



Grammer ve Diller

Prof.Dr. Banu Diri



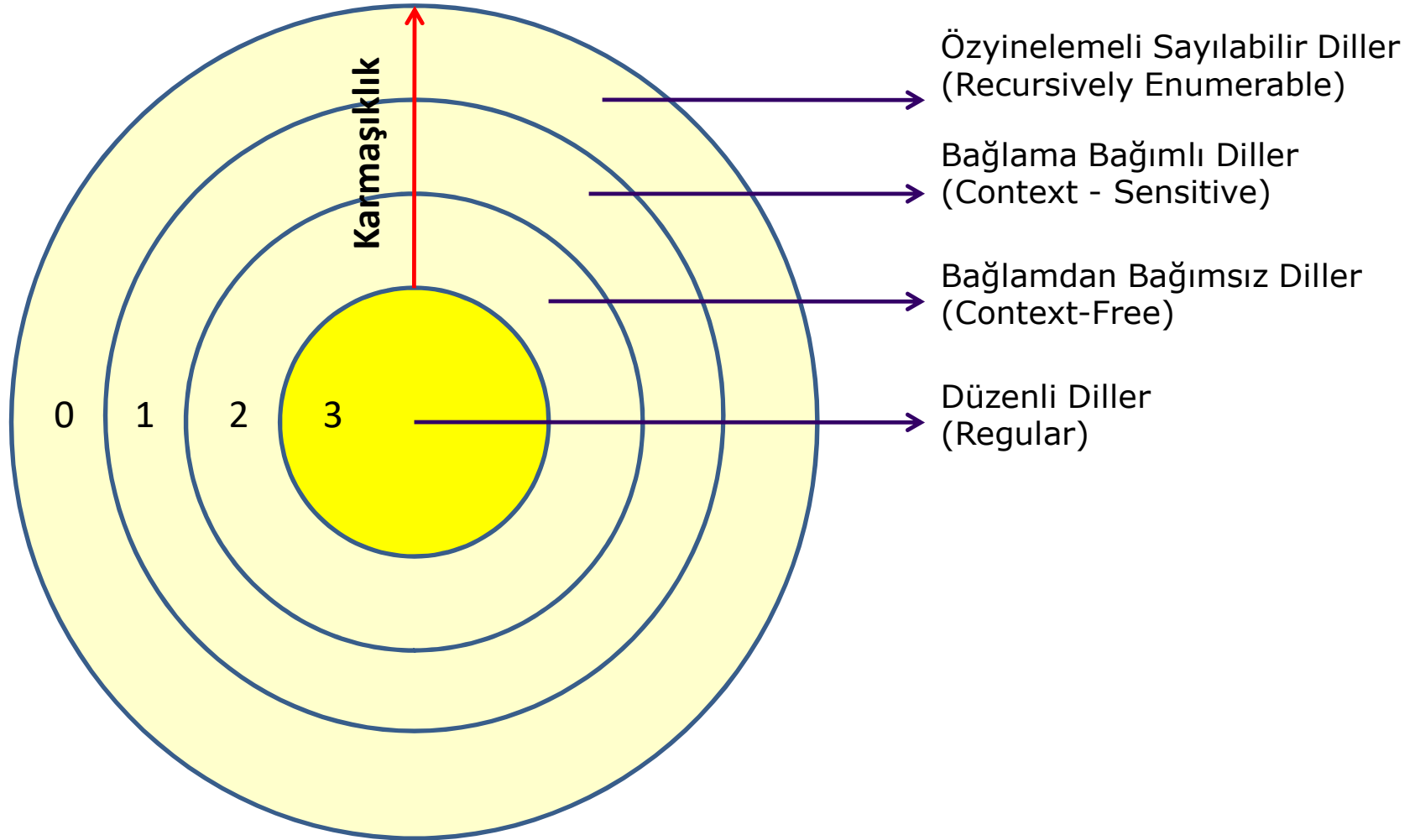
-
1. Her biçimsel dil belirli bir alfabe üzerinde tanımlanır.
 2. Alfabe sonlu sayıda simgelerden oluşan bir kümedir.
 3. Alfabadeki simgelerin arka arkaya getirilmesi ile dizgiler (string) oluşturulur.
 4. Biçimsel dil, bir alfabadeki simgelerden oluşturulan dizgilerin bir kümesidir. Bu kümeyi E ile 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 alfabadeki simgelerden oluşan dizgileri *geçerli* ve *geçersiz* diye ikiye ayırır. Dilde yer alan ve tümceleri oluşturan dizgiler *geçerli* tersi ise *geçersiz*'dir.

-
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, alfabadeki simgelerden oluşturulan ve dilde yer alan geçerli dizgiler dilin tümcelerini oluşturmaktadır.
 7. Dilin hangi tümcelerden oluştuğunu gösteren kurallar bütünü ise dilbilgisi (grammar) olarak adlandırılır.

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

Chomsky Hiyerarşisi



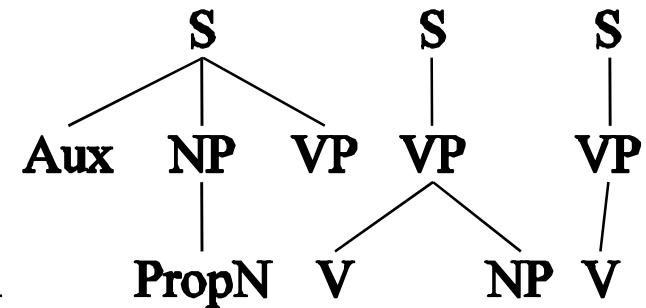
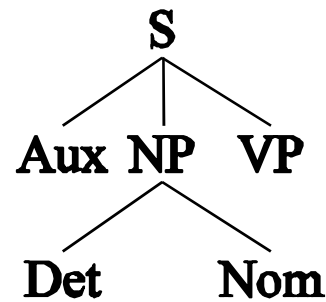
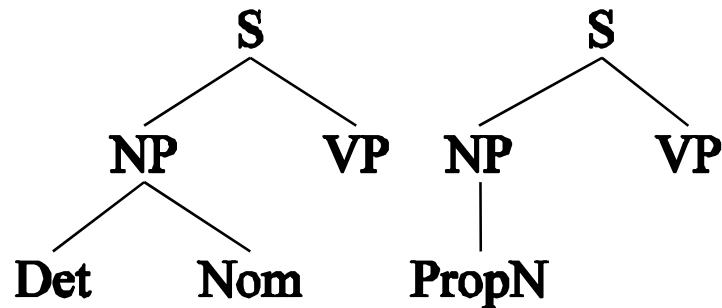
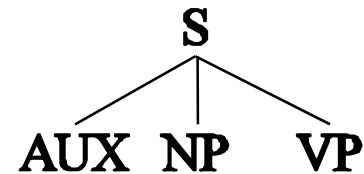
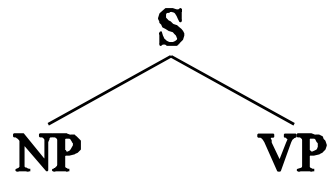
Sırasıyla;

