

Türkçe WordNet Oluşturulması ve Makine Çevirisinde Kullanılması

Program Kodu: 1001

Proje No: 116E104

Proje Yürütücüsü:
Prof. Dr. Olcay Taner Yıldız

Bursiyer(ler):

Özge BAKAY

Özlem ERGELEN

Ozan Mahir YILDIRIM

Razieh EHSANİ

Orçun ERKEK

Elif ESMER

Mehmet İleriş BOZKURT

Gökçe ULUDOĞAN

Rıza ÖZÇELİK

Selin YILDIRIM

ARALIK 2019

İSTANBUL

ÖNSÖZ

Türkçe WordNet oluşturulması ve Makine çevirisinde kullanılması başlıklı bu projede ana amaç daha önce BalkaNet'in bir parçası olarak 14000 sözcük öbeği ile küçük bir versiyonu oluşturulan Türkçe kelime ağının gerçekçi bir büyüklüğe ulaştırılmasıdır. Proje kapsamında 10 bursiyer ile çalışılmış, 1 dergi makalesi ile 10 konferans bildirisi üretilmiştir. Çok yüksek oranda emek yoğun bu çalışma sonucunda güncel Türkçe sözlüğün neredeyse tüm kelimelerini içeren 80.000 sözcük öbekli bir Türkçe kelime ağı oluşturulmuştur. Oluşturulan bu kelime ağı internet üzerinden erişime açılmıştır.

İÇİNDEKİLER

ÖNSÖZ	2
İÇİNDEKİLER	3
TABLO LİSTESİ	4
ÖZET	5
ABSTRACT	6
1. GİRİŞ	7
2. LİTERATÜR ÖZETİ	9
3. GEREÇ ve YÖNTEM	11
3.1. Sözcük Öbeklerinin Oluşturulması	11
3.1.1 Birleştirme İşlemi	13
3.1.2 Bölme İşlemi	14
3.2. Anlambilimsel İlişkiler	16
3.3. Diller Arası İlişkiler	20
4. BULGULAR	23
5. SONUÇ	24
KAYNAKÇA	26

TABLO LİSTESİ

- Tablo 3.1. Örnek Sözcük Öbekleri
- Tablo 3.2. Anlam-Tabanlı Kaymalar: Sıfat
- Tablo 3.3. Sözcük türü tabanlı Kaymalar: İsim ve Sıfatlar
- Tablo 3.4. Morfoloji-Tabanlı Kaymalar: Pasifler
- Tablo 3.5. Türkçe'den İngilizce'ye Bire-Çok Eşlemeler
- Tablo 3.6. İngilizce'den Türkçe'ye Bire-Çok Eşlemeler
- Tablo 3.7. KeNet'te Anlamsal İlişkiler
- Tablo 4.1. KeNet'teki Sözcük Öbeği Sayıları

ÖZET

Princeton kelime ağının üretilmesinden beri, diğer dillerde birçok kelime ağı üretilmiştir. Türkçe için en kapsamlı kelime ağı olan KeNet şu anda 80,000 synset içermektedir. Bu synsetler birbirleriyle anlamsal ve diller arası ilişkilerle bağlıdırlar. KeNet ayrıca Türkçe üzerinde Türkçe PropBank, alana özgü kelime ağları ve anlamsal işaretleme gibi birden çok Doğal Dil İşleme çalışmasında kullanılmıştır.

Türkçe kelime ağı KeNet iki platform ile kullanıcılara sunulmuştur. KeNet'in sunulduğu ilk platform, "<http://haydut.isikun.edu.tr/wordnet.ui-1.0/>" web sayfasıdır. Bu platform ile, kelime veya sözcük öbekleri araması yapılabilir. Arama sonuçlarına göre, bulunan kelimelerin hiyerarşileri ile anlamsal ilişkileri görülebilir. Ayrıca Türkçe kelime ağı KeNet'i github sitesi üzerinden de erişime açtık. Bu platform sayesinde, KeNet'in tüm dosyaları indirilebilir, Java, C++ ve Python gibi üç programlama dilinde KeNet kullanılabilir. KeNet'in github web sitesindeki web sayfaları Java, C++ ve Python programlama dilleri için sırasıyla "<https://github.com/olcaytaner/TurkishWordNet>", "<https://github.com/olcaytaner/TurkishWordNet-CPP>", "<https://github.com/olcaytaner/TurkishWordNet-Py>"dır.

Bu proje raporunda, KeNet'i meydana getirirken kullandığımız yöntemleri, bu yöntemlere ait basamaklara da referans vererek anlatacağız.

ABSTRACT

Since the creation of Princeton WordNet, several wordnets for other languages have been created. Being the most comprehensive wordnet for Turkish, KeNet currently includes 80,000 synsets, which are linked through their semantic as well as interlingual relations. KeNet has also been used in various NLP studies on Turkish, including Turkish PropBank, domain-specific wordnets and semantic annotations.

Turkish WordNet KeNet is accessible through two platforms. The first platform, where KeNet is served, is the web page, i.e. "<http://haydut.isikun.edu.tr/wordnet.ui-1.0/>". Through this platform, one can search words or lemma. According to the search results, one can view the hierarchy and semantic relations of the found words. We also serve Turkish WordNet KeNet through github site. Via this platform, one can download all the files of the KeNet, and may use KeNet in one of the three programming languages Java, C++, and Python. The web pages of KeNet in github site are "<https://github.com/olcaytaner/TurkishWordNet>", "<https://github.com/olcaytaner/TurkishWordNet-CPP>", "<https://github.com/olcaytaner/TurkishWordNet-Py>", for Java, C++, and Python languages respectively.

In this report, we present the procedure we have adopted in creating KeNet with reference to stages we have followed in its construction.

1. GİRİŞ

Sözcükler ve anlamları ile ilgili bilgiler geleneksel olarak sözlüklerde saklanmaktadır. Doğal dil işleme çalışmalarının ilerlemesiyle birlikte, makine tarafından okunabilir sözlüklere ihtiyaç duyulmaya başlanmıştır (MILLER, 1995). Bu ihtiyaca cevap vermek amacıyla, sözlükbilimsel bilgiyi modern bilişime uyarlanmış bir formatta depolayan Sözcük Ağları (wordnets) oluşturulmaya başlanmıştır. İlk sözcük ağı, İngilizce dili için 1995 yılında George Miller tarafından Princeton WordNet (PWN) adıyla oluşturulmuştur (MILLER, 1995).

