

# Veri Artırımı için Yarı-Denetimli Bağlamsal Anlam Belirsizliği Giderme

## Semi-Supervised Contextual Word Sense Disambiguation for Data Augmentation

Dilara TORUNOĞLU-SELAMET  
İstanbul Teknik Üniversitesi  
Bilgisayar Mühendisliği Bölümü  
torunoglud@itu.edu.tr  
ORCID: 0000-0002-8247-7552

Gülşen ERYİĞİT  
İstanbul Teknik Üniversitesi  
Yapay Zekâ ve Veri Mühendisliği Bölümü  
gulsenc@itu.edu.tr  
ORCID: 0000-0003-4607-7305

### Öz

Yapay zekâ alanında son dönemlerde öne çıkan derin öğrenme mimarilerinin, doğal dil işleme konusunun önemli problemlerinden biri olan Anlam Belirsizliği Giderme (ABG) çalışmalarında kayda değer gelişmelere yol açtığı gözlemlenmektedir. Denetimli yöntemler rakiplerine göre daha yüksek performans sergilemektedirler. Bunun en büyük nedeni kullanılan eğitim verilerinin büyüklükleridir. ABG problemi için İngilizce dili üzerinde elle-etiketlenmiş çok miktarda veri çevrim içi olarak erişilebilir durumdadır. Ancak düşük-kaynaklı diller (DKD'ler) probleme uygun veri eksikliği yaşamaktadırlar. Yeterli derecede probleme uygun veri toplamak ve etiketlemek vakit alıcı ve yüksek maliyet gerektiren bir iştir. Bu probleme değinmek ve aşmak üzere, bu çalışmada yarı-denetimli bağlamsal anlam belirsizliği giderme yaklaşımının veri artırımı için (daha sonra denetimli öğrenmede eğitim verisi olarak kullanılmak üzere) kullanılabilceğinin gösterilmesi amaçlanmıştır. Bu bağlamda özellikle DKD'lerde ABG problemi için test verisi bulmanın zor olması nedeniyle yaklaşımın doğruluğunu ve ilerleyen dönemlerde DKD'lerde kullanılabilirliğini ispatlamak amacıyla çevrimiçi bulunan elle-etiketlenmiş İngilizce ABG verisi kullanılmıştır. Oluşturulan yarı-denetimli yöntemde öbek kümesi (seed set) ve bağlam vektörleri (context embeddings) kullanılmaktadır. Yapılan çalışma 9 farklı bağlamsal dil modelinde (ELMo, BERT, RoBERTa vb.) test edilmiş ve her bir dil modelinin ABG problemi üzerindeki etkileri raporlanmıştır. İlk temel yaklaşıma göre sonuçlar üzerinde %28 doğruluk oranında performans artışı sağlanmıştır.

(ELMo ile ilk temel yaklaşım ile %50,39 ve ELMo Anlam Öbek Esaslı Ortalama Benzerlik Modeli ile %78,06). Alınan ilk sonuçlara neticesinde, önerilen yaklaşımın özellikle DKD'ler yönelik ABG veri kümesi oluşturmak için gelecek vaat eden ettiği gösterilmiştir. Bu makale [18]'deki çalışmamızın genişletilmiş bir versiyonudur.

**Anahtar Sözcükler:** veri artırımı, anlam belirsizliği giderme, derin öğrenme, bağlam vektörü

### Abstract

Recently, neural architectures play a significant role in the task of Word Sense Disambiguation (WSD). Supervised methods seem to be ahead of its rivals and their performance mostly depends on the size of training data. A numerous number of human-annotated data available for WSD task have been constructed for English. However, low-resource languages (LRLs) still face difficulty in finding suitable data resources. Gathering and annotating a sufficient amount of training data is a time consuming and labor-expensive work. To address and overcome this problem, in this paper we investigate the possibility of using a semi-supervised context based WSD approach for data augmentation (in order to be later used for supervised learning). Since, it is even difficult to find WSD evaluation datasets for LRLs, in this study, we use English datasets to build a proof-of-concept and to evaluate their applicability onto LRLs. Our semi-supervised approach uses a seed set and context embeddings. We test with 9 different context based language models (including ELMo, BERT, RoBERTa etc.) and investigate their impacts on WSD. We increased our baseline results up to 28 percentage point improvements (baseline with ELMo 50.39% and

Gönderme ve kabul tarihi: 04.12.2020 - 22.02.2021  
Makale türü: Araştırma

*ELMo Sense Seed Based Average Similarity Model 78.06%) in terms of accuracy. Our initial findings reveal that the proposed approach is very promising for the augmentation of WSD datasets of LRLs. This study is an extension version of the work from [18].*

**Keywords:** data augmentation, word sense disambiguation, deep learning, contextual embeddings

## 1. Giriş

Anlam Belirsizliği Giderme (ABG) doğal dillerde yaygın olarak gözlenen bir sözcüğün birden fazla anlama sahip olması ve doğru anlamının bağlam içerisinde anlaşılıp otomatik olarak seçilmesi işlemidir. ABG, popüler ve birçok araştırmacı tarafından incelenen bir problemdir. Bu konudaki çalışmaların başlangıcı, makine çevirisi araştırmalarının başlangıcına dayanır. Doğal dillerin gösterimlerdeki çeşitlilik ve zenginliği nedeniyle ABG Doğal Dil İşleme'nin (DDİ) zorlu yaklaşımlarındandır. Sözcükler cümle içerisinde birden fazla anlamda bulunabilirler; çokanlamlılık (sözcüklerin anlam genişlemesi yoluyla yeni anlamlar kazanmasına ve aynı sözcüğün cümle içerisinde birden çok anlamda kullanılabilmesine durumu) veya eş seslilik (yazılışları ve okunuşları aynı anlamları ise farklı sözcükleri içeren durumu).

ABG yaklaşımı için birden fazla yöntem öne atılmış ve test edilmiştir, bunlar: denetimli, denetimsiz ve bilgi-tabanlı [1], [2] yöntemlerdir. Yarowsky [3]'deki ABG araştırmasında denetimsiz öğrenme algoritması önermiştir [4], [5]. Yakın zamanda yapay zekâ ve derin öğrenme yöntemleri DDİ araştırmalarında da kayda değer rol oynamaktadır. Derin öğrenme yöntemlerinden elde edilen yüksek başarımlar araştırmacıları bu yöntem ve yaklaşımların DDİ problemlerine de uygulanabilirliğini denemeye teşvik etmiştir [6], [7]. Ancak derin öğrenmeye dayalı sınıflandırıcıların başarılı performans verebilmeleri için büyük miktarda etiketli veriye ihtiyaç duyulmaktadır. Ek olarak, sınıflandırıcıların genelleme başarımları eğitim verisinin büyüklüğüne ve niteliğine bağlıdır. İngilizce dilinde ABG yaklaşımına ait çok miktarda etiketli veri çevrim içi olarak bulunmakta. Bunlardan en popüler olanları SemCor ve OMSTI'dir [8], [9]. Ancak DKD'ler uygun etiketli veri bulma konusunda sıkıntı çekilmektedir.

Otomatik veri artırımı, özellikle Ses Tanıma [10], [11] ve Bilgisayarlı Görü [12]–[14] alanlarında kullanılan ve veri eksikliği durumları için kullanılabilir.

popüler yöntemlerden biridir. Ancak DDİ'de bu yöntemin uygulanabilirliği doğal dillerin zenginliğinden ve karmaşıklığından dolayı daha zor olmaktadır. Örneğin Bilgisayarlı Görü alanında bir fotoğrafın ters-yüz döndürülmüş olması o fotoğrafın ait olduğu etiket sınıfını değiştirecektir. Ancak DDİ'de semantik anlamını ve kalitesini kaybetmeden tüm sözcüklere uygulanabilecek evrensel bir kural çıkarmak çok zordur. Bu nedenle veri artırımının DDİ'de uygulanabilirliği kısıtlanmıştır. Zhang ve ark. [15] yaptıkları çalışmada WordNet [16] ontolojisini kullanarak veri artırımı yapabilmek için sözcüklerin anlamlarını anlamdaşlarıyla değiştirme yaklaşımını önermişlerdir. Benzer başka bir yaklaşım ise [17]'de yapılmış ve sözcük benzerliğini kullanarak sözcükler yer değiştirilmiştir.

