



Dilbilgisi ve Diller

Doç.Dr.Banu Diri



1. Her biçimsel dil belirli bir alfabe üzerinde tanımlanır.
2. Alfabe sonlu sayıda simgelerden oluşan bir kümedir.
3. Alfabedeki simgelerin arka arkaya getirilmesi ile dizgiler (string) oluşturulur.
4. Biçimsel dil, bir alfabedeki simgelerden oluşturulan dizgilerin bir kümesidir. Bu kümeyi E ile gösterirsek, bu alfabe üzerinde tanımlanan her dil E 'nin bir alt kümesidir. E deki her dizgi dilin tümcelerini (sentence) oluşturur.
5. Bir alfabe üzerinde tanımlanan biçimsel bir dil, bu alfabedeki simgelerden oluşan dizgileri *geçerli* ve *geçersiz* diye ikiye ayırır. Dilde yer alan ve tümceleri oluşturan dizgiler *geçerli* tersi ise *geçersiz*'dir.

Slide 1

-
6. Biçimsel dil açısından dizgi, tümce ve sözcük terimleri birbirlerinin yerine kullanılabilir. Tümce dilde yer alan dizgi veya sözcükleri anlatmak için kullanılır. Buna göre, bir alfabe ve bu alfabe üzerinde tanımlı bir dil düşünüldüğünde, alfabedeki simgelerden oluşturulan ve dilde yer alan geçerli dizgiler dilin tümcelerini oluşturmaktadır.
 7. Dilin hangi tümcelerden oluştuğunu gösteren kurallar bütünü ise dilbilgisi (grammar) olarak adlandırılır.

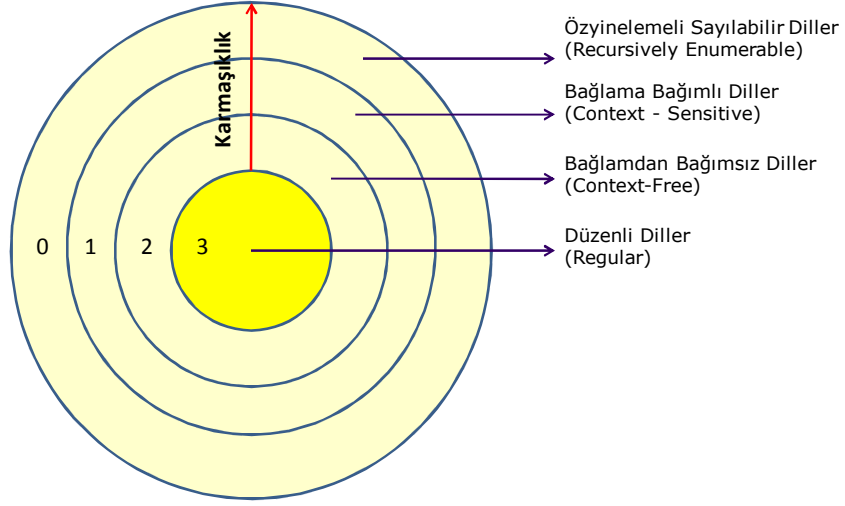
Slide 2

Biçimsel dilbilgisi ve dillerin incelenmesinde, değişik harf grupları değişik alanlarda kullanılır.

Harf Grubu	Örnekler	Kullanım alanları
Latin alfabesinin başındaki büyük harfler	A,B,C,...	Sözdizim değişkenleri
Latin alfabesinin başındaki küçük harfler ve rakamlar	a,b,c,...,0,1,2	Uç simgeler
Latin alfabesinin sonundaki büyük harfler	U,V,W,Y,...	Sözdizim değişkeni ya da uç simgeler
Latin alfabesinin sonundaki küçük harfler	u,v,w,y,...	Uç simge dizgileri (sözcükler)
Yunan alfabesinin başındaki küçük harfler	α , β , γ ,...	Tümcesel yapılar

