



Sözdizimsel Analiz (Syntactic Analysis)



Doç.Dr.Banu Diri

Morfolojik Çözümlemede Analitik Yaklaşımlar

Doğal Dil İşleme çalışmalarında anlam bütünsel çözümleme yapabilmek için, bazı yaklaşımlar mevcuttur.

Sözdizimsel Analiz

Anlambilimsel Analiz

Slide 1

Sözdizimsel analiz, sözdizimini (syntax) veya cümleyi oluşturan **morfolojik** öğelerin **hiyerarşik** kurallara uyumunu karşılaştırarak ölçümlemektir. Böylece söz dizimin anlamlı olup olmadığının ölçülebilmesi için düzenleyici bir süreç gerçekleşmiş olur.

Bu sayfaya ihtiyaç duymayan doğal diller olsa bile birçok doğal dil bu işlemi zorunlu kılar. Sözdizimsel analiz için bir çok çözüm düşünülmüş olsa da en çok tercih edilen çözüm şekilleri :

Dilin sözdizimsel yapısının gramer adı verilen tanımlayıcı şeklinde gösterilmesi

Dilin tanımlanmış grameri temelinde onun genel yapısının oluşturulmasıdır

Slide 2

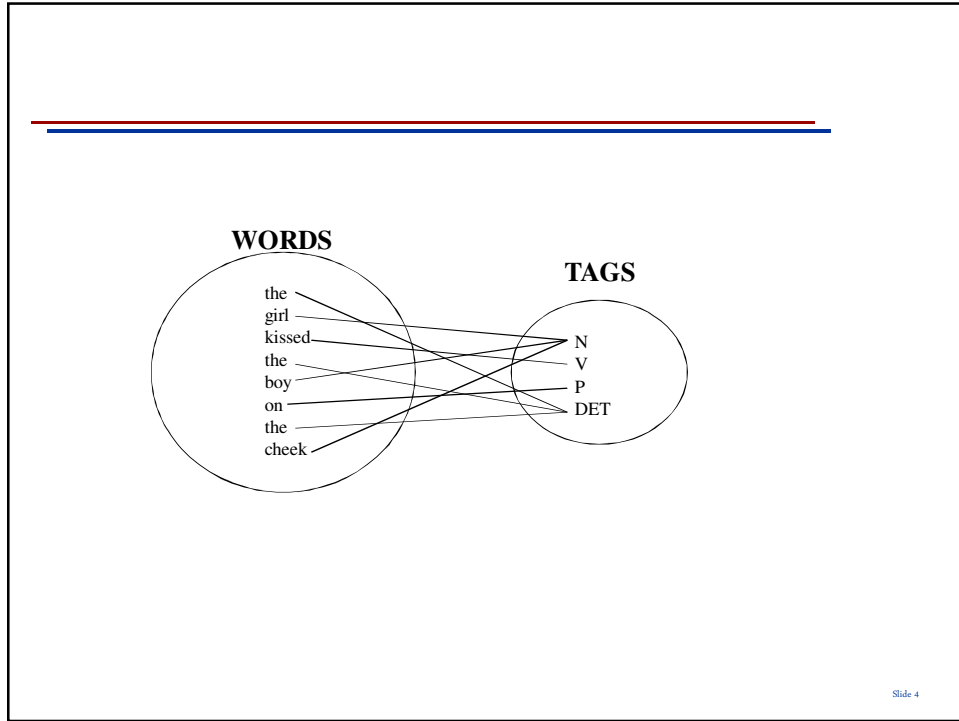
POS (Part of Speech)

Doküman içerisinde yer alan her bir kelime **POS** olarak adlandırılır. Her POS kendisine verilen **TAG**'ler (etiket) ile ifade edilir ve sözdizimsel yapı çıkarılırken kullanılır.

Benzer sözdizimsel davranışlar ile kelimelerin sınıflandırılması **syntactic** veya **grammatical categories** veya **parts of Speech (POS)** olarak adlandırılır.