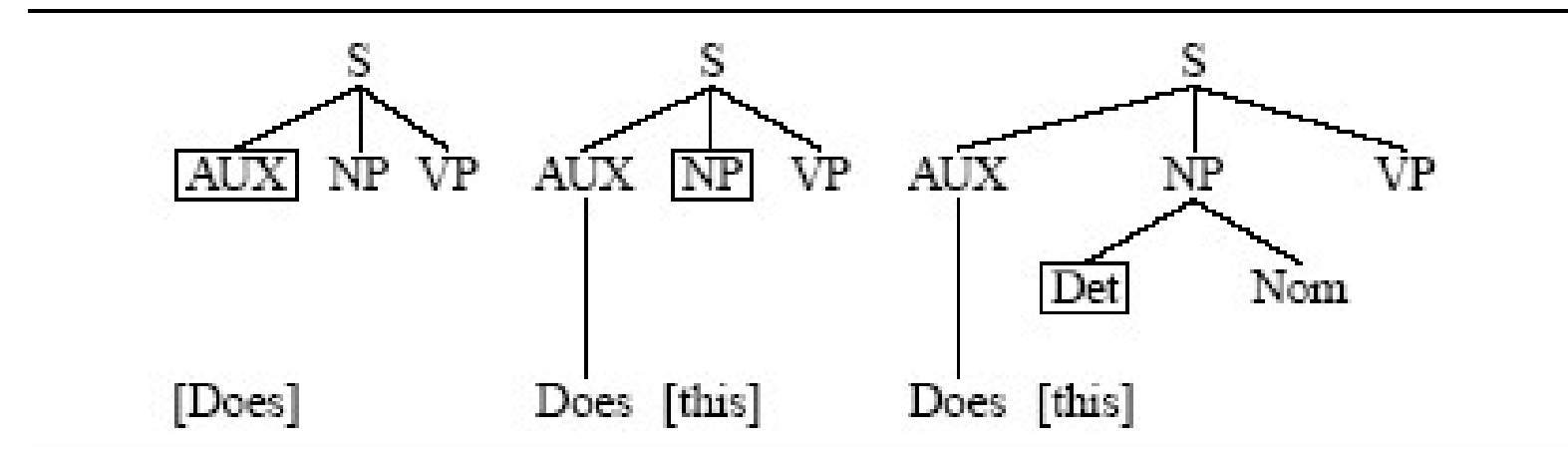
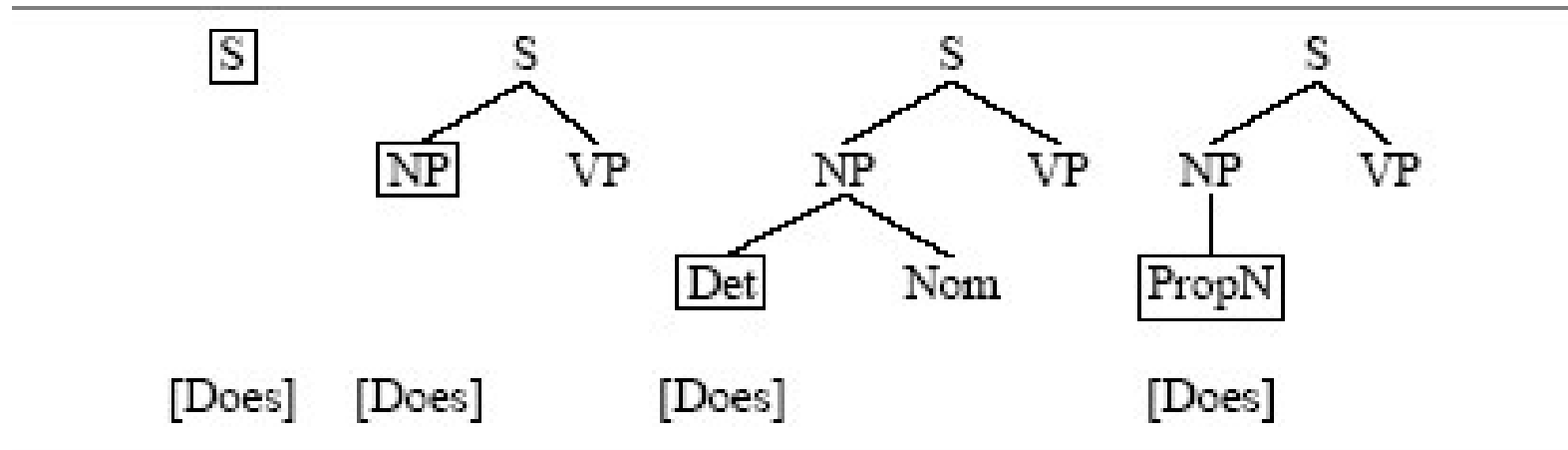
- Dilbilgisi ve Dilin Biçimsel Tanımı
- Dilbilgisi ve Dillerin Sınıflandırılması
tür-0, tür-1, tür-2, tür-3
- Sağ-doğrusal ve Sol-doğrusal Dilbilgisi
- Bağlamdan-Bağımsız Dilbilgisi (Context Free Grammar-CFG)
- Türetme/Ayrıştırma Ağacı
Chomsky Normal Biçimi
Yukarıdan-aşağıya Ayrıştırma (Top-down parsing)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