Sözcük ağı, daha geniş tanımıyla, farklı sözcük anlamları ve onların tanımlarına dayanan son derece kapsamlı bir sözlüktür. Bir sözcük ağındaki sözcüklerin çoğu isim, fiil, sıfat ve zarf gibi içerik sözcükleridir. Bir sözcük ağının temel yapı taşları, anlamlar arasındaki eş anlamlılık ilişkisine dayanarak oluşturulan sözcük öbekleridir (synsets). Başka bir deyişle, sözcük ağlarındaki ana birimler sözcük öbekleridir ve dil içi ve diller arası olanlar da dahil olmak üzere tüm eşlemeler bu sözcük öbeklerine dayalı olarak oluşturulur.

Sözcük öbeği oluşturmak için esas olarak kullanılan eş anlamlılık ilişkisine ek olarak, sözcük öbekleri arasındaki diğer anlamsal ilişkileri de not etmek mümkündür. Sözcüksel anlambilimde, sözcüklerin kendi aralarındaki ilişkilere dayanarak tanımlanabileceği öne sürülmektedir. Bu ilkeyi benimseyen sözcük ağları, belirli bir dildeki anlamsal ilişkileri, sözcük öbeklerini kullanarak eşler. Sözcük ağlarında eşlenen en yaygın anlamsal ilişkiler eş anlamlılık, alt/üst kavramlılık, parça-bütün ilişkisi ve zıt anlamlılıktır.

PWN ilk olarak bu amaç için oluşturulmamış olsa da (MILLER, 1995), bilişimsel dilbilimdeki doğal dil işleme (NLP) çalışmaları için oldukça popüler bir sözlük kaynağı haline gelmiştir. Zengin sözcüksel kaynaklar olan sözcük ağları, günümüzde sözcük algılamasının ortadan kaldırılması, metin analizi, duyarlılık analizi ve bilgi erişimi gibi çeşitli görevler için kullanılabilir.

Son yirmi yılda birçok dil için pek çok sözcük ağı oluşturulmuştur ve Türkçe için iki farklı sözcük ağı bulunmaktadır: BalkaNet (TR-wordnet) (BİLGİN, 2004) ve KeNet (EHSANİ, 2018; BAKAY,

2019). Bu raporda, Trke iin daha kapsamlı bir szck ađı olan KeNet'i oluřturma sreci anlatılacaktır. Literatr zeti blmnde szck ađları hakkında arka plan bilgisi, Gere ve Yntem blmnde szck beklerinin oluřturulma sreci ve dil ii ve diller arası anlamsal iliřkilerin eřlenmesi, Bulgular blmnde alıřma sonucunda elde edilen istatistiki verilerin sunumu ve son olarak Sonu blmnde alıřmanın nemi verilmiřtir.

2. LİTERATÜR ÖZETİ

İster sıfırdan ister bir öncekini genişleterek sözcük ağı oluşturmak, birkaç adım gerektiren ve hem insan emeğini hem de otomatik sistemleri kullanan yoğun bir süreçtir. 1995'te ilk sözcük ağı olan PWN'nin oluşturulmasından bu yana, birçok dilde sözcük ağları oluşturulmuştur. PWN, çeşitli doğal dil işleme çalışmalarında kullanılarak sözcük ağlarının kullanılabilirliğini göstermiş ve diğer dillerde yapılan çalışmalara bir örnek teşkil ederek yeni sözcük ağları oluşturulmasının yolunu açmıştır. Farklı diller için sözcük ağları oluşturulurken kullanılan bir yol, özellikle İngilizce'ye benzer yapıdaki diller için PWN'nin örnek alınması ve PWN'nin çalışmanın yapıldığı dile çevrilmesiyle yeni sözcük ağlarının oluşturulması olmuştur. Bu şekilde oluşturulan sözcük ağlarında kullanılan yaklaşım literatürde "genişletme" olarak adlandırılmaktadır. Buna karşın, çalışmanın yapıldığı dil için gerekli sözcük öbeklerinin ve anlamsal ilişkilerin sıfırdan oluşturulmasıyla üretilen sözcük ağları çalışmalarına da rastlamak mümkündür. Bu şekilde yapılan çalışmalarda kullanılan yaklaşım ise "birleştirme" olarak bilinmektedir.

Genişletme yaklaşımıyla sözcük ağları oluşturulan dillere örnek olarak Fince, Lehçe ve Fransızca verilebilir. Örneğin, Fince Sözcük Ağı, PWN ile aynı sayıda sözcük öbeğine sahiptir (117.000 sözcük öbeği ve 209.941 sözcük-anlam çifti) (LINDEN vd., 2012). Aynı sayıdaki sözcük öbeklerinden de anlaşılacağı üzere, Fince WordNet, PWN'deki sözcük öbeklerinin Fince'ye çevrilmesiyle oluşturulmuştur. Benzer şekilde, bazı araştırmacılar, PWN'yi hem ilgili dile çevirip hem de üzerine eklemeler yaparak yeni sözcük ağları oluşturmuşlardır. Bu şekilde yapılan çalışmalara örnek olarak pWordNet verilebilir. pWordNet, İngilizce'den çevrilen sözcük öbeklerine ek olarak PWN'den yaklaşık 1000 adet daha fazla sözcük öbeğine sahip olan Lehçe sözcük ağıdır (PIASECKI vd., 2016). Fransızca sözcük ağı WOLF da PWN'ye dayanarak ve aynı zamanda çeşitli çok dilli kaynaklardan yararlanılarak oluşturulmuştur (SAGOT, 2008). Bu sözcük ağı bir derlem kullanılarak oluşturulmuş ve daha sonra bu derlem İngilizce, Romence, Bulgarca ve Çekçe gibi diğer dillerle ilişkilendirilmek üzere işaretlenmiştir. Sözcük öbekleri oluşturulduktan sonra, araştırmacılar PWN'deki tek anlamlı sözcükleri Fransızca'ya çevirme yolunu kullanmış ve bu çalışma sonucunda WOLF'da 32.351 sözcük öbeği oluşturulmuştur.

