

# Türkçe Anlamsal Söylem ve Cümle Benzerliği Analizleri İçin Veri Kümesi Oluşturma Yöntemi

Gökhan Ercan  
Bilgisayar Mühendisliği Bölümü  
Işık Üniversitesi  
İstanbul, Türkiye  
gokhan.ercan@isik.edu.tr

Orçun Erkek  
Bilgisayar Mühendisliği Bölümü  
Işık Üniversitesi  
İstanbul, Türkiye  
orcun.erkek@isik.edu.tr

Onur Açıkgöz  
Bilgisayar Mühendisliği Bölümü  
Işık Üniversitesi  
İstanbul, Türkiye  
onur.acikgoz@isik.edu.tr

Rıza Özçelik  
Bilgisayar Mühendisliği Bölümü  
Boğaziçi Üniversitesi  
İstanbul, Türkiye  
riza.ozcelik@boun.edu.tr

Selen Parlar  
Bilgisayar Mühendisliği Bölümü  
Boğaziçi Üniversitesi  
İstanbul, Türkiye  
selen.parlar@boun.edu.tr

Olcay Taner Yıldız  
Bilgisayar Mühendisliği Bölümü  
Işık Üniversitesi  
İstanbul, Türkiye  
olcaytaner@isikun.edu.tr

**Abstract**—Doğal dil işleme çalışmamızın amacı Türkçe dili için paragraf-cümle düzeyinde anlamsal söylem analizi ve paragraf-cümle ve cümle-cümle düzeyinde metinsel benzerlik ölçümlemesi için bir veri kümesi hazırlamaktır. Girdi olarak kullanılan çoktan seçmeli sorular Türkiye Cumhuriyeti Ölçme, Seçme ve Yerleştirme Merkezi tarafından gerçekleştirilen sınavlarda çıkmış çoktan seçmeli Türkçe sorularıdır. Hedeflenen yaklaşımlar için iki kategoride dört farklı soru tipi belirlenmiştir: (i) paragrafın akışının bozulmasının tespit edilmesi, (ii) cümlelerin doğru sırasının bulunması, (iii) paragrafta geçen ifade ile anlatılmak istenen cümlelerin bulunması, (iv) anlamca en yakın cümlelerin bulunması. Tüm veri toplama, hazırlama, biçimbilimsel etiketleme ve biçim dönüştürme aşamaları sonucunda nihai olarak anlamsal söylem analizi için 434 soruluk, metinsel benzerlik analizi için de 539 soruluk veri kümesine ulaşılmıştır.

**Index Terms**—doğal dil işleme, anlamsal modelleme, söylem analizi, ölçümleme veri kümesi, cümle benzerliği, anlamsal ölçümleme, çoktan seçmeli

## I. GİRİŞ

Doğal dil işleme (DDİ) araştırmalarının temel amacı son derece karmaşık bir fenomen olan doğal dilin, anlaşılması, yorumlanması, sınıflandırılması, özetlenmesi gibi insanoğlunun yüksek bilişsel yeteneklerine ihtiyaç duyan görevlerin kısmen ya da tamamen bilgisayarlar ile desteklenmesini sağlamaktır. Önemli bir bilişsel yeteneği ölçümleyen söylemin tespiti problemi DDİ araştırmalarında sıklıkla çalışılan bir görev olmuştur.

Öğrencilerin ana dillerini okuma, yazma, anlama ve yorumlama yeteneklerinin tespit edilmesi, eğitim sisteminin hemen her aşamasında birincil ölçme ve seçme yöntemi olarak kullanılmaktadır. Ülkemizde de ilköğretimden, yüksek lisans seviyesine kadar öğrencinin okuduğunu anlaması ve yorumlamasını ölçümleyen Türkçe çoktan seçmeli sorularının sorulması değişmeyen bir uygulamadır. Bu çalışmada sunduğumuz 434 soruluk veri kümesi ile, ilgili DDİ modellerinin gelişimini raporlayan dışsal (extrinsic) bir ölçümleme veri kümesi oluşturulmuştur. Buna ek olarak, oluşturduğumuz veri kümesinin, çeşitli yaş gruplarından çeşitli zorluk seviyelerinde sınavlara

giren öğrencilerin bu soruları çözme başarıları ile modelleri karşılaştırabilme imkanının sağlanması hedeflenmiştir.