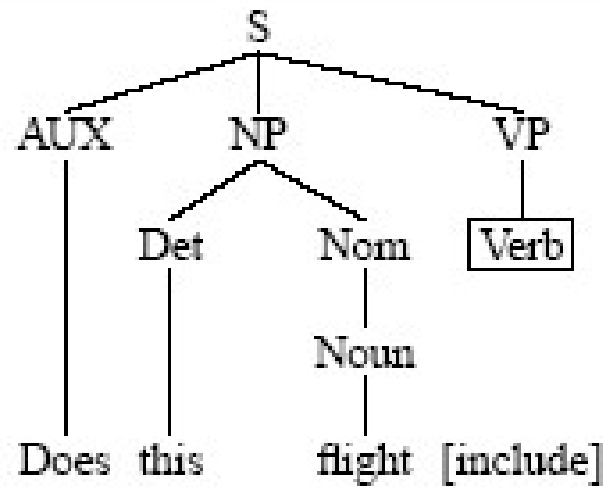
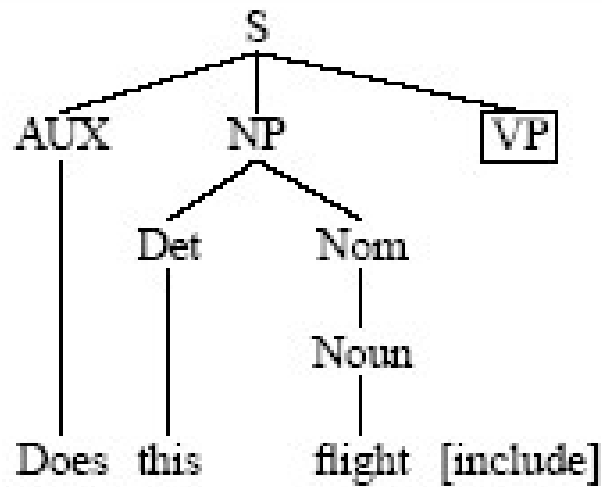
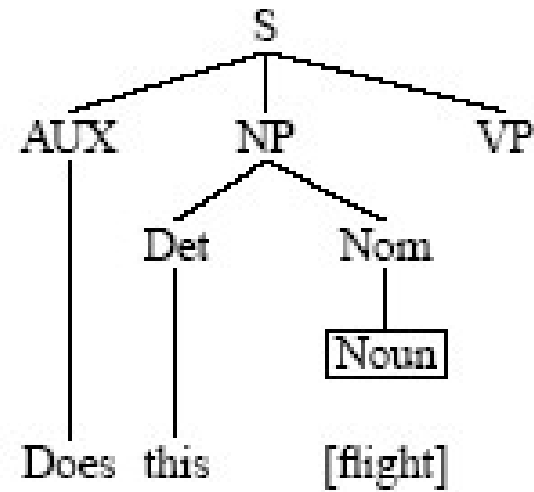
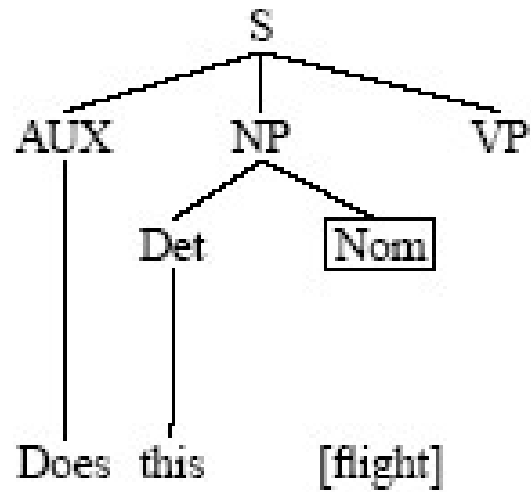
tür-0, tür-1, tür-2, tür-3
Örnekler ile anlatılacaktır

