lima kelompok: (1) preposisi, (2) konjungtor, (3) partikel penegas, (4) artikula, (5) interjeksi.

1. Preposisi

Preposisi (kata depan) menandai hubungan makna antara unsur di depan preposisi tersebut dengan unsur di belakangnya. Contoh dalam frasa pergi ke pasar, preposisi ke menyatakan hubungan makna arah antara pergi dan pasar. Peran semantis preposisi yang lazim dalam bahasa Indonesia adalah:

1. Penanda hubungan tempat, contoh: di, ke, dari, hingga, sampai, antara, pada
2. Penanda hubungan peruntukan, contoh: bagi, untuk, buat, guna
3. Penanda hubungan sebab, contoh: karena, sebab, lantaran
4. Penanda hubungan kesertaan atau cara, contoh: dengan, sambil, beserta, bersama
5. Penanda hubungan pelaku, contoh: oleh
6. Penanda hubungan waktu, contoh: pada, hingga, sampai, sejak, semenjak, menjelang
7. Penanda hubungan ihwal peristiwa, contoh: tentang, mengenai
8. Penanda hubungan milik, contoh: dari

2. Konjungtor

Konjungtor (kata sambung) adalah kata tugas yang menghubungkan dua satuan bahasa yang sederajat: kata dengan kata, frasa dengan frasa, atau klausa dengan klausa. Dilihat dari perilaku sintaksisnya dalam kalimat, konjungtor dibagi menjadi empat kelompok:

1. Konjungtor koordinatif adalah konjungtor yang menggabungkan kata atau klausa yang setara. Klausa yang dihubungkan dengan konjungtor koordinatif akan membentuk kalimat majemuk setara. Yang termasuk konjungtor koordinatif antara lain adalah dan, serta, atau, tetapi, melainkan, padahal sedangkan.
2. Konjungtor korelatif adalah konjungtor yang menghubungkan dua kata, frasa atau klausa yang memiliki status sintaksis yang sama. Konjungtor korelatif terdiri atas dua bagian yang dipisahkan oleh salah satu kata, frasa, atau klausa yang dihubungkan.

Contoh:

- (3) Baik Pak Anwar maupun istrinya tidak suka merokok
 - (4) Kita tidak hanya harus setuju, tetapi juga harus patuh
3. Konjungtor subordinatif adalah konjungtor yang menghubungkan dua klausa atau lebih, dan klausa itu tidak memiliki status sintaksis yang sama. Salah satu dari klausa itu merupakan anak kalimat. Penggabungan anak kalimat itu dengan induk kalimatnya menghasilkan kalimat majemuk bertingkat. Berikut ini adalah kelompok-kelompok konjungtor subordinatif:
 - a. konjungtor subordinatif waktu, contoh: sejak, ketika, setelah, hingga
 - b. konjungtor subordinatif syarat, contoh: jika, kalau, bila, asal
 - c. konjungtor subordinatif pengandaian, contoh: andaikan, seandainya
 - d. konjungtor subordinatif tujuan, contoh: agar, supaya, biar
 - e. konjungtor subordinatif konsesif, contoh: biarpun, meski(pun), walau(pun)
 - f. konjungtor subordinatif perbandingan, contoh: seperti, sebagai, daripada
 - g. konjungtor subordinatif sebab, contoh: sebab, karena, oleh karena
 - h. konjungtor subordinatif hasil, contoh: sehingga, sampai, maka
 - i. konjungtor subordinatif alat, contoh: dengan, tanpa

j. konjungtor subordinatif cara, contoh: dengan, tanpa

k. konjungtor subordinatif komplementasi,

contoh: bahwa

l. konjungtor subordinatif atributif, contoh: yang

m. konjungtor subordinatif perbandingan,

contoh: sama ... dengan ...

4. Konjungtor antarkalimat adalah konjungtor yang merangkaikan dua kalimat tetapi masing-masing merupakan kalimat sendiri-sendiri. Konjungtor ini selalu memulai suatu kalimat yang baru. Contoh: konjungtor antarkalimat: biarpun demikian, kemudian, sesudah itu, lagi pula, selain itu, sebaliknya, sesungguhnya, bahkan, (akan) tetapi, namun, kecuali itu, dengan demikian, oleh karena itu, sebelum itu.

3. Partikel Penegas

Kategori partikel penegas meliputi kata yang tidak tertakluk pada perubahan bentuk dan hanya berfungsi menampilkan unsur yang diiringinya. Ada empat macam partikel penegas yaitu -kah, -lah, -tah, dan pun.

4. Artikula

Artikula adalah kata tugas yang membatasi makna nomina. Dalam bahasa Indonesia dikenal kelompok artikula yang bersifat gelar (sang, sri, hang, dang), artikula yang mengacu ke makna kelompok (para), dan artikula yang menominalkan (si)

5. Interjeksi

Interjeksi (kata seru) adalah kata tugas yang mengungkapkan rasa hati pembicara. Contoh: bah, sialan, aduhai. aduh, astaga, ayo, hai, nah.

RINGKASAN:

- ✓ Tujuan dalam bidang *Natural Language* ini adalah melakukan proses pembuatan model komputasi dari bahasa, sehingga dapat terjadi suatu interaksi antara manusia dengan komputer dengan perantaraan bahasa alami.

- ✓ Pengolahan Bahasa Alami (Natural language processing) mengenal beberapa tingkat pengolahan yaitu fonetik, morfologi, sintaksis, semantik, pragmatik, discourse knowledge dan world knowledge
- ✓ Jenis aplikasi yang bisa dibuat pada bidang *Natural Language* adalah: *text - based application* dan *dialogue- based applications*.
- ✓ Grammar suatu bahasa dapat dilihat sebagai suatu aturan yang menentukan apakah suatu kumpulan kata dapat diterima sebagai kalimat oleh bahasa tersebut.
- ✓ Grammar dari Chomsky Hierarchy yaitu Context Free Grammar memiliki sifat lebih mudah dipahami perilakunya dan pengolahannya serta masih dapat diolah dalam bentuk program yang terstruktur.
- ✓ Parsing adalah suatu proses menganalisa suatu kumpulan kata dengan memisahkan kata tersebut dan menentukan struktur sintaksis dari tiap kata tersebut.. Gramatika yang dipakai juga sangat berkaitan dengan proses parsing apa yang digunakan.

LATIHAN SOAL :

- 1) Jelaskan pengertian dari Pengolahan Bahasa Alami !
- 2) Sebutkan jenis aplikasi yang bisa dibuat pada bidang Pengolahan Bahasa Alami. Jelaskan masing-masing secara singkat !
- 3) Bagaimana cara menginterpretasikan suatu kalimat dengan *Semantic Analyzer* ! Berikan contohnya.