Slide 3

Chomsky Hiyerarşisi



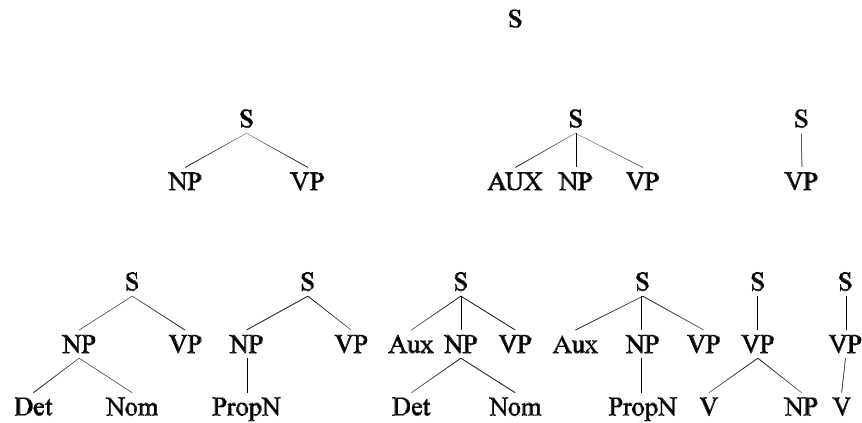
Slide 4

Sırasıyla;

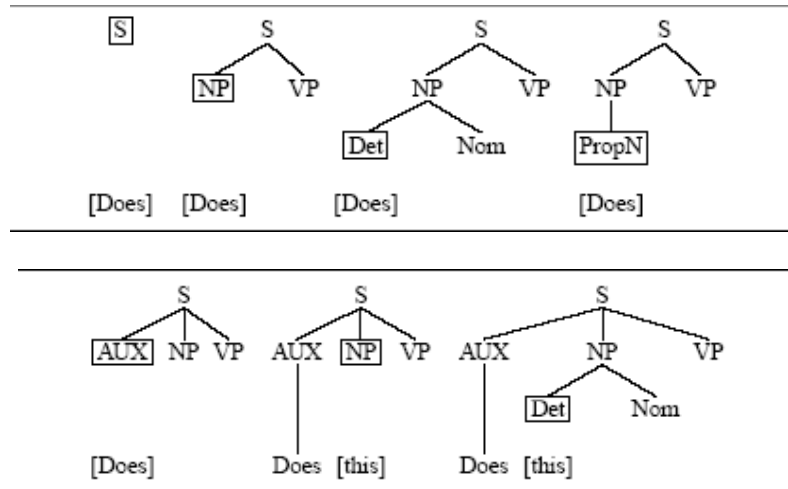
- Dilbilgisi ve Dilin Biçimsel Tanımı
- Dilbilgisi ve Dillerin Sınıflandırılması
tür-0, tür-1, tür-2, tür-3
- Sağ-doğrusal ve Sol-doğrusal Dilbilgisi
- Bağlamdan-Bağımsız Dilbilgisi (Context Free Grammar-CFG)
- Türetme/Ayrıştırma Ağacı
Chomsky Normal Biçimi
Yukarıdan-aşağıya Ayrıştırma (Top-down parsing)
Aşağıdan-yukarıya Ayrıştırma (Bottom-up parsing)

