

93760

YILDIZ TEKNİK ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ

TÜRKÇE METİNLERİN BİLGİSAYAR YARDIMI İLE  
ANALİZ EDİLMESİ VE YAZARININ BULUNMASI

Bilgisayar Müh. Şeniz GAYDE

F.B.E. Bilgisayar Bilimleri Mühendisliği Anabilim Dalında Hazırlanan

YÜKSEK LİSANS TEZİ

Tez Danışmanı : Prof. M. Yahya KARSLIGİL

Prof. M. Yahya Karşılığın İzniyle  
Prof. Dr. Oruç Bilgiç'in İzniyle  
Doç. Dr. Selim Akçelik'in İzniyle

İSTANBUL, 2000

## İÇİNDEKİLER

KISALTMA LİSTESİ.....	iii
ŞEKİL LİSTESİ.....	iv
ÇİZELGE LİSTESİ.....	v
ÖNSÖZ.....	vi
ÖZET.....	vii
ABSTRACT.....	viii
1. GİRİŞ.....	1
2. DOĞAL DİL İŞLEME.....	2
2.1 Doğal Dil İşleme Uygulamaları.....	3
3. TÜRKÇE.....	4
3.1 Biçimbilimsel Özellikleri.....	4
3.2 Sözcük Dizimi.....	6
3.2.1 Cümlelerin öğeleri.....	7
3.2.1.1 Yüklem.....	7
3.2.1.2 Özne.....	8
3.2.1.3 Nesne.....	9
3.2.1.4 Tümleç.....	10
3.2.1.5 Sorulan bir sorunun yanıtı olan öğeler.....	13
3.3 Cümle.....	13
3.3.1 Anlamlarına göre cümleler.....	13
3.3.2 Yüklem türüne göre cümleler.....	14
3.3.3 Yüklem yerine göre cümleler.....	15
3.3.4 Yapılarına göre cümleler.....	15
4. DOĞAL DİLLERİN İFADE EDİLME ŞEKİLLERİ.....	18
4.1 Gramerlerin Chomsky Hiyerarşisi.....	20
5. DOĞAL DİLLERİN ÇÖZÜMLENMESİ.....	23
5.1 Sonlu Durum Teknolojisi.....	24
5.2 Genişletilmiş Geçişli Ağlar.....	26
5.2.1 Dönüşümsel gramer.....	26
5.2.2 ATN'lerin tarihçesi.....	27
5.2.3 Tekrarlayan geçişli ağlar.....	28
5.2.4 Genişletilmiş geçişli ağlar.....	30

5.2.5	Açıklayıcı ve prosedürel şekilciliğin karşılaştırılması .....	32
6.	İSTATİSTİKSEL YAKLAŞIM.....	34
6.1	Olasılık ve Bayes Teoremi .....	34
6.2	Kesinlik Faktörleri ve Kural-Tabanlı Sistemler .....	36
6.3	Bulanık Mantık.....	38
6.3.1	Yazarların Bulanık Mantık kümeleri.....	39
6.3.2	Bulanık üyelik derecesi .....	40
6.3.3	Üyelik derecesi değerleri.....	41
6.3.4	Yazarların üyelik derecelerinin elde edilmesi .....	42
6.3.5	Netleştirme birimi(Defuzzification).....	43
7.	UYGULAMA.....	45
7.1	Sembollere Ayrıştırıcı .....	46
7.2	Normalize Edici.....	46
7.3	Biçimbilimsel Çözümleyici.....	47
7.4	ATN Çözümleyici .....	48
7.5	Yazarın Tanınması İçin Kullanılan Sistem .....	48
8.	SONUÇ.....	50
	KAYNAKLAR.....	53
	ÖZGEÇMİŞ.....	54

## KISALTMA LİSTESİ

ATN	Augmented Transition Network(Genişletilmiş Geçişli Ağlar)
BTE	Bilgisayar Tabanlı Eğitim
RTN	Recursive Transition Network(Tekrarlayan Geçişli Ağlar)



## ŞEKİL LİSTESİ

Şekil 3.1	Türkçe' de sözcüklerin diziliş sırası	5
Şekil 4.1	Acaba kelimesi için yapım ağacı	22
Şekil 5.1	Türkçe' de sıfat isim ilişkisi için sonlu durum makinesi	25
Şekil 5.2	Türkçe' de sıfat isim ilişkisi için iki seviyeli sonlu durum makinesi	25
Şekil 6.1	Belirsiz kuralları birleştirmek	37
Şekil 6.2	Yazarın tanınması için kullanılan bulanık mantık sistemi	39
Şekil 6.3	Bir üyelik derecesi fonksiyonu	41
Şekil 6.4	Üyelik fonksiyonları (a) sıcaklık (b) nemlilik	42
Şekil 7.1	Türkçe metni analiz eden ve yazarını tanıyan sistemin yapısı	45
Şekil 7.2	Türkçe kelimeler için biçimbilimsel çözümleyici	47
Şekil 7.3	ATN için kullanılan cümle ağı	48



## ÇİZELGE LİSTESİ

Çizelge 8.1	Yedi yazar için cümle türüne göre metin çözümleme sonuçları	50
Çizelge 8.2	Yedi yazar için sözcük türüne göre metin çözümleme sonuçları	51
Çizelge 8.3	Yazarı bilinmeyen makalenin diğer yazarlara göre üyelik derecesi	52



## **ÖNSÖZ**

**Değerli görüşleri ve gösterdiği anlayış için tez danışmanım Sayın Prof. M. Yahya Karşılıgil' e sonsuz teşekkürlerimi sunarım.**



## ÖZET

### TÜRKÇE METİNLERİN BİLGİSAYAR YARDIMI ile ANALİZİ ve YAZARININ TANINMASI

Günümüzde doğal dil işlemenin önemli bir yönü, dil bilim kuramlarına deney ortamı yaratarak daha kapsamlı ve hızlı bir şekilde sınanmalarını sağlamaktır. Bu çalışmada, Türkçe metinlerin bilgisayar yardımı ile incelenerek elde edilen verilerin bir veritabanında tutulması ve bu veriler yardımı ile yazarı bilinmeyen metinlerin yazar tahminin yapılabilmesi amaçlanmıştır. Verilerin belirlenebilmesi için çalışmada biçimbirimsel çözümleyici ve söz dizimi çözümleyicisi geliştirilmiştir.

Biçimbirimsel çözümleyicinin geliştirilmesi için sonlu durum iletim sistemleri kullanılmıştır. Söz dizimi analizi için, genişletilmiş geçiş ağları kullanılmıştır. Yazarın bulunması için, istatistiksel yaklaşım yöntemlerinden olan Bulanık Mantık teorisi kullanılmıştır.

Geliştirilen program sonucunda, kullanılan analiz yöntemlerinin olumlu sonuç verdiği ve metne ait verilerin büyük bir doğruluk payı ile bulunduğu görülmüştür. Yazar veritabanında yazarlara ait veriler oluştuğunda metne ait yazar tahmini doğru olarak yapılabilmektedir.



## **ABSTRACT**

### **A NATURAL LANGUAGE PROCESSING TOOL FOR THE ANALYSIS OF TURKISH TEXTS and AUTHOR RECOGNITION**

In this thesis, a system was developed for the analysis of Turkish texts and author recognition. Features extracted at the end of analysis is kept in the database and is used for author recognition if author of the text is unknown. For text analysis, a morphological analyzer and a syntactic analyzer were developed.

Morphological analyzer developed for the thesis, uses finite state transducers to encode variation of words. Syntactic Analyzer uses augmented transition networks which are known as most powerful analysis tools for syntax. For author recognition, Fuzzy Logic theory which is one of the statistical reasoning models is used.

As the number of texts increase, system becomes more powerful in author recognition.



## 1. GİRİŞ

Doğal Dil İşleme (Natural Language Processing) önümüzdeki yıllarda insanların bilgisayarlar ile etkileşimlerinde temel bir takım değişiklikler getirmeye aday teknolojilerden biridir. Bilgisayarlar ile Doğal Dil İşleme çok değişik alanlarda uygulaması olabilecek bir konudur. Örneğin, çoğumuzun kullandığı sözcük işlemci (word processor) gibi programlarda bulunan hatalı yazılmış sözcüklerin bulunması ve düzeltilmesi işlevi bu tip uygulamaların en basitlerindedir. Burada, bilgisayar çeşitli nedenlerle(Hızlı yazma sırasında hata, doğru yazımı bilmeme, vb. ) oluşan yazım hatalarını bulmakta ve eğer istenirse kullanıcıya düzeltmede kullanılmak üzere doğru sözcükler önermektedir. Daha karmaşık bir uygulama olarak bir veritabanına SQL ile değil de, örneğin Türkçe ile soru yöneltmeyi ve sistemin bunu çözümleyerek bir SQL sorgusuna dönüştürüp işledikten sonra sonuçları kullanıcıya vermesi gösterilebilir. Bilgisayar yardımı ile dilden dile (yarı-) otomatik bir şekilde metin çevirisi yapmak, bilgisayar yardımı ile dil öğretmek, bilgisayar yardımı ile tek veya çok dilli sözcüklere erişmek, doğal dilde cümle ve metin üretmek gibi uygulamaları doğal dil işlemenin en önemli örnekleri arasında görülebilir. Çok daha geniş bir bakış açısı ile de konuşma tanıma ve konuşma üretmeyi de - kullandıkları temel teknolojiler oldukça farklı olsa da- bu alan içinde görmek mümkündür. Örneğin, teknolojinin günümüzde geldiği noktada, A.B.D, Almanya ve Japonyada' ki araştırmacılar, telefonla konuşan iki kişinin konuşmalarını anında tanıyıp karşısındaki kişinin diline çeviren, onun anlayabileceği konuşmayı üreten sistemlerin prototiplerini gösterebilmişlerdir. Ancak bu gibi sistemlerin günlük hayatta etkin olarak kullanılmaları için daha uzun bir zamana gereksinim vardır. Doğal dil işlemenin bir diğer önemli yönü de dil bilim kuramlarına deney ortamı yaratarak daha kapsamlı ve hızlı bir şekilde sınanmalarını sağlamaktır. Bu açıdan, doğal dil işleme teknolojisi dilbilimcileri ile bilgisayar bilimcileri ortak çalışmaya yönlendirmektedir. Bu noktada, yapılabilecek en yararlı çalışmalardan birisi de metnin analiz edilmesi sonucunda elde edilecek dilbilimsel bulguların kullanılarak yazarı bilinmeyen metinlerin yazar tahminin yapılması veya yazarın tarzının ortaya konmasıdır.

## 2. DOĞAL DİL İŞLEME

Doğal Dil İşleme doğal bir dili (konuşma dili), çözümleyen, anlayan yorumlayan ve üreten bilgisayar sistemlerinin tasarımını ve gerçekleşmesini konu alan temel bir bilim ve mühendislik alanıdır. Doğal Dil İşleme yapay zeka ( Örneğin bilgi gösterimi, akıl yürütme, planlama vb.) , biçimsel diller teorisi ( dil çözümleme), kuramsal dilbilim, bilgisayar destekli dilbilim ve bilişsel psikoloji gibi birbirinden son derece farklı alanlarda geliştirilmiş kuram, yöntem, ve teknolojileri bir araya getirir. 1950 ve 1960' larda yapay zekanın bir alt alanı olarak görülen Doğal Dil İşleme çalışmaları bilimsel ve ticari içeriğinden dolayı yapay zekadan bağımsız olarak bilgisayar bilimlerinin temel bir disiplini olarak kabul edilmektedir. Doğal Dil İşleme alanındaki araştırmalarda temel amaçlar genellikle şunlar olmuştur :

- doğal dillerin işlev ve yapısını daha iyi anlamak
- bilgisayarlar ve insanlar arasındaki arabirim olarak doğal dil kullanmak ve bu şekilde bilgisayar ile insanlar arasında iletişimi kolaylaştırmak
- bilgisayar ile dil çevirisi yapmak

Japonya, ABD, İngiltere, Almanya, Hollanda, Fransa gibi ülkelerde bu teknolojiyi kullanan çeşitli yazılımlar ve bilgisayar sistemleri kullanıcıların hizmetlerine sunulmuştur. Bilim ve iş alanında geçerliliği en çok olan dil olması nedeniyle İngilizce bu gibi ürünlerin en çok uygulandığı dil olmuştur. Ancak bu teknolojileri Türkçe' ye uygulamak ve Türkçe' de bir araştırma altyapısı oluşturmak için daha çok çalışma yapılması gerekmektedir.

Doğal Dil İşleme ve yakın alanlarda yapılan çalışmalar, bir yanda işlenen dilin yapısından bağımsız olma iddiasında kuramlar geliştirirken, bir yandan da bunların geniş kapsamlı olarak uygulanabilmesi için işlenecek dillere özel kaynaklar üzerinde yoğunlaşmaktadır. Şu ana kadar geliştirilen kuramların çoğu genelde İngilizce ve benzeri dilleri temel uygulama alanı olarak aldığı için, çeşitli özellikleri ile bu dillerden farklı dillere uygulanmalarında sorunlar çıkabilmektedir.

Doğal dili anlamak için kullanılacak dört çeşit ana bilgi vardır:

- Morfolojik bilgi (Morphological knowledge)
- Sözdizimi (Sentaks) bilgisi(Syntactic knowledge)

- Anlambilim bilgisi (Semantic knowledge)
- Pragmatik bilgi (Pragmatical knowledge)

Morfoloji, sözcüklerin biçimleri ile ilgilidir. Sentaks, dilin yapısı olarak kabul edilebilir; cümleler kurabilmek için sözcükleri sıralamanın yollarını açıklar. Anlambilim, sözcükleri kavramlara ilişkilendirme yollarını açıklar. Pragmatik bilgi de bizim dünyayı nasıl gördüğümüzle ilgilidir. Anlambilim ve pragmatik bilgi bu çalışmanın dışında kaldığından detaylı bilgi için Doğal Dil İşleme kitaplarından yararlanılabilir.

Doğal Dil İşleme' nin kısa bir tanımından sonra Doğal Dil İşleme uygulamalarından kısaca bahsedilecek ve çalışmanın ana konularından biri olan doğal dillerin sözdizimi çözümlemesi detaylandırılacaktır.

## 2.1 Doğal Dil İşleme Uygulamaları

Bu bölümde, Doğal Dil İşleme uygulamalarının önemli alanları anlatılacaktır.

- **Makine Çevirisi:** Bir metni, bir dilden başka bir dile çevirmek için bilgisayar kullanımını amaçlayan uygulama alanıdır. 1950' li yıllardan beri çalışmalar süregelmektedir.
- **Dokümanların anlaşılması ve oluşturulması:** Bir bilgisayarın bir dokümanı okuyarak anlamasını ve dokümanının özetinin çıkarılmasını sağlayan araştırma alanıdır. Bilgisayar daha sonra, belirli sorulara yanıt verebilmelidir. Doküman oluşturma, doküman anlamaya bağlıdır. Bilgisayarın hafızasında saklanan bilgiyi biçimsel bir dilden doğal bir dile çevirir.
- **Veritabanları için doğal dil arabirimi :** Daha önce girilmiş bilgilerden sorulan sorunun cevabını bulan bir soru-cevap sistemidir. Bilgi çok büyük bir veritabanında saklanabilir ve kullanıcının ihtiyaç duyduğu bilgiye erişebilmesi için sistemi tanınması gerekir. Bu tipte bir uygulama, kullanıcıyı böyle bir yükten kurtarır. Kullanıcı SQL komutlarını yazmak yerine , konuşma dilinde bir cümle yazarak ihtiyaç duyduğu bilgiye ulaşabilir.
- **Bilgisayar Tabanlı Eğitim (BTE):** Bilgisayarlar, eğitim amacı ile, çoğu zaman öğrenci ile bilgisayar arasında soru cevap etkileşimi gerektiren çok farklı yollarla kullanılmışlardır. Belirli konuları öğretmek için tasarlanan bilgisayar programlarına doğal dil işlemenin entegre edilmesi bir BTE sistemini daha güçlü kılar. Yabancılar Türkçe' yi öğretmek için tasarlanan bilgisayar tabanlı eğitim araçları Doğal Dil İşleme için önemli bir uygulama alanıdır.

