

**T.C.  
YILDIZ TEKNİK ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ**

**TÜRKÇE DERLEMLER İÇİN SÖZDİZİMSEL GÖRSELLEŞTİRME VE  
SORGULAMA ARACI**

**CEM AGAN**

**YÜKSEK LİSANS TEZİ  
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI  
BİLGİSAYAR MÜHENDİSLİĞİ PROGRAMI**

**DANIŞMAN  
DOÇ. DR. BANU DİRİ**

**İSTANBUL, 2016**

**T.C.**  
**YILDIZ TEKNİK ÜNİVERSİTESİ**  
**FEN BİLİMLERİ ENSTİTÜSÜ**

**TÜRKÇE DERLEMLER İÇİN SÖZDİZİMSEL GÖRSELLEŞTİRME VE  
SORGULAMA ARACI**

Cem Ağan tarafından hazırlanan tez çalışması 14.04.2016 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı'nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

**Tez Danışmanı**

Doç. Dr. Banu DİRİ  
Yıldız Teknik Üniversitesi

**Jüri Üyeleri**

Doç. Dr. Banu DİRİ  
Yıldız Teknik Üniversitesi

Yrd. Doç. Dr. M. Fatih AMASYALI  
Yıldız Teknik Üniversitesi

Yrd. Doç. Dr. Tuğba YILDIZ  
Bilgi Üniversitesi

## ÖNSÖZ

---

Son yıllarda Türkçe Doğal Dil İşleme üzerine yapılan çalışmalar ve geliştirilen araçların sayısı artmaktadır. Bu durum dilimizin incelenip aydınlatılması ve elde edilen bulguların hayatımızı kolaylaştıracak uygulamalarda kullanılması açısından son derece umut vadedicidir. Bu tez çalışmasıyla amaçlanan, bu araçların arasına bir yenisini eklemektir.

Çalışmada her yönden desteğini esirgemeyen ve bana güvenen danışmanım Doç. Dr. Banu Diri'ye, çalışmam sırasında faydalandığım DDI araçlarının tasarlayıcıları sayın Prof. Dr. Kemal Oflazer ve Yrd. Doç. Dr. Gülşen Cebiroğlu Eryiğit'e yardımları için sonsuz teşekkürler ederim.

Şubat, 2016

Cem AGAN

## İÇİNDEKİLER

	Sayfa
SİMGE LİSTESİ .....	vi
KISALTMA LİSTESİ .....	vii
ŞEKİL LİSTESİ.....	viii
ÇİZELGE LİSTESİ .....	x
ÖZET.....	xi
ABSTRACT .....	xiii
<b>BÖLÜM 1</b>	
GİRİŞ.....	1
1.1    Literatür Özeti .....	1
1.1.1    Sözdizimsel Görselleştirme Araçları .....	2
1.1.2    Derlem Sorgulama Araçları.....	7
1.2    Tezin Amacı .....	10
1.3    Bulgular .....	11
<b>BÖLÜM 2</b>	
DERLEMLER HAKKINDA.....	12
2.1    Derlemler ve İşlevleri .....	12
2.2    Türkçe Derlem Çalışmaları .....	14
<b>BÖLÜM 3</b>	
GÖRSOR .....	17
3.1    Test Derlemlerinin Hazırlanıp Etiketlenmesi .....	17
3.2    İlişki Tablosu.....	20
3.3    İlişki Diyagramı .....	23
3.4    Graf Veritabanı ve Sorgulama Modülü .....	37

## BÖLÜM 4

SONUÇ VE ÖNERİLER .....	40
KAYNAKLAR.....	43
ÖZGEÇMİŞ.....	46



$n_i$	Sözdizim grafında bir düğüm
$n'$	Sözdizim grafında yeni oluşan bir düğüm
TMi	Tamlayan (isim)
TMs	Tamlayan (sıfat)
TNi	Tamlanan (isim)



## KISALTMA LİSTESİ

---

BOUN	Boğaziçi University
CoNLL	Conference on Natural Language Learning
EBA	En Büyük Ata
MWE	Multiword Expression
NLP	Natural Language Processing
OCR	Optical Character Recognition
ODTÜ	Orta Doğu Teknik Üniversitesi
PHP	Personal Home Page
PML	Prague Markup Language
regex	Regular Expression
TUD	Türkçe Ulusal Derlemi
XCES	Corpus Encoding Standard for XML
XML	Extensible Markup Language

## ŞEKİL LİSTESİ

	Sayfa
Şekil 1. 1	Bir cümle için ayrıştırma ağacı ve bağımlılık diyagramı [1]..... 2
Şekil 1. 2	Örnek cümle için Tree View diagramı ..... 4
Şekil 1. 3	Bir Dependence bağımlılık diyagramı ..... 4
Şekil 1. 4	Bir MaltDiver bağımlılık diyagramı..... 4
Şekil 1. 5	Örnek cümlelerin ayrıştırma ağacı için ParseViz'in ürettiği diyagram ..... 5
Şekil 1. 6	Treebank Viewer aracının ekran görüntüsü..... 5
Şekil 1. 7	Bir ayrıştırma ağacı için NlpViz'in ürettiği diyagram ..... 6
Şekil 1. 8	Bir ayrıştırma ağacı için phpSyntaxTree aracının ürettiği diyagram ..... 6
Şekil 1. 9	SearchTree'nin sorgulama arayüzü..... 7
Şekil 1. 10	Bir önceki şekildeki sorgu için Search Tree'nin cevabı ..... 8
Şekil 1. 11	Örnek bir sorgu ve Fangorn'un cevabı..... 8
Şekil 1. 12	TigerSearch aracının ekran görüntüsü..... 9
Şekil 1. 13	PLM Tree-Query aracının ekran görüntüsü ..... 10
Şekil 2. 1	Pen Treebank'ta kullanılan sözcük türü etiketleri [21] ..... 13
Şekil 2. 2	Bir ağaç yapılı derlemeden örnek bir cümle ..... 13
Şekil 2. 3	ODTÜ-Sabancı Ağaç Yapılı Derleminden örnek bir cümle ..... 14
Şekil 2. 4	Türkçe Ulusal Derleminin web sorgulama arayüzü ..... 15
Şekil 2. 5	TS Corpus'un web sorgulama arayüzü ..... 16
Şekil 3. 1	Bir cümlelerin NLP Pipeline ile analizi ..... 19
Şekil 3. 2	NLP Pipeline için örnek bir giriş dosyası..... 20
Şekil 3. 3	Örnek bir cümle için otomatik olarak üretilen ilişki tablosu..... 21
Şekil 3. 4	İlişki Katmanları..... 22
Şekil 3. 5	Örnek bir cümle için ilişki diyagramı ..... 23
Şekil 3. 6	Örnek grafin oluşturulması Adım 1 ..... 26
Şekil 3. 7	Örnek grafin oluşturulması Adım 2 ..... 26
Şekil 3. 8	Örnek grafin oluşturulması Adım 3 ..... 27
Şekil 3. 9	Örnek grafin oluşturulması Adım 4 ..... 27
Şekil 3. 10	Örnek grafin oluşturulması Adım 5 ..... 28
Şekil 3. 11	Örnek grafin oluşturulması Adım 6 ..... 28
Şekil 3. 12	Örnek grafin oluşturulması Adım 7 ..... 29
Şekil 3. 13	Örnek grafin oluşturulması Adım 8 ..... 29
Şekil 3. 14	Örnek grafin oluşturulması Adım 9 ..... 30
Şekil 3. 15	Örnek grafin oluşturulması Adım 10..... 30



Şekil 3. 16	Örnek grafin oluşturulması Adım 11.....	31
Şekil 3. 17	Derlem 1'e ait hata analizi grafiği .....	32
Şekil 3. 18	Derlem 2'ye ait hata analizi grafiği .....	32
Şekil 3. 19	Bağlama Hatası Örneği.....	33
Şekil 3. 20	İsim-fiil Eki Hatası Örneği .....	34
Şekil 3. 21	Kopulasız Fiil Hatası Örneği.....	34
Şekil 3. 22	CSD Hatası Örneği .....	35
Şekil 3. 23	Biçimbirimsel Bulanıklık Giderme Hatası Örneği .....	35
Şekil 3. 24	Sözdizimsel Analiz Hatası Örneği .....	35
Şekil 3. 25	Türemelerin olduğu yerlerde ilişkilerin gösterilmesi .....	36
Şekil 3. 26	Sözdizimsel Sorgulama Arayüzü .....	38



## ÇİZELGE LİSTESİ

---

	Sayfa	
Çizelge 1. 1	Örnek cümle için bağımlılık ayrıştırması	3
Çizelge 3. 1	Test Derlemlerinin Özellikleri	17
Çizelge 3. 2	İlişki Tablosundaki Kısaltmaların Açıklamaları	21
Çizelge 3. 3	Diyagramdaki Kısaltmaların Açıklamaları	24
Çizelge 3. 4	Öbekleme işlemi için graf oluşturma kuralları	25
Çizelge 3. 5	Öbek bağlama işlemi için graf oluşturma kuralları	25
Çizelge 3. 6	Öbek birleştirme işlemi için graf oluşturma kuralları	25
Çizelge 3. 7	Türetme işlemi için graf oluşturma kuralları	25
Çizelge 3. 8	Cümle bağlama işlemi için graf oluşturma kuralları	26
Çizelge 3. 9	Graf Oluşturma Kurallarının Başarımı	31
Çizelge 3. 10	Hata Kısaltmalarının Açıklamaları	33
Çizelge 3. 11	Sözdizimsel sorgulama örnekleri	38
Çizelge 3. 12	Karmaşık sorguya dönen örnek cevaplar	39

## TÜRKÇE DERLEMLER İÇİN SÖZDİZİMSEL GÖRSELLEŞTİRME VE SORGULAMA ARACI

Cem AGAN

Bilgisayar Mühendisliği Anabilim Dalı

Yüksek Lisans Tezi

Tez Danışmanı: Doç. Dr. Banu DİRİ

Derlemler bir dilin sözdizimsel ve anlambilimsel özelliklerinin keşfedilmesi ve elde edilen bulgulardan birçok araştırma alanında faydalanılması açısından önemli bilgi kaynaklarıdır. Derlemlere elle veya çözümleyici araçlar tarafından eklenen dilbilimsel açıklamalar (biçimbirimsel, sözdizimsel ve anlamsal) bu derlemlerdeki birimler ve bunlar arasındaki ilişkileri ortaya koyar.

Derlemler üzerinde araştırma yapılırken bu ilişkilerin takip edilmesini kolaylaştırmak için görselleştirme araçlarına başvurulur. Bunlar iç içe geçmiş olan karmaşık ilişkileri diyagramlar şeklinde özetler. Derlem araştırmalarında faydalanılan bir diğer araç türü ise sorgulama araçlarıdır. Derlem içersinden istenilen sözdizimsel koşulları sağlayan cümle yapılarının çıkartılmasında bu araçlardan faydalanılır.

İngilizce ve diğer bazı Batı dillerindeki derlemler için geliştirilmiş bu tür araçlar bulunmasına karşın, Türkçe derlemler için henüz böyle araçlara rastlanmamaktadır. Bu çalışmada Türkçe derlemler için sözdizimsel görselleştirme ve sorgulama imkanı sağlayan bir aracın geliştirilmesi üzerine çalışılmıştır. Bunun için mevcut çözümleyici araçlar kullanılarak bir test derlemi oluşturulmuş ve bu derlemdeki cümleler için sözdizimsel ilişkileri ve cümle yapılarını gösteren diyagramlar üreten bir araç geliştirilmiştir. Daha sonra araç yardımıyla derlemde elde edilen tüm yapılar bir graf veritabanına aktarılmıştır. Araca eklenen bir sorgulama modülü sayesinde isim ve fiil

kökünden anahtar sözcükleri içeren öntanımlı cümle yapılarının sorgulanması sağlanmıştır. Ayrıca, graf veritabanının web arayüzü ile daha karmaşık ve detaylı sorguların nasıl yapılabileceği araştırılıp derlem üzerinde örnek sorgular gerçekleştirilmiştir.

**Anahtar Kelimeler:** Türkçe, Derlem, Doğal Dil İşleme, Görselleştirme, Sorgulama



**SYNTACTIC VISUALISATION AND QUERY TOOL FOR TURKISH CORPORA**

Cem AGAN

Department of Computer Engineering

MSc. Thesis

Adviser: Assoc.Prof. Banu DİRİ

A corpus is an important source in terms of discovering syntactic and semantic features of a language and making use of these findings in various research areas. Annotations (morphologic, syntactic or semantic) which are added to a corpus manually or by analyzing tools show the relations between units in the corpus.

When making research on a corpus, visualization tools are used to ease the following of these relations. These tools summarize the complex relations as diagrams. Another tools used in corpus research are query tools. These are made use of in extraction of sentence structures laying down desired syntactic conditions.

Although there are such tools for corpora in English and some other Western Languages, it can not be run across such tools for Turkish corpora. In this work, we worked on developing a tool enabling syntactic visualization and querying for Turkish corpora. To achieve this, we first built a test corpus by the help of current analyzer tools and developed a tool that generates diagrams showing syntactic relations and sentence structures. After, the structures gotten out of the corpora were transferred to a graph database. By the help of a query module added to the tool it's enabled to query predefined sentence structures including noun or verb keywords. Besides we researched on how it can be queried more complex and detailed queries through the

web user interface of the graph database and gave some example queries on the corpus.

**Keywords:** Turkish, Corpus, Natural Language Processing, Visualization, Query



### GİRİŞ

Derlem bir dile ait metin parçalarının elektronik formdaki koleksiyonudur. Derlemler hesaplamalı dilbilim, makine çevirisi, doğal dil işleme vb. araştırma alanlarında kullanılırlar.

Bilgisayarla görselleştirme uygulamaları bir çok alanda olduğu gibi derlemlerdeki cümlelerin sözdizim ağaçlarının çıkartılmasında ve bunların sözcük türleriyle birlikte görüntülenmesinde de kullanılmaktadır.

Derlemler üzerinde belli sözdizimsel kalıplara uyan yapıları sorgulamak da mümkündür. Bu işlem derlem sorgulama araçları ile gerçekleştirilir.

Bu çalışmada Türkçe derlemlerdeki cümlelerin sözdizim ağaçlarını diyagram şeklinde üreten ve derlemler üzerinde sözdizimsel sorgulama yapabilen bir araç gerçekleştirilmiştir.

Birinci bölümde mevcut derlem görselleştirme ve sorgulama araçları tanıtılmış, çalışmanın amacı ve elde edilen bulgulardan bahsedilmiştir. İkinci bölümde derlemlerin işlevleri ve Türkçe derlem çalışmaları üzerinde durulmuştur. Üçüncü bölümde aracın geliştirme aşamaları anlatılmıştır. Dördüncü bölüm de ise sonuç ve öneriler ele alınmıştır.

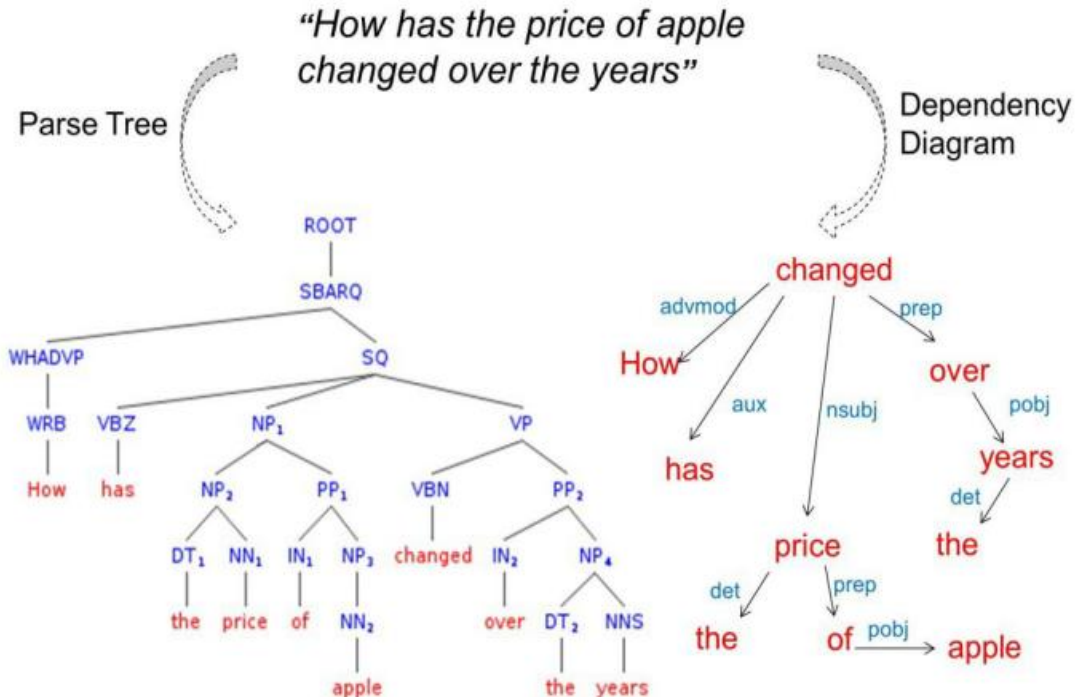
#### 1.1 Literatür Özeti

Bu çalışmada literatür özeti Sözdizimsel Görselleştirme Araçları ve Derlem Sorgulama Araçları olmak üzere iki başlık altında verilmiştir.

### 1.1.1 Sözdizimsel Görselleştirme Araçları

Doğal dildeki metinlerin dilbilimsel analizlerinin daha iyi takip edilebilmesi amacıyla görselleştirme araçları geliştirilmektedir. Bunun için bir çözümleyiciye (biçimbirimsel, sözdizimsel veya anlambilimsel) ve onun analizlerini yorumlayarak bir diyagrama dökülecek olan görselleştirme programına ihtiyaç duyulmaktadır. Bu tür görselleştirme uygulamalarının en bilinenlerine örnek olarak bağımlılık diyagramı (dependency diagram) ve ayrıştırma ağacı (parse tree) verilebilir.