Slide 5

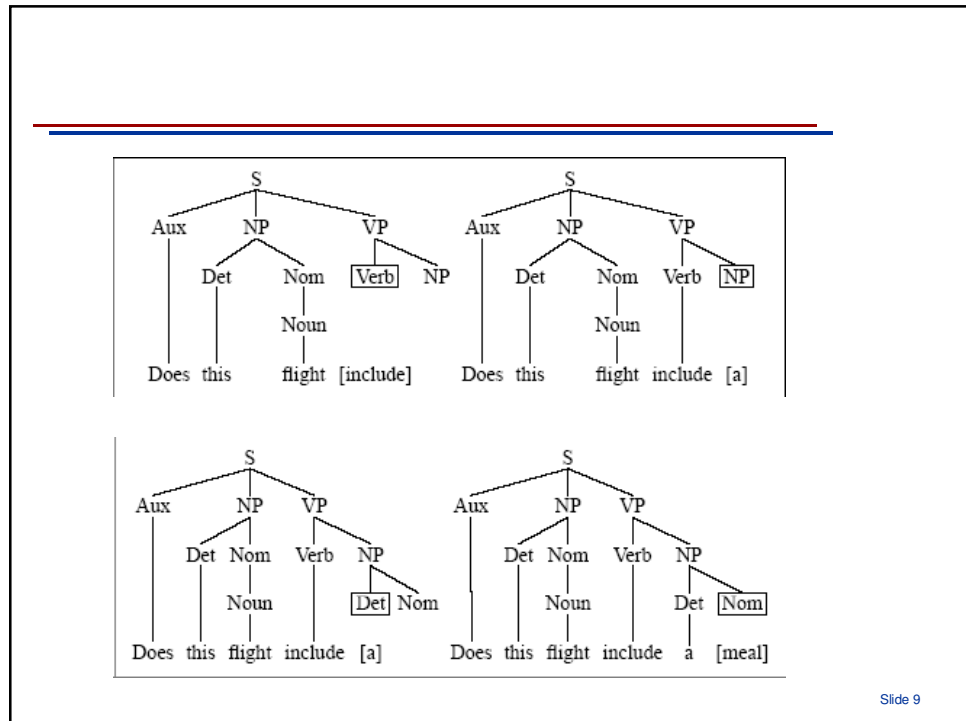
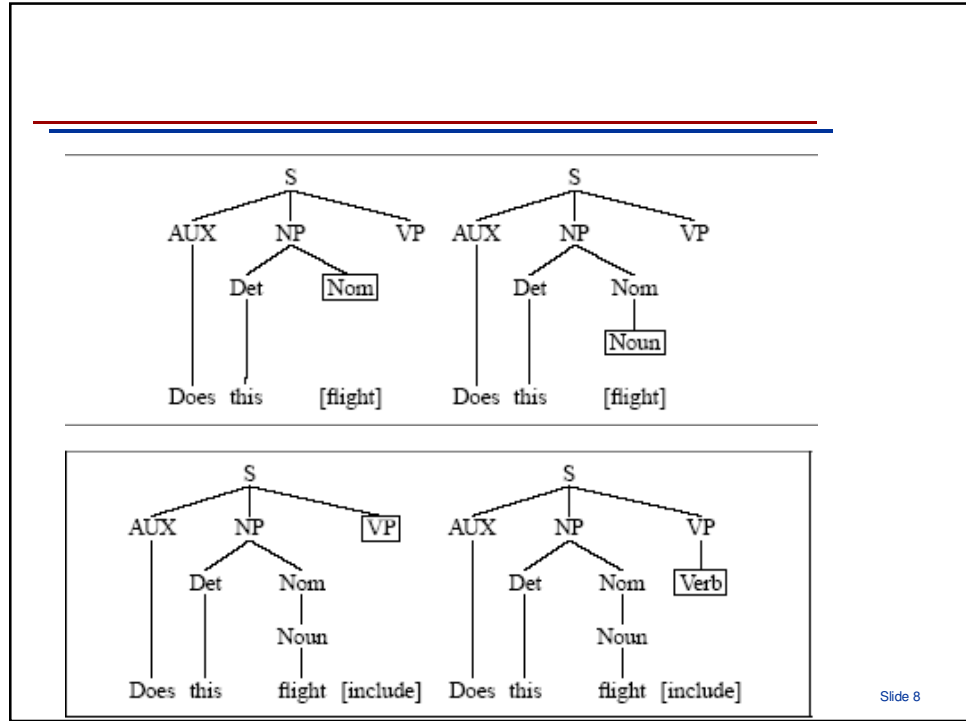
Yukarıdan-Aşağıya Ayrıştırma (Top-Down Parsing, Left-to-Right, Depth-First)