### 3. TÜRKÇE

Türkçe Ural-Altay dil grubuna giren bir dildir. Yapısal olarak Türkçe'nin Doğal Dil İşlemede sorunlar yaratan yapısal özellikleri vardır. Öncelikle, sözcük yapısı ve üretimi açısından bitişken (agglutinative) bir dildir. Bu açıdan Türkçe, Fince ve Macarca'ya benzemektedir. Bu tip dillerde sözcükler bir kök sözcüğe ardı ardına eklenen (eklenirken de ünlü uyumu, ünsüz değişmesi, ünlü ve ünsüz düşmesi gibi nedenlerle değişikliğe uğrayan) biçimbirimlerden (morpheme) oluşur. Bu biçimbirimler eklendikleri kök veya gövdenin anlamını, sözcük türünü, veya sözdizimsel işlevini değiştirirler. Bu nedenle Türkçe'de bu şekilde kurulan bir sözcük başka bir dilde bir cümle ile ifade ediliyor olabilir. Buna örnek olarak

#### **geçştirilemeyeceklerindendir**

sözcüğü gösterilebilir. Kökü geç olan bu sözcük şu şekilde biçimbirimlerine ayrılabilir:

#### **geç+iş+tir+il+e+meye+cek+ler+in+den+dir**

Türkçe'nin hemen her dilbilimi kitabında incelenmesinin nedeni, dilin birtakım dilbilgisel olgularda tipik bir örnek oluşturmasıdır .örneğin, ses uyumu, bitişken sözcük yapısı, sözdizimindeki serbestlik, ve öbek yapısında (phrase structure) tamlayanların her zaman tamlananlardan önce gelmesi gibi. Doğal Dil İşlemede karşılaşılan sorunlara ışık tutabilmesi açısından Türkçe'nin temel özellikleri şöyle sıralanabilir:

#### 3.1 Biçimbilimsel Özellikleri

Türkçe'nin sözcük yapısı köklere yapım (derivation) ve çekim (inflection) eklerinin sonek (suffix) olarak eklenmesine dayanır. Çekim soneklerinin çokluğu, yapım soneklerinin çok üreken olması ve soneklerin sözdizime (syntax) olan doğrudan etkisi, bilgisayarlı biçimbilim çözümlemesinde ilginç sorunlar ortaya çıkarmaktadır. Örneğin :

(1) kalemleri

<u>cözümleme</u>	<u>örnek kullanım</u>
a. kök+çoğul+3. tekil - iyelik	Ayşe'nin kalemlerini iade etmelisin.
b. kök+3. Çoğul - iyelik	Bunlar kardeşlerimin kalemleri.
c. kök+çoğul+durum	Şeniz kalemleri Deniz'e verdi.

Çözümlemedeki bu tip belirsizlikler bazen öbek ve cümle düzeyinde inceleme ile giderilebilir. Örneğin , bunlar kardeşlerimin kalemleri cümlesindeki kalemleri sözcüğü, öznedeki iyelik sonekinden dolayı kök+3.çoğul-iyelik olarak çözümlenebilir. Sözdizim düzeyinde inceleme sorunu her zaman gideremez. örneğin kalemleri kaybolmuş cümlesi 'Onun kalemleri kaybolmuş' anlamını taşıyabileceği gibi 'Onların kalemleri kaybolmuş' anlamına da gelebilir. Biçimbilim çözümlemesindeki belirsizlikler, sözcük türlerini bulmakta da sorun çıkarabilir: Örneğin,

(2) gelir

**çözümleme**

**örnek kullanım**

- a. ad Bu gelir düzeyindeki bir insan için çok lüks bir alışkanlık.  
b. eylem+zaman Her sabah saat sekizde gelir.

Çekim son eklerinin varlığı durumunda bu belirsizliklerin çözümü için daha çok bilgi vardır; örneğin gelirlerimiz sözcüğünde kök adıl olmak durumundadır çünkü ad çekimi almıştır.

Biçimbilimden başka, Türkçe'de bulunan belirsizlikler ve öbek yapısı belirsizlikleri de dil işlemede çözülmesi gereken sorunlardır. Örneğin, burada yiyebileceğim birşeyler bulabileceğimi düşünmekle hata etmişim cümlesinde burada belirteci yemek , bulmak ve hata etmek eylemlerini belirleyebilir ama en kuvvetli ihtimal bulmak eylemini belirlemesidir. Sokakta serbest bıraktığım köpeğimle karşılaştım cümlesinde sokakta ilgeç öbeği olarak serbest bırakmak veya karşılaşmak eylemlerini belirleyebilir.

Benzer belirsizlikler bileşik ad öbeklerinde de görülebilir .örneğin, güzel sınıf arkadaşımın annesi gibi. Bu tip sorunlar bilgisayar ile çözümlemede bir yandan hızlı çözümleme, diğer yandan sözcük dilbilgisinin zenginliği açısından çözüm bekleyen araştırma konularıdır.

	ÇOCUK	YETİŞKİN
ÖNY	%46	%48
ÖYN	%17	%25
NYÖ	%20	%13
YÖN	%10	%6
NÖY	%7	%8
YNÖ	%0	%0

Şekil 3.1 Türkçe'de sözcüklerin cümledeki diziliş sırası

### 3.2 Sözcük Dizimi

Türkçe' nin diğer bir özelliği de cümlelerde sözcük sırasının oldukça esnek olmasıdır. Şekil 3.1' de çeşitli sözcük dizimlerinin kullanım yüzdeleri verilmiştir. Cümle öğelerinin dizilişi açısından Türkçe bir **Özne-Nesne-Yüklem (ÖNY)** sıralı bir dil olarak kabul edilebilir. Adlar durum soneki almadığında ÖNY dizimi sayılır. Ancak Türkçe sözcük sıralarının değişmesine olanak tanır. Bu açıdan Türkçe Fince ve Japonca'ya benzer. Bu esnekliğin en önemli nedenlerinden biri Türkçe'nin özne, nesne diye adlandırdığımız sözdizimsel birimleri sözcüklere doğrudan durum ekleri getirerek göstermesidir. Bu işlevler Fince'de Türkçe' deki gibi eklerle, Japonca' da ise sözcüleri takip eden ilgeçlerle belirtilmektedir. Buna karşılık bu işlevi yüklenen sözcükler, örneğin İngilizce' de sözcüğün yeri ile belirlenmekte, ve dolayısı ile İngilizce'de sözcük sırası son derece katı olmaktadır. Bu nedenle İngilizce için geliştirilen sözdizim çözümlene yöntemleri Türkçe gibi sözcük sırası son derece esnek diller için kullanılamamaktadır.

Türkçe' de sözcük dizimindeki çeşitlenmeler vurgulama işlevini görür. Genellikle vurgulanmak istenen öbek yüklemden önce gelir.

- (3) **Ben bir arkadaşımı okul yolunda gördüm.**
- (4) **Bir arkadaşımı okul yolunda ben gördüm.**
- (5) **Ben okul yolunda bir arkadaşımı gördüm.**

Dizimdeki serbestlik durum sonekleri kullanılmadığında azalmaktadır.

Örneğin :

**Ben okul yolunda bir arkadaş gördüm**

cümlesinin öbekleri daha değişik bir şekilde sıralanamaz. Serbestliğin kısıtlı olduğu diğer bir nokta da belirteçlerdir. Sıra değişimi anlamı değiştirebilir. Örneğin,

- (6) **Sınavı zor yetiştirdim.**
- (7) **Zor sınavı yetiştirdim.**
- (8) **Zor sınav yetiştirdim.**

Ayrıca (7) de zor sıfat özelliği taşımaktadır.

Türkçe'nin sözdizimsel yapısal çözümlenmesinde sorun yaratan diğer bir olgu da, öbek öğelerinin ardışık olmama olasılığıdır. Örneğin, sahafılığın eskiden çok renkli yanları vardı



cümlesinde eskiden belirteci, tamladığı eylemin yanında değil, ad öbeğinin içinde yer almaktadır. Öbek yapısı gramerleri (Phrase structure grammars) bu çeşit örneklerde sorunlarla karşılaşmaktadır.

Biçimbilimsel ve sözdizimsel çözümlenmeleri gerçekleştiren ayrıştırıcılarda (parser) aranması gereken diğer bir özellik de, yanlış veri girildiğinde tanı yapabilmeleri ve mümkünse öneride bulunabilmeleridir. Örneğin,

veri	sorun
(9)* kötümüŖleŖti	sıra ve ses uyumu
(10)* kötümüŖleŖti	sıra
(11)* çiçeği	ünlü-ünsüz deęiŖimi

Benzer Ŗekilde, bir sözdizimsel ayrıştırıcı aŖaęıdaki hataları da bulmak zorundadır:

- (12)\* ArkadaŖımın annesine gördüm.  
 (13)\* Eski bir arkadaŖıma karŖılaŖtım.

### 3.2.1 Cümlelerin öęeleri

Cümlelerin temel öęeleri özne ve yüklemdir. Cümle yüklem üzerine kurulur. Bir söz öbeğinin cümle deęeri taşıması, yüklemine baęlıdır. Konuşmalı yazılardaki kimi özel durumları saymazsak, yalnız yüklemden oluşan cümleler kurabilmemize karŖın diğer öęelerle cümle oluşturamayız. Türkçe cümlelerin yardımcı öęeleri nesne (düz tümleç) ve dolaylı tümleçtir. Yardımcı öęelerin kullanımı seęimlik olup eylemin özelliklerine baęlıdır.

#### 3.2.1.1 Yüklem

Türkçe'de, tüm çekimli eylemler dışında, eylem almıŖ ad soylu tüm sözcükler, sözcük öbekleri yüklem görevinde kullanılabilir:

- Geceleri evime ayıŖıęında dönerim. (eylem)
- O gün olay yerindeydik. (ad tamlaması)
- Babası mühendistir. (ad)
- BaŖarının sırrı, çalışmaktır.(eylemsi)
- Daha deniz görmemiŖ bir çoban çocuęuyum. (iç içe girmiŖ tamlama)
- Çocuk yaramaz mı yaramaz. (pekiŖtirme öbeęi)



- **Siz kimsiniz? (zamir)**
- **Geceleyin bir ses böler uykumu : Nerdesin?(eylem ve zamir)**
- **Her şey tasarladığım gibiydi. (edat öbeği)**
- **O çocukları için hem bir ana hem bir babaydı.(bağlaç öbeği)**
- **Arkadaşına yardımcı oldu. (Bileşik eylem)**

Koşullu eylemler yalnız yan cümlecik oluşturur.

- **Hava güzel olursa, gezmeye gidilir.**

Cümlelerin yüklemi düşebilir.

- - **Nerden geliyorsun?**
- - **Evden (geliyorum)...**

Deyimler tek kavramı karşılayan söz öbeği olduğundan bölünmeksizin öge olur.

- **O karda yürür, izini belli etmez.**

Türkçede yüklem genelde sonda bulunur, ama devrik cümlelerde yüklem sabit bir yeri yoktur.

### 3.2.1.2 Özne

Özne, cümlelerin ikinci temel ögesidir. Eylemden sonra gelmesinin nedeni, gizli bir özne ile cümle kurulabilir, ama yüklemsiz bir cümle olmaz. Cümlede özneyi bulmak için, özne bir eylem yapıyorsa, eyleme "işini yapan kim, ne?" sorusu sorulur. Özne bir şey olursa, eyleme "olan kimdir, nedir?" diye sorulur.

Yüklemi etken çatılı eylemlerde, ya da eylem almış ad soylu sözcüklerden oluşmuş cümlelerin öznesi gerçek özne'dir. Yüklemi edilgen çatılı eylem olan cümleler, gerçek özne almaz. Bu cümlelerde özne, yapılan işin etkisine uğrayan ögedir. Yüklemi edilgen çatılı olan cümlelerde, eylemin kim tarafından yapıldığı da vurgulanabilir. Öge görevi, belirteç tümleci olmakla birlikte, gerçek özneyi belirten sözlere gizli özne denir. Edilgen geçişsiz çatılı fiillerin yüklem olduğu cümlelerde özne belirsizleşir. Özne, her zaman yalın halde bulunur.

- **Özne ile Yüklem Uyumluğu**

**a. Tekillik çoğulluk bakımından uyumluluk**

Türkçede, genel olarak özne tekil ise yüklem de tekil; özne çoğul ise yüklem de çoğul olur. Ancak, bu kurallara uymayan pek çok durum vardır. Cansız varlıklar, bitki ve hayvan adları çoğul özne görevindeyse yüklem tekildir. Organ adları çoğul özne görevindeyse yüklem tekildir. Özneler sayılarla belirtilirse yüklem tekildir. Örneğin , Kuşlar ötüşür, gözlerim kamaştı, Alanda sekiz kişi toplandı gibi.... Bu koşullar artırılabilir.

### **b.Özne ile yüklemün kişice uygunluğu**

1.Öznesi birinci, ikinci, üçüncü tekil ya da çoğul kişiler olan cümlelerde kişi bakımından özne yüklem uygunluğu kesindir. Ben bilirim, sen bilirsin, o bilir, biz biliriz, siz bilirsiniz, onlar bilirler gibi...

2.Öznesi ayrı ayrı kişilerden oluşan çok özneli cümlelerde değişik durumlar görülür.

- Özneyi oluşturan kümede, birinci tekil ya da çoğul kişi de varsa, yüklem birinci çoğul olur.
- Özneyi oluşturanlar yalnız ikinci ve üçüncü kişilerse yüklem çoğul olur.
- “Kimi, bazı “ belgisiz zamirleri birinci çoğul iyelik eki alarak özne olunca, yüklem üçüncü tekil kişi olabilir.
- Cümlede söz söyleyen kişi bir topluluk adına konuşuyorsa özne de yüklem de birinci çoğul olabilir.
- Kimi kez övünme, böbürlenme amacı ile birinci tekil kişi yerine birinci çoğul kişinin kullanıldığı görülür. yüklem de ona göre çoğullanır.

### **3.2.1.3 Nesne**

Cümlelerin üçüncü önemli ögesidir. Nesnelere, belirtili ve belirtisiz nesne diye ikiye ayrılır :

#### **1. Belirtili nesne**

Belirtme durum eki “-i” almış nesnelere, belirtili nesne, söz söyleyen ve dinleyence bilinen bir varlığı anlatır:

- Tatlđ dil yılanı deliğinden çıkarır. (ad)
- Biz biliriz bizim işlerimizi. (ad tamlaması)
- Deniz müzesini gezdim. (ad tamlaması)
- Gelen gideni aratır. (adlaşmış zarf)

#### **2.Belirtisiz nesne**

“-i” durum eki almamış olan ve “ne, neler?” sorularına yanıt veren nesnelere. Belirtisiz nesne öznenin yaptığı işten etkileneni açıkça anlatmaz. “Yatağına uzanmış, kitap okuyor.” cümlesinde, “kitap” belirtisiz nesnedir; dolayısıyla, bu kitabın hangi kitap olduğu tam olarak bilinmemektedir. Diğer örnekler:

- **İçsem su, susuzların elinden. (ad)**
- **Çocuğuna okul çantası aldı. (ad tamlaması)**
- **Durmadan konuşmak istiyordu.(ad eylem)**

Nesneler, geçişli etken eylemlerin yüklem olduğu cümlelerde bulunur. Geçişli edilgen eylemlerin yüklem olduğu cümleler nesnesiz sayılır; çünkü bu cümlelerde nesne özne ( sözde özne) görevinde bulunur. Dönüşlü eylemlerin kimileri nesne alabilir:

- **Adam suçu üstlenmiş.**

Ad cümlelerinde nesne yer almaz. Ancak yüklemi eylemsi olan ad cümleleri nesne alabilir:

- **O sırada ben gazete okumaktaydım.**
- **Sana düşen iş, sınavı kazanmaktır.**

Belirtisiz nesnelere her zaman yüklem yanında yer alır. Ancak “de” ile bağlanmış kimi söz öbeklerinde, belirtisiz nesne ile yüklem arasına başka sözcük girebilir:

- **Çarşıdan üç ekmek aldım.**
- **Bir kalem sana, bir kalem de arkadaşına aldım.**

#### 3.2.1.4 Tümleç

Tümleçler, dolaylı tümleç ve zarf tümleci olmak üzere ikiye ayrılır.

#### DOLAYLI TÜMLEÇ

“-e, -de, -den” durum ekleriyle çekimlenen ve yükleme sorulan “neye, nede, neden, nereye, nerede, nereden, kime, kimde, kimden?” sorularına yanıt vererek yüklemi yönelme, kalma(bulunma, çıkma, ayrılma) ilgisiyle tümleyen öğelerdir.“-e, -de, -den” eki alna sözcükler “zaman, neden, amaç” bildirirse dolaylı tümleç değil, zarf tümleci olur.

### Aldığı ekler bakımından dolaylı tümleçler :

a. “-e” durum ekiyle çekimlenen dolaylı tümleçler

- **Arkadaş, yurduma alçakları uğratma sakın !**
- **Soranlara selam olsun.**

b.“-de” ekiyle çekimlenen dolaylı tümleçler

Bu tümleçlerin temel işlevi, yüklem bildirdiği eylemi(bulunduğu, kaldığı, olduğu) yer bakımından tamamlamaktır.

“-de” li tümleçler yalnız eylem cümlelerinde bulunur. Ancak; “var, yok, gerek” gibi sözcüklerin yüklem olduğu ad cümlelerinde de “de”li tümleç bulunabilir.

- **Yolcu yolunda gerek**
- **Kurtuluşu kaçmakta buldu. (adeylem)**

c.”-den” eki ile çekimlenen dolaylı tümleçler

Bu tümleçlerin asıl işlevi, yüklem belirttiği iş, oluş, hareket ya da durumun nereden başladığını, nereden çıkıp uzaklaştığını bildirmektir.