Doğal dil işlemede söz konusu dil İngilizce olduğunda birden fazla sözdizimsel çözümleyiciye ve görselleştirme programlarına rastlamak mümkündür. Çok bilinen çözümleyiciler genellikle üniversitelerin araştırma gruplarınınca geliştirilmiştir. Bunlara örnek olarak Stanford Parser, Berkeley Parser, TurboParser (Carnegie Mellon) verilebilir. Bu çözümleyicilerin ürettikleri ayrıştırma ağaçları ve gerçekleştirdiği bağımlılık ayrıştırmaları çeşitli yazılımlar veya yazılım kütüphaneleri ile görselleştirilebilmektedir (Şekil 1.1).



Şekil 1. 1 Bir cümle için ayrıştırma ağacı ve bağımlılık diyagramı [1]

Bir cümlenin bağımlılık ayrıştırması, o cümledeki sözcükler arasındaki dilbilgisel ilişkileri gösterir. Bağımlılık dilbilgisi (dependency grammar), bağımlılık ilişkisini (dependency



relation) baz alan modern bir sözdizimsel teoriler sınıfıdır. Bağımlılık ilişkisi, fiili cümle/yan cümle yapısının merkezi olarak görür. Diğer tüm sözdizimsel birimler ise fiile doğrudan veya dolaylı olarak bağımlı kabul edilir. Bağımlılık dilbilgileri, öbek yapısı dilbilgilerinden (phrase structure grammar) farklıdır; çünkü öbekselsel düğümleri içermezler, yine de öbekler hakkında bilgi verirler. Baş sözcük (head word) ve onun bağımlılıkları arasındaki ilişki ile yapı belirlenir. İlişkisel yapılar ögesel yapılara göre daha düzdür, çünkü sonlu fiil öbeği ögesine sahip değildirler. Bu nedenle özgür sözcük düzenine sahip Türkçe, Çekçe vb. dillerin analizleri için daha uygundur [2].

Bağımlılık ayrıştırmalarını görselleştiren programlara verilebilecek örneklerden biri Carnegie Mellon Noah's Ark [3] grubunun Tree View'idir[4]. Aynı grubun Turbo Parser[5] isimli çözümleyicisi ile bir cümle için üretilen analiz, Tree View yardımıyla diagrama aktarılabilir. Çizelge 1.1'de "The cat in the hat saw him with a telescope." cümlesinin bağımlılık ayrıştırması görülmektedir.

Çizelge 1. 1 Örnek cümle için bağımlılık ayrıştırması

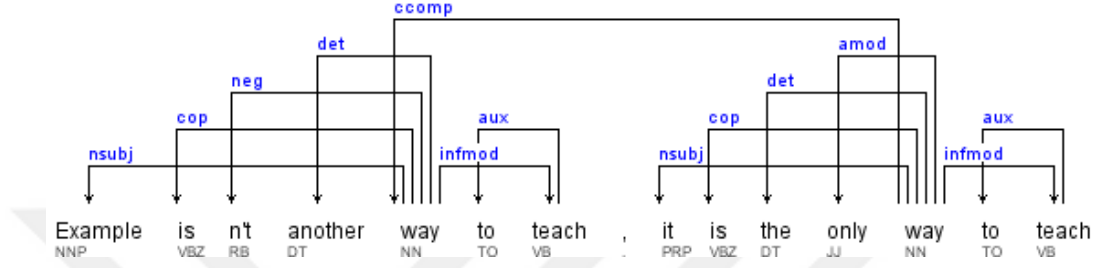
1	The	_	DT	DT	_	2	det	_	_
2	cat	_	NN	NN	_	6	nsubj	_	_
3	in	_	IN	IN	_	2	prep	_	_
4	the	_	DT	DT	_	5	det	_	_
5	hat	_	NN	NN	_	3	pobj	_	_
6	saw	_	VB	VBD	_	0	null	_	_
7	him	_	PRP	PRP	_	6	dobj	_	_
8	with	_	IN	IN	_	6	prep	_	_
9	a	_	DT	DT	_	10	det	_	_
10	telescope	_	NN	NN	_	8	pobj	_	_
11	.	_	.	.	_	6	punct	_	_

Bu ayrıştırmaya ait Tree View diyagramı ise Şekil 1.2'deki gibidir.



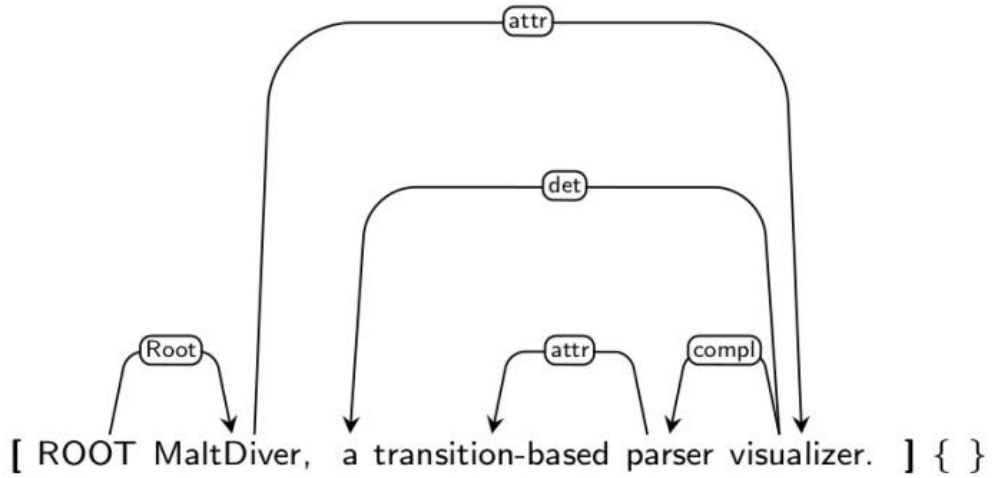
Şekil 1. 2 Örnek cümle için Tree View diagramı

Standford Parser'ın bağımlılık ayrıştırmalarını görselleştiren programlar da mevcuttur. Bunlara örnek olarak tydevi [6] ve DependenceSee [7] verilebilir (Şekil 1.3).



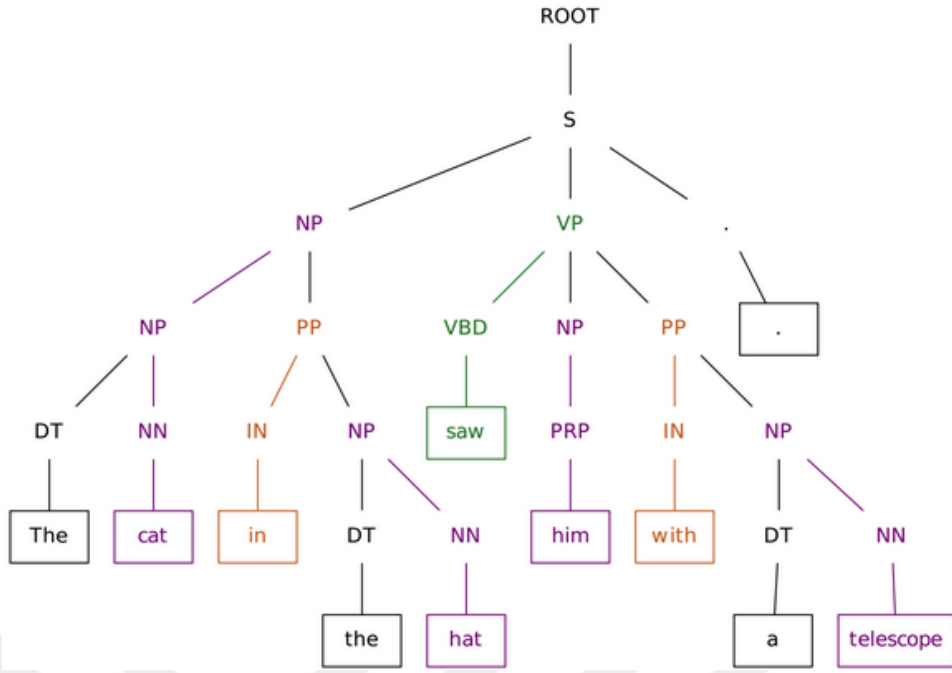
Şekil 1. 3 Bir DependenceSee bağımlılık diyagramı

Geçiş tabanlı bir çözümleyici olan Malt Parser'ın [8] analizleri de MaltDiver [9] aracı ile görselleştirilir (Şekil 1.4).



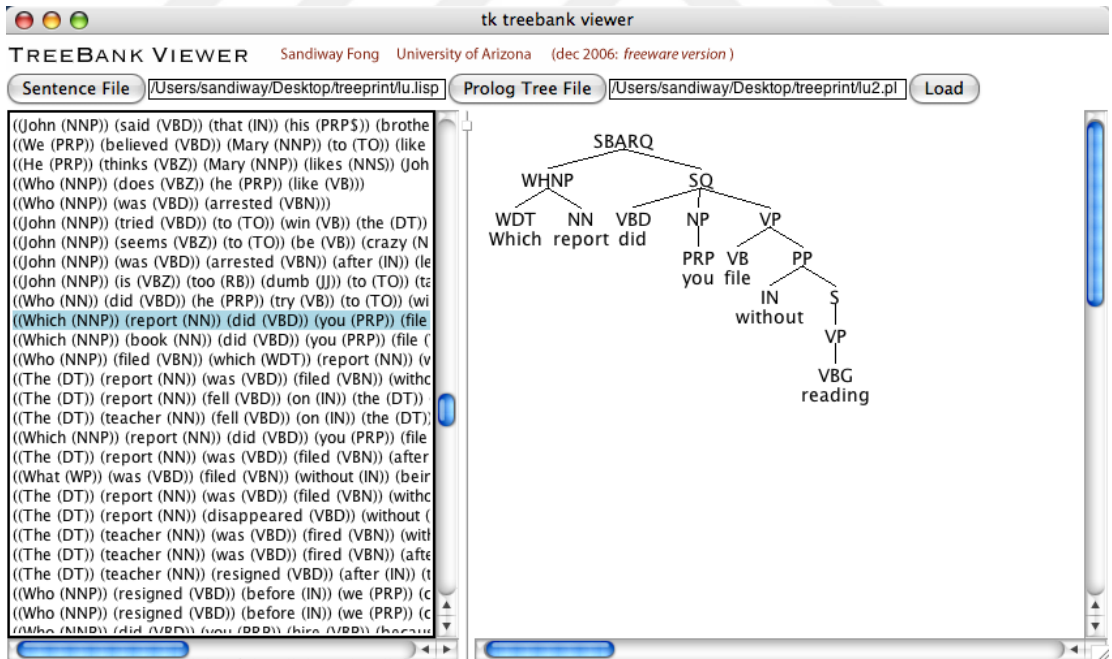
Şekil 1. 4 Bir MaltDiver bağımlılık diyagramı

Birçok çözümleyici için ayrıştırma ağaçlarının diyagramlarını üreten programlar da mevcuttur. Örneğin, Noah's Ark grubunun geliştirdiği ParseViz [10] adlı araç, ayrıştırma ağaçlarını görselleştirir (Şekil 1.5).



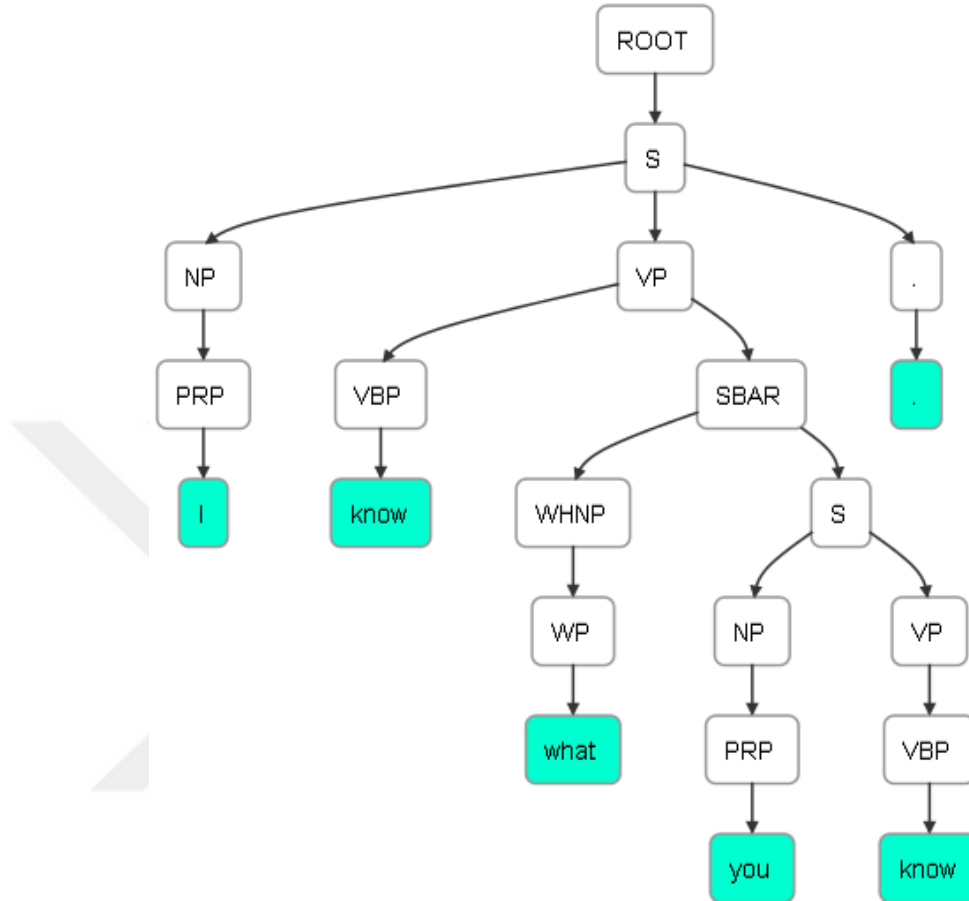
Şekil 1. 5 Örnek cümlenin ayrıştırma ağacı için ParseViz'in ürettiği diyagram

Treebank Viewer [11] programı ise, sözdizimsel olarak etiketlenmiş derlemlerdeki cümlelerin ayrıştırma ağaçlarını diyagram şeklinde görüntülemeye yarar (Şekil 1.6).



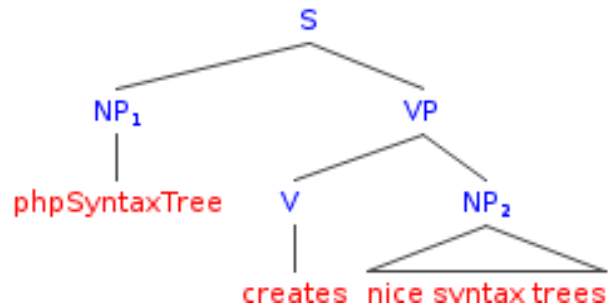
Şekil 1. 6 Treebank Viewer aracının ekran görüntüsü

NlpViz [12] ise Stanford'un CoreNLP [13] kütüphanesini kullanarak ayrıştırdığı cümlelerin analizlerini dagre [14] ve d3 [15] görüntüleme kütüphaneleriyle diyagrama dönüştürür (Şekil 1.7).



Şekil 1. 7 Bir ayrıştırma ağacı için NlpViz'in ürettiği diyagram

Ayrıştırma ağacı görselleştirme araçlarına başka bir örnek olarak PHP ile kullanılabilen phpSyntaxTree [16] de verilebilir (Şekil 1.8).



Şekil 1. 8 Bir ayrıştırma ağacı için phpSyntaxTree aracının ürettiği diyagram

### 1.1.2 Derlem Sorgulama Araçları

Ağaç yapılı derlemler dilbilimciler için oldukça değerli kaynaklardır. Bunlar üzerinde yapılacak karmaşık sorgular ile bir dilin sözdizimsel ve anlamsal özellikleri aydınlatılabilmektedir.

Liteatürde ağaç yapılı derlemlerin sorgulanması üzerine çeşitli çalışmalar bulunmaktadır. Bunlardan biri, herhangi bir sorgulama dili veya regex bilgisi gerektirmeyen Search Tree'dir. Search Tree basit ve kullanıcı dostu arayüzüyle cümle yapıları içerisinde sorgulama yapmaya imkan tanır (Şekil 1.9). Sistem herkese açıktır ve arayüze web tarayıcı üzerinden erişilir. Ağaç yapılı derlemlerdeki tüm etiket ve kategorilerin seçilmesini sağlayan seçeneklere sahiptir ve sorgu sonucu, sorgu alt ağacı ve bu ağacın ait olduğu cümleyle birlikte döndürülür (Şekil 1.10). [17]

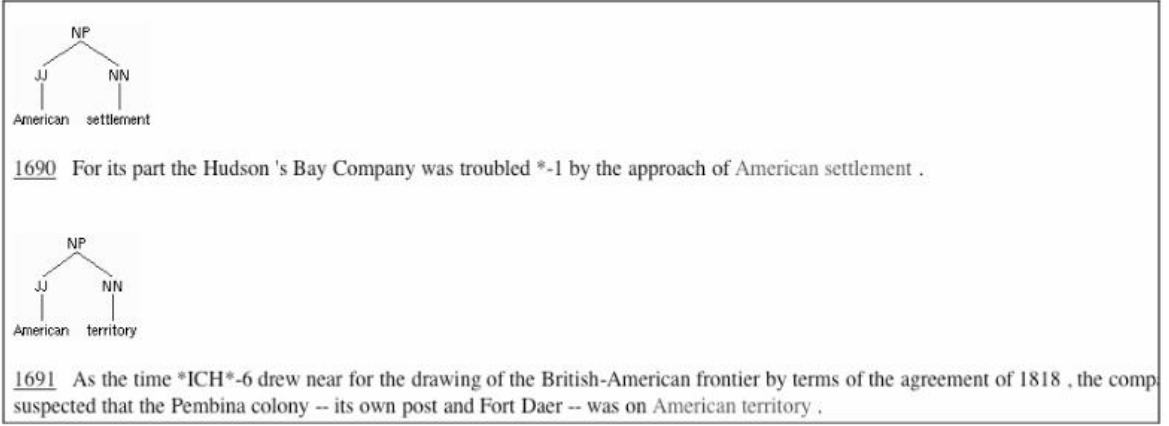
The screenshot displays the SearchTree search interface. It features several input fields and radio buttons for search criteria:

- Terminal:** A text input field containing the word "American".
- Node:** A dropdown menu, currently empty.
- Label:** A dropdown menu containing the tag "JJ".
- Relation to mother node:** A dropdown menu containing the relation "Daughter".
- Two radio buttons are present: "Start of word" and "End of word", both of which are currently unselected.