Birleştirme yaklaşımıyla oluşturulan sözcük ağlarına örnek olarak ise Arapça Sözcük Ağı (AWN) verilebilir. AWN, yukarıdan aşağıya prosedürüyle Arapça Temel Kavramlar baz alınarak

oluşturulmuş. Bu kavramlar büyük oranda EWN ve BalkaNet'teki Ortak Temel Kavramlara (CBC) dayanmaktadır. Temel kavramlara ek olarak, AWN'ye dile özgü kavramlar da eklenmiştir. Çeşitli dillerdeki sözcük ağlarının mevcut durumuna Global Sözcük Ağı Birliğinin (Global WordNet Association, GWA) internet sayfasından erişmek mümkündür.

Çeşitli yaklaşımlarla oluşturulmuş tek bir dile odaklanan sözcük ağlarına ek olarak, birden fazla dilin birbirleriyle ilişkilendirildiği çok dilli sözcük ağları projelerine de rastlamak mümkündür. Bu şekilde birden fazla dilin sözcüklerinin birbirleriyle ilişkilendirildiği sözcük ağlarına örnek olarak EuroWordNet ve BalkaNet verilebilir. EuroWordNet (EWN), birçok Avrupa dilinden (ör. Hollandaca, Almanca, Felemenkçe, İtalyanca ve İspanyolca) oluşan ve eş anlamlılık ilişkisinin sözcük öbeklerinde belirtildiği ve sözcük öbeklerinin temel anlambilimsel ilişkiler yoluyla işaretlendiği çok dilli bir sözcük ağıdır (VOSEN, 1997). EWN, PWN sözcük ağını temel alan diller-arası-dizin (Inter-Lingual-Index) sistemini kullanarak barındırdığı dilleri bu dizin aracılığıyla birbirine bağlamıştır. Bu dizin sayesinde EWN'de bir dildeki sözcükten başka bir dildeki benzer sözcüklere erişmek mümkün kılınmıştır. Sıkça işaretlenen anlamsal ilişkilere ek olarak, EWN'de aynı zamanda her anlam bir denklik ilişkisi ile ilişkilendirmiştir. EWN'nin mevcut kaynakları ve veritabanları çeşitli projelerde kullanılmıştır. BalkaNet, EWN'ye benzer şekilde, altı Balkan dilinden (Bulgarca, Çekçe, Yunanca, Rumence, Sırpça ve Türkçe) oluşan çok dilli bir wordnet projesidir (TUFIS vd., 2004). BalkaNet projesinde, Balkan dilleri için ayrı sözcük ağlarından oluşan çok dilli sözlük ağları geliştirilmiş ve bu sözcük ağları, daha sonra her bir Balkan dilinde sözcükler arasındaki anlamsal ilişkileri temsil etmek ve çok dilli bir anlambilim ağı geliştirmek amacıyla birbirine bağlanmıştır. Diller arası ilişkilendirme EuroWordNet'te kullanılan sisteme benzer bir şekilde tamamlanmıştır. BalkaNet, daha az çalışılmış Balkan dillerinin doğal dil işleme çalışmalarındaki yerini güçlendirmek amacıyla yapılmıştır. BalkaNet'te Türkçe için toplamda 14,626 sözcük öbeği ve 19,934 dil içi anlamsal ilişki bulunmaktadır.

3. GEREÇ ve YÖNTEM

3.1. Sözcük Öbeklerinin Oluşturulması

Diğer bütün sözcük ağlarında olduğu gibi, KeNet'i oluşturmak için atılması gereken ilk adım, sözcük öbeklerini oluşturmaktır. Bir sözcük öbeği, anlamı ve sözcük türü aynı olan bir grup sözcük olarak tanımlanabilir. Örnek bir sözcük öbeğinin yapısı aşağıdaki gibidir:

```
<SYNSET>  
  <ID>TUR10-0038510</ID>  
  <LITERAL>anne<SENSE>2</SENSE>  
  </LITERAL>  
  <POS>n</POS>  
  <DEF>...</DEF>  
  <EXAMPLE>...</EXAMPLE>  
</SYNSET>
```

KeNet'ten alınma örnek sözcük öbekleri Table 3.1'de gösterilmektedir. Bu tabloda, KeNet'te en sık görülen 4 sözcük türü için örnekler listelenmiştir; isim, fiil, sıfat ve zarf. Bu örneklerin her biri için, ilk sütun sözcük öbeğinin numarasını göstermektedir. Numaradan "-" ile ayrılmış karakterler sözcük öbeğinin türünü belirtmektedir (n: isim, v: fiil, a: sıfat adv: zarf). İkinci sütunda verilen sözcükler, aynı sözcük öbeğindeki eş anlamlı sözcükleri göstermektedir. Üçüncü sütun tanımları ve son olarak dördüncü sütun da eş anlamlı sözcüklerden birisini içeren örnek bir cümle göstermektedir.

Tablo 3.1. Örnek Sözcük Öbekleri

Numara	Sözcükler	Tanım	Örnek Cümle
TUR10-0000030-n	su ab âb	Hidrojenle oksijenden oluşan, oda sıcaklığında sıvı durumunda bulunan, renksiz, kokusuz, tatsız madde	
TUR10-0000220-a	abajurlu	Abajuru olan	Üstünde lacivert abajurlu, parlak bir madenden lamba
TUR10-0000350-v	abanmak	Eğilerek bir şeyin, bir kimsenin üzerine kapanmak	Efendi, sen de ne üstüme abanıyorsun?
TUR10-0000520-b	abartısız	Abartmadan, abartısız olarak, mübalağasız bir biçimde	

Bu sözcük öbeklerini oluşturma sırasında, veritabanının ilk sürümü Türk Dil Kurumu tarafından en son basılan Güncel Türkçe Sözlüğü (GTS) (2011 basımı) kullanılarak hazırlanmıştır (EHSANİ vd., 2018). Konvansiyonel olarak, GTS, sözcüklerin eş anlamlılarını sözcüğün tanımının ardından virgülle ayırarak işaretlemektedir:

affetmek: Herhangi bir kötü davranış için ceza vermekten vazgeçmek, **bağışlamak**.

Aynı sözcük öbeğinin içinde bulunması gereken gerçek eş anlamlı sözcüklere karar vermek için, ilk aşamada tanımlar virgüllerden bölünmüş ve virgül-ayrımli sözcükler eş-anlamlı adaylar olarak listelenmiştir. Daha sonra, bu listeler sözcükler ve tanımlarına göre insan işaretleyiciler tarafından eş anlamlılık ilişkilerine göre gruplandırılmıştır Bu aşama sonunda 49,774 sözcük ikilisi işaretlendi.