“den” li tümleçler eylem cümlelerinde, “var, yok” gibi eylem almış, eylemsi ile kurulan cümlelerde bulunur.

- **Nice taşlı yollardan geçtik.**
- **Evden geliyor.**

Kısaca, -e, -de, -den durum ekleriyle çekimlenen dolaylı tümleçler, eylemin

- **Nereye yöneldiğini ya da yaklaştığını**
- **Nerede bulunduğunu, olduğunu**
- **Nereden başladığını, çıktığını, uzaklaştığını**

anlatır.

Bir cümlede, birden çok ekle çekimlenen birden fazla dolaylı tümleç bulunabilir.

## **BELİRTEÇ (ZARF) TÛMLECİ**

Yüklemi zaman, durum, yer, yön, nicelik, soru ilgileriyle tamamlayan öğelerdir. Yükleme “ne zaman, nereye, ne kadar?” soruları yöneltilerek bulunur.

Belirteç tümleçleri yüklemi şu yönlerden tamamlar :

### a.Zaman ilgisiyle

- **Dün bir mektup aldım. (Ne zaman?)**
- **Benim doğduğum köyleri / geceleri eşkiyalar basardı.**

### b.Yer-yön ilgisiyle

- **Güneşe karşı oturmuş şarkı söylüyordu.**
- **Yukarı çıkınız.(Nereye?)**

İçeri, dışarı belirteçleri “-e, -de, -den” durum eklerini alırlarsa dolaylı tümleç olurlar.

### c.Durum ilgisiyle

- **Yağmur durmadan yağıyordu.**
- **Sebzeleri iyice yıkadı.**

### d.Neden ilgisiyle

- **Senin uğruna nice acılara katlandım. (Niçin?)**
- **Aşırı hız yüzünden yollarda birçok kaza oluyor.**

### e.Dilek ya da koşul ilgisiyle

- **Bari bu kez kazansam da emeklerim boşa gitmese.**
- **Şayet bulamazsan ben sana yardım ederim.**

### f.Yineleme ilgisiyle

- **Yine hazan mevsimi geldi.**
- **Tekrar görüşelim.**

### g.Nicelik ilgisiyle

- **Çok çalıştım.**
- **En çok annesini seviyormuş.**

h.Soru ilgisiyle

- **Ne bakıyorsun?**
- **Ah, neye doğdun sarı yıldız?**

Yüklemi “ile, için, yalnız, belki...” edatlarıyla tamamlanan öğelere ilgeç (edat) tümleci adı da verilmektedir. Ne var ki, bu gibi sözcüklerle oluşturulan öğelerin, cümlede belirteç görevinde olduğu açıktır. Edat tümleci biçimsel bir adlandırmadır.

### 3.2.1.5 Sorulan bir sorunun yanıtı olan öğeler

- Bir soru hangi öğeyi buldurmaya yönelikse o öğenin görevini üstlenir.
- “Nereye?” sorusunun yanıtı, dolaylı tümleç de belirteç tümleci de olabilir. Bu, soruyu yanıtlayan sözcüğün “-e” eki alıp almamasına bağlıdır.
- “Nereye?” sorusunun kendisi her zaman dolaylı tümleç görevindedir.
- “Ne?” sorusu özneyi bulmak için kullanılmışsa “özne”, nesneyi bulmak için kullanılmışsa “nesne” görevindedir.
- “Ne?” sorusu “niçin?” anlamı taşırsa belirteç tümleci görevini üstlenir.
- Ekeylem alan soru sözcüklerinin tümü yüklemi buldurmaya yöneliktir ve cümlede yüklem görevini üstlenir.
- Bir cümlede “mi?” sorusu hangi öğeden sonra kullanılmışsa, soru o cümleyi buldurmaya yöneliktir.

## 3.3 Cümle

Cümleler yüklemelerine, öğelerinin dizilişine, anlamlarına ve yapılarına göre olmak üzere dört ana grupta ele alınabilir.

### 3.3.1 Anlamlarına göre cümleler

#### 1. Olumlu Cümle

Yüklemin iletmediği yargının gerçekleştiğini, gerçekleşmekte olduğunu ya da gerçekleşeceğini bildiren cümlelerdir.

##### a.Anlamca ve biçimce olumlu cümleler

Yüklemleri olumluluk bildiren cümlelerdir.

### **b.Biçimce olumsuz, anlamca olumlu cümleler**

Yüklemleri “-me” ya da deęille olumsuz yapılan, ancak anlamca olumlu olan cümlelerdir.

## **2. Olumsuz Cümle**

Yüklemin bildirdięi yargının gerçekleşmedięini ya da gerçekleşmeyeceęini bildiren cümledir.

Türkçe’de, “-me” (Geniş zamanlı eylemlerde -mez) “-siz” ekleri ile “yok”, “deęil”, “ne...ne” sözcükleri cümleyi olumsuz yapmakta kullanılır.

Olumsuz cümleler ikiye ayrılır :

### **a.Anlamca ve biçimce olumsuz cümle**

### **b.Biçimce olumsuz, anlamca olumsuz cümleler**

## **3. Soru Cümlesi**

“mi” ya da soru sözcüklerinden biriyle kurulmuş, soru anlamı veren cümledir.

## **4. Ünlem Cümlesi**

Korkma, şaşma, acıma, kızma, sevinme, özlem, beęenme, övme gibi duyguları anlatan cümlelerdir Ünlem cümleleri, ünlemler, yansımalar, seslenme sözleri ve ünlem deęerindeki sözlerle kurulur.

### **3.3.2 Yüklem türüne göre cümleler**

#### **1. Eylem (Fiil) Cümlesi**

Yüklemi çekimli eylem olan cümlelerdir.

#### **2.Ad (İsim) Cümlesi**

Yüklemi, ad, ad soylu bir sözcük ya da tamlamadır. Bunlar eylem olarak yüklem görevi üstlenirler.

### 3.3.3 Yüklemin yerine göre cümleler

Yüklemin cümledeki yerine göre cümleler ikiye ayrılır:

#### 1.Kuralı (Düz) cümle

Yüklemi sonda bulunan cümlelerdir. Başka bir deyişle, yardımcı öğeleri yüklemden ya da yüklem öbeğinden önce gelen cümlelerdir. Bu tür cümlelerde genellikle Türkçe'nin sözdizimi olan Özne + Nesne (Tümleç)+ Yüklem (ÖNY) sıralaması korunur.

#### 2.Devrik Cümle

Yüklemin başta ya da ortada bulunduğu cümlelerdir. Söyleyişe belirli bir özellik katmak, söyleyişi konuşma diline yaklaştırmak için kullanılır. Şiir dilinde devrik cümlelere daha çok rastlanır.

Devrik cümlede yüklem dışındaki öğelerin dizilişi değişik özellikler gösterir. Bu dizilişler, anlam farklılıkları yaratır.

### 3.3.4 Yapılarına göre cümleler

#### 1.Basit (Yalın) Cümle

Tek bir yüklemle kurulan, bağımsız tek bir yargı (düşünce, duygu, istek...) bildiren cümlelerdir. Böyle cümlelerde yan cümlecik, başka bir cümleyi tamamlamaz ya da başka bir cümleyle tamamlanamaz.

Basit (yalın) cümledeki tek yüklem; çekimli bir eylem, ekeyleme çekimlenmiş ad soylu bir sözcük ya da ad tamlaması olabilir.

#### 2. Bileşik Cümle

İçinde birden çok yargı (duygu, düşünce, istek...) içerdikleri için yalın cümleye göre daha zengindir.

Bu cümleler bir temel cümleyle, yeteri kadar yan cümlecikten oluşurlar.

**Temel cümle** : Çekimli bir eylem ya da ekeyleme almış ad soylu bir sözcükle kurulan; tek başına yargı bildiren ve yan yargılarla ulaşılan temel yargıdır.



**Yancümlecik** : Eylemsilerle ve dilek-şart kipinde çekimlenmiş eylemlerle oluşturulur. Bu cümlede eylemsilerin ve dilek-şart kipinde çekimlenmiş eylemlerin sayısı kadar yancümlecik vardır.

Bileşik cümlelerde, temel cümle her zaman tektir. Yancümlecik(ler) ise birden çok olabilir.

- **Bileşik Cümlelerin Oluşum Biçimleri**

a.Eylemsilerle ve çekimlenmiş eylemlerle oluşanlar :

- **Yüzünün güldüğünü / görmedim.**
- **Arayan / bulur.**

b.Eylemsiler ve ekeylemlerle çekimlenmiş bir sözden oluşanlar :

- **Kapımızı çalan / sadece oydu.**
- **Konuması / bile bir başkaydı.**
- **Hepimizin inandığı / değerler var.**

Bileşik cümlelerde, yancümlecikler temel cümleciğin öznesi, tümleci, nesnesi olabilir.

### 3.Sıralı ve Bağlı Cümle

- **Sıralı Cümle**

Tek başlarına yargı değeri taşıyan birden çok cümlelerin, aralarına nokta konmayıp virgülle veya noktalı virgülle art arda sıralanmasıyla oluşan cümledir. Bu cümlelerin arasına nokta konursa bağımsız ayrı ayrı cümleler elde edilir.

- **Birden başım döndü. Gözüm karardı. (Bağımsız cümle)**
- **Birden başım döndü, gözüm karardı.**  
(sıralı cümle, ortak bir öğeleri yok, bağımsız)

- **Bağımlı Sıralı Cümle**

Sıralı cümlelerin, özneleri, yüklemeleri ya da diğer öğeleri ortak olabilir. Bu durumdaki cümleler “bağımlı sıralı cümle” olarak adlandırılır.

- **Ata dost gibi bakmalı, düşman gibi binmeli. (dolaylı tümleç ortak)**
- **Yunuslar; koya girdiler, atlaya zıplaya yüzdüler,**

### engine doğru gittiler. (özne ortak)

- **Anlamca Bağlı Cümle**

Böyle cümleler arasında, “bu yüzden”, “bunun için”, “dolayısıyla” gibi bağlayıcı sözlerle anlatılabilecek anlam ilişkileri vardır; ancak, bağlayıcı sözcükler kullanmadan, onların yerine noktalı virgül konarak da cümle bağlanabilir.

- **Aheste çek kürekleri; mehtap uyanmasın.**

- **“ki” ile bağlı cümle**

‘Ki’ bağlacı Türkçe değildir. Bu bağlaçla birbirine bağlanan cümleler “ki” kullanılmadan kurulacak olursa , ki’ den önceki çekimli eylem cümlelerin yüklemi olur.

- **Anladım ki sen bu işi başaramayacaksın.**
- **Bu işi başaramayacağını anladım.**

- **Eksilteli Cümle**

Kimi öğeleri, özellikle yüklemi , kendinden önceki ya da sonraki cümlede yer alan , yargısı tamamlanmadığı halde tam bir cümle anlamı verebilen sözdizimine “eksilteli cümle” denir.

Konuşma dilinde sıklıkla kullanılan bu tür cümleler yalnız öznel yüklem, dolaylı tümleş ya da nesneden ibaret olabilir.

- **İnsanlar söyleşe söyleşe, hayvanlar koklaşa koklaşa...(anlaşır)**
- **Nereye (gideceğiz) ?**
- **Okula (gideceğiz)**

Eksilteli cümleler yapıcı incelenirken eksik öge ya da öğeler zihinden tamamlanır.

#### 4. DOĞAL DİLLERİN İFADE EDİLME ŞEKİLLERİ

Bir dili oluşturan dilbilgisi (gramer) ve kurallardır. Diğer bir tanımla, gramer sonlu olmak zorundadır. Örneğin ,

Sadece { a,b } harflerinden oluşan bir alfabe düşünelim. Kural da cümlelerin ikinci yarısını oluşturan harfler, ilk yarısını oluşturan harflerin tersi olsun. Bu gramer ve kural doğrultusunda aşağıdakilerin hepsi bu dil için geçerli sözcükler olacaktır:

abba, bbaaabbaabb, bababbabab

Bu sıralamayı oluşturan kurallar şunlar olabilir :

$$\begin{array}{ll} S \rightarrow aa & S \rightarrow aSa \\ S \rightarrow bb & S \rightarrow bSb \end{array}$$

Bu kurallar bir cümle oluşturmak için nasıl kullanılabilir? İlk adım olarak başlangıç veya başlangıç sembolü olarak kullanılacak olan S yazılır. Daha sonra S dört seçenekten biri ile uzatılır. Yeni açıklama S içerdiği sürece, başka bir uzatma imkanı kalmayınca kadar, S yerine dört olasılıktan biri koyulur. Bu prosedür takip edilirse , bir cümle elde edebilmek için aşağıdaki gibi bir sıralama oluşturulabilir :

- (1) S
- (2) aSa
- (3) abSba
- (4) abbSbba
- (5) abbaSabba
- (6) abbabSbabba
- (7) abbabaababba

Sonraki adım olarak, gramerin ne olduğu ve bu bilginin bir dili tanımlayabilmek için nasıl kullanılabileceği şu şekilde özetlenebilir :

Tanım 1.1 Bir gramer,  $G$ , her biri sonlu olan dört elemanlı bir küme olarak tanımlanabilir.  $G = (N, \Sigma, P, S_0)$  olsun.

- (a)  $N$  terminal olmayan sembollerden oluşan bir kümedir.
- (b)  $\Sigma$  terminal sembollerden oluşan bir kümedir ve bu küme gerçek cümleleri oluşturmak için kullanılır.
- (c)  $P$  gramer kuralları kümesidir.
- (d)  $S_0$  sadece bir terminal olmayandan oluşur. Sözcüğü oluşturacak olan sıralama için başlangıç görevini görür.

Örnek 1 için,

$N = \{S\}$ ,  $\Sigma = \{a, b\}$ ,  $P = \{S \rightarrow aa, S \rightarrow bb, S \rightarrow aSa, S \rightarrow bSa\}$  ve  $S$  de başlangıç sembolüdür.

Tanımlanan dört kümeden her birinin oluşumunda takip edilmesi gereken sınırlamalar vardır.

$\Sigma$  ve  $N$ 'nin ortak elemanı olamaz. ( $\Sigma \cap N = \emptyset$ ,  $\emptyset$  boş kümeyi gösterir.) Değişik bir tanımla, sadece terminal olmayan semboller yeniden yazılabilir, böylece, terminal olmayan semboller yeniden yazma kuralının sağında da solunda da yer alabilir. Terminaller yeniden yazılamaz, ve sadece sağ tarafta yer alabilirler. Uygunluk açısından, terminal olmayan sembolün ( $X$ ) yeniden yazılması için kurallar vardır ve bu kurallar şu şekilde gösterilebilir :

$$X \rightarrow \alpha_1 | \alpha_2 | \alpha_3 | \dots | \alpha_n \quad (4.1)$$

$S \rightarrow aSa$  gibi tek bir kural şu şekilde çevrilebilir, nominal  $S$  in yerine sağ tarafta yer alan bir dizi sembol konulabilir.  $S$  için tanımlanan kurallar, yerine koyma işleminde birden çok seçeneği olduğunu gösterir, ama bu seçeneklerden sadece biri seçilebilir. Bu da, iyi oluşturulmuş bir cümle elde etmek için hangi seçeneğin seçilmesi gerektiği sorusunu ortaya çıkarır.

Bir cümle,  $\alpha$ , terminal veya terminal olmayan bir dizi sembolün arka arkaya sıralanmasıdır. Ayrıca, bir cümle,  $\alpha_{+1}$ ,  $\alpha$  den oluşturulabilir.

$\alpha \Rightarrow \lambda_1 X \lambda_2$ ,  $\alpha_{+1} \Rightarrow \lambda_1 \beta \lambda_2$ ,  $X \rightarrow \beta$   $P$  deki kurallardan biridir.

$\alpha_{+1}$ ,  $\alpha$  den tek bir ifade yerine konularak oluşturulduğu zaman  $\alpha \Rightarrow \alpha_{+1}$  şeklinde de yazılabilir. Daha genel olarak,

$$\alpha_i \Rightarrow \cdot \alpha_i \quad (4.2)$$

şeklinde yazılabilir.

$G = (N, \Sigma, P, S_0)$  olsun.  $\alpha_0$ , tanımdan dolayı, bir sıralamadaki ilk adım  $S_0$  a eşit olacağı için  $S_0$  a eşittir.  $L(G)$   $G$  ile tanımlanabilen ve oluşturulabilen sözcüklerden oluşan bir dil olsun.

$L(G) = \{w | S_0 \Rightarrow^* w \text{ ve } w \in \Sigma^*\}$  olarak gösterilebilir.

Daha genel olarak, her  $w$  aşağıdaki koşulları sağladığı sürece gramer açısından iyi bir oluşum sayılır:

(a)  $w$  yapım sıralamasının ilk formu olarak  $S_0$  dan türetilebilir.

