

**TÜRKÇE TÜMCELERİN YÜKLEM ODAKLI
ANLAM VE DİLBİLGİSİ ÇÖZÜMLEMESİ**

DOKTORA TEZİ

İlknur DÖNMEZ

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

HAZİRAN 2016

**TÜRKÇE TÜMCELERİN YÜKLEM ODAKLI
ANLAM VE DİLBİLGİSİ ÇÖZÜMLEMESİ**

DOKTORA TEZİ

**İlknur DÖNMEZ
(504052508)**

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

Tez Danışmanı: Prof. Dr. Eşref ADALI

HAZİRAN 2016

İTÜ, Fen Bilimleri Enstitüsü'nün 504052508 numaralı Doktora Öğrencisi İlknur DÖNMEZ, ilgili yönetmeliklerin belirlediği gerekli tüm şartları yerine getirdikten sonra hazırladığı “TÜRKÇE TÜMCELERİN YÜKLEM ODAKLI ANLAM VE DİLBİLGİSİ ÇÖZÜMLEMESİ” başlıklı tezini aşağıdaki imzaları olan jüri önünde başarı ile sunmuştur.

Tez Danışmanı : **Prof. Dr. Eşref ADALI**
İstanbul Teknik Üniversitesi

Jüri Üyeleri : **Yrd. Doç. Dr. Cüneyd TANTUĞ**
İstanbul Teknik Üniversitesi

Doç. Dr. Banu DİRİ
Yıldız Teknik Üniversitesi

Yrd. Doç. Dr. Yusuf YASLAN
İstanbul Teknik Üniversitesi

Yrd. Doç. Dr. Arzucan ÖZGÜR
Boğaziçi Üniversitesi

Teslim Tarihi : **8 Haziran 2016**
Savunma Tarihi : **24 Haziran 2016**





Eşime ve çocuklarıma,



ÖNSÖZ

Bilgisayar Mühendisliğimdeki doktoram boyunca çalışmalarımda beni yönlendiren ve insani ve ahlaki değerleri ile örnek edindiğim, Türkçe için çok emek vermiş olan hocam, Eşref Adalı'ya, Türkçe konusunda bize kaynak sağlayan ve bazı incelemelerde bize destek olan Türk Dil Kurumu'ndan Belgin Aksu'ya, emeği geçen tüm hocalarıma ve bu günlere gelmemde büyük pay sahibi olan aileme ve desteğini benden esirgemeyen eşime, çocuklarıma ve dostlarıma teşekkürlerimi sunarım.

Haziran 2016

İlknur DÖNMEZ
(Elektronik ve Haberleşme Mühendisi)



İÇİNDEKİLER

Sayfa

ÖNSÖZ	vii
İÇİNDEKİLER	ix
KISALTMALAR.....	xiii
ÇİZELGE LİSTESİ.....	xv
ŞEKİL LİSTESİ.....	xvii
ÖZET	xix
SUMMARY	xxi
1. GİRİŞ.....	1
1.1 Tezin Amacı.....	1
1.2 Literatür Özeti	1
1.2.1 Dil bilgisi çözümlemesi (Grammar Analyzer)	2
1.2.2 Türkçe anlam analizi çalışmaları.....	4
1.2.3 Diğer anlamsal kaynaklar ve araçlar	5
1.3 Tezin Katkısı.....	8
1.4 Tezin Kullanım Alanları	9
1.4.1 Metin özeti çıkarma (Text summarizer)	9
1.4.2 Bilgi çıkarımı (Knowledge extraction).....	10
1.4.3 Bilgi getirme sistemi (Information retrieval).....	10
1.4.4 Soru cevap sistemleri (Question answering system)	11
1.4.5 Makine çevirisi (Machine translation).....	11
1.4.6 Duygu analizi (Sentiment analysis).....	12
1.4.7 Metin tanılama (Paraphrase identification).....	12
1.4.8 Anlamsal rol etiketleme (Semantic role labelling SRL).....	13
1.4.9 Metin gerektirimlerinin çıkarılması (Recognizing textual entailment) ..	14
1.4.10Belge sınıflandırma (Document classification)	16
1.5 Tezin Kapsamı	16
1.6 Tezde Kullanılan Bazı Temel DDİ Kavramları	17
2. YAPILARINA GÖRE TÜMCELER VE ANLATIM BOZUKLUĞU	21
2.1 Tümceyi Oluşturan Öğeler	21
2.1.1 Yüklem	21
2.1.2 Özne.....	21
2.1.3 Nesne	22
2.1.4 Tümleçler.....	22
2.2 Türkçe’de Yapılarına Göre Tümce Türleri	22
2.2.1 Basit tümce	23
2.2.2 Bileşik tümce	23
2.2.3 Bağlı tümce.....	23

2.3 Yükleme Odaklı Yapısal Çözümleme için Yükleme Gelen Ekler	23
2.3.1 Şahıs ekleri	24
2.3.2 Zaman ve kip ekleri	24
2.3.3 Eylem çatıları.....	24
2.4 Türkçede Anlatım Bozukluğu Olayı	25
2.4.1 Sözcüksel anlatım bozuklukları.....	26
2.4.2 Yapısal ve anlama dayalı anlatım bozuklukları	27
3. TÜMCE ÇÖZÜMLEMEDE KURAMSAL YAKLAŞIMLAR	31
3.1 Tümce Öbeklerinin Çözümlemesi	31
3.2 Tümcenin Kavramsal Olarak Çözümlemesi.....	34
3.2.1 Kavram listelerinin oluşturulması	35
3.2.2 Yüklemin sınıflandırılması	38
3.2.3 Öbek kavram çiftlerinin Verb-Net tematik rolleriyle ilişkisi.....	38
3.3 Tümceleri Alt Tümcelere Ayırma.....	40
3.3.1 İsim eylemlerden alt tümce yapımı	42
3.3.2 Sıfat eylemlerden alt tümce yapımı	43
3.3.3 Zarf eylemlerden alt tümce yapımı.....	44
3.3.4 Alt tümcelere ait yeni yüklem oluşturulması	45
4. YÜKLEM ODAKLI ÇÖZÜMLEME MODELİ	49
4.1 Tümcenin Matris Gösterimi	49
4.2 Ana Model	50
4.3 Çatı Eklerinin Etkisi	53
4.3.1 Geçişli / geçişsiz yüklem	54
4.3.2 İsim kökünden türeyen veya isim yüklemeler.....	54
4.3.3 Oldurgan yüklem	54
4.3.4 Ettirgen yüklem	55
4.3.5 Edilgen yüklem.....	56
4.3.6 Dönüşlü yüklem.....	56
4.4 Yükleme Sonuna Gelen Şahıs Eklerinin Etkisi.....	57
4.5 Yükleme Sonuna Gelen Zaman Eklerinin Etkisi	57
4.6 Uygulama Çıktısı.....	59
4.6.1 Alt tümcelere ayırma örnekleri.....	60
4.6.2 Öbek kavram uyumsuzluğunu yakalama örneği	63
4.6.3 Zaman öbeği ve zaman eki uyumsuzluğunu yakalama örnekleri.....	64
4.6.4 Şahıs öbeği ve şahıs eki uyumsuzluğunu yakalama örnekleri.....	65
5. GELİŞTİRİLEN ARAÇLARIN DEĞERLENDİRİLMESİ.....	67
5.1 Türkçe Anlatım Bozukluğu Olayının İncelenmesi.....	67
5.1.1 Kullanılan kaynaklar ve özellikleri.....	67
5.1.2 Anlatım bozukluğu türleri ve görülme sıklığı	68
5.2 "Alt Tümce Ayırıcının" ve "Öbek Kavram Bulucunun" değerlendirilmesi....	69
5.2.1 "İTÜ DDİ Bağıllık Çözümleyicinin" değerlendirilmesi.....	69
5.2.2 "Alt Tümce Ayırıcının" değerlendirilmesi.....	70
5.2.3 "Öbek Bulucunun" değerlendirilmesi.....	71
5.2.4 "Kavram Bulucunun" değerlendirilmesi	72
5.2.5 Genel değerlendirme	73

5.3 Anlamsal ve Dilbilgisi Çözümleyicinin Değerlendirilmesi.....	73
6. ÖRNEK BİR UYGULAMADA TÜMCE MATRİSİ.....	77
6.1 Temel Model.....	77
6.1.1 Basit tümcenin matris gösterimi.....	79
6.1.2 Sınıflandırma ve özellik seçme algoritmaları	80
6.1.3 Deneysel sonuçlar.....	81
6.1.4 Özellik uzayında tamsayı değişken kullanımı	82
6.2 Belge Sınıflandırmada Tümce Matrisi Kullanımı Sonucu	84
7. SONUÇ VE ÖNERİLER	85
7.1 Tümce Ayırma, Öbek Bulma ve Kavram Bulmaya İlişkin Sonuçlar	85
7.2 Anlamsal ve Dilbilgisel Uyuşum Tespit Sonuçları.....	85
7.3 Kaba Anlamsal Matrisin Belge Sınıflandırmada Kullanımı Sonuçları	86
7.4 İleriki Çalışmalar	86
KAYNAKLAR.....	89
ÖZGEÇMİŞ	98



KISALTMALAR

DDİ	: Doğal Dil İşleme
NLP	: Natural Language Processing
ARE	: Anlamsal Rol Etiketleme
SRL	: Semantic Role Labelling
MGÇ	: Metin Gerektirimlerinin Çıkarılması
RTE	: Recognizing Textual Entailment
GC	: Grammar Checker





ÇİZELGE LİSTESİ

	<u>Sayfa</u>
Çizelge 1.1 : VerbNet vurmak sınıfı-18.1.	6
Çizelge 1.2 : FrameNet rolleri.....	7
Çizelge 1.3 : FrameNet yardımcı roller.....	7
Çizelge 1.4 : Sözcüklerin anlamsal ilişki türleri.	19
Çizelge 3.1 : Türkçedeki hal ekleri ve öbekler arasındaki ilişkilerin biçimsel dil temsili.	31
Çizelge 3.2 : Basit tümcedeki hal ekleri ve öbekler arasındaki ilişkilerin biçimsel dil temsili.	32
Çizelge 3.3 : Bileşik tümcedeki hal ekleri ve öbekler arasındaki ilişkilerin biçimsel dil temsili.	33
Çizelge 3.4 : Tüm tümce çeşitleri için hal ekleri ve öbekler arasındaki ilişkilerin biçimsel dil temsili.	34
Çizelge 3.5 : Analiz aşamasında kullanılan kavramlar.	35
Çizelge 3.6 : Bazı kavram listelerinden örnek elemanlar.....	37
Çizelge 3.7 : VerbNet tematik rollerinin öbek kavram karşılıkları.	39
Çizelge 3.8 : İsim eylemler için alt tümcelerin oluşturulması	42
Çizelge 3.9 : Sıfat eylemler için alt tümcelerin oluşturulması	43
Çizelge 3.10: Zarf eylemler için alt tümcelerin oluşturulması.....	44
Çizelge 3.11: Eylemden yüklem oluşturma değişkenleri.....	45
Çizelge 3.12: Geniş zamanlı ana eyleme sahip tümcede alt eylemin zaman ekinin bulunması.....	46
Çizelge 3.13: Geniş zamanlı ana eyleme sahip tümcede alt eylemin kişi ekinin bulunması.....	47
Çizelge 4.1 : Zaman öbeklerinin sınıflandırılması.	58
Çizelge 4.2 : Zaman eklerinin zaman kümelerine göre uyumluluğu.	58
Çizelge 5.1 : Günlük metin incelemesi.	68
Çizelge 5.2 : Anlatım bozukluğu türleri incelemesi.....	68
Çizelge 5.3 : "İTÜ DDİ Bağlılık Çözümleyicinin" tümce başına doğruluk değerleri.	70
Çizelge 5.4 : Alt tümce ayırma doğruluk değerleri.....	71
Çizelge 5.5 : Öbek bulmada doğruluk değerleri.	71
Çizelge 5.6 : Her bir öbek türü için doğruluk.	72
Çizelge 5.7 : Kavram bulmada doğruluk tablosu.....	73
Çizelge 5.8 : Doğruluk oranları.....	73
Çizelge 5.9 : Bütün tümce için modelin başarımı.	74
Çizelge 5.10: Alt tümceler için modelin başarımı.	75
Çizelge 6.1 : Yüklem bilgisinin vektör gösterimi.	79
Çizelge 6.2 : Sınıflandırma algoritmaları parametreleri.....	81

Çizelge 6.3 : Farklı veri kümeleri ve modeller için başarımlar oranları.....	82
Çizelge 6.4 : Dataset-II'ye özellik seçimi uygulandığında başarımlar oranı.	82
Çizelge 6.5 : Dataset-III'ye özellik seçimi uygulandığında başarımlar oranı.	82
Çizelge 6.6 : DataSet-IV'de yüklem için kullanılan özellikler.	83
Çizelge 6.7 : Dataset-IV başarımlar oranları.....	84



ŞEKİL LİSTESİ

	<u>Sayfa</u>
Şekil 3.1 : Kavram listeleri boyutları.	36
Şekil 3.2 : Alt tümcelerin bulunması.....	41
Şekil 4.1 : Tümcenin matris gösterimi.	50
Şekil 4.2 : Öbek Bulucu.	51
Şekil 4.3 : Kavram Bulucu.	51
Şekil 4.4 : Yükleme sınıf matrisi.	52
Şekil 4.5 : Ana model.....	53
Şekil 4.6 : Oldurganlık eki alan yüklem için Y matrisinin dönüşümü.....	55
Şekil 4.7 : Ettirgenlik eki alan yüklem için Y matrisinin dönüşümü.....	55
Şekil 4.8 : Edilgenlik eki alan yüklem için Y matrisinin dönüşümü.	56
Şekil 4.9 : Dönüşlü eki alan yüklem için Y matrisinin dönüşümü.	57
Şekil 4.10 : Uygulama çıktısı sorunsuz tümce örneği.....	59
Şekil 4.11 : Alt tümceye ayırma örneği (1).....	60
Şekil 4.12 : Alt tümceye ayırma örneği (2).....	61
Şekil 4.13 : Alt tümceye ayırma örneği (3).....	62
Şekil 4.14 : Öbek kavram uyumsuzluğu örneği.	63
Şekil 4.15 : Zaman uyumsuzluğu örneği.	64
Şekil 4.16 : Kişi uyumsuzluğu örneği.	65
Şekil 6.1 : Tümce kaba anlamsal temsil modeli.....	78
Şekil 6.2 : Tümcenin matris temsili.	79



TÜRKÇE TÜMCELERİN YÜKLEM ODAKLI ANLAM VE DİLBİLGİSİ ÇÖZÜMLEMESİ

ÖZET

Çalışmamız tümcelerin anlamsal ve dilbilgisi çözümlemesini içermektedir. Tümcenin anlamsal ve dilbilgisi açısından çözümlemesi Doğal Dil İşleme (DDİ)'nin ana konulardan biridir. Çalışmamızda, çözümleme yapılırken önce metin içindeki tümcelerin her biri basit tümce olacak şekilde alt tümcelere ayrıştırılmaktadır. Her bir alt tümceye ait öbek kavram çiftleri bulunmakta ve daha sonra her bir alt tümcedeki temel dilbilgisi ve anlamsal yanlışları saptamak için yüklemi temel alan yeni bir yöntem önerilmektedir.

Türkçe tümcede yüklem özne ve zaman bilgisi içerir. Ayrıca yüklem, o tümcenin hangi öbeklerden oluşabileceği konusunda da belirleyicidir. Örneğin, “büyümek” yüklemi tümce içinde nesne almazken, “-de” ekiyle biten dolaylı tümleş öbeğini alır. Örneğin “Ayşeyi büyüdü.” tümcesi sorunluyken, “Sokakta büyüdü.” tümcesi doğrudur. Yüklem ayrıca her bir öbeğin içereceği kavram hakkında da bilgi içermektedir. Örneğin “düşünmek” yüklemi insanlara özgüdür. Dolayısıyla özne olarak insan kavramıyla ilişkilidir. “Kapı bugün ne yapacağını düşündü.” tümcesi mantıklı değildir. Bu saptamalardan yola çıkarak çalışmamızda, tümcelerin öbekleri bulunmuş; her bir öbeğin hangi kavramla ilişkili olduğu belirlenmiş ve tümcenin dilbilgisi çözümlemesini ve anlam çözümlemesini yapan bir model tasarlanmıştır.

Çalışmamızda tümceler hal ekleri ve temel öbek yapıları kullanılarak 10 öbeğe ayrıştırılmıştır. Her bir öbeğin içerdiği kavramın 51 kavram sınıfından hangisine ait olduğu belirlenmiştir. Öbek-kavram çiftlerinin yüklemle uyumluluğu araştırılırken çatı, kişi ve zaman ekleri de değerlendirilmiştir.

Çalışmamızda, her öbek kavram bir matris elemanı ile temsil edilmektedir oysa birleşik tümcelerde öbekler içinde iç tümcelere sahip olabilmektedir. Örneğin “okula sevinçle gelen Ayşe” öznesi içinde farklı bir iç tümceyi içermektedir. Anlamsal ve dilbilgisi hatası bu iç tümcelerde bulunabilmektedir. Örneğin eğer özne öbeği “okulda sevinçle gelen Ayşe” olsaydı, gelmek fiili bulunma öbeğiyle uyumlu olmayacak dolayısıyla bu özneyi içeren tümce doğru bir tümce olmayacaktı. Bu nedenle çalışmamız iç tümceleri de içerecek şekilde genişletilmiştir. Bu amaçla tümceler içerdikleri sıfat fiil, zarf fiil yada mastar sayısının alt tümceye bölünmüştür ve ayrılan her bir alt tümce için çözümleme tekrarlanmıştır.

Günümüzde hala pek çok DDİ uygulamasında, tümcelerin içerdiği her bir kelime binler boyutunda temsil edilmekte, farklı kelime sayısına sahip tümcelerin boyutları sabit olmamakta ve tüm bu tümce temsili oldukça ayrı bir yapıya sahip olmaktadır. Çalışmamızda oluşturulan, tümcenin içerdiği öbek kavram türünden sabit uzunluklu, nispeten az boyutlu (10x51) kaba anlamsal matris temsili pek çok anlamsal DDİ çalışmasında kullanılabilecek özelliktedir. Çalışmamızın son bölümünde bu temsilin anlamsal uygulamalarda başarı sağladığı gösterilmiştir. Tümcenin yüzeysel anlamını

içeren matris yapısının son satırına yüklem özellikleri de eklenerek bir döküman sınıflama uygulamasında kullanılmıştır. WEKA paketi ile beş farklı çeşit sınıflandırma algoritması kullanılarak beş ayrı katagorideki dökümanlar sınıflandırılmış sonuçta 145 özelliikle 86.10 başarı elde edilmiştir. Modelimize ait özellikleri eski özelliklere eklediğimizde en yüksek başarı olan 97,12'lik en yüksek başarı değeri elde edilmiştir.

Sonuç olarak bu çalışmamızda tümcenin öbek kavram vektör temsili oluşturulmuş ve tümcenin dil bilgisel ve anlamsal olarak çözümlenmesi için vektör kıyaslanması kullanan yeni bir yöntem sunulmuştur. Bu yöntemle yapısal hatalardan hedeflenen %81,16'lık dilim içinden %64'lük hata tespit edilmiştir. Çalışmamız ayrıca %81,16 başarı ile tümcelerin alt tümcelerinin bulunduğu; %89 başarıyla tümcelerin kendilerinin ve alt tümcelerinin öbeklerine ayrıldığı, %82,8 başarıyla içerdiği kavramların bulunduğu, içerdiği zaman türünün incelenip yüklemle kıyaslandığı, öznesinin tipinin, tekil ya da çoğul olduğunun incelendiği Türkçe tümce çözümleme kaynağı olmak hedefindedir. Türkçenin düzenli tümce yapısı ve düzenli yüklem yapısı bu çalışmanın esin kaynağı olmasına karşın, öbek-kavram temsili tüm diller için kullanılabilecek bir yöntemdir.

GRAMMATICAL AND SEMANTIC ANALYSIS OF TURKISH SENTENCE BASED ON PREDICATE

SUMMARY

The grammatical and semantic analysis of the sentence is one of the main subjects of Natural Language Processing (NLP). In this study, the sentences are separated into their sub sentences, the related phrases and their concepts are found for each sentences and the coarse-grained semantic representation is done for each sub sentences. In this study, we present a novel method to detect basic grammatical and semantic disorders by concentrating on the predicate.

In Turkish, the predicate includes information about the subject and tense. The predicate also helps to identify the phrases which make up the sentence. For example, “*büyümek (to grow)*” does not take an object, but it can take a locative phrase ending with the suffix “-de”. The predicate is also informative about the semantic concept of a phrase. For example “*düşünmek (to think)*” is specifically an action performed by a human, so the subject will be related with the concept of a human. With these properties considered, a model has been designed to find phrases in a sentence, identify their relations to specific concepts, and analyse the sentences grammatically and semantically.

Because of analysing sentences grammatically and semantically, first of all sentence is divided into sub-sentences. The number of sub-sentences depends on the gerunds (verbal nouns), participles (verbal adjectives) and con-verbs (verbal adverbs) in the sentence. A compound sentence may have more than one complex sentence and each complex sentence may have more than one sub-sentence. If the sentence is compound, the first complex sentence is taken and the reminder part is stored. For the complex part the number of the light verbs gives the number of the sub sentences that we want to maintain. For each light verb form and their related phrases the sub sentences are generated with determined rules. After the all sub sentences of the complex sentence are generated, the process goes on from the starting point, first complex sentence of the reminder part is found and algorithm goes on until all sub sentences are found.

Grammatical analysis in our study involves the presence of argument phrases in the sub-sentences. İTÜ NLP dependency parser outputs, case markers in the sentence and formal language representation with phrases that we determined is used to find phrases in a sentence. Then the phrases of the sentence and the concept of each phrases is found. Maintained phrase-concept pairs are checked with predicate according to its compatibility for each sub-sentences.

The grammar checking problem has been studied with the development of the language technologies since the 1970s. Today for English a grammar checker (GC) program can detect various errors, such as agreement in tense, number, word order and in the last ten years GC recognize grammar errors based on the content of the surrounding words. Different rule based, statistical [1], [2], [3] and hybrid [4], [5], [6] methods have been

used for English grammar checking applications. Doğan ve Karaağaç in 1012 [7], İşgüder and Adalı in 2014 [8] and Aygöl [9] analyse Turkish sentences grammatically. There are also text spell error correction studies in Turkish [10], [11]. Despite the efficient GC applications, there are usually too many exceptions in real usage of a natural language.

In our study the sentences and text are represented as condensed vectors or matrices. Condensed vector representation of words, sentences and texts has become crucial because of big data processing issues. In most natural language applications, sparseness is one of the important issues. Vector representation of the words is done via deep learning in 2013 [12]. The distance between the word vectors can show the semantic and syntactic relations between the words. But the best Pearson correlation of the semantic relatedness of word vectors is about 75 % [13]. Meanings of larger units, calculated compositionally is still an issue for NLP and NLP deep learning applications [14].

The focus point of this study is predicates which are seen as relations or functions over arguments by Gottlob Frege [15]. To analyse the "*concept effect*" and the "*phrase effect*" separately different models are formed. In the first model, sentence is separated into phrases. Then the sentence is checked according to the predicate if it can take the phrase or not.

In our second model, the concepts of the phrases are found. Then the sentence is checked according to the predicate if it can take the phrase-concept pair or not. For example if the subject is a dog, it is in the animal concept class, the predicate of the sentence should be in the verb class which is suitable with animal concept class as subject. In this example predicate should not be the predicate "*akmak (to flow)*" which is in liquid concept verb class or the predicate "*düşünmek (to think)*" which is in human concept verb class. It can not be said directly that the concept is not suitable with the predicate.

Compatibility with the predicate can change according to phrase type. For example it is possible to say "*Ali thought the dog.*" but we can not say "*The dog thought.*". The predicate "*to think*" can not take "*dog*" (animal concept) as subject phrase but it can take "*dog*" as object phrase. We can give another example with "*river*", liquid concept. *Dere yavaşça akıyordu.* (*The river flows gently.*), *Dereye düştü.* (*He fell into the river.*), *Balık derede yüzyyor.* (*Fish swims in the river.*). Here "*dere (river)*" (liquid concept) is compatible with "*akmak (to flow)*" as subject phrase, is compatible with "*düşmek (to fall)*" as dative phrase and is compatible with "*yüzmek (to swim)*" as locative phrase. In our study, we represented the sentence as Cartesian product of phrase types and basic concept classes as 10x51 matrix.

In morphologically rich languages, the meaning of a word is strongly affected by the suffixes that are attached to it. Some suffixes and morphological structure give information about meaning [16]. In Turkish, especially the verb takes different types of suffixes [17]. The verb suffixes can affect the phrases that sentence can have and can give information about time, possession and valence. These suffixes are considered in the study. The verb root type also affects the phrases that sentence can have. In Turkish Possessive Suffixes of the predicate and subject of the sentence should be compatible. Tense and Mood Suffixes of the predicate should be compatible with the time phrase in the sentence. The verb valency changing suffix directly effects the phrases that the predicate can take. Verb valency refers to the number of arguments

controlled by a verbal predicate. It plays an important role in a number of the syntactic frameworks that have been developed in the last few decades. Basically VerbNet [18], FrameNet [19] and ProbeBank [20] define their arguments according to predicate. For ten years, concept relation is also studied with verb on Corpus pattern analysis [21].

In our study, decomposition of phrase and concept pairs overlap the roles of VerbNet at some points. For example if "*directs to the X (Turkish dative phrase)*" has location concept it is equalized with "*Goal*" role in VerbNet, if "*away from the X (Turkish ablative phrase)*" has location concept it is equalized with "*Source*" role in VerbNet.

In the coverage of this study, the concepts and verb classes are determined. The basic-level concept categorization depended on the nature of everyday human interaction both in a physical environment and in a culture [22]. In our study, for the concept base sentence analysing part, the concepts are selected from the ontological representations databases like WordNet [23]. We pay attention to meet the roles of sentences like VerbNet via concept selection. The concepts are determined through the guidance of the two points. One of them is good sentence representation which may vary according to the application and the other point is issue of revealing which concepts are determining factor for the predicate compatibility. For filling the related noun phrases lists for each concepts, Balkanet [24], [25] and some special databases and dictionaries [26] are used. Some categorical terms are maintained from the Turkish Dictionary [27].

From the viewpoint of predicate, predicates are categorized according to verb classes. The concept of the noun phrases are directly related with the verb classes. In our study we have 51 concepts for each phrases and 510 verb-classes. For example one of the verb classes has verbs that take time as object phrase and one of the other verb classes has verbs that take location as dative phrase. We had verb lists according to compatibility of each six phrases from Turkish Language Association (TDK) as a trustworthy source for lexical datasets and dictionaries. [28] Verb categorization according to the other phrases and concept is done by ourselves in this study.

When we find phrase-concepts pairs of the sentence, it is represented with matrix. One of the interests of the NLP community is to find representations to process the large amount of unlabelled language data. In our study syntactic and semantic information of a sentence is represented by 10x51 matrix (or 510x1 vector). Due to the concern of data and model visualization some concept classes are combined together under the 51 basic class so the concept space that are considered via the checker get smaller in this study.

The element of the matrix can take the value between 0-1. 1 means the sentence have related phrase-concept pair and 0 means the sentence does not have related phrase-concept pair. "*Ayşe*" is a person as subject, "*kırılan kalemi (broken pen)*" is an object as accusative phrase, "*sevdiği evinden (from her home that she loves)*" is a location as ablative phrase, "*okula (to the school)*" is location as dative phrase and "*sevinçle (with happiness)*" is instrument phrase.

In our model, automatic semantic labelling and grammatical and semantic checking of the sentence according to predicate is done simultaneously. In the model, first of all the sentence is preprocessed via İTÜ Turkish NLP Web Service [29]. Dependency Parsing result of the sentence is the input variable for our grammatical and semantic analyser.

As the basic model, after the sentence is preprocessed via ITU NLP Web Service, the Phrase Finder finds the 10 different phrases in the sentence via using dependency of words and case makers. After the phrases are found, all phrases are categorized according to the concepts. As a result, the matrix representation of the sentence is prepared as X matrix. If X matrix has the phrase concept pair, the related element of the matrix has 1, if X matrix does not have the phrase concept pair, the related element of the matrix has 0 value. On the other hand, verb is searched in the verb classes related with phrase concept pairs and predicate compatibility matrix is prepared as Y matrix. X matrix is generated by the observed phrase concept pairs of sentence and Y matrix shows the capacity of phrases-concept pairs that sentence may have.

The problem exists when the element of Y is equal to 0 but conjugate element of X is 1. It means the verb does not take a concept on that phrase but sentence has the concept. Result is calculated by the function $F = X' \vee Y$.

As a summary of grammatical and semantic analysis part, we divide our model to five parts to see each model's contribution separately. The first model checks the predicate according to phrase that sentence can have. The second model checks the predicate according to phrase-concept pairs that sentence can have. In the third model, valency changing suffixes effect is added onto the second model. In the fourth model, predicate possession suffixes and subject phrases compatibility effect is added onto the third model. And in the last model predicate time suffixes and time phrases compatibility effect is added onto the fourth model. These 5 model are regenerated with the way that the sub sentences are also considered. Determination of the structural error with our new method reached %64 accuracy at sixth model inside the %81.16 part of structural errors. It means we reached the %81.34 success inside the target part of structural errors.

As a result of this study a detailed semantic vector representation of sentence is formed and grammatical and semantic analysis of sentence is done with presented feature vector comparison process. The sentences are separated to its sub sentences with the 81.16% success ratio. The phrases of the sentences are found with the 89% success ratio. The concept of the sentences are found with the 82.8% success ratio. The time class of the sentence is determined and compared with the predicate according to its compatibility, the subject phrase type and singularity/plurality issues are searched. The study is aimed to be a resource for Grammatical and Semantic Analysis of Turkish Sentences and Texts.

Even though the study is done for Turkish, the method of representing the semantic arguments of the sentence with concept phrase pairs can be applied to all languages. Sentence is represented with its phrase-concept pairs as coarse grained semantic matrix. This coarse-grained semantic matrix representation of sentences (texts) can be used as an input for a great deal of semantic applications such as question answering, information extraction and text categorization.

1. GİRİŞ

Tezimiz Doğal Dil İşleme (DDİ) alanında Türkçe tümcelerın dil bilgisi ve anlam çözümlemesini içermektedir. Bilişim ortamlarında geçen "*dil bilgisi çözümlemesi*" sözcüklerin doğruluğı, sırlaması ve kurallar dahilinde sözcük ilişkileriyle ilgiliyken; "*anlamsal çözümleme*" içerilen kavramlar ve kavramsal ilişkilerle ilgilidir. Her ne kadar tümcelerın dil bilgisi çözümlemesi yaklaşık 50 yıldır çalışıyor olsa da metinlerin anlamsal değerlendirme çalışmaları ancak son 10-15 yıldır karşımıza çıkmaya başlamıştır. Türkçe DDİ konusunda ise son yıllarda yapılan anlamsal çözümleme çalışmaları mevcuttur ama henüz bir başlangıç niteliğindedir. Türkçe için yapılan anlamsal çalışmalardaki yetersizliğin nedeni başka diller için var olan *WordNet*, *FrameNet*, *VerbNet* ve *PropBank* gibi altyapıların Türkçe sözcükler için hazırlanmamış olması olabilir. Bu tezin, Türkçe tümcelerın dil bilgisi ve anlam çözümlemesi ile ilgili bir çalışma ve kaynak olarak bu alandaki bir eksikliği karşılaması umulmaktadır.

1.1 Tezin Amacı

Bu tezin amacı Türkçe tümcelerın söz dizim çözümleme sonuçlarını kullanarak sözcüklerin dil bilgisi ve anlam çözümlemesinin yapılması ve tümcenin dil bilgisi ve anlamsal açıdan yükleme odaklanarak uyumlu olup olmadığının bir algoritma aracıyla tespit edilmesidir.

1.2 Literatür Özeti

DDİ, doğal dillerin (insana özgü tüm diller) işlenmesi ve kullanılması amacı ile araştırma yapan bilim dalıdır. Doğal dillerin kurallı yapısının çözümlenerek anlaşılması veya yeniden üretilmesi amacını taşır. 1970'lerden sonra bilişim teknolojilerindeki gelişmeler, bilim insanlarını diller üzerinde çalışmaya teşvik etmiştir. İlk dönemlerde bilgisayar ile konuşmayı amaçlayan bilim insanları zamanla, konuşulanları veya yazılanları anlamaya ve değerlendirmeye çalışmıştır. Bu çalışmaların sonunda, bir sistemin bir konuşma veya yazıyı anlaması, sorulana yanıt

vermesi ve diller arası çeviri yapabilmesi hedeflenmektedir. Bu hedeflere ulaşabilmek için dillerin ses, biçim, dizilim, anlam gibi temel özellikleri incelenmektedir [30]. Çalışmamız hem dil bilgisi açısından hem de anlamsal açıdan tümceyi çözümlemeyi kapsadığı için bu konularda yapılan önceki çalışmaları iki alt kümeye ayırmakta fayda vardır.

1.2.1 Dil bilgisi çözümlemesi (Grammar Analyzer)

DDİ alanında metindeki dil bilgisi hatalarının denetimi iki kümede toplanabilir. İlk küme sadece sözcüklerin doğru yazımıyla ilgilenen çalışmaları içerir. İkinci küme ise yazımın doğruluğunun yanında tümce içi bağımlılık ilişkilerinin ve dizilimin dil kurallarına uyumluluğunu araştırmaktadır.

Yazım hatalarının giderilmesi çalışmaları: Yazma işleminin bilgisayar ortamına geçmesiyle birlikte bilgisayar ortamında bulunan bir metnin yazım hatalarını bulmak ve düzeltmek bilişimciler için bir araştırma alanı olmuştur. Metindeki yazım hatalarını bulmak için değişik yöntemler kullanılabilir. Yöntemlerden biri, metninlerin yazıldığı dilin sözlüğünün kullanımı olabilir. Bu yöntem bitişken olmayan diller için uygun olsa da Türkçe gibi bitişken diller için ilk seçenek değildir. Türkçe gibi kurallı bir dille yazılmış bir metin için ses uyum kuralları, hece yapıları, eklerin uyumu ve tümce yapısı gibi özelliklerin kullanılması önceliklidir.

Türkçe metinlerde yazım hatalarını bulmak için değişik yöntemler kullanılabilir. Bu yöntemler sözcük temelli olabileceği gibi kural tabanlı da olabilir. A. Solak ve K. Oflazer tarafından hazırlanmış olan Türkçe yazım denetimi çalışması sözlük temelli bir çalışmadır [31]. Bu çalışmada 4 bayrakla işaretlenmiş 23.000 kök sözcük kullanılmaktadır. Metin içindeki sözcükler sondan ekleri çıkarılarak bu sözlükte aranmaktadır. Sondaki harfler çıkarılırken ünsüz yumuşaması ve ses düşmesi de dikkate alınmıştır. Ayrıca ekler sınıflandırılmış ve çözümlemenin daha hızlı olması sağlanmıştır. Bu çalışma tek tek sözcükleri çözümlemekte tümce bazında denetleme yapmamaktadır.