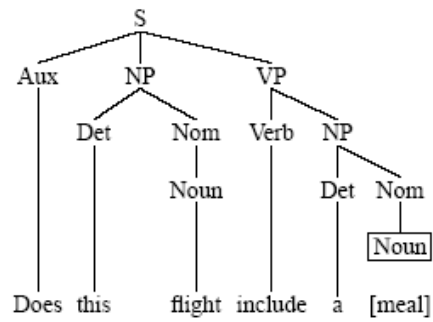


Slide 6

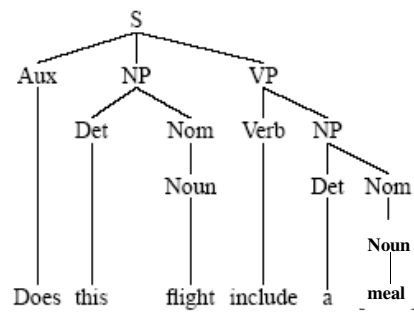


Slide 7



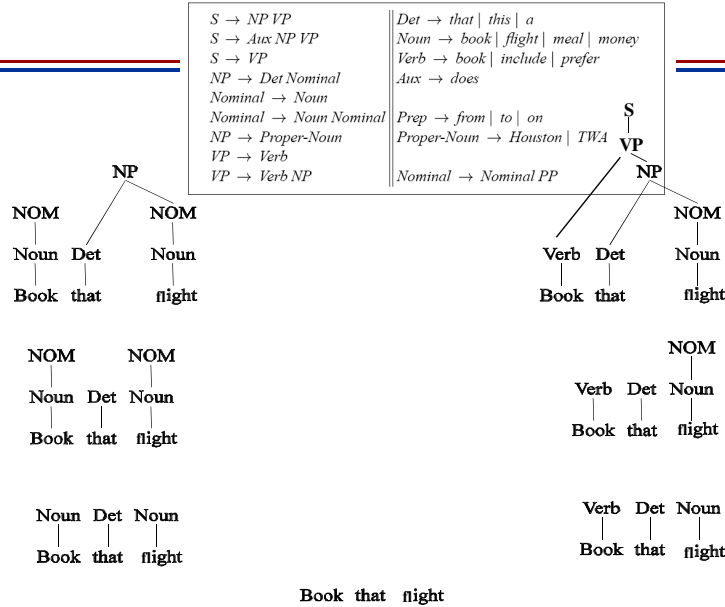


Slide 10



Slide 11

Aşağıdan-Yukarı Ayrıştırma (Bottom-Up Parsing)

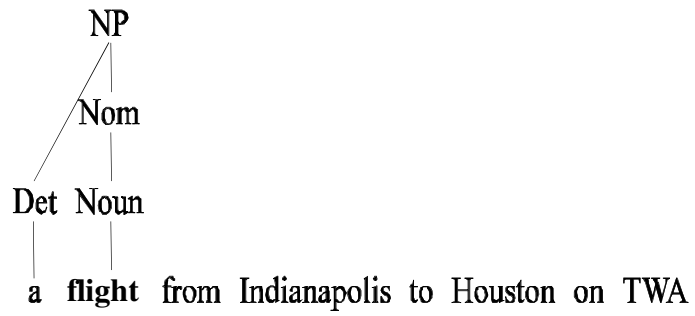


Slide 12

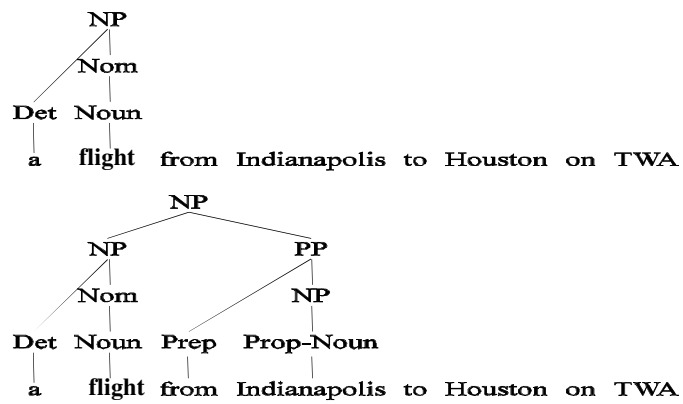
S -> Aux NP VP
 S -> NP VP
 NP -> Det Nominal
 NP -> NP PP