(b)  $w$  ya boş küme olmalıdır, ya da sadece terminal sembollerden oluşmalıdır.

$S_0 \Rightarrow^* w$ ,  $w$  nin oluşumundaki terminal olmayan semboller eklendiği zaman tamamen bir yapım sıralaması olarak çevrilebilmelidir.

#### 4.1 Gramerlerin Chomsky Hiyerarşisi

Bütün gramerler  $G = (N, \Sigma, P, S_0)$  genel tanımına uyarlar. Bununla birlikte, gramerler  $P$  'deki üretim şekillerine göre de sınıflandırılabilirler.

##### Tanım 1.4

Bir gramer,  $G$ ,  $P$ 'deki her üretim  $X \rightarrow aY$  veya  $X \rightarrow a$  şeklinde ise ve  $X$  ve  $Y \in N$  kümesinde terminal olmayan,  $a$  da  $\Sigma$  kümesinde bir terminal sembolse sağ-doğrusal (right-linear) adını alır.

Daha genel olarak,  $G = (N, \Sigma, P, S_0)$  grameri

(a) Şekilden bağımsız (Context free)

$$X \rightarrow a, X \in N \text{ ve } a \in (N \cup \Sigma) \quad (4.3)$$

$a$  terminal veya terminal olmayanlar kümesinden seçilmiş elemanların istenen şekilde sıralanmasıyla oluşturulabilir.

(b) Şekile duyarlı (Context-sensitive)

$$\gamma_1 X \gamma_2 \rightarrow \gamma_1 a \lambda_2, X \in N (\gamma_1, \gamma_2, a) \in (N \cup \Sigma) \quad (4.4)$$

$X$ , sadece ve sadece  $\lambda_1$  ve  $\lambda_2$  nin context'inde  $a$  olarak yeniden yazılabilir

(c)Sınırsız (Unrestricted)

$$\alpha \rightarrow \beta, \beta \in (N \cup \Sigma) \quad \alpha \in (N \cup \Sigma) \quad (4.5)$$

Örnekler

Alfabenin iki sesli ve dört sessiz harften oluştuğunu düşünelim.

$$\Sigma = \{a, o, b, c, d, f\}$$

W başlangıç kuralı olsun ve P' de de şu kurallar bulunsun:

$$W \rightarrow aX|oX|bY|cY|dY|fY$$

$$X \rightarrow b|c|d|f|bY|cY|dY|fY$$

$$Y \rightarrow a|o|aX|oX$$

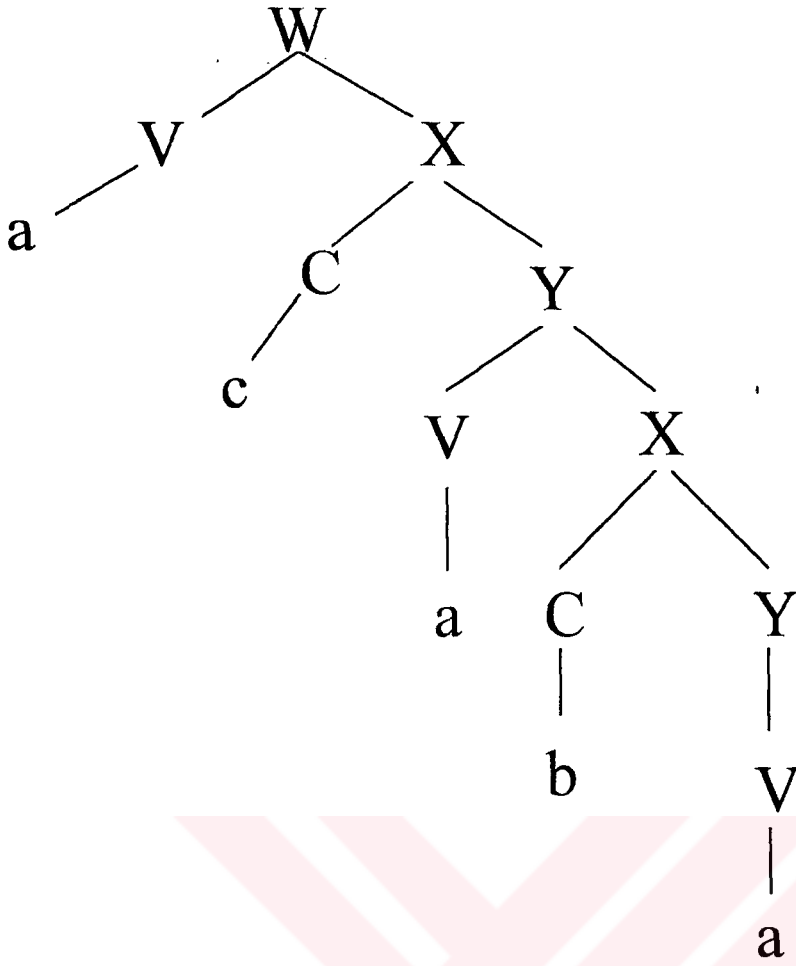
Acaba sözcüğü altı harfli alfabeden şu şekilde türetilebilir:

$$W \Rightarrow aX \Rightarrow acY \Rightarrow acaX \Rightarrow acabY \Rightarrow acaba$$

Kurallar sistemi oldukça karmaşık olmakla birlikte, bütün sesli ve sessiz harfler sisteme dahil olsaydı karmaşıklık artardı. Bu gramer için, orijinal tanıma eşdeğer daha basit bir tanım yapılabilir:

$V = \{a, b\}$ , ünlüler kümesi, ve  $C = \{b, c, d, f\}$ , ünsüzler kümesi olsun.  $V$  ve  $C$  terminal kümeleri veya sınıfları olarak tanımlanabilir, çünkü herbiri eşitliğin sol tarafında sınırlı bir şekilde yer alabilir. Gramer şu şekilde yeniden yazılabilir:

ACABA için yapım ağacı şekil 4.1' de gösterilmiştir:



Şekil 4.1 Acaba sözcüğü için yapım ağacı

$G, G = (N, \Sigma, T, P, S)$  şeklinde beş elemandan oluşan bir küme olsun.

- (a)  $N, \Sigma, P$  ve  $S$  tanımlarında daha önce yapılan tanımlarından bir farklılık olmamıştır.
- (b)  $T, T_1, T_2, \dots, T_n$  şeklinde gösterilecek olan bir terminal sınıfları kümesidir.
- (c)  $\Sigma, \Sigma = T_1 Y T_2 \dots Y T_n$  şeklinde tanımlanabilecek bir terminal sembolleri kümesidir.
- Terminal sınıfları,  $T_i, \Sigma$  ı sonlu sayıda alt kümeye parçalar.
- (d) Herhangi bir terminal sınıfı,  $T_i, T_i \rightarrow a_1 | a_2 | \dots | a_m$  şeklinde tek bir üretimin sadece sol tarafında yer alabilir.  $T_i$ , gerçek bir terminal olmayan değildir çünkü terminal alfabeti  $\Sigma$  dan alt küme oluşturmak için kullanılır.

Bu tanımın kullanılabileceği gramerler sınıflanmış gramerler adını alır.

## 5. DOĞAL DİLLERİN ÇÖZÜMLENMESİ

Bir cümlenin analizi, o cümleye sözdizimsel ve anlamsal yapı atanması gerekliliğini ortaya çıkarır. Sözdizim analizi, cümledeki sözcükler ve biçimbirimler arasındaki sözdizimi ilişkisi ile ilgilenirken, anlambilimsel analiz cümlenin anlam bütünlüğü ile ilgilenir. Bu bölümde biçimbilimsel çözümlene ve sözdizim analizi detaylandırılacaktır.

Biçimbilimsel çözümlene, bir sözcüğün yüzeysel şeklini, derin yapısına dönüştürme işlemidir. Bunun için birtakım kurallardan yararlanır.

- Yüzeysel şekil, bir sözcüğün metinde geçen şeklidir: Bu geçmiş zamanlı bir fiil, 3. kişiye çekimlenmiş bir fiil veya çoğul bir isim olabilir.
- Derin yapı, sözcüğün köklerine ve eklerine ayrılmış sözlükte bulunulabilecek halidir.

Biçimbilimsel çözümlenici, sözcükleri ayrıştırmak ve türlerini bulmak için sonlu durum makinelerini kullanır.

Doğal bir dilin sözdizimi için bir gramer yazmak, o dilde işlenecek bütün cümle yapılarını toplamak ve hepsini gramer-yazım kuarallarına uygun olarak sınıflamak anlamına gelir. Büyük bir sistemde, çok fazla kural yazılır. Kurallar arttıkça sistem yavaşlayacak ve gramerin hatalarını bulup gidermek daha da zorlaşacaktır. Doğal dillerdeki tanımsızlıklardan dolayı da işlem hızı azalacaktır. Sözcükler cümlede değişik şekilde dizilebilir ve değişik anlamlara sahip olabilir.

Gramerler, iyi oluşturulmuş cümleler elde etmek için sözcüklerin hangi biçimlerde biraraya getirilebileceği yollarını içeren kurallardan oluşur. Bir dil için, gramer ile çözümleniciyi birbirinden ayırmak çok önemlidir. Çözümlene, dilin makine işlemiyle ilgili olan parçasıdır. Her sözcüğü dilin grameriyle eşleştirilebilen yapılarla özdeşleştirilebildiği zaman bir cümle çözümlenmiş sayılabilir. Çözümlenici, grameri ve sözcük dizisini girdi olarak alır, sözcük dizisi gramere uygunsa sözcük dizisi için bir gramer yapısı çıktı olarak döner. Sözcük dizisi dilin gramerine uygun değilse, çıktı olarak hiçbirşey dönmez. Kavramsal olarak, çözümlenici (PARSER) ile gramer birbirinden farklıdır: Gramer, iyi oluşturulmuş yapısal nesnelere



kümesinin özet tanımıdır. Çözümleyici ise, gramerdeki nesnelere çıktı olarak veren bir algoritma olarak tanımlanabilir.

Recursive Transition Networks (Tekrarlayan Geçişli Ağlar) (RTN) gramer tanımlamanın yollarından birisidir. Bir RTN grameri, bir dizi ağın biraraya gelmesinden oluşur. Ağlar, söz dizimi kurallarına ait sınıflarla (TÜR) isimlendirilen, yaylarla bağlanan durumların (state) biraraya gelmesiyle oluşur.

Doğal Dil İşleme' de, sonraki yıllarda RTN'lerin daha ayrıntılı şekli olan ATN'ler (Augmented Transition Networks) (Genişletilmiş Geçişli Ağlar) kullanılmaya başlanmıştır. ATN'ler, RTN'lerle mümkün olabilen çözümleme tekniklerinden daha güçlü çözümleme tekniklerine duyulan ihtiyaç sonucu ortaya çıkmışlardır. 1970'li yıllarda Doğal Dil İşleme' de ATN gramerleri geniş bir şekilde kullanılmaya başlanmıştır. ATN'ler, ağdaki durumlar arasında geri dönüşlü graflarla dil tanıma veya dil çözümleme modelleri oluştururlar. ATN'ler hem bilgisayarla işlem yapılabilmesi açısından güçlüdür, hem de prosedürel yapıya sahiptir. Bu tezde yapılan çalışmalardan bir tanesi, ATN ağlarının Türkçe'ye uyarlanmasıdır.

### 5.1 Sonlu Durum Teknolojisi

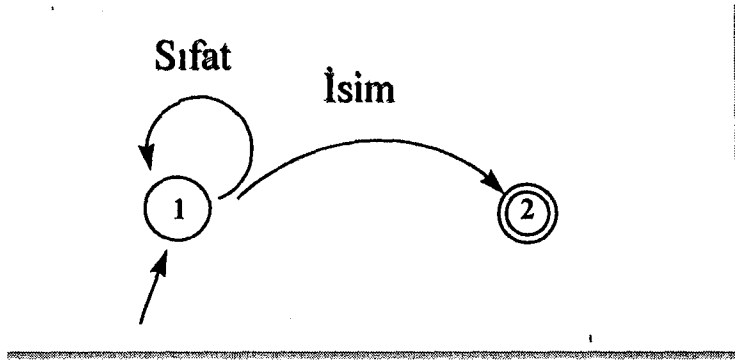
Sonlu durum teknolojisinin temeli otomataya dayanmaktadır. Otomata yaylarla ilişkilendirilen durumlar kümesinden oluşmaktadır. Yaylar bir durumdan diğer duruma geçmek için gerekli şartları göstermektedir. Otomata ifadelerle gösterilebilir ve doğal dil davranışını anlatabilecek kurallar yazmak için kullanılabilir : biçimbilimden (sözcük seviyesi) bütünsel çözümlmeye (cümle seviyesi) kadar.

Türkçe' de bir isim öbeği düzenli bir ifade olarak şöyle gösterilebilir:

Sıfat\* İsim

Türkçe' de sıfat mutlaka isimden önce gelir ve bir (veya birden çok ) sıfat ile onu (onları) takip eden isim öbeği meydana getirir. Yukarıdaki düzenli ifade (kural) Şekil 5.1' de gösterildiği gibi şu ifadeleri tanıyabilir : araba, kırmızı araba, büyük kırmızı araba, büyük ev.

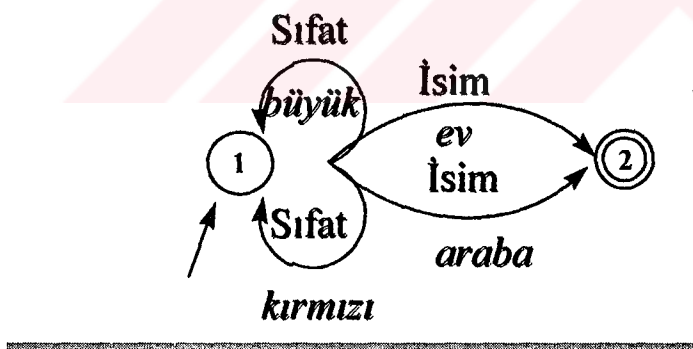
## Sıfat\* İsim



Şekil 5.1 Türkçe' de sıfat isim ilişkisi için sonlu durum makinesi

Şekil 5.1' de de görülebileceği gibi her yay sadece bir türle gösterilmiştir. Birden çok bilgi içeren yaylar da mümkündür. Şekil 5.2' de her yay birden çok bilgi içermektedir. Yayların altında sözcük , yayların üstünde de sözcüğün dahil olduğu tür gösterilmektedir. Bu çeşit otomataya iki seviyeli sonlu durumlu otomata veya iletme sistemi (transducer) adı verilir.

[Sıfat:[ büyük | kırmızı ]]\* İsim:[ ev | araba ]



Şekil 5.2 Türkçe' de sıfat isim ilişkisi için iki seviyeli sonlu durum makinesi

Sonlu-durum yaklaşımının teorik ve pratik açıdan bazı yararları vardır. Teorik açıdan iyi yani, sonlu-durum makinelerinin iyi bilinen özelliklere sahip iyi anlaşılabilir matematiksel varlıklar olmalarıdır. Sonlu-durum iletme sistemleri, birleştirilebilir, arakesitleri alınabilir. Pratik açıdan avantajları, sonlu durum hesaplama yöntemlerinin nerdeyse mükemmel yakın sonuç veren, hızlı ve dilden bağımsız mekanizmalar olmalarıdır.

## 5.2 Genişletilmiş Geçişli Ağlar

ATN' ler RTN' lerin genişletilmesiyle oluşturulan prosedürel mekanizmalardır.

Bu bölümde, önce Doğal Dil İşleme' nin temeli sayılan dönüşümsel gramer açıklanacaktır. Daha sonra ATN' lerin atası sayılan RTN' ler ele alınacaktır. Daha sonra RTN' ler esas alınarak , genişletme yolları açıklanacak ve ATN' l erle ilgili detaylı bilgi verilecektir. En sonunda da, ATN' lerle ilgili karşılaştırmalar yer alacaktır.

### 5.2.1 Dönüşümsel gramer

1960' lı yıllarda geliştirilen Chomsky'nin dönüşümsel grameri, karmaşık cümleleri basit hale getirmek için kullanıldı. Dönüşümsel gramerin iki ana bileşeni vardır: Birincisi , esas bileşen, diğeri de dönüşümsel bileşen. Esas bileşeni meydana getiren kurallar (öbek yapısı kuralları) cümleyi oluşturmak için kullanılır. Dönüşümsel bileşen, cümlenin oluşumunu tamamlamak için kullanılır. Dönüşümsel gramerde kullanılan başka bileşenler de vardır: Sözlük ve biçimbirimlerden yüzeysel yapıyı oluşturmak için kullanılan sözcükleri oluşturmak için gerekli morfolojik kurallar.