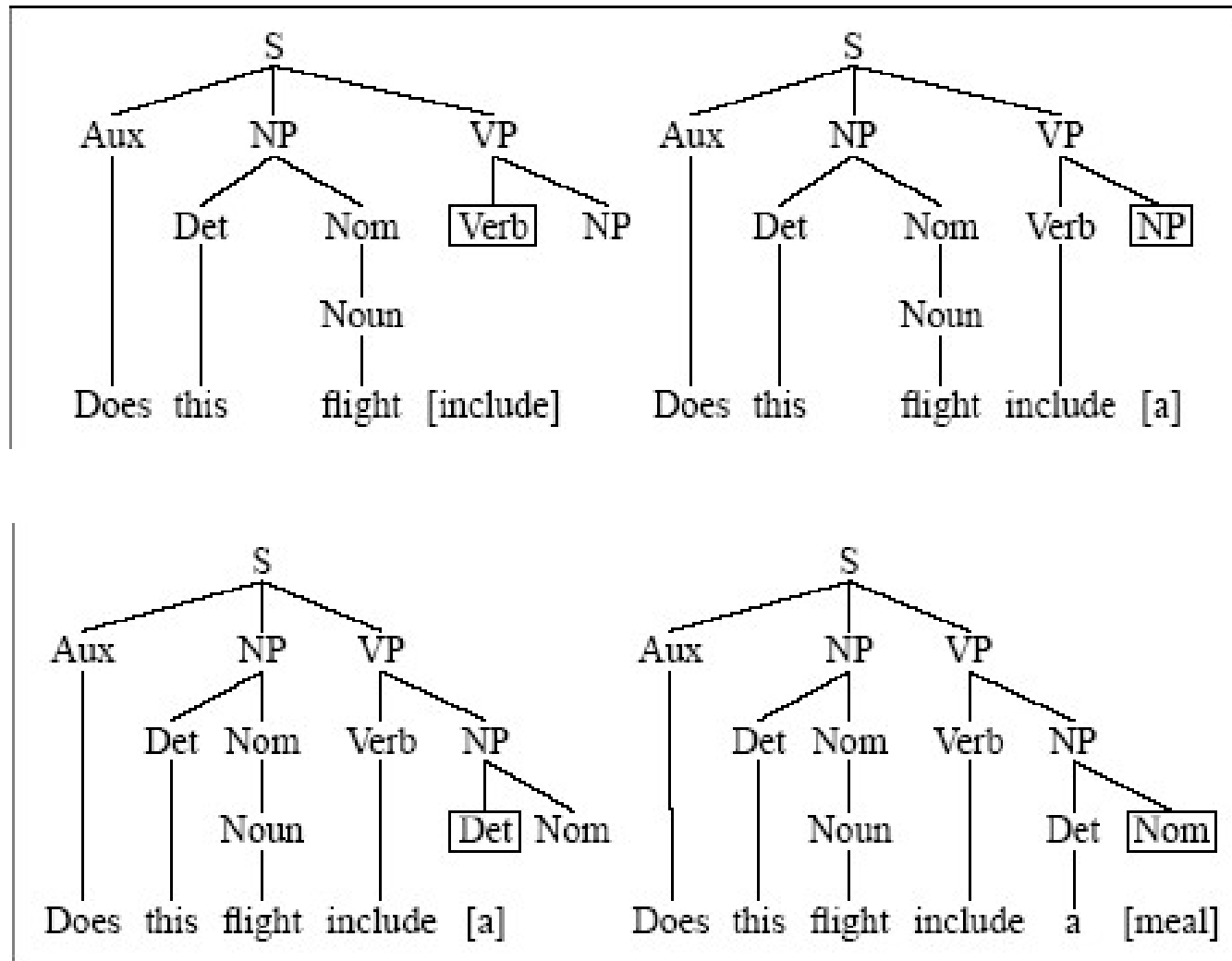
Yukarıdan-Aşağıya Ayırıştırma (Top-Down Parsing, Left-to-Right, Depth-First)

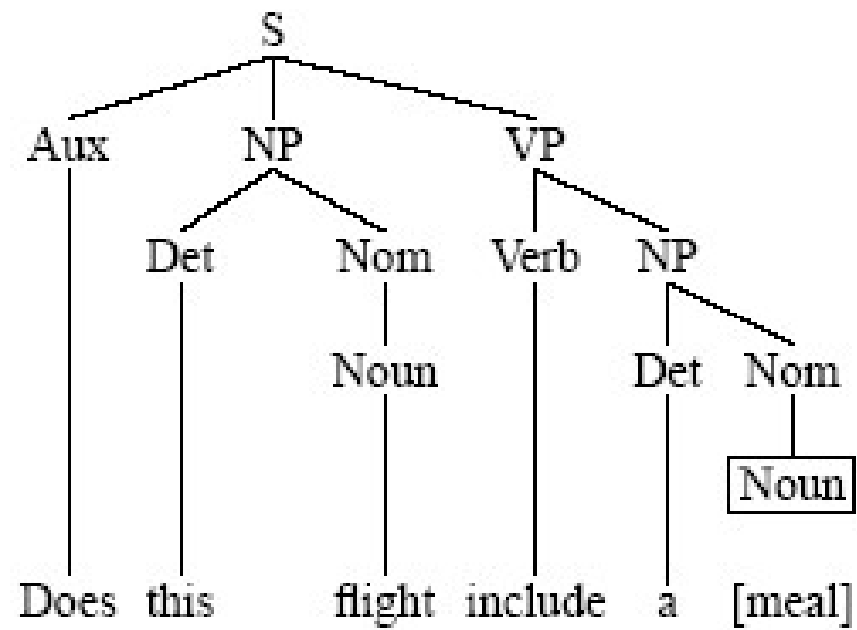
S

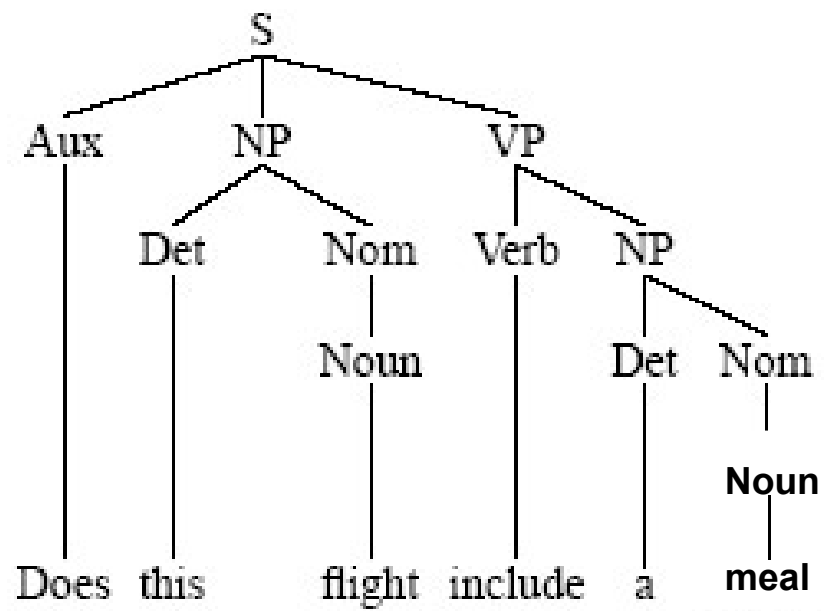












Aşağıdan-Yukarı Ayırıştırma (Bottom-Up Parsing)