POS'ları hem İngilizce hem de Türkçe için görelim.

Slide 3



Kelimelerin temel sözdizim fonksiyonları

Nouns (isimler) dünyadaki varlıkları tanımlar (e.g. 'people/insanlar', 'animals/hayvanlar', 'hat/şapka').

Determiners ismi belirler (e.g. 'the', 'a', 'an').

Adjectives(sıfatlar) isimleri niteleyen sözcükler olup isimlerle ortaya çıkar (e.g. 'red', 'long', 'intelligent'). Tek başlarına kullanılmazlar. Örneğin "kırmızı", bir renk ismidir, "iki", bir sayı ismidir. Ancak bu sözcükler isimlerin özelliklerini bildirecek duruma gelirse sıfat olur. "kırmızı top".

Verbs(fiiller) bir oluşu, bir durumu kişiye bağlayarak anlatırlar. (e.g. 'have/sahip olmak', 'threw/fırlatmak', 'walked/yürümek').

Adverbs(zarf/belirtec) fiillerin niteliğini bildirir (e.g. 'often', 'heavily'). "Güzel bir evde oturmak istiyorum." cümlesinde "güzel" sözcüğü "ev" ismini nitelediği için sıfat, "Bu ev uzaktan daha güzel görünüyordu" cümlesinde ise "görünmek" fiilinin nasıl olduğunu bildirdiği için zarftır.

Prepositions(edat/ilgeç) Tek başına bir anlam taşımayan, ancak kendinden önceki sözcükle birlikte kullanıldığında belirli bir anlamı olan sözcüklerdir (e.g. 'in', 'on', 'over')(gibi, için, ile, kadar, karşı, göre, üzere, doğru, sanki, beri, sadece, ancak, yalnız, sonra,...)

Conjunctions(bağlaç) Cümleleri veya aynı görevdeki sözcükleri birbirine bağlayarak aralarında anlam ilgisi kuran sözcüklere denir (e.g. 'and', 'or', 'but') (ile, ve, de, ama, fakat, ancak, yalnız, ki, hem...hem, ne...ne, çünkü, madem, ...).

Slide 5

WORD	LEMMA	TAG
the	the	+DET
girl	girl	+NOUN
kissed	kiss	+VPAST
the	the	+DET
boy	boy	+NOUN
on	on	+PREP
the	the	+DET
cheek	cheek	+NOUN

Slide 6

Kelime Sınıfları: TAG kümesi

Tag	Description	Example	Tag	Description	Example
CC	Coordn. Conjunction	<i>and, but, or</i>	SYM	Symbol	<i>+, %, &</i>
CD	Cardinal number	<i>one, two, three</i>	TO	"to"	<i>to</i>
DT	Determiner	<i>a, the</i>	UH	Interjection	<i>ah, oops</i>
EX	Existential 'there'	<i>there</i>	VB	Verb, base form	<i>eat</i>
FW	Foreign word	<i>mea culpa</i>	VBD	Verb, past tense	<i>ate</i>
IN	Preposition/sub-conj	<i>of, in, by</i>	VBG	Verb, gerund	<i>eating</i>
JJ	Adjective	<i>yellow</i>	VCN	Verb, past participle	<i>eaten</i>
JJR	Adj., comparative	<i>bigger</i>	VBP	Verb, non-3sg pres	<i>eat</i>
JJS	Adj., superlative	<i>widest</i>	VBZ	Verb, 3sg pres	<i>eats</i>
LS	List item marker	<i>1, 2, One</i>	WDT	Wh-determiner	<i>which, that</i>
MD	Modal	<i>can, should</i>	WP	Wh-pronoun	<i>what, who</i>
NN	Noun, sing. or mass	<i>llama</i>	WPS	Possessive wh-	<i>whose</i>
NNS	Noun, plural	<i>llamas</i>	WRB	Wh-adverb	<i>how, where</i>
NNP	Proper noun, singular	<i>IBM</i>	\$	Dollar sign	<i>\$</i>
NNPS	Proper noun, plural	<i>Carolinas</i>	#	Pound sign	<i>#</i>
PDT	Predeterminer	<i>all, both</i>	"	Left quote	<i>(' or ")</i>
POS	Possessive ending	<i>'s</i>	"	Right quote	<i>(' or ")</i>
PP	Personal pronoun	<i>I, you, he</i>	(Left parenthesis	<i>(, (, {, <)</i>
PPS	Possessive pronoun	<i>your, one's</i>)	Right parenthesis	<i>(,), }, >)</i>
RB	Adverb	<i>quickly, never</i>	,	Comma	<i>,</i>
RBR	Adverb, comparative	<i>faster</i>	.	Sentence-final punc	<i>(. ! ?)</i>
RBS	Adverb, superlative	<i>fastest</i>	:	Mid-sentence punc	<i>(; ; ... - -)</i>
RP	Particle	<i>up, off</i>			