Below the search parameters, a partial parse tree is shown:

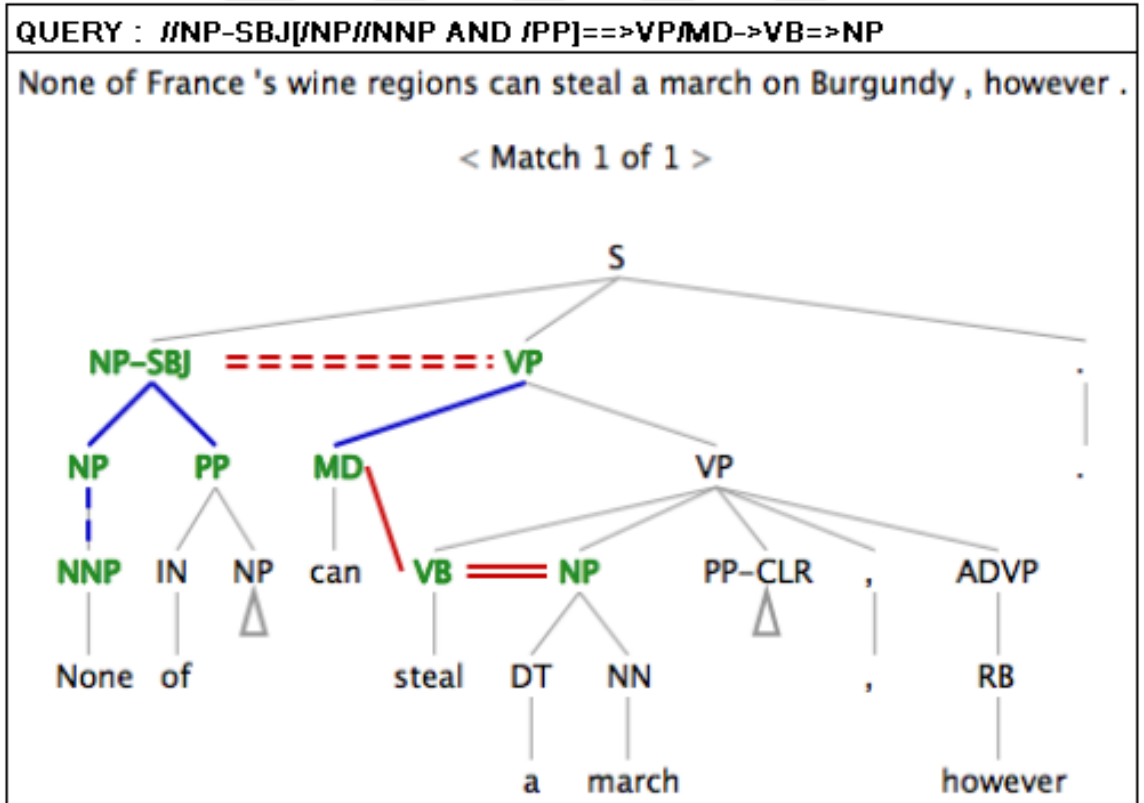
```
-NP
|
|-JJ / "American"
|
|-NN
```

Şekil 1. 9 SearchTree'nin sorgulama arayüzü



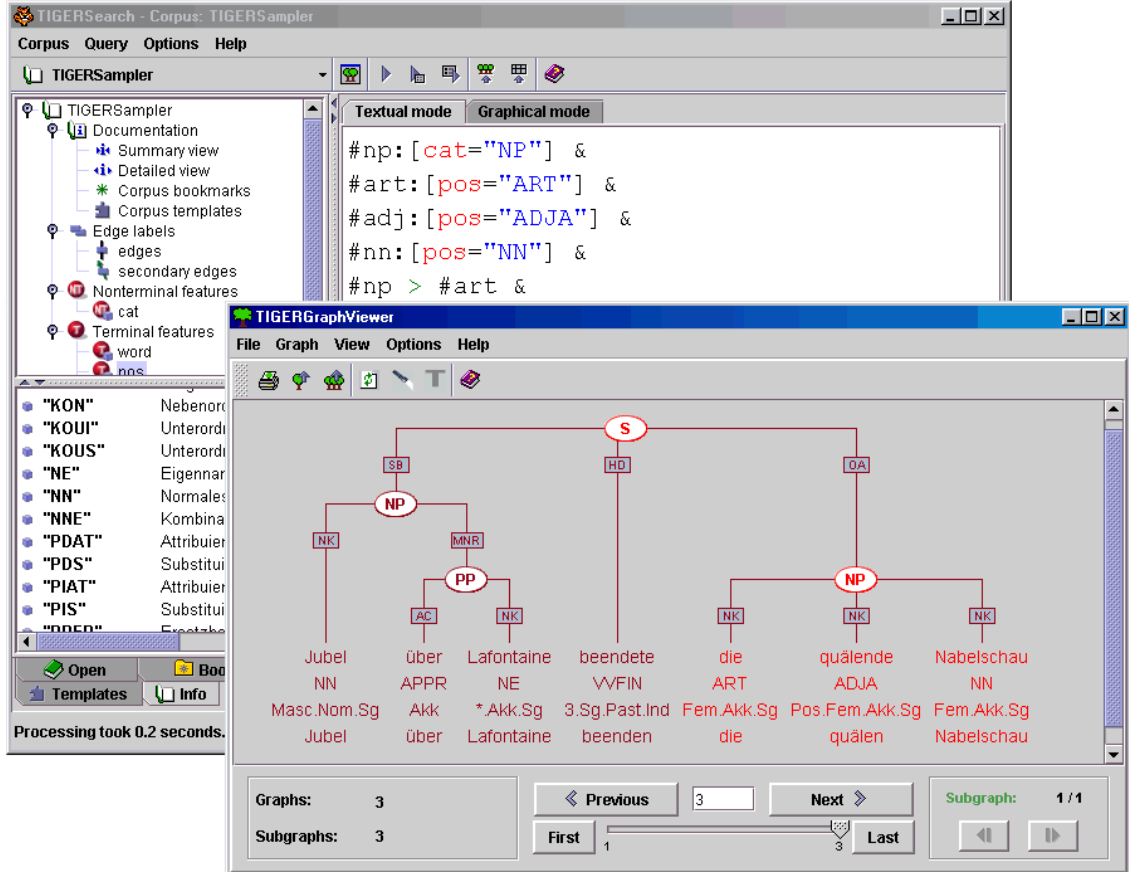
Şekil 1. 10 Bir önceki şekildeki sorgu için Search Tree'nin cevabı

Fangorn ise, özellikle büyük derlemlerin hızlı bir şekilde sorgulanması için tasarlanmıştır. Disiplinlerarası bir araştırma projesinin ürünüdür. Bu proje genel amaçlı yarı yapılandırılmış veritabanlarının ve bilgi erişimi motorlarının özelliklerini bir araya getirmektedir. Sorgular "yol ifadeleri" içerir. Bu ifadelerle bir düğümün atalarına, çocuklarına ve torunlarına doğru hareket ederek arama yapılabilir (Şekil 1.11)[18].



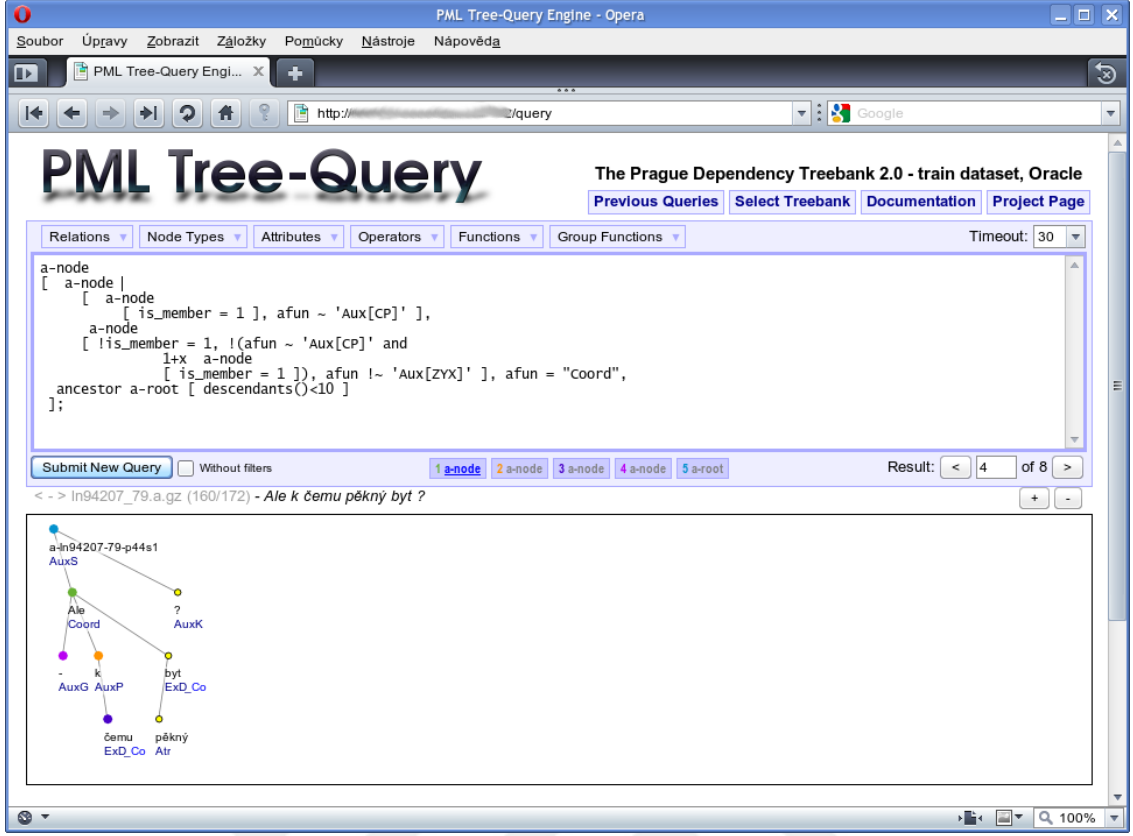
Şekil 1. 11 Örnek bir sorgu ve Fangorn'un cevabı

TIGERSearch [19] graf yapılı veritabanlarından (ağaç yapılı derlemeler) bilgi erişimi üzerine özelleşmiş bir programdır. Deneyimi olmayan kullanıcılar arayüz üzerinden kısmi graflar çizerek sorguları formülize edebilmektedirler. İleri seviye kullanıcılar ise programın sorgu dilini kullanarak sorgulama yapabilirler. Sorgu sonuçları grafiksel olarak görüntülenir (Şekil 1.12).



Şekil 1. 12 TigerSearch aracının ekran görüntüsü

PML-TQ [20] güçlü bir açık kaynaklı sorgu aracıdır ve her tür (karmaşık, bağımlılık, ögesel, çok katmanlı ve paralel) dilbilimsel olarak etiketlenmiş derlemin sorgulanmasında kullanılabilir. Derlem PML veri biçiminde kodlanmış olmalıdır, fakat başka formatlar için dönüştürme betikleri de mevcuttur (Şekil 1.13).



Şekil 1. 13 PLM Tree-Query aracının ekran görüntüsü

## 1.2 Tezin Amacı

Sözdizimsel çözümleme ile etiketlenmiş cümlelerin sözcükleri arasındaki sözdizimsel ilişkilerin ve bunların oluşturduğu yapıların görselleştirilmesi, bu tür cümlelerden oluşan derlemler üzerinde araştırma yapanlar için büyük kolaylık sağlamaktadır. Ayrıca, derlemlerde cümlenin her seviyesinden sorgulama yapılabilmesi dilin sözdizimsel ve anlambilimsel özelliklerinin keşfedilmesini sağlamaktadır.

Literatürde İngilizce ve diğer batı dillerinde sözdizimsel olarak etiketlenmiş derlemler için bir dizi görselleştirme ve sorgulama araçlarına rastlanmaktadır. Söz konusu Türkçe olduğunda bu tür araçları bulmak zorlaşmaktadır.

Bu çalışmanın amacı, Türkçe için geliştirilmiş mevcut sözdizimsel çözümleyici araçlarla etiketlenerek oluşturulan derlemlerdeki cümlelerin sözdizimsel ilişkilerinin görselleştirilmesini ve bu derlemler üzerinde karmaşık sorgular yapılmasını sağlayan bir aracın geliştirilmesidir.



### 1.3 Bulgular

Mevcut Türkçe sözdizimsel analiz araçlarıyla ilköğretim hayat bilgisi kitaplarından oluşan bir test derlemi analiz edilmiş ve %97.8 oranında isabetli analizler elde edildiği gözlemlenmiştir. Analizlerin doğrulukları sınanırken çözümleyici tarafından verilen etiketlerin program tarafından üretilecek tablolar üzerinden takip edilmesinin daha kolay olacağı düşünülerek bir ilişki tablosu modülü geliştirilmiştir.

Hatalı sözdizimsel analizlerin olduğu cümlelere bakıldığında bunların bir kısmına biçimbirimsel bulanıklığın bir kısmına da sözdizimsel bulanıklığın giderilememesinin yol açtığı görülmüştür. Hatalı analiz edilmiş cümlelerin bazılarında uygun yerlere virgül koyulup analiz tekrarlandığında düzelme olduğu görülmüştür. Hatalı analize sahip bazı karmaşık cümleler parçalanarak basitleştirildiğinde de yine doğru analizler elde edilebilmiştir.

Sözdizimsel etiketlerden cümleleri oluşturan yapılara ulaşabilmek için çözümleyici araç tarafından kullanılan etiketler ve bunların ne tür yapılar (öbek, yancümle veya cümle) oluşturduğu belirlenmiştir. Bir cümlenin yapısı çıkartılırken hangi adımların uygulanması ve yapıyı temsil edecek grafın hangi kurallara göre oluşturulacağı tespit edilmiştir. Geliştirilen bir program modülüyle, elde edilen graf yapılarının diyagram şeklinde görselleştirilmesi sağlanmıştır.

Cümle yapıları çıkartıldıktan sonra Neo4j graf veritabanına aktarılmıştır. Uygulama arayüzüne sorgulama kalıpları eklenerek isim ve fiil kökleri ile sorgulama yapılabilmesi sağlanmıştır. Ayrıca, veritabanının web arayüzündeki sorgulama kısmından çeşitli kompleks sorgular yapılarak sözcükler arasındaki dolaylı ilişkilerin sorgulanabileceği görülmüştür.

### DERLEMLER HAKKINDA

Bu bölümde derlem, derlem türleri ve derlemlerin işlevleri hakkında bilgiler verilmiştir. Ayrıca, bu zamana kadar yapılmış Türkçe derlem çalışmalarından bahsedilmiştir.

#### 2.1 Derlemler ve İşlevleri

Dilbilimde, büyük miktarda metin içeren yapılandırılmış doküman kümelerine derlem denir. Günümüzde çoğunlukla elektronik ortamda saklanır ve işlenirler. Derlemler dile ait istatistiksel çözümlenmeler ve hipotez testleri (bir istatistiksel çıkarım metodudur) yapmada kullanılırlar.

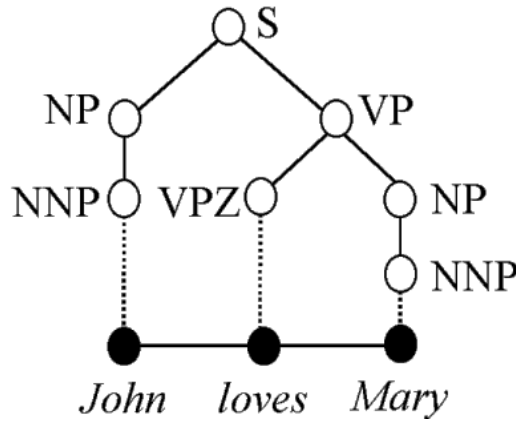
Derlemler, dilbilimsel araştırmalara elverişli hale getirilmek için etiketleme (annotation) denilen bir işleme tabi tutulurlar. Bu işlemlerle derlemdeki metinsel bilgiye dilbilimsel bazı açıklamalar eklenir. Etiketlemeye örnek olarak sözcük türü etiketleme gösterilebilir. Sözcük türü etiketlemede her bir sözcüğün türü (fiil, isim, sıfat, vs.) etiketler şeklinde derleme eklenir. Başka bir örnek olarak da sözcük köklerinin belirtilmesi verilebilir (Şekil 2.1).

**Table 2**  
The Penn Treebank POS tagset.

1. CC	Coordinating conjunction	25. TO	to
2. CD	Cardinal number	26. UH	Interjection
3. DT	Determiner	27. VB	Verb, base form
4. EX	Existential <i>there</i>	28. VBD	Verb, past tense
5. FW	Foreign word	29. VBG	Verb, gerund/present participle
6. IN	Preposition/subordinating conjunction	30. VBN	Verb, past participle
7. JJ	Adjective	31. VBP	Verb, non-3rd ps. sing. present
8. JJR	Adjective, comparative	32. VBZ	Verb, 3rd ps. sing. present
9. JJS	Adjective, superlative	33. WDT	<i>wh</i> -determiner
10. LS	List item marker	34. WP	<i>wh</i> -pronoun
11. MD	Modal	35. WP\$	Possessive <i>wh</i> -pronoun
12. NN	Noun, singular or mass	36. WRB	<i>wh</i> -adverb
13. NNS	Noun, plural	37. #	Pound sign
14. NNP	Proper noun, singular	38. \$	Dollar sign
15. NNPS	Proper noun, plural	39. .	Sentence-final punctuation
16. PDT	Predeterminer	40. ,	Comma
17. POS	Possessive ending	41. :	Colon, semi-colon
18. PRP	Personal pronoun	42. (	Left bracket character
19. PP\$	Possessive pronoun	43. )	Right bracket character
20. RB	Adverb	44. "	Straight double quote
21. RBR	Adverb, comparative	45. '	Left open single quote
22. RBS	Adverb, superlative	46. "	Left open double quote
23. RP	Particle	47. '	Right close single quote
24. SYM	Symbol (mathematical or scientific)	48. "	Right close double quote

Şekil 2. 14 Pen Treebank'ta kullanılan sözcük türü etiketleri [21]

Bazı derlemeler tam kapsamlı dilbilimsel incelemeden geçirilmiş olabilir. Bunlar ağaç yapılı derlem (Şekil 2.2) olarak da adlandırılır ve sözdizimsel veya anlambilimsel açıdan cümle yapıları etiketlenmiştir. Böyle derlemeler için bütün bir derlemin tamamen ve tutarlı bir şekilde etiketlendiğinin kontrol edilmesi zor olduğundan genellikle küçük boyutlu olmaları tercih edilir.