A flight from Indianapolis to Houston on TWA

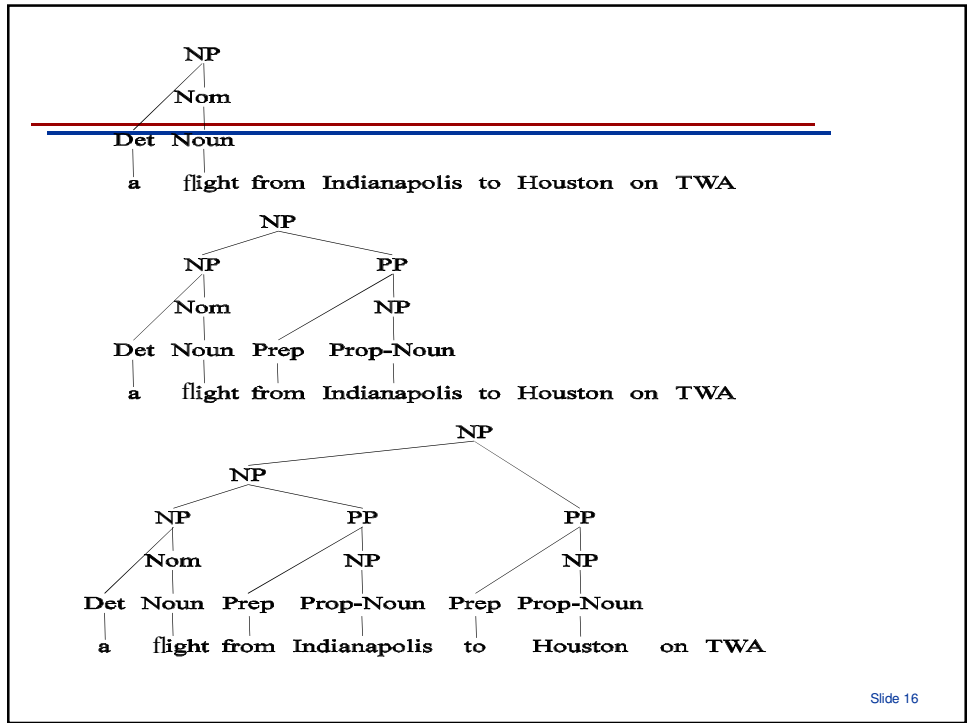
Slide 13



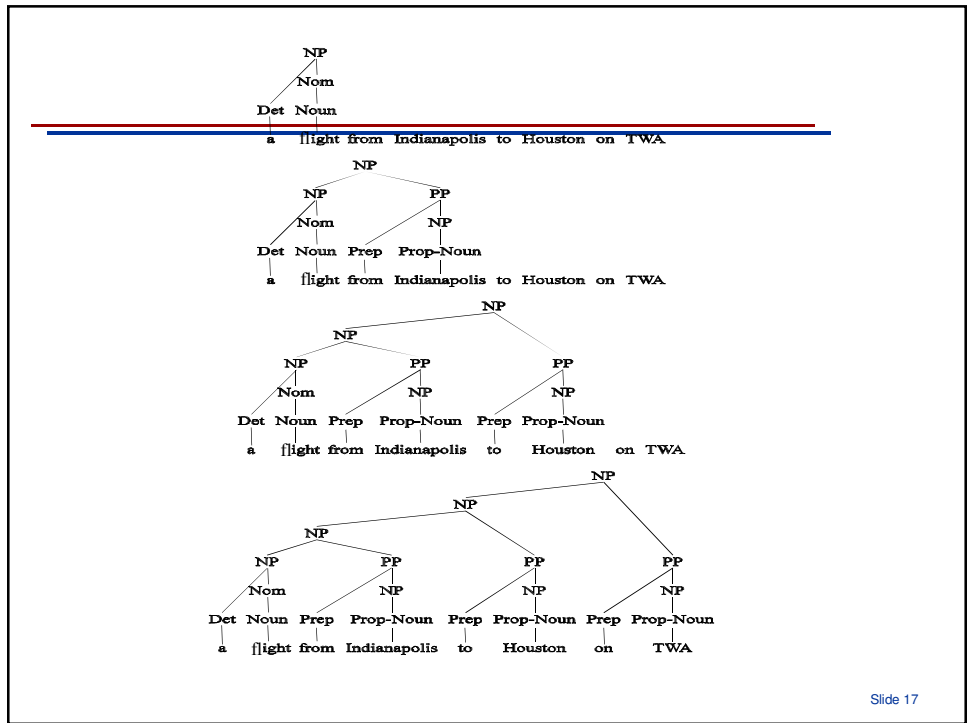
Slide 14



Slide 15



Slide 16



Slide 17

Örnek “Bir cümle isim veya fiil grubundan oluşur.”, “isim grubu isim ve/veya sıfat, fiil grubu da isim grubu ve/veya fiil den oluşur”.