Temel bileşen ve öbek yapısı kuralları ortak yanları olan cümlelerle daha kolay açıklanabilir:

1.
  - a. Çocuklar kitabı masaya koyar.
  - b. Çocuk kitabı masaya koyuyordu.
  - c. Çocuk kitabı masaya koydu.
  - d. Çocuk kitabı masaya yerleştirmeliydi.
2.
  - a. Çocuklar kitabı masaya koydu mu?
  - b. Çocuk kitabı masaya koyuyor muydu?
  - c. Çocuk kitabı masaya koymuş muydu?
  - d. Çocuk kitabı masaya koymalı mıydı?
3.
  - a. Kitabı masaya kim koydu?
  - b. Çocuğun masaya koyduğu neydi?
  - c. Çocuk kitabı nereye koydu?

4.
  - a. Kitap masaya çocuk tarafından konuldu.
  - b. Kitap masaya çocuk tarafından mı konuldu?
  - c. Kitap masaya çocuk tarafından mı konulur?
  - d. Kitap masaya çocuk tarafından konulmalı mıydı?
  - e. Çocuk tarafından masaya konulan neydi?
  
5.
  - a. Çocuk kitabı masaya koymadı.
  - b. Çocuk kitabı masaya koymamalıydı.
  - c. Çocuk kitabı masaya koymamalı mıydı?
  - d. Kitabı masaya kim koymadı?
  - e. Kitap masaya çocuk tarafından konulmadı mı?

Bu cümlelerde ortak olan, bütün cümlelerde kullanılan KOYMAK eylemidir. Bunun yanında, bütün cümlelerde eylemi yapan bir kişi, eylemin yapıldığı nesne ve yer vardır. Farkları ise eylemin gerçekleştiği zaman, şimdiki zaman veya geçmiş zaman ve eylemin yapısıdır. Şekil yönünden de cümleler sınıflanabilir. Bazısı olumlu , bazısı değişik türde soru, bazısı etken, bazısı edilgen, bazısı da eylemin henüz gerçekleşmediğini anlatan olumsuz cümlelerdir.

Bu cümlelerin, şekil açısından birbirinden farklı oldukları halde anlam yönünden birbirlerine benzedikleri nasıl anlaşılabilir? Şekil ve anlam (derin yapı) arasında bir ayırım yapılmalıdır. Bu cümleler şekil ( yüzeysel yapı) yönünden birbirinden farklıdır. Anlam yönünden ( derin yapı) ise birbirinden çok farklı değillerdir. Temel bileşenin, bir cümlenin derin yapısını (anamını) oluşturan bir öbek yapısı grameri, dönüşümsel bileşenin ise bir cümlenin derin yapısını yüzeysel yapısına(son şekil) dönüştürmek için kullanıldığı söylenebilir.

### 5.2.2 ATN' lerin tarihçesi

1960'lı yıllarda geliştirilen ve önceki bölümde anlatılan Chomsky'nin dönüşümsel grameri, karmaşık cümleleri basit hale getirmek için kullanılıyordu. Fakat , bilgisayar kullanarak yapılan dönüşümleri geri alıp , cümleleri çözümleyerek anlayabilen bir yapı oluşturmanın nerdeyse imkansız olduğu ispatlandı. ATN' ler , dönüşümleri işlem sırasında geri alan çözümleyiciler yazmak için kullanılacak bir programlama dili olarak geliştirildi.

Doğal dil için geçiş ağları çözümlenme (parsing) prosedürünün kullanılması fikri Thorne tarafından ortaya atıldı ve daha sonra Bobrow ve Fraser tarafından uygulaması yapıldı. Woods' un, Apollo'nun aya ait örnekleri arasından jeolojik verileri elde etmek için kullanılan bir Doğal Dil İşleme sistemine ön-uç (front end) olarak kullanılmalarıyla ATN'lerin popülaritesi arttı.

Bölümün sonunda anlatılacak bazı olumsuzluklarına rağmen, ATN'ler bugüne kadar geliştirilen en başarılı çözümlenme tekniklerindedir. Aya ait sistem için ilk kullanılışından bu yana, mekanizma birçok dil-anlama sisteminde kullanıldı. Bugüne kadar geçen zamanda, bir çok dil için ATN grameri geliştirildi.

### 5.2.3 Tekrarlayan geçişli ağlar

Tekrarlayan Geçişli Ağlar (RTN ler) bize doğal dilde tekrarlayan yapılarla uğraşma imkianı verir. RTN'ler güç açısından içerik-bağımsız gramerlere eşdeğerdir ve genişletilmiş geçiş ağları için de temel teşkil ederler.

RTN gramerleri, nitelendirilmiş yönlü yaylarla bağlanan sonlu sayıda düğümler kümesinden (durumları gösterir) oluşan nitelendirilmiş bir grup ağdan oluşur. Yaylar, gramerde bir ağ niteleyen normal bir sözcükle, sözlüksel veya biçimsel sınıfla nitelenebilir. Sonlu-durumlu geçiş ağları (Finite-State Transition Networks)(FSTNler) ile RTN ler arasındaki temel fark adlandırılmış alt ağlardır. Bir yayın bir alt ağı temsil etmesi mümkündür. Temel fikir, ortak kullanılan yaylardan isimlendirilmiş yeni bir ağın oluşturulmasıdır. Yeni oluşturulan ağ, diğer ağlar tarafından ihtiyaç duyulan yerde, gramerin her yerinde açık olarak yer almak yerine adı ile işaret edilebilir.

Yeni alt ağlar oluşturmak içerik açısından, FSTN'lere küçük bir değişiklik gibi görünse de, bir ağı ters çevirme prosedürüne hatırı sayılır bir karmaşıklık getirir. Esas problem, tek bir yayı ters çevirmenin, tüm alt ağı ters çevirmeyi gerektirebilmesidir. Alt ağ ters çevrilirken, orijinal yay hatırlanmalıdır ki ters çevirme işlemi daha sonra oradan devam edebilsin. RTN ler, pushdown otomatanın (PDA) bir spesifikasyonudur. Bir RTN' de bir sözcük veya sözlük sınıfı yerine bir alt ağ ismi ile nitelendirilen bir yayı ters çevirmek için ismini veren alt ağı ters

çevirmek gerekir , bu yapıldıktan sonra nerden devam edileceği hatırlanmalıdır. Bir PDA' de bu amaç için kullanılacak bir yığın tutulur.

Verilen bir dizi sözcüğün RTN' in gramer yapısına uygun olup olmadığını anlamak için, başlangıç durumlarından birinden başlayan ve sonuç durumlarından binde son bulabilen bir yol bulmak şarttır. Her safhada , gelen sözcüğün ait olduğu sınıf ile nitelenen yay takip edilmelidir. Sadece ve sadece tüm bu koşullar gerçekleşirse, başarılı bir yol bulunmuş demektir.

RTN' ler bir dili tanımlamak için kullanılan kuralların yerine kullanılabilirler. Bir RTN gramerinde bir küme ağ vardır ve dilin karmaşıklığı arttıkça ağların sayısı da artar. Bir ağ, yönlü yaylarla birbirine bağlanan bir dizi durumdan (düğüm) oluşur. Bazı durumlar, ağa giriş ve çıkış durumlarını belirtmek üzere başlangıç ve terminal durumları olarak belirtilir. Bir yay, başındaki durumdan sonundaki duruma izin verilebilir bir geçiş içerir ve bu geçişin olabilmesi için yayın etiketi giriş sembolü ile aynı olmalıdır. Bir ağ belirli bir çözümlemeyi saklamak için kullanılmaz. İşlenen dil için mümkün olabilecek cümleleri karşılayabilecek modeller içerir. Geçiş ağlarında, durumları göstermek için çemberler ve çemberleri birbirine bağlayıp yayları oluşturmak için de oklar kullanılır.

RTN' lerde kullanılan yayların çeşitleri etiketlerine göre değişiklik arz edebilir. Kullanılan dört çeşit yay vardır:

1. **Sınıf Yayları:** En çok kullanılan yaylar, etiketi bir sözlük sınıfı olan sınıf yaylarıdır. Eğer giriş yapılan sözcüğün ait olduğu sınıf yaydaki etiket ile aynı ise yayın durum değişimi gerçekleşir ve giriş cümlesindeki ilk sözcük atlanmış olur.
2. **Sözcük Yayları :** Yayın etiketi sözcüğün kendisidir. Doğal dilin gramerinde spesifik ve sabit bir biçimsel fonksiyonu olan sözcükler için uygundur. Giriş cümlesinden , sadece bir sözcüğü işlediği için sınıf yaylarına benzer.
3. **Atlama Yayları :** Girdi olarak sözcük almadan, mümkün olabilecek hareketlerle gramerde geçişe izin veren çok özel bir yaydır. FSR ve PDA deki NULL-TRANSITION (Boş geçiş) olayına eşdeğerdir. Atlama yayları ağın formal gücünü değiştirmemekle birlikte yazılımı kolay kılar.

4. **İtme Yayları :** Etket bir ağın adı ise çözümleyici o anki durumu yığına atarak diğer ağa geçer. Geçilen ağda, kabul durumuna gelinemezse yay red olur. Red olmazsa, giriş cümlesinden bir sözcük daha alınarak devam edilir..

Basit bir S ağı (Cümle ağı) , Türkçe cümlelerin özne + -i hali , belirtili nesne + eylem yapıları için kullanılır.Nesne ve özne kullanımı Türkçe'de seçimlik olduğu için, atlama yayları kullanılmıştır. Böylece özne+yüklem, nesne+yüklem veya sadece yüklem kabul edilebilir olması sağlanmıştır.

Ağın sorunu, RTN gramerinin sınırlamalarından dolayı sözcüklerin hallerini kontrol edememesidir. Türkçe' de özne ile belirtili nesneyi ayıran tek ayrıntı hal bilgisidir. Özne yalın halde, belirtili nesne -i halinde olur. RTN' lerin hafızası ve yaylarla kabul edilen yapıları kontrol etmek için mekanizması olmadığı için hal kontrolü RTN grameri ile yapılamaz.

Diğer problem ise, RTN ile eylemin geçişli etken olup olmadığı kontrolünün yapılamamasıdır. Geçişsiz etken eylemlerin ve geçişli edilgen eylemlerin yüklem olduğu cümlelerde nesne bulunmaz. Bu durum, Türkçe için yazılan tüm gramerlerde göz önünde bulundurulmalıdır. RTN ile daha önce de belirtilen problemlerden dolayı bu mümkün değildir. Örneğin, ben bu sorunun yanıtını bildim cümlesinin yüklemi geçişlidir, cümle yanıtı-nı sözcüğünü belirtili nesne olarak alır. Bu cümle Türkçe için geçerli bir cümledir. Örnekteki RTN grameri tarafından da geçerli sayılır. Diğer bir örnekte, Ben eve gittim cümlesinin yüklemi geçişsizdir, bu cümlede nesne bulunmamalıdır. Bu cümle Türkçe için geçersiz bir cümle olmasına rağmen , RTN grameri bunu geçerli olarak kabul edecektir.

RTN' ler bu gibi durumlarda yetersiz kaldığından, doğal diller için daha güçlü çözümlenme tekniklerine ihtiyaç duyulmaktadır.

#### 5.2.4 Genişletilmiş geçişli ağlar

Genişletilmiş Geçişli Ağlar (ATN' ler) , Turing makinelerinin oluşturma gücüne sahip, RTN' lere dayanarak 1970' lerde oluşturulan prosedürel mekanizmalardır.

Ağdaki yaylara şartların ve hareketlerin eklenmesi ve saklayıcı(register) kullanılması ATN' lerin RTN' lerde bulunmayan özellikleridir.



ATN' lerin üstün tarafları ve kullanılışları şöyledir :

**1.Saklayıcılar (Registers) :** Saklayıcılar ağın ters çevrilmesi sırasında değerlerin hatırlanmasını sağlarlar, programlama dillerindeki değişkenlere benzetilebilirler. Çözümlemede, o anki duruma ait bilgileri tutarlar.

Bir ATN de iki çeşit saklayıcı vardır. Birinci tip saklayıcılar kullanıcı-tanımlı olup sadece tanımlanmış oldukları ağ için kullanılabilirler. Bu saklayıcılar, sadece ve sadece dilin gramerini yazan kişi tarafından dikkatlice değiştirilmelidir. Bu saklayıcılara görev ataması gramer yazarı tarafından yayların hareketler kısmında yapılır.

Ağda özel kullanıcı tanımlı saklayıcılar dışında , iki çeşit saklayıcı daha vardır: Birincisi, o anki değerleri otomatik olarak tutan Star (Yıldız) dır. Diğer bir deyişle, bir sınıf yayı için sözcüğü, bir itme yayı için de ağdan dönen yapıyı tutar. Uzun-mesafe bağımlılıkların üstesinden gelmek için kullanılan diğer saklayıcıya Hold(Tut) adı verilir. Bir durum dışında Star' a benzer bir şekilde uygulanır. Aralarında farklılık sağlayan durum, Hold saklayıcısı gramer yazarı tarafından değiştirilebilirken Star saklayıcısı değiştirilemez.

**3.Şartlar (Conditions) :** Şartlar, hangi yayın takip edileceğini gösteren kısıtlamalardır. Bir şart, giriş sembolünün ( Star saklayıcıda tutulur) yerel saklayıcı içeriklerine dayanarak oluşturulan Boolean bir kombinasyonudur. Giriş sembolü yayın etiketi ile aynı olsa bile şart yanlış döndürürse (NIL ile gösterilir) yay takip edilmez. Bu, ilk olarak giriş sembolü ve yayın etiketinden daha kesin sınırlamalar getirir, ikincisi, önceki durumlar hakkındaki bilgileri tutmak ve saklayıcı tarafından yapıları gelecek geçişleri belirlemek için kullanılabilir.

Bir yayın takip edilmesi için, önce şartlar kontrol edilir ve şartların gerçekleşmesi halinde durum değişir, en sonunda da yayın hareketler kısmındaki işlemler gerçekleşir.

**4.Hareketler (Actions) :** Hareketler, saklayıcı atama ve yapı-tasarımı işlemlerini gerçekleştirir. Çözümlemenin durumu, yerel ve global saklayıcılardaki değişimle bu kısımda değişir .

RTN' ler için anlatılan yay tipleri ATN' lerde de kullanılır. Tezde uygulanan ATN çözümleyicisinde kullanılan ek şartlar ve hareketler aşağıda gösterilmiştir:



- **(KATEGORİ<KATEGORİ ADI><ŞART>>HAREKETLER)**

Sınıf yaylarında, giriş dizisindeki ilk sözcük isim, eylem gibi <sınıf adı> nın varlığı için kontrol edilir. Eğer sözcük o sınıfa aitse, şart test edilir. Şartın gerçekleşmesi için hareketler uygulanır ve en sonunda durum değişir.

- **(KELİME<KELİME> <ŞART><HAREKETLER>)**

Sözcük yayları giriş dizisindeki ilk sözcüğün <sözcük> olup olmadığını kontrol eder. Bu adımdan sonrası, sınıf yayları gibi işlem görür.

- **(ATLA<ŞART><HAREKETLER>)**

Atlama yayları, daha önceki yay tiplerine benzemekle birlikte herhangi bir sözcük tüketimine olanak sağlamazlar.

- **(İTME<ŞART1><AĞ><ŞART2><HAREKETLER>)**

İtme yaylarında, çözümlemenin içinde bulunduğu durum için testler içeren <ŞART1> öncelikle test edilir. Eğer şart sağlanırsa, içinde bulunulan durum yığına atıldıktan sonra kontrol <AĞ> a geçer. <AĞ> çekildikten sonra, <AĞ> dan dönen yapıya uygulanacak testlerin yer aldığı <ŞART2> test edilir. <ŞART2> de gerçekleşirse durum değişimine neden olacak <HAREKETLER> uygulanır.

### 5.2.5 Açıklayıcı ve prosedürel şekilciliğin karşılaştırılması

Bilgisayar programlama , bir bilgisayara bir görevi nasıl gerçekleeyeceğini anlatmak için bir bilgi kümesini verme aktivitesidir. İnsanların sahip olduğu bir çok bilgi prosedürel yolla ifade edilir. Diğer bir yol ise, kuralların ve prensiplerin, program tarafından değiştirilecek şekilde açıklayıcı sembolik yapılar şeklinde ifade edilmesidir.

ATN-tabanlı çözümleyiciler 1970' li yıllarda en çok kullanılan çözümleyiciler arasındaydı. Sonraki yıllarda kullanımlarında azalmalar görüldü. ATN' lerin birçok dil anlama sisteminde çok kullanışlı olduğu kanıtlanmasına rağmen bazı dezavantajları vardır :

1.Gramerin RTN kısmı açıklayıcı bir yapıya sahiptir ve anlaşılması kolaydır, fakat geçişlerin eklenmesiyle biçimin açıklayıcı yapısı bozulur.

2.Cümlenin bütün sözcükleri sistem tarafından bilinmezse ve cümlenin yapısı ağdaki bir yola karşılık gelmezse çözümleme işlemi başarısız olur. Kısmi eşleme yapma gibi bir yeteneği yoktur. Cümledeki problemin ne olduğunun anlaşılması için çok fazla bir yardımı yoktur.