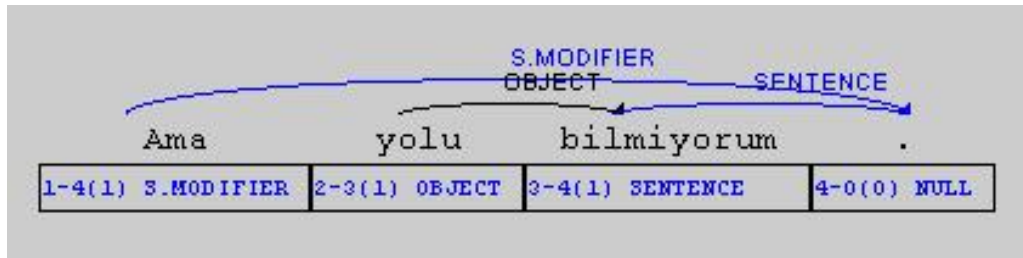
Şekil 2. 15 Bir ağaç yapılı derlemden örnek bir cümle

Derlemler, derlem dilbilimindeki temel bilgi tabanlarıdır. Farklı türlerdeki derlemlerin incelenmesi ve işlenmesi hesaplamalı dilbilim, konuşma tanıma ve makine çevirisi alanlarındaki bir çok çalışmanın konusudur. Genellikle Saklı Markov modellerinin oluşturulmasında kullanılırlar. Bu modeller sözcük türü etiketleme ve diğer amaçlar için kullanılır. Bunlardan türetilen yeni derlem ve frekans listeleri dil öğretiminde de kullanılır. Derlem aynı zamanda yabancı dilde yazı yazmaya yardımcı bir unsur olarak da düşünülebilir[22].

## 2.2 Türkçe Derlem Çalışmaları

Türkçe üzerine yapılmış derlem çalışmalarının ilklerinden biri ODTÜ Türkçe Derlemidir [23]. 1990 sonrasına ait yazılı Türkçe örneklerden oluşmuştur ve 2 milyon sözcüğe sahiptir. Dokümanlar 10 değişik türden alınmıştır. Derlem XCES [24] etiketleriyle işaretlenmiştir. Araştırmacılar kullanıcı anlaşması formunu doldurmak şartıyla derlemde faydalanabilmektedir [25].

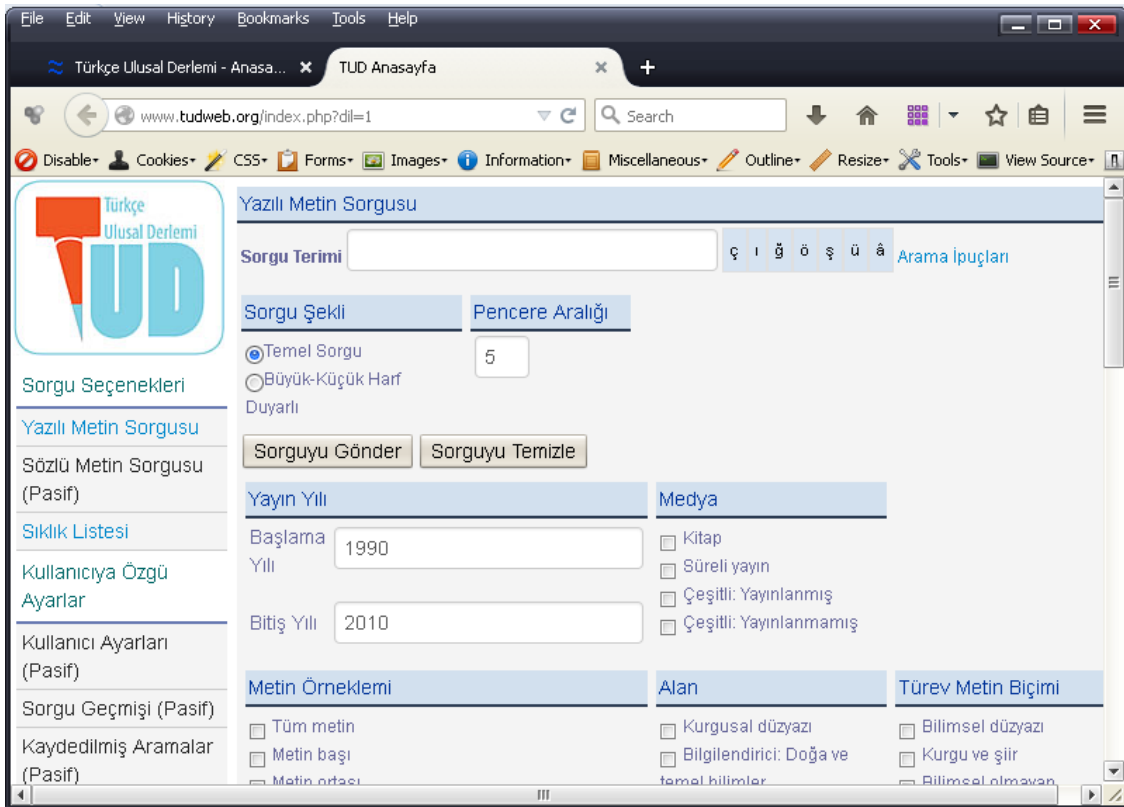
ODTÜ derleminin bir alt kümesi (7262 cümle), biçimbirimsel ve sözdizimsel etiketlerle işaretlenerek ODTÜ-Sabancı Ağaç Yapılı Türkçe Derlemi [26] oluşturulmuştur. Bu derlemde yazı türlerinin dağılım oranı ODTÜ Türkçe Derlemine yakın tutulmuştur. Derlemin yapısı XML tabanlıdır. Araştırmacılar, ODTÜ Türkçe derlemine benzer şekilde bu derlemde faydalanabilmektedir (Şekil 2.3).



Şekil 2.16 ODTÜ-Sabancı Ağaç Yapılı Derleminden örnek bir cümle

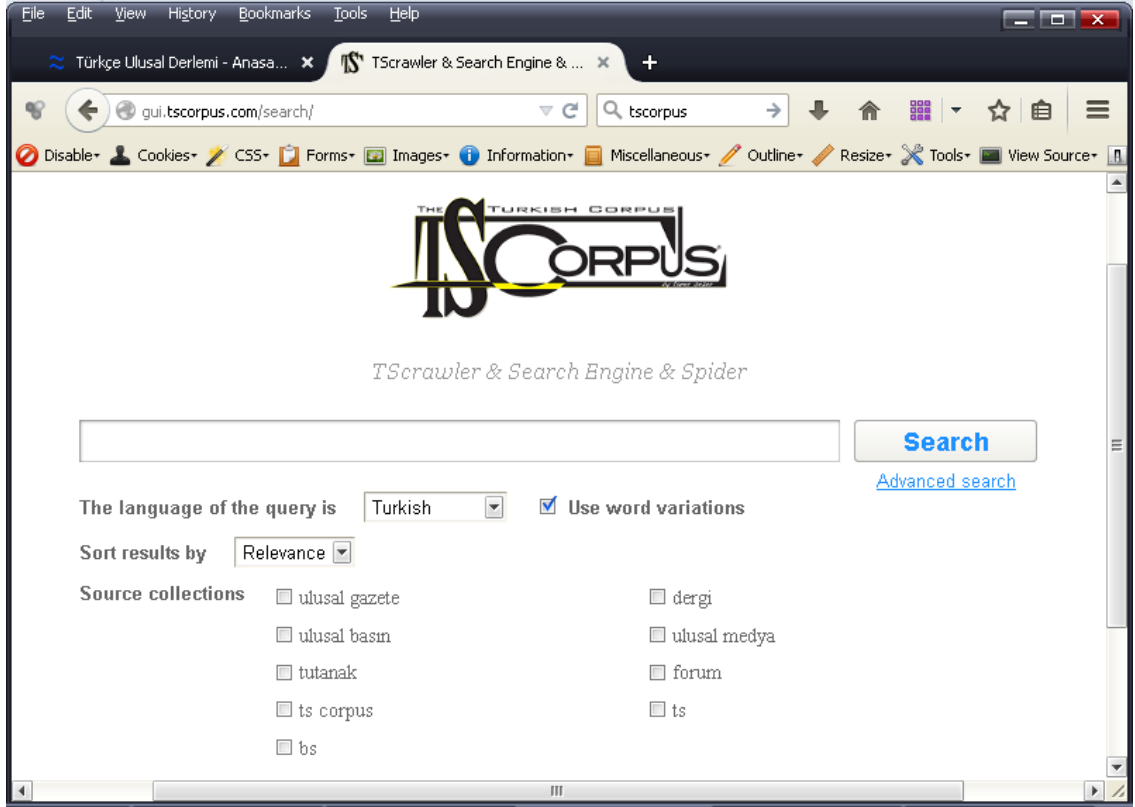
Literatürde geçen bir diğer derlem çalışması Boğaziçi Üniversitesi akademisyenlerince hazırlanan BOUN Corpus'tur [27]. Bu derlem dört alt derlemde meydana gelmektedir. Bunların üçü Türkçe gazetelerden (NewsCor), biri de Türkçe web sayfalarından (GenCor) örneklenmiştir. Derlem toplamda 423 milyon sözcükten oluşmaktadır.

Türkçe Ulusal Derlemi (TUD) [28] 50 milyon sözcükten oluşan, 20 yıllık bir dönemi (1990-2009) kapsayan, günümüz Türkçesinin çok sayıda farklı alan ve türlerinden yazılı ve sözlü örneklerini içeren, geniş kapsamlı, dengeli ve temsil yeterliliğine sahip, genel amaçlı bir referans derlemdir. TUD-Tanıtım Sürümü, 1990-2009 yıllarını kapsayan 4438 veri kaynağından seçilen, 9 konu alanını ve 34 dilsel türü içeren doküman örneklerinden oluşmaktadır. Kullanıcılar yaklaşık 48 milyon sözcük üzerinden, medya, metin örnekleme, konu alanı, türev metin biçimi, yazar cinsiyeti, yazar türü, hedef okur ve metin türü kısıtlama ölçütleriyle sorgularını gerçekleştirebilirler [29] (Şekil 2.4).



Şekil 2. 17 Türkçe Ulusal Derleminin web sorgulama arayüzü

İlk versiyonu 1 Mart 2012'de yayınlanan TS Corpus [30] ise, tamamı sözcük türü ve biçimbirimsel bazda işaretlenmiş toplam 491 milyon birimden (491.360.398 adet token) oluşan genel amaçlı bir Türkçe derlemdir. Web sayfasından çevrimiçi erişime açıktır (Şekil 2.5).



Şekil 2. 18 TS Corpus'un web sorgulama arayüzü

## BÖLÜM 3

### GÖRSOR

Bu bölümde, tez çalışması için oluşturulan test derleminin hazırlanması ve çalışmanın konusu olan GÖRSOR aracının geliştirilme safhaları açıklanmıştır. Ayrıca, uygulamanın kullanım alanları konusunda da örnekler verilmiştir.

#### 3.1 Test Derlemlerinin Hazırlanıp Etiketlenmesi

Çalışmaya ilk olarak küçük test derlemleri oluşturularak başlanmıştır. Derlemleri otomatik etiketleyeceğimiz ve etiketleme doğruluğunun yüksek olmasını istediğimiz için sözcük kökü çeşitliliğinin düşük ve cümle yapılarının göreceli olarak basit olduğu dokümanlardan örnekler kullanılmıştır. Buna göre Milli Eğitim Bakanlığı tarafından 2012 yılında yayınlanan Hayat Bilgisi 1 [31] ve Hayat Bilgisi 2 [32] ders kitapları OCR'den geçirilerek ham metinleri elde edilmiştir. Sonra her metin imla denetiminden geçirilmiş ve metinler cümlelere ayrılmıştır. Cümle segmentasyonunda basit regex ifadeleri yeterli olmuştur. Test derlemlerinin özellikleri Çizelge 3.1'de yer almaktadır.

Çizelge 3. 2 Test Derlemlerinin Özellikleri

Derlem	Sözcük Sayısı	Cümle Sayısı
Hayat Bilgisi 1	3207	655
Hayat Bilgisi 2	6576	1246

Çalışmamız için sözcük türü ve sözdizim etiketleri gerekmektedir. Bu ihtiyacımızı Turkish NLP Pipeline [33] araçlarını kullanarak giderdik. Çalışmamızı yaptığımız sırada

bu araç kümesinin 2013 sürümü elimizde mevcuttu. Kısa süre sonra bu araçlar <http://tools.nlp.itu.edu.tr/> adresinden çevrimiçi kullanıma açılmıştır.

Turkish NLP Pipeline üç ayrı modülden oluşmaktadır. İki seviyeli biçimbirimsel çözümleyici [34], algılayıcı tabanlı biçimbirimsel bulanıklık giderici [35] ve bağımlılık ayrıştırıcı [36].

NLP Pipeline ile derlemimizi etiketlerken öncelikle doküman içerisindeki her birimin (sözcük veya noktalama) bir satırda ve cümle sonlarında da üç adet yıldız karakteri olacak şekilde bir giriş dosyasına yazılması sağlanmıştır (Şekil 3.2). Bu dosya pipeline tarafından önce biçimbirimsel çözülemeye tabi tutulmuş, sözcüklerin kök ve ekleri belirlenmiştir. Bu işlemin çıktısı biçimbirimsel bulanıklık gidericiye giriş olarak aktarılmıştır. Her sözcük için en olası biçimbirimsel çözümleme seçilmiş ve bu işlemin çıktısı da sözdizimsel çözümleyiciye giriş olarak verilmiştir. Bu aşamada sözcüklerin hangi sözcükle sözdizimsel ilişki içerisinde olduğu referans numaraları ile belirtilmiştir.

Şekil 3.1’de ‘Ömer, çantasına görsel sanatlar dersi için resim defteri ve boya kalemlerini koydu.’ cümlesi için etiketleme işleminin basamakları gösterilmiştir. Şekil 3.1’deki son kutudaki analizler CoNLL [37] formatında verilmiştir. Sütunlarda sırasıyla birim numarası, sözcük-noktalama, sözcük kökü, sözcük türü, sözcük türü alt grubu, biçimbirimsel özellikler, referans edilen sözcük numarası ve bağımlılık ilişkisi yer almaktadır.

Tüm cümleler etiketlendikten sonra aralarına ayırıcı bir kod konularak özel bir formatta derlem dosyası halinde kaydedilmiştir.



Ömer , çantasına görsel sanatlar dersi için resim defteri ve boya kalemlerini koydu .

BIÇİMBİRİMSEL ANALİZ

Ömer Ömer+Noun+Prop+A3sg+Pnon+Nom  
 , , +Punc  
 çantasına çanta+Noun+A3sg+P3sg+Dat  
 görsel görsel+Adj  
 sanatlar sanat +Noun+A3pl+Pnon+Nom  
 sanatlar sanat +Noun+A3sg+Pnon+Nom^DB+Verb+Zero+Pres+A3pl  
 dersi de +Verb+Pos^DB+Adj+AorPart^DB+Adj+JustLike  
 dersi ders +Noun+A3sg+P3sg+Nom  
 dersi ders +Noun+A3sg+Pnon+Acc  
 için iç+Noun+A3sg+P2sg+Nom  
 için iç+Noun+A3sg+Pnon+Gen  
 için iç+Verb+Pos+Imp+A2pl  
 için için+Postp+PCGen  
 için için+Postp+PCNom  
 resim resim +Noun+A3sg+Pnon+Nom  
 defteri defter +Noun+A3sg+Pnon+Acc  
 defteri defter +Noun+A3sg+P3sg+Nom  
 ve ve +Conj  
 boya boy +Noun+A3sg+Pnon+Dat  
 boya boya +Noun+A3sg+Pnon+Nom  
 boya boya +Verb+Pos+Imp+A2sg  
 kalemlerini kalem +Noun+A3pl+P3sg+Acc  
 kalemlerini kalem +Noun+A3pl+P2sg+Acc  
 kalemlerini kalem +Noun+A3pl+P3pl+Acc  
 kalemlerini kalem +Noun+A3sg+P3pl+Acc  
 koydu koy +Noun+A3sg+Pnon+Nom^DB+Verb+Zero+Past+A3sg  
 koydu koy +Verb+Pos+Past+A3sg  
 . . +Punc

BIÇİMBİRİMSEL BLANIKLIK GİDERME

Ömer Ömer+Noun+Prop+A3sg+Pnon+Nom  
 , ,+Punc  
 çantasına çanta+Noun+A3sg+P3sg+Dat  
 görsel görsel+Adj  
 sanatlar sanat+Noun+A3pl+Pnon+Nom  
 dersi ders+Noun+A3sg+P3sg+Nom  
 için için+Postp+PCNom  
 resim resim+Noun+A3sg+Pnon+Nom  
 defteri defter+Noun+A3sg+P3sg+Nom  
 ve ve+Conj  
 boya boya+Noun+A3sg+Pnon+Nom  
 kalemlerini kalem+Noun+A3pl+P3sg+Acc  
 koydu koy+Verb+Pos+Past+A3sg  
 . .+Punc

SÖZDİZİMSEL ANALİZ

1	Ömer	Ömer	Noun	Prop	A3sg Pnon Nom	13	SUBJECT
2	,	,	Punc	Punc		3	notconnected
3	çantasına	çanta	Noun	Noun	A3sg P3sg Dat	13	DATIVE.ADJUNCT
4	görsel	görsel	Adj	Adj		5	MODIFIER
5	sanatlar	sanat	Noun	Noun	A3pl Pnon Nom	6	CLASSIFIER
6	dersi	ders	Noun	Noun	A3sg P3sg Nom	7	OBJECT
7	için	için	Postp	Postp	PCNom	13	MODIFIER
8	resim	resim	Noun	Noun	A3sg Pnon Nom	9	CLASSIFIER
9	defteri	defter	Noun	Noun	A3sg P3sg Nom	10	OBJECT
10	ve	ve	Conj	Conj		11	COORDINATION
11	boya	boya	Noun	Noun	A3sg Pnon Nom	12	CLASSIFIER
12	kalemlerini	kalem	Noun	Noun	A3pl P3sg Acc	13	OBJECT
13	koydu	koy	Verb	Verb	Pos Past A3sg	14	SENTENCE
14	.	.	Punc	Punc		15	notconnected

Şekil 3. 19 Bir cümleinin NLP Pipeline ile analizi

```
Adamın  
alındığı  
yer  
belli  
.  
***  
Hem  
neden  
alınmasın  
ki  
?  
***
```

Şekil 3. 20 NLP Pipeline için örnek bir giriş dosyası

### 3.2 İlişki Tablosu

Örnek derlemlerimiz için, etiketli derlem dosyaları oluşturulduktan sonra bu etiketlerin doğruluklarının sınamasına geçilmiştir. Bu iş yapılırken her cümle için Şekil 3.1’de en altta yer alan kutudaki biçimbirimsel ve sözdizimsel analizlerin doğrulanması gerçekleştirilmiştir. Biçimbirimsel analizler bu kutudan kolayca takip edilebilmesine karşın, sözdizimsel analizlerin takibi cümle yapısı karmaşıklaştıkça zorlaşmaktadır. Bu takibi kolaylaştırmak için aracımıza bir ‘ilişki tablosu’ modülü eklenmiştir. Bu sayede derlemden her hangi bir cümle seçildiğinde bilgisayar tarafından sözcükler arasındaki sözdizimsel ilişkileri kuş bakışı gösteren iki boyutlu bir grafik üretilmektedir.