Veri kümesini oluşturan bütün sorular Türkiye Cumhuriyeti Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı (ÖSYM) tarafından yapılmış resmi sınavlarda sorulmuş (çıkılmış) sorulardan seçilmiştir. Hedeflenen iki problem için 4 farklı soru tipi belirlenmiştir. Çeşitli kaynaklardan elde edilen sorular xlsx dosya biçiminde biriktirilmiş, nihai veri kümesi ise belirlediğimiz yapısal şablon dahilinde xml biçimine çevirerek sonlandırılmıştır. Soruları otomatik cevaplamayı hedefleyecek anlamsal modellerin Türkçe dili için kelime ve kelime-altı biçimbilimsel bilgilere ihtiyaç duyabileceği değerlendirilerek soru ve cevaplarda geçen bütün cümleler biçimbilimsel çözümleme ve biçimbilimsel anlamsızlık giderme işlemlerinden geçirilmiştir. Çalışmanın nihai çıktısı olan xml dosyası biçimbilimsel çözümleme sonuçlarını da içermektedir. Şekil 1’de çalışmanın tamamlanan ve planlanan aşamaları ana hatlarıyla gösterilmiştir.

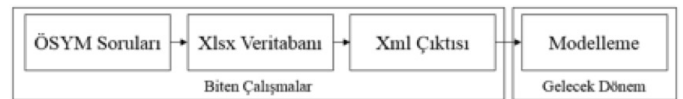


Figure 1. Çalışmanın tamamlanan ve planlanan aşamaları

## II. YAZIN

### A. Çoktan Seçmeli Sorular ile Ölçümleme

Çoktan seçmeli soru veri kümelerinin bir ölçümleme yöntemi olarak kullanılması yeni değildir. MSR (The Microsoft Research Sentence Completion Challenge Data) çalışması [1] cümlede-boş-bırakılan kelimenin-tahmini tipi beş seçenekli 1040 soruluk çoktan seçmeli veri kümesi sunmaktadır. MSR veri kümesi birçok sürekli ve ayrık dil modelleri ve kelime benzerlik modellerinin dışsal ölçülmesinde önemli bir destekleyici ölçümleme yöntemi olarak kullanılmaktadır [2],

[3]. MSR veri kümesinin gerçek insan değerlendirmelerini baz alarak ölçümleme yapması, kararsızlık (perplexity) gibi temel (intrinsic) istatistiksel ölçümleme yöntemlerinin açıklarının doldurulmasını sağlamıştır [2].

Benzer bir çalışmada [4] Landauer ve arkadaşları TOEFL (Test of English as a Foreign Language) sınavının kelimenin-eş-anlamlısını-bulma tipindeki 80 sorusunu veri kümesi olarak tanımlamışlardır. TOEFL veri kümesi 1997 yılından bu yana eş anlamlı kelime bulma modelleri için araştırmacılar tarafından karşılaştırma aracı olarak kullanılmıştır [5]. Bullinaria ve arkadaşları 2012 yılında TOEFL sorularını %100 oranında çözebilen modeller üretebildiklerini yayınlamışlardır [6].

Diğer bir örnek çalışmada [7] ise geliştirilen DDİ ve muhakeme (reasoning) modellerinin Tokyo Üniversitesi'ne giriş için yapılan sınavlarda sorulan matematik sorularının çözümünde ortalama öğrenci performansını yakalayabildiği raporlanmıştır. Araştırmacılar çalışmalarının bulgularını makine öğrenmenin bir başarısı olarak değil, eğitim sistemindeki yaratıcılık eksikliğinin bir sonucu olarak yorumlamışlardır.