Her ne kadar bazı fiillerin edilgen ve ettirgen çatıları GTS'de ayrı madde olarak verilse de, bu fiiller aynı eylem köküne sahip oldukları için KeNet'ten silindi.

Sözcük öbeklerinin büyük bir çoğunluğu bu aşama sonucunda oluşturulmuş olsa da, bu sözcük öbeklerinin organizasyonunu iyileştirmek için bazı ardışık işlemlere de gerek duyuldu. Sözcük öbeklerinin oluşturulmasında karşılaşılan en önemli sorun, sözcüklerin anlambilimsel ilişkileri olduğundan, sözcük öbekleri içindeki eş-anlamlılık ilişkisini kontrol etmek için diğer iki işlem daha gerçekleştirildi. Bu sebeple, sıradaki iki alt bölüm bu amaca ulaşmak için gerçekleştirilen bölme ve birleştirme işlemlerinden bahsedecektir.

3.1.1 Birleştirme İşlemi

Birleştirilmiş sözcük öbekleri, sözcükler birleştirilirken anlamlar da birleştirildiğinden, birleştirilen sözcük sayısı kadar farklı anlama sahiptir. Birleştirilmiş sözcük öbeklerinin tanımları başlangıçta aralarına "]" işareti konarak gösterilmiştir. Bu birleştirme işleminin esas amacı verilen bir sözcük öbeğini oluşturan sözcüklerin ortak bir tanımını oluşturmaktır. Bu tanımlar, sözcük öbeğindeki hiç bir sözcük oluşturulan tanımın içinde geçmeyecek şekilde tekrar yazıldı. Birleştirme işlemi sonucunda, 10.612 sözcük öbeğinin tanımı insan işaretleyiciler tarafından yeniden yazıldı.

Birleştirilmiş sözcük öbeklerinin büyük bir kısmı içlerindeki sözcüklerin eş-anlamlılık ilişkisi açısından doğru olsa da, bazılarının yanlış olduğu fark edildi ve bu sebeple, bu işlemin ikinci amacı olarak gerçek bir sözcük öbeği oluşturmadıkları için bölme ihtiyacı gerektiren sözcük öbeklerinin bir kısmı kenara ayrıldı. Bu biçimde olan 1.144 sözcük öbeği bölme işlemine transfer edildi. (bkz. Bölüm 3.1.2.).

Yanlış oluşturulan sözcük öbeklerinden başka, bazı sözcük öbeklerinde gereken bazı sözcüklerin eksik olduğu durumlar da görüldü. Eksik olan sözcüklerin bazıları KeNet'te olsa da, bazıları da KeNet'te yoktu. KeNet'te olan sözcükler sözcük öbeklerine eklenirken, olmayanlar farklı kelimeler olarak oluşturuldu ve sözcük öbeklerine eklendi. Bu eklemenin sonucunda, 122 sözcük öbeği yeni kelimelerle zenginleştirildi.

Son olarak, sözcük öbeklerindeki kelimelerin sırası da bu aşamada kontrol edildi. Zaman kısıtından dolayı, sadece ilk sözcük kontrol edildi. Aynı sözcük öbeğinin içindeki en sık kullanılan sözcük, sözcük öbeğinin görünümü daha güzel olması için sözcük öbeğinin ilk yerine yerleştirildi. Kalan sözcüklerin sırası sonra işlenmek üzere ileri bir tarihe bırakıldı.

3.1.2. Bölme İşlemi

KeNet'in anlambilimsel veya diller arası ilişkileri üzerinde çalışırken, bazı sözcük öbeklerinin eş anlamlı kelimeler içermediğini veya yanlış birleştirmelerden dolayı bazı sözcüklerin o sözcük öbeğine ait olmadıklarını gördük (BAKAY, 2019b). Diğer bir deyişle, bazı sözcük öbeklerinin organizasyonu onları oluşturan sözcüklerin eş anlamlılıkları açısından yanıltıcı olabiliyordu. Bu problemi çözmek için, diğer çalışmalarda yanlış görülen sözcük öbeklerini toplayarak bir havuz oluşturduk. Bu havuzdaki sözcük öbeklerini Google sheet'te anlamsal açıdan farklılıklarına göre ayrıştırdık. İnsan işaretleyiciler daha sonra bu yanlış birleştirilmiş sözcük öbeklerini böldüler ve yeni oluşturulan sözcük öbekleri için yeni tanımlar yazdılar.

Yanlış birleştirmelerin üç ana sebebi bulunmaktaydı: anlam-tabanlı kaymalar, sözcük türü tabanlı kaymalar and morfoloji tabanlı problemler (BAKAY, 2019b). Tablo 3.2, 3.3 ve 3.4 bu üç durumu sırasıyla örneklendirmektedir. Anlam-tabanlı kayma için, anlamsal olarak ilişkili olan sözcükler yanlışlıkla birleştirilmiştir. Örneğin, Tablo 3.2'de, sıfatlar arasında anlambilimsel olarak bir süreklilik bulunmaktadır. Bariz anlamsal farklılıktan dolayı bu sıfatlar birbirlerinin yerine yazılamamakta, sonuç olarak yanlış birleştirilmiş sözcük öbeği yeni sözcük öbeklerine bölünmektedir. Sözcük türü tabanlı kaymalar açısından, anlamsal olarak ilişkili fakat farklı sözcük türüne sahip olan sözcükler yanlışlıkla birleştirilmiştir. Fakat, farklı sözcük türlerine ait sözcükler birbirlerinin yerine kullanılmadığından, farklı sözcük türlerini içeren bu sözcük öbekleri Tablo 3.3'te gösterildiği gibi bölünmüştür. Son olarak, morfolojik olarak ilişkili olan sözcükler yanlışlıkla aynı sözcük öbeklerine yerleştirilmiştir. Dolayısıyla, morfolojik olarak ilişkili olan fiiller, örneğin aynı fiilin etken ve edilgen çatıları farklı sentaktik özelliklere sahip oldukları için eşit olarak görülemez. Bu sebeple, farklı argüman yapılarına sahip olan fiiller yeni sözcük öbeklerine bölünmüştür.

