

Kavramlar Arası Anlamsal İlişkilerin Türkçe Sözlük Tanımları Kullanılarak Otomatik Olarak Çıkarılması

Automatic Extraction of Semantic Relationships Using Turkish Dictionary Definitions

Emre Yazıcı¹, M. Fatih Amasyalı¹

¹Elektrik Elektronik Fakültesi, Bilgisayar Mühendisliği Bölümü

Yıldız Teknik Üniversitesi

emre@yazici.info, mfatih@ce.yildiz.edu.tr

Özet

Bilgisayarların doğal dil işleme alanındaki rolleri son yıllarda oldukça önem kazanmıştır. Otomatik özet çıkarımı, makine çevirisi, doğal dilde arama ve soru cevaplama gibi çalışma alanları; otomatik metin analizi ve yapay zeka uygulamalarını destekleyici anlamsal ilişki veritabanları oluşturma gerekliliğini doğurmuştur. Bu tür bir veritabanının oluşturulması zor ve zaman alıcı bir süreç olduğundan, böyle bir veritabanını elle oluşturmak arzu edilen bir iş olmayabilir. Bu belgede, Türkçe için iki farklı sözlük kullanılarak kavramlar arasındaki anlamsal ilişkileri bulmaya yönelik farklı otomatik yöntemler sunulmaktadır.

Anahtar kelimeler: Doğal Dil İşleme, Örüntü Eşleme, Anlamsal Ağ, Bilgi Çıkarımı

Abstract

In the recent years, the role of computers in the field of natural language processing has increased so much. Work areas such as automatic summarization, machine translation, natural language search and question answering have brought the need for semantic relation databases to support automatic text analysis and artificial intelligence applications. Since the development process is hard and time consuming, developing such databases manually may not be desirable. In this paper, we present different automatic methods to extract semantic relationships between concepts using two Turkish dictionaries.

Keywords: Natural Language Processing, Pattern Matching, Semantic Web, Information Extraction

1. Giriş

Bu çalışmanın amacı, sözcüklerin sözlükteki tanımlarından yola çıkarak bilgisayarlı sistemlerin kullanımına uygun anlamsal bilgi çıkarımının sağlanmasıdır. Sözlükler, barındırdıkları her bir kavram hakkında nitelikli anlamsal bilgi içermesine rağmen bu bilginin organizasyonu bilgisayar uygulamaları tarafından kullanıma elverişli değildir [1]. Bir dilde bulunan sözcükler arasındaki anlamsal ilişkilerin çıkarılmasıyla oluşturulan anlamsal bir veritabanı, bilgisayar uygulamalarının gündelik yaşamı ilgilendiren konularda daha etkin rol oynamalarını sağlayacaktır. Bundan başka, bilgisayar

destekli çeviri sistemlerinin eş sesli sözcüklerin varlığından kaynaklanan hatalı çevirilerinin en aza indirgenerek bu sistemlerin başarımlarının artırılması gibi konularda da kavramlar arasındaki anlamsal ilişkilerden faydalanılabilir. Bu amaca yönelik bir veritabanının çıkarılmasında üzerinde çalışılan dile ait sözlükten yola çıkmak, sözlüklerin dile ait sözcük hazinesinin önemli bir kısmını barındırması nedeniyle iyi bir başlangıç olacaktır.

Kapsamı Türkçe ile sınırlandırılan bu çalışmada Türk Dil Kurumu (TDK) tarafından hazırlanan Büyük Türkçe Sözlük [2] ile bir Wikimedia projesi olan Vikisözlük'ten [3] faydalanılmıştır. Vikisözlük'te yer alan Türkiye Türkçesi dışındaki maddeler değerlendirilmeye alınmamış olup Vikisözlük'teki Büyük Türkçe Sözlük kaynaklı sözcük tanımları sadece kategoriye dayalı anlamsal ilişki çıkarımında değerlendirilmeye alınmıştır. Çalışma kapsamında kullanılan her bir sözlük ve barındırdıkları sözcük sayıları Tablo 1'de gösterilmiştir.