### B. Söylem Analizi

Belirli bir konu ya da bağlam hakkında yazılmış olan doğal dil parçasına söylem adı verilir. Dilsel öğeler bir araya geldiklerinde anlamlı ve bütünsel bir mesajı ifade ediyorlar ise söylemi oluştururlar. Söylem analizi bir metindeki değişen farklı söylemlerin tespit edilmesi problemdir. Söylem analizi kökler, ekler, zamanlar, sayfa düzenleri gibi öğeler ile ilgili olabilir. Söylem, sıralı cümlelerden oluşur. Söylem içindeki her yeni cümle okuyucunun kendisinden önce gelen cümleleri okuduğu varsayımı ile yazılır ve bu cümlelere gönderme (atıf) yapar. Bu sözdizimsel ve anlamsal bağlantılar DDİ araştırmacısının söylemi tespit edebilmek için başvurduğu en önemli ipuçlarıdır. Örneğin "Köylü hepimizin velinimetidir. Bu necip unsurun refahını düşüneceğiz." ifadelerinde ikinci cümlede geçen "bu necip unsur" ifadesi birinci cümledeki "köylü" ifadesine atıf yapmaktadır. Bu örnekten anlaşılacağı gibi dilsel öğeler içinde buldukları söylem dışına çıkartılıp tek başlarına değerlendirdiklerinde anlam kaybedebilirler.

### C. Anlamsal Metin Benzerliği

Anlamsal metin benzerliği üzerinde çokça çalışılan bir DDİ problemdir. Görevin temel amacı iki metin arasındaki anlamsal benzerlikleri rakamsal olarak belirlemektir. Bu çalışmada Türkçe dili için metin benzerliğini paragraf-cümle ve cümle-cümle ikilileri düzeyinde modelleyen yöntemlerin araştırılması ile ilgilenilmektedir. Türkçe dilinin biçimbilimsel bilgilerini önsel bilgi olarak kullanan en iyi anlamsal modellerin, oluşturduğumuz veri kümesinden yüksek başarımla sağlanması beklenmektedir.

## III. YÖNTEM

### A. Veri Toplama

Bilindiği gibi ALES, DGS, LES, LYS, OKS, ÖSS, ÖYS, SBS, TEOG, YGS, YÖS sınavları ÖSYM tarafından düzenlenen farklı zorluk seviyelerinde sınavlardır. ÖSYM çoktan seçmeli sınavlarının ortak noktası genellikle tekrar

eden Türkçe soru tiplerinden oluşmalarıdır. Edebiyat ve dil bilgisi gibi bilgi ölçümleyen sorular çalışma kapsamı dışında bırakılmıştır. Yaptığımız ön araştırma sonucunda belirlediğimiz aday 32 soru tipinin 4 tanesi veri kümesi kapsamına alınmıştır. Veri kümesi sadece çıkmış ÖSYM sınav soruları ile kısıtlanmıştır. Sınav soruları hem ÖSYM'nin İnternet sitesinde paylaştığı belgelerden, hem de çeşitli İnternet siteleri ve çıkmış soru kitaplarından elde edilmiştir. Örnek olarak son 52 yılın paragraf soruları ve çözümlerini içeren soru kitabı [8] çalışmanın kaynaklarından bir tanesidir.

### B. Veri Hazırlama

Bu aşamada soru tipi yapıları xlsx dosya biçiminde hazırlanmıştır. Buna ek olarak ÖSYM sınav türleriyle birlikte sınavların hangi yıllarda yapıldığı bilgisi işaretlenmiştir. Oluşan veri havuzu hızlıca taranarak hangi sınavda hangi soru tipinden oransal olarak ne kadar olduğu tespit edilmiş ve hangi soru tipinden ne kadar soruya ulaşabileceğinin tahminlemesi yapılmıştır. Tahminlenen veriler ışığında belirlenen ana problem hedef alan iki çeşit soru tipi aşağıda anlatılmaktadır.