Türkçede yazım hatalarının düzeltilmesi ile ilgili diğer bir çalışma K. Oflazer tarafından yapılmıştır [32]. Türkçenin ek zengini olduğu bir gerçektir, bu nedenle metinlerin daha yetkin bir şekilde çözümlenebilmesi hatalarının tespit edilebilmesi için avantaj sağlamaktadır. Eklerin sözcük kök veya gövdesine eklenişlerinde

oluşturdukları desenlerin incelenmesi biçim bilimsel çözümlemenin konusudur. Türkçe dil bilgisi kurallarına göre yapım ekleri çekim eklerinden önce köke eklenmektedir. Dolayısıyla bir kez çekim eki alan sözcük, daha sonra sadece çekim eki alabilmektedir. Çekim eklerinin türüne göre sırası da kurallara bağlıdır. Yapım eklerinde ise farklı sıralardaki ekler farklı anlamlar oluşturabilmektedir. Örneğin farklı sıradaki aynı yapım ekleri “Pazarlıkçı” ve “Pazarcılık” gibi birbirinden farklı anlama gelen iki farklı sözcük oluşturabilmektedir. Türkçede sözcük sonuna gelen yapım ve çekim ekleri sesli ve sessiz harf kurallarına uyum gösterir. Biçimbilimsel çözümlemede tüm bu kurallar dikkate alınır. Bu çalışmada biçimbilimsel çözümleyici ve dinamik programlama tabanlı bir arama algoritması kullanılmaktadır. Bu çalışmada öncelikle yanlış yazılmış sözcük için tüm aday kökler bulunmuş daha sonra “*edit distance metric*” denen ölçüm kullanılarak tüm aday köklerden olası tüm sözcükler türetilmiştir. İki sözcüğün birbirine benzerliğinin bulunmasında “*q gram*” yöntemi kullanılmıştır.

Türkçe yanlış yazılmış sözcüklerin bulunması ile ilgili bir başka çalışmada sözcük bazlı denetim yapılmıştır. Sistem Türkçe metindeki sözcükleri giriş olarak almakta ve istatistiksel yöntemlerden n-gram sıklıklarına bakarak sözcüğün doğru yazılıp yazılmadığına karar vermektedir [33]. Çok büyük derleme çalışan bu sistem eğer bir sözcüğün olasılık dağılımı 0 veya sıfıra çok yakın ise o sözcüğü yanlış yazılmış olarak işaretlemektedir. Ayrıca Türkçe tümcelerde sözcükleri ek alarak son harflerini değiştirmelerine rağmen bulup değiştirebilen bir program Öztürk ve Adalı tarafından 2009’da tasarlanmıştır [34].

Türkçe yapısal hataların incelenmesi çalışmaları: Dil bilgisi çözümleme, yazımın doğruluğunun yanında tümce içi bağıllık ilişkilerinin dil kurallarına uyumluluğunu da araştırmaktadır. Türkçe tümcelerin dil bilgisi açısından çözümlenmesi ve öğelerinin incelenmesi konusunda 2012’de Doğan ve Karaağaç [7], 2014’te İşgüder ve Adalı [8] ve Aygül’ün [9] çalışmaları bulunmaktadır. Bu çalışmalar kapsamında tümceler söz kümelerine bölünmekte ve sözcüklere eklenen hal ekleri incelenmektedir. Türkçe metinlerdeki yazım yanlışlarını otomatik bulmaya yönelik çalışmaların varlığı da bilinmektedir [10], [11]. Bu iki çalışmada istatistiksel yöntemler kullanılmıştır.

Bu çalışmalardan 2012’deki Doğan ve Karaağaç’a ait olan dil bilgisel bağdaşmazlık ve anlamsal tutarsızlık adı altında öbek ve yüklemden kaynaklı uyumsuzlukları ve

bazı eklerin kullanımının getirdiği sorunları incelemiş ve sıralamıştır. 2014'te İşgüder ve Adalı çalışmalarında öbeklerin kümelenmesi konusunda çalışmıştır. 2014'te Aygül'ün yaptığı çalışmada koşullu rastgele alanlarla tümce öbekleri bulunmaktadır. Litaratürde yüklemi öbek kavram uyumuna göre inceleyen bir çalışma görülmemiştir.

1.2.2 Türkçe anlam analizi çalışmaları

Türkçe için son 10 yıl içinde tümce içindeki sözcükler arasındaki kavram ilişkilerini inceleyen çalışmalar mevcuttur. 2006'da Güngör ve arkadaşları Türkçe dil elemanlarının anlamsal kümelenmesi üzerine bir çalışma yapmıştır [35]. 2007'de Onur ve Tolga Güngör tarafından yapılan bir çalışma Türkçe bir sözlükteki tanımlardan kavramlar arasındaki üst-kavram ilişkilerinin çıkarılması ile ilgilidir [27]. 2007'deki bir başka çalışmada bazı zaman ve kip eklerinin getirdiği çok anlamlılık üzerinedir [36], 2011'de Şerbetçi ve arkadaşları [37] ve paralel olarak Amasyalı ve arkadaşları [38] Türkçe anlamsal ilişkileri bulmak üzerine bir çalışma yapmıştır. 2012'deki bir çalışma Türkçe alt anlam üst anlam ilişkilerinin otomatik çıkarılması ile ilgilidir [39]. 2013'te Fırat ve arkadaşlarının yaptığı çalışmada biçimsel bir ontoloji oluşturulmaktadır [40]. Yine 2013'te Yıldız ve arkadaşları Türkçe'deki parça bütün ilişkilerinin otomatik bulunması ile ilgili bir çalışma gerçekleştirmiştir [41]. Şahin ve arkadaşları tarafından 2015'te yapılan çalışma örgü ve anlamsal benzerlikleri kullanarak Türkçe alt anlam üst anlam ilişkilerini bulmaktadır [42].

2007'den bu yana anlamsal çözümleme uygulamalarda kullanılmaya başlamıştır. 2007'de yapılan bir çalışmada döküman sınıflandırırken vektör model benzerliği yerine wordnet kavram yakınlığını kullanan bir benzerlik yöntemi kullanmışlardır [43]. 2010'da gizli anlamsal inceleme (Latent Semantic Analysis) içeren belge özetleme çalışması [44]. 2012'de Twiter'daki Etkinlikleri Tespit eden uygulamada semantik benzerlikler sözcük ve sözcük dizisi benzerliklerinden faydalanmaktadır [45]. 2016'da anlamsal bilgi çıkarım sistemi kullanılarak iş başvurusunda bulunanlar ve işleri eşleştiren bir sistem tasarlanmıştır. Anlamsal ilişkilendirme için kavramsal benzerlik kullanılmıştır [46]. Bunun dışında soru cevap sistemleri üzerine kural tabanlı bir çalışma [47] ve Türkçe sözcüklerin anlamsal benzerliklerinin ölçülmesi ve metin sınıflandırmada kullanılması ile ilgili çalışma [48] mevcuttur.

Türkçe için 2005'lerden itibaren anlamsal alanda pek çok çalışma kavram ilişkileri çıkarma, alt anlam üst anlam bulma ya da ontoloji oluşturma ve anlamsal bilginin farklı DDİ uygulamalarında kullanılması şeklindedir. İngilizce için yapılan binlerce çalışmayla kıyaslandığında yeterince anlamsal çalışma yapılmamış olduğu daha iyi anlaşılır. Bu çalışmaların İngilizce'deki kadar yaygın olmamasının sebebi İngilizce anlamsal çalışmalarda sıklıkla kullanılan WordNet, FrameNet ve VerbNet gibi kaynak ve araçların henüz Türkçe için oluşturulmamış olması olabilir.

1.2.3 Diğer anlamsal kaynaklar ve araçlar

Özellikle İngilizce anlamsal araştırmalarda sıklıkla karşımıza çıkan bazı kaynaklar incelenmiştir.

WordNet, George Miller ve takımı tarafından Princeton Üniversitesinde tasarlanmış, uygulanabilir ilişkileri gösteren bir anlamsal ağ yapısından oluşan büyük boyutlu veri tabanıdır [23]. Kavramsal bir sözlük gibi düşünebileceğimiz WordNet'de isimler, eylemler, sıfatlar ve zarflar eş anlamlılarıyla birlikte kümelenmiştir (synset). Sözcükler ve dolayısıyla eş anlamlı sözcük kümeleri birbirine alt anlam / üst anlam, içerme / parçası olma, zıt anlam ve gerektirme gibi anlamsal ilişkilerle bağlanmıştır. Sonuçta bu ağ yapısı bir sözcüğün anlamını vermekten çok diğer anlam kümeleri ile ilişkisi ve hangi anlam kümesi içinde olduğu bilgisini vermektedir. 1991'de WordNet ile tüm DDİ alanlarında yaygın olarak kullanılan İngilizcenin anlamsal sözcük ağı oluşturulmuştur [49]. BalkaNet [24], [25] projesi WordNet'in 2004'te Türkçe'ye uyarlanmaya çalışılmış halidir. İngilizce sözcüklere karşılık düşen Türkçe sözcükler ve bunların alt küme, üst küme ilişkileri büyük oranda oluşturulmuş ama tamamlanmamıştır. Referans olarak İngilizceden başlandığı için Türkçeye uymayan durumlar da gözlenmiştir.

VerbNet, 2000 yıllarının başında Kipper, Dang, ve Palmer tarafından oluşturulan, teorik Levin (1993) eylem sınıfları alt yapısını kullanan İngilizce için hazırlanmış anlamsal rol sözlüğüdür [18]. VerbNet'de, eylemler sıra düzensel sınıflar olarak kümelendirilmiştir. Her bir sınıfın taşıdığı anlamsal özellikler söz dizimsel bir şekilde eşleştirilmiştir. FrameNet'e göre daha güçlü çıkarımlar yapmayı sağlar: Eylemlerin sadece anlamsal özellikleri değil söz dizimsel yapısı hakkında da bilgi verir. VerbNet, aktör, zaman, faydalanan, sebep gibi otuz altı tane temel anlamsal rol etiketi kullanır.

Çizelge 1.1 VerbNet'in kullandığı anlamsal rol etiketlerini örnek bir eylem sınıfı üzerinden göstermektedir.

Çizelge 1.1 : VerbNet vurmak sınıfı-18.1.

Sınıf Yapısı: İş yapan[+kontrol eden] Etkilenen[+somut] Araç[+somut]			
Sınıfa üye sözcükler: vurmak, yumruk atmak, çarpmak, tekmelemek, tartaklamak.			
Çerçeve:			
Tipi	Örnek	Yazım	Anlam İlişkileri
Nesne Alır	Ali	Özne	Sebepl olmak(Özne,Olay)
	topa	Nesne	Tavır(süresince(Olay), YönelmişHareket, Özne)
	vurdu.	Yüklem	Temas(bitti(Olay), Özne, Nesne)...

Çizelge 1.1’de görüldüğü üzere “*Vurmak (Hit)*” sınıfı üç rol almaktadır. Bunlar, eylemi yapan, eylemden etkilenen ve eylemde kullanılan araç rolleridir. Bu sınıf tekmelemek, vurmak, ateş etmek gibi sözcükleri içermektedir. Bu sınıftaki eylemler geçişlidir. Bu sınıftaki eylemler anlamsal olarak durum değişikliği meydana getirir. Bir temas vardır. Yönelmiş bir hareket vardır.

FrameNet anlamsal kalıplar ve bu kalıplar arası tanımlanan anlamsal ilişkiler topluluğudur. Bu kalıp elemanları çok iyi tanımlanmış anlamsal roller olarak kabul edilir. Anlamsal kalıplar birbirine bazı ilişki kümeleriyle bağlanmıştır. “*miras almak*” ve “*kullanmak*” örnek iki ilişkidir. Raporlamak kalıbı haberleşmek kalıbından tek yönlü olarak miras almaktadır, bir başka değişle haberleşme kalıbına ait özellikler raporlama kalıbı için de geçerlidir ama tersi doğru değildir. Bu şekilde FrameNet, gerektirme kuralları ve çıkarsama ontolojisi oluşturmaktadır [19]. Verilen sözcük birimine ait tümcelerın parçalara ayrılarak, her bir parçanın veya çerçeve elemanının anlamsal ve söz dizimsel özelliklerinin örneklenmesidir. Örnek vermek gerekirse;

“*O asla erkek kardeşi hakkında konuşmaz. (He never spoke about his brother.)*”

Çerçeve (Frame): Konuşmak (The Speaking)

Sözcük Birimi (Lexical Unit): konuş.eylem (spoke.verb)

Çerçeve Elemanı (Frame Element -partition1): konuşan, O (speaker, He)

Çerçeve Elemanı (Frame Element -partition2): konuşulan konu, erkek kardeşi (topic, about his brother)

Çerçeveler sözcük birimlerin anlamları ile tanımlanırlar. “*the hot of temperature*” ve “*the hot of taste*” “*hot*” sıfatı ile birlikte kullanılmış sözcük birimleridir. Bu iki

sözcük birimi anlamlarındaki farklılıklarından dolayı iki ayrı çerçeveyi çağrışırlar. Bu çerçeveler “*Temperature Scale*” ve “*Taste Experience*” çerçeveleridir.

FrameNet’de kullanılan ana tematik roller ve sadece ontolojik olarak gerekli, fakat merkezi roller gibi sipesifik olmayan çevresel veya temel olmayan (non-core) tematik roller vardır. Bu roller Çizelge 1.2 ve Çizelge 1.3’te görülmektedir.

Çizelge 1.2 : FrameNet rolleri.

Ana Roller	Örnek
Etken özne (Actor)	Farkındalık olmadan yapan. Taş camı kırdı.
Olaydan etkilenen (Actee)	Taş camı kırdı tümcesinde cam Actee rolündedir.
Neden olan (Causer)	Eyleme neden olan.
Deneyimleyen (Experiencer)	Öznenin eylemi yaparken farkında olmasıdır.
Farkında özne (Agent)	Bir eyleme sebep olan etken öznedir.
Farkında nesne (Patient)	Olaydan etkilenen. John Mary ’i öldürdü.

Çizelge 1.3 : FrameNet yardımcı roller.

Yardımcı Roller	
Kaynak (Source)	“[The girl GOAL] received a book from [the boy SOURCE].”
Hedef (Goal)	“[The girl GOAL] received a book from [the boy SOURCE].”
Yol (Path)	“John rolled the ball [down the hill PATH].”
Araç (Instrument)	“John cut his hair [with a knife INSTRUMENT].”
Konum (Location)	“John is [at home LOCATION].”

PropBank, FrameNet’den bir kaç açıdan farklıdır. PropBank, eylem odaklı bir anlamsal sözlüktür. FrameNet ise yüklem, isim ve diğer sözcüklerin tanımlamalarının genelleştirilmiş hali olan kalıpları kullanır. PropBank olayları ve isimlerin kullanıldığı halleri yorumlamaz. PropBank sözlükteki tüm eylemleri karşılama sözü verir, FrameNet sadece büyük bir derlem içinden bir küme örnekleme elde edilmiş eylemleri kapsar.

PropBank tarzı bir yorumlama daha çok sözdizimsel yaklaşıma yakındır. FrameNet tarzı yorumlama isim ve eylemlerin tümünü kapsayan anlamsal çözümlemeye daha yakındır. Başlangıçta, PropBank makine öğrenmesi ile anlamsal rol etiketi atamak amacıyla bir eğitim verisi olarak geliştirilmiştir.

Aşağıda bazı eylemler için PropBank’deki rol kümeleri örneklendirilmiştir.

1- “ O çantasını 100 sente Alice’e satmayı önerdi (He offered Alice to sell his bag for 100 cent.)” “Offer” eylemi için propbank’deki rol kümesi verilmiştir.

Arg0: öneren varlık, o (entity offering)

Arg1: önerilen şey, çanta (commodity)

Arg2: bedel, 100 sent (price)

Arg3: faydalanan, Alice (benefactive or entity offered to)

2-“*Başkan Bush belli tipteki saatler için gümrük kaldırma düzenlemesini onayladı. (President Bush has approved duty-free treatment for imports of certain types of watches.)*” “approved” eylemi için tümce propbank’dan faydalanılarak aşağıdaki rollere ayrıştırılmıştır.

Arg0: onaylayan varlık, başkan Bush (entity approving)

Arg1: önerilen şey, belli türdeki saatler için gümrük kaldırma düzenlemesi (entity approved)

SemLink; PropBank, VerbNet ve FrameNet’in ortak kullanımını sağlayan bir anlamsal sözlükler birleşimidir. İngilizce ve diğer diller için uluslararası ve yerel pek çok sözlük çalışması daha mevcuttur. Bu anlamsal sözlükler halen güncellenmektedir.

1.3 Tezin Katkısı

Günümüzde kullanılmakta olan dil bilgisi çözümleme programları ve yazım programları öbek kavram çiftlerinin yüklemle uyuşup uyuşmadığını denetlememektedir. Girişlerimizi oluşturan DDİ Bağlılık Çözümleyicinin doğru çalıştığı tümceler için bu çalışma yapısal hataların hedeflenen diliminin %81,34’ünü bulmada katkı sağlamaktadır. DDİ Bağlılık Çözümleyicinin doğruluğunun artmasıyla birlikte tüm uygulamalarda doğrudan kullanılabilir.

Tezin diğer önemli katkısı tümceleri matris şeklinde temsil edilmesini sağlayan bir kaynak oluşturmaktır. 2013’e kadar sözcükler sözlük sayısı kadar eleman boyutunda temsil edilmekte ve sadece ait olduğu sözcük 1 diğer tüm sözcükler 0 olarak gösterilmekteydi. Çok seyrek ve büyük boyutlu verilerle çalışmak DDİ alanının en büyük sıkıntılarından biridir. 2013’te derin öğrenme yöntemleri kullanılarak sözcüklerin genellikle 10 ile 500 arasında boyutla temsil edildiği sıkıştırılmış vektör temsili bu ihtiyaç için bir cevap niteliğindedir [12]. Bu yöntemle sözcük vektörleri arasındaki uzaklığa bakılarak sözcük benzerlikleri tahmin edilmeye çalışılmaktadır.

Sözcükler sıkıştırılmış vektörlere dönüştürüldüğünde; sözcük kümeleri arasındaki anlamsal benzerliğin nasıl değerlendirileceği, hesaplanacağı DDİ alanında ve DDİ derin öğrenme alanında çözüm bekleyen bir konudur [14]. Çalışmamızda on tür öbekten oluşan basit tümceler 510 boyutlu vektörler olarak gösterilmektedir ve yüzysel bir şekilde anlam bilgisini içermektedir. Tümcenin kaba anlam içeriğini sağlayan böyle bir vektör pek çok DDİ alanı için kaynak niteliğindedir.

Tezimiz ayrıca tümceleri alt tümcelerine ayıran, tümcelerin on çeşit öbeğini ve bu öbeklere ait ellibir kavramı bulan Türkçe için bir kaynak olma hedefindedir.

1.4 Tezin Kullanım Alanları

Tümcelerinin bazı öge bilgileri, Tümcenin kaba anlam içeriği ve öbeklerin içerdiği kavram; metin özeti çıkarma, soru cevap sistemleri, bilgi çıkarımı, belge sınıflama ve bilgi getirimi gibi pek çok DDİ alanlarında ihtiyaç ve iyileştirme niteliğindedir. Tezimizde tümcelerin öbekleri bulunmakta ve her bir öbeğin içerdiği kavramın belirlenmektedir. Sonuçta her tümce öbek kavram çiftleriyle matris biçiminde temsil edilmektedir. Bu özellikleriyle tezimiz aşağıda özetlenen pek çok DDİ alanında giriş verisi olarak kullanılabilecek bir kaynaktır.

1.4.1 Metin özeti çıkarma (Text summarizer)

Özetleme DDİ alanındaki en zorlayıcı işlerdendir. Çünkü metinler dilin zengin anlamsal yapısına bağlıdır ve kişiye özgü yorumlar içerir. Bu alanda pek çok çalışma mevcuttur. Anlamsal yöntemler kullanılarak 2008’de yapılan bir çalışmada metin tümcelere ayrılır, her bir tümcenin bir değer fonksiyonu vardır. Bu değer fonksiyonu belli özelliklerin ağırlıklandırılmış toplamıdır. WordNet’den faydalanılarak tümcelerin anlamsal benzerlikleri bulunur. Benzer tümcelerin çokluğuna ve değer fonksiyonuna bakılarak özet metin oluşturulur [50]. Bunun dışında sık kullanılan sözcük ve tümce sıklıklarından faydalanılan istatistiksel yöntemler de metin özeti çıkarımında sıklıkla karşılaşılmaktadır [51]. Bir metnin içerdiği tümcelerde yüklemnin hangi kavramla ilişkili olduğu, olayın nerede geçtiği, ne zaman olduğu, hangi kavramlara en çok yönelme olduğu; hangi kavramların kaynak niteliğinde olduğu tümce matrislerimiz içinde bulunacağından, yaptığımız çalışma metin özeti çıkarma alanında fayda sağlayacaktır.

1.4.2 Bilgi çıkarımı (Knowledge extraction)

Düzenli (XML verisi gibi) ya da düzensiz veriden istenen bilginin çıkarılmasıdır. Elde edilen bilginin bir makine tarafından okunabilir ve yorumlanabilir olması gerekir. Bu nedenle bilginin anlaşılmayı kolaylaştıran bir şekilde temsil edilmesi gerekir. Bu düzenli kayıtlı bilgi tanımlayıcıları ve kavram ilişkileri ile yeniden kullanılabilir veya şema tabanlı olarak üretilebilir olmalıdır. 2003'te yapılan bir çalışmada FramNet'deki yapı temel alınarak rol atama teknikleri ile düzensiz veriden çıkarım yapılmaktadır [52]. Yine 2003'te yapılan bir çalışmada istatistiksel yöntem ek olarak ontoloji üreten sözcük temelli terim açılımı kullanılmıştır [53]. Çalışmamızda sözcükler kavramlarla eşleştirildiği için benzer kavramlara sahip sözcüklerin kümelenmesi ve daha düzenli bilgi çıkarımı konusunda tezin katkısı olacağı düşünülmektedir.

1.4.3 Bilgi getirme sistemi (Information retrieval)

İnternetin yaygınlaşmasıyla daha da büyüyen veri havuzundaki bilginin çıkarılması günümüzün en güncel konularından biri olmuş ve bu bağlamda birçok İnternet arama motoru geliştirilmiştir. Ancak, arama motorları yazarların metinleri için belirttiği anahtar sözcükler ile kullanıcının aramada kullandığı anahtar sözcüklerinin çakışmamasından doğan etkin erişim problemiyle karşı karşıya kalmışlardır. Doğru bilgiyi, belgeyi getirmek için metin anahtar sözcükleriyle kullanıcı anahtar sözcükleri arasındaki anlamsal ilişkideki dikkate alan (eş anlam, zıt anlam) ya da aranan dökümanları anlamsal olarak da inceleyen çalışmalar yaygınlaşmaya başlamıştır. 2007'de IADIS Uluslararası Konferansında yayınlanan makale rol atama tekniğinin Bilgi Getirme Sistemleri alanında kullanımı ile ilgilidir. Bu makalede üzerinde arama yapılan belgeler önce tümcelere sonra rol atama uygulaması da kullanılarak özne, nesne ve etkilenen kişi diye üç bölüme ayrıştırılmıştır. Sözcük benzerliği ile geri getirme yöntemlerine göre ortalama doğru öngörü değeri yüzde 55'lerden yüzde 95'lere çıkmıştır. Anlamsal rol ataması kısmında PropBank kullanılmıştır [54]. Çalışmamız bu alanda hedef belgelerin içerdiği tümcelerin on çeşit öbeğe ve bu öbeklerin içerdikleri kavramlara göre daha detaylı aranmasını ve daha iyi sonuç alınmasını sağlayabilir.

1.4.4 Soru cevap sistemleri (Question answering system)

Soru Cevap sistemleri, DDİ ile üretilen bir sorunun cevabını bir algoritma yardımı ile araştırma, sentezleme ve cevabı üretmeyi amaçlar. Son yıllarda dünya da DDİ ve bilgi çıkarımı gibi konularda olan gelişmeler, ülkemizde soru cevaplama sistemlerine olan ilgiyi de arttırmıştır, öyle ki, müşteri hizmetleri gibi yoğun iletişim ağlarına sahip kurumlar bilgilendirme amaçlı ya da yarı-otomatik çevrim içi soru-cevap sistemleri oluşturulmaya başlanmıştır [47]. 2010'da yapılan bir çalışmada WordNet, EventNet gibi yapılar kullanılmış ve anlamsal bilginin soru cevap sistemleri üzerinde nasıl bir etkisi olduğu araştırılmıştır. Bu çalışmada metindeki çoklu anlamsal bilginin bir arada gösterilebildiği Lymba formatı kullanılmıştır [55]. Bir başka çalışmada anlamsal rol atama bilgisinin soru cevap sistemlerinin doğruluğuna katkısı olup olmadığı araştırılmıştır. Çalışmada FrameNet'deki anlamsal roller kullanılmıştır. Anlamsal rol atamaları biparite ağağdaki optimizasyon problemi kullanılarak yorumlanmıştır [56]. Yine başka bir çalışmada anlamsal rol atamanın Soru cevap sistemlerinde kullanıldığında doğruluk artışına rağmen performanstaki düşüşe vurgu yapılmakta ve anlamsal rol atamasının daha az performans yükü ile soru cevap sistemlerine nasıl uygulanacağına dair bir çözüm üretilmektedir. Çalışma, anlamsal rollerin atamasında FrameNet'e benzer ama soru cevap sistemleri için özelleştirilmiş bir anlamsal sözlük kullanmaktadır [57]. Tezimizde oluşturduğumuz tümcenin kaba anlamsal temsili özellikle soru cevap sistemleri için uygun bir kaynaktır. Öncelikle sorunun çözümlenmesinde faydalıdır. Sorunun özne ile ilgili mi, nesne, kaynak, hedef, edat ile ilgili mi olduğunu anlamayı sağlar. Her bir öbek kavramın ne olduğu belirlendikten sonra istenen soruya gerekli cevap'da eklenerek daha doğru cevap oluşturmak sağlanabilir.

1.4.5 Makine çevirisi (Machine translation)

Makine Çevirisi, bir dilden diğer dile dönüşüm yapmayı amaçlayan sistemdir. Burada esas olan diller arasındaki tümce yapısı bilgisi, ek bilgisi ve dil bilgisine dayalı bilgilerin doğru tanınması ve kayba uğramadan dönüştürülmesidir. Makina çevirisi kural tabanlı yöntemlerle yapılmaya başlanmıştır. 1994'te elektronik verinin artması ile birlikte IBM'in Candide sistemi kural tabanlı yaklaşımlardan daha iyi sonuç alınca,

makine çevirisinde istatistiksel yöntemlerin üstünlüğü başlamıştır [58]. Hesaplama gücünün artması ve paralel dil verisine ulaşımın kolaylaşması araştırmacıların bu alana olan eğilimlerini arttırmıştır. İstatistiksel yöntemlerle makine çevirisinde yapı bakımından benzer diller arasında çok büyük başarımlar elde edilmiştir. Buna rağmen sözcük dizilimleri ve yapıları (bükümlü, eklemeli vb) birbirinden farklı olan dil ikililerinde ve çok nadir karşılaşılan sözcüklerde sorunlar devam etmektedir. 2012’de yapılan bir çalışmada ingilizceden türkçeye çeviri yapılırken kök bilgisi de kullanılmaktadır [59]. 2015’te yine istatistiksel bir yöntem olan Sinir Ağları kullanılarak ingilizce türkçe çeviri çalışması yapılmıştır [60]. İki kaynak dil için Türkçe için oluşturduğumuz sistem yapılırsa sadece tümce matrisinin elemanları kavram olarak çevrilerek kaba anlamsal bir çeviri yapılabilir yada kaba anlamsal çeviri detaylı çevirinin doğruluğunun kontrol edilmesinde katkı sağlayabilir.

1.4.6 Duygu analizi (Sentiment analysis)

Duygu analizi; belli bir konuya ya da ürüne bağlı olarak metin içeriğinin olumlu, olumsuz veya tarafsız sınıflarından hangisine dahil olduğunu inceler. Başta twitter olmak üzere artan sosyal medya kullanımı, kişileri ve ürünleri hakkında kullanıcı yorumlarına sahip şirketleri bu alana yöneltmiş durumdadır. Kullanıcı verisinin doğru yoruma izin verecek kadar artması, sonuçların firmalar açısından müşteri odaklı ürün ve hizmet olarak görünür hale gelmesi bu alanda yapılan çalışmaları artmasına neden olmuştur. Duygu analizinin biraz daha ayrıntılandırılmış biçimi olan Düşünce Madenciliği (Opinion Mining) çalışmalarında aranan konular hakkındaki düşünceler bazı sıfat ve sözcüklerle ifade edilip derecelendirilir. Bo Pang ve Lillian Lee 2008’de duygu analizi ve düşünce madenciliği ile ilgili ayrıntılı bir çalışma yayınlamıştır [61]. Duygu analizi çalışmalarında öznenin ve nesnenin ne olduğu, hangi kavramla ilgili olduğu özellikle önemlidir çünkü duygu analizi çalışmaları genellikle bir konu veya ürün-üzün tipleri üzerinedir. Çalışmamız bu konu ve kavramların özne ve nesne gibi öbekler halinde çıkartılmasında kolaylık sağlayacaktır.

1.4.7 Metin tanılama (Paraphrase identification)

Metin Tanılama bir metnin anlamını daha anlaşılır bir biçimde (ya da farklı sözcüklerle) ifade etmektir. İki tümcenin bir birini tanılaması çift yönlü gerektirir

şeklinde. Bilgi çıkarımı, makine öğrenmesi, bilgi getirme sistemleri ve telif hakkı ihlallerinin otomatik tanılmasında metin tanılama sistemleri kullanılabilir.

Metin Tanılama ilk olarak McKeown (1979) tarafından bilgisayarın kullanıcı girişlerini doğru anlaması amacıyla kullanılmıştır [62]. Daha sonra Ravichandran ve Hovy [63], Barzilay ve Lee [64] ve Dolan ve Brockett [65] istatistiksel teknikler kullanarak metin tanılama yapmıştır. Bu çalışmalarda geniş miktardaki metne kaba benzerliklerine göre bakılarak aynı anlama sahip tümceler çıkarılır.

Bir diğer çalışmada Chang ve arkadaşları [66] kısıtlı gizli bilgi ile diziler arasındaki ilişkilerin modellenmesi için ayrışimsal bir teknik önermiştir. Bu uygulamada, gizli birliktelikler ikili sınıflandırıcıyı istatistiksel olarak modellemekte kullanılmıştır.

Heilman ve Smith [67], bir tümceden diğerine söz dizim ağacını dönüştürerek kullanmıştır. 2011’de Socher ve arkadaşları [68] sözcükler ve onları ifade eden sözcükler arasındaki ilişkilerini bulmak için sinir ağlarının özel bir yapısı olan yinelemeli otomatik çözümleyici (recursive autoencoder) kullanmıştır. Büyük miktarda işaretlenmemiş veriden sözcüklere bağılıkları ve vektör uzayındaki sözcüğün söz dizim ağaçlarındaki düğümleri öğrenilmiştir. Bu bilgiyi ikili sözcükler arasındaki ifade ilişkisini tespit etmek için kullanmışlardır.

2015’te Metin tanılama sonuçlarının iyileştirilmesi için makine öğrenmesi katsayılarının incelenmesi ile ilgili bir çalışma yapılmıştır [69]. Metin tanılamada iki taraflı gerektirme içerdiği için dilin mecazi, eş anlamlı kullanımlarından dolayı zorluklar içermektedir. İki tümcenin sözcük olarak birbiriyle eşleşmesinden önce öbek kavramı olarak eşleşme sağlayıp sağlamadığını görmek Metin Tanılama doğruluklarını artırabilir. Bu açıdan tezimiz bu alanda fayda sağlayabilecek bir kaynak niteliğindedir.

1.4.8 Anlamsal rol etiketleme (Semantic role labelling SRL)

Anlamsal Rol Etiketleme (ARE), kaba anlamsal ayrıştırma olarak da kullanılır. DDİ’de tümcenin yüklemine ve içerdiği öbekler için çeşitli rollerin belirlendiği alandır. Örneğin “*Ali arabasını Ayşe’ye sattı.*” tümcesinde sattı yüklemi oluşturmaktadır. Ali satıcıdır (iş yapandır). Araba satılan şeydir (işten etkilenen nesne) ve Ayşe alıcıdır. ARE bir tümceden anlam çıkarmak için önemli bir adımdır, Tümcenin, bir söz dizim ağacına göre biçimden bağımsız temsilidir. Örnek olarak “*Ali arabasını Ayşe’ye sattı.*”

tümcesi “Arabasını Ayşe’ye Ali sattı.” gibi farklı biçimlerde yazılmış olsa da anlamsal rolleri aynıdır. Anlamsal etiketleme pek çok DDI alanı için giriş verisi sağlamaktadır ve önemli bir konudur. 2004 yılında yapılan bir çalışmada roller otomatik olarak çıkarılmaya çalışılmıştır [70]. 2010 yılında yapılan bir başka çalışmada anlam çıkarımı için rol kütüphaneleri kullanılmıştır [71]. 3.2.3 bölümünde detaylandırıldığı üzere öbek kavram çiftleri Verb-Net tematik rolleriyle eşleşmekte daha doğrusu bu rolleri bir nevi kapsamaktadır. Tezimizde kullandığımız öbek kavram çiftleri rol etiketi olarak kullanılıp anlamsal işaretleme yapmada kullanılabilir.

1.4.9 Metin gerektirimlerinin çıkarılması (Recognizing textual entailment)

Metin Gerektirimi DDI’de metin parçaları arasındaki tek yönlü ilişkiyi bulma amacındadır. Metin gerektiriminde örnek tümce “*t*”, hipotez tümce “*h*” ile ifade edilir. Mantıksal gerektirimden farklı olarak daha gevşek bir ilişki vardır. Eğer “*t*” ifadesi “*h*” ifadesini gerektiriyorsa, “*t*”nin doğru olduğunu söyleyen bir okuyucu büyük oranda “*h*”nin de doğru olduğunu söyleyecektir.

DDİ çevrelerinde bu konudaki çalışmalar 2005’lerden itibaren yaygınlaşmaya başlamıştır. Metin Analizi Konferansı (Text Analysis Conference) ve Anlamsal Değerlendirme Çalışmaları (SEMEVAL Exercises) Metin Gerektirimlerinin Çıkarılması (MGÇ) hakkındaki çalışmaları teşvik etmektedir. Bu internet ortamlarında amaç sitede paylaşılan ortak referans tümcelerin gerektirim ilişkisini büyük doğrulukta bulabilmektir.