Bu gramerde “Küçük çocuk kırmızı top aldı” cümlesinin çözümü

$C \rightarrow \dot{I}G \text{ } FG$: (cümle isim ve fiil grubundan oluşur)

$\dot{I}G \rightarrow S \dot{I}$: (isim grubu \rightarrow sıfat+isim)

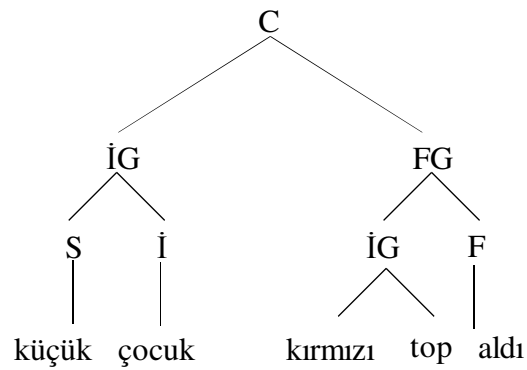
$FG \rightarrow \dot{I}G \text{ } F$: (fiil grubu \rightarrow isim grubu + fiil)

$S \rightarrow$ küçük|kırmızı

$\dot{I} \rightarrow$ çocuk|top

$F \rightarrow$ aldı

Slide 18



Slide 19

Chomsky Normal Form (CNF)

is one of the most basic Normal Forms.

In CNF each production has the form

$$A \rightarrow BC$$

$$A \rightarrow \alpha$$

where $A, B, C \in V$ (i.e., nonterminals)

and $\alpha \in \Sigma$ (i.e., a terminal)

This can be done by simple substitution.

Consider the following

$$S \rightarrow bA$$

$$S \rightarrow aB$$

$$A \rightarrow bAA$$

$$A \rightarrow aS$$

$$A \rightarrow a$$

$$B \rightarrow aBB$$

$$B \rightarrow bS$$

$$B \rightarrow b$$

Replace terminals with NEW Nonterminals

Plus a rule to generate the Terminal

$$C \rightarrow a$$

$$D \rightarrow b$$

Note these are already in standard form.

Terminals should now only appear in rules in CNF.

Rewriting the rules yields:

$$S \rightarrow DA$$

$$S \rightarrow CB$$

$$A \rightarrow DAA$$

$$A \rightarrow CS$$

$$A \rightarrow a$$

$$B \rightarrow CBB$$

$$B \rightarrow DS$$

$$B \rightarrow b$$

$$C \rightarrow a$$

$$D \rightarrow b$$

Slide 21

Only the following rules are a problem:

$A \rightarrow DAA$

$B \rightarrow CBB$

These can be rewritten as follows:

$A \rightarrow ZA$

$Z \rightarrow DA$

$B \rightarrow KB$

$K \rightarrow CB$

Now we have rules in CFN.

SO WHAT?

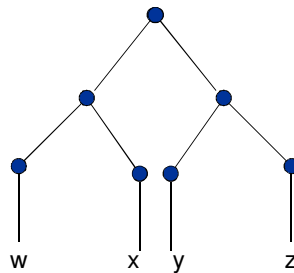
For grammars in Chomsky Normal Form the parse tree is always a binary tree.

We can talk about the relationship between:

- 1) the depth of the parse tree
- and 2) the length of its yield.

Slide 22

If a parse tree for a word w is generated by a CNF and the parse tree has a path length of at most i , then the length of w is at most 2^{i-1} .



A parse tree of depth 3 with a yield of at most 2^{3-1} .

Slide 23

Genişletilmiş Geçiş Ağları – GGA Augmented Network Transition Grammer- ATN

- Biçimsel dillerin, doğal dilleri tam olarak ifade edemediği durumlarda Genişletilmiş Geçiş Ağları tercih edilir.
- GGA, cümlelerin sözdizimsel analizinde kullanılır.
- GGA sonlu makinelere benzeyen durum ve bu durumlar arası geçiş kurallarından oluşmaktadır.