Bu makalede sunulanlar [18]'deki çalışmamızın genişletilmiş bir versiyonudur. DKD'ler için ABG eğitim verisi bulmanın çok zor olması nedeniyle, [18]'de veri artırımı için yarı-denetimli bağlamsal anlam belirsizliği giderme (daha sonra denetimli öğrenmede eğitim verisi olarak kullanılmak üzere) yaklaşımı önerilmiştir. Ve bu çalışmanın doğruluğunu ispatlamak üzere, önerilen model İngilizce veri kümesi üzerinde test edilmiştir. Ancak bu yöntemin başarılı olabilmesi için DKD'ler'in mutlaka WordNet gibi bir ontolojiye sahip olmaları gerekmektedir. Oluşturduğumuz yarı-denetimli yöntem öbek kümesi (seed set) ve bağlam vektörü (context embeddings) kullanmaktadır. [18]'de anlatıldığı üzere çalışma 9 farklı bağlam-tabanlı dil modelinde test edilmiştir (ELMo, BERT, RoBERTa vb.) ve etkileri incelenmiştir. Temel yaklaşımlarımıza göre %28 doğruluk oranında performans artırışı sağlanmıştır. Alınan ilk sonuçlarımıza göre önerilen yaklaşım özellikle DKD'ler için ABG veri kümesi oluşturmak için gelecek vaat ettiği görülmüştür üzerine çalışma ilerletilmiştir. Orijinal çalışmaya ek olarak, öbek kümeleri maksimum ve ortalama benzerlik modelleri belirli bir eşik değerine göre genişletilmiş ve en yüksek başarımları veren dil modelinde test edilmiştir. Ek olarak, [18]'de tek bir iterasyon üzerinde gösterilen yöntem burada çoklu iterasyon ile tekrarlanarak önerilen çevrim tamamlanmış ve değerlendirilmesi yapılmıştır. Bu ek çalışmanın ek testleri ve sonuçları Bölüm 6'da detaylı olarak anlatılmaktadır.

Makalenin geri kalan kısmı şu şekilde düzenlenmiştir: Bölüm 2'de mevcut çalışmalar özetlenmiştir. Bölüm 3'de yarı-denetimli bağlamsal ABG modelleri detaylarıyla anlatılmaktadır. Bölüm 4'de deney düzeni anlatılmaktadır. Bölüm 5'de deneysel sonuçlar ve başarımlar değerlendirilmiştir.

Bölüm 6'da ise ek yapılan çalışmalar raporlanmıştır. Ve Bölüm 7'da sonuçlar ve gelecek çalışmalar paylaşılmıştır.

## 2. İlgili Çalışmalar

DDİ'de veri artırımı genelde veride aşırı uyuma (overfitting) problemi yaşandığında ve derin öğrenme modellerinin genelleme başarımlarını arttırmak için kullanılmaktadır. Son zamanlarda önerilen bağlamsal veri artırım yaklaşımları derin öğrenme yöntemleriyle veriyi sözcüklerin rasgele uygun başka sözcüklerle yer değiştirmesiyle zenginleştirmeye çalışmaktadır. Literatürde yapılan bazı çalışmalarda kodlayıcı-kod çözücü modeller [19-21] ile oto-kodlayıcıların yapay kod çözücüleri kullanılarak yapılan modeller [22-24] kullanılmıştır. Aynı zamanda yine veri artırımı çalışmalarında eşanlımlılar listesi [15], [17], gramer indüksiyonu [25] ve eyleme özgü buluşsal kural ekleme gibi [26]–[28] yaklaşımlar da kullanılmıştır. Melamud ve ark. [29] ve Peters ve ark. [30] çalışmalarında çift yönlü dil modellerinde bağlam kullanımlarının DDİ'de etkili olduğunu göstermişlerdir. Kobayashi ve ark. [30], [31] yaptığı çalışmalarda ise boşluk doldurma yaklaşımı yine yapay sinir ağları modellerinde incelenmiştir. [33]'de yapılan çalışmada ise veri artırımı yöntemi için bağlamsal artırım yaklaşımı önerilmiştir. Yaklaşımlarında cümlelerin, içerisindeki sözcüklerin başka sözcükler ile yatay ilişkiler bağlamında yer değiştirmesi halinde dahi hala doğal dil olarak algılanacağı farz edilmiştir. Sözcükleri cümle içerisindeki pozisyon bilgilerine göre çift-yönlü dil modeli tarafından tahmin edilen sözcüklerle yer değiştirmişlerdir. Çalışmaları, Kolomyets ve ark. [34] Fadaee ve ark. [35] yaptığı çalışmalara benzerlik göstermektedir. Ayrıca bu son üç çalışma dil modellerini veri artırımı çalışmalarında kullanarak ortak özellik sergilemektedir. Lala ve ark. [36] veri artırımı yaklaşımını İngilizce-Çek dili arasında çeviri yapabilmek adına kullanmıştır.

## 3. Yarı-Denetimli Bağlamsal ABG Modelleri

Önerilen yaklaşım bilgi-tabanlı Lesk algoritmasındaki [1] yaklaşım ile Yarowsky [3]'nin yarı-denetimli modelindeki yaklaşım baz alınarak oluşturulmuştur. Lesk algoritması sözcüklerin verilen “komşu bölge”

(metnin bir kesiti) ile ortak bir konu ile ilgili olma eğilimine sahip olma varsayımı üzerine dayanmaktadır. Algoritmada bir sözcüğün doğru anlamı, o sözcüğün sözlükteki her bir anlamının ayrı ayrı sözcük ile yer değiştirerek ve sözlük içerisindeki tanımı ile verilen bağlam karşılaştırıldığında en çok örtüşmeyi veren sözcüğün seçilmesiyle belirlenmektedir. Yarowsky [3] çalışmasında, bir sözcüğün farklı iki anlamı için temsili bir grup öbek örnekleri belirlenirse bu öbek örnekleri, sözcüklerin bağlam içerisinde bir anlam-söylem ve anlam-sıralama eğilimlerinden dolayı aşamalı olarak artırım sağlanabileceğini göstermiştir. Bu makalede, benzer şekilde bağlam vektörü kullanılarak öbek örnekleri artırılmıştır.