Metin Gerektirimlerinin Çıkarılmasının çok bilinen bir uygulaması ikili sınıflandırma problemidir. Burada sistem bir gerektirme ilişkisi var mı, yok mu diye bir tahminde bulunmaya çalışır. Gerektirmenin bilinmeme durumunu da kapsayan üçlü sınıflandırma uygulaması da vardır.

Aşağıdaki iki örnek RTE3 çalışmalarından alınmış gerektirme ilişkisinin sağlandığı referans-hipotez çifti örneğidir.

Referans Metin: “Sıradışı Kız” Sue Graham adında (Mabel Normand tarafından oynanmıştır) bir küçük kasaba kızının hikâyesini anlatan bir Hollywood filmidir. Bu film Mack Sennett tarafından yönetilmiştir. “Vidor’un

kralı” ve “İnsanlara Göster” gibi Hollywood hakkında yapılan benzer birçok filme esin kaynağı olmuştur.

Hipotez Metin: *“Sıradışı Kız” Sennett tarafından yönetilmiştir.*

Dikkat edilmelidir ki referans metin birkaç tümceyi içerecek şekilde oldukça uzun oysa hipotez metin kısadır. MGÇ modellerinde ilk çalışılan modellerden biri sözcük kesesi (bag-of-words) sınıflandırıcılarıdır [72], [73] ve [74]. Bu yöntemde referans ve hipotez metinler öncelikle özel isim algılayıcısından geçirilir. Daha sonra tüm sözcüklerin referanslandırıldığı (örnek: İngilizcede hayvan, hayvan değil, dişi, erkek, nötr, çoğul, tekil vb şekilde işaretlendiği) bir ayrıştırma sisteminden geçirilir. Gerektirme ilişkisini tahmin etmek adına sözcük kümeleri modelinin iki tümcedeki sözcüksel ve anlamsal örtüşmeleri bulmak için kullanır. İki söz dizisi arasındaki ilişkili ifadeleri sırasal olarak eşleştirilir ve gerektirme olup olmadığına karar verilir. Bu yaklaşım iki tümce arasında hiçbir yapısal anlaşma içermez ama ortalama % 59 doğruluk sağlamaktadır.

Aşağıdaki örnek yine RTE3 veri kümesinden alınmıştır:

Referans Metin: *Dev Procter ve Gamble. 1.8 ABD Dolarlık ArGe bütçesiyle, çoğu küçük şirket olmak üzere 500 aktif şirketi yönetmektedir.*

Hipotez Metin: *500 küçük şirket Procter ve Gamble’ın ortağıdır.*

Yukardaki örneğin referanslandırma ve eşleşme ile çıkarsama yapılabilecek bir örnek olmadığı açık olarak görülmektedir. Ana tümce hipotez tümceyi belirtmemektedir ve bu kararı vermek bir DDİ sistemi için de zordur çünkü anlamsal çözümlemeye ve mantıksal durumlara ihtiyaç vardır.

Birinci model, zıt anlamlılık, yükleme göre kapsanabilecek rollerin çeşitliliği ve olumsuzlama gibi anlamsal özellikleri kapsamadığından tümceleri daha ayrıntılı inceleyen modeller üzerinde çalışmalar yapılmıştır. 2007’li yıllarda mantıksal gerektirim kurallarının ve dilin matemetiksel temsiline RTE alanında kullanımı ile ilgili çalışmalar artmıştır [75], [76]. Bos ve Markert (2005) mantıksal yapılar içeren derin anlamsal çözümleme kullandılar ve sonuçlarını kuram kanıtlayıcısından geçirdiler. Bu yöntemle doğruluk değerini yüzde 50’lerden yüzde 61’lara kadar çıkardılar ama MGÇ1 sınama kümesinde basit yaklaşımlar kadar iyi sonuç vermemiştir.

Daha yakın zamanlarda MacCartney ve Manning (2007) [77] daha gevşek bir yapısalılık içeren Doğal Mantık diye adlandırdıkları bir yöntem kullanmıştır. Öncelikle ilk yöntemdeki gibi referans tümce ve hipotez tümce hizalanmış ve bazı sözcüksel, söz dizimsel ve anlamsal özellikleri kullanan bir sınıflandırıcı tarafından yerel gerektirme kararları verilmiştir. Yerel kararlar birleştirme kuralları kullanılarak toplu gerektirme kararı verildi. Bu sistem, MGÇ3 ortak verileri üzerinde hem yüksek doğruluğu bakımından hem de basit MGÇ sistemlerinin sonuçlarıyla kıyaslanması bakımından yüzde 69 oranında performans göstermiştir. Berant ve arkadaşlarının 2011’de önerdiği bir yöntemde bir graf yapısında sistem tüm gerektirme kurallarını içeren genel yapıyı öğrenmektedir [78]. Metin gerektirimlerinin çıkarılması tek yönlü olduğu için bir paragraf ve tümce içinden esnek çıkarsamalar yapmaya izin vermektedir. Bu anlamda orjinal bir tümcenin kaba anlamsal çıkarsaması yapan sistemimiz MGÇ için sonuç oluşturabilecek özelliktedir.

1.4.10 Belge sınıflandırma (Document classification)

Belge sınıflandırma bir belgeyi bir yada birden fazla katagoriye bir algoritma yardımı ile atama işidir. Belgeler konusuna göre, yazar tipine göre, yayınlanma yılına göre gibi değişik özellikleri göz önüne alınarak sınıflandırılmak istenebilir. Altıncı bölümde kaba anlamsal tümce matrisimiz bir belge sınıflandırma uygulamasında denenerek iyi sonuçlar elde edilmiştir.

1.5 Tezin Kapsamı

Tez anlatım bozukluğu sorununa yapısal bir anlatım bozukluğu konusu olan öbek kavram çiftleri ve yüklem uyumu açısından yaklaşmaktadır. Yüklemin tümce içinde geçen öbek kavram çiftleriyle uyumlu olduğunu ya da uyumlu değilse hangi öbek kavram çiftinin uyumlu olmadığını göstermektedir. Bunun dışında yüklem zaman ekinin tümce içinde geçen zaman öbeğiyle uyumlu olup olmadığı, yüklem şahıs ekinin özne ile uyumlu olup olmadığı bu kapsamda incelenmektedir. Yüklemin aldığı çatı eklerinin öbek kavram çifti ile yüklem arasındaki uyumu nasıl etkilediği ayrıca değerlendirilmiştir.

Yüklem ile öbek kavram çifti uyumuna göre anlatım bozukluğu tespiti çalışmasında basit tümce, bileşik ve bağlı tümce gibi tüm tümce çeşitleri kapsamaktadır. Bağlı

tümcelerde birbirine vağılı birden fazla yüklem bulunduğı için öncelikle bağılı tümceler ayrılarak tek tek değerdendirilir. Bileşik tümce önce olduğı haliyle öbek kavram çiftlerine ayrılır ve yüklem ile öbek kavram çifti uyumu açısından anlatım bozukluğu olup olmadığı incelenir. Daha sonra içerdığı alt tümcelerde sorun olabileceğinden bileşik tümce alt tümcelere ayrılır ve her bir alt tümce için yüklem ile öbek kavram çifti uyumu kontrol edilir. Örneğın “Okulda koşarak gelen Ayşe, bugün derslere katılmadı.” bileşik tümcesi için tümce tümüyle değerdendirildiğinde “Okulda koşarak gelen Ayşe” özne-insan çiftiyle temsil edilmektedir ve katılmadı yüklemi öznenin insan olduğı durumla uyuşma gösterdiğinden sorun bulunamamaktadır. İkinci aşamada bu tümce “Ayşe okulda koşarak geldi” ve “Ayşe bugün derslere katılmadı” şeklinde iki alt tümceye ayrılmakta ve bu iki alt tümce için yüklem ile öbek kavram çifti uyumu kontrol edilmektedir ve bu örnekte ilk alt tümce için bulunmak-organizasyon ile geldi yüklemi arasında uyuşmazlık olduğı tespit edilmektedir.

Tezimizde uygulama kapsamında her çeşit tümce alt tümcelere ayrıştırılarak her bir alt tümce için kaba anlamsal matris temsili yapılmaktadır.

Bu çalışma yedi bölümden oluşmaktadır. İkinci bölümde bir tümcenin yapısını oluşturan temel kavramlar ve anlatım bozukluğu kavramı incelenmektedir. Üçüncü bölümde türkçe tümce çözümleme uygulamamızdaki teorik yaklaşımlar anlatılmaktadır. Dördüncü bölümde yüklem odaklı dil bilgisi ve anlamsal çözümleme modelimiz ayrıntılandırılmaktadır. Beşinci bölüm türkçe anlatım bozukluğu olayının çeşitlerinin ve görülme sıklığının incelenmesi, alt tümce ayırıcı, öbek ve kavram burucu araçlarının başarı değerdeleri incelenmesi ve anlamsal ve dil bilgisi çözümleme aracının değerdendirilmesi üzerinedir. Altıncı bölümde tümcenin kaba anlamsal temsili bir sınıflandırma uygulamasında kullanılmaktadır. Yedinci ve son bölümde sonuçlar aktarılmış ve çalışma sonlandırılmıştır.

1.6 Tezde Kullanılan Bazı Temel DDİ Kavramları

Tez içinde kullanılan DDİ’nin bazı temel kavramları ayrıntılandırılmıştır.

Derlem Bir dilin türlü kullanım alanlarından belli kurallarla derlenmiş örneklerinin dil bilgisi ve kuramsal dil bilimi araştırmalarında kullanılmak üzere bilgisayar tarafından okunabilecek biçimde bir araya getirilmiş kümesidir. Derlemler amaca

göre farklı özellikler içermektedir. Örneğin bir dilin farklı konulardaki tüm sözcüklerini ve ardışıl sözcük kümelerini içerebilecek kadar büyük derlemeler oluşturulup bu derlemelerden dil modelleri çıkarılmaktadır. Yine anlamsal ya da biçimsel çözümlemede işaretli derlemeler kullanılmaktadır. İşaretli derlem, derlemi oluşturan tüm sözcüklerin, tümcelerın ya da metnin daha büyük parçaların elle etiketlendiđi derlemelerdir. Burada amaç işaretli büyük boyuttaki derlemelerden bilgisayarın öğrenmesini sağlayarak, bilgisayarın öğrendiđi bilgiyi işaretsiz söz veya sözcüklerin işaretlenmesinde kullanmasıdır. Türkçe için söz varlıklarının incelendiđi çeşitli derlem konusunda çeşitli çalışmalar mevcuttur [79].

Bağlamdan Bağımsız Dil Modeli Bu model 1950’lerde Noam Chomsky tarafından geliştirilmiştir [80]. Bir dilin bağlam bağımsız olduğunu bağlam bağımsız bir dil modeliyle ifade edilip edilemeyeceğinden anlaşılmaktadır. Bağlamdan bağımsız dil modelinde sözcük ya da söz kümeleri arasındaki tüm dönüşümler, A bir değişken ve a uç eleman olmak üzere, $A \rightarrow Xa$ şeklindedir. Yani dönüşümün sol tarafında sadece bir değişken vardır. Bu aslında A gördüğümüz her yerde, herhangi bir başka duruma (yani aradığımız parçanın çevresindeki birimlere) bakmaksızın dönüşümü yapabilmek anlamına gelmektedir. Bağlam bağımlı olma durumunda ise $xAb \rightarrow xab$ örneğindeki gibi A değişkenini dönüştürebilmek için, etrafını da kontrol etmek gerekmektedir. Tezimizde türkçe için öbekleri değişkenler olarak kabul eden bağlamdan bağımsız bir dil modeli önerilmektedir.

Bağlılık Çözümleyici (Dependency Parser) Güncel bağlılık dil modeli kullanımı, Tesniere’in 1959’daki çalışmalarına dayanmaktadır [81]. Tesniere’ye göre tümce kendisini oluşturan öğeleri sözcükler olan düzenli bir topluluktur. Zihin tümceyi oluşturan sözcükler ve komşuları arasında ilişkileri bulur ve bu ilişkilerin bütünü tümcenin iskeletini oluşturur. Her bir ilişki bir alt terimi bir üst terime bağlamaktadır. Günümüzde bağlılık grameri tabanlı metin ayrıştırmasının amacı metin içersinde geçen her tümce için tümceyi oluşturan sözcükler arasındaki alt terim, üst terim ilişkisini bulmaktır.

Sıradüzeli avram sözlüğü (Ontoloji) Temel varlık kategorilerini listeyen, özelliklerini çıkaran ve bunlar arasındaki ilişkileri belirleyen kavramsal sözlüklerdir.

Kavramlar sıra düzensel bir yapıda birbirine bağlanmıştır. Kavramları bağlayan temel ilişki türleri eş anlamlılık, çok anlamlılık, eş seslilik, alt anlam, üst anlam, karşıt anlam, alt parça ve üst parça Çizelge 1.4’de tanımlarıyla birlikte gösterilmektedir. Bu sözlükler her bir ilişki tipi için bir sözcükten diğer sözcüklere dallar oluşturan ağaç yapılı derlemelerdir. Burada alt anlamla, alt parça; üst anlamla üst parça karıştırılabilmektedir. Alt anlam kavramsal olarak kapsanan anlam olmakla birlikte alt parça bir varlığın fiziksel olarak oluştuğu parçalarla ilgilidir.

Çizelge 1.4 : Sözcüklerin anlamsal ilişki türleri.

<i>İlişki Türü</i>	<i>Açıklama</i>
<i>Eş anlamlılık</i> <i>(Synonymy)</i>	Eş anlamlı (anlamdaş) sözcükler, yazılışları ve okunuşları farklı olmasına rağmen aynı anlamı taşıyan sözcüklerdir. Bu sözcükler birbirlerinin yerine kullanılabilir.
<i>Çok anlamlılık</i> <i>(Polysemy)</i>	Çok seslilikten farklı olarak bir sözcüğün birden fazla anlamı çağrıştırması ve aynı anda içermesi anlamına gelir. Örneğin Türkçe dil sözcüğü hem bir organ olarak kullanılır hem de bu organın da işin içinde yer aldığı lisan anlamı vardır.
<i>Eş seslilik</i> <i>(Homonymy)</i>	Yazılış ve okunuşları aynı olan; ama anlamları birbirinden farklı olan sözcüklere eş sesli (sesteş) sözcükler denir.
<i>Alt anlam</i> <i>(Hyponymy)</i>	Bir sözcüğün bir kavram kümesinin örneği olma durumudur. Örneğin “menekşe”,çiçek’in alt anlamları arasındadır. Ama çiçek dediğimizde tüm çiçek türleri aklımıza gelir.
<i>Üst anlam</i> <i>(Hypernymy)</i>	Bir kavramın başka bir kavramın genel adı olma durumudur. Çiçek, menekşenin üst anlamıdır.
<i>Karşıt anlam</i> <i>(Antonymy)</i>	Anlamca birbirinin karşıtı olan, birbiriyle çelişen sözcüklere zıt anlamlı sözcükler adı verilir. Zıt anlamlı sözcükler genellikle sıfat ve zarf özelliğindeki sözcüklerde bulunur.
<i>Alt parça</i> <i>(Meronymy)</i>	Parça ile bütün arasındaki anlamsal ilişkiyi gösterir. Örneğin “göz” sözcüğü “vücut” sözcüğünün alt parçasıdır.
<i>Üst parça</i> <i>(Holonymy)</i>	Parça ile bütün arasındaki anlamsal ilişkiyi gösterir. Örneğin “vücut” sözcüğü “göz” sözcüğünün üst parçasıdır.



2. YAPILARINA GÖRE TÜMCELER VE ANLATIM BOZUKLUĞU

Biçim açısından eklemeli bir dil olan Türkçe, akrabalık ilişkisi açısından Ural-Altay dil ailesinin Altay koluna girmektedir. Türkçe, bulundurduğu sesli sayısı ile doğru orantılı olarak, ses ahangı açısından gelişmiş bir dildir. Yapısı bakımından kural tabanlı bir dildir. Köklere sondan eklenen eklerle yeni sözcükler oluşturulur. Köke eklenen yapım ve çekim ekleriyle sözcükler pek çok farklı anlam ifade edebilmektedir. Günlük kullanımda TDK güncel sözlüğündeki yaklaşık 100.000 sözcükten kat kat fazla sözcük türetilmektedir. Yabancı sözcükler ve dile sonradan girmiş sözcükler dışında Türkçe bir kurallar dizisi şeklinde yapılandırılabilir. Çalışmamızda yüklemın öbek kavram çiftleriyle uyumluluğunu incelerken kural tabanlı bir yöntem geliştirilmiştir. Bu bölümde tezimizle bağlantılı konular kısaca özetlenip örneklendirilmiştir.

2.1 Tümceyi Oluşturan Öğeler

Duygu ve düşüncelerimizi anlatan ve içinde yargı bulunan sözcük dizisine tümce denir. Tümcede yüklem, özne, nesne ve tümleç olmak üzere 4 temel öge vardır. Bu dört temel öge içinde, en önemlisi yüklem ve öznedir.

2.1.1 Yüklem

Bir tümcede iş, oluş, yargı bildiren ve çekimli eylemden oluşan sözcüğe yüklem denir. Yüklemın en önemli özelliği tümceyi bir yargıya bağlamasıdır. Bu nedenle yüklem tümcenin temel ögesidir. Bütün yalın, türemiş ve bileşik eylemler, adlar ve ad soylu sözcükler ve sözcük öbekleri yüklem olabilir.

2.1.2 Özne

Tümcede işi yapan, eylemi oluşturan kişi, hayvan, bitki ya da canlı, cansız tüm varlıklara özne denir. Özne bir ya da birden çok sözcükten oluşabilmektedir. Birden çok sözcükten oluşan öznelere özne öbeği denir.

2.1.3 Nesne

Öznenin yaptığı işten etkilenen ve geçişli eylemi tümleyen sözcüklere nesne denir. Nesne görevini üstlenen sözcüklere düz tümleş adı da verilir. Nesnenin görevi, öznenin ve yüklemnin yaptığı işi belirtmektir. Bu durumu bir tümce içerisinde şöyle gösterebiliriz. Nesne “-i” ekini alıp almamasına göre belirtili nesne ya da belirtisiz nesne olarak adlandırılır. Belirtili nesnelerde nesne bilinen belli bir şeydir, belirtisizde ise bilinen bir şey değildir. Örneğin *Ayşe kedi gördü.*” tümcesinde görülen herhangi bir kedir. Ama *Ayşe kediye gördü.*” tümcesinde daha önce bahsi geçen, görülen yada bilinen bir keden bahsedilmektedir.

2.1.4 Tümleşler

Eylemin anlamını tümleyen, ad durum ekleri ve ilgeç alan sözcüklere tümleş denir. Tümleşler eylemin anlamını pekiştirir ve tümcenin anlamını tamamlar. Tümleşler dolaylı tümleş, ilgeç tümleşleri ve belirteç tümleşleri olarak üçe ayrılır.

Dolaylı tümleşler: Ad durum eklerinden “-a(-e), -da(-de), -dan(-den)” eklerini alan tümleşlere dolaylı tümleş denir. Dolaylı tümleşler ad durum eklerini aldıkları için yönelme, çıkma ve bulunma bildirirler. Örneğin; “Ali dedesinin evinden geldi” tümcesinde “dedesinin evinden” öbeği dolaylı tümleş öbeğidir.

İlgeç tümleşleri: İlgeçlerle kurulmuş tümleşlere ilgeç tümleşleri denir. Tümleşler “ile, için, gibi, göre, kadar” gibi ilgeçlerle birleşerek ilgeç tümleşlerini oluşturur. “Sen gelene kadar burada bekleyeceğim.” tümlecinde “Sen gelene kadar” öbeği ilgeç tümlecidir.

Belirteç tümleşleri: Yüklemnin anlamını zaman, yer, yön, durum, nicelik ilgisiyle tamamlayan sözcüklere belirteç tümşeleri denir. Belirteç tümleşleri bazen yalın olarak kullanılır, bazen de ad durum ekleri alır. “Sabah gelirim.” tümcesinde “Sabah” belirteç tümlecidir.

2.2 Türkçe’de Yapılarına Göre Tümce Türleri

Türkçede yapılarına göre tümceler temel olarak üçe ayrılır.

2.2.1 Basit tümce

Bir tek çekimli eylemle kurulan ve içinde bir tek bağımsız yargı bulunan tümceye basit tümce denir. Basit tümcenin içinde herhangi bir eylemsi bulunmaz. Basit tümcede tek bir bağımsız yargı ve tek bir çekimli eylem bulunur. Basit tümceler bildirme, istek, emir, soru ve ünlem tümceleri biçiminde olabilir. Örneğin, “*Çok güler yüzlü ve iyi niyetli bu çocuk ile çok asabi ve somurtkan abisi bu konuda bir türlü karar veremedi.*” tümcesi oldukça uzun bir tümce olmasına rağmen basit bir tümcedir.

2.2.2 Bileşik tümce

İçinde birden çok çekimli eylem ve birden çok yargı bulunan tümcelere bileşik tümce denir. Bileşik tümcelerde bir çekimli eylem, bir temel tümce ve bu temel tümceyi tamamlayan yan tümceler bulunur. Örneğin “*Akşam birden başlayan yağmur sokakta oynayan çocukları sırlıklam etti.*” tümcesi bileşik bir tümcedir. “*Akşam yağmurun başlaması*” yan tümce, “*yağmurun sokakta oynayan çocukları sırlıklam etmesi*” temel tümcedir.

2.2.3 Bağlı tümce

Anlam yakınlığı nedeniyle birbirine bağlanan tümcelere bağlı tümce denir. Bazı sıralı tümceler anlam yönünden birbirine tamamen bağlıdır. Bazıları ise anlam bakımından birbirine bağımlı değildir. Genellikle anlatıma kıvraklık ve canlılık kazandırabilmek için yalın ve bileşik tümceler birbirine bağlanır ve sıralı tümceleri oluşturur. Örneğin “*Ali geldi; okula gitti.*” tümcesi bağlı tümcedir. “*Ali geldi.*” ve “*Ali okula gitti.*” tümcelerinden oluşur.

2.3 Yükleme Odaklı Yapısal Çözümleme için Yükleme Gelen Ekler

Bir tümcede yükleme tümceyi yargıya bağlayan en temel ögedir. Bu yüzden yükleme eklenen eklerde tümce hakkında doğrudan bilgi içerir. Yükleme eklenen temel ekleri şahıs eki, zaman eki, kip eki ve çatı eki olarak kümeleyebiliriz.

2.3.1 Şahıs ekleri

Şahıs ekleri, eylemi şahsa bağlayan; eylemdeki işi, kılışı, hareketi yapanı; oluşa ve duruma sahip olanı bildiren eklerdir. Şahıs eklerinin tekil ve çoğul şekilleri vardır. Toplam altı şahıs eki çeşidi vardır. Zaman ve kip eklerine göre şahıs ekleri değişiklik gösterir; Ayrıca eylem kök veya gövdesinin ünlüyle veya ünsüzle birleşmesine göre şahıs ekleri değişir. Türkçedeki şahıs ekleri “-m, -im, -n, -in, -sin, -k, -iz, -niz, -iniz, -sınız, -ler ve -sinler”’dir.

2.3.2 Zaman ve kip ekleri

Yüklemler sonlarına gelen zaman ve kip ekleriyle eylemin, yargının ya da oluşun zamanını belirtir ya da gerçekleşip gerçekleşmemesini dilek, istek, gereklilik veya emir kavramları içerisinde verirler. “-di, -dı, -du, -dii, -tı, -ti, -tu, -tü, -muş, -miş, -muş, -müş, -r, -(i)r, -(ı)r, -(ü)r, -(u)r, -(e)r, -(a)r, -yor, -(i)yor, -(ı)yor, -(ü)yor, -(u)yor, -ecek, -acak” zaman ekleri; “-meli, -malı, -se, -sa, -e, -a” kip ekleridir.

2.3.3 Eylem çatıları

Eylem çatı ekleri eylemin öbeklerle ilişkilerini etkileyen, değiştiren eklerdir.

Etken eylem: İş yapan gerçek bir öznenin bulunduğu eylemlere etken eylem denir. Etken eylemler çatı eki almazlar.

Edilgen eylem: İş yapan gerçek bir öznesi olmayan eylemlere edilgen eylem denir. Edilgen çatılı eylemlerde özne işi yapan değil, yapılan işten etkilenen durumundadır. Edilgen eylemler, eylem kök veya gövdelerine “-l” veya “-n” eklerinin getirilmesiyle oluşturulur.

Dönüşlü eylem: Öznenin işi yaparken aynı zamanda o işten etkilendiğini gösteren eylemlerdir. Yani eylemi yapan da ondan etkilenen de öznedir. Dönüşlü eylemlerin olduğu tümcelerde özne gerçek öznedir. Nesne yoktur. Eyleme “-l” veya “-n” ekleri getirilerek yapılır. Bu eylemler nesne alamazlar; geçişsizdirler. “Kızlar süslendi; delikanlılar güzelce giyindi.” tümceleri dönüşlü eylem bulunduran tümcelerdir.

İşteş eylem: Eylemde bildirilen işin birden fazla kişi tarafından yapıldığını; işin beraber ya da karşılıklı yaptıklarını bildiren eylemlerdir. “-ş” ekiyle yapılır. “*Dövüşmek, uçuşmak, gülüşmek, görüşmek*” gibi eylemler işteş eylemlerdir.

Geçişli eylem: Belirtili ya da belirtisiz nesne alabilen eylemlerdir. Bu eylemlere “*ne, neyi, kimi*” soruları sorulduğunda belirtili ya da belirtisiz nesne bulunur. İş, kılış eylemleri geçişlidir. “*Titizlikle elindeki yazıları inceliyordu.*” tümcesi geçişli eylem bulunduran bir tümcedir.

Geçişsiz eylem: Nesne alamayan eylemlerdir. Oluş ve durum eylemleri geçişsizdir. Yükleme nesneyi bulmak için sorulan “*ne, neyi, kimi*” sorularının cevabı yoktur. “*Kar yağdı, tren durdu, ben uyudum, kartallar uçtu, dışarıda kaldı, o da yoruldu..*” tümceleri geçişsiz eylem bulunduran tümcelerdir.

Oldurgan eylem: Geçişsiz bir eyleme “-dir, -t, -r” eklerinden biri getirilerek eylem geçişli yapılırsa buna oldurgan eylem denir. “*Çocuğunu sevgi ve emekle büyüttü.*” tümcesinde “*büyümek*” geçişsiz bir eylem olmasına rağmen aldığı “-t” ekiyle geçişli hale gelmiştir. Bu durumda “*büyütmek*” oldurgan bir eylemdir.

Ettirgen eylem: Geçişli olduğu hâlde “-dir, -t, -r” ekleriyle tekrar geçişli yapılan fillerdir. Geçişlilik dereceleri artırılmıştır. Eylemi bir başkasına yaptırma söz konusudur. Oldurgan eylemler ettirgen hâlde getirilebilir. “*Ayşe kapağı Ahmet’e açtırdı.*” tümcesi ettirgen bir eyleme sahiptir.

2.4 Türkçede Anlatım Bozukluğu Olayı

Duygu ve düşüncelerimizi karşımızdakine aktarırken kurduğumuz tümcelerin açık ve anlaşılır olması, gereksiz unsurlar taşımaması, çelişkili anlatımlardan uzak olması ve dil bilgisi açısından doğru olması gerekir. Tümcelerimiz bu özellikleri taşımadığında, anlatım bakımından bozuk olur; iletişim tam olarak gerçekleşmez. Anlatım bozuklukları; sözcüksel anlatım bozuklukları ve yapısal anlatım bozuklukları olmak üzere iki temel başlıkta incelenebilir.

2.4.1 Sözcüksel anlatım bozuklukları

Gereksiz sözcük kullanılması: İyi ve sağlam bir tümcede gereksiz sözcük bulunmaz. Tümcede gereksiz sözcüğün kullanılması, anlatım bozukluğuna yol açar. Tümcede düşüncenin belirtilmesinde belli bir görevi olmayan sözcükler gereksizdir. Bu tür sözcükler, tümceden çıkarılmalıdır. Bir sözcük tümceden çıkarıldığında, tümcenin anlam ve anlatımında bir bozulma, daralma olmuyorsa, o sözcük gereksizdir. Çıkarıldığında tümcenin anlamı ve anlatımı bozuluyorsa, o sözcük gereklidir. Gereksiz sözcük kullanımından kaynaklanan anlatım bozuklukları, eş anlamlı sözcüklerin bir arada kullanılması ve anlamca birbirini kapsayan sözcüklerin bir arada kullanılması şeklinde olabilir. Örneğin “*Bari hiç olmazsa sen yanımızda kal.*” ve “*Kardeşim soruları hemen çözüverdi.*” tümceleri gereksiz sözcük kullanımına örnek tümcelerdir.

Anlamca çelişen sözcüklerin kullanılması: İyi bir tümce, karşıladığı yargıyı tam olarak anlatmalıdır. Yani tümceden bir anlam çıkarılmalıdır. Böyle olmaz da tümce çeşitli anlamlara gelirse; hem öyle bir anlam, hem böyle bir anlam çıkarsa ve birden çok yoruma yol açarsa, o tümcede çelişkili anlatım söz konusudur. İyi bir tümce açık olmalıdır. Tümcedeki açıklık ise anlamın kolayca anlaşılır olması demektir. Anlamca birbiri ile uyuşmayan sözcüklerin bir arada kullanılması, tümcede çelişkili ifadenin doğmasına neden olur. “*Tam üç yıla yakın bir zaman insanlık dramı yaşandı burada.*” tümcesi buna örnektir.

Sözcüğün yanlış anlamda kullanılması: Sözcüklerin karşıladığı anlam iyi bilinmelidir. Bu olmazsa, anlatmak istediğimiz düşünce ile ortaya çıkan düşünce farklı olur. Bu nedenle konuşurken ya da yazarken, düşüncelerimizi tam ifade edecek sözcükleri kullanmalıyız. Aksi hâlde düşüncelerimizi iyi anlatamayız, hatta sözümüz yanlış anlaşılabilir. “*Saat üçten dörde kadar olan süreçte işimi yaptım.*” tümcesinde “*süreçte*” değil “*sürede*” sözcüğü kullanılmalıdır.

Sözcüğün yanlış yerde kullanılması: Tümcedeki sözcüklerin yerinde kullanılması, söylenmek istenenin karşıtı bir anlamın ortaya çıkmasına ya da tümcenin anlaşılmasına yol açar. “*Yeni okula geldim ders zili çaldı.*” tümcesinde “*okula yeni gelindiği*” belirtilmek istenirken “*okulun yeni olduğu*” gibi bir anlam oluşmaktadır. Burada “*yeni*” sözcüğünün kullanım yeri yanlıştır.

Bağlaç yanlışları: Bağlaçlardan bazıları olumlu ve olumsuz yargıları birbirine bağlar. Bu duruma uymayan kullanımlarda anlatım bozukluğu meydana gelir. Örneğin, “*Ben yarın dışarı çıkamam ama evde yapılacak bir sürü işim var. ve Ahmet Bey oğlunu çok seviyor fakat bir dediğini iki etmiyordu.*” tümcelerinde bağlaç yanlışlığı vardır.

Eklerle ilgili yanlışlar: Tümcede eklerin eksik ya da fazla kullanılması anlatım bozukluğuna yol açar. “*Biz konuşmasını sevmeyen bir milletiz.*” tümcesinde “*konusmasını*” sözcüğünde iyelik eki gereksiz ve yanlış kullanılmıştır. Tümcenin doğrusu; “*Biz konuşmayı sevmeyen bir milletiz.*” olmalıdır.

Tamlama yanlışları: Çoğunlukla ad ve sıfatların aynı tamlanana bağlanması sonucu oluşan bir anlatım bozukluğudur. Bu nedenle isimlerle sıfatların aynı tamlanana bağlandığı kullanımlara dikkat etmek gerekir. “*Özel ve kamu kuruluşları iki gün tatil edildi.*” tümcesinde “*kamu kuruluşları*” ifadesi doğrudur. Çünkü bu, isim tamlamasıdır. Ancak “*özel*” sözcüğü “*kuruluşları*” tamlananına bağlanamaz. Çünkü “*özel*” sözcüğü sıfattır. Bu nedenle “*özel kuruluşları*” ifadesi yanlıştır. Tümcedeki bozukluğu gidermek için “*özel*” sözcüğünden sonra “*kuruluşlar*” sözü getirilmelidir ki doğru tümce “*Özel kuruluşlar ve kamu kuruluşları iki gün tatil edildi.*” şeklindedir.

2.4.2 Yapısal ve anlama dayalı anlatım bozuklukları

Sözcükler ve eklerin yazımına, tekrarından çok yapısal uyumsuzluklara dayanan ve sözcüklerin ve öbeklerin anlamındaki uyumsuzlukları içeren anlatım bozukluğu türüdür.

Deyimin yanlış anlamda kullanılması: Deyimler, dilin anlatım gücünü ve söyleyiş güzelliğini zenginleştiren unsurlardır. Deyimler, kısa ve özlü sözlerdir. Deyimlerin kalıplaşmış anlamları vardır ve deyimler bu kalıplaşmış anlamları çerçevesinde kullanılır. Kalıplaşmış belli bir anlamı karşılayan deyimlerin başka bir anlamda kullanılması, anlatım bozukluğuna yol açar. Örneğin “*Onun bize yaptığı iyiliklere hep göz yumduk.*” tümcesinde anlatım bozukluğu vardır.

Anlam belirsizliği: Kişilerden ya da onlarla ilgili durumlardan söz ederken, o kişilerin yerini tutan zamirleri kullanmayız. Bundan dolayı tümcede kişi bakımından bir belirsizlik ortaya çıkar. Anlam belirsizliği dediğimiz bu ifade bozukluğunu gidermek için tümcede sözünü ettiğimiz kişinin yerini tutacak zamiri mutlaka kullanmalıyız.

“Okula gitmediğini bugün öğrendim.” tümcesinin “onun okula gitmediğini” mi yoksa “senin okula gitmediğini” mi demek istendiği belirsizdir.

Mantık ve sıralama hatası: Tümcede verilen kavramların önem sırasının karıştırılması ya da tümcenin mantık açısından yanlış oluşturulması sonucunda ortaya çıkan anlatım bozukluklarıdır. “*Bırak patates doğramayı, yemek bile yapamaz o.*” tümcesinde sıralamadan kaynaklı mantık hatası vardır.

Özne yüklem uyumsuzluğu: Özne yüklem uyumsuzluğu kişi bakımından, tekillik çoğulluk bakımından ve özne eksikliği bakımından olmak üzere üç kümede incelenir:

Kişi bakımından uyumsuzluk; İyi bir tümcede özne ve yüklem arasında kişi bakımından uyum olmalıdır. “*Ben ve Ayhan buraya daha önce gelmişti.*” tümcesinde özne 1. çoğul (biz) yüklem 3. tekil (o) olduğundan özne ile yüklem arasında uyumsuzluk söz konusudur. Özne 1. çoğul olduğuna göre yüklem de birinci çoğul olması gerekir ki doğru tümce “*Ben ve Ayhan buraya daha önce gelmiştik.*” dir.