3.ATN gramerinin gelişimi modüleriteden yoksundur. Gramerin herhangi bir kısmındaki değişiklikler istenmeyen yan etkilere neden olabilir.

Yukarıdaki problemlerden dolayı, insanlar gramer tanımlamak için daha açıklayıcı biçimlere dönüş yapmışlardır. Prosedürel biçimlerin aksine açıklayıcı biçimler dil işleme modellerine referans olmadan kolayca anlaşılabilirler.

Fakat, bütün bu dezavantajlarına rağmen, ATN her zaman için kullanışlı bir mekanizma olarak dil işleme çalışmalarında yer almıştır.



## 6. İSTATİSTİKSEL YAKLAŞIM

Bazı problemlerin çözülmesi için kesin olmayan ama desteklenen yanları olan bazı doktrinlerin ortaya atılması gerekmektedir. Bu problemler iki sınıfa ayrılabilir.

Birinci sınıf; dünyada rasgelelik gösteren problemleri içerir. Tüm kart oyunları buna örnek gösterilebilir. Bu tipte problemler için kesin birşey söylemek mümkün olmazken, daha önce çıkan çeşitli sonuçlarla dayanarak çeşitli varsayımlarda bulunulabilir.

İkinci sınıf; rasgelelik içermez, sadece belli durumlarda ortaya çıkabilecek bazı istisnai durumlar sözkonusudur. Zor alan, birer birer sayabileceğimizden daha fazla istisnainin var olmasıdır. Uzman bilgisine ihtiyaç duyulabilecek tüm konular, tıpta teşhis ve edebiyatta yazar sınıflaması da dahil olmak üzere, bu sınıfa girer. Bu konularda, istatistiksel ölçüler çözümlenen dünyanın özetini oluşturduğu için, mümkün olabilecek bütün istisnai durumların dikkate alınmasına gerek kalmamaktadır. Bütün istisnaları dikkate almak yerine, ne kadar sıklıkta bir istisnaya karşılaşılmışının beklenildiğini söyleyen nümerik bir özet kullanılabilir.

### 6.1 Olasılık ve Bayes Teoremi

Çoğu problem-çözme sistemi için, sistem çalıştığı sürece veri toplamak ve bu verilere dayanarak sistemin davranış şekline karar vermesini sağlamak çok önemlidir. Bu davranışı medelleme için, verilerin istatistiksel teorisine ihtiyaç duyulmaktadır. Bayesian istatistiği böyle bir teoridir. Bayesian istatistiğinin ana notasyonu şarta bağlı olasılıkla aynıdır:

$$P(H|E) \quad (6.1)$$

Yukardaki ifade, E durumu olduğu zaman H hipotezinin olasılığını göstermektedir. Bunu hesaplamak için, H' nin durum meydana gelmeden önce sahip olduğu olasılık değerine ihtiyaç duyulmaktadır. Bunu sağlamak için, birbirinden bağımsız  $H_i$  değerleri içeren bir uzay tanımlanmalıdır.

$P(H_i|E)$  = E durumunda  $H_i$  hipotezinin gerçekleşme olasılığı

$P(E|H_i)$  = Hipotez  $i$ ' nin doğru olması durumunda E durumunu sağlama olasılığı

$P(H_i)$  = Spesifik bir durumun eksikliği sonucunda E durumunu sağlama olasılığı.

Bu tip olasılıklar öncelik olasılığı adını alır.

$k$  = olası hipotez sayısı

Bayes teoremi aşağıdaki denklemde görüldüğü gibi ifade edilir:

$$P(H_i) = \frac{P(E|H_i) \cdot P(H_i)}{\sum_{n=1}^k P(E|H_n) \cdot P(H_n)} \quad (6.2)$$

Örneğin, belirli bir yerin istenen mineralleri bulmak için kazılmak üzere seçilmesi için iyi bir yer olup olmadığını belirleyen jeolojik olayı düşünelim. Eğer, minerallerin önceki bulunma olasılıkları ve minerallerin o anda bulunma olasılıkları biliniyorsa ve yer hakkında belli fiziksel özellikler çıkarılabiliyorsa, minerallerin bulunma olasılıkları Bayes teoremi ile hesaplanabilir.

Bayes teoremi hesaplanmasındaki karmaşıklıktan dolayı dezavantajı büyük olan bir teoremdir. Durumlar arttıkça hesaplama karmaşıklığı artacağından ve bir yazar tanıma sisteminde durum sayısı çok fazla olduğundan tez için tercih edilmemiştir ama doğal dil işleme uygulamalarından olan sözcük yazımı hatalarını düzeltmede ve doküman erişiminde kullanılan bir yöntemdir. Bayes teoreminin bu dezavantajından dolayı aşağıdaki yöntemler geliştirilmiştir :

- Kurallara kesinlik faktörleri atamak
- Bayesian Ağları
- Demster-Shafer teorisi

Ayrıca tezde kullanılmak üzere seçilen ve belirsizliğe bu yöntemlerden oldukça farklı bir nümerik yaklaşım gösteren Bulanık Mantık yöntemi vardır.

## 6.2 Kesinlik Faktörleri ve Kural-Tabanlı Sistemler

Bayesian ağları; bakteriyel enfeksiyonu olan hastalara uygun terapiler önermesi için tasarlanan MYCIN sisteminde kullanılmıştır. [Shortliffe,1976;Buchanan ve Shortliffe, 1984; ShortLiffe ve Buchanan,1975)

Sistemin ihtiyaç duyduğu fiziksel bilgilerin uzman bir kişi tarafından girilebilmesi için etkileşimli bir sistem olarak tasarlanmıştır. MYCIN uzman sisteme örnek olarak gösterilebilir; çünkü ancak bir uzmanın yapabileceği işi yapmak üzere tasarlanmıştır. Burada olasılık usulundan bahsedilecektir.

MYCIN teşhiste yardımcı olacak verileri bir kurallar kümesi içinde tutar. Her kuralın bir kesinlik faktörü vardır. Kesinlik faktörü, koşulların gerçekleşmesi halinde kuralda ortaya konan olayın gerçekleşme ölçüsüdür. Tipik bir MYCIN kuralı aşağıdaki gibidir:

- Eğer
- (1) Organizma lekeli ise
  - (2) Organizmanın iç çekirdeği varsa
  - (3) Organizmanın büyümesi yavaşsa

Bu organizmanın staphylococcus olma olasılığı 0.7 dir.

MYCIN bu kuralları, varolan klinik veriyle sebep ilişkisi kurarak hastalığa neden olan organizmayı bulmak için kullanır. Organizmanın kimliğini bulduktan sonra, hastalığın iyileşmesini sağlayacak bir tedavi yöntemi önermeye çalışır. MYCIN' ın kesin olmayan bilgiyi nasıl kullandığını anlamak için şu sorulara cevap vermek gerekmektedir: “Kesinlik faktörleri ne anlama gelir?” ve “MYCIN bir sonuç faktörü elde etmek için bütün kurallarına ait kesinlik faktörlerini nasıl birleştirir?” Bölümün geri kalanında bu soruların cevapları üzerine yoğunlaşılacaktır.

Bir kesinlik faktörü ( $CF[h, e]$ ) iki bileşen kullanılarak ifade edilebilir:

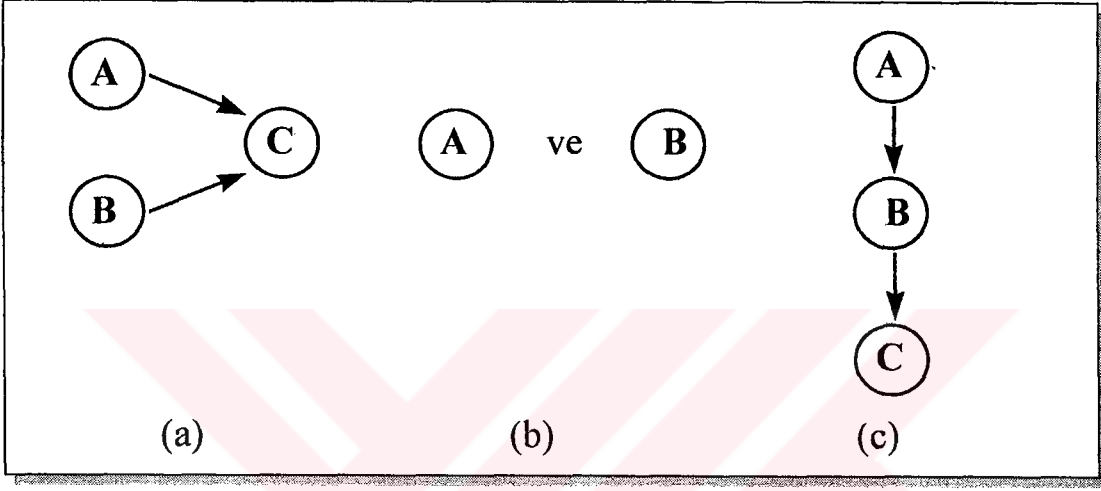
- $MB[h, e]$  : h hipotezinin e durumunu ortaya çıkarmasına duyulan inancı gösteren 0 ile 1 arasında bir ölçüdür.
- $MD[h, e]$  : h hipotezinin e durumunu ortaya çıkarmamasına duyulan inancı gösteren 0 ile 1 arasında bir ölçüdür.

Bu iki ölçü kullanılarak kesinlik faktörü şöyle tanımlanabilir.

$$CF[h,e] = MB[h,e] - MD[h,e] \quad (6.3)$$

Herhangi bir durum bir hipotezi ya destekleyeceği ya da reddedeceği için ve her MYCIN kuralı bir (veya) birden çok duruma eşleştirebileceği için, her kural için tek bir sayı MD'yi, MB'yi ve CF'yi göstermek için yeterlidir.

MYCIN'ın CF'leri kuralları yazan uzmanlar tarafından sağlandı. Probleme uygulanan kuralların hepsinin CF'sini birleştirerek sonuca gitmek için tek bir CF elde etmek gerekmektedir.



Şekil 6.1 Belirsiz kuralları birleştirmek

Şekil 6.1 gözönünde bulundurulması gereken üç birleştirme senaryosunu göstermektedir. Şekil 6.1 (a) da tek bir hipotez için kanıt sağlayan birden çok kural gösterilmiştir. Şekil 6.1 (b) birden çok önerme biraraya getirerek sonuca gitme yolunu göstermektedir. Şekil 6.1 (c) bir kuralın çıktısı diğer kurala girdi hazırlamaktadır.

Bu kombinasyonları gerçeklemek için ne gibi formüller kullanılması gerektiğinden bahsedilmeden önce fonksiyonların biraraya getirilmesi sonucunda sağlanmak istenen özellikler anlatılmalıdır:

- Elde edilen sonuçların bir sırası olmadığı için birleştirme fonksiyonlarının birleştirme ve yer değiştirme özelliği olmalıdır.
- Kesinliğe ulaşıncaya kadar ek destekleyici kanıtlar MB'yi artırmalıdır. Aynı şekilde desteklemeyen kanıtlar MD'yi azaltmalıdır.
- Belirsiz çıkarımlar zincirleme olarak biraraya getirilirse, belirsizlik tek bir çıkarıma göre artacaktır.

Bu özellikler dikkate alınarak Şekil 6.1 (a)'daki senaryo şöyle değerlendirilebilir:

Hipoteze inanma ve inanmama ölçüleri aşağıdaki gibi hesaplanan  $s_1$  ve  $s_2$  gözlemlerini ortaya çıkarır:

$$MB[h, s_1 \wedge s_2] = \begin{cases} 0 & MD = [h, s_1 \wedge s_2] = 1 \text{ ise} \\ MB[h, s_1] + MB[h, s_2] * (1 - MB[h, s_1]) & \text{Diğer durumlarda} \end{cases} \quad (6.4)$$

$$MD[h, s_1 \wedge s_2] = \begin{cases} 0 & MB = [h, s_1 \wedge s_2] = 1 \text{ ise} \\ MD[h, s_1] + MD[h, s_2] * (1 - MD[h, s_1]) & \text{Diğer durumlarda} \end{cases} \quad (6.5)$$

Örnek olarak;  $h$ 'nin gerçekleşeceğine olan inanç  $MB=0.3$  olarak gösterilirse,  $MD[h, s_1] = 0$

$MB[h, s_1] = 0.3$  olur. Ayrıca,  $MB[h, s_2] = 0.2$  olan bir kural daha konulsun.

$$MB[h, s_1 \wedge s_2] = 0.3 + 0.2 * 0.7 = 0.44$$

$$MD[h, s_1 \wedge s_2] = 0$$

$$CF[h, s_1 \wedge s_2] = 0.44$$

Yukarıdaki örnekten de görülebileceği gibi, uygulanan hesaplama kesinlik faktörünü artırmıştır.

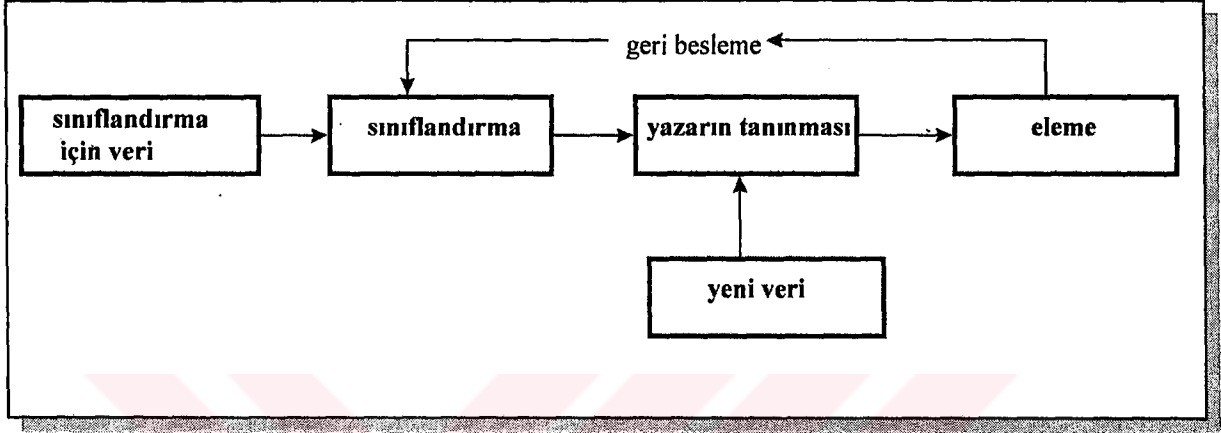
Şekil 6.1 (b)'de birkaç hipotezi birleştirerek kesinlik faktörü hesaplama ihtiyacı duyulmaktadır. Bu ihtiyaç, kural sayısı ikiden fazla olduğu zaman ortaya çıkmaktadır. Kesinlik faktörünün kombinasyonu MB ve MD kullanılarak hesaplanabilir.

Şekil 6.1 (c)'de kurallar birbirine zincirlenmiştir ve birinin çıktısı diğerine girdi olarak kullanılmaktadır.

### 6.3 Bulanık Mantık

Daha önce anlatılan tekniklerde; küme teorisi ve mantıkla sağlanan matematiksel bilgiler kullanılmamıştır. Bu bölümde tezde yazarın tanınması için kullanılan bulanık mantık teorisi ve tanımlanan üyelik derecesi fonksiyonları anlatılmıştır.

Bir metnin yazarının tanınabilmesi için, doğal dil işleme sonucunda ortaya çıkan metnin dilbilgisel özelliklerine dayanarak metni uygun olduğu sınıfa yerleştirmek gerekmektedir. Metnin çözümlenmesi sonucunda elde edilen veriler kullanılarak bir sınıflandırma işlemi yapılır. Sınıflandırma ve yazar tanıma sistemini Şekil 6.2' deki gibi göstermek mümkündür. Buna göre sınıflandırma işleminde metnin çözümlenmesiyle elde edilen veriler kullanılmaktadır. Yazarın tanınması işlemi ise, metnin yazarının sınıflandırma işlemi sonucunda ortaya çıkan sınıflardan hangisine ait olduğunu bulur.