Tablo 3.2. Anlam-Tabanlı Kaymalar: Sıfat

Birleşik Sözcük Öbeği No	Literaller	Yeni Sözcük Öbeği No	Tanım
TUR10-0054930	ateşin ateşli coşkulu hararetili heyecanlı coşkunca vectli koyu nefes kesici deli huruşan	TUR10-0054930	Coşkusu olan Heyecan veren Azgın
	azgın celalli coşkun	TUR10-0065120	Coşmuş, taşmış
	derin	TUR10-0196230	Ayrıntıya önem verilerek hazırlanan
	lirik	TUR10-0512300	Coşkun, ilhamla dolu

Tablo 3.3. Sözcük türü tabanlı Kaymalar: İsim ve Sıfatlar

Birleşik Sözcük Öbeği No	Literaller	Yeni Sözcük Öbeği No	Tanım
TUR10-0312230	bilek çelim erk	TUR10-0102350	Bir işi yapabilme gücü, kudret, iktidar
	can dirilik	TUR10-0131370	Güç
	fer	TUR10-0268090	Güç, kuvvet, nüfuz
	güç	TUR10-0312170	Ağır ve yorucu emekle yapılan, müşkül
	güç	TUR10-0312230	Büyük etkinliği ve önemi olan nitelik
	erk kuvvet nüfuz	TUR10-0250960	Sözü geçerlik, istediğini yaptırabilme gücü

Tablo 3.4. Morfoloji-Tabanlı Kaymalar: Pasifler

Birleşik Sözcük Öbeği No	Literaller	Yeni Sözcük Öbeği No	Tanım
TUR10-1156950	dayak atmak dövme kötek atmak kötek çekmek marizlemek sıkıdan geçirmek tekme tokat girişmek tepelemek	TUR10-0889290	Tokat, yumruk, tekme vurarak canını acıtmak
	sopalamak	TUR10-0703060	Sopa ile vurmak, dövme
	sopalanmak	TUR10-0703080	Sopa ile vurulmak, sövülmek
	sövülmek	TUR10-0706470	Sövme işine konu olmak
	tepelemek	TUR10-0765510	Bozguna uğratmak, hırpalamak

3.2. Anlambilimsel İlişkiler

Anlamsal ilişki işaretlemelerinin ilk aşamasında odaklanılan iki ana ilişki şunlardı: zıt anlamlılık ve alt/üst kavramlılık (EHSANİ vd., 2018). TDK ve Türkçe Vikipedi'nin kurallara dayalı işlenmesiyle bir dizi aday çifti belirlendi ve bu aday çiftler daha sonra insan işaretleyiciler tarafından kontrol edildi. Zıt anlam adayları TDK'da karşıt kalıplar aranarak çıkarıldı. Zıt anlamlılık için kullanılan kalıplardan en yaygın olanı en yaygın olanı "a'nın karşıtı" idi. Alt ve üst kavramlar için alt kavram aracılığıyla üst kavramlara ulaşma yolu tercih edildi. Zıt anlamlılarda olduğu gibi, üst kavramları bulmak için "a'dan biri" şeklinde yapılan aramalarla a'ya, yani üst kavrama, ulaşıldı. Bu aşamanın sonunda toplam 4560 anlambilimsel ilişki elde edildi.

Anlambilimsel ilişkilerin sayısını ve kapsamını artırmak amacıyla bazı destekleyici prosedürler uygulandı. En önemli anlambilim ilişkisi üst kavramlılık olduğundan, zamanın ve kaynakların çoğunu bu ilişkiyi tamamlamak üzere kullanıldı, KeNet'in şu anki halinde sadece isim kategorisi

için bir anlamsal hiyerarşi ağacı var. Tüm düğümlerin üst kavramlılığa dayalı olarak birbirine bağlandığı kapsayıcı bir anlamsal hiyerarşik ağaçta, ağacın en üst kısımlarında yer alan üst kavramlar en önemli olanlardır. Bu nedenle, anlamsal ağaca ilk olarak en üstte yer alan üst kavramlar belirlenerek başlandı. Ağacın en üst kısmında yer alan bu kavramları bulmak için, İngilizce Emlak Sözcük Ağı'nda en çok alt kavramın bağlandığı, dolayısıyla en çok tekrar eden üst kavramlar Google Sayfalar üzerinde listelendi. İşaretleyiciler daha sonra bu üst kavramların Türkçe karşılıklarını KeNet'teki tanımlardan bularak hiyerarşik ağacın en üst kısmını oluşturdular. Bu aşamadan sonra, tüm hiyerarşik ağacın tamamlanması için dışarıda kalmış sözcük öbekleri işaretleyiciler tarafında bağlanmaları gereken üst kavramların altına taşındı. Anlamsal ağacın şu anki durumunda 23,075 sözcük öbeği üst kavramları ile eşleştirilmiş durumdayken, geriye kalan 20,233 sözcük öbeği bağımsız bir şekilde ağacın dışında yer almaktadır.

Tablo 3.5. Türkçe'den İngilizce'ye Bire-Çok Eşlemeler

Türkçe	İngilizce
---------------	------------------

Sözcük Öbeği No	Literaller	Tanım	Sözcük Öbeği No	Literaller	Tanım
TUR10-0322 140	haksız nahak	Hak ve adalete uygun olmayan	ENG31-0195 1967-s	undue unjustified unwarranted	lacking justification or authorization
			ENG31-0096 1901-a	inequitable unjust	not equitable or fair
			ENG31-0096 0366-a	unfair unjust	not fair
			ENG31-0204 4679-a	unrighteous	not righteous
			ENG31-0141 1393-s	unlawful wrongful	having no legally established claim

Tablo 3.6. İngilizce'den Türkçe'ye Bire-Çok Eşlemeler

Türkçe	İngilizce
---------------	------------------

Sözcük Öbeği No	Literaller	Tanım	Sözcük Öbeği No	Literaller	Tanım
TUR10-0320 470	hafiflemek	Herhangi bir sebeple eski ağırlığı azalmak	ENG31-0024 5945-v	abate let up slack off slack die away	become less in amount or intensity
TUR10-0320 480	hafiflemek azalmak	Etkisi, gücü azalmak			
TUR10-0451 010	kırılmak yatışmak	Soğuk, rüzgâr vb. eski gücü kalmamak, azalmak			

Son olarak, KeNet'e Tablo 3.7.'de örnekleri ve sayıları verilen 5 farklı anlamsal ilişki daha eklendi. Tüm bu ilişkiler için işaretlenen ana sözcük türü isimlerken zıtlık ve özellik ilişkileri esas olarak sıfatlar için işaretlenmiştir.