$S \rightarrow NP VP$

$S \rightarrow Aux NP VP$

$S \rightarrow VP$

$NP \rightarrow Det Nominal$

$Nominal \rightarrow Noun$

$Nominal \rightarrow Noun Nominal$

$NP \rightarrow Proper-Noun$

$VP \rightarrow Verb$

$VP \rightarrow Verb NP$

$Det \rightarrow that | this | a$

$Noun \rightarrow book | flight | meal | money$

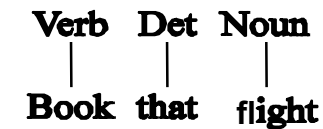
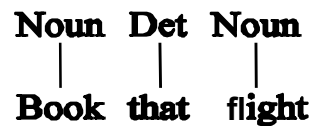
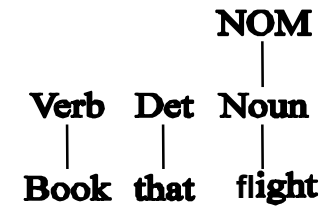
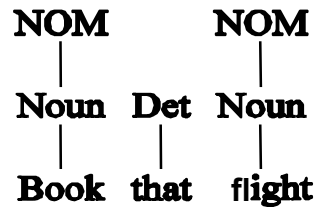
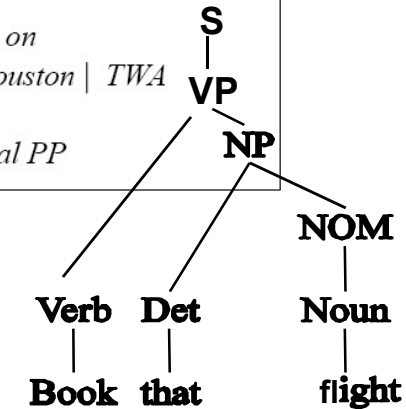
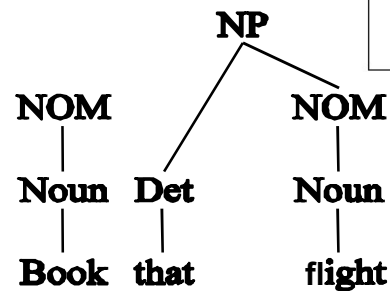
$Verb \rightarrow book | include | prefer$