### C. Soru Tipleri

1) *Söylem Analizi:* Birinci soru tipi paragrafın akışının bozulmasının tespit becerisini sorgulamaktadır. Bu soru aşağıdaki biçimlerde de sorulabilmektedir:

- ... anlam bütünlüğünü bozan cümle hangisidir?
- ... yazarın düşüncesi hangi cümlede değişmiştir?

**1. - 5. sorularda, numaralanmış cümlelerden hangisinin anlam bütünlüğünü bozduğunu bulunuz.**

**(I) Sanayi devrimiyle birlikte insanın çevre üzerindeki etkisi hız kazanmıştır. (II) Sanayileşme çabası içindeki insanın fosil yakıtlara yönelmesi, nüfus artışı, şehirleşme, orman yangınları vb. sonucunda atmosferde artan sera gazları, günümüzde küresel ısınma problemini ortaya çıkarmıştır. (III) Dünyada en çok karşılaşılan durum, insanların çeşitli etkinliklerinin atmosferin güneş enerjisini yutma ve yayma şeklini değiştirmesidir. (IV) Bu problemin varlığı, sanayi devriminin başlangıcından günümüze değin geçen süre içinde, ortalama küresel hava sıcaklığının artmasıyla kanıtlanmıştır. (V) Güvenilir meteorolojik kayıtların tutulmaya başlandığı 1886'da 14,5 °C olan dünyanın ortalama sıcaklığı 1995'te 15,4 °C ye yükselmiştir.**

A) I.      B) II.      C) III.      D) IV.      E) V.

Figure 2. ALES 2012 sonbahar dönemi sınavından örnek bir paragrafın akışının bozulmasının tespiti sorusu (Soru tipi 1).

İkinci soru tipi verilen cümlelerin doğru sıralamasının bulunmasını içermektedir. Aşağıdaki farklı biçimlerde de sorulabilmektedir:

- ... olayların oluş sırasına göre nasıl sıralanmalıdır?
- ... giriş cümlesi hangisi olur?
- ... hangisi baştan üçüncü cümle olur?
- ... hangisi son cümle olur?

**32. – 33. SORULARDA, NUMARALANMIŞ SÖZLERİN ANLAMLI VE KURALLI BİR CÜMLE OLUŞTURAN SIRLANIŞINI BULUNUZ.**

I. eskiden yüz yüze sağlanan iletişim  
 II. uzaktan iletişime bıraktı  
 III. günümüzde yerini  
 IV. her yerde, her zaman  
 V. öyle ki herkes, herkese  
 VI. ulaşmaya başladı

A) I., III., VI., IV., V., II.      B) I., III., II., V., IV., VI.  
 C) IV., II., I., III., V., VI.      D) V., IV., VI., III., I., II.  
 E) V., VI., III., II., I., IV.

Figure 3. YÖS 2009 sınavından örnek bir doğru sıralama bulma sorusu (Soru tipi 2).

Soru tipi 1 ve soru tipi 2 için örnekler Şekil 2 ve Şekil 3'te sırasıyla verilmiştir. Her iki soru tipinin otomatik çözümü için de söylem analizi modeli gerekmektedir. Verilen bir cümle için verilen söylemi devam ettirme olasılığını hesaplayabilen her modelin kolaylıkla bu soruları çözen bir modele çevirebileceği öngörülmüştür. Tablo I, nihai veri kümesinde seçilen her soru tipinden kaç soru üretildiğini ve farklı ÖSYM sınavı türlerine göre dağılımını göstermektedir.