Tekillik-çoğulluk bakımından uyumsuzluk; Özne ile yüklem arasında belli bir uyum söz konusudur. Özne insan ve çoğul ise yüklem tekil ya da çoğul olabilir. Ancak insan dışındaki varlıkların (hayvan, bitki, eşya ve benzeri) çoğul şekilleri özne olduğunda yüklem daima tekil olur. “*Kuşlar ne de güzel uçuyorlar.*” tekillik çoğulluk bakımından uyumsuz bir tümce; “*Kuşlar ne de güzel uçuyor.*” uyumlu bir tümcedir. Yine “*Ağaçlar çiçek açmışlar.*” tekillik çoğulluk bakımından uyumsuz bir tümce; “*Ağaçlar çiçek açmış.*” uyumlu bir tümcedir.

Özne eksikliği bakımından uyumsuzluk; Özne ile yüklem arasında tekillik, çoğulluk ve kişi uyumsuzluğunun yanında, özne eksikliği de anlatım bozukluğuna yol açar. Özne, tümcenin temel ögesidir. Yüklemde bildirilen iş, oluş ya da hareketi yapan durumundadır. Yüklemdeki eyleme göre öznenin olmaması ya da bir öznenin birden fazla yükleme bağlanması anlatım bozukluğuna yol açar. Bu, daha çok sıralı ve bağlı tümcelerde karşımıza çıkan bir bozukluktur. “*Bir milletin dili ve edebiyatı hür olmadıkça yükselemez.*” tümcesinde özne eksikliğinden kaynaklanan bir bozukluk söz konusudur. Bu tümcede özne belirtilmemiş, bu da ifadeyi eksik bırakmıştır. Tümceye özne

getirerek bozukluk “*Bir milletin dili ve edebiyatı hür olmadıkça o millet yükselmez.*” şeklinde düzeltilebilir.

Öge eksikliği: Tümcede kullanılması gereken bir ögenin bulunmaması, anlatım bozukluğuna yol açar. Tümcenin temel öğeleri özne ve yüklem eksikliğini işlediğimizden, burada nesne ve dolaylı tümleç eksikliği üzerinde duracağız. Öge eksikliği sıralı ve bağlı tümcelerde karşımıza çıkar. Genellikle ortak kullanılan öğelerin yükleme bağlanamamasından kaynaklanır. “*Ben öğretmenime inanır ve severim.*” tümcesinde “*öğretmenime*” dolaylı tümlecinin “*inanır ve severim*” yüklemlerine bağlanmak istenmesi anlatım bozukluğuna yol açmıştır. “*Ben öğretmenime inanırım.*” olur; ama “*Ben öğretmenime severim.*” olmaz. Bu nedenle ikinci tümceye “*nesne*” getirilerek “*Öğretmenime inanır, onu severim.*” şeklinde anlatım bozukluğu giderilebilir:

Yüklem eksikliği: Sıralı ve bağlı tümcelerde iki tümcenin bir yükleme bağlanması sonucu anlatım bozukluğu meydana gelir. Yüklem eksikliği, bazen ikinci bir eylemin kullanılmaması ya da ek eylemin ortak kullanılması ile oluşur. Örneğin, “*Beşiktaş iskelesine geldiğimizde o işine, ben evime gittim.*” tümcesinde “*yüklem*” eksikliği anlatım bozukluğuna yol açmıştır. Her iki tümceyi “*gittim*” yüklemine bağlayamayız. “*Ben gittim*” olur ama “*o gittim*” olamaz. Bu nedenle tümcedeki yüklem eksikliğini giderirsek “*Beşiktaş iskelesine geldiğimizde o işine gitti; ben evime gittim.*” haliyle tümce anlamlı hâle gelir.



3. TMCE MLEMEDE KURAMSAL YAKLAIMLAR

alımamızda tmcelerin dil bilgisi ve anlam lemesi “bek” ve “kavram” olarak iki ana balık altında incelenmektedir. Trke tmceler lenirken ncelikle tmceler bir btn olarak ele alınmı ve bekleri bulunmutur. Daha sonra daha ayrıntılı bir leme yapabilmek iin baėlı ve bileik tmceler paralanarak basit tmcelere dntrlm ve her bir basit alt tmce iin leme yapılmıtır.

3.1 Tmce beklerinin lenmesi

Yklemle uyumluluėu denetlenecek bekler, izelge 3.1’deki gibi belirlenmitir. Bu sekiz beėin yedi tanesi ismin durum ekleriyle elde edilmitir. Bu eklerin oluturduėu bekler dıında alımamız kapsamında -le, -la ekli ara beėi ve ek almayan zarf beėi de kullanılmıtır.

izelge 3.1 : Trkedeki hal ekleri ve bekler arasındaki ilikilerin biimsel dil temsili.

bek	Ek
P_1 zne	-
P_2 Belirtisiz Nesne	-
P_3 Belirtili Nesne	-(y)ı, -(y)i, -(y)u, -(y)
P_4 Ynelme beėi	-(y)e, -(y)a
P_5 Bulunma beėi	-de, -da, -te, -ta
P_6 Ayrılma beėi	-den, -dan, -ten, -tan
P_7 Ara beėi	-le, -la
P_8 Zarf beėi	-
P_9 Niteleme Zarf beėi	-
Y Yklem	-

Trkede bek sırası esnektir dolayısıyla S tmcesi yukarda bahsedilen P_1 , P_2 , P_3 , P_4 , P_5 , P_6 , P_7 , P_8 , P_9 szdizim deėikenlerinin tm permtasyonları ve Y eklinde oluturulabilir. Yklem beėi de diėer beklerle yer deėitirdiėinde tmce hala anlaılabilir. Yklem beėinin sonda bulunmadıėı bu kullanım ancak bazı gnlk konumalarda ve iirlerde karılaılan yazı dilinde pek grlmeyen bir kullanımdır.

Türkçedeki hal ekleri ve öbekler arasındaki ilişkilerin biçimsel dil temsil yöntemiyle gösterimi Çizelge 3.2’de verilmiştir. Bu temsil P üretim kuralı, N sözdizim değişkeni, T uç sembol ve S başlangıç sembolü olmak üzere P, N, T, S şeklindedir [80], [82].

Çizelge 3.2 : Basit tümcedeki hal ekleri ve öbekler arasındaki ilişkilerin biçimsel dil temsili.

İfade	Tanım
$S \leftarrow p Y$	S: Basit Tümce
$P_1 \leftarrow X \mid \lambda$	P_1 : özne öbeği
$P_2 \leftarrow X \mid \lambda$	P_2 : X’i gösterme durumu
$P_3 \leftarrow X_i \mid \lambda$	P_3 : X’i gösterme durumu
$P_4 \leftarrow X_e \mid \lambda$	P_4 : X’e yönelme durumu
$P_5 \leftarrow X_{de} \mid \lambda$	P_5 : X’de bulunma durumu
$P_6 \leftarrow X_{den} \mid \lambda$	P_6 : X’den ayrılma durumu
$P_7 \leftarrow X_{le} \mid \lambda$	P_7 : X’i kullanma durumu
$P_8 \leftarrow \text{dün} \mid \text{sabaha karşı vb} \mid \lambda$	P_8 : zarf veya zarf öbeği
$P_9 \leftarrow \text{güzelce} \mid \text{biraz} \mid \text{yavaş vb} \mid \lambda$	P_9 : niteleme zarfı veya öbeği
$X \leftarrow \text{çocuk} \mid \text{eski ev vb.}$	X: isim yada isim öbeği
$Y \leftarrow \text{gelecekti} \mid \text{gitmeliydin} \mid \text{Ayşedir vb.}$	Y: yüklem

Basit tümcedeki hal ekleri ve öbekler arasındaki ilişkilerin biçimsel dil temsili Çizelge 3.2’de gösterilmektedir. “i”, “e”, “de”, “den”, “le” (hal ekleri), çocuk, eski ev vb. (isim ve isim öbekleri), dün, sabaha karşı vb. (zarf ve zarf öbeği), güzelce, biraz, yavaş vb. (niteleme zarfı ve niteleme zarf öbeği), gelecekti, gitmeseydin, Ayşedir vb. (Yüklem) uç sembollerdir. S, P_1 , P_2 , P_3 , P_4 , P_5 , P_6 , P_7 , P_8 , P_9 , X ve Y sembolleri ise sözdizim değişkenleridir. λ boş karakteri ifade etmektedir.

$$P = \{P_i : 1 \leq i \leq 8\} \quad (3.1)$$

kümesi üzerindeki tüm permütasyonların kümesi Π olsun. Türetim kuralı $P \in \Pi$ için $S, S \leftarrow pZ$ şeklinde ifade edilir. p yüklemden önceki öbeklerin tüm permütasyonlarını ifade eder.

Eğer tümce sadece isim öbeklerinden oluşan basit bit tümce değil de isim eylem, zarf eylem ve/veya sıfat eylem içeren birleşik bir tümceyse tümceler iç içe bir yapı oluşturabilirler. Görüldüğü üzere burada bir iç tümce içeren bir tümcenin iç tümcesi de bir başka iç tümce içermektedir.

Çizelge 3.3 : Bileşik tümcedeki hal ekleri ve öbekler arasındaki ilişkilerin biçimsel dil temsili.

İfade	Tanım
$S \leftarrow p Y$	S: Bileşik Tümce
$P_1 \leftarrow X \mid \lambda$	P_1 : özne öbeği
$P_2 \leftarrow X \mid \lambda$	P_2 : X'i gösterme durumu
$P_3 \leftarrow Xi \mid \lambda$	P_3 : X'i gösterme durumu
$P_4 \leftarrow Xe \mid \lambda$	P_4 : X'e yönelme durumu
$P_5 \leftarrow Xde \mid \lambda$	P_5 : X'de bulunma durumu
$P_6 \leftarrow Xden \mid \lambda$	P_6 : X'den ayrılma durumu
$P_7 \leftarrow Xle \mid \lambda$	P_7 : X'i kullanma durumu
$P_8 \leftarrow p E_{zarfeylem} \mid \text{dün} \mid \text{sabaha karşı vb.} \mid \lambda$	P_8 : zarf veya zarf öbeği
$P_9 \leftarrow \text{güzelce} \mid \text{biraz} \mid \text{yavaş vb.} \mid \lambda$	P_9 : niteleme zarfı veya öbeği
$X_{isimeylemle} \leftarrow p E_{isimeylem}$	İçinde mastar bulunduran isim öbeği
$X_{sifateylemle} \leftarrow p E_{sifateylem} X \mid p E_{sifateylem}$	İçinde ortağ bulunduran isim öbeği
$E_{isimeylem} \leftarrow \text{gelmek} \mid \text{gidiş} \mid \text{örme vb}$	Mastar (isim eylem)
$E_{sifateylem} \leftarrow \text{gelen} \mid \text{öpülesi} \mid \text{yapacak vb}$	Ortağ (sıfat eylem)
$E_{zarfeylem} \leftarrow \text{koşarak} \mid \text{kayıp} \mid \text{gelmeden vb}$	Ulaç (zarf eylem)
$X \leftarrow X_{isimeylemle} \mid X_{sifateylemle} \mid \text{çocuk} \mid \text{eski ev vb.}$	X: isim yada isim öbeği
$Y \leftarrow \text{gelecekti} \mid \text{gitmeliydin} \mid \text{Ayşedir vb.}$	Y: yüklem

Bileşik tümcedeki hal ekleri ve öbekler arasındaki ilişkilerin biçimsel dil temsili Çizelge 3.3'de gösterilmektedir. S bileşik tümcesi yukarıda bahsedilen $P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9$ sözdizim değişkenlerinin tüm permütasyonları ve Y şeklinde oluşturulabilir, burada basit tümcedeki gösterime ek olarak X isim öbeği sıfat eylem içeren bir isim öbeği yada zarf eylem içeren bir isim öbeği de olabilir. Ayrıca zarf öbeği, zarf eylem içeren bir zarf öbeği olabilir.

Çizelge 3.3'de görüldüğü üzere "i", "e", "de", "den", "le" (hal ekleri), çocuk, eski ev vb. (isim ve isim öbekleri), dün, sabaha karşı vb. (zarf ve zarf öbeği), güzelce, biraz, yavaş vb. (niteleme zarfı ve niteleme zarf öbeği), gelecekti, gitmeseydin, Ayşedir vb. (Yüklem) uç sembollerdir. S, $P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9$, X, Y, $E_{zarfeylem}$, $E_{sifateylem}$, $E_{isimeylem}$, $X_{sifateylemle}$ ve $X_{isimeylemle}$ sembolleri ise sözdizim değişkenleridir. λ boş karakteri ifade etmektedir.

Örnek olarak, iç içe öbek yapıları içeren "Başından geçenleri anlatan adam ağladıkça ben ağladım." tümcesi $S \leftarrow p Y$; $p = P_8 P_1$ şeklindeki ana tümcenin zarf öbeği $P_8 \leftarrow p E_{zarfeylem}$ şeklindedir. "Başından geçenleri anlatan adam ağladıkça" sözcük grubuna karşılık gelir. Bu öbekte "ağladıkça" $E_{zarfeylem}$ dir. "Başından geçenleri

"anlatan adam" sadece $p = P_1$ 'den oluşan birinci iç öbektir. P_1 özne öbeği ek almamış isim öbeğinden oluşmuştur ($P_1 \leftarrow X$); İsim öbekleri sıfat eylemleri içerebilir ($X \leftarrow X_{sifateylemle}$); İsim öbekleri sıfat eylem içerdiğinde, isim öbeği öbek permütasyonları, isim eylem ve isim öbeği şeklinde sıralanarak elde edilir ($X_{sifateylemle} \leftarrow p E_{sifateylem} X$). Bu şekilde ifade ilerlendiğinde ikinci iç tümcede; $E_{sifateylem}$ "anlatan"; X "adam" ve p 'de "başından geçenleri" söz öbeklerine karşılık gelmektedir. Bu şekilde çözümleme devam eder.

Çizelge 3.4 : Tüm tümce çeşitleri için hal ekleri ve öbekler arasındaki ilişkilerin biçimsel dil temsili.

İfade	Tanım
$S_{Tum} \leftarrow S B S \mid S$ $B \leftarrow ama \mid ve \mid veya \mid , \mid ; vb$	S_{Tum} : Basit Tumce, Bilesik Tumce, Bagli Tumce

Çizelge 3.4'deki ifadeler Çizelge 3.3'de gösterilen dil modeli üzerine eklendiğinde S_{Tum} basit, bileşik ve bağlı tüm tümceler için hal ekleri ve öbekler arasındaki ilişkilerin biçimsel dil temsili oluşturulmuş olur. Çizelge 3.4 ifadeleri S bileşik veya basit tümcesini bağlı tümceye dönüşümünü içeren ifadeleri göstermektedir.

Örnek olarak, "Buradan koşarak uzaklaştı ve bir daha gelmedi" tümcesi bağlacı "ve" olan $S_{Tum} \leftarrow S$ ve S şeklinde hem bileşik hem de basit tümcenin bağlanmasıyla oluşmaktadır.

Türkçedeki öbek sıralamasındaki serbestlikle ilgili çalışma Hoffman tarafından 1995'te yapılmıştır [83]. Tümce öbeklerine ayrılırken tümcenin bağıllık çözümlemesi sonuçları ve tümcenin öbekleri içeren biçimsel temsili kullanılmıştır. Çalışmamız kapsamında tümcenin içerdiği sekiz tür öbeği bulan "**Öbek Bulucu**" programı geliştirilmiştir.

3.2 Tümcenin Kavramsal Olarak Çözülmesi

Tümce sekiz öbeğe göre ayrıştırıldıktan sonra her bir öbek için kavram çözümlemesi yapılmıştır. Her bir öbeğin hangi kavramla ilgili olduğu saptanmıştır. Varlıkların kavramsal olarak sınıflandırılması son zamanlarda ontolojilerle birlikte sıklıkla karşımıza çıkmaktadır. Lakoff'a göre varlıkları sınıflandırmak, fiziksel çevreye, kültüre bağlı farklı değişkenleri içeren kapsamlı bir süreçtir [22].

Çalışmamızda tümceleri kavram uyumluluklarına göre inceleyebilmek için, dört önemli nokta göz önünde bulundurularak kavram seçimi yapılmıştır. Bunlardan ilki tümceyi hangi kavramların en iyi temsil edebileceği, ikincisi kavramların birbirinden yeterince ayrık olması, üçüncüsüye hangi kavramların yüklem uyumluluğuna göre belirleyici olduğu ve son olarak elimizde hazır bulunan terim sözlükleridir. Kavramlar seçilirken WordNet [49] gibi değişik ontolojiler incelenmiş ve sıra düzensel yapının en tepesinden en uygun kavramlar seçilmiştir. Çalışmamızda kullanılan kavramlar Çizelge 3.5’de gösterilmektedir.

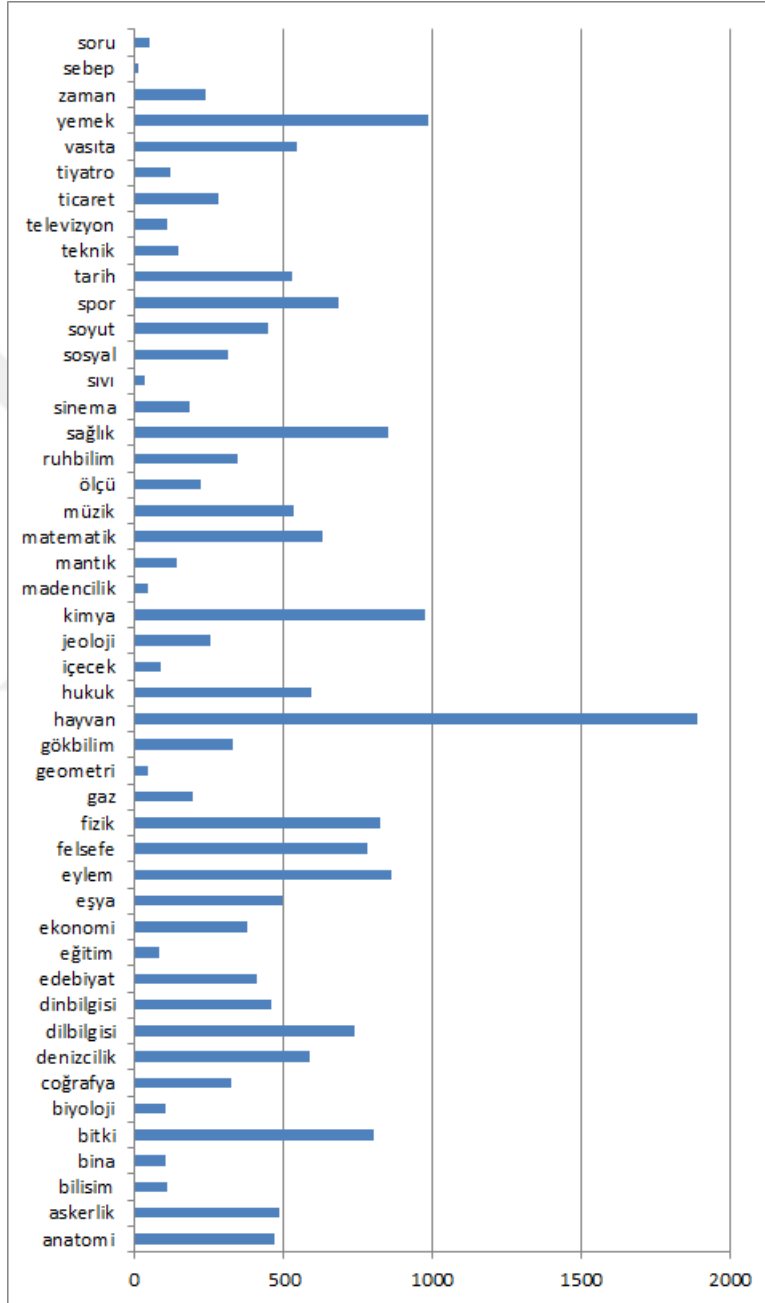
Çizelge 3.5 : Analiz aşamasında kullanılan kavramlar.

1 İnsan	11 Yemek	21 Fizik	31 Edebiyat	41 Denizcilik
2 Anatomi	12 İçecek	22 Kimya	32 Eğitim	42 Tiyatro
3 Hayvan	13 Gaz	23 Madencilik	33 Ekonomi	43 Sosyal
4 Sıvı	14 Eylem	24 Matematik	34 Felsefe	44 Televizyon
5 Bitki	15 Organizasyon	25 Mantık	35 Gökbilim	45 Sinema
6 Yer	16 Bilişim	26 Geometri	36 Ruhbilim	46 Spor
7 Zaman	17 Ölçü	27 Müzik	37 Sağlık	47 Tarih
8 Vasıta	18 Soyut	28 Hukuk	38 Jeoloji	48 Teknik
9 Eşya	19 Biyoloji	29 Dilbilgisi	39 Ticaret	49 Soru
10 Bina	20 Coğrafya	30 Dinbilgisi	40 Askerlik	50 Sebep

3.2.1 Kavram listelerinin oluşturulması

Kavram başlıkları belirlendikten sonra, her bir öbeğin hangi kavram kümesinde olduğunu anlamak için “*kavram listeleri*” oluşturulmuştur. Bu listeler, kavramı içeren isim ve isim öbeklerini bulundurulur. Bu listeler oluşturulurken BalkaNet’den [24], [25], Türk Dil Kurumunun terim sözlüklerinden, klasik sözlüklerin tanımlarda geçen açıklamalardan ve özel terim sözlüklerinden yararlanılmıştır [26]. TDK tarafından hazırlanmış olan bu kaynakların bazıları bize hazır sözcük listeleri olarak sunulmuştur. Bazı kaynakların ise işlenip, gerekli verinin çıkarılması gerekmiştir. İşlemler sonucunda elde edilen listelerden bazıları için örnek elemanlar Çizelge 3.6’de gösterilmektedir. Bu 50 türdeki kavrama ek olarak eğer aranan sözcük hiç bir kavram listesinde bulunamamışsa "bulunamadı" olarak işaretlenmiş ve toplamda 51 tane kavram türü kullanılmıştır.

Kavram dosyalarının içerdiği isim ve isim öbeği sayıları Şekil 3.1’de gösterildiği gibidir. Bu kavramlar dışında "yer" kavramı listesi 34790 tane eleman, organizasyon kavramı listesi 10777 tane eleman ve insan kavramı listesi 9061 tane eleman içermektedir. Bu iç kavram listesi diğerlerinden orantısal olarak çok büyük olduğu için aynı tabloda gösterilmemiştir.



Şekil 3.1 : Kavram listeleri boyutları.

Çizelge 3.6 : Bazı kavram listelerinden örnek elemanlar.

Kavram Sınıfı	Elemanları
Zaman	Saat 5, bu sabah, akşam üstü, öğleden sonra, bugün akşam v.b.
Yer	Dağ, yer, yerleşke, Kuzey yarımküre, doğu Asya, Türkiye v.b.
Gaz	Hava, su buharı, helyum gazı, karbon dioksit, hidrojen v.b.
Sıvı	Su, yağmur, sel, krema, jel, civa, kan, akışkan v.b.
İçecek	Kapıçino, süt, çay, içecek, kahve, ayran, içki, şarap v.b.
Yemek	Et, ekmek, hamburger, pilav, makarna, pasta, kek, kurabiye v.b.
Yapı	Bina, inşaat, gökdelen, köprü, heykel, site, villa, ev v.b.
Araç	Araba, bisiklet, motobisiklet, kamyon, otobüs, itfaiye arabası v.b.
Eşya	Kalem, sandalye, televizyon, tabak, çatal, kaşık, bilgisayar v.b.
İnsan	Ali, ayşe, öğretmen, başkan, sanatçı, baş danışman, bahçıvan v.b.
Hayvan	Köpek, aslan, deniz kaplumbağası, katil balina, kobra yılanı v.b.
Bitki	Çiçek, papatya, gül, ağaç, kayın, çalı, çam ağacı, çim, fidan v.b.
Org.	Üniversite, şirket, kuruluş, kurum, topluluk, vakıf v.b.
Soyut	Sevgi, akıl, hırs, istek, arzu, öfke, kıskançlık, zeka, yaratıcılık v.b.
Ölçü	Birim, metre, kilogram, candela, Japon yeni, türk lirası, uzunluk, v.b.
Eylem	Tüm eylemler

BalkaNet Kaynağının Kullanımı

BalkaNet Türkçe dahil Balkan dillerinin terimlerini kavram sınıfları altında sıra düzensel olarak kümelemeyi hedefleyen bir ontoloji kütüphanesidir. Word-Net yapısı temel alınarak oluşturulmuştur. Akademik çalışmalar için paylaşımına açık olan bu kütüphane kullanılarak önce çözümleme kapsamında belirlediğimiz kavramlar seçilmiş daha sonra bu kavramların alt kavramlarına kadar inilerek liste genişletilmiştir.

Türkçe Büyük Sözlükten Veri Üretme

TDK'dan sağlanan büyük sözlük, sözcükler ve sözcüklerin açıklamalarını içermektedir. Büyük sözlükten bilgi çıkarımı esnasında, sözcüklerin açıklamalarında geçen özel sözcüklerden faydalanılmıştır. Örneğin Büyük Sözlükte “gül” sözcüğü “*Gülgillerin örnek bitkisidir (Rosa)*” ve “*Bu bitkinin katmerli, genellikle kokulu olan çiçeğidir.*” olarak tanımlanmaktadır. Bitki kavram sınıfı oluşturulurken tanımdaki “*örnek bitkisidir*” ve “*çiçeğidir*” özel sözcükleri aratılarak gül’ün bitki kavram sınıfının bir elemanı olduğu bulunmuştur. Bu amaçla aranan her bir kavram için belirleyici ifadeler seçilmiş ve bu ifadeler sözcük tanımlarında aratılarak kavramlara ait sözcüklere ulaşılmıştır.

3.2.2 Yüklemin sınıflandırılması

Tümcenin öbek kavram çiftlerini yüklem durumuna göre inceleyebilmek için yüklem uygun oldukları öbek kavram çiftlerine göre sınıflandırılmıştır. İsim öbekleri için belirlenen kavram sayısı, yüklem için oluşturulan öbek kavram sınıfları sayısı ile doğrudan bağlantılıdır. Çalışmamızda her bir öbek için 51 kavram kullanılmıştır. Her bir yüklem 51 kavram x 10 öbek çiftinden hangisiyle ilişkilendirilebileceğine göre sınıflandırılmıştır ve böylece 510 yüklem sınıfı elde edilmiştir. Örneğin, zaman kavramını nesne olarak alan yüklem, insan kavramını nesne olarak alan yüklem, yer kavramını yönelme olarak alan yüklem gibi yüklem kümeleri oluşturulmuştur. Yüklem sınıflarına örnek olarak, öznesi insan olan yüklem sınıfında “*okumak, düşünmek, yemek, ağlamak, yakınmak*” gibi yüklem bulunmektedir. Belirtili nesnesi bitki olan yüklem sınıfında “*sulamak, koparmak, ekmek, dikmek*” gibi yüklem bulunmektedir.

Yüklem hangi durum ekini aldığına göre yüklem kümeleri listeleri Türk Dil Kurumundan sağlanmıştır. Türk Dil Kurumundan alınan yaklaşık 7500 yüklem kökü eylem olan ve çatı eki almamış yaklaşık 1000 tanesi tek tek incelenerek hangi öbek kavram sınıfına gireceği çalışmamız kapsamında belirlenmiştir. Yüklem çatı eki aldığıda nasıl davranacağı da çalışmamız kapsamında kurallarla belirlenmiştir

3.2.3 Öbek kavram çiftlerinin Verb-Net tematik rolleriyle ilişkisi

Bir tümcede öbek kavram çiftlerinin bulunması o tümcenin anlamsal çözümlemesinde ilk aşamayı oluşturmaktadır. Türkçede öbek kavram çiftleri kim, ne, neyi, nede, neyle, nerede, nereyi, kaçta, kimde, kimle, neyle, ne kadar gibi sorulara cevap vermektedir. Örneğin eylemi kim yaptı sorusunun cevabı bir insan, bir hayvan, bir bitki olabilir. Doğrudan etkilenen nesne bir bitki olabildiği gibi bir ölçü olabilir. Örneğin Ayşe 5 cm uzadı. Bulunma (de hali) öbeğinin karşılığı bir yer olabileceği gibi bir eşya olabilir. Örneğin, “*Tabağında böcek var.*” Kaynak (den hali) öbeğinin karşılığı bir insan olabileceği gibi bir zaman da olabilir. Örneğin, “*Ayşeden sıkıldım.*” veya “*Sabahtan yola çıktı.*” Eğer Araç öbeğinin kavramı bir vasıta ise aksiyonun o vasıta ile yapıldığı, eğer araç öbeğinin kavramı bir kişi ise aksiyonun bir kişiyle yapıldığını, eğer araç öbeğinin kavramı soyut ise aksiyonun soyut bir duyguyla ya da nedenle yapıldığı anlamına gelmektedir.

Çizelge 3.7 : VerbNet tematik rollerinin öbek kavram karşılıkları.

Tematik Rol	Eşlenik Öbek Kavram	Örnek
Yapan <i>Agent</i>	özne-insan; özne-hayvan	Suzan fısıldadı. <i>Susan whispered.</i> Agent V
Birlikte Yapan <i>Co-agent</i>	araç-insan; araç-hayvan	Brendayla Molly tanıştı <i>Brenda met with Molly.</i> Agent V fwithg Co-Agent
Değer <i>Asset</i>	özne-ölçü	\$100,000 bu evi alır. <i>\$100,000 will build you a house.</i> Asset V Beneciary Product
Faydalanan <i>Beneficiary</i>	e hali (hedef) -insan; e hali (hedef)-hayvan	Carmen Mary’e bir elbise aldı <i>Carmen bought a dress for Mary.</i> Agent V Theme fforq Beneciary
Sebep <i>Cause</i>	den hali (kaynak)-soyut	O suçluluktan kurtardı <i>It freed him of guilt.</i> Cause V Source fofg Theme
Hedef <i>Destination</i>	öbek:e hali(hedef)	Colorado’ya geldi. <i>He came to Colorado.</i> Theme V f+pathg Destination
Kaynak <i>Source</i>	öbek:den hali (kaynak)	A’dan B’ye dönmüş. <i>He converted from A to B.</i> Patient V ffromg Source ftog Goal
Bulunma <i>Location</i>	öbek:de hali (bulunma)	Tek boynuzlu atlar dünyada yoktur. <i>Unicorns don’t exist on Earth.</i> Theme V f+locg Location
Zaman <i>Time</i>	de hali-zaman	John partiye 5’te başladı. <i>John started the party at 5 oclock.</i> Agent V Theme V Time
Araç <i>Instrument</i>	öbek: le hali	Paula topa deynekke vurdu. <i>Paula hit the ball with a stick.</i> Agent V Patient fwithg Instrument
Etkilenen <i>Patient</i>	b’li nesne-insan; b’li nesne-hayvan	Celia bebeğin saçını taradı. <i>Celia brushed the baby’s hair.</i> Agent V Patient
Tema <i>Theme</i>	nesne-hayvan ve insan olmayan tüm kavramlar	Turistler resimlere bayıldı. <i>The tourists admired the paintings.</i> Experiencer V Stimulus

Kolayca anlaşılabilceği gibi öbek kavram çiftleri bir tümce ile ilgili en temel bilgileri içermektedir. Bazı durumlarda Türkçedeki öbek kavramlar Anlamsal Rol Atama (ARA) uygulamalarındaki tematik rollerle örtüşmektedir. Anlamsal roller ile tematik roller arasındaki eşleşmeler Çizelge 3.7’de gösterilmektedir.

3.3 Tümceleri Alt Tümcelere Ayırma

Çalışmamızda tümceler öbek ve kavram çiftlerine göre çözümlenmektedir. Bu amaçla daha ayrıntılı bir çözümleme yapmak için tümceler birer basit tümce olan alt tümcelere ayrıştırılmaktadır. Örneğin “*Dün koşarak okula gelecek Ayşe, bugün çok isteksiz.*” tümcesinde eğer alt tümce bulunmadan çözümleme yapılırsa, “*Dün koşarak okula gelen Ayşe*” özne-insan çiftiyle temsil edilmektedir. Oysa eğer tümce alt tümcelere ayrılırsa “*Ayşe dün okula koşarak gelecek.*” alt tümcesi için zaman uyumsuzluğu olduğu anlaşılacaktır.

Diğer dillerde olduğu gibi Türkçede de yüklem bir tümcenin en temel ögesidir. Türkçede eylemin aldığı çatı, kişi ve zaman ekleri bu önemi artırmaktadır. Bu yüzden bir tümce alt tümcelere ayrılırken içerdiği yüklem sayısınca alt tümceye ayrılmaktadır. Örneğin “*Ayşe’nin bu sabah okula geldiğini gördüm*” tümcesinde iki ayrı olay vardır.

1: Ayşe bu sabah okula geldi.; 2: Ben gördüm.

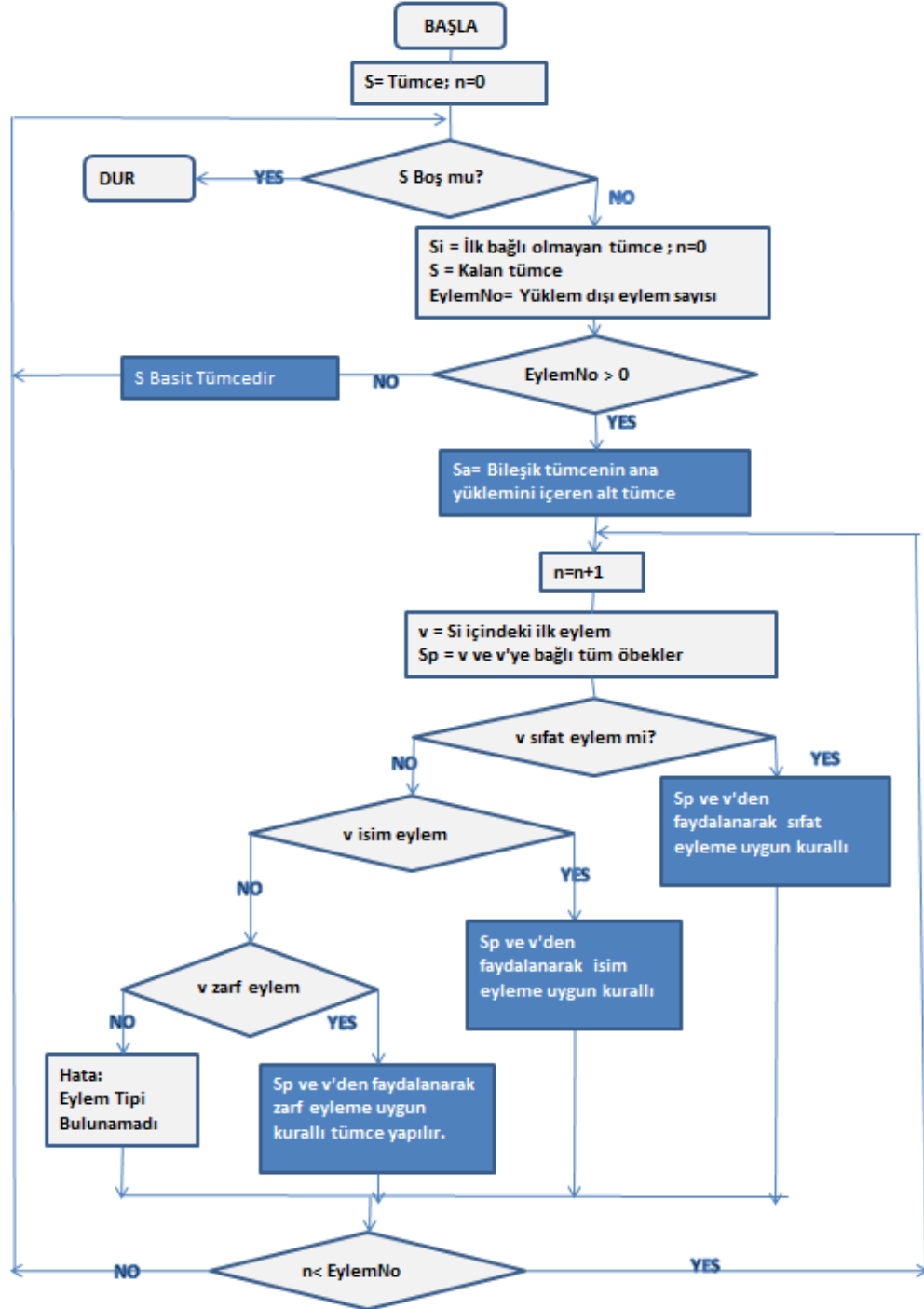
Bir başka örnek olarak, “*Ayşe’nin parlayan sarı saçları ve Fatma’nın ışıldayan mavi gözleri var.*” tümcesi 3 tümceye ayrıştırılmıştır.