$Aux \rightarrow does$

$Prep \rightarrow from | to | on$

$Proper-Noun \rightarrow Houston | TWA$

$Nominal \rightarrow Nominal PP$



Book that flight

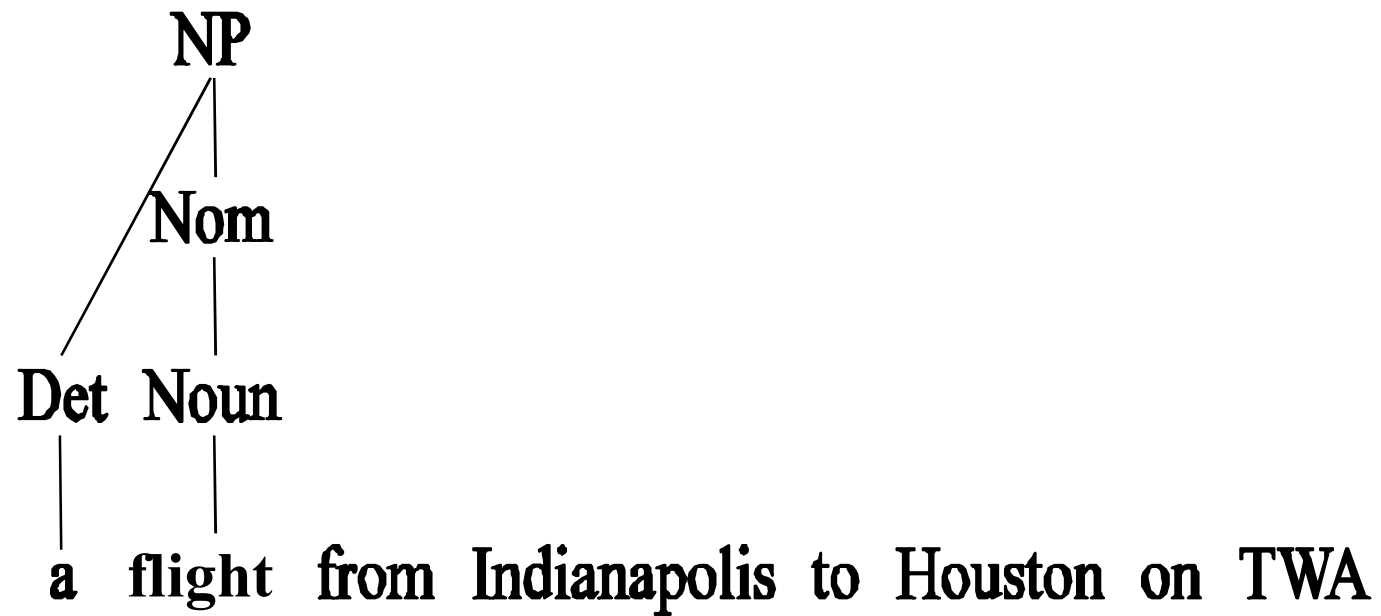
S -> Aux NP VP

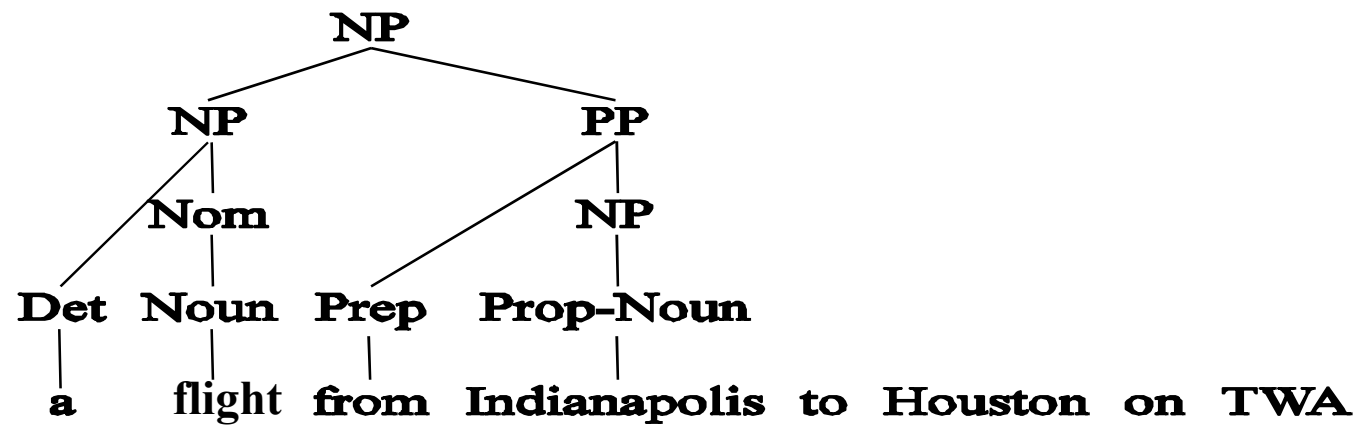
S -> NP VP

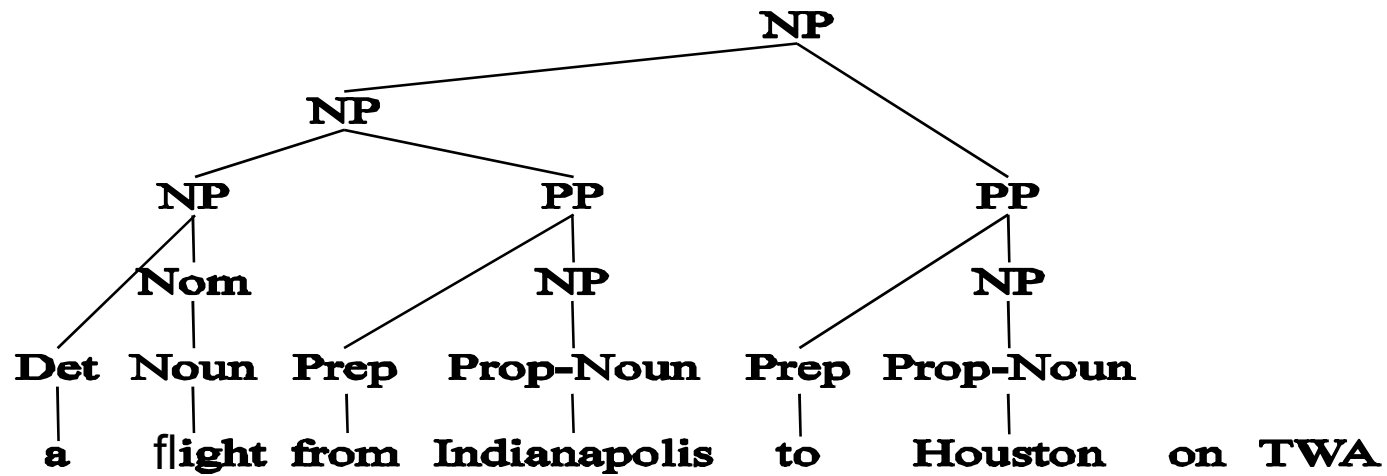
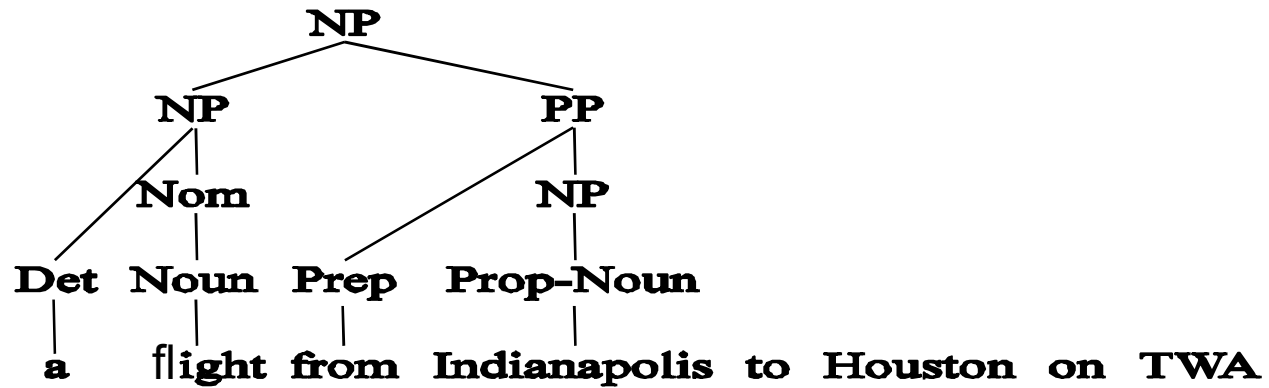
NP -> Det Nominal

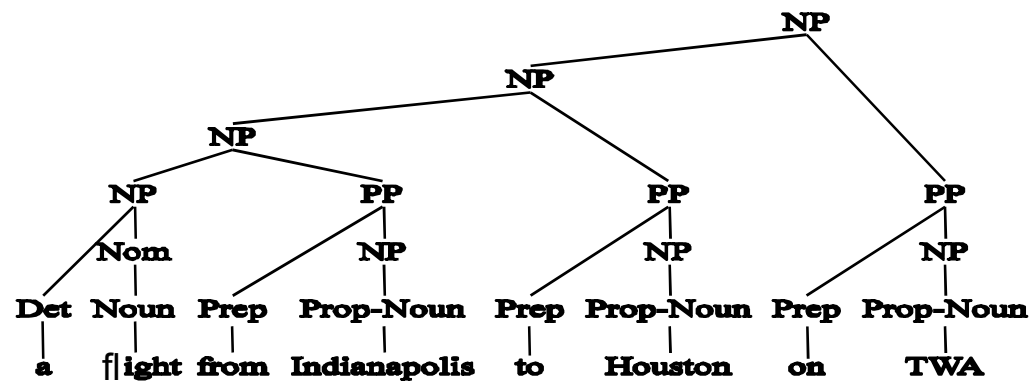
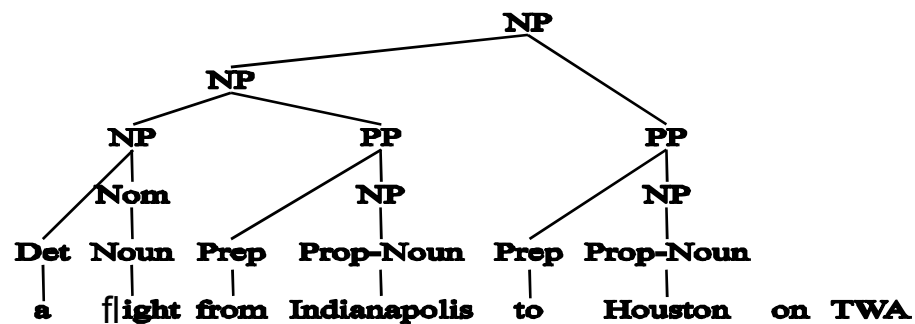
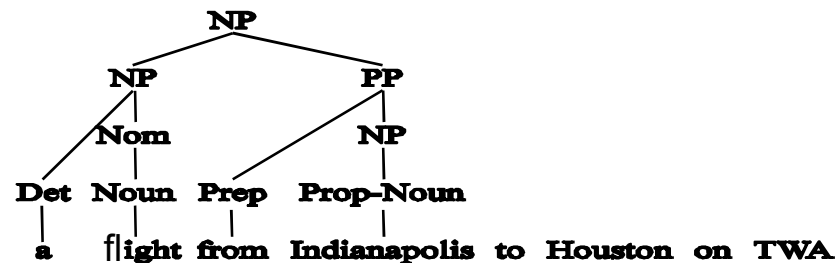
NP -> NP PP