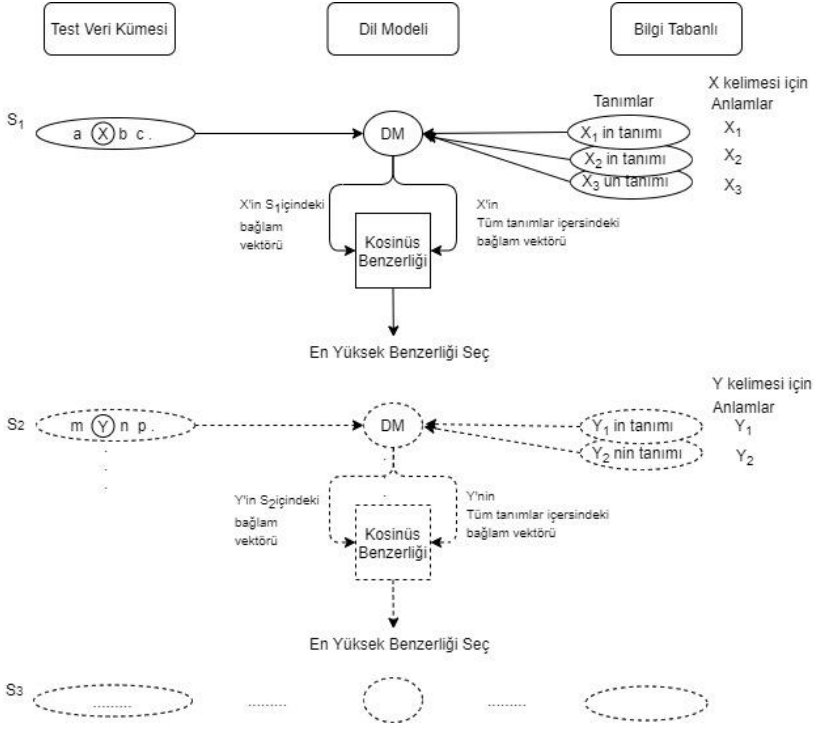
### 3.1 Tanım Esaslı Mod

Temel bir yaklaşım olması adına ilk olarak bilgi tabanlı (WordNet'deki tanım bilgileri, detaylar için Bölüm 4.1 bakınız) sistemden elde ettiğimiz girdilerle tanım esaslı bir model oluşturulmuştur. Hedef sözcüğün doğru anlamını bulabilmek için sözcüğün anlamlarının tanımları (bilgi tabanlı sistemden gelen) ile hedef sözcüğün cümle içerisindeki bağlamı (hedef sözcüğün test kümesinde yer aldığı cümle) önceden eğitilmiş dil modeline verilmiştir. Sonrasında kosinüs benzerliğine göre en yüksek olasılığı veren anlam hedef sözcüğün doğru anlamı olarak belirlenmiştir.

Model detaylı olarak Şekil-1'de gösterilmektedir. X hedef sözcüğünün bilgi tabanlı sistemden elde edilen tanımları ( $X_1$ ,  $X_2$  ve  $X_3$ ) bağlam vektörleri elde edilmek üzere dil modeline verilmiştir.  $X$ 'in test cümlesi ( $S_1$ ) içindeki bağlam vektörü ile  $X$ 'in tüm tanımlar içerisindeki bağlam vektörleri kosinüs benzerliği ile karşılaştırılmış ve en yüksek benzerlik olasılığını veren anlam  $X$  hedef sözcüğünün tahmin edilen anlamı olarak seçilmiştir.

### 3.2 Anlam Öbek Esaslı Modeller

Tanım esaslı modele karşıt bir model olarak, Anlam Öbek Esaslı modeller etiketli örnek cümleleri kullanmaktadır ve hedef sözcüklerin tanımları yerine bu örnek cümleler ile bağlam test cümlesini kosinüs benzerliğine göre karşılaştırmaktadır. Bu modeller için, her bir hedef sözcük için (eğer varsa) öbek kümesi oluşturulmuştur.



Şekil-1: Tanım Esaslı Model

### 3.2.1 Anlam Öbek Esaslı Maksimum Benzerlik Modeli

modeline verilmektedir. Bu dil modelinden gelen bağlam vektörleri sonrasında kosinüs benzerliği ile karşılaştırılmaktadır. Bu benzerlik karşılaştırılmasında maksimum benzerliği veren bağlam vektörü X hedef sözcüğünün tahmin edilen anlamı olarak seçilmiştir. Başka bir deyişle en benzer örnek cümlelerin hedef sözcüğünün doğru öbek kümesinde olduğu varsayılmaktadır. Modelin metodolojisi Şekil-2'nin üst kısmında verilmiştir.

### 3.2.2 Anlam Öbek Esaslı Ortalama Benzerlik Modeli

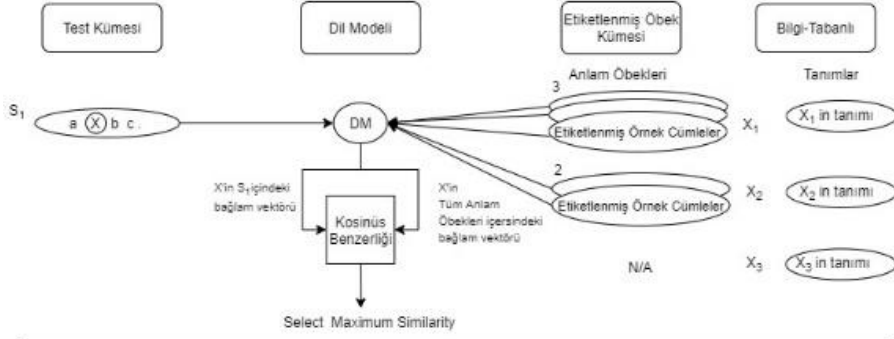
Bir önceki modelden (maksimum benzerlik) farklı olarak tahmin edilen anlam, benzerlik skorunun her bir öbek kümesinden gelen benzerlik oranının ortalaması alınarak belirlenmektedir. Modelin metodolojisi Şekil-2'nin alt kısmında verilmiştir. Test cümlesindeki ( $S_2$ ) birden fazla anlamı olan ( $Y_1$  ve  $Y_2$ )

Y hedef sözcüğünün doğru anlamının bulunabilmesi için tüm öbek kümelerindeki etiketlenmiş örnek cümleler ( $Y_1$  anlamı için 3 örnek,  $Y_2$  anlamı için 2 örnek) seçilen dil modeline bağlam vektörleri çıkarılmak üzere verilmektedir. Elde edilen bağlam vektörleri test cümlesinin bağlam vektörü ile kosinüs benzerliğine göre karşılaştırılmaktadır. Sonra bu benzerlik skorları her bir öbek kümesinin içerisinde ortalaması alınarak ortalama benzerlik elde edilmektedir. Son olarak her anlamdan gelen ortalama benzerlik skorları arasından maksimum benzerliği veren anlam tahmin edilen anlam olarak seçilmektedir.

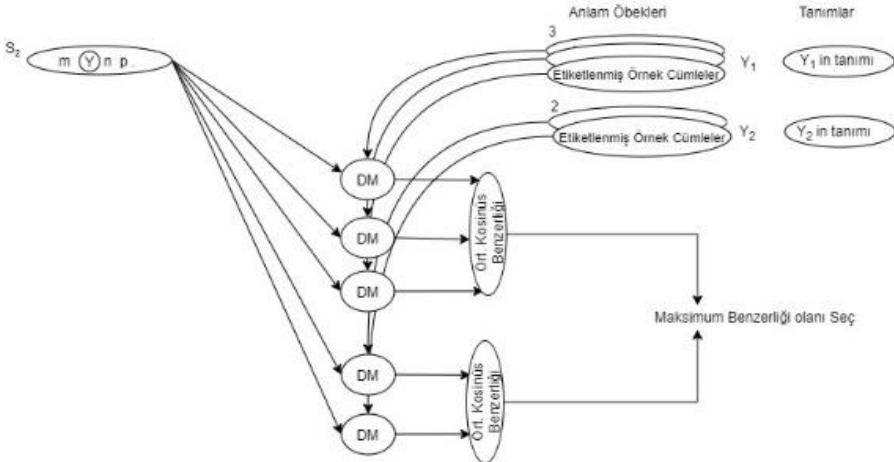
## 4. Deney Düzenegi

WordNet birden fazla anlama sahip her sözcüğe ait tanım bilgisini ve/veya örnek cümle bilgisini içeren bir ontolojidir. Bu ontoloji birden fazla dil (bir kısmı da DKD'ler için) için üretilmiş ve araştırmacılarının kullanımına sunulmuştur. Bu nedenle yaklaşımımızda tanımlara ve etiketli örnek cümlelere ait bağlam vektörleri kullanılmıştır (detaylar için

### Anlam Öbek Esaslı Maksimum Benzerlik Modeli



### Anlam Öbek Esaslı Ortalama Benzerlik Modeli



Şekil-2: Anlam Öbek Esaslı Modeller

Bölüm-3'e bakınız). Ancak bazı durumlarda, tanım bilgisi kaliteli bir etiketleme yapılabilmesi için yeteri kadar detaylı belirtilmemiştir. WordNet'de tüm sözcüklere ait tanım bilgisi olmasına rağmen örnek cümle içeren sözcük sayısı oldukça düşüktür. Ek olarak, eldeki örnek cümleler kaliteli bağlam vektörlerinin elde edilebilirliği için de yetersizdir (örn. verilen örnek cümle ya sözcüğü doğrudan içermemekte yada tam formunu buldurmamaktadır). Bu nedenle, WordNet ontolojisi çalışmamızda doğrudan kullanılamamaktadır. Bu problemi aşmak için, Yarowsky[3]'de yapılan çalışmadan esinlenerek SemCor veri kümesinden elde edilen elle işaretlenmiş örnek cümleler kullanılıp öbek kümeleri oluşturulmuştur. Yaklaşımımız, insanlar tarafından