1:Sarı saçları ve mavi gözleri var; 2:Ayşe’nin sarı saçları parlar.; 3:Fatma’nın mavi gözleri ışıldar.

Türkçenin bu yapısından dolayı tümceleri öbeklerine ayırırken önce tümceler alt tümcelere ayrılmış daha sonra her alt tümcenin öbekleri ayrı ayrı incelenmiştir. Bir tümce ana yükleme sahip tümceye ek olarak içerdiği eylem sayısı kadar tümceye ayrıştırılmıştır.

Şekil 3.2’de Tümcenin alt tümcelere ayırması aşamasında kullanılan akış diyagramı görülmektedir. Tümceler alt tümcelere ayrılırken İTÜ DDİ Bağlılık Çözümleyici çıktıları giriş verilerini oluşturmaktadır. Tümceler ayrıştırılırken önce bağlı tümceler

sonra bileşik tümceler ayrıştırılır. Yüklemelerin birbirine “sonra, ama” gibi bağlaçlarla bağlandığı ilk kısım, kalan kısımdan işlenmek üzere ayrılır. Artık bağlı olmayan bu tümcenin bileşik tümce olup olmadığı içinde bulundurduğu yüklem dışındaki eylemlere bağlıdır. Eğer içinde yüklemden başka eylem bulunduruyorsa eylemin isim eylem, sıfat eylem ya da zarf eylem olma durumuna göre uygun yapı kullanılarak basit alt tümceler oluşturulur.



Şekil 3.2 : Alt tümcelerinin bulunması.

İlk aşamada bağlı tümcenin kalan kısmı ile tüm işlemler tekrar edilir. Yani eğer hala bağlı tümce ise “*sonra, ama*” gibi bağlaçlarla bağlandığı ilk kısım, kalan kısımdan işlenmek üzere ayrılır. İlk kısım içinde yüklemden başka eylem bulunduruyorsa eylemin isim eylem, sıfat eylem ya da zarf eylem olma durumuna göre alt tümcelerine ayrılır. Bu işlem tüm tümce işlenene kadar, yani kalan kısım boşalana kadar devam eder.

3.3.1 İsim eylemlerden alt tümce yapımı

Eylemin adını bildiren, tümleç olan ve ad gibi kullanılan sözcüklere isim eylem ya da mastar denir. Mastarlar kök ya da gövdelerine “-mak, -ma, -ış” ekleri getirilerek türetilir. Mastarlar adların ve ad soylu sözcüklerin tüm özelliklerini gösterir. Tümce içersinde özne ya da tümleç olarak kullanılır. Ad durum eklerini, çatı neklerini ve olumsuzluk eklerini alabilir. Mastarlar, zaman kiplerine bağlı değildir. Eylem çekimine girmez. Bu yönleriyle eylemlerden ayrılırlar.

Çizelge 3.8 : İsim eylemler için alt tümcelerin oluşturulması

Ek	Türü	Örnek Tümce	Beklenen Sonuç
-mek	Inf1	Bu konuda daha fazla düşünmekte yarar var.	1-Bu konuda daha fazla düşündür. 2-Bunda yarar var.
-me		Her işini not alma iyi bir alışkanlıktır.	1-Her işini not almadır. 2-Bu iyi bir alışkanlıktır.
-ış	Inf2	Senin her sabah 5’de kalkışına şaşıyorum.	1-Senin her sabah 5’te kalkışındır. 2-Buna şaşıyorum.

Mastar içeren eylemler zaman eki yada şahıs eki almazlar, eylemin mastar eki aldığı durumlarda, çoğu zaman eylemden soyut olarak bahsedilmektedir ve aslında eylemin zamanı ve şahısı belirsizdir. İsim eylem içeren öbekler tümceye dönüştürülürken bu eylemlere ana yüklem zaman ve şahıs ekini vermek hatalı anlamlara neden olabilmektedir. Örneğin, “*İnsan beynini çözmek hiç bu kadar yakın olmadı.*” tümcesine çözmek mastarına ana tümcenin geçmiş zaman eki yada 3. tekil şahıs eki eklendiğinde “*İnsan beynini çözdü.*” tümcesi çıkarılır ki bu doğru bir çıkarım değildir. Bu tümcede soyut bir eylemden söz edilmektedir. Bu örnekte birinci tümce ana tümcenin yüklemine sahip ikinci tümcenin öznesi durumundadır. Elimizdeki örnekleri değerlendirdiğimizde mastar eki almış bir eylem için artık isme dönüşmüş haliyle

sonuna “-dir” eki getirilerek kullanmanın daha doğru bir yaklaşım olduğuna karar verdik, İsim eylemlerden alt tümce yapımı Çizelge 3.8’de gösterilmektedir.

3.3.2 Sıfat eylemlerden alt tümce yapımı

Eylemin sıfat olarak kullanılmasına sıfat eylem ya da ortaç denir. Ortaçlar bazen ad olarak da kullanılır. Ortaçlar çatı eklerini alır, olumsuz ve edilgen biçimleri yapılır ve zaman kavramı taşır. Bu nedenle diğer sıfatlardan ve mastarlardan farklı özellik gösterirler.

Sıfat eylem öbeği; türüne ve ana tümcenin yüklemine aldığı zaman ve şahıs ekine göre ana tümceden koparılıp alt tümce haline dönüştürülür. Sıfat eylemler adı üzerinde sıfat gibi davranarak bir ismi nitelerler. sıfat eylemin yüklem olacağı tümcede nitelediği isim özne olur.

Çizelge 3.9 : Sıfat eylemler için alt tümcelerin oluşturulması

Ek	Türü	Örnek Tümce	Beklenen Sonuç
-an	PresPart	Kışın açan çiçeklerin ömrü az olur	1- Çiçeklerin ömrü az olur. 2- O çiçekler kışın açar.
-ası	Feellike	Analarımızın öpülesi elleri vardır	1-Analarımızın elleri vardır. 2-O eller öpülür.
-mez	PresPart	Senin ne bitmez çilen varmış.	1-Senin ne çilen varmış. 2-O çile bitmez.
-ar	PresPart	Bu işte düzlüğe çıkar bir yol yok.	1-Bu işte bir yol yok. 2-O yol düzlüğe çıkar.
-dik	PastPart	Sizi biraz gidilmedik, yollardan götürüyim.	1-Sizi yollardan götürüyim. 2-O yollar gidilmedi.
-ecek	FutPart	Bana gazetemi getirecek biri yok mu?	1-Biri yok mu? 2-O biri gazetemi getirecek.
-mış	NarrPart	Onda hiç görülmemiş güzel elbiseler var.	1-Onda güzel elbiseler var., 2-O elbiseler hiç görülmemiş.
-dik + -üm	PastPart	Çözdüğün sorularla neden hala uğraşıyorsun?	1-Sorularla neden hala uğraşıyorsun, 2-O soruları çözdü+n.
-ecek + -üm	FutPart	Bana tekrar geleceğin günü iple çekiyorum.	1-Günü iple çekiyorum. 2-O gün bana gelecek+sin

Sıfat eylemler bazen şahıs eki ve zaman eki alabilirler, bu gibi durumlarda türüne yada ana tümcenin yüklemine bakılmaksızın sonuna gelen çekim ekleriyle çekimlenirler. Örneğin “Ayşe geleceği zaman bana haber ver.” tümcesinde sıfat eylem, ana tümcenin 2. tekil şahıs ekinden ve emir kipinden bağımsız olarak gelecek zaman ve 1. tekil

şahıs eki almıştır. Farklı sıfat eylem türlerine göre alt tümce elde edilişi Çizelge 3.9’de gösterilmektedir.

3.3.3 Zarf eylemlerden alt tümce yapımı

Eylemden türeyen ve belirteç görevi yapan sözcüklere zarf eylem ya da ulaç denir. Ulaçlar eylem çekimine girmez, kişi kavramı taşımaz. Ulaçlar tümcede belirteç olarak tümleç görevini üstlenirler. Bu yönleri ile de sıfat eylemlerden ayrılırlar. olumsuzluk eki alabilirler. Genellikle sözcükler arasında ilgi kurar. Eylemin belirttiği işi başka bir eyleme bağlayan, eylemin belirttiği işin durumunu, sebebini ve zamanını gösteren, birden çok eylemin belirttiği işler arasında nicelik ve nitelik yönünden kıyaslama yapan kullanım alanları mevcuttur.

Çizelge 3.10 : Zarf eylemler için alt tümcelerin oluşturulması

Ek	Türü	Örnek Tümce	Beklenen Sonuç
-maksızın	Without HaveDoneSo	Ders çalışmaksızın sınavı kazanamazsın.	1-Ders çalış.(Ana eylem zamanı). 2-Yoksa sınavı kazanamazsın.
-ıp	After DoingSo	Bu kağıdı müdüre imzalayıp geri getireceğim	1-Kağıdı müdüre imzalatacağım. 2-Sonra geri getireceğim.
-arak	ByDoingSo	Sadece kitap okuyarak bir ömrü geçiremezsin.	1-Sadece kitap okursun. 2-Böyle bir ömrü geçiremezsin
-ınca	When	Kapıyı açınca karşımda onu gördüm.	1-Kapıyı açtım. 2-O anda onu gördüm.
-madan	Without HaveDoneSo	Bize haber vermeden gitmeyin sakın.	1-Bize haber verin. 2-Önceden gitmeyin sakın.
-dıkça	AsLongAs	Sorularıçözdükçe konuyu daha iyi anlıyorum.	1-Soruları çözüyorum. 2-Böylece konuyu daha iyi anlıyorum.
-alı	Since DoingSo	Köyden ayrılalı yaklaşık 10 yıl oldu.	1-Köyden ayrıldık. 2-Ondan sonra 10 yıl oldu.
-ken	PastPart	Televizyon seyredirken çoğu kez uyuyakalırdı.	1-Televizyon seyredirdi. 2-Bu sırada çoğu kez uyuyakalırdı.

Zarf eylem öbeği; türüne ve ana tümcenin yüklemine aldığı zaman ve şahıs ekine göre ana tümceden koparılıp alt tümce haline dönüştürülür. Zarf eylemler bazen şahıs eki ve zaman eki alabilirler, bu gibi durumlarda türüne yada ana tümcenin yüklemine bakılmaksızın sonuna gelen çekim ekleriyle çekimlenirler. Örneğin “*Hep böyle bağırıcaksan ben gidiyorum.*” tümcesinde zarf eylem, ana tümcenin 1. tekil şahıs ekinden ve şimdiki zaman ekinden bağımsız olarak gelecek zaman ve 3. tekil

şahıs eki almıştır. Farklı zarf eylem türlerine göre alt tümce elde edilişi Çizelge 3.10'de gösterilmektedir.

3.3.4 Alt tümcelere ait yeni yüklem oluşturulması

Birleşik tümceler alt tümcelere ayrılırken isim eylemlerin, sıfat eylemlerin ve zarf eylemlerin bir yüklem içerdiği beklenen şahıs eki bilgisi ve zaman eki bilgisi olmayabilir. Öncelikli olarak isim eylem, sıfat eylem ve zarf eylem kendine bağlı olan öbeklerle birlikte bir alt tümce oluşturmak üzere gruplanır. Eğer eylemsi herhangi bir zaman ve şahıs eki içermiyorsa kendine bağlı öbeklerdeki özne ve zaman bilgisine bakılır. Eğer alt tümce özne içermiyorsa ana tümcenin yüklemine içerdiği kişi ekine bakılır. Yine eğer alt tümce öbekleri zaman bilgisi içermiyorsa ana tümcenin yüklemine içerdiği kişi ekine bakılır.

Kök durumundaki eylemin alacağı zaman ve şahıs eki belliyse alt tümce oluşturmak için bu ekleri eylemle birleştirmek gerekir. Birleştirme yapılırken türkçenin büyük ve küçük sesli uyumu, sert sessizlerin benzeşmesi ve yumuşama gibi kurallara uyulması gerekmektedir. Bu yüzden kök eylemin bazı özellikleri ekin eylem sonuna eklenme şeklini belirler. Bu özellik değişkenleri Çizelge 3.11'de gösterilmektedir.

Çizelge 3.11 : Eylemden yüklem oluşturma değişkenleri

Değişkenler
1-Eylem kökünün sesli ile bitip bitmemesi
2-Eylemin son seslisinin "a,ı"; "e,i"; "o,u" veya "ö,ü" oluşu
3-Eğer sessiz ile bitiyorsa son sessizin sert olması yani "p,ç,t,k" olması
4-Eylem kökünün tek hecelik kelime olup olmama durumu
5-Eylemin olumsuzluk eki alıp almaması
6-Eylemin aldığı zaman ekinden sonra sesli ile bitip bitmeme durumu
7-Eylemin aldığı zaman ekinden sonra yeni son seslinin "a,ı"; "e,i"; "o,u" veya "ö,ü" oluşu
8-Eylemin aldığı zaman ekinden sonra sessiz ile bitiyorsa son sessizin sert olması yani "p,ç,t,k" olması

Birleşik zamanları da içeren 31 çeşit zaman ve kip eklerinin 6 çeşit şahıs eki eşleştirmeleri için eylemin sahip olduğu özelliklere göre nasıl zaman ve şahıs ekine sahip doğru bir yükleme dönüştürüleceği tez kapsamında yazılımsal olarak kodlanmıştır. Her bir zaman ve şahıs ikilisi için eyleme eklenecek eklerle ilgili kuralalrın gösterimi mümkün olmayacağı için sadece ana tümcenin geniş zaman

olduğu durumda alt eylemin zaman eki ve şahıs ekinin bulunması ile ilgili detaylar Çizelge 3.13 ve Çizelge 3.12’de gösterilmektedir.

Çizelge 3.12 : Geniş zamanlı ana eyleme sahip tümcede alt eylemin zaman ekinin bulunması.

Tek hece çok hece	Sesli veya sessizle biten	son sesli	Ek	Yeni Son Sesli	Örnek
Tek	Sesli ile bitiyor	aei oöü	"r"	Eski son sesli	ye-r
Tek	Sessiz ile bitiyor	aiou	"ar"	"a"	yık-ar soy-ar
Tek	Sessiz ile bitiyor	eiöü	"er"	"e"	bez-er büz-er
Çok	Sessiz ile bitiyor	ai	"ır"	"ı"	bırak-ır kaşın-ır
Çok	Sessiz ile bitiyor	ei	"ir"	"i"	üşen-ir biçil-ir
Çok	Sessiz ile bitiyor	ou	"ur"	"u"	soyun-ur
Çok	Sessiz ile bitiyor	öü	"ür"	"ü"	üzül-ür

Çizelge 3.13’de görüldüğü üzere geniş zaman eki alan bir eylem için tek hece yada çok hece olması, sesli ile bitip bitmem durumu ve son seslinin “a,ı; e,i; o,u veya ö,ü” gruplarından hangisine dahil olduğu bilinmelidir. Bu üç parametre bilinirse eyleme hangi ekin geleceği ve yeni son seslinin ne olacağı kurallar dahilinde belli olur. Örneğin “sun” eylem kökü tek hecelidir, sessiz ile biter ve son seslisi “u”dur. Buna göre alacağı ek “ar” ve yeni son seslisi “a” olacaktır. Eylemin zaman eki almış son hali “sunar” şeklindedir.

Eğer eylem kökü geniş zaman eki alacaksa eylemin zaman eki almış son halinde son harf sessiz bir harf olan “r” olacaktır ve şahıs eki sadece eylemin son hecesi olan r”den önceki sesli harfin ne olduğuna göre belirlenecektir. Az önceki “sunar” örneği için eğer şahıs eki 2. çoğul şahıs (siz) olacaksa son sesli “a” olacağı için kişinin siz ve son seslinin “ar” grubuna dahil olduğu durumda ek “sınız” olacaktır.

Çizelge 3.13 : Geniş zamanlı ana eyleme sahip tümcede alt eylemin kişi ekinin bulunması.

Kişi	Son sesli	Ek	Örnek
ben	ai	"ım"	yık-ar-ım; sun-ar-ım; bırak-ır-ım
ben	ei	"im"	ye-r-im; beyen-ir-im; böl-er-im
ben	ou	"um"	bul-ur-um; yorul-ur-um
ben	öü	"üm"	üzül-ür-üm; bölün-ür-üm
sen	ai	"sın"	yık-ar-sın; sun-ar-sın
sen	ei	"sin"	beyen-ir-sin; böl-er-sin
sen	ou	"sun"	bul-ur-sun; yorul-ur-sun
sen	öü	"sün"	üzül-ür-sün; bölün-ür-sün
o	ai	-	yık-ar
o	ei	-	ye-r
o	ou	-	bul-ur
o	öü	-	üzül-ür
biz	ai	"ız"	yık-ar-ız
biz	ei	"iz"	ye-r-iz
biz	ou	"uz"	bul-ur-uz
biz	öü	"üz"	üzül-ür-üz
siz	ai	"sınız"	yıka-ar-sınız
siz	ei	"siniz"	ye-r-siniz
siz	ou	"sunuz"	bul-ur-sunuz
siz	öü	"sünüz"	üzül-ür-sünüz
onlar	ai	"lar"	yık-ar-lar
onlar	ei	"ler"	ye-r-ler
onlar	ou	"lar"	bul-ur-lar
onlar	öü	"ler"	üzül-ür-ler



4. YÜKLEM ODAKLI ÇÖZÜMLEME MODELİ

Çalışmamızda yüklemle öbekler arasındaki ilişkileri inceleyen bir çözümleme yöntemi geliştirilmiştir. Çalışmamız yüklem odaklı dilbigisi çözümlemesini incelemektedir çünkü yüklem uyumlu olmadığı hal ekleri gibi biçimsel hataları tespit edebilmektedir, yüklem odaklı anlamsal çözümlemeyi içermektedir çünkü "*Çanta köpeği düşündü*" gibi bir tümcenin anlamsal olarak doğru olmadığını, düşünmek ile çanta öznesinin uyumlu olmadığını tespit edebilmektedir. Bu yöntem uygulanırken çalışmamız kapsamında kurguladığımız yeni bir tümce gösterimi olan, tümcelerin sabit uzunluklu matrislerle gösterimi kullanılmaktadır. Bu yöntem ve tümcenin matris gösterimi ilgili bölümlerde anlatılacaktır.

4.1 Tümcenin Matris Gösterimi

Günümüzde bilişim ortamlardaki veri boylarının büyümesi, DDİ alanında sözcüklerin, tümcelerin ve metinlerin; alışılmış olduğu üzere sadece geçerli sözcük için 1 ve sözlükteki diğer tüm sözcükler için 0 olacak şekilde seyrek ve büyük boyutlu gösterimi yerine sıkıştırılmış vektörlerle gösterimine ihtiyaç duyulmaktadır. Bu güne kadarki çalışmalarda sadece bir sözcük gösterimi için tüm sözlük ele alınırsa yaklaşık 500.000, Türkçe için sadece sözcük kökü kullanılırsa yaklaşık 2000 tane özellik uzayı kullanmak gerekmektedir. 2013'te Mikolov derin öğrenme yöntemleri kullanılarak sözcüklerin 10 ile 500 arasında boyutlarda gösterim etmiştir [12]. Bu yöntemle sözcük vektörleri arasındaki kosinüs uzaklığına bakılarak sözcük benzerlikleri tahmin edilmeye çalışılmaktadır. Sözcük vektörleri arasındaki anlamsal ilişki % 75 Pearson korelasyon değerine sahiptir [13]. Bir başka deyişle sözcük vektörleri arasındaki yakınlık, sözcük anlamlarının birbiri arasındaki yakınlıkla % 75 ilişki göstermektedir. Burada bahsedilen sadece tek başına sözcükler arasındaki anlamsal benzerliktir. Sözcükler sıkıştırılmış vektörlere dönüştürüldüğü durumda; sözcük kümeleri arasındaki anlamsal benzerliğin nasıl değerlendirileceği, hesaplanacağı DDİ

alanında ve DDİ derin öğrenme alanında çözüm bekleyen bir konudur ve üzerinde çalışılmaktadır [14].

Çalışmamızda Mikolov'un sözcük vektör gösteriminden farklı olarak tümcenin vektör veya matris gösterimi söz konusudur. Çalışmamızda bir tümcenin kaba (coarse-grained) matris gösterimi yapılırken, öncelikle tümce birer basit tümce olacak şekilde alt tümcelerine ayrıştırılmıştır. Tümcedeki alt tümce sayısı tümce içinde geçen isim eylem, sıfat eylem ve zarf eylem sayısının toplamı kadardır. Çalışmamızda her bir alt tümce için tümce 10x51'lik öbek kavram matrisiyle ya da 510x1'lik öbek kavram vektörüyle ifade edilmektedir.

	zaman	yer	gaz	sıvı	içecek	yemek	bina	vasıta	eşya	insan	hayvan	bitki	org.	soyut	ölçü	eylem	...	bilinmiyor
özne	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	...	0
belirtisiz nesne	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	...	0
belirtili nesne	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	...	0
yönelme öbeği	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	...	0
bulunma öbeği	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	...	0
ayrılma öbeği	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	...	0
araç öbeği	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	...	0
zaman zarfı	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	...	0
niteleme zarfı	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	...	0
edat	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	...	0

Şekil 4.1 : Tümcenin matris gösterimi.

Şekil 4.1'de görüldüğü üzere “Ayşe, kırılan kalemi sevdiği evinden okula sevinçle götürdü.” tümcesi için; “Ayşe” “özne-insan” çiftiyle eşleşmekte, “kırılan kalemi” “belirtili nesne öbeği-eşya” çiftiyle eşleşmekte, “sevdiği evinden” “ayrılma öbeği-yer” çiftiyle eşleşmekte, “okula” “yönelme öbeği-yer” çiftiyle eşleşmekte ve son olarak “sevinçle” “araç öbeği-soyut” çiftiyle eşleşmektedir.

Matris elemanları 0 ya da 1 değerlerini almaktadır. Öbek kavram çiftinin varlığı 1 ile bu çiftin tümcede bulunmaması 0 ile gösterilmektedir. Bu matris gösterimi sayesinde, tümce ve yüklem matrisleri oluşturulduktan sonra sadece matris elemanı sayısınca kıyaslama yapılarak $O(1)$ karmaşıklığında yüklem odaklı dil bilgisi ve anlamsal hata tespiti yapılabilmektedir.

4.2 Ana Model

Çalışmamızda giriş verileri olacak tümceler öncelikle İTÜ DDİ Bağlılık Çözümleyici aracı [29] kullanılarak bağlılık çözümlemesi yapılmış tümceler elde edilmektedir.

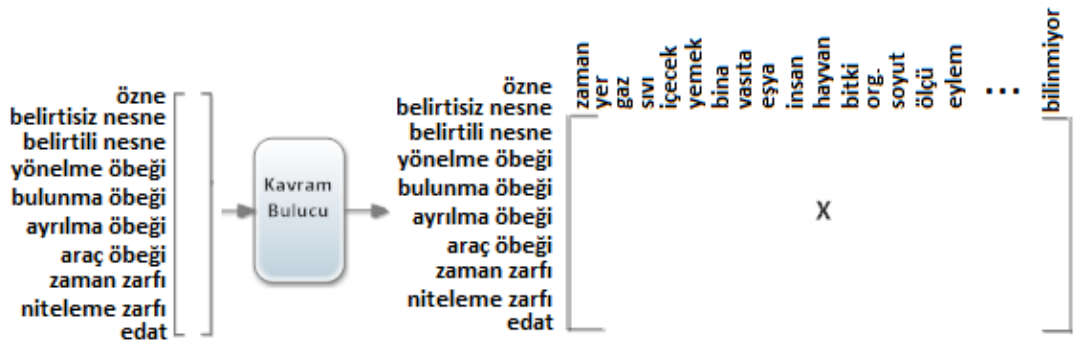
Tümceler tek tek ekran arayüzünden girilebildiği gibi toplu olarak bir metin dosyası üzerinden de girilebilmektedir. Giriş olarak bir metin verilse bile çözümleme tümce bazında ardışıl olarak yapılmaktadır. Bağlılık Çözümleyicinin her bir tümce için vermiş olduğu bağlılık ilişkileri çalışmamız kapsamında geliştirilen programın giriş verisi niteliğindedir.

Bağlılık çözümlemesi yapılmış giriş tümcelerimiz çalışma kapsamında geliştirilen “Öbek Bulucu” aracımız kullanılarak, varsa içerdiği on öbek bulunmaktadır. Öbekler bulunurken bağlılık çözümlemesi sonuçları, bağlamdan bağımsız dil gösterimi ve durum ekleri kullanılmaktadır. “Öbek Bulucu” aracı hem ikili hem de karakter katarı olacak şekilde 10x1 boyutlu iki ayrı çıkış üretmektedir. “Öbek Bulucu” aracının genel yapısı Şekil 4.2’de gösterilmektedir.



Şekil 4.2 : Öbek Bulucu.

Daha sonra “Öbek Bulucu” aracının çıktıları “Kavram Bulucu” aracının giriş verisi olarak kullanılır. Her bir öbeğin içerdiği kavram sınıfı, 51 kavram listesinde taranarak belirlenmektedir. Sonuç olarak, tümce Şekil 4.3’de gösterilen ve “Kavram Bulucu” aracımızın çıktısı olan X matrisine dönüşmektedir.

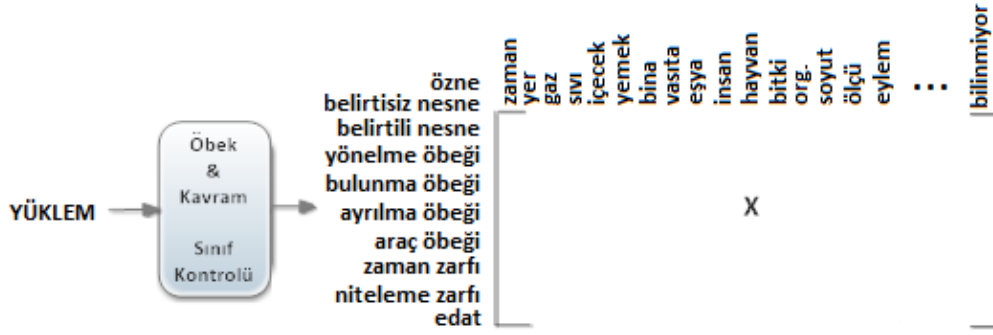


Şekil 4.3 : Kavram Bulucu.

Diğer taraftan, Tümcenin yüklemi, 510 ayrı öbek kavram sınıflarında aratılmaktadır. 510 yüklem sınıfı; özne olarak insan kavramıyla uyumlu eylem sınıfı; özne olarak eşya

kavramıyla uyumlu eylem sınıfı; belirtili nesne olarak bitki kavramıyla uyumlu eylem sınıfı; bulunma öbeği olarak soyut kavramla uyumlu eylem sınıfı gibi sınıflardır ve her sınıf eylem kökleri listelerinden oluşmaktadır. Analiz edilecek tümcenin yüklemi bulunduğundan sonra yüklem kökü eylemse bu eylem 510 eylem sınıfında aratılır. Şekil 4.4’de görüldüğü üzere yüklem uyumlu olduğu öbek kavram sınıfları 1, uyumlu olmadığı öbek kavram sınıfları 0 olarak işaretlenerek Y matrisi oluşturulmaktadır.

Eğer yüklem kökü bir isimse Y matrisi belirtili nesne ve belirtisiz nesne almayacak şekilde yani ikinci ve üçüncü satırı 0 olacak ve tüm diğer öbekleri alabilecek yani diğer tüm elemanları 1 olacak şekilde oluşturulur.



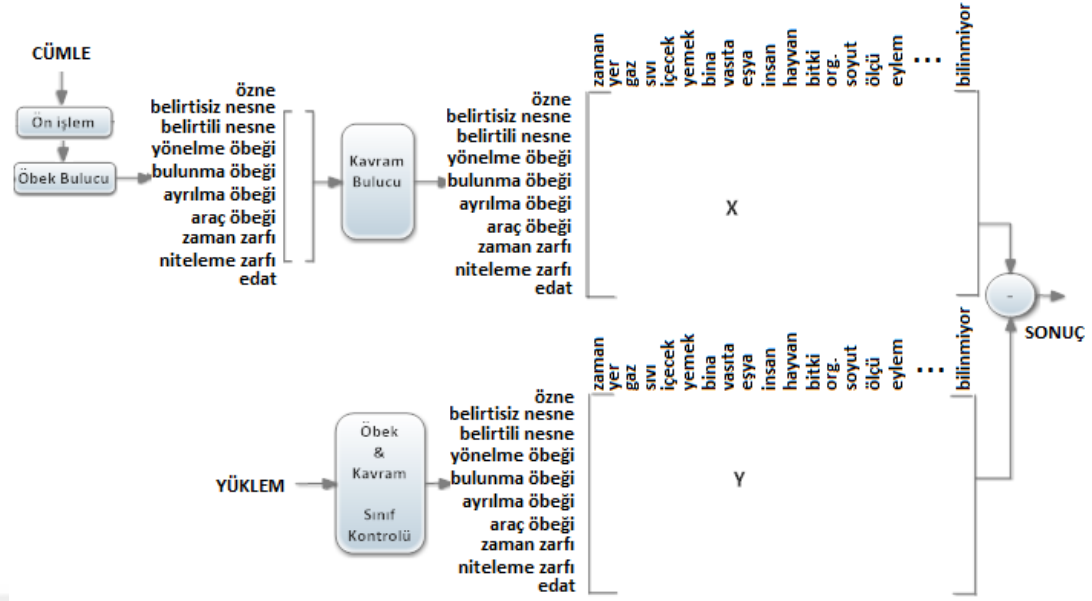
Şekil 4.4 : Yüklem sınıf matrisi.

X matrisi tümcenin gözlenen durumunu; Y matrisi ise tümcenin içermesi olası en genel halini gösterir. Yüklem odaklı dil bilgisi ve anlamsal hata tespiti yapılırken yükleme ait Y matrisi elemanları ile tümcenin öbek kavram ilişkilerini içeren X matrisinin elemanları bire bir kıyaslanmaktadır. Şekil 4.5’de tüm parçalarıyla birlikte ana model gösterilmektedir.

Eğer belli bir öbek kavram çifti için Y elemanı 0, aynı öbek kavram çiftindeki X elemanı 1 ise tümce yüklem ile öbek kavram uyumu bakımından sorunludur. Bu yüklem bu öbek kavram çiftini kabul etmediği halde, bu öbek kavram çiftinin tümcede bulunduğu anlamına gelir. Tümcenin doğruluğu (4.1) eşitliği ile gösterildiği üzere X matrisinin tümleyeni ve Y matrisi toplanarak bulunur.

$$F = X' \vee Y \quad (4.1)$$

Çalışmamızda sadece öbeklerin yüklemle uyumluluğunun denetimi birinci modeli, öbek kavram çiftlerinin yüklemle uyumluluğunun denetimi ikinci modeli oluştur-



Şekil 4.5 : Ana model.

maktadır. Yüklemin aldığı eklerin etkisi de üst üste konarak toplamda beş ayrı model oluşturulmuştur. Böylece her bir etmenin katkısının ayrı ayrı değerlendirilmesi sağlanmıştır.

4.3 Çatı Eklerinin Etkisi

Ana modeldeki Y matrisi yüklem kökünün uyumlu olduğu öbek kavram çiftlerini ve Y_n matrisi yüklem çatı ekleriyle birlikte uyumlu olduğu öbek kavram çiftlerini göstermektedir. Her bir çatı eki için Y matrisinin nasıl Y_n matrisine dönüşeceği dil bilgisi kurallarına göre bellidir. Örneğin edilgen bir yükleme sahip tümcede nesne sözde özneye dönüşür ve tümce nesne içermez. "Ayşe camı kırdı" tümcesindeki kırmak yüklemi "-ıl" edilgenlik eki alırsa "cam" sözde özneye dönüşür ve tümce nesne alamaz hale gelir. Edilgen yüklem içeren bir tümcede "Ayşe camı kırıldı." gibi hem özne hem de nesnenin bir arada kullanımı mümkün değildir. Bu örnekte edilgen Y_n matrisi elde edilirken matrislerin her satırı bir öbeği gösterdiği için Y matrisinin nesneyi içeren ikinci ve üçüncü satırı sıfırla çarpılmalıdır ve dönüşüm matrisinin ilgili öbeğe ait satırları sıfır olmalıdır. Ayrıca Y matrisinin nesnesi, Y_n matrisinin öznesi (sözde özne) olacağından "etkenden edilgene dönüşüm matrisi" bunu yapabilecek şekilde ayarlanmalıdır.

Çalışmamızda üçüncü model olarak, Y matrisi çatı eklerinin etkisiyle Y_n matrisine dönüşmekte ve tümcenin doğruluğu (4.2) eşitliği ile ifade edilmektedir.

$$F = X' \vee Y_n \quad (4.2)$$

Ana modelde oluşturduğumuz Y matrisinden yola çıkarak her bir çatı eki için Y_n matrisinin nasıl elde edebileceği; Y_n matrisinin elde edilmesi için kullanılan T_n dönüşüm matrisinin ne olması gerektiği anlatılmıştır.