A flight from Indianapolis to Houston on TWA









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

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

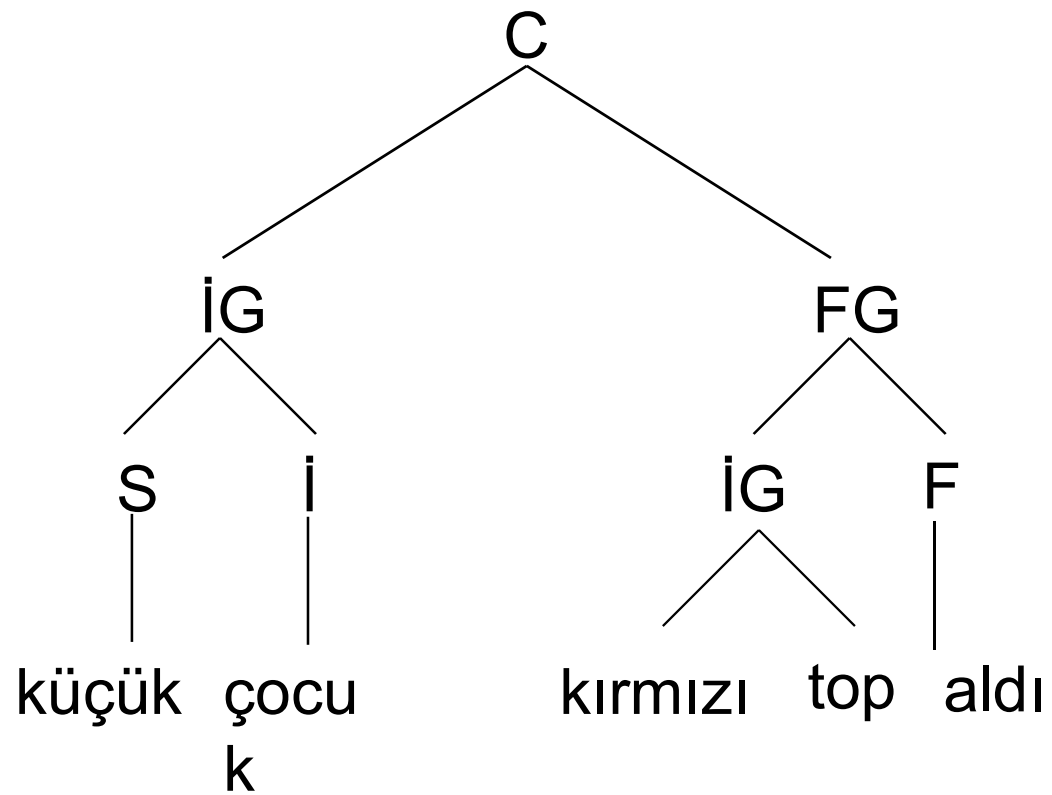
$\dot{I}G \rightarrow S \ \dot{I}$: (isim grubu $\rightarrow$ sıfat+isim)

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

$S \rightarrow$ küçük|kırmızı

$\dot{I} \rightarrow$ çocuk|top

$F \rightarrow$ aldı



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:

$$S \rightarrow DA$$
$$S \rightarrow CB$$
$$A \rightarrow DAA$$
$$A \rightarrow CS$$
$$A \rightarrow a$$
$$B \rightarrow CBB$$
$$B \rightarrow DS$$
$$B \rightarrow b$$
$$C \rightarrow a$$
$$D \rightarrow b$$

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.

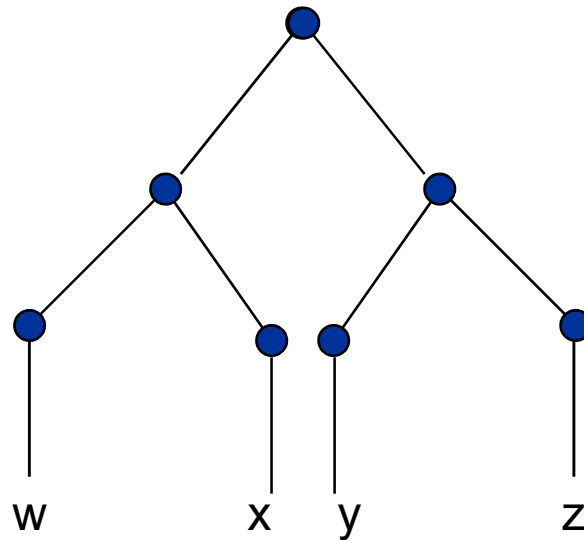
SO WHAT?

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.

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} .