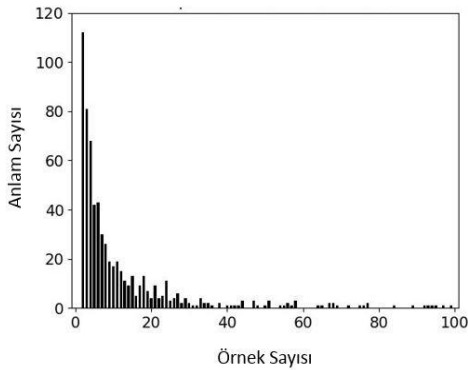
oluşturulmuş ontolojideki (WordNet) her anlam için elle işaretlenmiş öbek kümelerini, daha sonra denetimli öğrenmede kullanılmak amacıyla, genişletmeyi ve yarı-denetimli bağlamsal anlam belirsizliği giderme yaklaşımını üretmeyi amaçlamaktadır. Bu bölümde yaklaşımımızda kullanılan veri kümeleri ve önceden eğitilmiş dil modelleri anlatılmaktadır.

## 4.1 Veri Kümeleri

WordNet [16] İngilizce için hazırlanmış çok büyük ve sözcüksel bir veri tabanıdır. İsimler, filler, sıfatlar ve zarflar farklı kavramları belirten eş kümelerce gruplanmıştır. WordNet aynı zamanda birden fazla anlama sahip her sözcüğe ait tanım bilgisini ve/veya örnek cümle bilgisini içermektedir. Ancak

WordNet’de her bir anlam için yeterli örnek cümle bulunmamaktadır. SemCor [8], bildiğimiz kadarı ile, WordNet anlamları kullanarak insanlar tarafından elle oluşturulmuş en büyük veri kümesidir. Ayrıca, Semcor İngilizce denetimli ABG problemi için kullanılan ana veri kümelerinden biridir. SemCor 226.040 anlam etiketi içermektedir. Önerdiğimiz modelde ise SemCor’un 3.0 versiyonu WordNet’in 3.0 versiyonu ile birlikte kullanılmıştır.

Yaklaşımımızda WordNet anlamları/tanımları bilgi tabanlı sistem görevi görecektir şekilde tasarlanmıştır. Bilgi tabanlı sistemin detayları Bölüm 3’de anlatılmıştır. Ancak modellerin performansının ölçülebilmesi için, bir sözcüğün bir anlamın en az iki etiketli örnek cümlesi (SemCor’un bir alt kümesi ile WordNet’deki örnekler) olması beklenmektedir ve çalışmada bu sözcüklere odaklanılmıştır. Bu sayede örneklerden en az biri test kümesinde diğeri ise öbek kümesinde yer alabilecektir. Sonuç olarak, test kümemizde bu şarta uyan (en az iki etiketli cümle içeren anlam) anlamlar içerisinde 629 toplam örnek bulunmaktadır. Öbek kümesinde ise her bir anlam için bir veya birden fazla örnek (11398 örnek) olabilmektedir. Şekil-3 öbek kümemizin dağılımı verilmektedir.



Şekil-3: Öbek Kümelerinin Dağılımı

Y-ekseni toplam anlam sayısını (öbek kümesi) göstermektedir ve daha önce belirtildiği üzere 629’dur. X-ekseni ise bu kümelerdeki örnek sayısını göstermektedir. Şekil-3’ten anlaşılacağı üzere pek çok farklı anlam kümesi toplamda 6 örnek cümleden daha az örnek içermektedir. Örneğin, toplam 112 anlam kümesinde (629’un içerisinde) sadece 2 örnek cümle yer almaktadır. Çizelge-1’de bu seçim stratejisine bir örnek verilmiştir. Çizelge-1’de WordNet’den elde

edilen birden fazla anlama sahip “zero” (sıfır) sözcüğünün anlamları ve bu anlamlara ait etiketli örnek sayıları verilmiştir. WordNet’de anlamlar belli bir düzene göre kodlanmaktadır ve temsil edilmektedir.

Örneğin, “zero%1:06:00” kodlamasında “zero” kelimenin kök halini, “%1” kelimenin kaçınıcı ana anlam kümesine ait olduğunu, “:” sonra gelen “06” hangi alt anlam kümesine ait olduğunu, “00” ise anlam anahtarını göstermektedir.

Çizelge-1’den görüldüğü üzere, “zero” sözcüğü WordNet’de 9 farklı alt anlama sahiptir. Ancak sadece 4 farklı anlam için toplamda 15 tane etiketlenmiş örnek cümleye sahiptir. Yaklaşımımızda bu 4 anlam kümesinden sadece 3 farklı anlamı çıkarılmıştır, çünkü anlamlardan biri (“zero%5:00:02” ait) sadece bir tane etiketli örnek cümlesi içermektedir.

Çizelge-1: Birden Fazla Anlama Sahip “zero” Sözcüğünün Anlamları ve Örnek Sayıları

Sözcük	Anlamları	Etiketlenmiş Örnek Cümle Sayısı	Seçim
“zero”	zero%1:06:00	0	X
	zero%1:09:00	0	X
	zero%1:23:01	3	✓
	zero%1:23:02	6	✓
	zero%2:30:00	0	X
	zero%2:30:01	0	X
	zero%3:01:00	0	X
	zero%5:00:00	5	✓
	zero%5:00:02	1	X

## 4.2 Önceden Eğitilmiş Dil Modelleri

Çalışmadaki yaklaşımımıza en uygun olanı bulabilmek için testler 9 farklı bağlamsal dil modelin (farklı parametrelerle eğitilmiş modeller göz önüne alındığında toplam 16 farklı model) üzerinde yapılmıştır. Bu dil modelleri açık erişim kullanımlı olarak çevrim içi bulunmaktadır ve yaklaşımımızda sözcük temsillerini daha anlamlı yapabilmek adına bağlam vektörlerini elde etmek üzere kullanılmıştır. BERT [37], transformatörlerden elde edilen çift yönlü kodlayıcı temsilleridir ve önceden eğitilmiş derin öğrenme temsillerini etiketsiz veri üzerinde sağdan sola ve soldan sağa bağlamları çıkararak ilerlemektedir. ELMo [30] ise yeni geliştirilen derin öğrenme sözcük temsillerini kapsamaktadır. Hem



sözcüklerin kompleks özelliklerini hem de dilsel bağlamları modelleyerek kullanmaktadır. Bunlara ek olarak OPENAI-GPT [38], OPENAI-GPT2 [39], Transformer-XL [40], XLNet [41], XLM [42], RoBERTa [43] ve DistilBERT [44] gibi diğer dil modelleri de çalışmamızda kullanılmış ve test edilmiştir.

## 5. Deney Sonuçları ve Tartışma

Tanım Esaslı temel yaklaşım modeli 9 farklı dil modeli kullanılarak test edilmiştir. Bu testlerin doğruluk sonuçları Çizelge-2’de verilmiştir. Bazı dil modellerinin birden fazla alt modeli bulunmaktadır. Örneğin, BERT dil modelinin çevrim içi olarak kullanılabilen 22 farklı değişken ve dillerde eğitilmiş modeli bulunmaktadır. Bu çalışmada sadece iki farklı modeli için test yapılmıştır: *base\_cased* ve *large\_cased*.