Table I  
SÖYLEM ANALIZI SORU TIPLERİNİN SINAVLARA GÖRE DAĞILIMI

Sınavlar	Tip 1	Tip 2	Toplam
ALES	72	125	197
DGS	40	82	122
LES	11	22	33
ÖSS	20	1	21
ÖYS	10	2	12
Diğer	12	37	49
Toplam	165	269	434

2) *Anlamsal Metin Benzerliği*: Üçüncü soru tipinde, paragrafta ya da paragrafın içinde geçen bir ifade ile anlatılmak istenilenin aşağıdaki cümlelerin hangisine benzediğinin bulunması istenmektedir. Bu soru tipi öğrencilere aşağıdaki gibi farklı ifadelerle sorulabilmektedir:

- ... asıl anlatılmak istenen nedir?
- ... metnin konusu nedir?
- ... hangisi söylenebilir/söylenemez?
- ... ana düşünce nedir?
- ... değişilmiştir/değişilmemiştir?

Dördüncü soru tipi, verilen cümleyle anlamca en yakın cümlelerin bulunmasını beklemektedir. Soru tipi aşağıdaki farklı

Hakkımdaki övgüleri de yergileri de pek ciddiye aldığımı söyleyemem. Övgüleri ilk duyduğumda yurtdışındaydım. Bana moral verdi, beni güdüledi bu övgüler. Övülmek elbette iyi bir şeydir; ama yalnızca bununla yaşanmaz ki. Düşünsenize, "Ben geleceğin yazaryım." dersen, bunu ilke edinip oturursan, kendini yenilemekten de kaçırırsan sonun ne olur? Ortada kalırsın; bırak geleceği, bugünü bile görmezsin.

**Bu parçada anlatılmak istenen aşağıdakilerden hangisidir?**

A) Yazarların, okurlardan gelecek tepkileri umursamadığı  
 B) Beğenilen yazarların da kalıcılığının, kendilerini sürekli geliştirmelerine bağlı olduğu  
 C) Sanatçılara olumlu eleştiriler yönelmek gerektiği  
 D) Kendini belirli ölçütlerle sınırlayan yazarların özgün ürünler veremeyeceği  
 E) Yazarların, okurlardan, yazdıklarını değerlendirmelerini beklemediği

Figure 4. ÖSS 2006 sınavından örnek bir-paragrafta-geçen-ifade-ile-anlatılmak-istenilen-cümle sorusu (Soru tipi 3).

biçimlerde de sorulabilmektedir:

- ... anlamca en yakındır?
- ... anlamca birbiriyle uyuşmamaktadır?
- ... çıkarılabilecek yargıyı bulunuz?
- ... hangisi verilen cümleyle aynı anlamdadır?
- ... sözüyle anlatılmak istenen nedir?

Buraya geldiğin zaman beni ara.

**Aşağıdakilerden hangisi bu cümleyle aynı anlamdadır?**

A) Buraya gelirken beni ara.  
 B) Buraya gelirken beni ara.  
 C) Buraya gelmeden önce beni ara.  
 D) Buraya gelince beni ara.  
 E) Buraya geleceksen beni ara.

Figure 5. YÖS 2009 sınavından örnek bir anlamca-en-yakın-cümlelerin bulunması sorusu (Soru tipi 4).

Soru tipi 3 ve soru tipi 4 için örnekler Şekil 4 ve Şekil 5'te sırasıyla verilmiştir. Soru tipi 3 için paragraf-cümle ve cümle-paragraf çiftleri seviyesinde, soru tipi 4 için ise cümle-cümle

çiftleri seviyesinde metinsel benzerliği bulan modellerin uygun olduğu öngörülmüştür. Soru metinlerindeki olumsuzluk ifade eden “uyuşmamaktadır”, “yakın değildir” gibi ifadeler sorunun doğru cevabını değiştirdiğinden bu gibi sorular veri setinde olumsuz soru olarak ayrı bir şekilde işaretlenmiştir. Tablo II, nihai veri kümesinde seçilen her soru tipinden kaç soru üretildiğini ve farklı ÖSYM sınavı türlerine göre dağılımını göstermektedir.