Tablo 1: Kullanılan sözlükler ve barındırdıkları sözcük sayıları

Sözlük	Kavram Sayısı
TDK Büyük Türkçe Sözlük	98107
Vikisözlük	160049

Çalışma dahilinde sözlük verilerinden çıkarılması hedeflenen anlamsal ilişkiler seçilirken, sözlükte yer alan kavramlara ilişkin şu sorulara cevaplar bulabilmek hedeflenmiştir:

- Bu kavramla anlamdaş olan diğer kavramlar nelerdir?
- Bu kavramla karşıt anlamlı olan diğer kavramlar nelerdir?
- Bu kavram hangi bütünün bir parçasıdır?
- Bu kavram hangi alt ve üst kavramlara sahiptir?
- Bu kavram hangi madde veya bileşenlerden yapılmıştır?
- Bu kavram ne amaçla kullanılmaktadır?
- Bu kavram neye benzemektedir?
- Bu kavram nerelerde bulunabilmektedir?

- Eylem, hangi hedefe ulaşmak için yapılmaktadır?
- Bu kavram ne ya da kim tarafından oluşturulmuştur?

Bu sorular yardımıyla çıkartılan her bir ilişki türü ve bunlarla ilgili eşleşme kuralları Bölüm 3.1 altında incelenmiştir. İlişki türleri seçilirken literatürde var olan ilişkilerin yanında metin işlemede faydalı olabileceği düşünülen farklı ilişki türlerine de yer verilmiştir. Söz konusu ilişkilerin yapılacak çalışmalarda ne şekilde kullanılabileceğine dair örnekler de yine Bölüm 3.1 altında yer almaktadır.

2. Önceki Çalışmalar

Bilgisayarlı sistemlerin kullanımına yönelik anlamsal veritabanı oluşturma konusunda farklı ve benzer yaklaşımları benimseyen birçok çalışma yapılmıştır. Yaklaşımlardan ilki veritabanlarının insanlarca elle oluşturulmasıdır ki bunun en iyi bilinen örneği George A. Miller tarafından oluşturulmaya başlanmış WordNet projesidir [4]. İsim, fiil, sıfat ve zarf türü sözcük ya da sözcük gruplarına ilişkin eş anlam kümeleri (*synset*) ve bu kümeler arası bazı anlamsal ilişkileri barındıran WordNet, bu belgenin kaleme alındığı tarihteki son sürümü olan 3.0 sürümünde 117.659 eş anlam kümesi içinde organize edilen 155.287 farklı sözcük barındırmaktadır. WordNet'e ait veri tabanına <http://wordnetweb.princeton.edu/perl/webwn> adresinden erişilebilir. WordNet'in az sayıda uzman kişi tarafından elle oluşturulması nedeniyle bu büyüklüğe erişmesi 10 yıldan fazla zaman almıştır. Aynı yaklaşımı (veritabanını az sayıda kişi ile oluşturmak) benimseyen bir başka proje ise Cyc [5]'tir. Lenat tarafından 1990 yılında oluşturulmaya başlanmıştır. Günümüzde ticari bir kuruluş haline gelen Cyc'in içerisinde yüz binlerce kavramın milyonlarca ilişkisinin olduğu söylenmektedir. Geliştirme sürecinde veri tabanının bir kısmı halka açılmış, web kullanıcılarının da veri tabanına katkıda bulunmaları amaçlanmıştır. Sistemin veri tabanına <http://www.cycfoundation.org/concepts> adresinden erişilebilir. OpenMind [6] projesi ise yeni bir yaklaşım sergilemiştir. OpenMind veritabanı, az sayıda uzman kişi yerine çok fazla sayıda uzman olmayan insanın Internet üzerinden katılımıyla oluşturulmuştur. OpenMind veritabanına katılımcılar bilgileri basit cümle formunda girmektedir. Bu basit yaklaşım sayesinde 1 yıl içinde veritabanında 1 milyona yakın cümle biriktirilmiştir. Bu büyük veritabanı tahmin edilebileceği üzere oldukça fazla gürtültü ve tekrar içermektedir. Bu soruna odaklanan araştırmacılar basit şablonlar kullanarak OpenMind projesinde toplanan cümlelerden otomatik olarak oluşturulmuş yaklaşık 200 bin kavram içeren bir anlamsal ağ (ConceptNet [7]) oluşturmuşlardır. Kavramlar arası ilişkiler ve bu ilişkilerin işlenmemiş OpenMind veri tabanındaki frekanslarından elde edilen güvenilirlik ölçümleri ConceptNet veri tabanında yer almaktadır. Veritabanına <http://web.media.mit.edu/~hugo/conceptnet/> adresinden erişilebilmektedir.