Slide 7

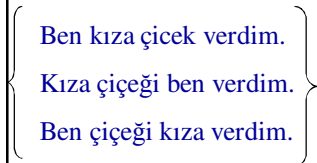
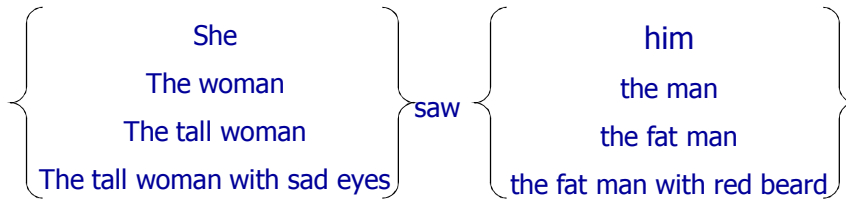
Phrase structure (Öbek yapısı-Sözcük dizimi)

- Kelimeler öbek içerisinde belli bir sırada yer alır.
- Türkçe’de cümleler en genel şekliyle **özne**, **nesne** ve **yüklem** bileşenlerinden oluşur. Cümledeki öğelerin yerini değiştirebiliriz (Fince ve Japonca’ya benzer).
- Cümleye eklenmek istenen anlamlar arttıkça cümleler, özne, yer tamlayıcısı, **zarf** tamlayıcısı, nesne ve yüklem gibi bileşenleri de içerir.
- Cümlelerin anlamını kuvvetlendiren cümle dışı bileşenler de (**bağlaç**, **edat**, vb) cümlede bulunabilir.
- Türkçe’de özne ile yüklem cümlelerin temel bileşenleridir ve genelde tüm cümlelerde yer alırlar. Yer tamlayıcısı, zarf tamlayıcısı, nesne gibi bileşenler bazı cümlelerde yer almayabilirler veya bazı cümlelerde sadece biri, bazılarında sadece ikisi bulunabilir. Bu bileşenlerin cümle içindeki sıralanışları da değişebilir.

Slide 8

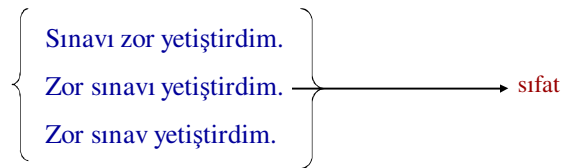
Phrase structure (Öbek yapısı-Sözcük dizimi)

- İngilizce’de sözcük sırası (word order) son derece katıdır.



Slide 9

Serbestliğin kısıtlı olduğu durumlar mevcuttur. Belirteç'lerin (zarf) kullanımı gibi.



Slide 10

The/DT grand/JJ jury/NN commented/VBD on/IN a/DT
number/NN of/IN other/JJ topics/NNS ./.

VB DT NN .
Book that flight .

VBZ DT NN VB NN ?
Does that flight serve dinner ?

Slide 11

Problem ??