Bu grafikte sütünlarda sözcükler, satırlarda da (ilişki katmanlarında) ilişki tipleri yer almaktadır. İlişkiler ise baklava parçaları ile gösterilmektedir. Baklavaları oluşturan ilk üçgene (hücrede sağa yapışık) ait sözcük referans veren, ikinci üçgene (hücrede sola yapışık) ait sözcük de referans alan sözcüktür.

Şekil 3.3’de bu grafiğe bir örnek verilmiştir. Satırlardaki ilişki tipi isimleri kısaltma şeklindedir ve Çizelge 3.2’de kısaltmaların açıklamaları verilmiştir.

Öner	çantasına	görsel	sanatlar	dersi	için	resim	defteri	ve	boya	kalemlerini	koydu	
							1 1	2 2				CRD
				1 1								OBJ-P
			1 1			2 2			3 3			CLA
		1 1										MOD-P
1												SUB
										1 1		OBJ-S
	1											JNT-S
					1							MOD-S

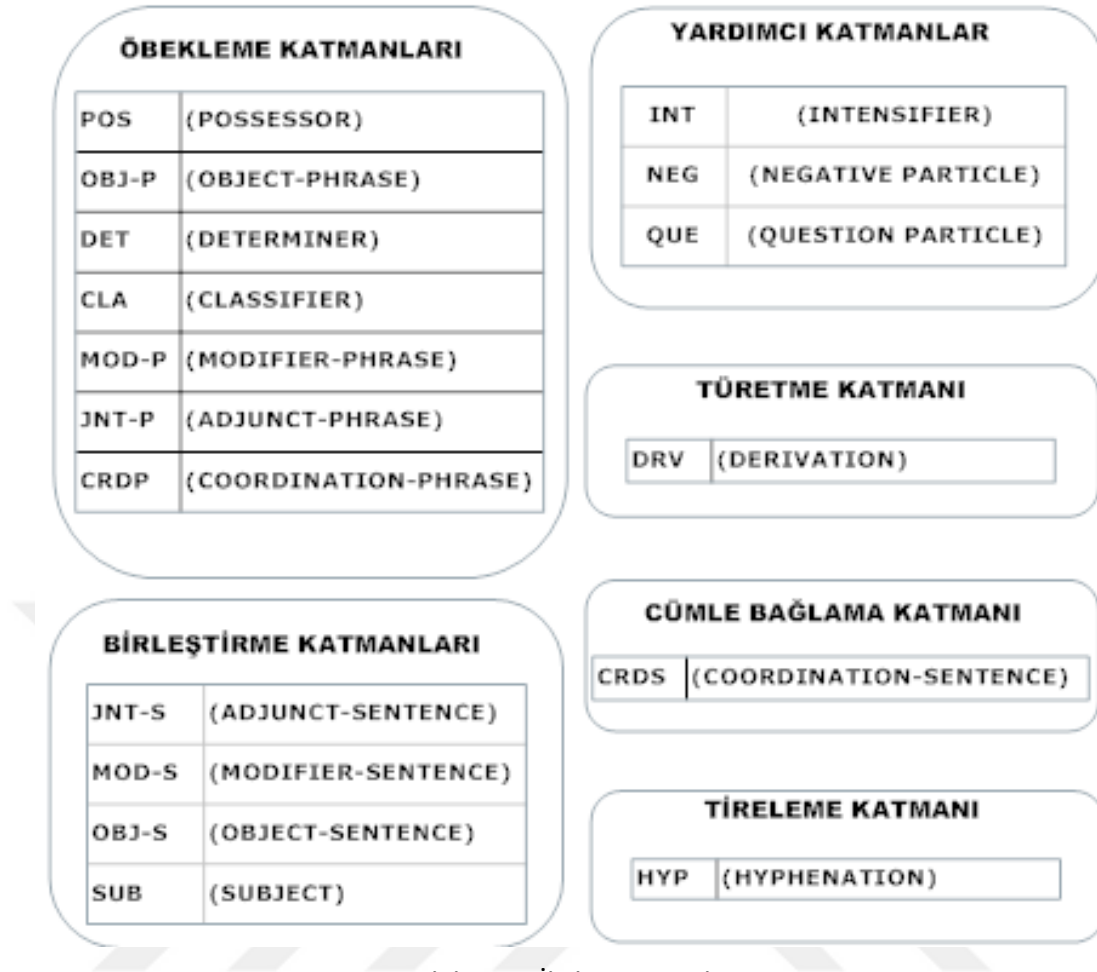
Şekil 3. 21 Örnek bir cümle için otomatik olarak üretilen ilişki tablosu

Çizelge 3. 3 İlişki Tablosundaki Kısaltmaların Açıklamaları

Kısaltma	Açıklama
CRD	Coordination
CLA	Classifier
SUB	Subject
OBJ	Object
JNT	Adjunct
MOD	Modifier

Ayrıca, bazı ilişki tipleri için hem TİP-P hem de TİP-S versiyonları mevcuttur. İlişkinin bir öbek (P) mi yoksa bir cümle (S) mi oluşturduğuna bakılarak hangi versiyonun kullanılacağı belirlenir.

Bir ilişki tablosu için olası ilişki katmanları grupları ve bunlara ait ilişki katmanları Şekil 3.4'deki gibi belirlenmiştir. Öbeikleme katmanındaki ilişkiler sözcük öbeklerini oluşturmaktadır. Birleştirme katmanındaki ilişkiler ise cümleyi meydana getiren en büyük yapıları birleştirerek cümleyi oluşturmaktadır. Türetme katmanı sözcüklerdeki türeme ilişkilerini göstermekte, cümle bağlama katmanı ise cümleleri bağlaçlarla birbirine bağlamaktadır.



Şekil 3. 22 İlişki Katmanları

Örnek derlemimizdeki cümlelerin analizleri, otomatik oluşturulan ilişki tabloları yardımıyla kontrol edilmiştir. Basit yapılarından dolayı örnek derlemlerimizdeki cümlelerin söz dizimsel analizlerinin hata oranları düşük çıkmıştır.

Sözdizimsel etiketlemede karşılaştığımız hataların en önemli nedeni yanlış biçimbirimsel çözümlemenin seçilmiş olmasıdır. Örneğin ‘yazınız’, ‘veriniz’ ve ‘düşünürler’ gibi fiillerin isim olarak algılanması sözdizimsel analizin yanlış yapılmasına neden olmuştur:

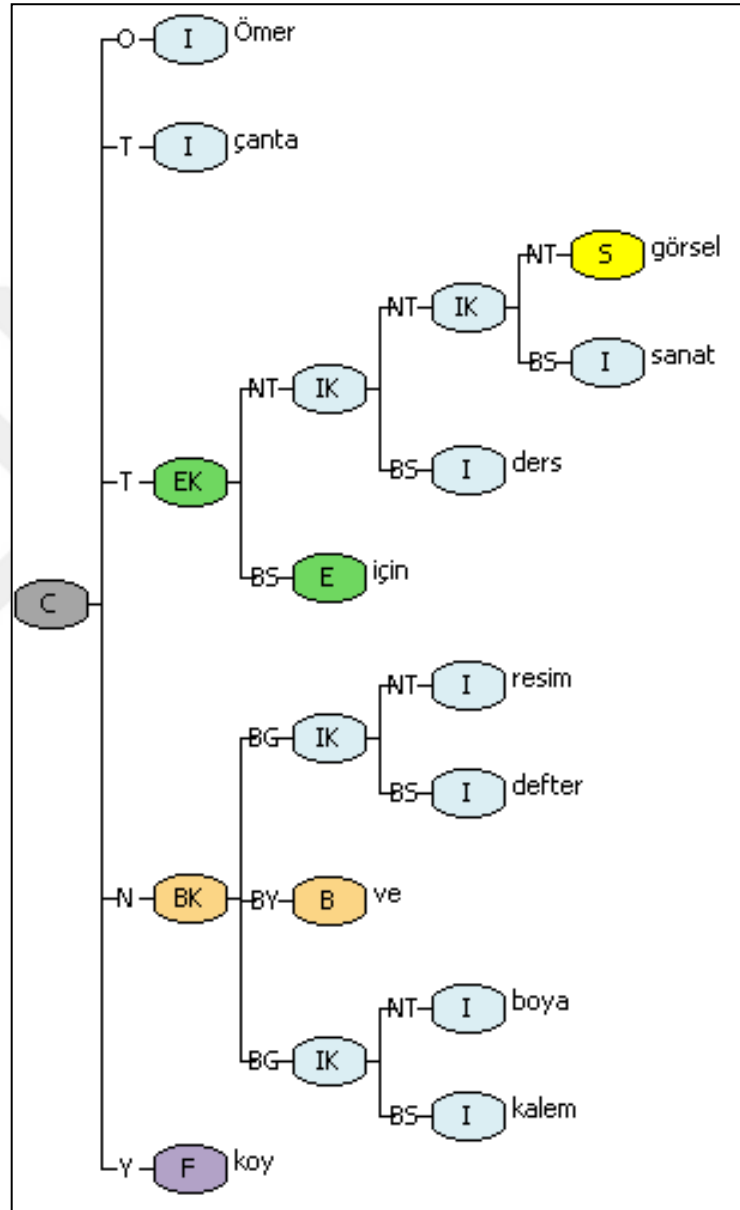
yazınız = yazı+Noun+A3sg+P2pl+Nom

veriniz =veri+Noun+A3sg+P2pl+Nom

düşünürler = düşünür +Noun+A3pl+Pnon+Nom

### 3.3 İlişki Diyagramı

İlişki tablolarından sonra sözcükten öbeklere, cümleciklere ve cümlelere kadar her sözdizimsel düzeyden elde edilen yapıları gösteren bir ilişki diyagramı modülü oluşturulmuştur. İlişki diyagramları oluşturulurken ilişki tablosundan faydalanılmıştır. Şekil 3.5’de bir önceki örnek cümlemiz için üretilen ilişki diyagramı görülmektedir.



Şekil 3.23 Örnek bir cümle için ilişki diyagramı

Diyagramda dört isim öbeği, bir edat öbeği (tümleç durumunda), bir bağlaç öbeği (nesne durumunda) vardır. Kısaltmaların anlamları Çizelge 3.3’de verilmiştir.

Çizelge 3. 4 Diyagramdaki Kısaltmaların Açıklamaları

Kısaltma	Açıklama
O	Özne
N	Nesne
Y	Yüklem
T	Tümleç
I	İsim
S	Sıfat
B	Bağlaç
F	Fiil
C	Cümle
İK	İsim Öbeği
EK	Edat Öbeği
BK	Bağlaç Öbeği
NT	Niteleyici
BS	Baş sözcük
BG	Bağlanan
BY	Bağlayıcı

Şekil 3.5'deki gibi bir graf yapısı elde edilirken aşağıdaki işlemler gerçekleştirilir:

- 1) Her sözcük için öbeleme katmanlarındaki tüm katmanlar yukarıdan aşağıya doğru taranır, eğer katmanda sözcüğün referans verdiği bir ilişki varsa bu ilişkiyi oluşturan sözcükler öbeleme kurallarına göre öbeklenir.
- 2) Birleştirme katmanlarındaki tüm ilişkiler işlenerek birleştirme kurallarına göre cümle oluşturulur.
- 3) Türetme katmanındaki tüm ilişkiler işlenir.
- 4) Yardımcı katmanlardaki tüm ilişkiler işlenerek yardımcı katman kurallarına göre gerekli öbekler oluşturulur.
- 5) Cümle bağlama katmanındaki tüm ilişkiler işlenerek cümleler bağlaçlarla birbirine bağlanır.

Sözü geçen kurallar graf oluşturma kurallarıdır. Bir cümle için oluşturmakta olduğumuz grafa, a ve b düğümlerini (sözcükleri) ekleyeceğimizi ve bunların  $a \rightarrow b$  şeklinde bir sözdizimsel ilişkiye sahip olduğunu varsayalım. Buna göre kurallar Çizelge 3.4, 3.5, 3.6, 3.7 ve 3.8'deki gibidir. Burada n, n1, n2 herhangi bir düğümü (her düğüm bir sözdizimsel birimi temsil eder), n' yeni oluşan bir düğümü, kırmızı (koyu renk) ok iptal

edilen bir ilişkiyi, yeşil ok (açık renk) yeni kurulan bir ilişkiyi, EBA(x) ise x düğümünün en büyük atası olan düğümü ifade etmektedir.

Çizelge 3. 5 Öbekleme işlemi için graf oluşturma kuralları

ÖBEKLEME KURALLARI	
Durum	İşlem
$a \rightarrow b$ bağlantısı yapılacak. a ve b grafta yok.	$a \rightarrow n', b \rightarrow n'$
$a \rightarrow b$ bağlantısı yapılacak. grafta a yok, $b \rightarrow n$ var.	$b \rightarrow n, a \rightarrow n', b \rightarrow n', n' \rightarrow n$
$a \rightarrow b$ bağlantısı yapılacak. grafta $a \rightarrow n$ var, b yok.	$EBA(a) \rightarrow n', b \rightarrow n'$
$a \rightarrow b$ koordinasyon bağlantısı yapılacak. grafta $a \rightarrow n_1$ (koordinasyon öbeği) ve $b \rightarrow n_2$ var.	$b \rightarrow n_2, b \rightarrow n_1, n_1 \rightarrow n_2$

Çizelge 3. 6 Öbek bağlama işlemi için graf oluşturma kuralları

ÖBEK BAĞLAMA KURALLARI	
Durum	İşlem
$a \rightarrow b$ bağlantısı yapılacak. a ve b grafta yok.	$a \rightarrow n', b \rightarrow n'$
$a \rightarrow b$ bağlantısı yapılacak. grafta $a \rightarrow n$ var, b yok. a'nın tipi noktalama veya bağlaç değilse: aksi durumda:	$EBA(a) \rightarrow n', b \rightarrow n'$ $b \rightarrow n$
$a \rightarrow b$ bağlantısı yapılacak. grafta $a \rightarrow n_1$ ve $b \rightarrow n_2$ var.	$EBA(b) \rightarrow n_1$

Çizelge 3. 7 Öbek birleştirme işlemi için graf oluşturma kuralları

BİRLEŞTİRME KURALLARI	
Durum	İşlem
$a \rightarrow b$ bağlantısı yapılacak. a ve b grafta yok.	$a \rightarrow n', b \rightarrow n'$
$a \rightarrow b$ bağlantısı yapılacak. grafta $a \rightarrow n$ var, b yok.	$EBA(b) \rightarrow n$
$a \rightarrow b$ bağlantısı yapılacak. grafta a yok, $b \rightarrow n$ var.	$a \rightarrow n$
$a \rightarrow b$ bağlantısı yapılacak. grafta $a \rightarrow n_1$ ve $b \rightarrow n_2$ var.	$EBA(a) \rightarrow n_2$

Çizelge 3. 8 Türetme işlemi için graf oluşturma kuralları

TÜRETME KURALLARI	
Durum	İşlem
$a \rightarrow b$ bağlantısı yapılacak. grafta a yok, $b \rightarrow n$ var.	$a \rightarrow b$
$a \rightarrow b$ bağlantısı yapılacak. grafta $a \rightarrow n_1$ ve $b \rightarrow n_2$ var.	$n_1 \rightarrow b$

Çizelge 3. 9 Cümle bağlama işlemi için graf oluşturma kuralları

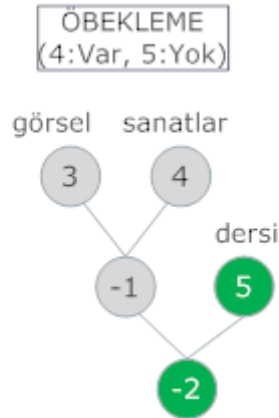
CÜMLE BAĞLAMA KURALLARI	
Durum	İşlem
$a \rightarrow b$ bağlantısı yapılacak. grafta $a \rightarrow n_1$ ve $b \rightarrow n_2$ var. eğer $EBA(b) \rightarrow C$ ise; değilse;	$EBA(b) \rightarrow EBA(a)$ $EBA(a) \rightarrow EBA(b)$
$a \rightarrow b$ bağlantısı yapılacak. grafta $a \rightarrow n$ var, $b$ yok.	$EBA(a) \rightarrow n', b \rightarrow n'$

Aşağıda bu kuralların nasıl çalıştığını gösteren bir örnek verilmiştir. Şekil 3.6 ile Şekil 3.16 arasındaki şekillerde 'Ömer çantasına görsel sanatlar dersi için resim defterini ve boya kalemlerini koydu' cümlesi için adım adım  $a \rightarrow b$  ilişkilerinin graf üzerinde hangi değişimleri gerçekleştirdiği görülmektedir.