4.3.1 Geçişli / geçişsiz yüklem

Geçişli bir yükleme sahip tümce nesne ve belirtili nesne alabilirken; geçişsiz bir yükleme sahip tümce nesne alamaz. Geçişli bir yüklem nesne kavram sınıflarında bulunurken geçişsiz yüklem nesne kavram sınıfında bulmaz. Yüklemin geçişli veya geçişsiz olma durumu eylem köküyle ilgili bir özellik olduğu durumlarda yüklem kökü eylem sınıf listelerinde aranırken belirtili nesne ve tüm kavram çiftlerine ait sınıflarda o eylem bulunmayacaktır. Örneğin almak geçişli bir eylemken olmak geçişsiz bir eylemdir. "*Deprem saat beşte oldu*" tümcesinin yüklemi kavram sınıflarında aratılırken belirtisiz nesne eşya sınıfında bulunamayacağı için bu sınıfı gösteren matris elamanı 0 olacaktır. Bu durum tüm belirtisiz nesne ve 51 kavram çifti ; tüm belirtili nesne ve 51 kavram çifti için geçerli olacaktır. Nesne dışındaki diğer öbek kavram çiftlerinin hangi değeri alacağı eylem köküne bağlıdır.

4.3.2 İsim kökünden türeyen veya isim yüklemeler

Eğer tümcede yüklem kökü “-len, -leş” eklerini almış isim köklü eylemse veya sadece isimden oluşmuşsa tümce nesne içermez. Bu tarz yüklemelerin saptanması için uygulamamızda yüklem kökü incelenmekte ve eğer yüklem kökü isim türünde ise Y matrisinin ikinci ve üçüncü satırları sıfırlanmaktadır. Örneğin “*Bu sabah bir baktım evimin önündeki ağaç çiçeklenmiş.*”, “*Zaman geçtikçe pantolon siyahlaştı.*”; “*Ayşe çok güzel bir kızdır.*” veya “*Bu her halde hayatımda gördüğüm en şiddetli rüzgardır*” tümcelerinin nesne almıyacağı açıktır.

4.3.3 Oldurgan yüklem

Oldurganlık ekiyle geçişsiz yüklem geçişli yükleme dönüşmektedir. Oldurganlık ekiyle birlikte geçişsiz haldeki özne, nesne olur. Bir başka deyişle Y matrisinin birinci satırı, Y_1 matrisinin ikinci ve üçüncü satırına yani belirtili ve belirtisiz nesneye

geçmektedir. Yeni matrisin öznesi sadece eylemi yaptırtacak özellikte olduğu için insan, hayvan ve bitki kavramını içermektedir.

“Ağaç büyüdü.” → Ağacı büyüttü. “Ağaç büyüdü.” → Ağaç büyüttü.

Şekil 4.6’deki gibi Y matrisi, T_1 dönüşüm matrisi ile çarpılıp C_1 sabit matrisle toplanarak oldurgan Y_1 matrisi elde edilmektedir.

$$\begin{bmatrix} \begin{pmatrix} 0000000000 \\ 1000000000 \\ 1000000000 \\ 0001000000 \\ 0000100000 \\ 0000010000 \\ 0000001000 \\ 0000000100 \\ 0000000010 \\ 0000000001 \end{pmatrix}_{10 \times 10} \times \begin{pmatrix} a_1b_1c_1d_1e_1f_1g_1 \\ a_2b_2c_1d_2e_2f_2g_2 \\ a_3b_3c_3d_3e_3f_3g_3 \\ a_4b_4c_4d_4e_4f_4g_4 \\ a_5b_5c_5d_5e_5f_5g_5 \\ a_6b_6c_6d_6e_6f_6g_6 \\ a_7b_7c_7d_7e_7f_7g_7 \\ a_8b_8c_8d_8e_8f_8g_8 \\ a_9b_9c_9d_9e_9f_9g_9 \\ a_0b_0c_0d_0e_0f_0g_0 \end{pmatrix}_{10 \times 7} + \begin{pmatrix} 0001110 \\ 0000000 \\ 0000000 \\ 0000000 \\ 0000000 \\ 0000000 \\ 0000000 \\ 0000000 \\ 0000000 \\ 0000000 \end{pmatrix}_{10 \times 7} = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ a_1b_1c_1d_1e_1f_1g_1 \\ a_1b_1c_1d_1e_1f_1g_1 \\ a_4b_4c_4d_4e_4f_4g_4 \\ a_5b_5c_5d_5e_5f_5g_5 \\ a_6b_6c_6d_6e_6f_6g_6 \\ a_7b_7c_7d_7e_7f_7g_7 \\ a_8b_8c_8d_8e_8f_8g_8 \\ a_9b_9c_9d_9e_9f_9g_9 \\ a_0b_0c_0d_0e_0f_0g_0 \end{pmatrix}_{10 \times 7} \end{bmatrix}$$

Şekil 4.6 : Oldurganlık eki alan yüklem için Y matrisinin dönüşümü.

Çatı ekleri için kullanılan tüm şekillerde yer sıkıntısından dolayı 10x51 boyutlu olan Y ve Y_n matrisleri 10X7 boyutlu gösterilmektedir.

4.3.4 Ettirgen yüklem

Ettirgenlik ekiyle yüklem yönelme öbeği alabilecek hale gelir. Yönelme öbeğinin alabileceği kavramlar ise yüklem eski halinin öznesinin kavramlarıdır.

$$\begin{bmatrix} \begin{pmatrix} 1000000000 \\ 0100000000 \\ 0010000000 \\ 1000000000 \\ 0000100000 \\ 0000010000 \\ 0000001000 \\ 0000000100 \\ 0000000010 \\ 0000000001 \end{pmatrix}_{10 \times 10} \times \begin{pmatrix} a_1b_1c_1d_1e_1f_1g_1 \\ a_2b_2c_1d_2e_2f_2g_2 \\ a_3b_3c_3d_3e_3f_3g_3 \\ a_4b_4c_4d_4e_4f_4g_4 \\ a_5b_5c_5d_5e_5f_5g_5 \\ a_6b_6c_6d_6e_6f_6g_6 \\ a_7b_7c_7d_7e_7f_7g_7 \\ a_8b_8c_8d_8e_8f_8g_8 \\ a_9b_9c_9d_9e_9f_9g_9 \\ a_0b_0c_0d_0e_0f_0g_0 \end{pmatrix}_{10 \times 7} = \begin{pmatrix} a_1b_1c_1d_1e_1f_1g_1 \\ a_2b_2c_1d_2e_2f_2g_2 \\ a_3b_3c_3d_3e_3f_3g_3 \\ a_1b_1c_1d_1e_1f_1g_1 \\ a_5b_5c_5d_5e_5f_5g_5 \\ a_6b_6c_6d_6e_6f_6g_6 \\ a_7b_7c_7d_7e_7f_7g_7 \\ a_8b_8c_8d_8e_8f_8g_8 \\ a_9b_9c_9d_9e_9f_9g_9 \\ a_0b_0c_0d_0e_0f_0g_0 \end{pmatrix}_{10 \times 7} \end{bmatrix}$$

Şekil 4.7 : Ettirgenlik eki alan yüklem için Y matrisinin dönüşümü.

Şekil 4.7’de görüldüğü üzere Y matrisi, T_2 dönüşüm matrisi ile çarpılarak ettirgen Y_2 matrisi elde edilmektedir. Sonuç olarak Y matrisinin birinci satırı Y_2 matrisinin

dördüncü satırına geçer. Burada yönelme satırında yüklemle ilgili işi doğrudan yapabilecek eski öznenin sahip olduğu kavramlar kullanılmalıdır. Görüldüğü üzere özne değişmemektedir.

Ali ekmeği aldı. → Ali ekmeği Ayşe'ye aldırdı.

4.3.5 Edilgen yüklem

Edilgenlik eki almamışken nesne diye işaretlenen öbek, yüklemın edilgenlik eki aldığı tümcelerde özne satırına geçer (sözde özne). Şekil 4.8'de görüldüğü üzere Y matrisi ile T_3 matrisinin çarpımı Y_3 matrisini vermektedir. Edilgen bir yükleme sahip tümcede belirtili yada belirtisiz nesne bulunamaz.

Ali ekmeği aldı. → Ekmek alındı.

$$\begin{bmatrix} 0110000000 \\ 0000000000 \\ 0000000000 \\ 0001000000 \\ 0000100000 \\ 0000010000 \\ 0000001000 \\ 0000000100 \\ 0000000010 \\ 0000000001 \end{bmatrix}_{T_3}^{10 \times 10} \times \begin{bmatrix} a_1 b_1 c_1 d_1 e_1 f_1 g_1 \\ a_2 b_2 c_2 d_2 e_2 f_2 g_2 \\ a_3 b_3 c_3 d_3 e_3 f_3 g_3 \\ a_4 b_4 c_4 d_4 e_4 f_4 g_4 \\ a_5 b_5 c_5 d_5 e_5 f_5 g_5 \\ a_6 b_6 c_6 d_6 e_6 f_6 g_6 \\ a_7 b_7 c_7 d_7 e_7 f_7 g_7 \\ a_8 b_8 c_8 d_8 e_8 f_8 g_8 \\ a_9 b_9 c_9 d_9 e_9 f_9 g_9 \\ a_0 b_0 c_0 d_0 e_0 f_0 g_0 \end{bmatrix}_Y^{10 \times 7} = \begin{bmatrix} a_2+a_3 & b_2+b_3 & c_2+c_3 & d_2+d_3 & e_2+e_3 & f_2+f_3 & g_2+g_3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ a_4 & b_4 & c_4 & d_4 & e_4 & f_4 & g_4 \\ a_5 & b_5 & c_5 & d_5 & e_5 & f_5 & g_5 \\ a_6 & b_6 & c_6 & d_6 & e_6 & f_6 & g_6 \\ a_7 & b_7 & c_7 & d_7 & e_7 & f_7 & g_7 \\ a_8 & b_8 & c_8 & d_8 & e_8 & f_8 & g_8 \\ a_9 & b_9 & c_9 & d_9 & e_9 & f_9 & g_9 \\ a_0 & b_0 & c_0 & d_0 & e_0 & f_0 & g_0 \end{bmatrix}_{Y_3}^{10 \times 7}$$

Şekil 4.8 : Edilgenlik eki alan yüklem için Y matrisinin dönüşümü.

4.3.6 Dönüştürme yüklem

Dönüştürme yüklemlerde olaydan etkilenen nesne ve olayı gerçekleştiren özne aynı olduğu için daha fazla bilgiyi Y matrisine aktarmak adına hem özne alanına hem de belirtisiz nesne alanına eski özneye ait kavramlar yazılmalıdır. Dönüştürme yüklemın olduğu tümcelerde nesne belirtisiz nesne olarak kullanılır, belirtili nesne kullanılmaz.

Ayşe bebeğin saçlarını taradı → Ayşe tarandı.

Örnek tümcesinde tarayan kişi Ayşe ve taranan kişi de Ayşe'dir. Şekil 4.9'da görüldüğü üzere Y matrisi ile T_4 matrisi çarpılarak Y_4 matrisi elde edilmektedir.

$$\begin{matrix}
\begin{pmatrix} 1000000000 \\ 1000000000 \\ 0000000000 \\ 0001000000 \\ 0000100000 \\ 0000010000 \\ 0000001000 \\ 0000000100 \\ 0000000010 \\ 0000000001 \end{pmatrix} & \times & \begin{pmatrix} a_1b_1c_1d_1e_1f_1g_1 \\ a_2b_2c_1d_2e_2f_2g_2 \\ a_3b_3c_3d_3e_3f_3g_3 \\ a_4b_4c_4d_4e_4f_4g_4 \\ a_5b_5c_5d_5e_5f_5g_5 \\ a_6b_6c_6d_6e_6f_6g_6 \\ a_7b_7c_7d_7e_7f_7g_7 \\ a_8b_8c_8d_8e_8f_8g_8 \\ a_9b_9c_9d_9e_9f_9g_9 \\ a_0b_0c_0d_0e_0f_0g_0 \end{pmatrix} & = & \begin{pmatrix} a_1b_1c_1d_1e_1f_1g_1 \\ a_1b_1c_1d_1e_1f_1g_1 \\ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \\ a_4b_4c_4d_4e_4f_4g_4 \\ a_5b_5c_5d_5e_5f_5g_5 \\ a_6b_6c_6d_6e_6f_6g_6 \\ a_7b_7c_7d_7e_7f_7g_7 \\ a_8b_8c_8d_8e_8f_8g_8 \\ a_9b_9c_9d_9e_9f_9g_9 \\ a_0b_0c_0d_0e_0f_0g_0 \end{pmatrix} \\
\mathbf{T}_4 \quad 10 \times 10 & & \mathbf{Y} \quad 10 \times 7 & & \mathbf{Y}_4 \quad 10 \times 7
\end{matrix}$$

Şekil 4.9 : Dönüştürücü eki alan yüklem için Y matrisinin dönüşümü.

4.4 Yüklem Sonuna Gelen Şahıs Eklerinin Etkisi

Tümcedeki tüm öbeklerin yüklemle uyumluluğunun denetimini tamamlamak için zaman bildiren öbeğin yüklemle uyumluluğu ve şahıs bildiren öbeğin yüklemle uyumluluğu çalışmamız kapsamında incelenmiştir. Türkçede yüklem aldığı şahıs ekleriyle tümcenin öznesi hakkında bilgi vermektedir ve yüklem aldığı şahıs ekinin özne ile uyumlu olması beklenmektedir. Örneğin, “Ayşe ile ben okula gittim.” tümcesinde yüklem aldığı “-m” eki ile “Ayşe ile ben” öznesi arasında uyumsuzluk olduğu görülmektedir. Doğru olan “Ayşe ile ben okula gittik.” tümcesidir. Daha önce bahsedilen öbek kavram denetimine ve çatı eklerine, yüklem aldığı ekle şahıs bildiren öbeğin uyumluluğu eklenerek dördüncü model elde edilmiştir.

4.5 Yüklem Sonuna Gelen Zaman Eklerinin Etkisi

Çalışmamızda son olarak, tümcenin zaman bildiren öbeğinin, yüklem aldığı zaman ekiyle uyumluluğunun denetlenmektedir. Bu inceleme yapılırken, zaman kavramına sahip öbekler yedi farklı biçimde kümelendirilmiştir [84].

K1: Geçmiş zamanda bir an; **K2:** Şu an; **K3:** Gelecek zamanda bir an; **K4:** Sıklık; **K5:** Geçmiş zamanda bir zaman dilimi; **K6:** Bugün; **K7:** Gelecek zamanda bir zaman dilimi

Tablo 4.1’de görüldüğü üzere “zaman” kavramı içeren her bir öbek bu kümelerden hangisine girdiğine göre işaretlenmiştir. Daha sonra Tablo 4.2’de görüldüğü üzere bir

Çizelge 4.1 : Zaman öbeklerinin sınıflandırılması.

Örnek Sözcük	Nokta Zaman			Periyodik Zaman			K4
	K1	K2	K3	K5	K6	K7	
dün 5'te	x	-	-	-	-	-	-
şu an	-	x	-	-	-	-	-
haftaya 4'te	-	-	x	-	-	-	-
evelde	-	-	-	x	-	-	-
bugün	-	-	-	-	x	-	-
yaza	-	-	-	-	-	x	-
ara sıra	-	-	-	-	-	-	x
demin	x	-	-	-	-	-	-

Çizelge 4.2 : Zaman eklerinin zaman kümelerine göre uyumluluğu.

Zaman Türü	Örnek	K1	K2	K3	K4	K5	K6	K7
Bilinen geçmiş	bildim	1	1	0	1	1	1	0
Duyulan geçmiş	bilmişim	1	1	0	1	1	1	0
Şimdiki zaman	biliorum	0	1	1	1	0	1	1
Gelecek zaman	bileceğim	0	1	1	1	0	1	1
Geniş zaman	bilirim	0	1	1	1	0	1	1
İstek kipi	bileyim	0	1	1	1	0	1	1
Şart kipi	bilsem	1	1	1	1	1	1	1
Emir kipi	bileyim	0	1	1	1	0	1	1
Gereklilik kipi	bilmeliyim	0	1	1	1	0	1	1
Hikaye bilinen geçmiş	bildiydim	1	1	0	1	1	1	0
Hikaye duyulan geçmiş	bilmiştim	1	1	0	1	1	1	0
Hikaye Şimdiki zaman	biliyordum	1	1	0	1	1	1	0
Hikaye Gelecek zaman	bilecektim	1	1	1	1	1	1	1
Hikaye Geniş zaman	bilirdim	1	1	1	1	1	1	1
Hikaye İstek kipi	bileydim	1	1	1	1	1	1	1
Hikaye Şart kipi	bilseydim	1	1	1	1	1	1	1
Hikaye Gereklilik kipi	bilmeliydim	1	1	1	1	1	1	1
Rivayet Duyulan Geçmiş	bilmişmişim	1	1	0	1	1	1	0
Rivayet Şimdiki Zaman	biliyormuşum	1	1	1	1	1	1	1
Rivayet Gelecek Zaman	bilecekmişim	1	1	1	1	1	1	1
Rivayet Geniş Zaman	bilirmişim	0	1	1	1	1	1	1
Rivayet İstek kipi	bileymişim	1	1	1	1	1	1	1
Rivayet şart kipi	bilseymişim	1	1	1	1	1	1	1
Rivayet Gereklilik kipi	bilmeliymişim	1	1	1	1	1	1	1
Şart Bilinen Geçmiş	bildiysem	1	1	0	1	1	1	0
Şart Duyulan Geçmiş	bilmişsem	1	1	0	1	1	1	0
Şart Şimdiki Zaman	biliyorsam	0	1	1	1	1	1	1
Şart Gelecek Zaman	bileceksem	0	1	1	1	0	1	1
Şart Geniş Zaman	bilirsem	0	1	1	1	1	1	1
Şart Gereklilik kipi	bilmeliysem	1	1	1	1	1	1	1

ön çalışmayla otuz bir zaman ve kip ek çeşitlerinin bu yedi farklı zaman kümesinden hangileriyle birlikte kullanılabileceği belirlenmiştir.

4.6 Uygulama Çıktısı

Çalışmamız kapsamında oluşturduğumuz uygulama bir tümcelik girişleri ya da tümcelerden oluşan metinleri giriş olarak kabul etmektedir. Şekil 4.10'da ekran arayüzünden girilen sorunsuz bir tümce için uygulama çıktısı görünmektedir.

Cümle giriniz: Ali bugün kırmızı araba aldı.

Response string:

1	Ali	Ali	Noun Prop	A3sg Pnon Nom	5	SUBJECT
2	bugün	bugün	Adverb	Adverb	5	MODIFIER
3	kırmızı	kırmızı	Adj	Adj	4	MODIFIER
4	araba	araba	Noun Noun	A3sg Pnon Nom	5	OBJECT
5	aldı	al	Verb Verb	Pos Past A3sg	0	PREDICATE
6	.	.	Punc Punc		5	PUNCTUATION

CÜMLE ÖBEKLERİ VE ANALİZİ:

Özne Bilgisi:

Öbek: ali

Aslı: ali

Türü: Noun

Kavram: insan

Bsiz Nesne Bilgisi:

Öbek: kırmızı araba

Aslı: araba

Türü: Noun|

Kavram: vasıta

Zarf Bilgisi:

Öbek: bugün

Aslı: bugün

Türü: Adverb

Kavram: zaman

Yüklem Bilgisi:

Öbek: aldı

Aslı: al

Türü: Verb

Kavram: eylem

Fill Zamanı: Past

Zaman öbeği:

Zaman Uyuşmazlığı yok!!!

Kişi eki: A3sg; 0

Özne: ali

Kişi Ekiyle Uyuşuyor...

Alt Ana Cümle: ali kırmızı araba bugün aldı

Tüm Alt Cümleler

Alt Cümle0: Ali kırmızı araba bugün aldı.

SONUÇ: Cümle uyumlu!!!!

ANALİZ TAMAMLANDI

Şekil 4.10 : Uygulama çıktısı sorunsuz tümce örneği.

Uygulama çıktısında tümceye ait öbekler,öbeğin dil bilgisi açısından türü ve hangi kavram veya kavramları içerdiği gösterilmektedir. Ayrıca zaman öbeğinin ne olduğu bu öbeğin eylem zaman ekiyle uyumlu olup olmadığı, öznenin ne olduğu, öznenin şahıs ekiyle uyumlu olup olmadığı ve en sonunda tümcenin yüklem odaklı anlamsal ve dil bilgisel uyumluluğa sahip olup olmadığı gösterilmektedir.

4.6.1 Alt tümcelere ayırma örnekleri

Alt tümceye ayırma işlemi bileşik ve basit alt tümce içeren bağlı bir tümce üzerinden Şekil 4.11’de gösterilmektedir.

```
Cümle giriniz: Ayşe tatile gitti, dinlenerek geldi.
Bağlı cümle, (İşlenen): Ayşe tatile gitti.
Bağlı cümle, (Kalan): Dinlenerek geldi .
Response string:
1  Ayşe Ayşe Noun Prop A3sg|Pnon|Nom 3  SUBJECT
2  tatile tatil Noun Noun A3sg|Pnon|Dat 3  MODIFIER
3  gitti git Verb Verb Pos|Past|A3sg 0  PREDICATE
4  . . Punc Punc _ 3  PUNCTUATION

CÜMLE ÖBEKLERİ VE ANALİZİ:

Özne Bilgisi:
Öbek: ayşe
Aslı: ayşe
Türü: Noun
Kavram: insan

E Hali Öbeği Bilgisi:
Öbek: tatile
Aslı: tatil
Türü: Noun
Kavram: bilinmiyor

Yüklem Bilgisi:
Öbek: gitti
Aslı: git
Türü: Verb
Kavram: eylem

Fiil Zamanı: Past
Zaman öbeği:
Zaman Uyuşmazlığı yok!!!

Kişi eki: A3sg; O
Özne: ayşe
Kişi Ekiyle Uyuşuyor...

Alt Ana Cümle:ayşe tatile gitti
Tüm Alt Cümleler
Alt Cümle0: Ayşe tatile gitti.
SONUÇ: Cümle uyumlu!!!!
```

Şekil 4.11 : Alt tümceye ayırma örneği (1).

Tümceler alt tümcelerine ayrılırken öncelikle tümce bağlı tümce içeriyor mu diye bakılır. Bağlı tümce "ve", "ama", "," gibi bağlaçlarla ayrılan tümcedir ve bağlaç içermeyen ilk kısmı alınır. Alınan kısım bağlı tümce kurallarına yüklem içeren bir tümcedir. Bu tümce bileşik tümce mi yani içinde isim eylem, zarf eylem yada sıfat eylem içeriyor mu diye bakılır. Bileşik tümceyse alt tümcelerine ayrılır ve ayrılan her bir basit tümce için öbek kavramı çözümlemesi yapılır zaten bağlı tümceden ayrılan kısım basit tümce ise doğrudan öbek ve kavram çözümlemesi yapılır.

```
Response string:
1      _      dinle Verb Pass Pos 2      DERIV
2      Dinlenerek      Adverb ByDoingSo      3      MODIFIER
3      geldi gel Verb Verb Pos|Past|A3sg 0      PREDICATE
4      .      .      Punc Punc      3      PUNCTUATION
```

CÜMLE ÖBEKLERİ VE ANALİZİ:

Zarf Bilgisi:
Öbek: dinlenerek
Aslı: dinle
Türü: Verb
Kavram: eylem

Yüklem Bilgisi:
Öbek: geldi
Aslı: gel
Türü: Verb
Kavram: eylem

Fil Zamanı: Past
Zaman öbeği:
Zaman Uyumsuzluğu yok!!!

Kişi eki: A3sg; O
Kişi Ekiyle Uyuşuyor...

Alt Ana Cümle: geldi

Alt Zarf Cümlesi ÖbekNo-7:

Alt zarf cümlesi: dinlendi.

Bağlılık Bilgisi: Ana eylem yan eylemin gerçekleşmesiyle gerçekleşir

Alt Cümle Bağlılık tipi: Zarf Öbeği

Tüm Alt Cümleler
Alt Cümle0: Geldi.
Alt Cümle1: dinlendi
SONUÇ: Cümle uyumlu!!!!

Şekil 4.12 : Alt tümceye ayırma örneği (2).

Bağlı tümcenin kalan kısmı için başa dönülür ve tüm adımlar tekrar edilir. Sonuç olarak her bir bağlı tümce parçası (bu ya bileşik yada basit tümcedir) ve her bir bileşik tümce parçası (basit tümcedir) öbek kavramı açısından incelenir. Tüm bu çözümleme işlemi kalan kısım boş olana kadar devam eder.

Şekil 4.12’de görüldüğü üzere "*Ayşe tatile gitti, dinlenerek geldi*" tümcesinde önce ",,"den önceki kısım alınır. Bu tümce bileşik tümce mi diye bakılır. Örneğimizde "*Ayşe tatile gitti*" basit tümcedir. Bu yüzden doğrudan kavram ve öbek çözümlemesi yapılır, yüklem odaklı dil bilgisi ve anlam olarak doğruluğu incelenir, daha sonra ilk bağlaca yada sonlanma işaretine kadar olan kalan kısım alınır. Burada "*dinlenerek geldi*" tümcesidir. Tümce bileşik tümce mi diye bakılır. İçinde dinlenerek zarf eylemi içerdiğinden bileşik tümce olduğu anlaşılır. Bileşik tümce olduğu için öncelikle "*Geldi*" ve "*Dinlendi*" olarak alt tümcelerine ayrılmış ve Şekil 4.13’de görüldüğü üzere bu iki alt tümce için çözümleme tekrarlanmıştır.

```
Response string:
1   Geldi gel   Verb Verb Pos|Past|A3sg   0   PREDICATE
2   ..   ..   Punc Punc _   1   PUNCTUATION

CÜMLE ÖBEKLERİ VE ANALİZİ:

Yüklem Bilgisi:
Öbek: geldi
Aslı: gel
Türü: Verb
Kavram: eylem

Fiil Zamanı: Past
Zaman öbeği:
Zaman Uyuşmazlığı yok!!!

Kişi eki: A3sg; O
Kişi Ekiyle Uyuşuyor...

SONUÇ: Cümle uyumlu!!!!

Response string:
1   dinlendi   dinle Verb Verb Pos|Past|A3sg   0   PREDICATE
2   ...   ...   Punc Punc _   1   PUNCTUATION

CÜMLE ÖBEKLERİ VE ANALİZİ:

Yüklem Bilgisi:
Öbek: dinlendi
Aslı: dinle
Türü: Verb
Kavram: eylem

Fiil Zamanı: Past
Zaman öbeği:
Zaman Uyuşmazlığı yok!!!

Kişi eki: A3sg; O
Kişi Ekiyle Uyuşuyor...

SONUÇ: Cümle uyumlu!!!!

ANALİZ TAMAMLANDI
```

Şekil 4.13 : Alt tümceye ayırma örneği (3).

4.6.2 Öbek kavram uyumsuzluğunu yakalama örneği

Eğer tümcede yükleme uyumsuz öbek kavram varsa sorunlu öbek gösterilir.

Şekil 4.14’de “çıkartıldı” eylemiyle belirtili nesne olan “köpeği” öbeği uyumsuzdur.

Cümle giriniz: Ali bugün köpeği sokağa çıkartıldı.

Response string:

1	Ali	Ali	Noun Prop	A3sg Pnon Nom	5	SUBJECT
2	bugün	bugün	Adverb	Adverb	5	MODIFIER
3	köpeği	köpek	Noun Noun	A3sg Pnon Acc	5	OBJECT
4	sokağa	sokak	Noun Noun	A3sg Pnon Dat	5	MODIFIER
5	çıkartıldı	çıkart	Verb Verb	Pass Pos Past A3sg	0	PREDICATE
6	.	.	Punc Punc	_	5	PUNCTUATION

CÜMLE ÖBEKLERİ VE ANALİZİ:

Özne Bilgisi:

Öbek: ali

Aslı: ali

Türü: Noun

Kavram: insan

Bli Nesne Bilgisi:

Öbek: köpeği

Aslı: köpek

Türü: Noun

Kavram: hayvan

E Hali Öbeği Bilgisi:

Öbek: sokağa

Aslı: sokak

Türü: Noun

Kavram: yer

Zarf Bilgisi:

Öbek: bugün

Aslı: bugün

Türü: Adverb

Kavram: zaman

Yüklem Bilgisi:

Öbek: çıkartıldı

Aslı: çıkart

Türü: Verb

Kavram: eylem

Fiil Zamanı: Past

Zaman öbeği:

Zaman Uyuşmazlığı yok!!!

Kişi eki: A3sg; O

Özne: ali

Kişi Ekiyle Uyuşuyor...

Alt Ana Cümle: ali köpeği sokağa bugün çıkartıldı

Tüm Alt Cümleler

Alt Cümle0: Ali köpeği sokağa bugün çıkartıldı.

SONUÇ: Cümle uyumsuz!!!

Sorunlu öbek: köpeği

ANALİZ TAMAMLANDI

Şekil 4.14 : Öbek kavram uyumsuzluğu örneği.

4.6.3 Zaman öbeği ve zaman eki uyumsuzluğunu yakalama örnekleri

Eğer tümcede yüklem ekiyle zaman öbeği arasında uyumsuzluk varsa arayüzde gösterilir. Şekil 4.15'deki örnekte "*dün*" zaman öbeği ile "*bitecek*" yüklemi arasında uyumsuzluk vardır.

Cümle giriniz: Müzik yayını dün bitecek.

Response string:

1	Müzik	müzik	Noun	Noun	A3sg Pron Nom	2	POSSESSOR
2	yayını	yayın	Noun	Noun	A3sg P3sg Nom	4	SUBJECT
3	dün	dün	Adverb	Adverb	_	4	MODIFIER
4	bitecek	bit	Verb	Verb	Pos Fut A3sg	0	PREDICATE
5	.	.	Punc	Punc	_	4	PUNCTUATION

Cümle içindeki alt cümle sayısı:1

CÜMLE ÖBEKLERİ VE ANALİZİ:

Özne Bilgisi:
Öbek: müzik yayını
Aslı: yayın
Türü: Noun
Kavram: bilinmiyor

Zarf Bilgisi:
Öbek: dün
Aslı: dün
Türü: Adverb
Kavram: zaman

Yüklem Bilgisi:
Öbek: bitecek
Aslı: bit
Türü: Verb
Kavram: eylem

Fiil Zamanı: Fut
Uyuşmayan öbek: dün
Zaman Uyuşmazlığı var!!!

Kişi eki: A3sg; O|
Özne: müzik yayını
Kişi Ekiyle Uyuşuyor...

Alt Ana Cümle:müzik yayını dün bitecek
Tüm Alt Cümleler
Alt Cümle0:Müzik yayını dün bitecek.
SONUÇ: Cümle uyumsuz!!!

ANALİZ TAMAMLANDI

Şekil 4.15 : Zaman uyumsuzluğu örneği.

4.6.4 Şahıs öbeği ve şahıs eki uyumsuzluğunu yakalama örnekleri

Eğer tümcede yüklemın ekiyle özne arasında uyumsuzluk varsa gösterilir. Şekil 4.16'deki örnekte "biz" öznesi ile "çıkıttım" yüklemi arasında uyuşmazlık vardır.

Cümle giriniz: Biz sabah denize doğru yola çıkıttım.

Response string:

1	Biz	biz	Pron	Pers	A1pl Pnon Nom	6	SUBJECT
2	sabah	sabah	Noun	Noun	A3sg Pnon Nom	3	POSSESSOR
3	denize	deniz	Noun	Noun	A3sg Pnon Dat	4	ARGUMENT
4	doğru	doğru	Postp	PCDat		6	MODIFIER
5	yola	yol	Noun	Noun	A3sg Pnon Dat	6	MWE
6	çıkıttım	çık	Verb	Verb	Pos Past A1sg	0	PREDICATE
7	.	.	Punc	Punc		6	PUNCTUATION

CÜMLE ÖBEKLERİ VE ANALİZİ:

Özne Bilgisi:

Öbek: biz

Aslı: #biz

Türü: Pron|

Kavram: insan

E Halı Öbeği Bilgisi:

Öbek: yola

Aslı: #yol

Türü: Noun

Kavram: yer

Edat Bilgisi:

Öbek: sabah denize doğru

Aslı: #doğru

Türü: Postp

Kavram: soyut mantık sıfat

Yüklem Bilgisi:

Öbek: çıkıttım

Aslı: #çık

Türü: Verb

Kavram: eylem

Fil Zamanı: Past

Zaman öbeği:

Zaman Uyuşmazlığı yok!!!

Kişi eki: A1sg; Ben

Özne: biz

Kişi Ekinde Uyuşmazlık var!!!

Alt Ana Cümle: Biz yola sabah denize doğru çıkıttım.

Tüm Alt Cümleler

Alt Cümle0: Biz yola sabah denize doğru çıkıttım.

SONUÇ: Cümle uyumsuz!!!

ANALİZ TAMAMLANDI

Şekil 4.16 : Kişi uyumsuzluğu örneği.



5. GELİŞTİRİLEN ARAÇLARIN DEĞERLENDİRİLMESİ

Geliştirilen araçlar için yapılan değerlendirilmeler ilgili bölümlerde anlatılmaktadır. Ama araçlar değerlendirilmeden önce Anlamsal ve Dilbilgisel Hataları tespit eden aracımızın tespit edeceği anlatım bozukluklarının tüm hatalar içindeki yeri belirlenmelidir.

5.1 Türkçe Anlatım Bozukluğu Olayının İncelenmesi

Günlük hayatımızdaki metinlerde anlatım bozuklukları çok sık karşımıza çıkmamaktadır. Sıklığı az da olsa eğitim görmüş ve dil konusunda özenli insanlar bile özellikle uzun tümcelerde anlatım hataları yapabilmektedir. Modelimizi değerlendirmeden önce değişik kaynaklardan aldığımız tümceler üzerinde anlatım bozukluklarının günlük hayatımızda hangi sıklıkla görüldüğü incelenmiştir.

5.1.1 Kullanılan kaynaklar ve özellikleri

Anlatım bozukluğunun resmi yazılara göre daha sık rastlandığı "*Film alt yazıları*", anlatım bozukluğunun neredeyse hiç rastlanmadığı "*Akademik yayınlar*" ve anlatım bozukluğunun az rastlandığı "*Vikipedi*"den alınan 300 tümcelik eşit boyda metin elle sınıflandırılmıştır. Çizelge 5.1'de görüldüğü gibi bu metinler; tümcelerin içerdiği ortalama sözcük sayısı, tümcelerin içerdiği ortalama hal eki ve zarf eylem eki (-i, -e, -de, -den, -le, ip, -arak, -ınca vb.) sayısı, tümcelerin içerdiği ortalama öbek sayısı ve tümce türüne göre incelenmiştir.

Sonuç olarak günlük hayattaki metinlerde tümcedeki ortalama sözcük sayısı 10 ve ortalama ek sayısı 6 olarak bulunmuştur. Tümceler ortalama olarak yüklem dışında 3 öbek daha içermektedir. Ortalama %51'lik bir oranla en sık karşılaşılan tümce türü bileşik tümce ve %14,25'lik bir oranla en az rastlanan tümce türü bağlı tümcedir.