Table II  
ANLAMSAL METİN BENZERLİĞİ SORU TIPLERİNİN SINAVLARA GÖRE DAĞILIMI

Sınavlar	Tip 3	Tip 4	Toplam
ALES	69	62	131
DGS	57	37	94
LES	20	18	38
ÖSS	73	43	116
ÖYS	58	20	78
Diğer	53	29	82
Toplam	330	209	539

#### D. Veri Girişi

Soru verileri üretilirken öncelikle Internet üzerinde bulduğumuz kopyalanabilir özelliği olan sayısal kaynaklar kullanılmıştır. Eğer soru kopyalanabilir değil ise elle metin girişi yapılmıştır. Kullandığımız kaynak kitaplardaki sorular ise bir OKT (Optik Karakter Tanımlama) yazılımından geçirilerek sayısal metin haline çevrilmiştir. Verileri OKT işlemine hazırlamak için bütün sorular yüksek çözünürlüklü bir telefon kamerası ile çekilmiştir. OKT işleminden sonra birçok veride oluşan tarama hataları tespit edilip elle müdahale edilerek düzenlenmiştir.

#### E. XML Dönüştürme

Derlenen soruların farklı programlama platformlarından kolaylıkla okunabilmesini sağlamak amacıyla xml formatı kullanılmıştır. İlk aşamada xlsx biçiminde hazırlanmış olan veri kümesinin xml biçimine çevrilmesini sağlamak için Java platformunda bir çevirici program geliştirilmiştir. Apache POI bileşeni aracılığı ile xlsx dosyaları belleğe yüklenerek her elemanı (element) bir sınıf (class) ile temsil edilmek üzere belirlenen şablon yapı xml dosya biçimine ihraç edilmiştir.

Veri kümesinin nihai yapısındaki bazı önemli eleman (element) ve nitelikler (attributes) aşağıda açıklanmıştır:

- qtypes: Farklı soru tiplerini gruplayan elemandır.
- etCode: Sorunun ait olduğu sınav kodudur.
- id: Her soru için belirlenmiş olan tekil numaradır.
- question: Her bir soru örneğini temsil eden elemandır.
- answer: Sorunun doğru cevabıdır. Örneğin soru tipi 2 için “S4,S2,S3,S5,S1” değeri söylemi oluşturan doğru cümle sıralamasını temsil ederken, soru tipi 4 için “S1” değeri A seçeneğindeki cümlenin doğru cevap olduğunu temsil etmektedir.
- p1Segment: Soruda altı çizili olan kelimeyi veya cümleyi temsil eder.
- paragraph: Soruyu cevaplamak için sunulan cümleler bütünüdür.

- sentence: Sorudaki her cümleyi temsil eder. Birden fazla kelime (token) barındırır.
- token: Her bir kelimenin biçimbilimsel çözümlemesi ile birlikte temsil edildiği elemandır.