Şekil 3. 24 Örnek grafın oluşturulması Adım 1

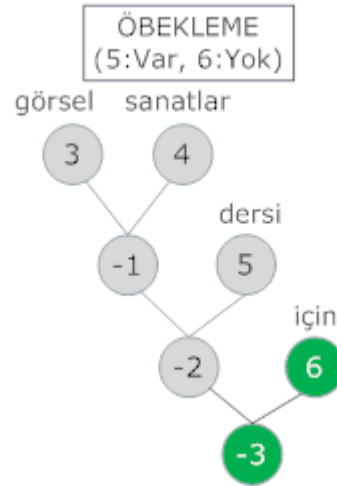
Başlangıçta grafımız boştur. 3 ve 4 nolu sözcükler öbekleme kurallarına göre öbeklendiğinde grafta 3, 4 ve -1 nolu düğümler oluşur. -1 nolu düğüm bir sıfat tamlamasıdır (Şekil 3.6).



Şekil 3. 25 Örnek grafın oluşturulması Adım 2

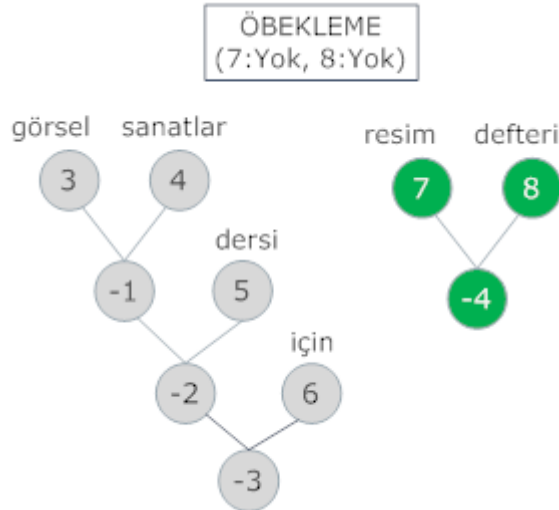


4 ve 5 nolu sözcükler öbeklenirken 4'ün grafta olduğu 5'inse olmadığı gözlemlenir. 5 düğümü oluşturur. 4 ile ilişki içersindedir. Kurala göre 4'ün atasıyla birleşerek -2 düğümünü oluşturur. -2 düğümü de bir isim öbeğidir (Şekil 3.7).



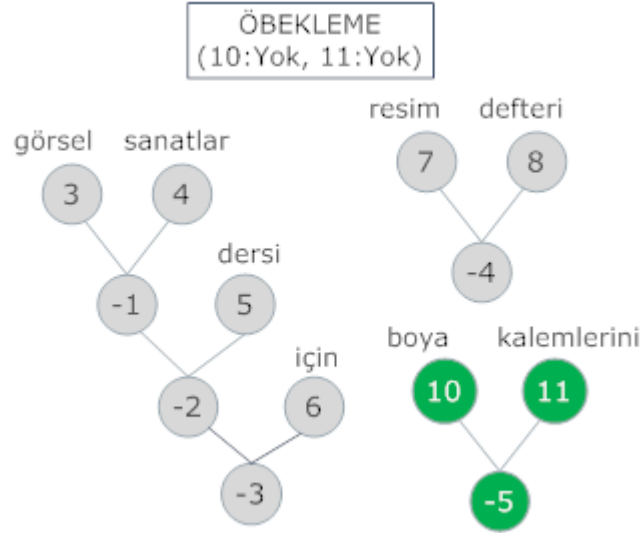
Şekil 3. 26 Örnek grafin oluşturulması Adım 3

6'ıncı düğüm grafta yoktur bu yüzden oluşturulur. 5'inci düğüm grafta vardır ve kurala göre bu düğümün atası ile 6 nolu düğüm birleşerek -3 nolu düğümü oluştururlar. 6 nolu düğüm bir edat olduğundan -3 nolu düğüm bir edat öbeğidir (Şekil 3.8).



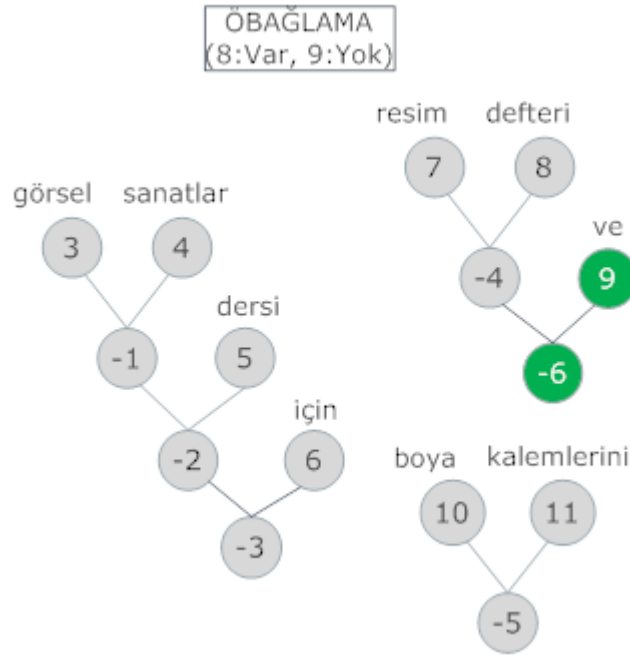
Şekil 3. 27 Örnek grafin oluşturulması Adım 4

7 ve 8 nolu sözcükler grafta yoktur. Bu yüzden 7 ve 8 nolu düğümler oluşturulur ve birbirlerine bağlanarak -4 nolu düğüm meydana gelir. -4 nolu düğüm bir isim tamlamasıdır (Şekil 3.9).



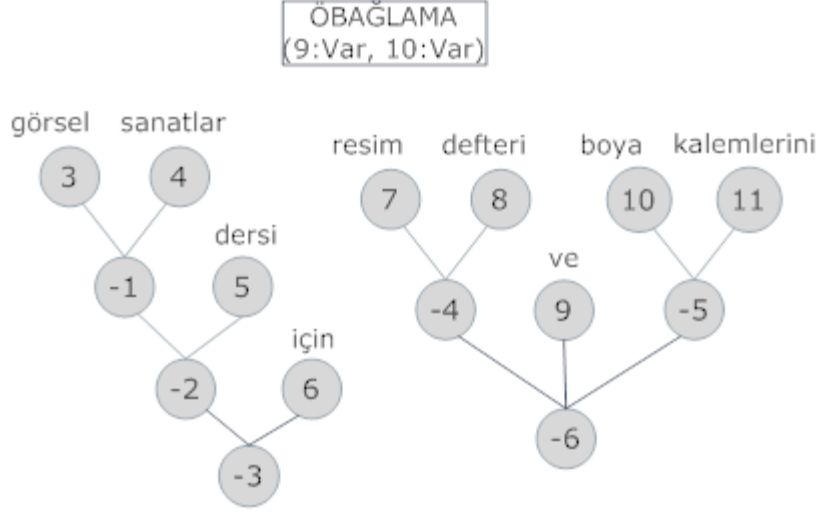
Şekil 3. 28 Örnek grafin oluşturulması Adım 5

10 ve 11 nolu sözcükler grafta yoktur. Bu yüzden 10 ve 11 nolu düğümler oluşturulur ve birbirlerine bağlanarak -5 nolu düğüm meydana gelir. -5 nolu düğüm de bir isim tamlamasıdır (Şekil 3.10).



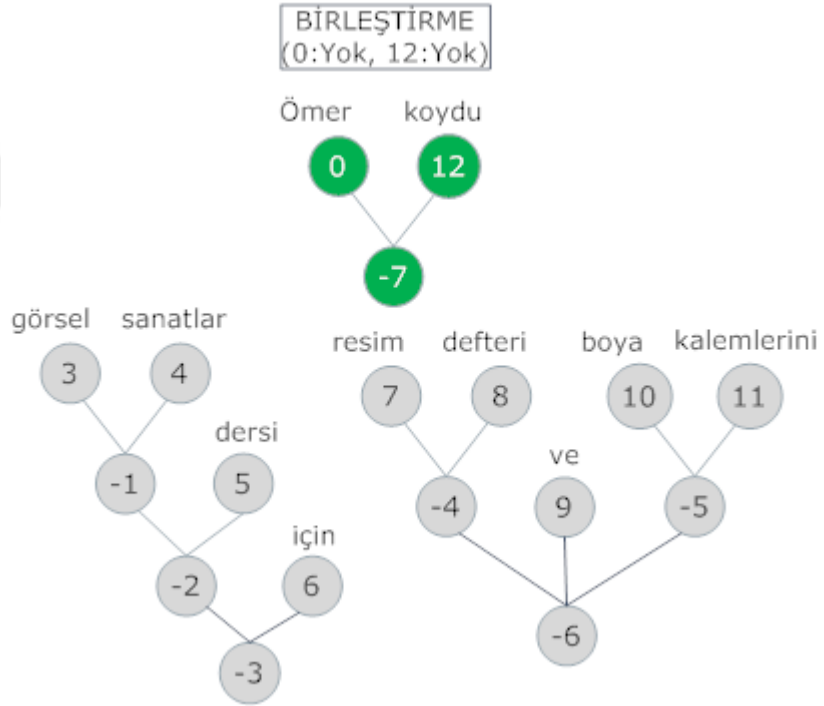
Şekil 3. 29 Örnek grafin oluşturulması Adım 6

9 nolu sözcük grafta yoktur bu yüzden 9 nolu düğüm oluşturulur. Öbek bağlama kurallarına göre 8 nolu düğümün atasıyla birleştirilerek -6 nolu düğüm oluşturulur. 9 nolu düğüm bir bağlaç olduğundan -6 nolu düğüm bir bağlaç öbeğidir (Şekil 3.11).



Şekil 3. 30 Örnek grafin oluşturulması Adım 7

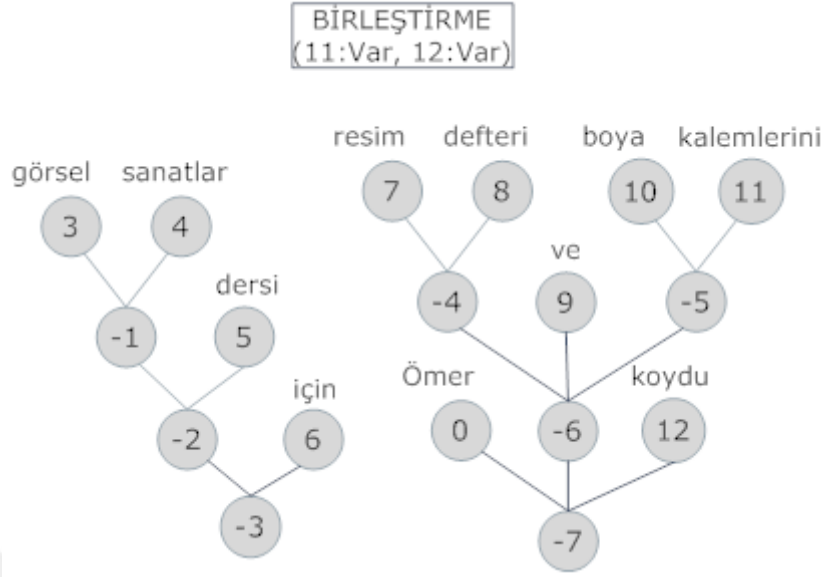
9 ve 10 nolu sözcükler grafta vardır. Kurala göre 9 nolu düğümün bağlı olduğu düğüme 10 nolu düğümün atası da bağlanır ve bağlaç öbeği tamamlanır (Şekil 3.12).



Şekil 3. 31 Örnek grafin oluşturulması Adım 8

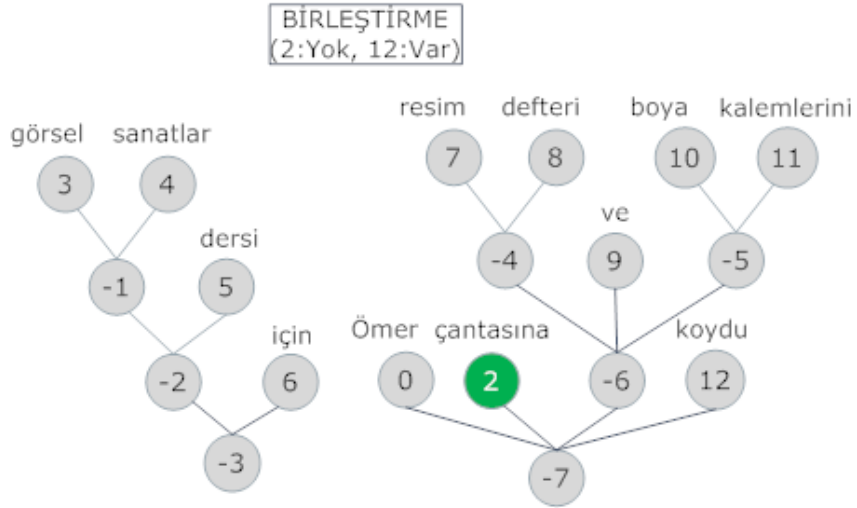
0 ve 12 nolu sözcükler grafta yoktur, bu yüzden 0 ve 12 nolu düğümler oluşturulur. 0 ve 12 nolu sözcükler arasında özne-yüklem ilişkisi vardır. Bu yüzden birleştirme kurallarına

göre yeni bir düğüm altında birleştirilirler. Yeni oluşan -7 nolu düğüm bir cümledir (Şekil 3.13).



Şekil 3. 32 Örnek grafın oluşturulması Adım 9

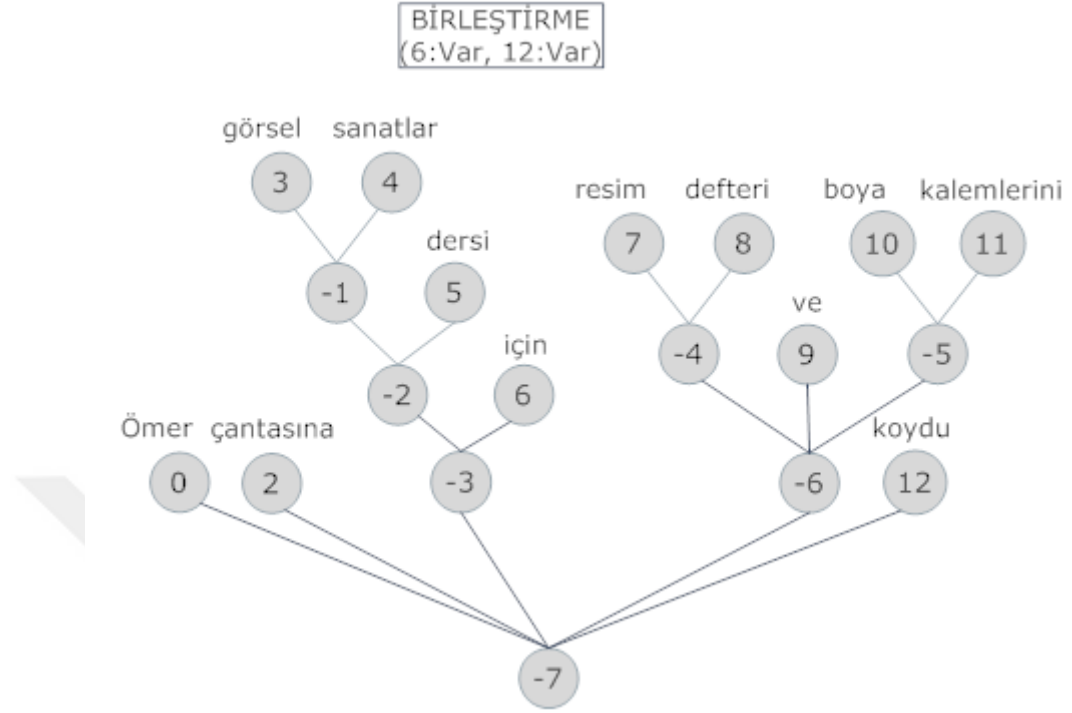
11 ve 12 nolu düğümler arasında nesne yüklem ilişkisi vardır. Kurala göre 11 nolu düğümün en büyük atası 12 nolu düğümün bağlı olduğu düğüme bağlanır. Böylece cümlenin nesnesi yerine yerleştirilmiş olur (Şekil 3.14).



Şekil 3. 33 Örnek grafın oluşturulması Adım 10

2 nolu sözcük grfta yoktur bu yüzden 2 nolu düğüm oluşturulur. 2 ve 12 nolu sözcükler arasında tümleç-yüklem ilişkisi vardır. Kurala göre 2 nolu düğüm 12 nolu

düğümün bağlı olduğu düğüme bağlanır. Böylece cümlenin dolaylı tümleci yerine yerleştirilmiş olur (Şekil 3.15).



Şekil 3. 34 Örnek grafin oluşturulması Adım 11

6 ve 12 nolu sözcükler arasında da tümleç-yüklem ilişkisi vardır. Kurala göre 6 nolu düğümün atası 12'nin bağlı olduğu düğüme bağlanır. Böylece cümlenin edat tümleci de yerine yerleştirilmiş olur (Şekil 3.16).