Çizelge 5.1 : Günlük metin incelemesi.

Kaynak Metin	Ortalama Sözcük Sayısı	Ortalama Ek Sayısı	Ortalama Öbek Sayısı	Tümce Türü
Film Alt Yazıları	7.93	5.3	2.31	%58,32 Basit %36,37 Bileşik %5,31 Bağlı
Akademik Yayınlar	13.2	7.5	3.25	%27,21 Basit %60,55 Bileşik %12,24 Bağlı
Vikipedi	11.3	5.83	2.58	%18,74 Basit %56,12 Bileşik %25,14 Bağlı

5.1.2 Anlatım bozukluğu türleri ve görülme sıklığı

Çizelge 5.2’de anlatım bozukluğu çeşitlerinin, günlük hayatta kullanılan metinlerde ve İngilizceden Türkçeye Google çeviri aracıyla çevrilen metinlerde görülme sıklığı karşılaştırılmıştır.

Çizelge 5.2 : Anlatım bozukluğu türleri incelemesi.

Anlatım Bozukluğu Türleri	Günlük Metin %	Google Çeviri %
Sözcük Temelli Hatalar (STH)		
- Yanlış sözcük kullanımı	2,68	5
- Yanlış sözcük sıralaması	1,33	3
- Fazla sözcük kullanımı	0,61	2
STH Toplam oranı	4,62	10
Yapısal Hatalar (YH)		
- Bağlaçlar ve iç tümcelerde hata *	0,83	15
- İsim tamlamaları hataları	0,92	6
- Mantıksızlık	0,54	13
- Öbek kavram çiftlerinin yüklem uyumsuzluğu *	2,12	28
- Kişi ve zaman öbeklerinin yüklem uyumsuzluğu *	1,91	9
- Çoğul ekinin yanlış kullanımı *	1,43	6
YH Toplam oranı	7,75	77
Toplam anlatım bozukluğu oranı	12,37	87

Çizelge 5.2’de görüldüğü üzere, günlük hayatta kullanılan metinlerde sadece %12,37 oranında anlatım bozukluğu görülmektedir ve yapısal anlatım bozuklukları, sözcüksel anlatım bozukluklarının yaklaşık 1,5 katıdır.

Günlük metinlerdeki düşük anlatım bozukluğu oranının aksine, Google çevirisinde çeviri tümcelerinin %87'si anlatım bozukluğu içermektedir. Ayrıca bu tümcelerdeki yapısal anlatım bozuklukları sözcüksel anlatım bozukluklarının 7 katından fazladır. Sonuç olarak günlük metinlerin %87,63 düzgün ifadelerden oluşmaktadır. Google çevirinde, düzgün ifadelerin oranı %13 olarak bulunmuştur.

Çizelge 5.2'de görüldüğü gibi çalışmamız "*" ile işaretlenen "*öbek kavram çiftlerinin yüklem uyumsuzluğu*", "*Çoğul ekinin yanlış kullanımı*", "*kişi ve zaman öbeklerinin yüklem uyumsuzluğu*" ve "*bağlaçlar ve iç tümceler*" maddelerini kapsamaktadır. Bu da günlük metinlerdeki yapısal hataların %81,16'sına eşittir.

5.2 "Alt Tümce Ayırıcının" ve "Öbek Kavram Bulucunun" değerlendirilmesi

Çalışmamız kapsamında oluşturulan Alt tümce ayırıcı, tümce öbek bulucu ve kavram bulucu araçlarının doğruluğu değerlendirilecektir.

Başarım ölçülürken; basit, bileşik ve bağlı rastgele 1000 tümce geliştirilen araçlar tarafından işlenerek değerlendirilmiştir. Her aşamada doğruluk oranlarını bulunmuştur. Hatasız ve hatalı sonuçlar incelenmiştir. Hatalı işaretlenen kısımların kaçta kaçının geliştirilen araçtan kaynaklandığı, kaçta kaçının İTÜ DDİ Bağlılık Çözümleyici aracından kaynaklandığı hesaplanmıştır. Sonuçlar ilgili çizelgelerde gösterilmiştir.

5.2.1 "İTÜ DDİ Bağlılık Çözümleyicinin" değerlendirilmesi

Çalışmamızda hem alt tümce ayırıcımız, hem öbek kavram bulucumuz hem de yüklem odaklı anlamsal ve dil bilgisi çözümleyicimiz İTÜ DDİ Bağlılık Çözümleyicinin çıktılarını giriş olarak kullanmaktadır. Bu nedenle çalışmamızın doğruluğunu incelemek için İTÜ DDİ Bağlılık Çözümleyicinin tümce başına doğruluğu da bizim için önem taşımaktadır.

Mayıs 2016'da güncellenmiş olduğu internet sayfasında yayınlanan DDİ Bağlılık çözümleyici aracının tümce bazında doğrulukları incelenmiştir. Bu güncelleme ile birlikte tümce içindeki tüm parçaların (tokenların) işaretleme ve bağlamalarının doğru olması şeklindeki bağlılık çözümleyicinin tümce için doğru sonuç vermesi yüzdesi %38,7'lerden %56,8'lere çıkmıştır. Toplam %29 oranında görünen yüklem bulamama

"ROOT" diye işaretleme hatası ve ağ kaynaklı bir tümceye başka bir tümcenin bağıllık bilgisini gönderme gibi sorunlar ortadan kalkmıştır.

DDİ Bağıllık çözümleyiciyi tümce bazında değerlendirdiğimiz için Bağıllık çözümleyici bazen ayrıntılarda hata verdiğinde hata küçük bile olsa oluşturduğumuz araçları etkileyebilmektedir. Ana bağıllıklar doğru olmasına karşın bir sözcüğü yanlış bir sözcüğe bağlayabilmekte (*DDİ Bağıllık Çözümleyici sözcük bağlantısında hata*) ya da özne, nesne gibi sözcüğün türünü ya da sıfat, zarf gibi özelliğini yanlış işaretleyebilmektedir (*DDİ Bağıllık Çözümleyici sözcük işaretlemesinde hata*). Bir tümce için bağıllık çözümleyici hem tüm bağlantılarında hem de işaretlemeye doğrusu hatasız olarak değerlendirilmektedir. Çizelge 5.3'de görüldüğü üzere 178 tane örnekte hem bağlantıda hata hem de işaretlemeye hata bulunmaktadır.

Çizelge 5.3 : "İTÜ DDİ Bağıllık Çözümleyicinin" tümce başına doğruluk değerleri.

Tür	Sayısı
Tümce için DDİ B. Ç. hatasız	568
Tümce için DDİ B. Ç. sadece sözcük işaretlemesinde hata	105
Tümce için DDİ B. Ç. hem işaretlemesinde hem bağlantıda hata	178
Tümce için DDİ B. Ç. sadece sözcük bağlantısında hata	149

Sonuç olarak elimizdeki tümceler için %56,8 tümcede bağıllık çözümleyici doğru çalışmaktadır. %14,9 tümcede sadece bağlantıdan kaynaklı hataya rastlanmıştır. %10,5 tümcede sadece işaretlemeye kaynaklı hataya rastlanmıştır. Tümce başına hem bağlantı hatası hem de işaretleme hatasının bir arada görülme değeri %17.8 olarak bulunmuştur.

5.2.2 "Alt Tümce Ayırıcının" değerlendirilmesi

DDİ Bağıllık Çözümleyiciden geçen tümce öncelikle bağlı alt tümcelere ayrıştırılmakta ve her bir parça, içerdiği isim eylem, zarf eylem ve zarf eylem kadar alt tümceye ayrıştırılmaktadır. Çizelge 5.4'de görüldüğü üzere değerlendirme aşamasında tüm alt tümceleri doğru ayrılan tümce "*Doğru Ayrılan Tümce Sayısı*" olarak ifade edilmektedir. Bir tümcenin 3 alt tümcesi varsa ve 2 tanesi doğru ayrılmış ve bir tanesi yanlışsa doğru ayrılan 2 alt tümceyi "*Doğru Ayrılan Alt Tümce Sayısı*" içinde değerlendirmekteyiz. Alt tümce ayırıcında, "*DDİ Bağıllık Çözümleyici Kaynaklı*

Yanlış Ayırma" veya *"Tümce Aracımız Kaynaklı Yanlış Ayırma"* sayılarında ayrıca incelenmektedir.

Çizelge 5.4 : Alt tümce ayırma doğruluk değerleri.

Hata Türü	Sayısı
Doğru Ayrılan Tümce Sayısı	580
Doğru Ayrılan Alt Tümce Sayısı	903
DDİ Bağlılık Çözümleyici Kaynaklı Yanlış Ayırma	290
Tümce Ayırma Programımız Kaynaklı Yanlış Ayırma	129

Çizelge 5.4'de görüldüğü üzere alt Tümce Ayırma 1000 örnekle sınıandığında 580 tümcenin tüm alt tümceleriyle doğru ayrıldığı görülmektedir. 290 tane tümce DDİ bağlılık çözümleyici kaynaklı yanlış ayrılmıştır, toplam 129 tane tümce tümce ayırma programımız kaynaklı yanlış ayrılmıştır. Bir tümcenin birden fazla alt tümcesi olduğu için ve bazı hatalı ayrılan tümcelerin doğru ayrılan alt tümcelerinin olduğu için toplam 903 tane doğru alt tümce vardır.

Sonuç olarak doğru ayrılan tümce sayısı %58'dir. *"DDİ Bağlılık Çözümleyici Kaynaklı Yanlış Ayırma* %29 ve *"Tümce Aracımız Kaynaklı Yanlış Ayırma"* %13'dür.

DDİ bağlılık çözümleyicinin doğru çalıştığı durumlar için Tümce Ayırma Aracı'mız %81,6 doğru sonuç vermektedir.

5.2.3 "Öbek Bulucunun" değerlendirilmesi

Tezimizde tümceler on öbek olarak incelenmektedir. Çizelge 5.5'de görüldüğü üzere eğer bir tümcenin tüm öbekleri doğruysa *"Tüm Öbekleri Doğru Tümce"* olarak ifade edilir, Değerlendirme aşamasında *"DDİ Bağlılık Çözümleyici Kaynaklı Öbekte Yanlışlık Olan Tümce* sayısı ve *"Öbek Bulucu Aracımız Kaynaklı Öbekte Yanlışlık Olan Tümce* sayısı incelenmiştir.

Çizelge 5.5 : Öbek bulmada doğruluk değerleri.

Hata Türü	Sayısı
Tüm Öbekleri Doğru Bulunan Tümce	566
Tüm Öbekleri Doğru Alt Tümce	790
DDİ Bağlılık Çözümleyici Kaynaklı Öbekte Yanlışlık Olan Tümce	356
Öbek Bulucu Aracımız Kaynaklı Öbekte Yanlışlık Olan Tümce	77

Çizelge 5.5’de görüldüğü üzere alt Tümce Ayırma 1000 örnekle sınıdığında 566 tümce tüm öbekleriyle ve alt tümcelerinin öbekleriyle doğrudur. 356 tümcede en az bir alt tümcesinde DDİ Bağlılık Çözümleyici kaynaklı öbek hatası vardır. 77 tümcede en az bir alt tümcesinde "*Öbek Bulucu*" programımız kaynaklı öbek hatası vardır.

Sonuç olarak, tüm öbekleri doğru bulunan tümce sayısı %56,6’dır. "*DDİ Bağlılık Çözümleyici Kaynaklı Yanlış Ayırma* %35,6 ve "*Tümce Aracımız Kaynaklı Yanlış Ayırma* %7,7’dir. Toplam 790 tane alt tümcenin öbekleri doğru bulunmuştur. DDİ bağlılık çözümleyicinin doğru çalıştığı durumlar için "*Öbek Bulucu*" aracımız %89 doğru sonuç vermektedir.

Çizelge 5.6’de her bir öbek türü için doğruluk değerleri gösterilmektedir. Aracımız kaynaklı hataların daha çok özne, nesne kaynaklı olduğu görülmektedir.

Çizelge 5.6 : Her bir öbek türü için doğruluk.

Öbek	Doğru Öbek	DDİ B. Ç. Öbek Yanlış	Aracımız Kaynaklı Öbek Yanlış
Özne	89	48	26
B’siz nesne	35	32	17
B’li nesne	58	46	16
-e hali	84	21	0
-de hali	23	22	0
-den hali	36	15	1
-le hali	7	12	0
Zarf	93	37	3
Nit. zarfı	20	14	5
Edat	10	18	3
Yüklem	111	91	6

5.2.4 "Kavram Bulucunun" değerlendirilmesi

Tezimizde tümceler elli bir kavram olarak incelenmektedir. Çizelge 5.7’de görüldüğü üzere eğer bir tümcenin tüm öbekleri doğruysa "*Tüm Kavramları Doğru Tümce*" olarak ifade edilir, Değerlendirme aşamasında kavramları yanlış bulunan tümce sayıları da incelenmiştir.

DDİ Bağlılık Çözümleyicinin tümünden hatalı sonuç vermediği tümceler için, tüm kavramları doğru bulunan tümce sayısı %56,5’dir. *DDİ Bağlılık Çözümleyici kaynaklı kavram hatası* %25,8 ve "*Öbek Bulucu aracımız kaynaklı kavram hatası*" %17’dir.

Çizelge 5.7 : Kavram bulmada doğruluk tablosu.

Hata Türü	Sayısı
Tüm Kavramları Doğru Bulunan Tümce	565
Tüm Kavramları Doğru Alt Tümce	951
DDİ Bağlılık Çözümleyici Kaynaklı Kavramda Yanlışlık Olan Tümce	258
Öbek Bulucu Aracımız Kaynaklı Kavramda Yanlışlık Olan Tümce	177

DDİ bağlılık çözümleyicinin doğru çalıştığı durumlar için kavram bulucu aracımız **%82,8** doğru sonuç vermektedir.

Aracımız öbekleri kavram sözlük listelerinde aratmakta eğer kavramı arattığı listede bulursa sözcüğü o kavramla etiketlemektedir. Eğer sözcük birden fazla kavram listesinde bulunuyorsa her bir listedeki kavram o öbeğe atanmaktadır. Böylece bir öbeğe birden fazla kavram atanabilmektedir. Değerlendirmemizde atanan kavramlar içinde doğru kavram da varsa kavramın doğru atandığı kabul edilmiştir. Eğer atanan kavramlar arasında doğru kavram yoksa öbeğin kavramının yanlış olduğu dolayısıyla tümce kavram atamasının yanlış olduğu düşünülmüştür.

5.2.5 Genel değerlendirme

Çizelge 5.8’den görüldüğü üzere şu anki doğruluk oranları ortalama **%57** civarındadır. DDİ Bağlılık Çözümleyicinin sadece doğru çalıştığı örnekler göze alındığında doğruluk oranı tümce ayırma ve kavram bulmada **%80**’lere öbek bulmada **%89**’lara ulaşmaktadır. DDİ Bağlılık Çözümleyicinin sorunlarının giderileceği bir gelecekte aracımız kullanılarak daha anlamlı sonuçlar elde edilebilecektir.

Çizelge 5.8 : Doğruluk oranları.

	Doğru	DDİ BC K. Yanlışlık	Aracımız K. Yanlışlık	DDİ BC Doğru İken Başarı Yüzdesi
Tümce Ayırma	%58	%29	%13	%81,6
Öbek Bulma	%56,6	%35,6	%7,7	%89
Kavram Bulma	%56,5	%25,8	%17	%82,8

5.3 Anlamsal ve Dilbilgisi Çözümleyicinin Değerlendirilmesi

Anlamsal ve Dilbilgisi Çözümleyicinin Değerlendirilmesi DDİ Bağlılık Çözümleyicinin hatasız çalıştığı 500 tümce üzerinde yapılmıştır. Türkçe Radyo Televizyon

(TRT) Kurumu kayıtlarından elde edilen yapısal anlatım bozukluğuna sahip 250 tümce ve yine TRT kayıtlarından elde edilen 250 sorunsuz tümce kullanılmıştır. İlk aşamada tümce alt tümcelerine ayrılmadan modeller sınanmıştır. Yani ilk aşamada, "*Sabah başlayan yağmur gün boyu dinmedi.*" Tümcesinde "*Sabah başlayan yağmur*" bir özne sıvı çifti olarak ele alınmış ve ayrıştırılmamıştır. Çalışmamızda her bir etmenin katkısını ayrı ayrı ölçebilmek amacıyla beş ayrı model oluşturulmuştur.

M1 modelinde yüklem tümce içindeki öbeklerle uyumluluğuna bakılmaktadır.

M2 modelinde öbek kavram çiftlerinin yüklemle uyumluluğuna bakılmaktadır.

M3 modelinde, yüklem aldığı çatı eklerinin etkisi ikinci modelle birleştirilmiştir.

M4 modelinde, üçüncü modele ek olarak yüklem aldığı kişi ekleriyle özne uyumluluğunun denetimi yapılmaktadır ve son olarak,

M5 modelinde yüklem zaman eklerinin zaman öbekleriyle uyumluluğunun denetimi dördüncü modele eklenmiştir.

Çizelge 5.9 : Bütün tümce için modelin başarımı.

	P	R	F	A
M1	0,58	0,87	0,69	0,58
M2	0,59	0,85	0,70	0,60
M3	0,61	0,86	0,72	0,63
M4	0,65	0,84	0,73	0,66
M5	0,66	0,84	0,74	0,68

Modellerin hata bulmada gösterdiği başarı; Tutturma değeri (P), Bulma değeri (R), F-değeri (F) ve Doğruluk (A) değerleri olarak Çizelge 5.9'de gösterilmektedir. Çizelgeden görüldüğü üzere öbek yüklem uyumunun denetlendiği birinci model üzerine eklenen her bir etmen doğruluk değerimizi artırmıştır. M3 ve M4 modelleri %3 artış sağlarken, M2 ve M5 modelleri %2 artış sağlamıştır.

Bir sonraki aşamada, tümce alt tümcelerine ayrılmış yani yukardaki örnek üzerinden devam edersek; "*Sabah başlayan yağmur gün boyu dinmedi.*" tümcesi "*Sabah yağmur başladı*" ve "*Yağmur gün boyu dinmedi*" diye iki alt tümceye ayrıştırılmıştır. Her tümceye ait tüm alt tümceler değerlendirilerek model başarımları ölçümü tekrarlanmıştır. Eğer bir tümcenin tüm alt tümceleri yüklem öbek uyumluluğu içeriyorsa o tümce

doğru, eğer alt tümcelerinden herhangi biri yüklem öbek uyumluluğu içermiyorsa yanlış olarak değerlendirilmiştir.

Çizelge 5.10 : Alt tümceler için modelin başarımı.

	P	R	F	A
M1	0,58	0,92	0,71	0,60
M2	0,61	0,90	0,72	0,63
M3	0,63	0,94	0,75	0,67
M4	0,61	0,92	0,73	0,64
M5	0,64	0,95	0,76	0,69

Alt tümcelerin de değerlendirildiği durumda, Çizelge 5.10'de öbek-yüklem uyumunun denetlendiği birinci model üzerine eklenen *"kişi ekleriyle özne uyumluluğu"* haricindeki her bir etmen doğruluk değerimizi artırmıştır. M2 modeli %3, M3 modeli %4 ve M5 modelleri %5 artış sağlamıştır.

Burada tümceler sadeleştiğinden dolayı dil bilgisi ve anlamsal hata daha kolay bulunacakmış gibi görünse de tümce ayırmadan kaynaklı sorunlar için içine girmiştir.

Akademik yayınlar arasında dil bilgisi çözümleme konusunda değişik çalışmalar mevcuttur fakat bu çalışmaların birbiriyle mantıklı bir şekilde kıyaslanması zordur. Bu çalışmalar aynı derlem üzerinde yapılırsa dahi (biz Türkçe için bir çalışma yapmaktayız) eğitim verilerindeki hatalı tümcenin doğrusunun ne şekilde işaretlendiği ve aynı tümcenin farklı şekillerde ifade edilebilmesi başarı değerlerini etkilemektedir (Chodorow ve ark. 2012). Türkçe dil bilgisi alanında yapılan çalışmalar ise yazım hatalarının düzeltilmesi ile ilgili olduğundan çalışmamızla kesişmemektedir.



6. ÖRNEK BİR UYGULAMADA TÜMCE MATRİSİ

Çalışmamız tümcelerin öbek kavram şeklinde kaba anlam gösterimini oluşturmaktadır. Bu öbek kavram gösteriminin yanı sıra yüklem anlamsal sınıfı da eklendiğinde tümceyi anlamsal olarak temsil eden vektörel bir gösterimdir. Yüklem anlamsal sınıfı ile kastedilen her bir yüklemi içerdiği anlama göre kümeleyen İngilizcedeki Levin (1993) sınıflarının [85] Türkçeye uyarlanmış halidir. Levin sınıfları, iletişim bildiren eylemler, sahiplik değiştiren eylemler ve hareket içeren eylemler gibi eylem sınıflarını içermektedir.

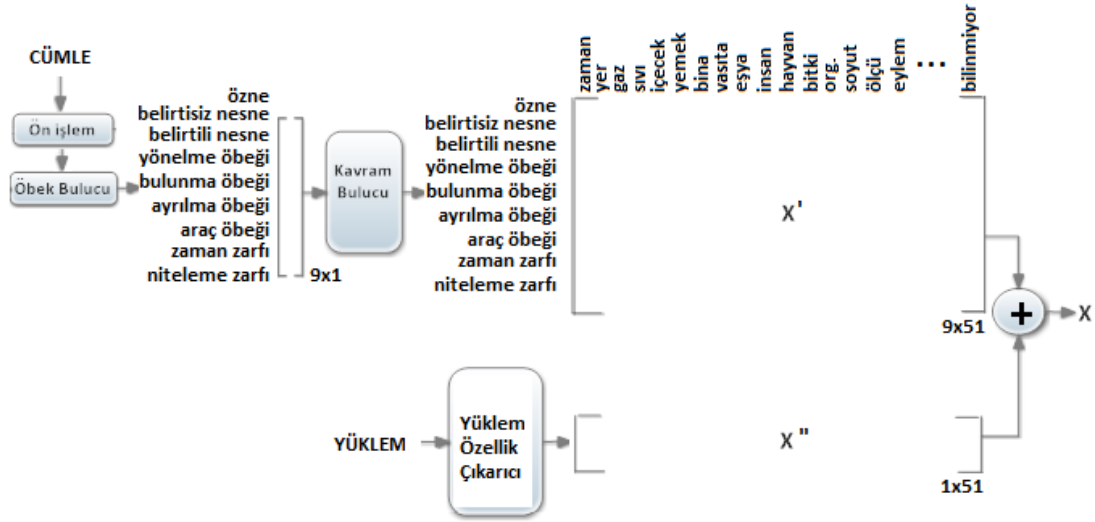
Tümcelerin matris formundaki kaba anlam gösterimi belge sınıflama, soru cevap sistemleri, bilgi çıkarımı ve özet çıkarımı gibi pek çok alanda kullanılabilecek bir yapıdadır. Çalışmamızda oluşturduğumuz tümce matrisi, belge sınıflandırma uygulamasında kullanılarak, belge sınıflandırma uygulamasına katkısı incelenmiştir.

Son yıllarda K-en yakın komşu [86] [87], Naive Bayes [88], Karar Ağacı [89], sinir ağları [90] [91], üretimsel olasılık sınıflandırıcı [92], çoklu regresyon modelleri [87] ve karar destek makinaları [93] [94] gibi değişik makina öğrenmesi yöntemleri belge sınıflandırma uygulamalarında kullanılmıştır. Türkçe içinde belge sınıflandırma alanında değişik uygulamalar mevcuttur [95], [96], [97], [98] ve [99]. Çalışmamızda %96,25'lik yüksek bir başarımla Yıldız ve ark. tarafından farklı makina öğrenmesi yöntemleri kullanılarak gerçekleştirilen; veriyi paylaşarak bir kıyaslama ortamı oluşturan çalışma verileri kullanılmıştır [100].

6.1 Temel Model

Belge sınıflandırma uygulamasında kullanılan temel modelde, metin veya tümce İTÜ DDİ Bağlılık Çözümleyicisinde işlenmektedir. Bağlılık çözümlemesi yapılmış tümceler bizim giriş tümcelerimizi oluşturmaktadır. Şekil 6.1'de görüldüğü üzere bağlılık çözümlemesi yapılmış tümceler önce alt tümce ayırıcıdan geçirilmekte basit tümcelere ayrılmaktadır. Her bir basit tümce için öbekler bulunmaktadır. Bulunan her bir öbeğe ait kavram belirlenmekte ve 9×51 'lik X' öbek kavram matrisi

oluşturulmaktadır. Bazı öbeklerin içerdiği isimler eş sesli özellikte olup birden fazla kavram listesinde bulunabilir. Burada tüm kavramlar dikkate alınmaktadır. O tümcenin gerçek kavramının, metin içinde geçen aynı kavrama sahip diğer sözcüklerle birlikte karşılaşımla sıklığının artacağı varsayılmaktadır. Örneğin belgede "kanun" sözcüğü geçiyorsa bunun müzikle mi yoksa hukukla mı ilgili olduğu belge içinde geçen diğer sözcüklerin kavram sınıflarına göre belli olacaktır. Böylece tümcedeki her bir öbek kavram ikilisi için varlık/yokluk bilgisi oluşturulmaktadır.



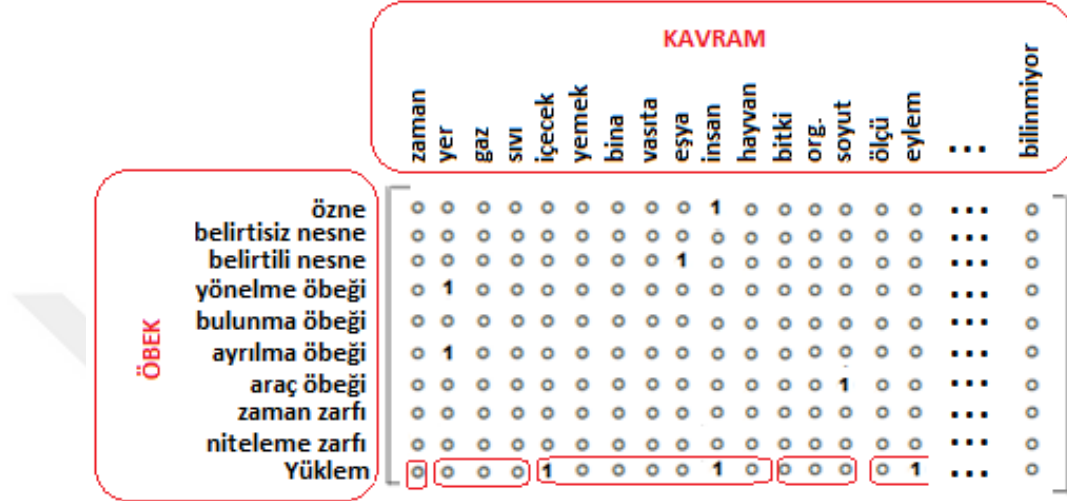
Şekil 6.1 : Tümce kaba anlamsal temsil modeli.

Paralel olarak, tümcenin yüklemi de özelliklerine göre ikilik sistemde kodlanmaktadır. Kodlamada ilk olarak yüklem kökü isim ya da eylem mi diye bakılmaktadır. Eğer yüklem isim köklüyse, bu isim kökün 51 kavram listesinden hangisiyle ilişkili olduğu bulunmakta ve ilişkili kavram listesi numarası ikili olarak kodlanmaktadır. Eğer eylem köklü bir yükleme sahipse, eylemin hangi Levin benzeri "Anlamsal Eylem Sınıfı"na ait olduğu bulunmaktadır [18]. Çatı eki türü, zaman ve kip ekleri türü, şahıs eki türü ikilik olarak kodlanarak 1x51 X'' matrisi elde edilmektedir.

Örneğin tümcenin yüklemi "doldurdum" ise bu yüklem eylem köklüdür. Levin sınıflarından "koymak (put verbs)" sınıfına aittir. Zamanı geçmiş zamandır. Şahıs eki olarak birinci tekil kişiyi bildirmektedir. Çalışmamızda türkçe eylemler Levin sınıfları göz önüne alınarak elle sınıflandırılmıştır. Sonuçta 9x51'lik X' matrisi ve 1x51'lik X'' matrisi 10x51'lik X matrisini oluşturmak üzere birleştirilir. X matrisi her bir alt tümce için kaba anlamsal tümce gösterimini vermektedir.

6.1.1 Basit tümcenin matris gösterimi

Bu çalışmada birleşik, bağlı tüm cümleler öncelikle basit tümceye dönüştürülür. Basit tümceler için matris gösterimi oluşturulurken, matrisin her bir elemanı 0 ya da 1'den oluşmaktadır. İlk dokuz satır için; 0 öbek kavram çiftinin yokluğunu, 1 öbek kavram çiftinin varlığını belirtir. Son satır yüklem özellikleri ile ilgilidir.



Şekil 6.2 : Tümcenin matris temsili.

Şekil 6.2’de görüldüğü üzere, “Ayşe” özne öbeğinde ve insan kavramında, “kalemimi” belirtili nesne öbeğinde ve eşya kavramında, “evinden” kaynak öbeği ve yer kavramında, “okula” hedef öbeği ve yer kavramında ve “sevinçle” araç öbeğinde ve soyut kavramındadır. Yüklem öbeği kategorisi de son satır olarak matrise eklenmiştir. Bu son satırın ilk elemanı yüklem kökünün isim ya da eylem olmasıyla ilgilidir.

Çizelge 6.1 : Yüklem bilgisinin vektör gösterimi.

Fiil Kökü	Tür No	Levin SınıfNo	Çatı Türü	Zaman Türü	Özne Türü	Diğer Öz.
0	000	XXXXXXXX	XXX	XXXXX	XXX	00...
İsim Kökü	Tür No	Kavram No	Çatı Türü	Zaman Türü	Özne Türü	Diğer Öz.
1	001	XXXXXXXX	XXX	XXXXX	XXX	00...
1	010	XXXXXXXX	XXX	XXXXX	XXX	00...
1	011	XXXXXXXX	XXX	XXXXX	XXX	00...
1	100	XXXXXXXX	XXX	XXXXX	XXX	00...

Çizelge 6.1’de görüldüğü üzere, eğer yüklem kökü isim ise, ikinciden dördüncüye kadar elemanlar ismi yükleme dönüştüren özel eklerin türünü belirtmektedir. Burada eğer yüklem kökü isimse ve kök “-len” eki almışsa, bu köke “sahip olmak” anlamı

katar ve “001” bitleri ile ifade edilir; kök “-leş” eki almışsa, bu köke “*olmak, dönüşmek*” anlamı katar ve “010” bitleri ile ifade edilir, kök “-le” eki almışsa, bu köke “*sahip etmek*” anlamı katar ve “011” bitleri ile ifade edilir. Eğer kök isimse ve bu özel eklerden hiçbirini almamışsa “100” bitleri ile ifade edilir. Beşinciden onbirinci bite kadar olan kısım isim olan kökün hangi kavram listesine ait olduğu ile ilgilidir. Örneğin tümcenin yüklemi “*hüzünlendim*” ise kökü “*hüzün*”dür; bu da “*soyut*” kavram listesi sınıfındadır. Türkçede kökü isim olan yüklemelerde çatı eki, zaman eki ve şahıs eki alabilmektedir. on ikinci bitten on dördüncü bite kadar olan elemanlar çatı eki türüyle, on beşinci bitten on dokuzuncu bite kadar olanlar zaman ve kip eki türüyle ve yirminci bitten yirmi ikinci bite kadar olan elemanlar şahıs eki türüyle ilişkilidir.

Eğer yüklem kökü eylem ise, beşinci bitten on birinci bite kadar olan elemanlar Levin Anlamsal eylem sınıflarıyla, on ikinci bitten on dördüncü bite kadar olanlar çatı eki türüyle, on beşinci bitten on dokuzuncu bite kadar olanlar zaman ve kip eki türüyle, yirminci bitten yirmi ikinci bite kadar olan elemanlar şahıs eki türüyle ilgilidir.

Uygulama “*Politika*”, “*Spor*”, “*Sağlık*”, “*Ekonomi*” ve “*Magazin*” sınıflarına sahip belge listelerinin sınıflandırılması amacıyla. Her bir sınıftan 230 makale içeren toplam 1150 makale arşivi için denetimli öğrenme yöntemi kullanılmıştır. Bu veri Türkçe Metin Sınıflandırma Uygulamasında kullanılmış olan ve bildiğimiz kadarıyla en yüksek başarı değerinin alındığı veridir [100].

6.1.2 Sınıflandırma ve özellik seçme algoritmaları

Bu çalışmada, sınıflandırma problemi için WEKA Naïve Bayes, LibSVM, budanma uygulanmayan bir çeşit karar ağacı olan J48 C 4.5 ve K- En Yakın Komşu sınıflandırma algoritmaları kullanılmıştır. Her bir sınıflandırma algoritmasında kullanılan WEKA parametreleri Çizelge 6.2’de gösterildiği gibi Naive Bayes için standart ayarlar kullanılmıştır. K- en yakın komşu algoritmasında 5 sınıf için öklid uzaklığı kullanılmıştır. Lib-SVM için doğrusal SVM kullanılmıştır. Karar ağacı için budanmamış karar ağacı 0,25 güven kasayısı kullanılmıştır.

Özellik seçim algoritması olarak “*Chi square*” özellik çıkarım modeli “*ranker search*” yöntemi ile birlikte kullanılmıştır. Başarım hesabında 10 katlı çapraz doğrulama (10-fold cross validation) kullanılmıştır.

Çizelge 6.2 : Sınıflandırma algoritmaları parametreleri.

WEKA NaiveBayes Parametreleri	
Modelin eski formatta gösterimi	Yok
Kernel Estimator	Yok
Denetimli Ayırıştırma (Discretization)	Yok
WEKA lazy.IBk (KNN) Parametreleri	
K (Sınıf sayısı)	5
Öklid uzaklığı	Evet
Uzaklıkların ağırlıklandırılması	Yok
Pencere boyu	Yok
WEKA Lib-SVM Parametreleri	
SVM Tipi	C-SVC
Kernel fonksiyon tipi	linear: $u' * v$
Kernel fonksiyon küme derecesi	3
WEKA J48 C 4.5 Parametreleri	
Budanmamış	Evet
Güven katsayısı	0,25
Hatalı Budamanın Azaltılması	Yok
Uç düğüm başına min örnek sayısı	2

6.1.3 Deneyisel sonuçlar

Deneyimizde, oluşturduğumuz Türkçe kaba anlamsal matris gösterim modelinin belge sınıflarını belirlemede faydalı bir yaklaşım olup olmadığı sınanmaktadır.