Bu alt modellerden, *base\_cased* modeli 12-katmanlı, 768-gizli katman, 12-baş, 109 milyon değişkenden oluşmakta, *large\_cased* modeli ise 24-katmanlı, 1024-gizli katman, 16-baş, 335 milyon değişkenden oluşmaktadır. İki dil modeli de İngilizce dili üzerinde eğitilmiş modellerdir. Diğer dil modellerine ait alt modellerin de detaylı anlatımları Huggingface<sup>1</sup> web sitesinde bulunmaktadır.

Çizelge-2’den görüldüğü üzere temel yaklaşım olarak kullandığımız modelde en yüksek sonucu **%50,39** ile diğer dil modellerine kıyasla ELMo vermektedir. Aynı zamanda anlam öbek esaslı modeller de yüksek performans veren dil modelleriyle denenmiş ve sonuçlar Çizelge-3’de verilmiştir.

Çizelge-3’den de anlaşıldığı üzere, Anlam Öbek Esaslı Modellerden Ortalama Benzerlik Modeli Maksimum Benzerlik Modellerinden neredeyse tüm dil modellerinde (GPT-2<sub>large</sub> ve BERT<sub>large</sub> modelleri hariç) daha yüksek sonuç vermiştir.

**Çizelge-2: Tüm Dil Modellerindeki Tanım Esaslı Model Sonuçları**

Dil Modeli	Model	Doğruluk
BERT	<i>base_cased</i>	%45,62
	<i>large_cased</i>	%42,13
ELMo	<i>base</i>	<b>%50,39</b>
GPT	<i>openai_gpt</i>	%49,44
GPT2	<i>gpt2</i>	%43,24
	<i>gpt2_medium</i>	%41,17
	<i>gpt2_large</i>	%47,85
Transformer-XL	<i>transfo_xl_wt103</i>	%48,17
XLNet	<i>xlnet_base_case_d</i>	%42,76
	<i>xlnet_large_cased</i>	%42,44
XLM	<i>xlm_mlm_en_2048</i>	%44,51
	<i>xlm_mlm_100_1280</i>	%41,97
RoBERTa	<i>roberta_base</i>	%45,15
	<i>roberta_large</i>	%42,76
DistilBERT	<i>base</i>	%46,10
	<i>base_distilled</i>	%43,40

Ancak yine de tüm dil modelleri tanım esaslı modele kıyasla doğruluk skorlarını artırmıştır (BERT %45,62’den %58,35’e (ortalama benzerlik modeli), Transformer-XL %48,17’den %71,06’e çıkarmıştır). Ve en yüksek skorlar ise ELMo dil modelinde **%78,06** ile ortalama benzerlik modelinden ve **%74,00** maksimum benzerlik modelinden elde edilmiştir.

BERT ve ELMo dil modellerinin arasındaki en büyük fark BERT çift yönlü dönüştürücü kullanmaktır. Ancak ELMo ayrı olarak eğitilmiş soldan sağa ve sağdan sola uzun kısa süreli belleklerin ard arda kullanılmasıyla çalışmaktadır [37].

Clark’ın çalışmasında da bahsedildiği üzere [45], BERT daha çok söz dizimsel ve eşgöndergelere öncelik vererek çalışmaktadır. Ancak Anlam Belirsizliği Giderme çalışmaları sözcüklerin söz dizimsel bakış açılarından çok anlamsal yaklaşımları üzerinde odaklanmaktadır.

1

[https://huggingface.co/transformers/pretrained\\_models.html](https://huggingface.co/transformers/pretrained_models.html)

**Çizelge-3: Tüm Dil Modellerindeki Anlam Öbek Esaslı Modeller**

Dil Modeli	Maksimum Benzerlik Modeli	Ortalama Benzerlik Modeli
BERT <sub>base</sub>	%56,44	%58,35
BERT <sub>large</sub>	%48,64	%45,79
ELMo	<b>%74,00</b>	<b>%78,06</b>
GPT	%54,05	%54,68
GPT2 <sub>large</sub>	%52,94	%52,62
Transformer-XL	%67,24	%71,06
DistilBERT <sub>base</sub>	%56,75	%59,30

Bu nedenle BERT veya BERT tabanlı dil modelleri ELMo'ya kıyasla ABG yaklaşımında daha düşük performans sonuçları vermektedir. Bu testlerin sonucunda ileriki çalışmalar için şu notlar alınmıştır:

- 1) Sözcüklerin sadece tanım bilgileri kullanarak bağlam vektörlerinin çıkarılması yeterli değildir,
- 2) BERT ve BERT tabanlı diğer dil modelleri daha çok söz dizimsel bakış açılarına odaklanmaktadır, bu nedenle ABG yaklaşımı için ELMo dil modelini kullanmak daha uygundur,
- 3) Performans sonuçlarının %28 oranında artması anlam öbek kümelerinin kullanımının bağlam bazlı yaklaşımlar için uygun olduğunu göstermektedir,
- 4) İlk sonuçlar göstermektedir ki, en azından bir WordNet ontolojisine sahip DKD'ler için bu yaklaşımı kullanmak gelecek vaat etmektedir.

Anlam Öbek Esaslı Modellerin veri artırımı konusundaki başarımlarını daha derin inceleyebilmek adına farklı bir kaç test daha yapılmıştır. Deneylerde en başarılı sonuç veren dil modeli (ELMo) kullanılmıştır. Bu deneylerde farklı eşik değerlerde (0'dan 0,9'a kadar) benzerlik skorları kullanılmıştır ve bu benzerlik eşik değerinin üzerinde kalan örnek sayıları listelenmiştir (629 test örneği 0 eşik değerinin üzerindedir). Bu testlerin sonuçları Çizelge-4 ve Çizelge-5'de verilmiştir. Örneğin, 0,90 benzerlik eşik değerine sahip bir değerde, önerilen model verilen iki bağlam vektöründe kosinüs benzerliğine göre %90 oranında benzerlik gösterdiğini garanti etmektedir.

Çizelge-4 ve Çizelge-5'de görüldüğü üzere, 0,90 benzerlik eşik değerine sahip etiketli örnek sayısı oldukça azdır (Maksimum benzerlik modeli için 18,

Ortalama benzerlik modeli için 4). Ancak yüksek eşit değerlerinde elde edilen yüksek performans, öbek kümelerinin bu yöntem ile veri artırımı yapılabileceğini göstermektedir. ELMo Maksimum Benzerlik modelinde 18 örnek cümle içerisinde %9 performans başarımları ile skor %83,33 olarak alınmıştır. Veri artırımı için öbek kümelerini doldurmak adına yüksek bir eşik değeriyle başlanması özellikle DKD'ler üzerinde kullanılabilir anlamında teşvik edicidir.

**Çizelge-4: ELMo Anlam Öbek Esaslı Maksimum Benzerlik Modeli**

Eşik Değeri	Test Örnek Sayısı	Doğruluk
0,90	18	% 83,33
0,80	100	%79,00
0,70	320	%76,25
0,60	502	%74,90
0,50	594	%73,74
0,40	621	%73,91
0	629	%74,00

**Çizelge-5: ELMo Anlam Öbek Esaslı Ortalama Benzerlik Modeli**

Eşik Değeri	Test Örnek Sayısı	Doğruluk
0,90	4	% 75,00
0,80	94	%76,60
0,70	315	%78,10
0,60	499	%78,16
0,50	592	%78,88
0,40	624	%78,52
0	629	%78,06

## 6. Yapılan Ek Çalışmalar

Daha önce yapılan testler sonucunda, veri artırımı sağlamak için öbek kümelerini doldurulması yaklaşımında belirli bir eşik değeriyle başlanması teşvik edici olması nedeniyle bu çalışmada daha da detaylı testler yapılmıştır. Daha önce yaptığımız çalışmalara ek olarak, WordNet'de bulunan bir sözcüğün tüm anlamları için öbek kümeleri oluşturulmuş ve otomatik olarak anlam etiketlemesi yapılmış OMSTI [9] veri kümesi kullanılarak öbek kümeleri benzer örneklerle doldurulmuştur. Bu amaçla en mantıklı eşik değeri belirlenmek üzere birden fazla ve birkaç tur test yapılmıştır. Öbek kümeleri 4 farklı yaklaşım modeli üzerine doldurulmuştur. Birinci modelde öbek kümeleri ortalama kosinüs benzerlik yaklaşımı kullanılarak eğitim kümesi oluşturulması sağlanmıştır ve 629'luk test kümesinde yine ortalama kosinüs benzerlik kullanılarak doğru anlam ataması yapılmışa sağlanmıştır. İkinci modelde ise yine ortalama



benzerlik yaklaşımı ile eğitim kümesi oluşturulmuş ancak bu sefer test aşamasında ise maksimum benzerlik kullanılarak atama yapılmıştır. 3. ve 4. modelde bu yaklaşımların tam terslerinden oluşmaktadır. Bu dört test içinde eğitim kümesi sadece bir tur oluşturulmuş ve belirlenen eşik değerlerine göre atamaları yapılmıştır. Bu çalışmanın sonuçları Çizelge-6'da verilmiştir.