Şekil 6.2 Yazarın tanınması için kullanılan bulanık mantık sistemi

### 6.3.1 Yazarların Bulanık Mantık kümeleri

Yazarlarda, ayırt edici özellikler olarak, Doğal Dil İşleme sonucu elde edilen şu veriler kullanılmıştır:

- Cümle başına sözcük sayısı
- Metindeki Türkçe sözcüklerin tüm sözcüklere oranı
- Metindeki yabancı sözcüklerin tüm sözcüklere oranı
- Tekrar eden sözcüklerin tüm sözcüklere oranı
- Ortalama sözcük uzunluğu
- Sıfat sayısının tüm sözcüklere oranı
- İsim sayısının tüm sözcüklere oranı
- Eylem sayısının tüm sözcüklere oranı
- Zarf sayısının tüm sözcüklere oranı
- Edat sayısının tüm sözcüklere oranı
- Bağlaç sayısının tüm sözcüklere oranı
- Sıfat sayısının tüm sözcüklere oranı



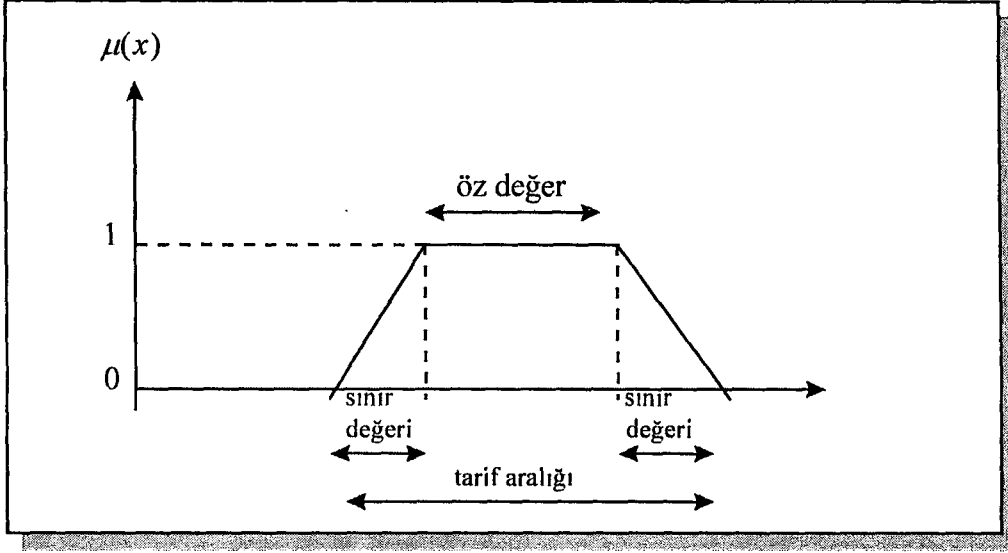
- Ünlem sayısının tüm sözcüklere oranı
- Cümledeki ortalama isim sayısı
- Cümledeki ortalama sıfat sayısı
- Cümledeki ortalama zarf sayısı
- Cümledeki ortalama eylem sayısı
- Cümledeki ortalama edat sayısı
- Cümledeki ortalama bağlaç sayısı
- Cümledeki ortalama ünlem sayısı
- Soru cümlelerinin tüm cümlelere oranı
- Devrik cümlelerinin tüm cümlelere oranı
- Eksilteli cümlelerin tüm cümlelere oranı

### 6.3.2 Bulanık üyelik derecesi

Bulanık Kümeler Teorisinde, belli bir özelliğe göre oluşmuş bir sınıfın bir elemanı esas alındığında diğer üyelerin bu elemana olan benzerliği ( $\mu$ ) üyelik derecesi ile gösterilir. İncelenen küme bir (A) kümesi ise bu kümedeki bir x elemanının üyelik derecesi  $\mu(x)$  olarak gösterilir. Tam benzeme durumunda  $\mu=1$ , hiç benzememe durumunda  $\mu=0$  olarak alınır.

$\mu$  0 ile 1 arasında değerler alır ( $0 \leq \mu \leq 1$ ). ( $\mu$ ) üyelik derecesi  $\mu(x) = f(x)$  şeklinde bir sürekli fonksiyonla ifade edilebildiği gibi (x)' in çeşitli değerlerine karşılık kesikli olarak da gösterilebilir.

$\mu(x) = f(x)$  sürekli fonksiyonunda (x) bağımsız değişkeninin belli değerler (sınır değerleri) aralığında tarifli olması gerekmektedir. Şekil 6.3' de bir  $\mu(x)$  fonksiyonunun sınır değerleri, plato şeklindeki ( $\mu(x) = 1$ ) öz değer ve sınır değerleri görülmektedir.



Şekil 6.3 Bir Üyelik Derecesi Fonksiyonu

### 6.3.3 Üyelik derecesi değerleri

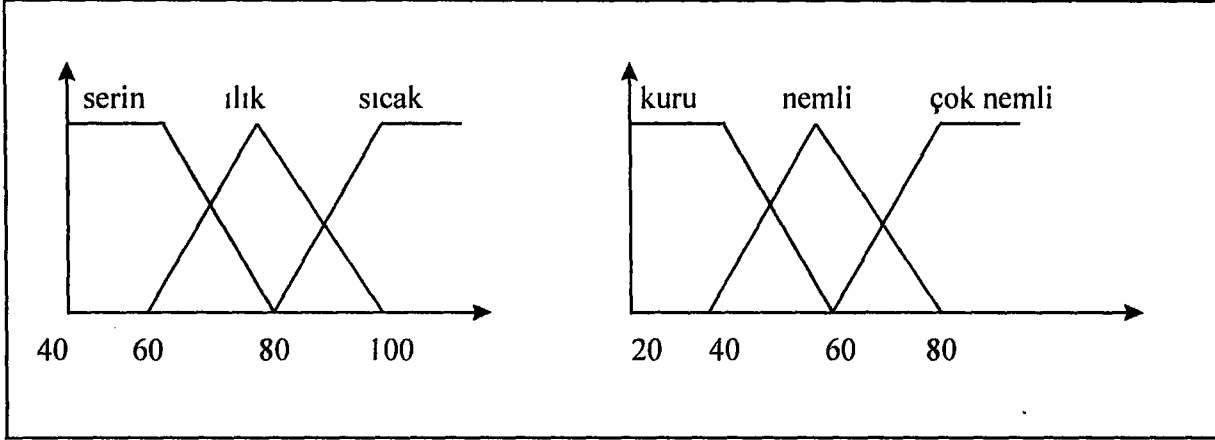
Bulanık mantık değişkenlerine üyelik dereceleri atamaları çok değişik şekillerde yapılır. Bu değer atamaları sezgisel yoldan olabildiği gibi bazı algoritmik ve mantıksal işlemlerle de gerçekleştirilebilir. Bulanık mantık değişkenlerine üyelik derecesi değeri atamak için kullanılan bazı yöntemler şöyle sıralanabilir:

- Sezgisel Yaklaşım
- Çıkarım
- Tümevarım
- Genetik Algoritmalar
- Bulanık mantık istatistiği
- Yapay sinir ağları
- Derecelendirme

Bu tez çalışmasında, yazarın tanınmasını sağlamak için sezgisel yaklaşım ve bulanık mantık istatistiği yöntemleri kullanılmıştır.

Sezgisel yaklaşımda, uygulanacak alanının durumuna göre ( $x$ ) değişkeninin fiziksel yapısına bağlı olarak değişim gösterecek bir  $\mu(x)$  fonksiyonu tarif edilir. Örneğin; sezgisel yaklaşım

kullanılarak hava sıcaklığına göre nemlilik ve sıcaklık üyelik derecesi fonksiyonları tarif edilebilir. Tarif edilen üyelik derecesi fonksiyonları Şekil 6.4' te gösterilmektedir.



Şekil 6.4 Üyelik fonksiyonları (a) sıcaklık (b) nemlilik

#### 6.3.4 Yazarların üyelik derecelerinin elde edilmesi

Doğal Dil İşleme işlemi sonucunda elde edilen ve bölüm 6.3.1' de açıklanan, çözümlenen metne ait on beş özellik için on beş tane üyelik fonksiyonu tarif edilmektedir. Üyelik fonksiyonları tanımlanırken, yazarlara ait istatistiksel veriler kullanılmaktadır. Sistem yazarı bilinen metinler geldikçe, yazar veritabanındaki on beş özelliğe ait verileri güncellemektedir. Yazarı bilinmeyen bir metin sisteme girdi olarak verilirse, üyelik derecesi fonksiyonları kullanılarak yazarı bilinmeyen metin için 0 ile 1 arasında bir rakam hesaplanır; en büyük değere sahip yazar metnin sahibi olma olasılığı en çok olan yazardır.

On beş özellik için de aynı üyelik fonksiyonu kullanılacağı için sadece metinde cümle başına düşen sözcük sayısı için üyelik derecesi fonksiyonu örnek olarak tanımlanacaktır.

$ck_i$  = Metinde cümle başına düşen sözcük sayısı

$n$  = Veritabanında bulunan yazar sayısı olsun.

Bu durumda, cümle başına düşen sözcük sayısı ( $ck_i$ ) olan bir metnin , cümle başına düşen sözcük sayısı kullanımı ( $ck_j$ ) olan bir yazara ait olma olasılığı ( metinle yazara ait metinlerin benzerliği) , ( $ck_i$ ) esas alındığında

$(ck_i > ck_j)$  için  $\mu_{ij} = \frac{ck_j}{ck_i}$  olur.

Eğer seçilen (j) yazarına ait cümle başına sözcük sayısı ( $ck_j > ck_i$ ) ise

$\mu_{ij} = \frac{ck_i}{ck_j}$  olarak yazılır.

Cümle başına sözcük oranları için üyelik dereceleri genel olarak

$$\mu_{ij(ck)} = \begin{cases} \frac{ck_j}{ck_i} & ck_i > ck_j \text{ ise} \\ \frac{ck_i}{ck_j} & ck_i < ck_j \text{ ise} \end{cases} \quad (6.6)$$

yazılır.

Bu fonksiyonda  $ck = ck_i$  için  $\mu_{ck} = 1$  olur.  $j=1,2,3,\dots,n$  için  $ck_i, ck_j$  benzerliklerine bakılırsa;

1.  $ck_{\min} < ck_j < ck_i$  aralığında  $\mu_{ij} = \frac{ck_j}{ck_i}$  olup, bu aralıkta  $ck_i$  sabittir ve değişim  $f(x) = m * x + n$  tarzında olup lineerdir.
2.  $ck_i < ck_j < ck_{\max}$  aralığında  $\mu_{ij} = \frac{ck_i}{ck_j}$  olup buradaki değişim  $f(x) = \frac{m}{x}$  tarzındadır.

Metnin çözümlenmesi sonucunda elde edilen diğer veriler için de aynı şekilde üyelik fonksiyonları tanımlanır ve her bir veri için yazarın üyelik derecesi hesaplanır.

### 6.3.5 Netleştirme birimi(Defuzzification)

Sınıflandırma işlemi sonucu elde edilen benzerlik değerlerinin, kesin çıkış ve sayısal değerlere dönüştürüldüğü kısımdır. Bu birimde kullanılan üç yöntem vardır:

- a. Maksimum ölçüt metodu: Netleştirme biriminde elde edilen benzerlik değerlerinin en büyük üyelik derecesi alınır.
- b. Maksimum ortalama metodu: Üyelik fonksiyonlarından elde edilen maksimum benzerlik değerlerinin ortalamasıdır ve maksimum ortalama aşağıdaki formülle hesaplanır:

$$Y_0 = \frac{\sum_{j=1}^n W_j}{I} \quad (6.7)$$

$W_j, \mu_j(w_j)$  : üyelik fonksiyonundan elde edilen maksimum değer ve

I : üyelik fonksiyonlarından elde edilen değerlerin sayılarıdır.

c. Ağırlık merkezi ortalaması (Center of Gravity) : Uygulamalarda yaygın bir şekilde kullanılan yöntemlerden biri olup üyelik fonksiyonlarının kesin çıkış değerlerinin ağırlık merkezidir. Buna göre

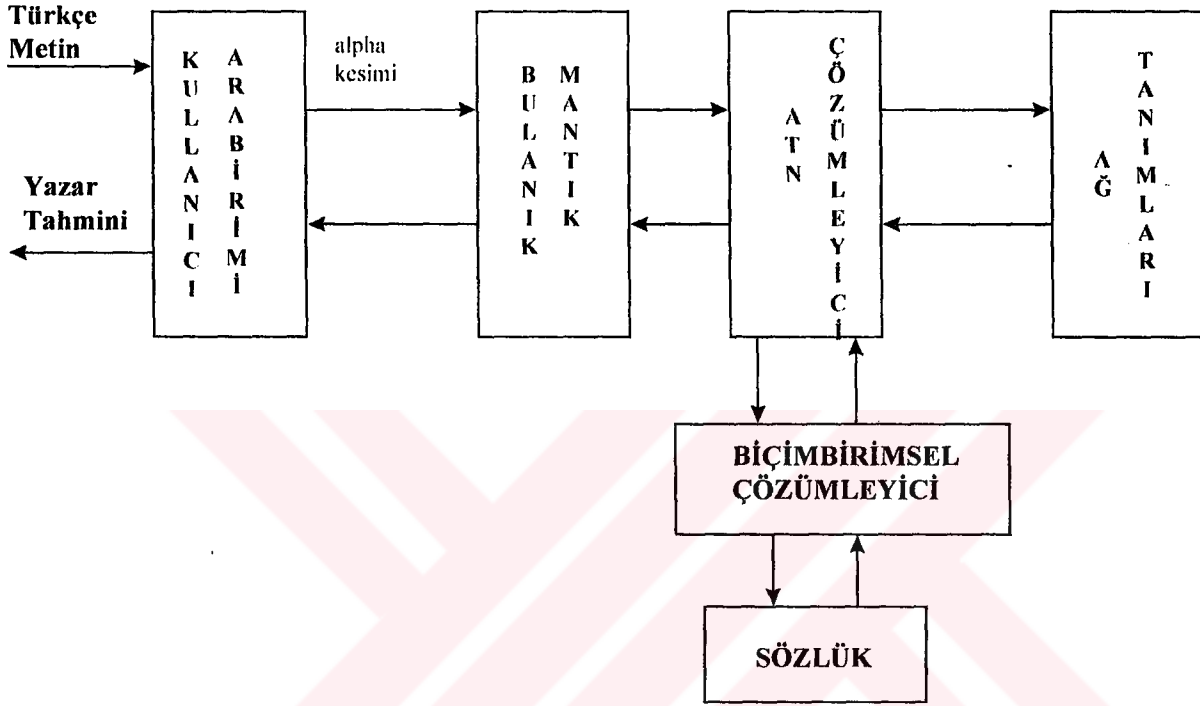
$$Y_0 = \frac{\sum_{j=1}^n \mu_j(W_j) \cdot W_j}{\sum_{j=1}^n \mu_j(W_j)} \quad (6.8)$$

Burada n: Üyelik fonksiyonlarından elde edilen değerlerin sayısıdır.

Geliştiren sistemde elde edilen üyelik derecesi fonksiyonlarının çıkışlarını tek bir sayısal değere dönüştürmek için Ağırlık Merkezi Ortalaması yöntemi kullanılmıştır.

## 7. UYGULAMA

Çalışmamamızın amacı, girilen metne ait gramer açısından ayırt edici bilgilerin çıkarılmasıdır. Bu bilgiler bir veritabanında tutularak, istatistiklerden yararlanarak yazarı bilinmeyen metinlere ait yazar tahmini yapılması için kullanılacaktır.



Şekil 7.1 Türkçe metni çözümleyen ve yazarını tanıyan sistemin yapısı

Şekil 7.1 sistemin yapısını göstermektedir. Çalışmada kullanılan bileşenler birimlere ayrıştırıcı, normalize edici, biçimbilimsel çözümleyici, sözcük veritabanı, ATN çözümleyici ve yazarın tanınması için Bulanık Mantık yöntemidir.

Çalışmanın basit bir arabirimi vardır. Çözümlemesi istenen metinler text dosyası olarak saklanmalıdır. Kullanıcı metni seçtikten sonra, yazılan programda biraz önce sözü geçen bütün bileşenler sırası ile çalışmakta ve metne ait bilgiler ekrana geldiği gibi, text dosyalarında saklanmaktadır. Ayrıca, yazarlarla ilgili veritabanı oluşturulduğunda metne ait yazar tahmini de yapılmaktadır.

## 7.1 Sembollere Ayrıştırıcı

Bilgisayarlar aptal makinelerdir. Yeni bir metin girdi olarak verildiğinde ilk istenen metnin önce cümlelere, sonra da sözcüklere ayrıştırılmasıdır. Türkçe’ de cümleler birbirinden boşlukla ayrılan sözcüklerden oluşur. Çoğu dilde olduğu gibi, Türkçe’de de kesintisiz devam eden alfabetik karakterler sözcüğü oluşturur. Cümle sonlarında ise noktalama işaretlerinden nokta, ünlem ve ya soru işareti olabilir. Ama bu üç noktalama işaretinden biriyle karşılaşılması cümlenin bittiği anlamına gelmemelidir. Örneğin, cümle içinde karşılaşılan vb. birimindeki noktanın cümlenin sonu anlamına gelmediği anlaşılmalıdır. Geliştirilen sembollere ayrıştırıcı bu kriterler göz önüne alınarak tasarlanmıştır.

Sembollere ayrıştırıcı, sözcük , cümle ve paragraf sınırlarını tanımaktadır. Bu amaç için sonlu durum teknolojisi kullanılmıştır.

Dr. Ahmet Yılmaz biraz önce geldi.

cümlesi girildiği zaman sembollere ayrıştırıcı aşağıdaki (birimleri) döndürmelidir :

Dr.