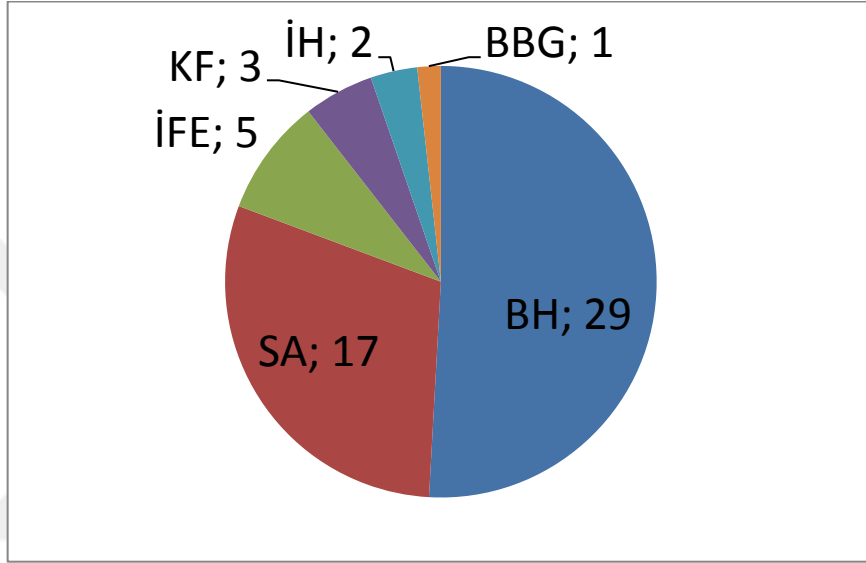
Graf oluşturma kurallarının başarımları test derlemleri üzerinde cümle bazında ölçülmüştür. Çizelge 3.9' da diyagramı doğru çıkarılan cümlelerin sayıları ve oranları verilmiştir.

Çizelge 3. 10 Graf Oluşturma Kurallarının Başarımları

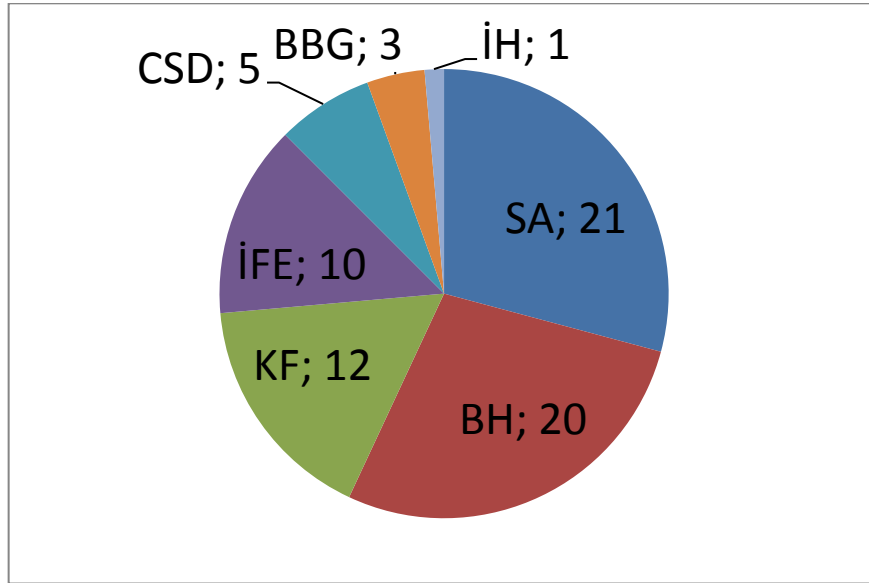
Derlem	Doğru Diyagram Sayısı	Başarı Oranı
Hayat Bilgisi 1	594	%90.68
Hayat Bilgisi 2	1165	%93.49

Tablodaki sayılar sözdizimsel diyagramı tamamıyla doğru çıkarılmış cümle sayılarını göstermektedir. Bir cümlenin diyagramında kısmi bir hata dahi olsa bu diyagram hatalı kabul edilmiştir.

Diyagramı hatalı çıkarılan cümlelerdeki hatalar incelendiğinde bu hataların belirli başlıklar altında toplanabildiği gözlenmiştir. Şekil 3.17 ve 3.18’de her iki derlemdeki hata türleri ve gözlenme sayıları verilmiştir. Çizelge 3.10’da hata isimlerinin kısaltmalarının açıklamaları yer almaktadır.



Şekil 3. 35 Derlem 1’e ait hata analizi grafiği

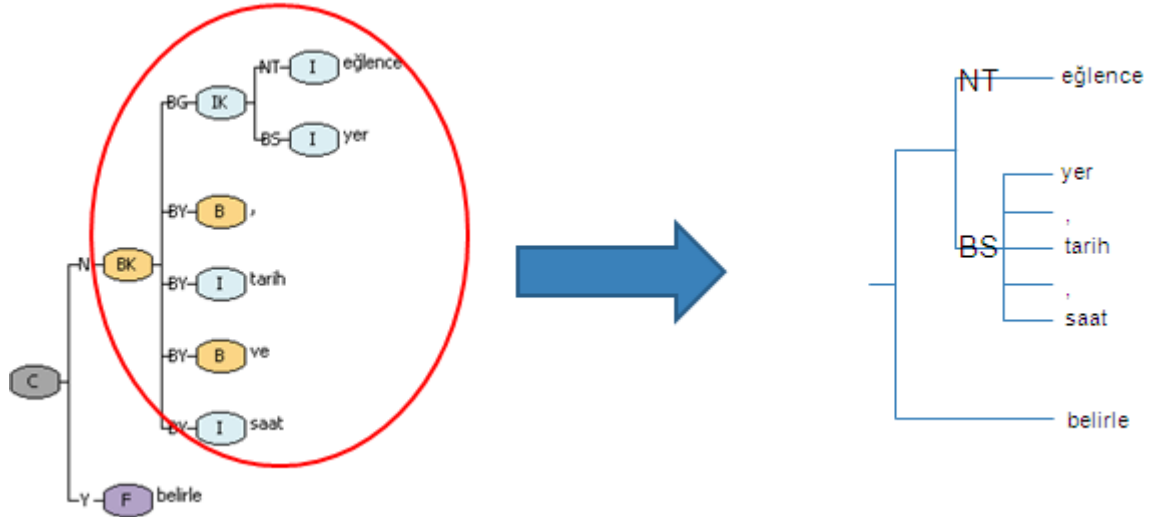


Şekil 3. 36 Derlem 2’ye ait hata analizi grafiği

Çizelge 3. 11 Hata Kısaltmalarının Açıklamaları

Kısaltma	Açıklama
BH	Bağlama Hatası
SA	Sözdizimsel Analiz Hatası
İFE	İsimfiil Eki Hatası
KF	Kopulasız Fiil Hatası
İH	İmla Hatası
BBG	Biçimbirimsel Analiz Hat.
CSD	Cümleden Sonra Düğüm

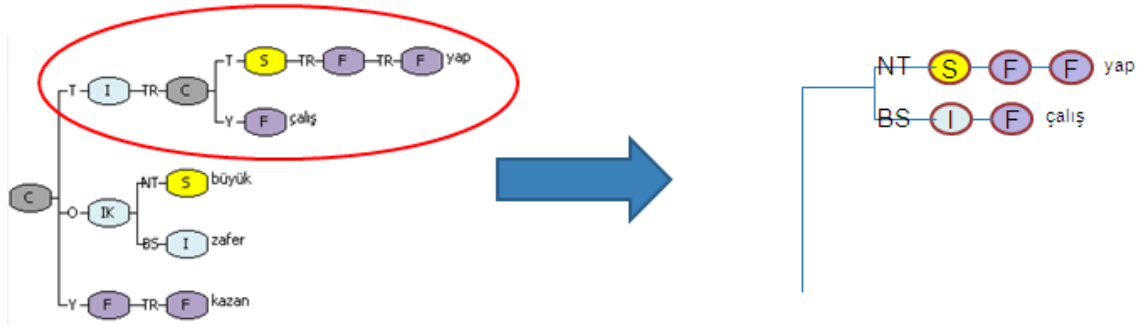
Aşağıda çizelgedeki hataların kaynakları açıklanmış ve her birine birer örnek verilmiştir.



Şekil 3. 37 Bağlama Hatası Örneği

Şekil 3.19'da sol tarafta 'Eğlencenin yerini , tarihini ve saatini belirlemeliyiz.' cümlesinin GÖRSOR aracı tarafından üretilen ilişki diyagramı yer almaktadır. Kırmızı halka içersine alınmış kısım hatalı üretilmiştir. Hata, bağlama grubunun yanlış oluşturulmasından kaynaklanmıştır. (Bağlama Hatası)

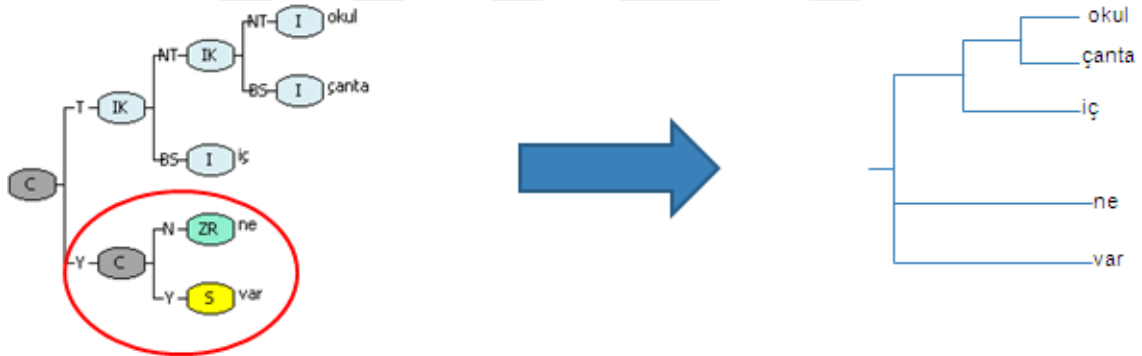
Doğru diyagram sağdaki gibi olmalıdır. Bu diyagrama göre 'eğlencenin' ortak tamlayan 'yerini', 'tarihini' ve 'saatini' sözcüklerinin oluşturduğu bağlama grubu ise tamlanandır.



Şekil 3. 38 İsim-fiil Eki Hatası Örneği

Şekil 3.20’de sol tarafta ‘Yapılan çalışmalarla büyük zaferler kazanıldı.’ cümlesinin GÖRSOR aracı tarafından üretilen ilişki diyagramı yer almaktadır. Kırmızı halka içersine alınmış kısım hatalı üretilmiştir. Hata, isim-fiil ekinin burada yan cümle oluşturduğu yanlışlığından kaynaklanmıştır. (İsim-fiil Eki Hatası)

Doğru diyagram sağdaki gibi olmalıdır. Bu diyagrama göre ‘yapılan çalışmalar’ bir sıfat tamlaması oluşturmaktadır.

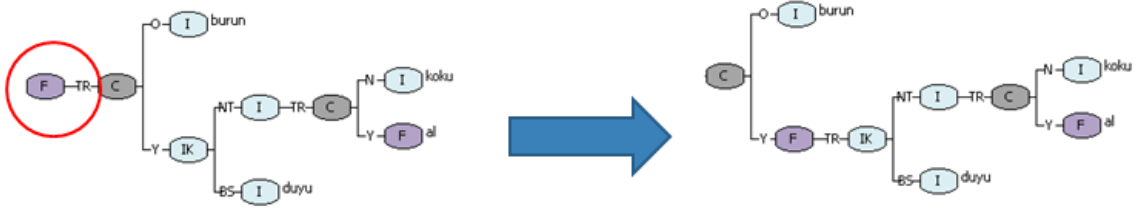


Şekil 3. 39 Kopulasız Fiil Hatası Örneği

Şekil 3.21’de sol tarafta ‘Okul çantanızın içinde neler var ?’ cümlesinin GÖRSOR aracı tarafından üretilen ilişki diyagramı yer almaktadır. Kırmızı halka içersine alınmış kısım hatalı üretilmiştir. Hata, ‘ne’ ve ‘var’ sözcüklerinin bağlı olduğu fazladan bir yan cümlecik üretilmesidir. (Kopulasız Fiil Hatası)

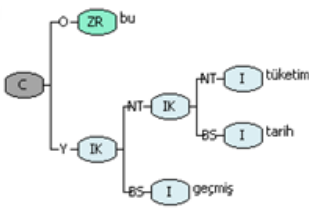
Doğru diyagram sağdaki gibi olmalıdır. Bu diyagrama göre ‘ne’ ve ‘var’ sözcükleri ana cümlecığe bağlı olmalıdır.





Şekil 3. 40 CSD Hatası Örneği

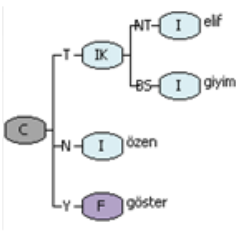
Şekil 3.22’de sol tarafta ‘Burun koku alma duyusudur.’ cümlesinin GÖRSOR aracı tarafından üretilen ilişki diyagramı yer almaktadır. Kırmızı halka içersine alınmış kısım hatalı üretilmiştir. Hata, kök düğümün fiil olarak belirlenmesidir. (Cümleden Sonra Gelen Düğüm Hatası) Doğru diyagram sağdaki gibi olmalıdır. Bu diyagrama göre fiil cümleye bağlıdır.



Analiz							
ID	FORM	LEMMA	CPOSTAG	POSTAG	FEATS	HEAD	DEPREL
0	Bunun	bu	Pron	DemonsP	A3sg Pnon Gen	3	SUBJECT
1	tüketim	tüketim	Noun	Noun	A3sg Pnon Nom	2	CLASSIFIER
2	tarihi	tarih	Noun	Noun	A3sg P3sg Nom	3	OBJECT
3	geçmiş	geçmiş	Noun	Noun	A3sg Pnon Nom	6	OBJECT
4	.	.	Punc	Punc	-	5	notconnected

Şekil 3. 41 Biçimbirimsel Bulanıklık Giderme Hatası Örneği

Şekil 3.23’de sol tarafta ‘Bunun tüketim tarihi geçmiş.’ cümlesinin GÖRSOR aracı tarafından üretilen ilişki diyagramı yer almaktadır. Sağdaki tabloda ‘geçmiş’ sözcüğünün türünün isim olarak belirlendiği görülmektedir. Oysa ki sözcüğün bu cümledeki türü fiildir. (Biçimbirimsel Bulanıklık Giderme Hatası) Biçimbirimsel bulanıklık gidermedeki hata sözdiziminin belirlenmesine de yansımıştır.



Analiz							
ID	FORM	LEMMA	CPOSTAG	POSTAG	FEATS	HEAD	DEPREL
0	Elif	elif	Noun	Noun	A3sg Pnon Nom	1	CLASSIFIER
1	giyimine	giyim	Noun	Noun	A3sg P3sg Dat	3	DATIVE.ADJUNCT
2	özen	özen	Noun	Noun	A3sg Pnon Nom	3	OBJECT
3	gösterir	göster	Verb	Verb	Pos Aor A3sg	5	SENTENCE
4	.	.	Punc	Punc	-	5	notconnected

Şekil 3. 42 Sözdizimsel Analiz Hatası Örneği

Şekil 3.24’de sol tarafta ‘Elif giyimine özen gösterir.’ cümlesinin GÖRSOR aracı tarafından üretilen ilişki diyagramı yer almaktadır. ‘Elif’ sözcüğünün bağımlılık ilişkisinin yanlış belirlenmesi sözdiziminin yanlış çıkarılmasına sebep olmuştur. Bu sözcüğün bağımlılık ilişkisinin tipi CLASSIFIER değil SUBJECT olmalıdır.

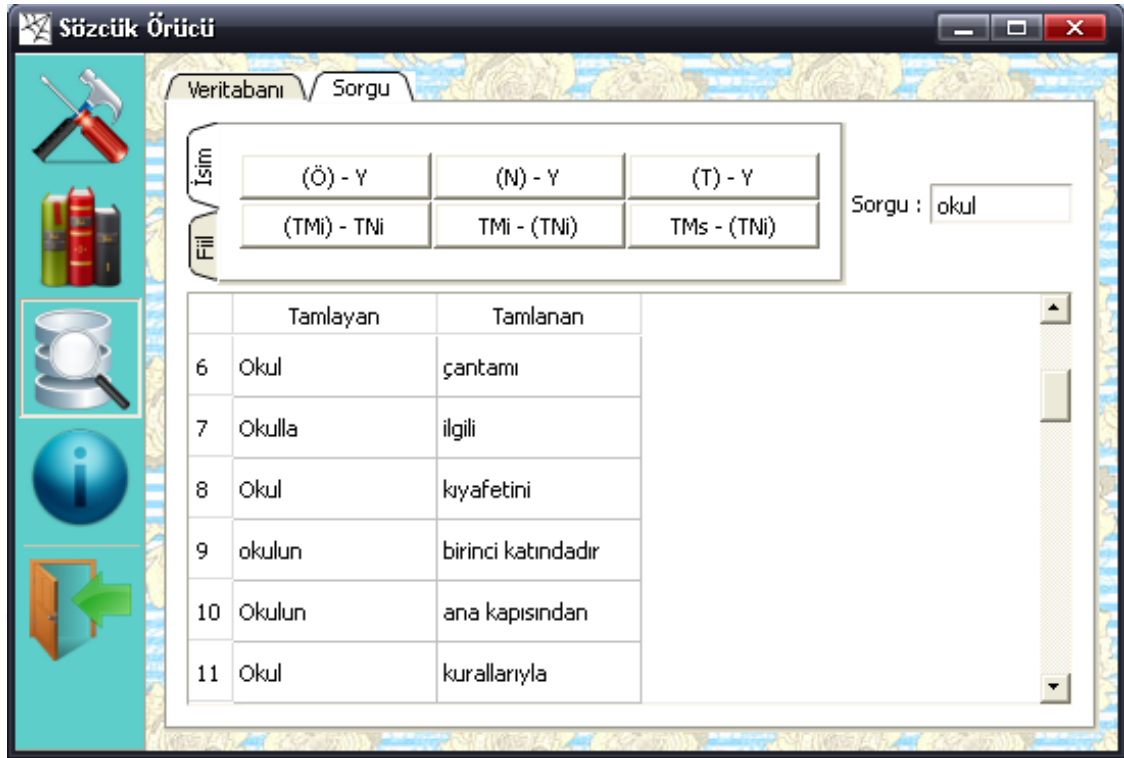


### 3.4 Graf Veritabanı ve Sorgu Modülü

Derlemler üzerinde sözdizimsel sorgular yapılabilmesi için derlemlerdeki cümlelerin ilişki grafları bir graf veritabanına kaydedilmiştir. Bunun için yüksek oranda bağlantılı verileri saklamakta kullanılan bir graf veritabanı olan Neo4j [38] kullanılmıştır. Neo4j, verileri düğümler ve bağlantılar şeklinde saklayan açık kaynaklı bir graf veritabanı olup, düğümlere ve bağlantılara etiket ve özellikler verebilmektedir.