Bu modellerle en yüksek doğruluk sonuçlarını 1. ve 4. model %90 benzerlik ile (%78,06) şeklinde vermektedir. Ancak sonuçlar detaylı incelendiğinde, öbek kümeleri doldurulurken eşik değerinin 0,9 seçilmesiyle (sadece %90 benzerlik olan cümlelerin kümeye dahil edilmesi) yeterli kadar örnek cümle gelmediği görülmektedir. Örneğin Maksimum kosinüs benzerliği ile eğitim kümesi oluşturulurken 0,9 eşik değeri ile toplam 9, ortalama benzerlik ile de toplam 75 cümle örneği gelmiştir. Bu cümle sayıları 0,8 eşik değerinde maksimum benzerlik ile toplam 9,601 cümle iken ortalama benzerlik ile toplam 3,614'tür. Bu nedenle sonraki deneylerde eşik değerinin en az 0,8 olması gerektiği kabul edilmiş ve bu eşik değeriyle en yüksek sonucu 4. modelde (Maks. Benz. Eğitim Ort. Benz.Test) **%77,26** olarak elde edilmiştir.

Yani, eğitim kümesi maksimum benzerlik kullanılarak oluşturulacak test kümesi ise ortalama benzerlik kullanılarak atama sonuçlandırılacaktır. Bu model seçilerek maksimum benzerlik eğitim küme oluşturulması 5 tur olarak tekrar edilmiştir. Yani, 1. turda %80 maksimum benzerlik kullanılarak öbek kümeleri doldurulmuş ve buna göre eğitim kümesi oluşturulmuş ve 2. tura geldiğinde ise 1. turdan elde edilen eğitim modelinin üzerine tekrar %80 benzerlik ile öbek kümeleri doldurulmuştur. Her bir turda eski eğitim kümesinin üzerine yeni örnekler katılarak öbek kümeleri genişletilmiştir. Bu eğitim kümeleri kullanılarak maksimum benzerlik ve ortalama benzerlik yaklaşımı ile de sözcük anlamları etiketlenmiş ve sonuçlar raporlanmıştır. Bu testlerin sonuçları Çizelge-7'de verilmiştir. Bu yaklaşım ile en yüksek sonuçlar tur sayısının 2 kez olması ile **%77,42** ile sonuçlanmıştır. Daha önceki yaklaşımımızda maksimum benzerlik modeli ile en yüksek sonucu %74,00 olarak elde edilmişti, bu çalışma ile %3'lük bir iyileştirme sağlanmış olundu. Ve sonuçlar detaylı incelendiğinde 2.turdan sonra başarımın düşmesinin en büyük nedeni ise belirli bir tur sayısından sonra eğitim kümesinin çok fazla yakınsamasından kaynaklandığı tespit edilmiştir.

**Çizelge-6: Ortalama ve Maksimum Benzerlik Modelleriyle Eğitim ve Test Sonuçları**

Model No	Açıklama	Eşik Değeri	Doğruluk
1	Ort. Benz. Eğitim Ort. Benz.Test	0,9	<b>%78,06</b>
		0,8	%77,12
		0,7	%76,47
2	Ort. Benz. Eğitim Maks. Benz.Test	0,9	%66,29
		0,8	%68,04
		0,7	%65,18
3	Maks. Benz. Eğitim Maks. Benz.Test	0,9	%66,29
		0,8	%68,36
		0,7	%66,93
4	Maks. Benz. Eğitim Ort. Benz.Test	0,9	<b>%78,06</b>
		0,8	<b>%77,26</b>
		0,7	%77,06

OMSTI veri kümesi kullanılarak yapılan otomatik veri artırımı yöntemi detaylı incelendiğinde 2. tur sonucunda 629 (test kümesi) anlam içinden aslında 154 tane anlam için yeni örnek geldiği görülmüştür. Bu 154 anlam içerisinde ise örnek sayısı 5'den az olup artan 14 anlam bulunmakta, örnek sayısı 10'dan az olup artan ise 32 anlam bulunmaktadır. Bu tur sayısı 5 olduğu halde bile toplam eklenen anlam sadece bu 154 tanesi içindir. 5.Tur sonucunda bu 154 anlam içerisinde örnek sayısı 5'den az olup artan 15 anlam (3 turda sadece 1 anlam için örnek artışı oldu) bulunmakta, örnek sayısı 10'dan az olup artan ise 38 anlam (3 turda sadece 6 anlam için örnek artışı oldu) bulunmaktadır. Başarımları daha detaylı incelemek üzere de sadece artışı sağlanan bu 154 anlam kümesi üzerinde detaylı bir analiz yapılmıştır. Sonuçları Çizelge-8'de gösterilmiştir.

**Çizelge-7: Maksimum Benzerlik ile Eğitim Ortalama Benzerlik ve Maksimum Benzerlik İle Test Sonuçları (Eşik Değeri = 0,8)**

Tur Sayısı	Maks. Benz. Eğitim Ort. Benz.Test	Maks. Benz. Eğitim Maks. Benz.Test	Cümle Sayısı
1	%77,26	%68,36	~ 9,601
2	<b>%77,42</b>	%67,72	~ 46,969
3	%76,78	%67,40	~ 54,306
4	%76,62	%67,09	~ 59,411
5	%76,31	%67,40	~ 63,244

Çizelge-4 ve Çizelge-5’de görüldüğü üzere, 0,90 benzerlik eşik değerine sahip etiketli örnek sayısı oldukça azdır (Maksimum benzerlik modeli için 18, Ortalama benzerlik modeli için 4). Ancak yüksek eşit değerlerinde elde edilen yüksek performans, öbek kümelerinin bu yöntem ile veri artırımı yapılabileceğini göstermektedir. ELMO Maksimum Benzerlik modelinde 18 örnek cümle içerisinde %9 performans başarımları ile skor %83,33 olarak alınmıştır. Veri artırımı için öbek kümelerini doldurmak adına yüksek bir eşik değeriyle başlanması özellikle DKD’ler üzerinde kullanılabilir anlamında teşvik edicidir.

Eğitim kümesinin iyileştirilmesi ile örnek cümle eklenen 154 anlam kümesi için yapılan testlerde görülmüştür ki, en yüksek sonuçlar **%60,38** eğitim kümesinin 2 tur döndürülmesi ile elde edilmiştir. Çizelgede 0 tur ile gösterilen kısım OMSTI veri kümesi kullanılmadan sadece SEMCOR veri kümesi kullanılarak elde edilen eğitim kümesinin sonuçlarıdır. Bu 154’lük anlam kümesi için toplam 4,484 adet cümle bulunmaktadır, tur sayıları arttıkça gelen cümle sayısı da artmakta ancak bir süre sonra fazla yakınsamasından dolayı başarımlar düşmektedir. Örneğin, 1. tur ile yeni 3,416 cümle gelerek bu 154’lük kümenin toplam örnek sayısını 7,900 çıkarmıştır. Aynı şekilde 2. turda 15,248 yeni cümle gelerek toplam cümle sayısı 23,148 çıkmıştır. Ancak bu turdan sonra başarımlar zirve yaptığı orandan düşmekte yine de elde edilen ilk sonuca göre (%57,79) başarımlar yaklaşık olarak %3 artmıştır. Bu çalışmalar göstermiştir ki, 0,8 eşik değeri ve 2. tur ile kayda değer başarımlar alınabilecektir.