Genişletilmiş Geçiş Ağları – GGA

Augmented Network Transition Grammer- ATN

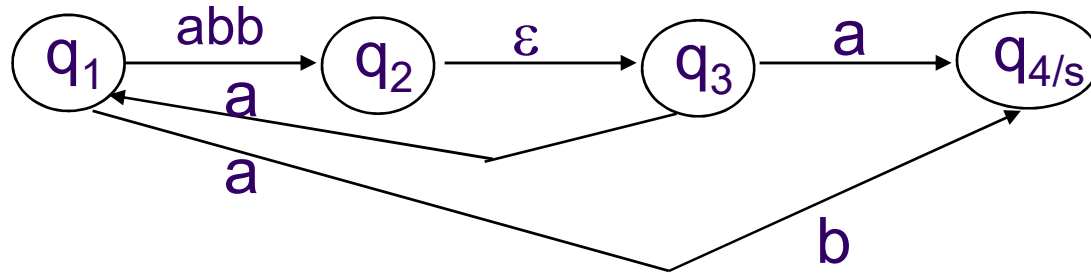
- Biçimsel dillerin, doğal dilleri tam olarak ifade edemediği durumlarda Genişletilmiş Geçiş Ağları tercih edilir.
- GGA, cümlelerin sözdizimsel analizinde kullanılır.
- GGA sonlu makinelere benzeyen durum ve bu durumlar arası geçiş kurallarından oluşmaktadır.

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

-
- ❖ 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.

Genişletilmiş Geçiş Ağı örneği



Bu ağda tanınabilecek birkaç metin örneği:

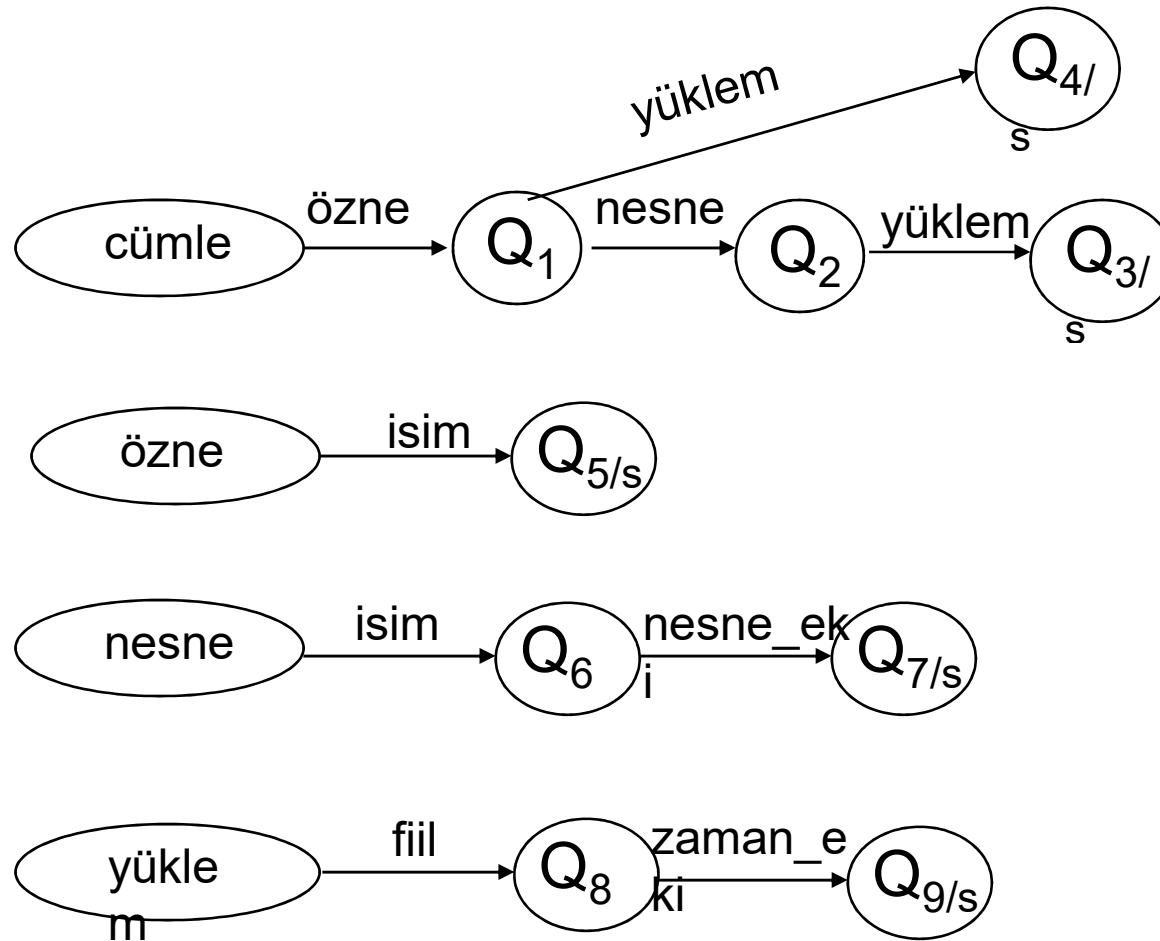
abba, abbaaabba, abbaab, b

Fakat *abbab, baab* yi tanıyamaz.

-
- ❖ Dilin grameri, durumlara karşı düşen yönlü graflar şeklinde ifade edilir.
 - ❖ Bağlarda ise, çoğalmakta olan durumlar arasındaki geçişleri tanımlayan etiket sınıfları bulunmaktadır.
 - ❖ Bu etiketler özel kelimeler, kelime kategorileri veya cümlelerin önemli parçalarını tanıyan diğer ağlara yönlendirme gibi değerler almaktadır.
 - ❖ GGA, dil tanımında belirtilen cümle birimlerini, hiyerarşik yapı içerisinde yukarıdan aşağıya doğru çözümlemeye çalışır.

Basit bir Türkçe sözdizimi kural kümesinin GGA yapısı

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



-
- ❖ 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 artırabilir.

“yırtık kırmızı büyük top”