Ahmet

Yılmaz

biraz

önce

gel

di

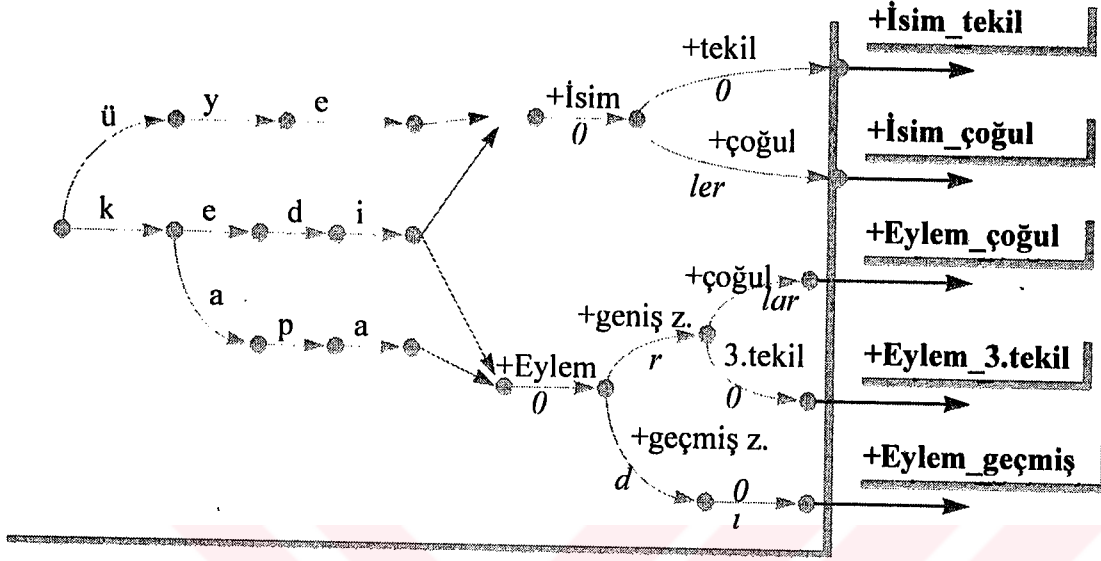
Birimlere ayrıştırıcının kullandığı sözlüklere, karşılaşılan yeni sözcükler eklendiği için sözcük ile sonraki karşılaşmada ayrıştırıcı sözcüğü tanımaktadır.

## 7.2 Normalize Edici

Sözcüklerin standard hale gelmesini sağlamaktadır. Örneğin, başlıkta büyük harfle yer alan sözcüğü ve ya özel adların ilk harflerini küçük harfe çevirerek sözlükte bulunmalarını sağlamaktadır.

### 7.3 Biçimbilimsel Çözümleyici

Çalışmada geliştirilen biçimbilimsel çözümleyici, sözcükleri ayırtırmak ve türlerini bulmak için sonlu durum makinelerini kullanmaktadır.



Şekil 7.2 Türkçe sözcükler için biçimbilimsel çözümleyici

Şekil 7.2 'de üye, kedi ve kapa sözcükleri bir sonlu-durumlu biçimbilimsel sistemde gösterilmektedir. Bu şekilde, sözcüklerin yüzeysel şekilleri yayların üstünde, biçimsel açıdan yapıları da yayların altında gösterilmiştir. Değişik yollar takip edilerek kedi ve üye isimlerinin çoğul hallerine ve ya kapamak eyleminin geniş zamanına veya geçmiş zamanına ulaşılabilir. Bir sözlüğün iletme sistemi ile gerçekleştirilmesinin birkaç avantajı vardır.

İletme sistemleri, sonlu durum, makineleri olarak bilginin sıkıştırılmasına olanak sağlarlar. İletme sistemleri iki yönlüdür ve sözcüklerin çekimlenmiş halinden kökünü elde etmek için kullanılabilirler gibi, yalın halden çekimli halini elde etmek için de kullanılabilirler. Bir avantajları da hızlı olmalarıdır.

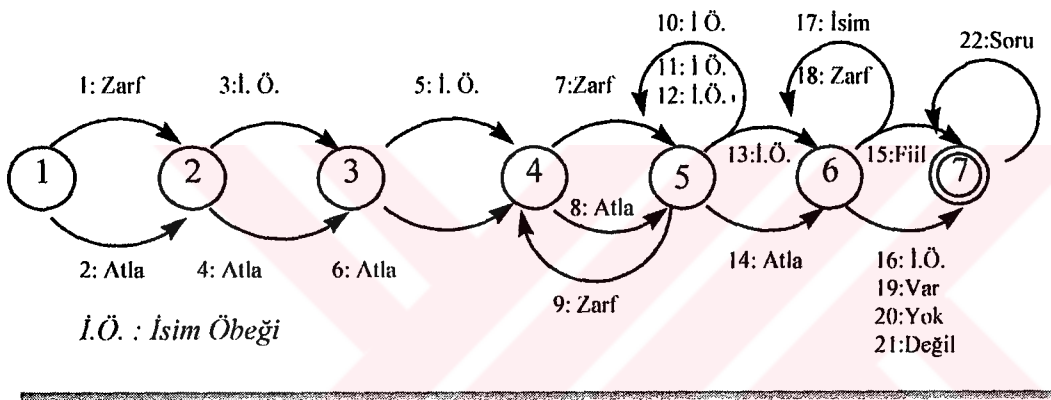
Tezde kullanılma nedenleri, sözcük türlerini bulabilmek için sözcüklerin yalın hallerine duyulan ihtiyaçtan kaynaklanmaktadır. Çünkü sözcüklerin ve türlerinin tutulduğu sözlükte sözcükler yalın halleri ile yer almaktadır.



## 7.4 ATN Çözümleyici

Sözcük türlerini ayırt etmekte bazı zamanlarda biçimbilimsel çözümleyici yetersiz kalmaktadır. Bu durumda cümlenin söz dizimi çözümlemesi önem kazanmaktadır. Örneğin, Türkçe’ de aynı sözcük yerine göre sıfat ve ya zarf olabilmektedir. Bu durumun anlaşılabilmesi için çalışmada ATN çözümleyici kullanılmıştır.

Çalışmada kullanılan ATN çözümleyici depth-first arama algoritmasını kullanan yukarıdan aşağıya soldan sağa çalışan bir çözümleyicidir. Türkçe’nin karmaşık yapısını göz önüne alınarak tasarlanan ağlar kullanılmıştır.



Şekil 7.3 ATN çözümleyici için kullanılan cümle ağı

Şekil 7.3’ te ATN çözümleyicinin kurallı cümleler için kullandığı cümle ağı yer almaktadır. Özellikle son dönem Türk Edebiyat’ ında sıkça kullanılan devrik cümle yapısı için herhangi bir ağın geliştirilmesi imkansızdır. Bu tür cümlelerle, yüklemi sonda olmayan, ve eksiltili cümlelerle karşılaşıldığında tür belirlemesi konusundaki yanılgılar artabilmektedir.

## 7.5 Yazarın Tanınması İçin Kullanılan Sistem

Yazarın tanınması için, önceki bölümde anlatılan ve istatistiksel yaklaşım yöntemlerinden biri olan Bulanık Mantık yöntemi kullanılmıştır. Sistemin daha etkili çalışması için alfa

kesimi ekrandan girilmekte ve benzerlik deęerleri bu girilen alpha kesiminin altında olan veriler dikkate alınmamaktadır.



## 8. SONUÇ

Bu tez çalışmasında, Türkçe metinlerin Doğal Dil İşleme yöntemleri kullanılarak çözümlenmesi ve yazarının tanınmasına yönelik bir sistem geliştirilmiştir. Yazarın tanınması için kullanılan yöntem istatistiksel yaklaşım yöntemlerinden biri olan Bulanık Mantık Teorisidir. Geliştirilen sistem öğrenen bir sistemdir. Tür bilgisi sözlükte bulunmayan sözcüğe rastlandıkça sistem bu sözcükleri bir dosyada saklamakta, daha sonra da kullanıcı tarafından eksik bilgilerin tamamlanmasına olanak sağlamaktadır. Çözümlenen makale sayısı arttıkça sistemin mükemmelliyeti de artmaktadır.

Sistem, yazar tanıma işlemi açısından da öğrenen bir sistemdir. Çözümlenen metin için kullanıcı yazar adını girdikçe, sistem yazara ait istatistiğini güncellemektedir. Yazarı bilinmeyen bir metin geldiği zaman, sistem metnin çözümlenmesi sürecinden sonra elde ettiği verileri ve yazarlar için tutulan istatistikleri kullanarak bir sınıflandırma işlemi yapmaktadır. Burada, Bulanık Mantık teorisi devreye girmektedir.

Geliştirilen sistem hızlı çalışan ve doğruluk payı yüksek olan sonuçlar üreten bir sistemdir. Yazar veritabanında, yazarlara ait veriler çoğaldıkça yazarının tanınma olasılığı artmaktadır.

Çizelge 8.1' de değişik gazetelerin yedi köşe yazarının, değişik günlerde yayınlanan on makalesinin ortalaması olan cümle türüne göre çözümlenme sonuçları yer almaktadır.

Çizelge 8.1 Yedi yazar için cümle türüne göre metin çözümlenme sonuçları

<b>YAZAR</b>	<b>Cümledeki ortalama kelime sayısı</b>	<b>Devrik cümle sayısı / Metindeki cümle sayısı</b>	<b>Soru cümlesi sayısı / Metindeki cümle sayısı</b>	<b>Diğer cümle türlerinin sayısı / Metindeki cümle sayısı</b>
<b>A</b>	11.47	46.44	22.58	30.98
<b>B</b>	11.63	13.33	16.63	70.04
<b>C</b>	6.32	2.75	3.29	93.96
<b>D</b>	8.79	4.25	2.13	93.62
<b>E</b>	11.40	7.66	14.61	77.73
<b>F</b>	9.88	28.58	3.70	67.72
<b>G</b>	11.22	2.76	5.17	92.07

Çizelge 8.2' de yine aynı köşe yazarlarının aynı makalelerinin sözcük türüne göre çözümleme sonuçları yer almaktadır.

Çizelge 8.2 Yedi yazar için sözcük türüne göre metin çözümleme sonuçları

<b>YAZAR</b>	<b>Sıfat sayısı/ Metindeki sözcük sayısı</b>	<b>İsim sayısı / Metindeki Sözcük sayısı</b>	<b>Fiil sayısı / Metindeki sözcük sayısı</b>	<b>Metindeki diğer sözcük türlerinin sayısı / Metindeki Sözcük sayısı</b>
<b>A</b>	13.12	31.58	21.32	33.98
<b>B</b>	6.72	42.04	17.77	33.47
<b>C</b>	16.33	28.70	25.78	29.19
<b>D</b>	11.34	54.11	18.12	16.43
<b>E</b>	4.35	47.07	19.88	28.70
<b>F</b>	16.86	41.02	21.28	20.84
<b>G</b>	14.25	35.77	13.67	36.31

Yazarının bulunması için girilen X makalesinin çözümlenmesi sonucunda aşağıdaki veriler elde edilmiştir:

Makalenin bir cümlesindeki ortalama sözcük sayısı	:	11.21
Devrik cümle sayısının makaledeki cümle sayısına oranı	:	%33.37
Soru cümlesi sayısının makaledeki cümle sayısına oranı	:	%9.15
Diğer cümle türü sayısının makaledeki cümle sayısına oranı	:	%57.48
Makaledeki sıfat sayısının makaledeki sözcük sayısına oranı	:	%12.28
Makaledeki zarf sayısının makaledeki sözcük sayısına oranı	:	%31.33
Makaledeki fiil sayısının makaledeki sözcük sayısına oranı	:	%24.60
Diğer sözcük türü sayısının makaledeki sözcük sayısına oranı	:	%31.79

X metni çözümlendikten sonra elde edilen veriler ve (Çizelge 8.1) ile (Çizelge 8.2) ' de gösterilen yazarlara ait veriler kullanılarak üyelik dereceleri elde edilmektedir. Daha sonra bu üyelik derecelerinin ortalaması alınarak  $y^*$  elde edilmektedir.  $y^*$ , X metninin veritabanındaki yazarlara ait olma olasılığını göstermektedir.  $\mu_i$  ise, X metnine ait özelliklerin veritabanındaki yazara ait özelliklere göre üyelik derecesini göstermektedir. (Çizelge 8.3)

- $\mu_{11}$  : Cümledeki ortalama sözcük sayısının üyelik derecesi  
 $\mu_{12}$  : Devrik cümle oranının üyelik derecesi  
 $\mu_{13}$  : Soru cümlesi oranının üyelik derecesi  
 $\mu_{14}$  : Makaledeki sıfat sayısı oranının benzerlik derecesi  
 $\mu_{15}$  : Makaledeki zarf sayısı oranının benzerlik derecesi  
 $\mu_{16}$  : Makaledeki fiil sayısı oranının üyelik derecesi

Çizelge 8.3 Yazarı bilinmeyen metnin veritabanındaki diğer yazarlara göre üyelik derecesi

YAZAR	$\mu_{11}$	$\mu_{12}$	$\mu_{13}$	$\mu_{14}$	$\mu_{15}$	$\mu_{16}$	$y^*$
A	0.98	0.72	0.90	0.93	0.99	0.86	0.90
B	0.96	0.40	0.81	0.55	0.75	0.72	0.70
C	0.56	0.82	0.16	0.75	0.92	0.95	0.76
D	0.78	0.13	0.10	0.92	0.58	0.74	0.54
E	0.98	0.23	0.71	0.35	0.67	0.80	0.62
F	0.88	0.87	0.18	0.73	0.76	0.87	0.86
G	0.99	0.08	0.25	0.86	0.86	0.55	0.59

Girilen X makalesi A yazarına aittir ve netleştirme sonucunda elde edilen sonuçlardan da görülebileceği gibi (Çizelge 8.3) X makalesinin %90 A yazarına ait olduğu sonucunun alınması çalışmanın başarılı bir sonuç verdiğini göstermektedir.

**KAYNAKLAR**

Bauer, D., Segond, F. ve Zaenen, A.,1995,LocoLex: Translation rolls off your tongue, Proceedings of the Association for Computers and the Humanities-Association for Literary and Linguistic Computing (ACH-ALLC), Santa Barbara.

Breidt, E., Segond, F., Valetto, G.,1994, Multi-word lexemes and their automatic recognition in texts, Proceedings of Computational Lexicography (COMPLEX), Budapest.

Cutting, D., Kupiec Julian, Pedersen Jan, Sibun Penelope,1992, A practical part-of-speech tagger, Proceedings of Applied Natural Language Processing, Trento.

Grefenstette, G.ve Segond, F.,1997,Multilingual Natural Language Processing, International Journal of corpus linguistics.

Grefenstette, G., Schulze, M., Heid U., Fontenelle, T. ve Gérardy, C.,1996, The DECIDE project: Multilingual collocation extraction,Proceedings of the European Association for Lexicography ( EURALEX), Gothenburg.

Grefenstette, G.,1995,Low level parsing applied to technical manuals,Proceedings of Industrial Parsing of Software Manuals, Limerick.

Grefenstette, G. ve Tapanainen, P.,1994, What is a word, what is a sentence: problem of tokenization, Proceedings of Computational Lexicography (COMPLEX), Budapest.

Jurafsky D.,Martin H. J.,Speech and Language Processing, 2000, Prentice Hall, New Jersey.

Karlıgil E. M.,1998,G-Banlı Kromozomların Bulanık Mantık Sınıflandırma ve Şekil Tanıma Yöntemleri ile Tanınmasına Yönelik Bir Sistem,Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Doktora Tezi .İstanbul.

Karttunen, L. ve Yampol, T.,1993, Interactive finite state calculus, Technical report, ISTL-NLTT 93-04-01, PARC, Palo Alto.

Karttunen, L., Kaplan, R. ve Zaenen, A.,1992,Two level morphology with composition, Proceedings of Computational Linguistics (COLING), Nantes.

Krusee, K. G., 1991,Computer Processing of Natural Languages, Prentice Hall ,New Jersey.

Rich E., Knight K.,1991, Artificial Intelligence, McGraw-Hill International Edition, 1991.

Wang L., A Course in Fuzzy Systems and Control, Prentice Hall International Edition.

Şimşek R.,1987,Örneklerle Türkçe Sözdizimi,Kuzey Matbaacılık.

**ÖZGEÇMİŞ**

Doğum tarihi	22.10.1976	
Doğum yeri	K.K.T.C.	
Lise	1987-1993	Güzelyurt Türk Maarif Koleji
Lisans	1993-1997	Yıldız Teknik Üniversitesi Elektrik-Elektronik Fak. Bilgisayar Bilimleri Müh. Bölümü
Yüksek Lisans	1997-2000	Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Bilimleri Müh. Anabilim Dalı, Bilgisayar Bilimleri Müh. Programı
Çalıştığı kurumlar		
	1997-1998	NETAŞ Ar-Ge
	1998-Devam ediyor	AKNET Bilgi İşlem