Kelimeler birden fazla sınıfa dahil olabiliyor: *this*

This is a nice day = PRP (*personal pronoun*)

This day is nice = DT (*determiner*)

You can go *this* far = RB (*adverb*)

Slide 12

Prepositions (edatlar)

of	540,085	through	14,964	worth	1,563	pace	12
in	331,235	after	13,670	toward	1,390	nigh	9
for	142,421	between	13,275	plus	750	re	4
to	125,691	under	9,525	till	686	mid	3
with	124,965	per	6,515	amongst	525	o'er	2
on	109,129	among	5,090	via	351	but	0
at	100,169	within	5,030	amid	222	ere	0
by	77,794	towards	4,700	underneath	164	less	0
from	74,843	above	3,056	versus	113	midst	0
about	38,428	near	2,026	amidst	67	o'	0
than	20,210	off	1,695	sans	20	thru	0
over	18,071	past	1,575	circa	14	vice	0

Slide 13

Pronouns (zamirler/adıllar)

it	199,920	how	13,137	yourself	2,437	no one	106
I	198,139	another	12,551	why	2,220	wherein	58
he	158,366	where	11,857	little	2,089	double	39
you	128,688	same	11,841	none	1,992	thine	30
his	99,820	something	11,754	nobody	1,684	summat	22
they	88,416	each	11,320	further	1,666	suchlike	18
this	84,927	both	10,930	everybody	1,474	fewest	15
that	82,603	last	10,816	ourselves	1,428	thyslf	14
she	73,966	every	9,788	mine	1,426	whomever	11
her	69,004	himself	9,113	somebody	1,322	whosoever	10
we	64,846	nothing	9,026	former	1,177	whomsoever	8
all	61,767	when	8,336	past	984	wherefore	6
which	61,399	one	7,423	plenty	940	whereat	5
their	51,922	much	7,237	either	848	whatsoever	4
what	50,116	anything	6,937	yours	826	whereon	2
my	46,791	next	6,047	neither	618	whose	2
him	45,024	themselves	5,990	fewer	536	ought	1
me	43,071	most	5,115	hers	482	howsoever	1
who	42,881	itself	5,032	ours	458	thrice	1
them	42,099	myself	4,819	whoever	391	wheresoever	1
no	33,458	everything	4,662	least	386	you-all	1
some	32,863	several	4,306	twice	382	additional	0
other	29,391	less	4,278	theirs	303	anybody	0
your	28,923	herself	4,016	wherever	289	each other	0
its	27,783	whose	4,005	oneself	239	once	0
our	23,029	someone	3,755	thou	229	one another	0
these	22,697	certain	3,345	'un	227	overmuch	0
any	22,666	anyone	3,318	ye	192	such and such	0
more	21,873	whom	3,229	thy	191	whate'er	0
many	17,343	enough	3,197	whereby	176	whenever	0
such	16,880	half	3,065	thee	166	whereof	0
those	15,819	few	2,933	yourselves	148	whereto	0
own	15,741	everyone	2,812	latter	142	whereunto	0
us	15,724	whatever	2,571	whichever	121	whichsoever	0

Slide 14

Conjunctions (bağlaçlar)

and	514,946	yet	5,040	considering	174	forasmuch as	0
that	134,773	since	4,843	lest	131	however	0
but	96,889	where	3,952	albeit	104	immediately	0
or	76,563	nor	3,078	providing	96	in as far as	0
as	54,608	once	2,826	whereupon	85	in so far as	0
if	53,917	unless	2,205	seeing	63	inasmuch as	0
when	37,975	why	1,333	directly	26	insomuch as	0
because	23,626	now	1,290	ere	12	insomuch that	0
so	12,933	neither	1,120	notwithstanding	3	like	0
before	10,720	whenever	913	according as	0	neither nor	0
though	10,329	whereas	867	as if	0	now that	0
than	9,511	except	864	as long as	0	only	0
while	8,144	till	686	as though	0	provided that	0
after	7,042	provided	594	both and	0	providing that	0
whether	5,978	whilst	351	but that	0	seeing as	0
for	5,935	suppose	281	but then	0	seeing as how	0
although	5,424	cos	188	but then again	0	seeing that	0
until	5,072	supposing	185	either or	0	without	0