**Çizelge-8: 154 Ek Anlam Kümesi için Doğruluk Sonuçları (Eşik Değeri = 0,8)**

Tur Sayısı	Doğruluk	Turda Gelen Örnek Sayısı	Toplam Örnek Sayısı
0	%57,79	0	4,484
1	%58,44	3416	7,900
2	<b>%60,38</b>	15248	23,148
3	%59,09	19598	42,746
4	%58,44	22728	65,474
5	%59,09	24527	90,001

Son olarak daha da detaylı bir analiz yapabilmek adına, SemCor’da 5’den az örneği olan anlamlar ile 10’dan az örneği olan anlamlar incelenmiştir. Sonuçları Çizelge-9 da verilmiştir. Çizelgede 0. tur olarak gösterilen kısım sadece SemCor kullanılarak elde edilen eğitim kümesi içerisinde 5’den az örneği olan anlamlar ve 10’dan az örneği olan anlamlar için doldurulmuştur. 5’den az örneği olan toplam 9 anlam için bu testler 6 turda denemiştir (0,1,2,3,4 ve 5.tur). Bu 9 anlamın başlangıçta toplam 33 örneği bulunmaktadır, en yüksek sonuçların elde edildiği 2.tur sonucunda ise örnekler 197’ye çıkmıştır. 10’dan az örneği olan toplam 25 anlam bulunmaktadır. Bu anlamlarında ilk başta 175 örneği bulunmaktadır, 2. tur sonucunda ise bu sayı 4,079 örneğe düşmüştür. Sonuçlar, 2. turdan sonra yeni örnekler gelmesine rağmen öbek kümelerinin yakınsamasından dolayı başarımlarının daha artmadığını göstermektedir. 9 anlam kümesi için 2. tur sonunda performansı %22 oranında artmış, 25 anlam kümesi için ise %8 oranında artmıştır. Bu şekilde eşik değerleri yüksek tutulup küçük kümelerde kontrollü bir başarımlar oranı artabileceği öngörülmektedir. Ve DKD’ler gibi özellikle her anlam için çok az örnek bulunabilecek diller de bu yaklaşım ile başarılı sonuçlar alınabileceği gösterilmektedir. Bu yöntem kullanılarak çok az örneklili anlam kümeleri doldurulabilir ve 2.turdan sonra en başarılı sonuçlar elde edilebilecektir.

DKD’ler için ABG eğitim verisi bulmanın çok zor olması nedeniyle, [18]’de veri artırımı için yarı-denetimli bağlamsal anlam belirsizliği giderme (daha sonra denetimli öğrenmede eğitim verisi olarak kullanılmak üzere) yaklaşımı önerilmiştir.

**Çizelge-9: 5'den ve 10'dan Az Örneği Olan Anlamların Doğruluk Sonuçları**

Tur Sayısı	5'den az örnek (9 anlam)	Örnek Sayısı	10'dan az örnek (25 anlam)	Örnek Sayısı
0	%44,44	33	%64,00	175
1	%55,55	63	%68,00	1,249
2	<b>%66,66</b>	197	<b>%72,00</b>	4,079
3	%66,66	359	%72,00	7,770
4	%66,66	534	%72,00	12,289
5	%66,66	724	%72,00	17,183

Ve bu çalışmanın doğruluğunu ispatlamak üzere, önerilen model İngilizce veri kümesi üzerinde test edilmiştir. Alınan ilk sonuçlarımıza göre önerilen yaklaşım özellikle DKD'ler için ABG veri kümesi oluşturmak için gelecek vaat ettiği görülmesi üzerine çalışma ilerletilmiştir. İlerletilen bu çalışmamızda da görüldüğü üzere bu yaklaşım DKD'ler için umut vaat etmektedir. Bu nedenle bir sonraki aşamada ilk olarak DKD'ler için bu yaklaşım test edilecektir. Ancak bu yöntemin başarılı olabilmesi için DKD'ler'in mutlaka WordNet gibi bir ontolojiye sahip olmaları gerekmektedir. DKDler arasından Türkçe de WordNet ontolojisine sahip bir dildir. Bu sebeple bir sonraki çalışmada yaklaşımlar Türkçe üzerinde test edilecektir.

## 7. Sonuç ve Gelecek Çalışmalar

Düşük-kaynaklı diller (DKD'ler) için doğal dil problemlerine uygun veri bulmak çok zordur. Anlam Belirsizliği Giderme (ABG) problemi de özellikle DKD'lerde veri eksikliği sorunu yaşamaktadır. Ancak ABG problemi için İngilizce dili üzerinde elle-etiketlenmiş çok sayıda veri çevrim içi olarak bulunmaktadır. DKD'lerde yeterli derecede probleme uygun veri toplamak ve etiketlemek vakit alıcı ve yüksek maliyet gerektiren bir iştir. DKD'ler için ABG eğitim verisi bulmanın çok zor olması nedeniyle, veri artırımı için yarı-denetimli bağlamsal anlam belirsizliği giderme (daha sonra denetimli öğrenmede eğitim verisi olarak kullanılmak üzere) yaklaşımı önerilmiştir. Bu çalışmanın doğruluğunu göstermek üzere, önerilen model İngilizce veri kümesi üzerinde test edilmiştir. Oluşturduğumuz yarı-denetimli yöntem öbek kümesi (seed set) ve bağlam vektörü (context embeddings) kullanmaktadır. Çalışma 9 farklı bağlam-tabanlı dil modelinde test edilmiştir (ELMo, BERT, RoBERTa vb.) ve etkileri incelenmiştir. Temel yaklaşımlarımıza göre %28

doğruluk oranında performans artışı sağlanmıştır. Öbek kümeleri maksimum ve ortalama benzerlik modelleri belirli bir eşik değerine göre genişletilmiş ve en yüksek başarımları veren dil modelinde test edilmiştir. En başarılı eğitim kümesi maksimum benzerlik modeli ile ve test kümesi ortalama benzerlik modeli kullanılarak yapılan yaklaşım ile elde edilmiştir. Ayrıca yine en yüksek sonuçlar 0,8 Eşik değeri ve eğitim veri kümesinin sadece 2. tur oluşturulmasıyla elde edilmiştir.

Gelecek çalışmalarda özellikle veri artırımı çalışmaları DKD'ler arasından Türkçe üzerinde denenmesi ve eğitilmiş modellerde kullanılmak üzere eğitim kümesi oluşturulması planlanmaktadır. Belirlenecek eşik değerleri yüksek tutulup küçük kümelerde kontrollü bir büyüme sağlanarak başarımın artabileceği öngörülmektedir.