Üç bileşenden oluşur:

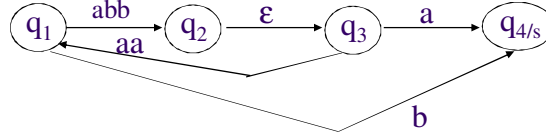
- En az bir başlangıç ve son durumu olan sonlu sayıdaki durumlar kümesi
- Belli bir metindeki mümkün olan harflerden oluşan küme
- Sonlu sayıdaki bir durumdan diğer bir duruma geçişi sağlayan geçişler kümesi

Slide 24

-
- ❖ Geçiş ağlarında bir durumdan diğer bir duruma geçmek için gerekli harf okunur ve bu harf geçilecek olan duruma geçmek için gereken harfle karşılaştırılır. Uygun olması durumunda geçilir.
 - ❖ Geçiş ağlarında doğru bir yol, bir başlangıç durumundan başlayıp, son duruma ulaşan geçişler sağlandığında tamamlanır.
 - ❖ Geçişlerdeki harflerin birbirine eklenmesiyle oluşan metin, ağın kabul etmesi için verilen metinle aynı ise, bu metin ağ tarafından kabul edilmiş demektir.

Slide 25

Geniřletilmiş Geiř Ađı rneđi



Bu ađda tanınabilecek birkaç metin rneđi:

abba, abbbaabba, abbaab, b

Fakat *abbab, baab* yi tanıyamaz.

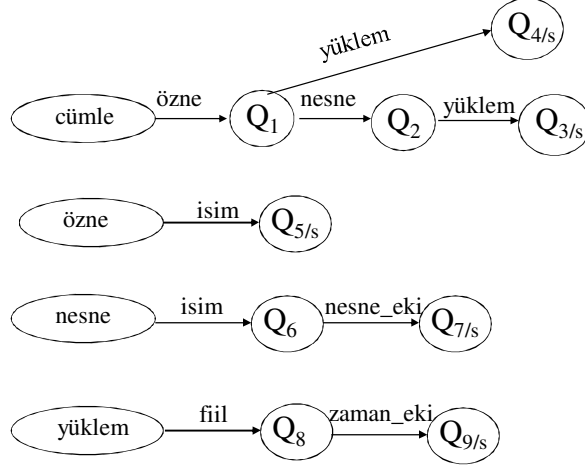
Slide 26

- ❖ Dilin grameri, durumlara karřı dūřen ynlū graflar řeklinde ifade edilir.
- ❖ Bađlarda ise, ođalmakta olan durumlar arasındaki geiřleri tanımlayan etiket sınıfları bulunmaktadır.
- ❖ Bu etiketler zel kelimeler, kelime kategorileri veya cūmlenin nemli paralarını tanıyan diđer ađlara ynlendirme gibi deđerler almaktadır.
- ❖ GGA, dil tanımında belirtilen cūmler birimlerini, hiyerarřik yapı ierisinde yukarıdan ařađıya dođru zmlmeye alıřır.

Slide 27

Basit bir Türkçe sözdizimi kural kümesinin GGA yapısı

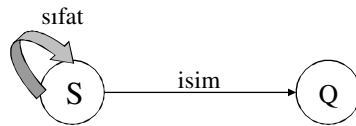
“Ali camı kırdı” cümlesi morfolojik analiz sonucunda “Ali cam + ı kır + dı”



Slide 28

- ❖ Bir sözdizimi kural kümesinin GGA biçiminde ifade edilmesin yararı, cümle birimlerinin bir kez tanımlandıktan sonra sonsuz kez kullanılıyor olmasıdır.
- ❖ Bir sıfat tamlaması bir kez tanımlanıp birçok yerde kullanılabilir.
- ❖ GGA'da döngüler bulunabilir. Bu özellik ile dilin kabul ettiği cümle yapıları sayısı sonsuza kadar artırılabilir.

“yırtık kırmızı büyük top”



Slide 29