Tablo 3.7. KeNet'te Anlamsal İlişkiler

Anlamsal İlişki	Türkçe Örnek	Sayı
-----------------	--------------	------

Alt/Üst Kavram	gürgen - ağaç	21,198
Alan Bilgisi	işlemci - bilgisayar bilimi	15,398
Parça-Bütün İlişkisi	kulp - bardak	2,761
Zıt Anlamlılık	sıcak - soğuk	1,908
Örnek-Üst Kavram	Antartika - kıta	1,421
Üye-Bütün	ebeveyn - aile	873
Madde-Bütün	hidrojen - su	367
Özellik	ılık - sıcaklık	226

3.3. Diller Arası İlişkiler

İngilizce için PWN'nin oluşturulmasından sonra, diğer diller için sözcük ağları oluşturmak için çeşitli çalışmalar yapılmıştır (PEIK, 1997; BLACK, 2006). Birkaç dilde sözcük ağlarının oluşturulmasıyla, bu sözcük ağlarını İngilizce ile eşleştirme fikri önem kazanmıştır çünkü sözcük ağlarının diller arasında bağlanması bu kaynakların makine çevirisinde kullanılmasının önünü açmıştır.

Çok dilli kelime ağlarının oluşturulmasında birkaç farklı yöntem kullanılmaktadır. Literatürde en sık kullanılan yaklaşımlardan ikisi, Literatür Özeti bölümünde de bahsedildiği üzere, *birleştirme* ve *genişletme* yöntemleridir. Birleştirme yöntemine göre oluşturulan çok dilli sözcük ağlarında, ilk olarak üzerinde çalışılan dil için İngilizce ile herhangi bir bağlantı olmaksızın sadece dil içi ilişkilere odaklanarak bir sözcük ağı oluşturulur. Bu tek dilli sözcük ağı daha sonra İngilizce sözcük ağı ile ilişkilendirilir. Genişletme yaklaşımda ise, anlamsal ilişkileri de dahil olmak üzere PWN'den bir dizi sözcük öbeği önce hedef dile çevrilir, daha sonra İngilizceden aktarılan bu anlamsal ilişkiler manuel olarak kontrol edilir. Bu iki yaklaşım karşılaştırıldığında, ikincisinin daha pratik ve daha az zaman alıcı olduğu iddia edilmektedir çünkü bu yaklaşımda PWN'den otomatik olarak çekilen anlamsal ilişkiler sayesinde birçok tek dil içi ilişki hedef dile doğru bir şekilde

aktarılmış olur. PWN'den ilişkilerin bu şekilde otomatik olarak çekilmesi, özellikle Fransızca gibi İngilizce'ye benzer diller için oldukça kullanışlı bir yöntemdir (SAGOT, 2008).

Giriş bölümünde tartışıldığı gibi, Türkçe için ihali hazırda iki sözcük ağı mevcuttur; BalkaNet'in Türkçe kısmı ve bu proje çerçevesinde oluşturulan KeNet. BalkaNet'in bir parçası olarak oluşturulan Türkçe sözcük ağı (TUFIS vd., 2004), EWN'de olduğu gibi, diller-arası- indeks (ILI) aracılığıyla PWN'ye entegre edilmiştir (BİLGİN vd., 2004). ILI kullanımı ile, BalkaNet'in Türkçe sözcük ağı ve PWN arasında sadece birebir eşlemeler dahil edilmiştir. BalkaNet'in aksine, KeNet ve PWN arasında diller arası ilişkiler işaretlenirken *genişletme* yaklaşımı benimsenmiştir. Ayrıca, eşleştirme işlemi için sözlük çevirilerine dayanarak, birden çoğa eşlenebilecek sözcük öbeği adayları da çıkarılmış ve bu ilişkiler Google E-Tablolar'da listelenmiştir. Dilbilimci işaretleyiciler daha sonra bu birden çoğa eşleme adaylarını tek tek kontrol etmiş ve doğru olan eşlemeleri kaydetmiştir.

KeNet ve PWN arasındaki birçok interlingual ilişki bire bir eşlemeler şeklinde işaretlenmiştir. Fakat KeNet'teki diller arası eşlemelerle ilgili olarak dikkat edilmesi gereken önemli bir nokta, KeNet'te, BalkaNet'teki gibi bire bir eşleşmelere ek olarak, aynı zamanda, birden çoğa eşlemelere de yer verilmiş olmasıdır. Bu şekilde birden çoğa eşlemelere yer verilmiş olmasının sebebi ise sözcük öbeklerini diller arasında eşlerken her zaman bir eşlemelerin mümkün olmayışıdır. Başka bir deyişle, BalkaNet'te bir sözcük öbeği PWN'de her zaman bir sözcük öbeği ile eşleştirilirken, KeNet'te gerektiğinde birden çoğa eşlemelere de yer verilmiştir (BAKAY vd., 2019a).

Birden çoğa eşleşmenin gerekli olduğu en yaygın iki durum (i) İngilizce'deki bir anlam ayrımının Türkçede bulunmadığı ya da (ii) tam tersi olarak, Türkçede mevcut olan bir anlam ayrımının İngilizcede var olmadığı durumlardır. Örnek vermek gerekirse, Tablo 3.5.'te bir Türkçe sözcük öbeği birden fazla İngilizce sözcük öbeğiyle eşleştirilmiştir. Bu birden çoğa eşleşmenin nedeni, İngilizcede 5 ayrı sözcük öbeği ile verilen anlamların Türkçede tek bir sözcük öbeğinde karşılık buluyor olmasıdır. Benzer şekilde, Tablo 3.6.'da tek bir İngilizce sözcük öbeği, Türkçede 3 farklı sözcük öbeği ile eşleştirilmiştir. Bu tür durumlarda, birden çoğa ihtiyaç duyulmasının sebebi, Türkçedeki anlam ayrımının İngilizcede var olmayışıdır.

El ile işaretlenen bu diller arası eşlemeler sonucunda, toplamda 19,398 Türkçe sözcük öbeği 19,208 İngilizce sözcük öbeği ile ilişkilendirilmiştir. Diller arası eşlemelerin 3500'ü bire bir eşlemeyken, 1250'si birden çoğa eşlemelerdir. Ayrıca, İngilizcede en sık kullanılan 5000 anlamdan 4417 tanesi (%88), en sık kullanılan anlamları diller arası eşlemelerde kapsamak amacıyla, Türkçe karşılıklarıyla eşlenmiştir.