## Kaynakça

- [1] Lesk, M. *Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone*, in Proceedings of the 5th annual international conference on Systems documentation. ACM, 1986, pp. 24-26.
- [2] Agirre, E. ve Soroa, A. *Personalizing pagerank for word sense disambiguation*, in Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, 2009, pp. 33-41.
- [3] Yarowsky, D. *Unsupervised word sense disambiguation rivaling supervised methods*, in 33rd annual meeting of the association for computational linguistics, 1995, pp. 189-196.
- [4] Zhong, Z. ve Ng, H. T. *It makes sense: A wide-coverage word sense disambiguation system for free text*, in Proceedings of the ACL 2010 system demonstrations, 2010, pp. 78-83.
- [5] Iacobacci, I., Pilehvar, M. T. ve Navigli, R. *Embeddings for word sense disambiguation: An evaluation study*, in Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), vol. 1, 2016, pp. 897-907.
- [6] Raganato, A., Bovi, C. D. ve Navigli, R. *Neural sequence learning models for word sense disambiguation*, in Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017, pp. 1156-1167.
- [7] Luo, F., Liu, T., Xia, Q., Chang, B. ve Sui, Z. *Incorporating glosses into neural word sense*

- disambiguation*, arXiv preprint arXiv:1805.08028, 2018.
- [8] Miller, G. A., Chodorow, M., Landes, S., Leacock, C. Ve Thomas, R. G. *Using a semantic concordance for sense identification*, in Proceedings of the workshop on Human Language Technology. Association for Computational Linguistics, 1994, pp. 240–243.
- [9] Taghipour, K. ve Ng, H. T. *One million sense-tagged instances for word sense disambiguation and induction*, in Proceedings of the nineteenth conference on computational natural language learning, 2015, pp. 338–344.
- [10] Jaitly, N. ve Hinton, G. E. *Vocal tract length perturbation (vltf) improves speech recognition*, in Proc. ICML Workshop on Deep Learning for Audio, Speech and Language, vol. 117, 2013.
- [11] Ko, T., Peddinti, V., Povey, D. Ve Khudanpur, S. *Audio augmentation for speech recognition” in Sixteenth Annual Conference of the International Speech Communication Association*, 2015.
- [12] Simard, P. Y., LeCun, Y. A., Denker, J. S. ve Victorri, B. *Transformation invariance in pattern recognition? tangent distance and tangent propagation*, in Neural networks: tricks of the trade. Springer, 1998, pp. 239–274.
- [13] Krizhevsky, A., Sutskever, I. ve Hinton, G. E. *Imagenet classification with deep convolutional neural networks*, in Advances in neural information processing systems, 2012, pp. 1097–1105.
- [14] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. ve Rabinovich, A. *Going deeper with convolutions*, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1–9.
- [15] Zhang, X., Zhao, J. ve LeCun, Y. *Character-level convolutional networks for text classification*, in Advances in neural information processing systems, 2015, pp. 649–657.
- [16] Miller, G. A. *Wordnet: a lexical database for English*, Communications of the ACM, vol. 38, no. 11, pp. 39–41, 1995.
- [17] Wang, W. Y. Ve Yang, D. *That’s so annoying!!!: A lexical and framesemantic embedding based data augmentation approach to automatic categorization of annoying behaviors using# petpeeve tweets*, in Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015, pp. 2557–2563.
- [18] Torunoğlu-Selamet, D., İnceoğlu, A. ve Eryiğit, G. *Preliminary Investigation on Using Semi-Supervised Contextual Word Sense Disambiguation for Data Augmentation*, 2020 5th International Conference on Computer Science and Engineering (UBMK), Diyarbakır, Turkey, 2020, pp. 337-342, doi: 10.1109/UBMK50275.2020.9219389.
- [19] Kim, Y. ve Rush, A. M. *Sequence-level knowledge distillation,” arXiv preprint arXiv:1606.07947*, 2016.
- [20] Sennrich, R., Haddow, B. ve Birch, A. *Improving neural machine translation models with monolingual data*, arXiv preprint arXiv:1511.06709, 2015.
- [21] Xia, Y., Qin, T., Chen, W., Bian, J., Yu, N. ve Liu, T.-Y. *Dual supervised learning*, in Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. org, 2017, pp. 3789–3798.
- [22] Bergmanis, T., Kann, K., Schutze, H. ve Goldwater, S. *Training data augmentation for low-resource morphological inflection*, in Proceedings of the CoNLL SIGMORPHON 2017 Shared Task: Universal Morphological Reinflection, 2017, pp. 31–39.
- [23] Xu, W., Sun, H., Deng, C. ve Tan, Y. *Variational autoencoder for semi-supervised text classification*, in Thirty-First AAAI Conference on Artificial Intelligence, 2017.
- [24] Hu, Z., Yang, Z., Liang, X., Salakhutdinov, R. ve Xing, E. P. *Toward controlled generation of text*, in Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR.org, 2017, pp. 1587–1596.
- [25] Jia, R. ve Liang, P. *Data recombination for neural semantic parsing*, arXiv preprint arXiv:1606.03622, 2016.
- [26] Furstenu, H. ve Lapata, M. *Semi-supervised semantic role labeling*, in Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, 2009, pp. 220–228.
- [27] Kafle, K., Yousefhusien, M. ve Kanan, C. *Data augmentation for visual question answering*, in Proceedings of the 10th International

- Conference on Natural Language Generation, 2017, pp. 198–202.
- [28] Silfverberg, M., Wiemerslage, A., Liu, L. ve Mao, L.J. *Data augmentation for morphological reinflection*, Proceedings of the CoNLL SIGMORPHON 2017 Shared Task: Universal Morphological Reinflection, pp. 90–99, 2017.
- [29] Melamud, O., Goldberger, J. ve Dagan, I. *context2vec: Learning generic context embedding with bidirectional lstm*, in Proceedings of the 20th SIGNLL conference on computational natural language learning, 2016, pp. 51–61.
- [30] Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K. ve Zettlemoyer, L. *Deep contextualized word representations*, arXiv preprint arXiv:1802.05365, 2018.
- [31] Kobayashi, S., Tian, R., Okazaki, N. ve Inui, K. *Dynamic entity representation with max-pooling improves machine reading*, in Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016, pp. 850–855.
- [32] Kobayashi, S., Okazaki, N. ve Inui, K. *A neural language model for dynamically representing the meanings of unknown words and entities in a discourse*, arXiv preprint arXiv:1709.01679, 2017.
- [33] Kobayashi, S. *Contextual augmentation: Data augmentation by words with paradigmatic relations*, arXiv preprint arXiv:1805.06201, 2018.
- [34] Kolomiyets, O., Bethard, S. ve Moens, M.-F. *Model-portability experiments for textual temporal analysis*, in Proceedings of the 49<sup>th</sup> Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2. Association for Computational Linguistics, 2011, pp. 271–276.
- [35] Fadaee, M., Bisazza, A. ve Monz, C. *Data augmentation for lowresource neural machine translation*, arXiv preprint arXiv:1705.00440, 2017.
- [36] Lala, C., Madhyastha, P. S., Scarton, C. ve Specia, L. *Sheffield submissions for wmt18 multimodal translation shared task*, in Proceedings of the Third Conference on Machine Translation: Shared Task Papers, 2018, pp. 624–631.
- [37] Devlin, J., Chang, M.-W., Lee, K. ve Toutanova, K. *Bert: Pre-training of deep bidirectional transformers for language understanding*, arXiv preprint arXiv:1810.04805, 2018.
- [38] Radford, A., Narasimhan, K., Salimans, T. ve Sutskever, I. *Improving language understanding by generative pretraining*, URL <https://s3-us-west-2.amazonaws.com/openaiassets/researchcovers/languageunsupervised/languageunderstandingpaper.pdf>, 2018.
- [39] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D. ve Sutskever, I. *Language models are unsupervised multitask learners*, OpenAI Blog, vol. 1, no. 8, 2019.
- [40] Dai, Z., Yang, Z., Yang, Y., Cohen, W. W., Carbonell, J., Le, Q. V. ve Salakhutdinov, R. *Transformer-xl: Attentive language models beyond a fixed-length context*, arXiv preprint arXiv:1901.02860, 2019.
- [41] Yang, Z., Dai, Z., Yang, Y., Carbonell, R., Salakhutdinov, R. ve Le, Q. V. *Xlnet: Generalized autoregressive pretraining for language understanding*, arXiv preprint arXiv:1906.08237, 2019.
- [42] Lample, G. ve Conneau, A. *Cross-lingual language model pretraining*, arXiv preprint arXiv:1901.07291, 2019.
- [43] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. ve Stoyanov, V. *Roberta: A robustly optimized bert pretraining approach*, arXiv preprint arXiv:1907.11692, 2019.
- [44] Sanh, V., Debut, L., Chaumond, J. ve Wolf, T. *Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter*, arXiv preprint arXiv:1910.01108, 2019.
- [45] Clark, K., Khandelwal, U., Levy, O. ve Manning, C. D. *What does bert look at? an analysis of bert’s attention”* arXiv preprint arXiv:1906.04341, 2019.