Sınıflandırma için, dört değişik sınıflandırma algoritması ve üç farklı veri kümesi kullanılmıştır. Bugüne kadar yapılmış en iyi modelde Yıldız ve arkadaşları 39699 boyutlu bir veriyi bazı eşitliklerle kullanmıştır. Bizim oluşturduğumuz model ise belgelerin içerdiği beş alt tümcenin kaba anlamsal matris gösterimi olup 2550 boyutludur. Her bir boyut matrisin bir elemanı ile ilişkilidir. Üçüncü modelde eski en iyi modele kaba anlamsal matris gösterimi eklenmiştir. Standart performans tüm örneklerin hedef sınıf olarak işaretlendirildiğindeki başarı yüzdesini göstermektedir.

DataSet-I veri kümesi Türkçe belge sınıflandırma alanında 94,75 % ile en yüksek başarıya değerine sahip veri kümesidir. Bu başarıyı SVM algoritması kullanılarak sağlanmıştır. Dataset-II veri kümesi belgeler içindeki en çok kavram içeren 5 alt tümcenin kaba anlamsal matris gösterimini kullanır ve matris elemanları veri kümesinin özelliklerini oluşturur. Her bir alt tümce $51 \times 10 = 510$ boyutla gösterim

edildiğinden 5 alt tümce için 2550 boyut kullanılır. DataSet-III veri kümesi DataSet-I ve DataSet-II veri kümelerinin birleşimidir.

Üç değişik veri kümesi ve modeli için başarımları Çizelge 6.3’de görülmektedir.

Çizelge 6.3 : Farklı veri kümeleri ve modeller için başarımları.

standart performans:16,6%	C 4.5 %	NB %	SVM %	KNN %
DataSet-I (39699 özellik)	76	92,25	94,75	65
DataSet-II (2550 özellik)	72,5	87,60	85,84	54,4
DataSet-III (42249 özellik)	81	97,12	96,54	68,5

“*Chi square*” Özellik seçim algoritmasının “*ranker search*” yöntemi ile birlikte DataSet-II veri kümesine uygulanmasından sonra başarımları Çizelge 6.4’de görülmektedir. Özellik seçim algoritması ile boyut 145’e indirilirken % 86,10 başarımları sağlanmıştır.

Çizelge 6.4 : Dataset-II’ye özellik seçimi uygulandığında başarımları.

standart performans:16,6%	C 4.5 %	NB %	SVM %	KNN %
DataSet-II (2550 özellik)	72,5	87,60	85,84	54,4
DataSet-II (145 özellik)	71,2	86,10	83	51,24

“*Chi square*” Özellik seçim algoritmasının “*ranker search*” yöntemi ile birlikte DataSet-III veri kümesine uygulanmasından sonra başarımları Çizelge 6.5’de görülmektedir.

Çizelge 6.5 : Dataset-III’ye özellik seçimi uygulandığında başarımları.

standart performans:16,6%	C 4.5 %	NB %	SVM %	KNN %
DataSet-III (42249 özellik)	81	97,12	96,54	68,5
DataSet-III (405 özellik)	79,5	96,60	93,7	67,2

Özellik seçim algoritması ile boyut 405’e indirilirken % 96,60 başarımları sağlanmıştır.

6.1.4 Özellik uzayında tamsayı değişken kullanımı

DataSet-IV veri kümesinde özellikler için ikili (0/1) değerler yerine tamsayı değerler kullanılmaktadır. Bu sayede özellik uzayının boyutlarının nasıl etkileneceği görülmek istenmiştir. Burada her bir alt tümce 10x51 yani 510 özellik yerine sadece 15 özellik

temsil edilmektedir. Her bir alt tümcede 10 öbekten yüklem hariç 9 öbeğin her biri için bir özellik kullanılmaktadır. Bu özellik değeri öbek 51 kavramdan hangisini içeriyorsa o kavramın indisi kadardır ve 0 yada 50 yada 0 ile 50 arasında değer almaktadır. Yüklem öbeği içinse 6 özellik kullanılmaktadır. Yüklem için kullanılan özellikler Çizelge 6.6’de görülmektedir. Sınıflandırma algoritmasında kullanılırken bu değerler 0-1 aralığında normalize edilmiştir.

Çizelge 6.6 : DataSet-IV’de yüklem için kullanılan özellikler.

Yükleme Ait Özellikler	Alabileceği değer
Yüklem kökü isim/fiil	0-fiil kökü 1-isim kökü
Sınıf numarası	Fiil köklü ise 0-108 arası İsim köklü ise 0-50 arası
Yüklem ek türü	0-fiil kökü 1-len eki almış isim kökü 2-len eki almış isim kökü 3-len eki almış isim kökü 4-ek almamış isim kökü
Yüklem çatı türü	0 ile 5 arasında tam sayı
Yüklem zaman türü	0 ile 30 arasında tam sayı
Yüklem şahıs türü	0 ile 5 arasında tamsayı

DataSet-IV veri kümesi DataSet-II veri kümesinde olduğu gibi metindeki 5 alt tümcenin kaba anlamsal temsilini tam sayı özellikler olarak içermekte ve her bir alt tümcenin temsili için 15 boyutlu vektör kullanılmaktadır. Metnin en çok kavram içeren 5 alt tümcesiyle temsili söz konusu olduğundan, metin 15x5 yani 75 boyutta temsil edilmektedir. Bu veri kümesi için WEKA paketi ile NaiveBayes sınıflandırma algoritması, K- En Yakın Komşuluk sınıflandırma algoritması, LSVM ve C 4.5 Karar Ağacı kullanılmıştır. Özellik seçim algoritması olarak “*Chi square*” özellik çıkarım modeli “*ranker search*” yöntemi ile birlikte kullanılmıştır. Başarım hesabında 10 katlı çapraz doğrulama (10-fold cross validation) kullanılmıştır.

Çizelge 6.7’de ilk satır sınıflandırma algoritmalarının başarımlarını göstermektedir. Sınıflandırma algoritmaları için diğer DataSet’lerde kullanılan Sınıflandırma Algoritma Parametrelerinin aynısı kullanılmıştır. “*Chi square*” Özellik seçim

algoritmasının “*ranker search*” yöntemi ile birlikte DataSet-IV veri kümesine uygulanmasından sonra başarımlar değerleri ikinci satırda gösterilmektedir.

Çizelge 6.7 : Dataset-IV başarımlar oranları.

standart performans:16,6%	C 4.5 %	NB %	SVM %	KNN %
DataSet-IV (75 özellik)	54,23	68,31	65,98	49,5
DataSet-IV (48 özellik)	51,98	67,05	63,02	48,11

DataSet-IV veri kümesi kullanıldığında sınıflandırma değerlerinde düşüş görülmektedir. Bunun nedeni anlamsal bakımdan tümüyle ayrık kavramların sıra numarası ile temsil edildiğinde, sıra numarası yakın olan kavramların anlamları yakın olmasa dahi benzer kabul edilmesi; sıra numarası uzak olan kavramların anlamları yakın olsa dahi farklı olarak kabul edilmesi olabilir. Özellik uzayını küçülten bu yaklaşım başarımlar değerleri olarak istediğimiz sonucu vermemiştir.

6.2 Belge Sınıflandırmada Tümce Matrisi Kullanımı Sonucu

Bu çalışmada İTÜ Türkçe DDİ Servisinin Tümcelerini bağluluk çözümleme uygulaması ve WEKA paketi içindeki sınıflandırma ve özellik seçimi algoritmaları kullanılmıştır. Bu çalışmada kaba anlamsal ve dil bilgisel gösterim matrisi kullanılarak % 86,10 doğruluk ile boyut 145'e düşürülmüştür. Eski veriye kaba anlamsal bilgilerin eklendiği bir başka değişle DataSet-I ile DataSet-II'nin birleştirildiği DataSet-III modelinde Naive Bayes yöntemi ile daha önceki en yüksek doğruluk değeri aşılarak 97,12 doğruluk elde edilmiştir.

7. SONUÇ VE ÖNERİLER

Çalışmamız;

Tümcenin alt tümcelerinin bulunduğu,

Tümcelerin öbeklerine ayrıştırıldığı,

Öbeklerin içerdiği kavramların bulunduğu,

Tümcenin içerdiği zaman türünün incelenip yüklemle kıyaslandığı,

Tümcenin öznesinin türünün, tekil ya da çoğul olduğunun incelendiği ve

Yüklemin öbekleriyle hem ek bakımından hem de kavram bakımından uyumluluğunun incelendiği

Türkçe anlamsal ve dil bilgisi tümce çözümleme kaynağı olmak hedefindedir. Ayrıca çalışmamızda tümcelerin kaba anlamsal gösterimi oluşturulmaktadır. Tümcenin bu kaba temsilinin farklı DDİ uygulamalarında kullanılmasının metinler içindeki anlamsal özelliklerden faydalanan uygulamalarda fayda sağlayabileceği görülmektedir.

7.1 Tümce Ayırma, Öbek Bulma ve Kavram Bulmaya İlişkin Sonuçlar

"Tümce Ayırma", "Öbek Bulma" ve "Kavram Bulma" uygulamalarında girilen herhangi bir tümce için doğruluk oranları ortalama %57 civarındadır. DDİ Bağlılık Çözümleyicinin sadece doğru çalıştığı örnekler göze alındığında doğruluk oranı "Tümce Ayırma" ve "Kavram Bulma"da %80'lere "Öbek Bulma"da %89'lara ulaşmaktadır.

7.2 Anlamsal ve Dilbilgisel Uyuşum Tespit Sonuçları

Tümcenin Anlamsal ve Dilbilgisel Uyuşumunu tespit etmek için DDİ Bağlılık Çözümleyicinin hatasız çalıştığı Türkçe Radyo Televizyon (TRT) Kurumu kayıtlarından elde edilen yapısal anlatım bozukluğuna sahip 250 tümce ve yine TRT kayıtlarından elde edilen 250 sorunsuz tümce içeren toplam 500 tümce kullanılmıştır. Anlamsal ve dil

bilgisi analizi çalışmamızın hedefi yapısal hataların “*öbek kavram çiftlerinin yüklem uyumsuzluğu*”, “*Çoğul ekinin yanlış kullanımı*”, “*kişi ve zaman öbeklerinin yüklem uyumsuzluğu*” ve “*bağlaçlar ve iç tümceler*” maddeleridir. Anlatım bozukluğu türlerini incelediğimizde hedeflediğimiz bu maddelerin günlük metinlerdeki yapısal hataların %81,16’lık dilimini oluşturduğu görülmektedir. Çalışmamız sonucunda %81,16’lık dilim içinden %64’lük hatayı tespit edebildik. Bu da hedeflediğimiz dilimde %81,34 başarı elde ettiğimiz anlamına gelmektedir.

Çalışmamız yapısal anlatım bozukluklarının saptanması ve düzeltilmesiyle ilgilidir. Her ne kadar farklı kaynak dillerde çalışıldığı için kıyaslanması doğru olmasa da yapısal hatalar kapsamında İngilizce için yapılan çalışmalarda 0,60’lara varan doğruluk değerleri elde edilmiştir (Felice ve ark. 2014), (Lee ve Lee 2014), (Ng ve ark. 2014). Bu çalışmaların hiç biri yüklem öbek kavram ilişkisinin kontrolü şeklinde değildir. Türkçe için ise yapısal anlatım bozukluklarının saptanması ile ilgili bir makale bulunamamıştır. Türkçe yapısal anlatım bozukluklarının saptanması için %68 doğruluk değeri iyi bir başlangıç değeridir.

7.3 Kaba Anlamsal Matrisin Belge Sınıflandırmada Kullanımı Sonuçları

Bu çalışmada İTÜ Türkçe DDİ Servisinin Tümceler bağlılık çözümleme uygulaması ve WEKA paketi içindeki sınıflandırma ve özellik seçimi algoritmaları kullanılmıştır. Metin Sınıflandırma Çalışmalarında karşılaşılan en önemli sorunlardan biri metin verilerinin gösterimindeki boyut sayısının büyüklüğüdür. Geleneksel yaklaşımlarda, metinler o dilin farklı sözcükleri sayısınca boyutlarla gösterim edilmektedir. Türkçe gibi eklemeli dillerde sözcük kökleri kullanılarak boyut azaltılabilse de 1000’ler seviyesinde kalmaktadır. Bu çalışmada kaba anlamsal ve dil bilgisel gösterim matrisi kullanılarak % 86,10 doğruluk ile boyut 145’e düşürülmüştür. Eski veriye kaba anlamsal bilgilerin de eklendiği bir başka değişle DataSet-I ile DataSet-II’nin birleştirildiği DataSet-III modelinde Naive Bayes yöntemi ile daha önceki en yüksek doğruluk değeri aşılarak 97,12 doğruluk elde edilmiştir.

7.4 İleriki Çalışmalar

Çalışmamız giriş tümcelerini DDİ Bağlılık çözümleyiciden almaktadır. Eryiğit ve takımı İTÜ DDİ bağlılık çözümleyicinin yeni güncellemeleri üzerine çalışmaktadır.

Bağlılık Çözümleyicideki iyileşmeler doğrudan doğruluk değerlerimizi etkileyecektir. Diğer yandan kavram sözlüklerinin kapsamının artırılması ve yeni konu başlıkları eklenmesi kavram bulmadaki doğrulukları ve kaba anlamsal temsilin gerçek anlamla olan benzerliğini artırabilir. Tezimizde Levin sınıfları sadece eylemler çevirilerek oluşturulmuştur. Bunu bir dil bilimci ile Türkçeye özgü bir biçimde oluşturmak gerekmektedir. Bu haliyle de başarılı sonuçlar almada yeterli olan çalışmamız bahsedilen değişiklikler yapıldığında daha da yüksek doğruluk değerlerine ulaşabilecektir.

Tümcenin kaba anlamsal matrisi pek çok uygulamada kullanılarak katkısı değerlendirilebilir. Türkçenin düzenli tümce yapısı ve düzenli yüklem yapısı bu çalışmanın esin kaynağı olmasına karşın, öbek kavram temsili tüm diller için kullanılabilecek bir yöntemdir.



KAYNAKLAR

- [1] **Atwell, E.S.** (1987). How to detect grammatical errors in a text without parsing it, *Proceedings of the third conference on European chapter of the Association for Computational Linguistics*, Association for Computational Linguistics, s.38–45.
- [2] **Izumi, E., Uchimoto, K., Saiga, T., Supnithi, T. ve Isahara, H.** (2003). Automatic error detection in the Japanese learners' English spoken data, *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 2*, Association for Computational Linguistics, s.145–148.
- [3] **Sjöbergh, J. ve Knutsson, O.** (2005). Faking errors to avoid making errors: Very weakly supervised learning for error detection in writing, *Proc. RANLP*, cilt2005.
- [4] **Domeij, R., Knutsson, O., Carlberger, J. ve Kann, V.** (2000). Granska-an efficient hybrid system for Swedish grammar checking, *I Proceedings of the 12 th Nordic Conference in Computational Linguistics, Nodalida-99*, Citeseer.
- [5] **Ehsan, N. ve Faili, H.** (2013). Grammatical and context-sensitive error correction using a statistical machine translation framework, *Software: Practice and Experience*, 43(2), 187–206.
- [6] **Wang, Y., Wang, L., Wong, D.F., Chao, L.S., Zeng, X. ve Lu, Y.** (2014). Factored statistical machine translation for grammatical error correction, *CoNLL-2014*, 83.
- [7] **Doğan, S. ve Karaağaç, G.** Dilbilgisel Bağdaşmazlık ve anlamsal tutarsızlık.
- [8] **Sahin, G. ve Adali, E.** (2014). A pilot study on automatic inference rule discovery from Turkish text, *Application of Information and Communication Technologies (AICT), 2014 IEEE 8th International Conference on*, IEEE, s.1–5.
- [9] **Aygül, M., Karaalioglu, G. ve Amasyalı, F.** (2014). Prediction of function tags of the simple Turkish sentences by conditional random fields, *Sigma*, 32, 23–30.
- [10] **Delibas, A.** (2008). *Doğal dil işleme ile Türkçe Yazım hatalarının denetlenmesi, (Doktora Tezi)*, Fen Bilimleri Enstitüsü.
- [11] **Dilsiz, S.** (2015). *Bulanık mantık ve yapay sinir ağları ile Türkçe yazım denetleyicisi, (Doktora Tezi)*, Fen Bilimleri Enstitüsü.

- [12] Mikolov, T., Chen, K., Corrado, G. ve Dean, J. (2013). Efficient estimation of word representations in vector space, *arXiv preprint arXiv:1301.3781*.
- [13] Marelli, M., Bentivogli, L., Baroni, M., Bernardi, R., Menini, S. ve Zamparelli, R. (2014). Semeval-2014 task 1: Evaluation of compositional distributional semantic models on full sentences through semantic relatedness and textual entailment, *SemEval-2014*.
- [14] Socher, R., Huval, B., Manning, C.D. ve Ng, A.Y. (2012). Semantic compositionality through recursive matrix-vector spaces, *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, Association for Computational Linguistics, s.1201–1211.
- [15] Haaparanta, L. ve Hintikka, J. (1986). *Frege synthesized: essays on the philosophical and foundational work of Gottlob Frege*, cilt181, Springer.
- [16] Bybee, J.L. (1985). *Morphology: A study of the relation between meaning and form*, cilt 9, John Benjamins Publishing.
- [17] Korkmaz, Z. (1994). *Türkçede eklerin kullanılış şekilleri ve ek kalıplaşması olayları*, cilt598, Türk Dil Kurumu.
- [18] Schuler, K.K. (2005). VerbNet: A broad-coverage, comprehensive verb lexicon.
- [19] Fillmore, C.J., Johnson, C.R. ve Petruck, M.R. (2003). Background to framenet, *International journal of lexicography*, 16(3), 235–250.
- [20] Giuglea, A.M. ve Moschitti, A. (2006). Semantic role labeling via framenet, verbnet and propbank, *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, s.929–936.
- [21] Hanks, P. (2004). Corpus pattern analysis, *Euralex Proceedings*, cilt 1, s.87–98.
- [22] Lakoff, G. (1990). *Women, fire, and dangerous things: What categories reveal about the mind*, cilt 1, Cambridge Univ Press.
- [23] Fellbaum, C. (1998). *WordNet*, Wiley Online Library.
- [24] Stamou, S., Oflazer, K., Pala, K., Christoudoulakis, D., Cristea, D., Tufis, D., Koeva, S., Totkov, G., Dutoit, D. ve Grigoriadou, M. (2002). Balkanet: A multilingual semantic network for the balkan languages, *Proceedings of the International Wordnet Conference, Mysore, India*, s.21–25.
- [25] Bilgin, O., Çetinoğlu, Ö. ve Oflazer, K. (2004). Building a wordnet for Turkish, *Romanian Journal of Information Science and Technology*, 7(1-2), 163–172.
- [26] Baytop, T. (1994). *Türkçe bitki adları sözlüğü*, cilt578, Türk Dil Kurumu.

- [27] **Güngör, O. ve Güngör, T.** (2007). Türkçe bir sözlükteki tanımlardan kavramlar arasındaki üst-kavram ilişkilerinin Çıkarılması, *Akademik Bilişim Konferansı*, 1(1), 1–13.
- [28] **TDK** (2010). Büyük türkçe sözlük, *Ankara: TDK*.
- [29] **Eryigit, G.** (2014). ITU Turkish NLP web service, *EACL 2014*, 1.
- [30] **Adalı, E.** (2012). Doğal Dil İşleme (Natural Language Processing), *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 6(6).
- [31] **Solak, A. ve Oflazer, K.** (1993). Design and implementation of a spelling checker for Turkish, *Literary and Linguistic Computing*, 8(3), 113–130.
- [32] **Oflazer, K.** (1996). Error-tolerant finite-state recognition with applications to morphological analysis and spelling correction, *Computational Linguistics*, 22(1), 73–89.
- [33] **Aşliyan, R., Günel, K. ve Yakhno, T.** (2007). Detecting misspelled words in Turkish text using syllable n-gram frequencies, *Pattern Recognition and Machine Intelligence*, 553–559.
- [34] **Şentürk, F.** (2009). *Biçimbirimsel Bul Ve Değiştir*, (Doktora Tezi), Fen Bilimleri Enstitüsü.
- [35] **Güngör, O.** (2006). Semantic Categorization of Turkish Language Elements.
- [36] **Temürcü, C.** (2007). A semantic framework for analyzing tense, aspect and mood: an application to the ranges of polysemy some suffixes in Turkish.
- [37] **Ayşe, Ş., Zeynep, O. ve İlknur, P.** (2011). Extraction of semantic word relations in Turkish from dictionary definitions, *Proceedings of the ACL 2011 Workshop on Relational Models of Semantics*, Association for Computational Linguistics, s.11–18.
- [38] **Amasyalı, M.F. ve diğerleri** (2011). Kavramlar Arası Anlamsal İlişkilerin Türkçe Sözlük Tanımları Kullanılarak Otomatik Olarak Çıkarılması/matic Extraction of Semantic Relationships using Turkish Dictionary Definitions, *EMO Bilimsel Dergi*, 1(1), 1–14.
- [39] **Yıldırım, S. ve Yıldız, T.** (2012). Automatic Extraction of Turkish Hypernym-Hyponym Pairs From Large Corpus, *24th International Conference on Computational Linguistics*, s.493.
- [40] **Fırat, Y., Kılıçaslan, Y. ve Uçar, Ö.** (2013). Bilgisayar Ortamında Biçimsel Ontoloji Oluşturulması, *Journal of Higher Education & Science/Yükseköğretim ve Bilim Dergisi*, 3(1).
- [41] **Yıldız, T., Yıldırım, S. ve Diri, B.** (2013). Extraction of part-whole relations from Turkish corpora, *Computational Linguistics and Intelligent Text Processing*, Springer, s.126–138.

- [42] **Sahin, G., Diri, B. ve Yildiz, T.** (2015). Pattern and semantic similarity based automatic extraction of hyponym-hypernym relation from Turkish corpus, *Signal Processing and Communications Applications Conference (SIU), 2015 23th*, IEEE, s.674–677.
- [43] **Yücesoy, B. ve diğerleri** (2007). Comparison of semantic and single term similarity measures for clustering turkish documents, *Machine Learning and Applications, 2007. ICMLA 2007. Sixth International Conference on*, IEEE, s.393–398.
- [44] **Güran, A., Bekar, E. ve Akyokus, S.** (2010). A comparison of feature and semantic-based summarization algorithms for Turkish, *International Symposium on Innovations in Intelligent Systems and Applications*.
- [45] **Ozdikis, O., Senkul, P. ve Oguztuzun, H.** (2012). Semantic expansion of tweet contents for enhanced event detection in twitter, *Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on*, IEEE, s.20–24.
- [46] **Çelik, D.** (2016). Towards a semantic-based information extraction system for matching résumés to job openings, *Turkish Journal of Electrical Engineering & Computer Sciences*, 24(1), 141–159.
- [47] **Derici, C., Çelik, K., Özgür, A., Güngör, T., Kutbay, E., Aydın, Y. ve Kartal, G.** Türkçe Soru Cevaplama Sistemlerinde Kural Tabanlı Odak Çıkarımı Rule-Based Focus Extraction in Turkish Question Answering Systems.
- [48] **Amasyalı, M.F. ve Beken, A.** Türkçe Kelimelerin Anlamsal Benzerliklerinin Ölçülmesi ve Metin Sınıflandırmada Kullanılması Measurement of Turkish Word Semantic Similarity and Text Categorization Application.
- [49] **Miller, G.A. ve Fellbaum, C.** (1991). Semantic networks of English, *Cognition*, 41(1), 197–229.
- [50] **Bawakid, A. ve Oussalah, M.** (2008). A semantic summarization system: University of Birmingham at TAC 2008, *Proceedings of the first text analysis conference (TAC 2008)*, Citeseer.
- [51] **Hu, M. ve Liu, B.** (2004). Mining and summarizing customer reviews, *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, s.168–177.
- [52] **Moschitti, A., Morarescu, P. ve Harabagiu, S.M.** (2003). Open Domain Information Extraction via Automatic Semantic Labeling., *FLAIRS Conference*, s.397–401.
- [53] **Alani, H., Kim, S., Millard, D.E., Weal, M.J., Hall, W., Lewis, P.H. ve Shadbolt, N.R.** (2003). Automatic ontology-based knowledge extraction from web documents, *Intelligent Systems, IEEE*, 18(1), 14–21.
- [54] **Lin, C.H., Yen, C.W., Hong, J.S. ve Cruz-Lara, S.** (2007). Event-based textual document retrieval by using semantic role labeling and coreference resolution, *IADIS international conference WWW/Internet 2007*.

- [55] **Moldovan, D., Tatu, M. ve Clark, C.** (2010). Role of semantics in question answering, *Semantic Computing*, 373–419.
- [56] **Shen, D. ve Lapata, M.** (2007). Using Semantic Roles to Improve Question Answering., *EMNLP-CoNLL*, s.12–21.
- [57] **Pizzato, L.A. ve Mollá, D.** (2008). Indexing on semantic roles for question answering, *Coling 2008: Proceedings of the 2nd workshop on Information Retrieval for Question Answering*, Association for Computational Linguistics, s.74–81.
- [58] **Hutchins, J.** (1994). *Research methods and system designs in machine translation: a ten-year review, 1984-1994*, na.
- [59] **Görgün, O. ve Yıldız, O.T.** (2012). Using morphology in English-Turkish statistical machine translation, *Signal Processing and Communications Applications Conference (SIU), 2012 20th*, IEEE, s.1–4.
- [60] **Gulcehre, C., Firat, O., Xu, K., Cho, K., Barrault, L., Lin, H.C., Bougares, F., Schwenk, H. ve Bengio, Y.** (2015). On Using Monolingual Corpora in Neural Machine Translation, *arXiv preprint arXiv:1503.03535*.
- [61] **Pang, B. ve Lee, L.** (2008). Opinion mining and sentiment analysis, *Foundations and trends in information retrieval*, 2(1-2), 1–135.
- [62] **McKeown, M.G.** (1985). The acquisition of word meaning from context by children of high and low ability, *Reading Research Quarterly*, 482–496.
- [63] **Echihabi, A., Hermjakob, U., Hovy, E., Marcu, D., Melz, E. ve Ravichandran, D.**, (2006). How to Select an Answer String?, *Advances in open domain question answering*, Springer, s.383–406.
- [64] **Barzilay, R. ve Lee, L.** (2003). Learning to paraphrase: an unsupervised approach using multiple-sequence alignment, *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, Association for Computational Linguistics, s.16–23.
- [65] **Dolan, B., Quirk, C. ve Brockett, C.** (2004). Unsupervised construction of large paraphrase corpora: Exploiting massively parallel news sources, *Proceedings of the 20th international conference on Computational Linguistics*, Association for Computational Linguistics, s.350.
- [66] **Chang, C.Y. ve Clark, S.** (2010). Linguistic steganography using automatically generated paraphrases, *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, Association for Computational Linguistics, s.591–599.
- [67] **Heilman, M. ve Smith, N.A.** (2010). Tree edit models for recognizing textual entailments, paraphrases, and answers to questions, *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, Association for Computational Linguistics, s.1011–1019.

- [68] **Socher, R., Huang, E.H., Pennin, J., Manning, C.D. ve Ng, A.Y.** (2011). Dynamic pooling and unfolding recursive autoencoders for paraphrase detection, *Advances in Neural Information Processing Systems*, s.801–809.
- [69] **Madnani, N., Tetreault, J. ve Chodorow, M.** (2012). Re-examining machine translation metrics for paraphrase identification, *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Association for Computational Linguistics, s.182–190.
- [70] **Swier, R.S. ve Stevenson, S.** (2004). Unsupervised semantic role labelling, *Proceedings of EMNLP*, cilt 95, s.102.
- [71] **Palmer, M., Gildea, D. ve Xue, N.** (2010). Semantic role labeling, *Synthesis Lectures on Human Language Technologies*, 3(1), 1–103.
- [72] **Glickman, O., Dagan, I. ve Koppel, M.** (2005). A probabilistic classification approach for lexical textual entailment, *Proceedings of the National Conference On Artificial Intelligence*, cilt 20, Menlo Park, CA; Cambridge, MA; London; AAI Press; MIT Press; 1999, s.1050.
- [73] **Jijkoun, V. ve de Rijke, M.** (2005). Recognizing textual entailment using lexical similarity, *Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment*, s.73–76.
- [74] **MacCartney, B., Grenager, T., de Marneffe, M.C., Cer, D. ve Manning, C.D.** (2006). Learning to recognize features of valid textual entailments, *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, Association for Computational Linguistics, s.41–48.
- [75] **Bos, J. ve Markert, K.** (2005). Recognising textual entailment with logical inference, *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, s.628–635.
- [76] **Roth, D. ve Sammons, M.** (2007). Semantic and logical inference model for textual entailment, *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, Association for Computational Linguistics, s.107–112.
- [77] **MacCartney, B. ve Manning, C.D.** (2007). Natural logic for textual inference, *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, Association for Computational Linguistics, s.193–200.
- [78] **Berant, J., Dagan, I. ve Goldberger, J.** (2011). Global learning of typed entailment rules, *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, Association for Computational Linguistics, s.610–619.
- [79] **AKSAN, Y. ve YALDIR, Y.** (2011). Türkçe Sözvarlığının Nicel Betimlemesi, *TNC Org Publications*, 24, 377–387.

- [80] **Chomsky, N.** (1956). Three models for the description of language, *Information Theory, IRE Transactions on*, 2(3), 113–124.
- [81] **Nivre, J.** (2005). Dependency grammar and dependency parsing, *MSI report*, 5133(1959), 1–32.
- [82] **Backus, J.W.** (1959). The syntax and semantics of the proposed international algebraic language of the Zurich ACM-GAMM conference, *Proceedings of the International Conference on Information Processing*, 1959.
- [83] **Hoffman, B.** (1995). The computational analysis of the syntax and interpretation of "free" word order in Turkish, *IRCS Technical Reports Series*, 130.
- [84] **Güneş, Demir, S.** (2008). Türkçedeki Zaman Belirteçlerinin Sınıflaması ve Dökümü Üzerine Bir Deneme Çalışması, *Turkish Studies*, 2, 278–293.
- [85] **Levin, B.** (1993). *English verb classes and alternations: A preliminary investigation*, University of Chicago press.
- [86] **Tan, S.** (2006). An effective refinement strategy for KNN text classifier, *Expert Systems with Applications*, 30(2), 290–298.
- [87] **Yang, Y. ve Chute, C.G.** (1994). An example-based mapping method for text categorization and retrieval, *ACM Transactions on Information Systems (TOIS)*, 12(3), 252–277.
- [88] **Kim, S.B., Rim, H.C., Yook, D. ve Lim, H.S.**, (2002). Effective methods for improving naive bayes text classifiers, *PRICAI 2002: Trends in Artificial Intelligence*, Springer, s.414–423.
- [89] **Wu, M.C., Lin, S.Y. ve Lin, C.H.** (2006). An effective application of decision tree to stock trading, *Expert Systems with Applications*, 31(2), 270–274.
- [90] **Li, C.H. ve Park, S.C.** (2006). Text categorization based on artificial neural networks, *Neural information processing*, Springer, s.302–311.
- [91] **Nakayama, M. ve Shimizu, Y.** (2003). Subject Categorization for Web Educational Resources using MLP., *ESANN*, s.9–14.
- [92] **Lewis, D.D.**, (1998). Naive (Bayes) at forty: The independence assumption in information retrieval, *Machine learning: ECML-98*, Springer, s.4–15.
- [93] **Joachims, T.** (1998). *Text categorization with support vector machines: Learning with many relevant features*, Springer.
- [94] **Tsochantaridis, I., Hofmann, T., Joachims, T. ve Altun, Y.** (2004). Support vector machine learning for interdependent and structured output spaces, *Proceedings of the twenty-first international conference on Machine learning*, ACM, s.104.
- [95] **Çataltepe, Z., Turan, Y. ve Kesgin, F.** (2007). Turkish document classification using shorter roots, *Signal Processing and Communications Applications, 2007. SIU 2007. IEEE 15th*, IEEE, s.1–4.

- [96] **Torunoğlu, D., Çakırman, E., Ganiz, M.C., Akyokuş, S. ve Gürbüz, M.Z.** (2011). Analysis of preprocessing methods on classification of Turkish texts, *Innovations in Intelligent Systems and Applications (INISTA), 2011 International Symposium on*, IEEE, s.112–117.
- [97] **Amasyalı, M.F. ve Diri, B.**, (2006). Automatic turkish text categorization in terms of author, genre and gender, *Natural Language Processing and Information Systems*, Springer, s.221–226.
- [98] **Yıldız, H., Gençtav, M., Usta, N., Diri, B. ve Amasyalı, M.** (2007). Metin Sınıflandırmada Yeni Özellik Çıkarımı, *IEEE SIU 2007 15. Sinyal İşleme, İletişim ve Uygulamaları Kurultayı*.
- [99] **Amasyalı, M. ve Beken, A.** (2009). Türkçe kelimelerin anlamsal benzerliklerinin ölçülmesi ve metin sınıflandırmada kullanılması, *IEEE Signal Processing and Communications Applications Conference, SIU-2009*.
- [100] **Yıldız, H.K., Gençtav, M., Usta, N., Diri, B. ve Amasyalı, M.F.** (2007). A New Feature Extraction Method for Text Classification, *Signal Processing and Communications Applications, 2007. SIU 2007. IEEE 15th*, IEEE, s.1–4.

ÖZGEÇMİŞ



Ad Soyad: İlknur Dönmez

Doğum Yeri ve Tarihi: Uşak 1979

Adres: Cumhuriyet Mah. Erguvan Sok. Enez Sitesi 1/3/8 Küçükçekmece, İstanbul

E-Posta: ilknur.buyukkuscu@hotmail.com ; buyukkuscu@itu.edu.tr

Lisans: İstanbul Teknik Üniversitesi Elektronik ve Haberleşme Mühendisliği

Y. Lisans: Ege Üniversitesi Matematik Bölümü

Mesleki Deneyim ve Ödüller: 5 yıl Vestek/Vestel Elektronik A.Ş.'de ArGe departmanında gömülü TV yazılımı alanında uzman yazılımcılık ve proje liderliği, 5 yıl Petrol Ofisi Genel Müdürlüğünde Bilgi Teknolojileri bölümünde havacılık otomasyonu ve positive kart projelerinde proje yöneticiliği yaptım. Halen İstanbul Teknik Üniversitesinde DPT Destekli İngilizce Türkçe Makina çevirisi projesinde çalışmaktayım.

TEZDEN TÜRETİLEN YAYINLAR/SUNUMLAR

- **Dönmez, İ.**, Adalı, 2015: Extracting phrase-content pairs for Turkish sentences. *In Application of Information and Communication Technologies (AICT), 2015 9th International Conference* October 2015 pp. 128-132. IEEE.
- **Dönmez, İ.**, Adalı, E. 2015: Türkçe Tümce Çözümlemede Vektör Yaklaşımı. *Afyon Kocatepe University Journal of Science Engineering*, 15(3).
- **Dönmez, İ.**, Adalı, E. 2016: Turkish Document Classification with Coarse-grained Semantic Matrix. *17th International Conference on Intelligent Text Processing and Computational Linguistics; April 3–9, 2016, Konya, Turkey*