Slide 15

Auxiliaries (yardımcı fiiller)

can	70,930	might	5,580	shouldn't	858
will	69,206	couldn't	4,265	mustn't	332
may	25,802	shall	4,118	'll	175
would	18,448	wouldn't	3,548	needn't	148
should	17,760	won't	3,100	mightn't	68
must	16,520	'd	2,299	oughtn't	44
need	9,955	ought	1,845	mayn't	3
can't	6,375	will	862	dare	??
have	???				

Slide 16

Kelime Sınıfları: TAG kümesi -TR

	TAG		TAG
Özne	Ö	Zaman Belirteci	ZB
Dolaylı Tümleç	D/DT	Tamlayan	T
Zarf Tümlenci	Z	Tamlanan	TN
Nesne	N	Zamir	ZM
Yüklem	Y	Nesne Eki	NE
İsim Grubu	İG	Tamlayan Eki	TE
Sıfat Grubu	SG	Tamlanan Eki	TNE
İsim Nesnesi	İN	Kip Eki	KE
Sıfat Nesnesi	SN	Zaman Eki	ZE
Diğer Zarflar	DZ	Dolaylı Tümleç Eki	DE
Sıfat	S	Ek Fiil	EF
İsim	İ		

Slide 17

Bu TAG'lerden oluşabilecek cümleler; öğelerinin dizilişlerine göre bir çok varyasyonlar oluşturabilirler. Bu varyasyonlar;

❖ Ö-D-Z-N-Y, Ö-D-N-Z-Y, Ö-Z-D-N-Y, Ö-Z-N-D-Y, Ö-N-Z-D-Y, Ö-N-D-Z-Y olabilir.

❖ Bunlara tüm öğeleri içermeyen iki, üç veya dört öğeden oluşan cümleler de eklenebilir.

İki unsur için: Ö-Y

Üç unsur için Ö-N-Y, Ö-D-Y, Ö-Z-Y

Dört unsur için Ö-N-D-Y, Ö-N-Z-Y, Ö-D-N-Y, Ö-D-Z-Y, Ö-Z-N-Y, Ö-Z-D-Y

❖ Bu varyasyonların hepsi {ÖλλλY} şeklinde gösterilebilir.

Slide 18

Cümlelerin sentaks analizine yönelik olan yöntemler ideal sonuçlar vermemesine rağmen %90'ların üzerinde başarı sağlamaktadır. Bu yöntemlerin ağaç formunda bir çok modeli vardır. En çok tercih edilenleri;

Yukarıdan aşağıya (top-bottom) süzgeçleme

Ağacın son elemanlarındaki semboller, çözümlenen cümlelerin parçalarına karşılık düşene kadar gramer kurallarına uygun olarak ayrıştırmaya devam eden bir yöntemdir.

Aşağıdan yukarıya (bottom-up) çözümleme

Çözümlenmiş cümle ile başlanır ve son elemanları cümlelerin kelimeleri olan ve başlangıç sembolüne kadar tek bir ağaç oluşacak şekilde gramer kurallarını geriye doğru uygulayan yöntemdir.

Slide 19

Syntax (sözdizim) veya Phrase Structure(Öbek yapısı): Basit *context-free* grammar (CFG)

Cümleyi oluşturan her bir POS'u *Backus-Nour* formunda gösterelim.

Bu ifade biçiminde "," mantıksal "ve"

"|" mantıksal "veya"

anlamını taşır.