Graf veritabanından sorgulama yapabilmek için bir sorgulama modülü (Şekil-6) yazılmıştır. Bu modülle isim ve fiil kökleri anahtar sözcük olmak üzere aşağıdaki aramalar yapılabilmektedir:

- 1) (Ö) - Y : anahtar sözcüğün özne olduğu cümlelerin fiilleri
- 2) (N) - Y : anahtar sözcüğün nesne olduğu cümlelerin fiilleri
- 3) (T) - Y : anahtar sözcüğün tümleç olduğu cümlelerin fiilleri
- 4) (TMi) - TNi : anahtar sözcüğün tamlayan olduğu isim tamlamalarının tamlananları
- 5) TMi - (TNi) : anahtar sözcüğün tamlanan olduğu isim tamlamaların tamlayanları
- 6) (TMs) - TNi : anahtar sözcüğün tamlanan olduğu sıfat tamlamalarının tamlananları
- 7) Ö - (Y) : anahtar sözcüğün yüklem olduğu cümlelerin özneleri
- 8) N - (Y) : anahtar sözcüğün yüklem olduğu cümlelerin nesneleri
- 9) T - (Y) : anahtar sözcüğün yüklem olduğu cümlelerin tümleçleri



Şekil 3. 44 Sözdizimsel Sorgulama Arayüzü

Şekil 3.26'daki sorgulama arayüzünden öntanımlı sorgular yapılabilmektedir. Çizelge 3.11'de örnek sorgulamalar ve sistemin döndürdüğü cevaplara birer örnek verilmiştir.

Çizelge 3. 12 Sözdizimsel sorgulama örnekleri

Anahtar Sözcük	Sorgulama Türü	Dönen Cevap
okul	(Ö) - Y	Okul bizim ikinci evimiz gibidir.
okul	(N) - Y	Okulumuzu koruyalım.
okul	(T) - Y	Okuldan dönüyordu.
okul	(TMi) - TNi	okul bahçesinin
okul	TMi - (TNi)	harp okuluna
okul	TMs - (TNi)	Atatürk'ün okuduğu okulların
al(mak)	Ö - (Y)	Ayşe teyze alacak
al(mak)	N - (Y)	Sayı boncuğunu aldı.
al(mak)	T - (Y)	Matematik dersi için aldı.

Graf veritabanı servisinin web arayüzündeki sorgulama ekranından Neo4j'in sorgulama dili olan Cypher [39] ile oldukça detaylı sorgular da yapmak mümkündür. Örneğin,

'Öğretmenlerin yaptığı eylemleri (olmak ve var hariç) yapan diğer özneleri bul' sorgusu Cypher dilinde 'MATCH (a)-[:O]-(cumle)-[:Y]-(b), (c)-[:Y]-(cumle2)-[:O]-(d) WHERE a.kok="öğretmen" and NOT b.kok IN ["\_", "ol", "var"] and c.kok=b.kok return b,c,d.kok;' şeklinde yazılabilir. Bu sorguya cevap olarak Çizelge 3.12'deki gibi kayıtlar dönmektedir.

Çizelge 3. 13 Karmaşık sorguya dönen örnek cevaplar

b	c	d.kok
çalışır	çalıştı	Atatürk
çalışır	çalışır	Hakim
kullanmalıdır	kullanırlar	Doktor
...	...	...

Bu sorgu "Şu beş koşulu sağlayan, b ve c sözcüklerinin kendisi ile d sözcüklerinin kökünü bul. 1. a özne olarak, b de yüklem olarak bir cümleye bağlı 2. c özne olarak, d de yüklem olarak bir cümleye bağlı 3. a'nın kökü öğretmen 4. b'nin kökü {olmak,var} kümesinde yok 5. c'nin kökü b'nin kökü ile aynı." şeklinde de okunabilir.

### SONUÇ VE ÖNERİLER

Çalışma boyunca Türkçe bir derlemin sözdizimsel olarak görselleştirilmesi ve derlem üzerinde karmaşık sorgulamalar yapılabilmesi üzerinde çalışılmıştır. Bu amaçla iki ilköğretim hayat bilgisi kitabından, görece basit cümle yapılarına sahip cümleler içeren iki test derlemi oluşturulmuştur.

Derlemlerdeki cümlelerin sözcükleri arasındaki bağımlılık ilişkilerini özetleyen bir ilişki Tablosu programı hazırlanmıştır. Bu program yardımıyla derlemlerdeki etiketlerin doğruluğu takip edilmiştir. Bazı hatalı analizler düzeltilerek derlemlerin yeni versiyonları elde edilmiştir.

Derlemlerdeki cümlelerin yapısal ilişkilerini bir ağaç diyagramı şeklinde görselleştiren bir ilişki diyagramı programı yazılmıştır. Bu programın birinci test derlemi için %90, ikinci test derlemi için de %93 oranında başarı elde ettiği gözlenmiştir. Daha sonra derlemlerdeki tüm yapısal ilişkiler graflar halinde bir graf veritabanına aktarılmıştır.

Görselleştirme ve sorgulama işlemlerinin tek bir arayüzden kontrol edilmesi için bir uygulama hazırlanmıştır. ilişki tablosu ve ilişki diyagramı modülleri bu uygulamaya eklenmiştir. Uygulamaya ayrıca derlem üzerinde sorgulama yapılabilmesi için bir sorgu arayüzü eklenmiştir. Bu sorgu arayüzü ile graf veritabanı üzerinde öntanımlı cümle yapılarının sorgulanabilmesi mümkün kılınmıştır.

Program daha güncel analiz araçlarıyla etiketlenmiş bir derlem olan ODTÜ-SABANCI ağaç yapılı derlemiyle çalıştırıldığında bazı cümleler için hiç, bazı cümleler için kısmi,

bazı cümleler içinse tam olarak diyagramların elde edilebildiği görülmüştür. Bunun birincil nedeni etiketleme sistemindeki farklılıklardır. ODTÜ-SABANCI ağaç yapılı derleminde COLLOCATION VE MWE gibi yeni etiketler mevcuttur. Diyagramı görüntülenebilen cümlelerdeki hatalar ise bazen hatalı etiketlemeler, bazen de oldukça uzun ve karmaşık cümlelerden kaynaklanmaktadır. Uygulamanın bu tip derlemler üzerinde başarılı olabilmesi için karmaşık yapılarıdaki cümleleri oluşturan bileşenler ve bunlar arasındaki ilişkiler tam olarak yakalanmalıdır.

Uygulamanın geliştirilmesi için güncel analiz programlarının takip ettiği etiket sistemlerinin yorumlanmasını sağlayan kodların uygulamaya eklenmesi gerekmektedir. Ayrıca, uygulamanın oldukça karmaşık ve uzun cümlelere kendini adapte edebilmesi için yeni yaklaşımlar geliştirilmelidir.

Ne kadar çok ham veri detaylı dilbilimsel çözülemeye tabi tutulup sözcükler ilişkilendirilirse bu ilişkilerin frekansları gerçek hayattaki doğal değerlerine yaklaşır. Bu ilişkiler daha gerçekçi Doğal Dil İşleme uygulamalarının geliştirilmesine imkan verecektir.

Sözcükler arası ilişkilerle örülmüş büyük bir veri yığınının pratikte bir çok faydası vardır. Bunlara ilk olarak bağlama duyarlı uygulamalar örnek verilebilir. Bir paragraf veya yazının içerisindeki cümlelerde geçen sözcüklerin bağlam modelleme için oluşturulmuş bir graf üzerinde birleştiği yollara yüksek frekanslarla bağlı olan sözcükler, semantik yorumlama sırasında arama alanını daraltmakta kullanılabilir.

Sözcükler arası sözdizimsel ilişkileri içeren graf veritabanlarına karmaşık sorgular yaparak birbirine benzer kavramları gruplamak ontoloji çıkartmaya yardımcı olabilir. Cümlelerde geçen varlıklara ait özellikler ve bu varlıkların sergiledikleri eylemler göz önüne alınarak varlıklar arasındaki benzerlikler ölçüldüğünde kümeleme yöntemiyle, aynı hiyerarşik seviyede olan kavramlar tespit edilebilir. Ayrıca, sözcük ağı oluşturma projelerinde belirli semantik ilişkilerin yakalanmasında graf veritabanlarından faydalanılabilir. Bazı işlevsel (dilbilgisel) sözcüklerin oluşturdukları öbekler sorgulanarak sözcükler arasındaki semantik ilişkiler ortaya çıkartılabilir.

Başka bir uygulama alanı olarak da akıllı sözcük önerileri veren ve yazımdaki bazı mantıksal hataları düzelteren uygulamalar verilebilir. Örneğin, metin girişi içeren bir uygulamada sözcükler girilirken dinamik olarak bu sözcüklerle yüksek frekansta sözdizimsel ilişkiye sahip olan sözcükler uygulama tarafından kullanıcıya önerilerek hem hızlı hem de yerinde kelime seçimleri mümkün kılınabilir. Ayrıca, metinlerdeki dilbilgisel yönden doğru fakat anlamsal yönden gerçekten uzak olan hatalı sözcükler tespit edilip yerine doğru sözcük önerileri verilebilir. Örneğin, 'Güneşi atmayan imparatorluk' ifadesinde geçen atmak fiili imla yönünden doğru olmakla birlikte anlamsal yönden gerçeğe uygun değildir. Güneş varlığının özne olarak geçtiği cümle veya cümleciklerin fiilinin 'atmak' olma olasılığı oldukça düşüktür. Sözdizimsel ilişkileri barındıran derlemlerden üretilen graf veritabanlarına yapılacak sorgulamalarda bu durum tespit edilip, olası eylemlerden bu fiile biçimsel olarak en yakın ve frekans olarak en yüksek olan 'batmak' kelimesi yazım önerisi olarak verilebilir.



## KAYNAKLAR

---

- [1] Sun, Y., Articulate: Creating Meaningful Visualizations from Natural Language, [https://www.evl.uic.edu/documents/sun\\_yiwen\\_articulatethesis\\_3-12.pdf](https://www.evl.uic.edu/documents/sun_yiwen_articulatethesis_3-12.pdf), 03.01.2016
- [2] Wikipedia, Dependency Grammar, [http://en.wikipedia.org/wiki/Dependency\\_grammar](http://en.wikipedia.org/wiki/Dependency_grammar), 29.12.2014
- [3] Carnegie Mellon University, Noah's Ark Research Group, <http://www.ark.cs.cmu.edu/>, 29.12.2014
- [4] Carnegie Mellon University, Tree View, <http://demo.ark.cs.cmu.edu/parse>, 29.12.2014
- [5] Carnegie Mellon University, Turbo Parser, <http://www.ark.cs.cmu.edu/TurboParser/>, 29.12.2014
- [6] SourceForge, Tydevi, <http://tydevi.sourceforge.net/>, 29.12.2014
- [7] Chaoticity, DependenceSee, <http://chaoticity.com/dependensee-a-dependency-parse-visualisation-tool/>, 29.12.2014
- [8] MaltParser, Malt Parser, <http://www.maltparser.org/>, 29.12.2014
- [9] Pompeu Fabra University, MaltDiver, <http://taln.upf.edu/pages/MaltDiver/>, 29.12.2014
- [10] Carnegie Mellon University, ParseViz, <http://www.ark.cs.cmu.edu/parseviz/>, 29.12.2014
- [11] The University of Arizona, TreebankViewer, <http://dingo.sbs.arizona.edu/~sandiway/treebankviewer/>, 29.12.2014
- [12] Ben Podgursky, Nlpviz, <http://nlpviz.bpodgursky.com/home>, 29.12.2014
- [13] Stanford University, CoreNLP, <http://nlp.stanford.edu/software/corenlp.shtml>, 29.12.2014
- [14] Github, Dagre, <https://github.com/cpetitt/dagre>, 29.12.2014
- [15] D3js, D3, <http://d3js.org/>, 29.12.2014

- [16] IronCreek Software, PhpSyntaxTree, <http://ironcreek.net/phpsyntaxtree/>, 29.12.2014
- [17] Nygaard, L. ve Johannessen, J. B. (2004). SearchTree-a user-friendly treebank search interface. In Proceedings of the 3rd Workshop on Treebanks and Linguistic Theories (pp. 183-189)
- [18] GHODKE, S. ve BIRD, S., Fangorn: A System for Querying very large Treebanks. Proceedings of COLING 2012: Demonstration Papers, pages 175–182
- [19] Stuttgart University, TIGERSearch, <http://www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/tigersearch.en.html>, 29.12.2014
- [20] Charles University in Prague, PML-TQ, <http://ufal.mff.cuni.cz/pmltq/>, 29.12.2014
- [21] University of Pennsylvania, Building a large annotated corpus of English: the Penn Treebank, <https://catalog.ldc.upenn.edu/docs/LDC95T7/cl93.html>, 03.01.2016
- [22] Wikipedia, Text Corpus, [http://en.wikipedia.org/wiki/Text\\_corpus](http://en.wikipedia.org/wiki/Text_corpus), 30.11.2014
- [23] Say, B., Zeyrek, D., Oflazer, K. ve Özge, U., 2002. "Development of a corpus and a treebank for present-day written Turkish", Proceedings of The Eleventh International Conference of Turkish Linguistics, s.183-192.
- [24] XCES, Corpus Encoding Standard for XML, <http://www.xces.org/>, 30.11.2014
- [25] Middle East Technical University, METU Turkish Corpus <http://ii.metu.edu.tr/corpus>, 30.11.2014
- [26] Oflazer, K., Say, B., Hakkani-Tür D.K. ve Tür, G., 2003. "Building a Turkish Treebank", Invited chapter in "Building and Exploiting Syntactically-annotated Corpora", Anne Abeille Editor, Kluwer Academic Publishers.
- [27] Sak, H., Güngör, T. ve Saraçlar, M., 2011. "Resources for Turkish morphological processing", Language Resources and Evaluation, 45(2), s.249-261.
- [28] Aksan, Y., 2012. "Construction of the Turkish National Corpus (TNC)", In Proceedings of the Eight International Conference on Language Resources and Evaluation, İstanbul, Türkiye, s. 3223-3227.
- [29] Turkish National Corpus, Türkçe Ulusal Derlemi (TUD), <http://www.tnc.org.tr>, 30.11.2014
- [30] Sezer, B. ve Sezer, T., 2013. "TS Corpus: Herkes için Türkçe Derlem", Proceedings of the 27th National Linguistics Conference, Antalya, Hacettepe University, Linguistics Department, s. 217-225.
- [31] Dalkılıç H. ve Gölge N., 2012. MEB İlköğretim Hayat Bilgisi Ders Kitabı 1
- [32] Özemir A. ve Çınar F., 2012. MEB İlköğretim Hayat Bilgisi Ders Kitabı 2
- [33] İstanbul Teknik Üniversitesi, Turkish NLP Pipeline, <http://web.itu.edu.tr/gulsenc/pipeline.html>, 30.11.2014

- [34] Güngördü, Z. ve Oflazer, K., 1994. "Parsing Turkish using the Lexical-Functional Grammar Formalism", In Proceedings of COLING'94, The 15th Conference on Computational Linguistics, Kyoto, Japan, s. 494–500.
- [35] Sak, H., Güngör, T. ve Saraçlar, M., 2007. "Morphological Disambiguation of Turkish Text with Perceptron Algorithm", In CILing, LNCS 4394, s.107-118.
- [36] Eryiğit, G., Nivre, J. ve Oflazer, K., 2008. "Dependency Parsing of Turkish", Computational Linguistics, 34 (3), 357-389.
- [37] Tilburg University, CoNLL-X Shared Task: Multi-lingual Dependency Parsing, <http://ilk.uvt.nl/conll/>, 30.11.2014
- [38] Neo4j, Graph Database, <http://neo4j.org/>, 30.11.2014
- [39] Neo4j, Cypher Query Language, <http://neo4j.com/developer/cypher-query-language/>, 20.11.2014



## ÖZGEÇMİŞ

---

### KİŞİSEL BİLGİLER

**Adı Soyadı** : Cem AGAN  
**Doğum Tarihi ve Yeri** : 17.07.1985 ÜSKÜDAR  
**Yabancı Dili** : İngilizce  
**E-posta** : [cemagan@hotmail.com](mailto:cemagan@hotmail.com)

### ÖĞRENİM DURUMU

Derece	Alan	Okul/Üniversite	Mezuniyet Yılı
Lisans	Bilgisayar Mühendisliği	Yıldız Teknik Üni.	2011
Lise	Sayısal	Turhan Tayan A.L.	2004

## YAYINLARI

### Makale

1. Agan, C., Diri, B., **“Türkçe Derlemler İçin Sözdizimsel Görselleştirme ve Sorgulama Aracı”**, TBV Bilgisayar Bilimleri ve Mühendisliği Dergisi (Revize edilmektedir.)

### Bildiri

1. Agan, C., Diri, B., **“Aritmetik Bir Problemi Anlama ve Çözme: APAÇ”**, Akıllı Sistemler ve Yenilikler Uygulamaları, ASYU 2012, Trabzon, Türkiye, 2012
2. Agan, Cem, and M. F. Amasyali. "Active learning with committees and the selection of starting sets." Signal Processing and Communications Applications Conference (SIU), 2013 21st. IEEE, 2013.