4.BULGULAR

Şu anda KeNet'te 76,436 sözcük öbeği, 108,411 literal ve 78,611 farklı literal bulunmaktadır. İçerilen sözcük türleri isim, fiil, sıfat, zarf, ünlem, zamir, edat ve bağlaçtır (sayılar için bkz. Tablo 4.1.). Literallerin sözcük sayılarına ilişkin, literallerin büyük bir çoğunluğu bir (72,073 - %66,48) veya iki sözcükten (31,831 - %29,36) oluşmasına rağmen rağmen, toplamda yedi sözcüğü bulan literallere rastlamak mümkündür. Son olarak, 18,842 sözcük öbeği için örnek cümleye yer verilmiştir (%24,65).

Tablo 4.1. KeNet'teki Sözcük Öbeği Sayıları

Sözcük Kategorisi	Sözcük Öbeği Sayısı
İsim	43,307
Fiil	17,853
Sıfat	12,262
Zarf	2,512
Ünlem	324
Zamir	68
Edat	27
Bağlaç	58
Boş	25
Toplam	76,436

5. SONUÇ

Bu raporda, Türkçe kelime ağı KeNet'i oluşturma ve genişletme süreci anlatılmıştır. Rapor, KeNet'in mevcut durumu hakkında bilgi sağlamanın yanı sıra, KeNet'in Türkçe için yapılan diğer doğal dil işleme çalışmalarıyla nasıl bağlantılandırıldığına da değinmektedir. Giriş bölümünde bahsedildiği gibi sözcük ağları çeşitli doğal dil işleme çalışmalarında kullanılabilir. Türkçe söz konusu olduğunda, KeNet şu ana kadar birçok dil kaynağını geliştirmek için kullanılmıştır. İlk olarak, KeNet'ten gelen anlamlarla, toplamda 9,500 cümle içeren tüm sözcükleri işaretlenmiş bir Türkçe veritabanı işaretlenmiştir (ARICAN vd., 2019), (AKÇAKAYA vd., 2018). İnce anlam ayrımlarına dayanarak oluşturulan sözcük öbeklerinin düzgün yapısı sayesinde zengin bir anlambilimsel Türkçe kaynağı yaratılmıştır. İkinci olarak, bu proje kapsamında sözcük ağlarının kullanımı alana özgü çalışmalara genişletilmiştir. Daha belirgin bir biçimde, KeNet'e kullanıcı yorumları ve tercihleri eklenerek emlak alanı için bir sözcük ağı oluşturulmuştur (PARLAR vd., 2019). Emlak-alan sözcük ağının anlam ayrıştırma problemini iyileştirdiği ve bu alanda yapılan çalışmalarda biçimbilimsel analizleri geliştirdiği gösterilmiştir. Üçüncüsü, KeNet, Türkçe PropBank'taki argüman yapısı ilişkilerinin işaretlenmesinde kullanılmaktadır (AK vd., 2019). Kısaca bahsetmek gerekirse, PropBank bir dilin argüman-yüklem ilişkilerini barındıran bir kaynaktır ve makine çevirisi, bilgi çıkarma, soru cevaplama ve bilgi alma gibi sayısız doğal dil işleme uygulamalarında yaygın olarak kullanılmaktadır. Dördüncüsü, KeNet, pozitif, negatif ve objektif olmak üzere üç polarite noktasıyla oluşturulan bir polarite sözlüğü ile Türkçenin duygusal analizleri için kullanılmaktadır (ARICAN vd., 2019). Beşinci olarak, KeNet için sözcük öbeklerini bulmak için literalleri aramamızı sağlayan bir kullanıcı arayüzü oluşturduk (ÖZÇELİK vd., 2019). Bu kullanıcı arayüzünde literaller; numaraları, sözcük türleri, eş anlamlı sözcükleri ve semantik olarak ilişkili oldukları sözcük öbekleri ile görüntülenmektedir. geliştirilen bu kullanıcı arayüzü KeNet'in web sayfasında kullanıma açıktır (<http://haydut.isikun.edu.tr/wordnet.ui-1.0/>). Son olarak, KeNet'teki anlamsal işaretlemeler ağaç-odaklı bir makine çevirisi geliştirmek için kullanıldı (BAKAY vd., 2019c). Bu çalışmada sözcük- ya da kalıp-odaklı çeviri sistemi yerine ağaç-odaklı bir çeviri sisteminin tercih edilmesinin sebebi Türkçenin biçimbilimsel özellikleridir. Türkçede ağaç-odaklı bir çeviri sistemiyle sözdizimsel bağlantılar daha iyi temsil edildiğinden bu çalışmayla 12.8 olan BLEU skorları 21.4'e yükselmiştir.

Bu projede doğal dil işleme alanında az çalışılmış bir dil olan Türkçe için kapsamlı bir sözcük ağı oluşturulduğundan, sözcük ağının oluşturulma sürecinde Türkçe diline has özelliklerden kaynaklanan zorluklara değinmek gerekir. Bahsedilmesi gereken en büyük zorluk, yapısal olarak farklı bir dil üzerinde çalışmaktı çünkü alandaki çalışmaların çoğu İngilizce gibi sıkça çalışılan dillere odaklanmaktadır. Türkçenin yapısının bu dillerden farklı olması sebebiyle, bu diller üzerine yapılan çalışmalardaki yaklaşımların çoğu mevcut çalışmaya aktarılamamıştır. Yapısal farklılıklar çoğunlukla biçimbilimsel analizler için sorun yaratmış olsa da, İngilizcedeki anlambilimsel eşlemelerin ve/veya ilişkilerin doğrudan Türkçe'ye kopyalanamadığı durumlarla da karşılaşmıştır. Bu tutarsızlık hem dil içi hem de diller arası anlambilim ilişkilerinde gözlenmiştir. Genel olarak, Türkçeden yapısal olarak farklı olan diller üzerine yapılan çalışmaları birebir örnek olarak almak yanıltıcı olacağından ve Türkçe'ye kolayca uyarlanabilecek kaynakların bulunmadığından, bu çalışmada otomatik işaretlemelerin yanı sıra büyük miktarda el ile işaretlemeye de yer verilmiştir. Bu da, bu şekilde yapısal olarak farklı olan diller üzerine yapılan öncü çalışmalarda el ile işaretlemelerin yapılan çalışmanın doğruluğu ve güvenilirliği için ne denli önemli olduğunu ortaya koymaktadır.