Slide 20

Brown Tags

NN – singular noun
NNP – proper nouns
NNS – plural nouns
NR – adverbial nouns (e.g. 'home')
JJ - adjective
AT – articles
VB – verb, base form
VBD – verb third person singular (e.g. 'likes')
RB – adverbs
IN - preposition

- S --> NP VP
- NP --> AT NNS |
AT NN |
NP PP
- VP --> VP PP |
VBD |
VBD NP
- PP --> IN NP

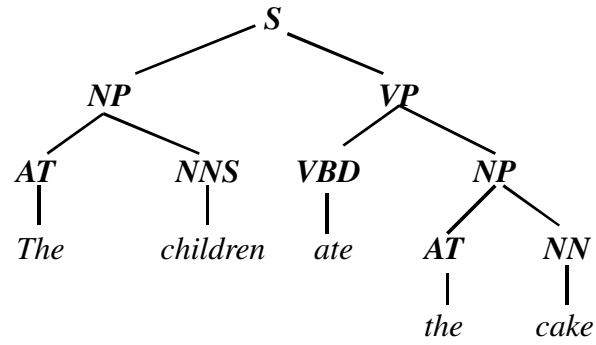
The Grammar

- AT --> the
- NNS --> children |
students |
mountains
- VBD --> slept |
ate |
saw
- IN --> in |
of
- NN --> cake

The Lexicon

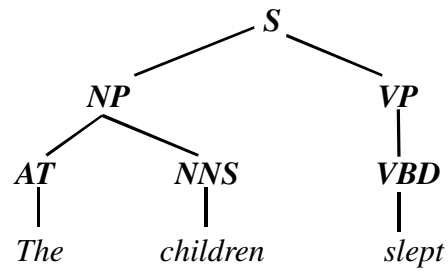
Slide 21

Parse Tree - 1 (Ayrıştırma Ağacı)



Slide 22

Parse Tree - 2 (Ayrıştırma Ağacı)



Slide 23

Örnek:

<sıfat tamlaması> ::= <sıfat>, <isim>

şeklinde kısaca tanımlanabilir. Türkçe cümleler için yapılacak bu şekildeki gösterimde tüm elemanlar bu şekilde ifadelendirilmelidir.

Slide 24

**<cümle> ::= <özne>, <dolaylı tümleş>, <zarf tümleci>, <nesne>, <yüklem> |
<özne>, <nesne>, <zarf tümleci>, <dolaylı tümleş>, <yüklem>**

<özne> ::= <isim grubu> | <sıfat grubu> | <isim>

<dolaylı tümleş> ::= <isim grubu>, <dolaylı tümleş_eki> | <sıfat grubu>, <dolaylı tümleş_eki>

**<zarf tümleci> ::= <zaman belirteçleri> | <isim grubu>,
<diğer_zarflar>**

<nesne> ::= <isim nesnesi> | <sıfat nesnesi>

<yüklem> ::= <fiil>, <zaman_eki>, <kip_eki> | <isim>, <ek_fiil>

Slide 25

Ana unsurları oluşturan öğeler:

<isim grubu>::= <tamlayan>, <tamlanan> | <isim> | <zamir>
<sıfat grubu>::= <sıfat>, <isim grubu>
<isim nesnesi>::= <isim grubu>, <nesne_eki>
<sıfat nesnesi>::= <sıfat grubu>, <nesne_eki>
<tamlayan>::= <isim>, <tamlayan_eki> | <zamir> , <tamlayan_eki>
<tamlanan>::= <isim>, <tamlanan_eki>

Öğeleri oluşturan kelimeler:

<zaman belirteçleri>::= şimdi | bugün | yarın | bu sabah | geçen yıl | dün
<diğer_zarflar>::= için | ile
<sıfat>::= beyaz | kırmızı | küçük | şirin | güzel | uzun | sarı
<isim>::= koyun | araba | köy
<fiil>::= getir

Slide 26

✓ Bilgisayarla doğal dilin modellenmesinde anlamsal analizden önce kelimelerden oluşturulan yapının cümle olup olmadığı test edilmelidir.

✓ Cümle oluşturmaya ifadelere anlam çıkarmak imkansızdır. "Kalem sarı koş kitap" ifadesi bir anlam ifade etmez. Rastgele kelimelerin yan yana gelmesiyle cümle oluşturulamaz. Cümleleri oluşturan kelimelerin bir sonu olmalıdır.