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<ns2:challenge xmlns:ns2="com.codenotfound.jaxb.model">
  <qtypes>
    <instances>
      <etCode>SBS</etCode><id>1.0</id><p>1.0</p><q>4.0</q>
      <question>
        <answer>S3</answer>
        <p1Segment></p1Segment>
        <paragraph>
          <sentence>
            <text>İnsan tarafından fark edilmediği sürece ,
              <tokens>
                <token>
                  <an>insan+NOUN+A3SG+PNON+NOM</an>
                  <form>insan</form>
                  <pos>NOUN</pos>
                  <text>insan</text>
                </token>
                <tokens>
                  <an>taraf+NOUN+A3SG+P3SG+ABL</an>
                  <form>taraf+sh+nDAn</form>
                  <pos>NOUN</pos>
                  <syllable></syllable>
                  <text>tarafından</text>
                </token>
              </tokens>
            </text>
          </sentence>
        </paragraph>
      </question>
    </instances>
  </qtypes>
</ns2:challenge>
```

Figure 6. Veri kümesinin nihai çıktısından örnek bir parça.

#### F. Biçimbilimsel Çözümleme

İlk aşamada çalışma grubumuzun DDİ kütüphanesi kullanılarak tüm paragraflar cümle ayırma (sentence segmentation) algoritmasından geçirilmiştir. İkinci olarak cümlelerine ayrılmış olan soruların her cümlesi biçimbilimsel çözümlemeye geçirilmiş ve aday çözümleme dalları elde edilmiştir [9]<sup>1</sup>. Son olarak ise aday çözümler biçimbilimsel bulanıklık çözümlemesi işleminden geçirilmiş ve en iyi dal çözümlemesi kelimeler halinde (tokens) ayrıştırılarak kayıt edilmiştir. Biçimsel çözümleme aşamalarında elde edilen tüm biçimbilimsel etiketlerin xml dosyası yapısına kaydedilmesi sağlanmıştır. Şekil 7’te “edilmediği” kelimesi için yapılan bir biçimbilimsel çözümleme örneği verilmiştir.

```
<tokens>
  <an>et+VERB+DB+VERB+PASS+NEG+DB+NOUN+PASTPART+A3SG+P3SG+NOM</an>
  <form>et+Hl+mA+DEk+sh</form>
  <pos>NOUN</pos>
  <text>edilmediği</text>
</tokens>
```

Figure 7. ‘Edilmediği’ kelimesi için yapılan bir biçimbilimsel çözümleme.

## IV. SONUÇ

Bu çalışmada söylem ve benzerlik anlamsal problemlerini hedefleyen iki farklı kategori dört farklı tipte toplam 973 çoktan seçmeli sorudan oluşan bir veri kümesi oluşturulmuş, Türkçe dilinin biçimbilimsel çözümleme etiketleri de eklenerek farklı platformlarda modelleme çalışmaları için kullanıma hazır hale getirilmiştir. Gelecek çalışmalarda geliştirilecek anlamsal söylem ve metin benzerliği modellerinin bu çalışmada üretilen veri seti üzerinde ölçümlendirilip sonuçlarının raporlanması amaçlanmaktadır. Ek olarak oluşturulan veri seti

<sup>1</sup><https://github.com/olcaytaner/MorphologicalAnalysis>

sorularının farklı öğretim seviyelerindeki öğrencilere (ilköğretim, lise, üniversite) çözdürülerek modellerin performansları ile insan performansları arasında karşılaştırma yapılması hedeflenmektedir.

#### REFERENCES

- [1] G. Zweig and C. Burges, "The microsoft research sentence completion challenge," Tech. Rep. MSR-TR-2011-129, 2011.
- [2] T. Mikolov, "Statistical language models based on neural networks," in *Presentation at Google, Mountain View*, 2012.
- [3] —, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.
- [4] T. K. Landauer, P. W. Foltz, and D. Laham, "An introduction to latent semantic analysis," *Discourse processes*, pp. 259–284, 1998.
- [5] [https://aclweb.org/aclwiki/TOEFL\\_Synonym\\_Questions\\_\(State\\_of\\_the\\_art\)](https://aclweb.org/aclwiki/TOEFL_Synonym_Questions_(State_of_the_art)).
- [6] J. A. Bullinaria and J. P. Levy, "Extracting semantic representations from word co-occurrence statistics: stop-lists, stemming, and svd," *Behavior research methods*, pp. 890–907, 2012.
- [7] N. Arai, "The impact of ai-can a robot get into the university of tokyo?" *National Science Review*, pp. 135–136, 2015.
- [8] S. Korkmaz, *Son 52 Yıl YGS-LYS Paragraf Soruları ve Çözümleri*. Akıllı Adam Yayınları.
- [9] O. Görgün and O. T. Yıldız, "A novel approach to morphological disambiguation for turkish," in *Computer and Information Sciences II*. Springer, 2011, pp. 77–83.