KAYNAKÇA

1. AK, K., Yıldız, O. T., Esgel, V., and Toprak, C. (2019). Construction of a Turkish proposition bank. *Turkish Journal of Electrical Engineering and Computer Science*, 26:570 – 581.
2. AKÇAKAYA, S. and Yildiz, O. T. (2018). An all-words sense annotated Turkish corpus. In *Proceedings of the International Conference on Natural Language and Speech Processing*.
3. ARICAN, B. N., Bakay, O., Avar, B., Yildiz, O. T., and Ergelen, O. (2019). English-turkish parallel semantic annotation of penn-treebank. In *Proceedings of the Global WordNet Conference*.
4. BAKAY, O., Ergelen, O., and Yildiz, O. T. (2019a). Integrating Turkish wordnet kenet to princeton wordnet: The case of one-to-many correspondances. In *ASYU*.
5. BAKAY, O., Ergelen, O., and Yildiz, O. T. (2019b). Problems caused by semantic drift in wordnet synset construction. In *UBMK*.
6. BAKAY, O., Avar, B., and Yildiz, O. T. (2019c). A tree-based approach for English-to-Turkish translation. *Turkish Journal of Electrical Engineering & Computer Sciences*, 27:1, 427-452.
7. BİLGİN, O., Cetinoglu, O., and Oflazer, K. (2004). Building a wordnet for Turkish. *Romanian Journal of Information Science*, 7:163–172.
8. BLACK, W., Elkateb, S., Rodriguez, H., Alkhalifa, M., Vossen, P., Pease, A., and Fellbaum, C. (2006). Introducing the Arabic wordnet project. In *International Wordnet Conference*, pages 295–300. Masaryck University, Brno, Czeck Republic.

9. EHSANİ, R., Solak, E., and Yildiz, O. (2018). Constructing a wordnet for Turkish using manual and automatic annotation. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 17(3).
10. MILLER, G. (1995). Wordnet: a lexical database for english. *ACM Communications*, 38:39–41.
11. ÖZÇELİK, R., Parlar, S., Bakay, O., Ergelen, O., and Yildiz, O. T. (2019). User interface for Turkish word network KeNet. In *Signal Processing and Communication Applications Conference*.
12. PARLAR, S., Arican, B. N., Erkek, M., Cayirli, K., and Yildiz, O. T. (2019). Emlak alanına Özgü kelime ağı. In *Proceedings of the Signal Processing and Communication Applications Conference*.
13. PEIK, V. (1997). Eurowordnet: a multilingual database for information retrieval. In *DELOS workshop on Cross- language Information Retrieval*. Vrije Universiteit, Amsterdam, Czech Republic.
14. SAGOT, B. (2008). Building a free French wordnet from multilingual resources. page 24. ACM.
15. TUFIS, D., Cristea, D., and Stamou, S. (2004). Balka-net: Aims, methods, results and perspectives. a general overview. *Romanian Journal of Information Science*, 7:9–43.
16. VOSSEN, P. (2007). Eurowordnet: A multilingual database for information retrieval. In *DELOS workshop on Cross- Language Information Retrieval*.

TÜBİTAK
PROJE ÖZET BİLGİ FORMU

Proje Yürütücüsü:	Prof. Dr. OLCAY TANER YILDIZ
Proje No:	116E104
Proje Başlığı:	Türkçe Wordnet Oluşturulması Ve Makine Çevirisinde Kullanılması
Proje Türü:	1001 - Araştırma
Proje Süresi:	24
Araştırmacılar:	
Danışmanlar:	
Projenin Yürütüldüğü Kuruluş ve Adresi:	IŞIK Ü.
Projenin Başlangıç ve Bitiş Tarihleri:	01/05/2017 - 01/11/2019
Onaylanan Bütçe:	215500.0
Harcanan Bütçe:	196876.31
Öz:	<p>Princeton kelime ağının üretilmesinden beri, diğer dillerde birçok kelime ağı üretilmiştir. Türkçe için en kapsamlı kelime ağı olan KeNet şu anda 80,000 synset içermektedir. Bu synsetler birbirleriyle anlamsal ve diller arası ilişkilerle bağlıdırlar. KeNet ayrıca Türkçe üzerinde Türkçe PropBank, alana özgü kelime ağları ve anlamsal işaretleme gibi birden çok Doğal Dil İşleme çalışmasında kullanılmıştır.</p> <p>Türkçe kelime ağı KeNet iki platform ile kullanıcılara sunulmuştur. KeNet'in sunulduğu ilk platform, http://haydut.isikun.edu.tr/wordnet.ui-1.0/ web sayfasıdır. Bu platform ile, kelime veya sözcük öbekleri araması yapılabilir. Arama sonuçlarına göre, bulunan kelimelerin hiyerarşileri ile anlamsal ilişkileri görülebilir. Ayrıca Türkçe kelime ağı KeNet'i github sitesi üzerinden de erişime açtık. Bu platform sayesinde, KeNet'in tüm dosyaları indirilebilir, Java, C++ ve Python gibi üç programlama dilinde KeNet kullanılabilir. KeNet'in github web sitesindeki web sayfaları Java, C++ ve Python programlama dilleri için sırasıyla https://github.com/olcaytaner/TurkishWordNet?, https://github.com/olcaytaner/TurkishWordNet-CPP?, https://github.com/olcaytaner/TurkishWordNet-Py??dir.</p> <p>Bu proje raporunda, KeNet'i meydana getirirken kullandığımız yöntemleri, bu yöntemlere ait basamaklara da referans vererek anlatacağız.</p>
Anahtar Kelimeler:	Kelime Ağı, Makine Çevirisi
Fikri Ürün Bildirim Formu Sunuldu Mu?:	Evet
Projeden Yapılan Yayınlar:	1- A tree-based approach for English-to-Turkish translation (Makale - Diğer Hakemli Makale), 2- A Multilayer Annotated Corpus for Turkish (Bildiri - Uluslararası Bildiri - Sözlü Sunum), 3- An All-Words Sense Annotated Turkish Corpus (Bildiri - Uluslararası Bildiri - Sözlü Sunum), 4- Comparison of Turkish Proposition Banks by Frame Matching (Bildiri - Uluslararası Bildiri - Sözlü Sunum), 5- AnlamVer: Semantic Model Evaluation Dataset for Turkish - Word Similarity and Relatedness (Bildiri - Uluslararası Bildiri - Sözlü Sunum),