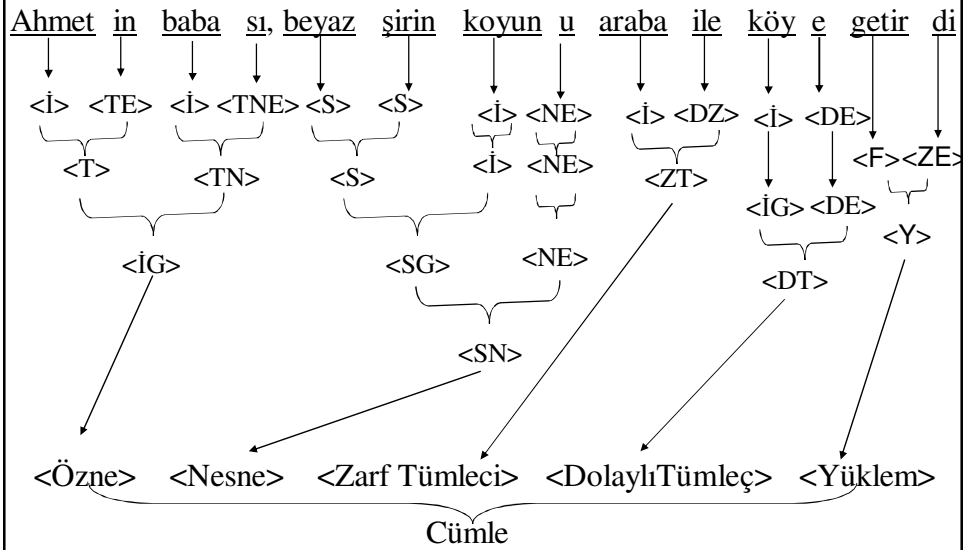
✓ Cümleler de sonlu yapılar olduğuna göre Sonlu Otomata ile gösterilebilirler. Bunun amacı, ardarda gelen kelimeler yığını bu kelimelerin ifade ettiği cümle öğelerini tanımlayan bir yapıya dönüştürmektir. Türkçe'deki cümleler için mümkün olan tüm sözdizimsel unsurlar dikkate alınmalıdır.

Slide 27

✓ Cümle unsurları, kelimeler, tamlamalar veya buna benzer cümle parçacıkları olabilir. Bu cümle parçacıklarının tespiti, anlamsal analiz için yararlı ipuçları vermektedir. Genellikle bir cümlede cümle parçacıkları ortak görev yaparlar.

Slide 28

Cümlelerin Hiyerarşik Yapısı



Slide 29

"Ahmet'in babası beyaz şirin koyunu araba ile köye getirdi."

Cümleinin çözümlenmesi:

1. Cümle durumundan başlanır.
2. Cümleden bir sonraki duruma geçmenin şartı <özne> isimli duruma ait alt GGA'dan başarıyla dönülmesidir. Bu nedenle <özne>ye gidilir.
3. <özne> durumundan bir sonraki duruma geçmenin şartı kelimelerin <isim grubu> veya <sıfat grubu> oluşturmasıdır. "Ahmet": isim, "in": tamlayan_eki, "baba": isim, "sı": tamlanan_eki olduğu tanımlanmış olan veritabanından anlaşılır. Bu kelimelerden oluşan "Ahmet'in babası" kelime grubu <isim grubu> alt GGA'na uygundur. "Ahmet'in babası" kelime grubu <isim grubu> olarak atanır ve sonraki duruma geçilir.
4. Diğer durumlara geçişin bir çok farklı şartı vardır. Bu şartlardan herhangi biri sağlandığında diğer durumlara geçilecektir. "beyaz şirin koyunu" kelime grubunda; "beyaz": sıfat, "şirin": sıfat, "koyun": isim olduğu veritabanından anlaşılır ve bu üç kelime <sıfat grubu> testini başarıyla geçer. "koyunu" kelimesindeki "u" nesne_eki tanımına uygundur. "beyaz şirin koyun" kelime grubu, "u" ekiyle birleşince <sıfat nesnesi> tanımına uyar. <nesne> testini geçmenin şartlarından biri de <sıfat nesnesi> olarak tanımlandığından "beyaz şirin koyunu" kelime grubu <nesne> olarak atanır ve bir sonraki duruma geçilir. Bir sonraki durum yüklem, zarf tümleci veya dolaylı tümleç olmalıdır.

Slide 30

5. "araba" kelimesi <isim> olarak tanımlanan kelimeler arasındadır. <isim grubu> olmanın şartlarından biri de <isim> olmaktır. "ile" kelimesi <diğer_zarflar> arasında olduğundan "araba ile" kelime grubu <zarf tümleci> testini başarıyla geçer ve "araba ile" kelime grubu <zarf tümleci> olarak atanır. Bu durumdan sonra GGA'na göre yüklem veya dolaylı tümleç gelmelidir.
6. "köy" kelimesi <isim> olarak tanımlanan kelimeler arasındadır. "e" eki ise <dolaylı tümleç_eki> dir. Dolayısıyla "köye" kelimesi <dolaylı tümleç> olarak tanınır ve atanır. Bir sonraki durum GGA'na göre <yüklem> olmalıdır.
7. "getir" kelimesi veritabanına göre <fiil> dir. "di" eki de <zaman_eki> olarak tanımlandığından "getirdi" kelimesi <yüklem> olarak kabul edilir ve atanır.
8. Böylelikle cümleinin unsurları olan <özne>, <nesne>, <zarf tümleci>, <dolaylı tümleç>, <yüklem> sırasıyla kabul edilmiş olur ve bu unsurlardan oluşan yapı cümle olarak kabul edilir ve atanır.

Slide 31

Sözdizimsel analiz, cümlelerin yapısal tanımını oluşturabilmek için morfolojik analizin sonuçlarını kullanır.

Amacı, arka arkaya gelen kelimeler yığını bu kelimeler yığınının ifade ettiği cümle birimlerini tanımlayan bir yapıya dönüştürmektir.

Cümle parçalarının doğru tespiti, anlam analizi için yararlı sonuçlar verir.

Gramer ve sözlükler, bilgisayardaki dilbilgisi modelini karakterize eder, algoritmik olarak işlenmesi de problemi matematiksel açıdan ifade eder.

Slide 32

Semantic (Anlamsal) Analiz

- Anlamsal analiz, kelimelerin ayrı ayrı veya birbirine bağlanmış kelimelerin oluşturduğu anlama karşılık doğru modelleme yapıları oluşturmaktır.
- Anlamsal Analiz iki parçaya ayrılır: sözcüksel (lexical) anlam ve birleşim (combination) anlamı.

Sözcüksel anlam: üst kavram (*hyperonymy*), alt kavram (*hyponymy*), zıt anlamlılık (*antonymy*), parçanın bütünü (*holonymy*), eş anlamlılık (*synonymy*), bütünün parçası (*homonymy*), çok anlamlılık (*polysemy*) and eş seslilik (*homophony*).

Gül, sevdiğim bir çiçektir.

Fıkıraya herkes güldü.

Compositionality: bütünün anlamı, parçaların anlamından daha farklı olabilir.

Deyimler: En az iki sözcükten meydana gelen ve sözcüklerden en az birisi mecaz anlamıyla kullanılan, cümlede eylem bildiren söz öbekleridir. Deyimi oluşturan sözcükler çoğu zaman kendi anlamlarından uzaklaşmıştır. “Haber duyunca etekleri zil çaldı.” cümlesinde “etekleri zil çalmak” çok sevinmek anlamına gelen bir deyimdir. Ancak burada etek, zil, çalmak sözlerinin sevinmekle bir ilgisi yoktur.

Slide 33