Konuyla ilgili çalışmalar son zamanlarda ülkemizde de yaygınlık kazanmıştır. Balkanet projesi dahilinde geliştirilen otomatik Türkçe WordNet oluşturma çalışması [8] sonucunda 11.628 eş küme ve bunlar arası 17.550 ilişki içeren bir anlamsal veritabanı oluşturulmuştur. Orijinal WordNet'le karşılaştırıldığında veritabanının oldukça küçük olduğu görülmektedir ancak veritabanındaki bilgilerin güvenilirliği

oldukça yüksektir. Literatürde, sözlük tanımlarını kaynak olarak kullanan ve bu açıdan burada sunulan çalışmaya benzer olan iki çalışma daha mevcuttur. İlkinde [9] WordNet'teki ilişkilerin büyük bir kısmını oluşturan alt kavram – üst kavram (*is a*) ilişkisine sahip kavram ikililerinin kelimelerin aldıkları eklerle bakılarak otomatik olarak bulunması gerçekleştirilmiştir. Diğerinde [10] ise alt kavram – üst kavram ve eşanlam ilişkileri otomatik olarak çıkartılmıştır. Burada sunulan çalışmanın, mevcut çalışmalardan iki temel farkı bulunmaktadır. İlki, diğerlerine göre çok daha fazla türde anlamsal ilişkinin çıkartılması ve bu sayede daha kapsamlı bir anlamsal veritabanı elde edilmesidir. İkincisi ise kaynak olarak iki farklı sözlük üzerinde çalışılmış olmasıdır.

3. Yöntem

Bu çalışmada kavramlar arası ilişkilerin belirlenmesinde sözcüğün sözlükte yer alan tanımının yanında gerekli olduğu durumlarda sözcüğün ve sözcüğün ilişkili olduğu kavramın sözcük türü bilgisinden de faydalanılmıştır. Bu açıdan bakıldığında sözlükte yer alan terimlerin söz dizimsel ve biçim bilimsel analizinin yanında anlam bilimsel analizi de yapılmaktadır. İlişki aranan sözcüğün sözcük türü bilgisi kullanılmakta olan sözlüklerde yer alırken, olası ilişkili sözcüğün sözcük türü bilgisi Zemberek [11] uygulaması kullanılarak edinilmektedir.

İlişki aranan sözcüğün sözcük türü bilgisinin kullanılmakta olan sözlükte araştırılması yerine Zemberek kullanılarak elde edilmesinin nedeni, sözcük tanımında yer alan olası ilişkili sözcüklerin çoğunlukla yalın halde bulunmasıdır. Sözlüklerde bulunan sözcükler ise çoğunlukla çekim eki almamış yalın halleriyle yer almaktadır. Zemberek, verilen bir sözcüğü yalın halde olmasa bile eklerinden arındırarak türünü tespit edebilmektedir. Ayrıca sözcüğün sözlük tanımında yer alan ve ilişkili sözcük olarak tespit edilen her sözcük sözlükte yer almamaktadır, dolayısıyla bu gibi sözcüklerin tür bilgisine sözlük tarafından sağlanan veri yardımıyla ulaşmak mümkün değildir. Zemberek, sözlükte tanımları yer alan sözcüklerden sayıca daha fazla sözcüğün türünü tespit edebilmektedir.

Sözlükte yer alan kavramlar arasındaki anlamsal ilişkilerin araştırılmasında sözcüğün türü ve sözlük tanımının yanında sözlükte bulunduğu kategori ve yüzey biçimi bilgilerinden de faydalanılabilmektedir [10]. Bölüm 3.1 altında sözcüğün türü ve sözlük tanımından faydalanılan kurallar yardımıyla ilişki çıkarımı yöntemi incelenirken, sözcüğün sözlükteki kategorisinden faydalanılarak alt kavram – üst kavram ilişkilerinin araştırılması konusuna Bölüm 3.2 altında değinilmiştir.

3.1. Kurallar Yardımıyla Anlamsal İlişki Çıkartımı

Sözlük verisi kullanılarak çıkartılması hedeflenen anlamsal ilişki türleri ile bu ilişkilerin saptanmasında kullanılan kurallar, örnekleriyle birlikte bu bölümün devamında incelenecektir. Çıkarılan kurallar, geliştirilen bir uygulama yardımıyla çalışma kapsamında kullanılan sözlükler üzerinde uygulanmış olup her bir sözlükteki eşleşme sayısı Tablo 2 ile Tablo 21 arasında yer alan tablolar üzerinde örnekleriyle birlikte gösterilmiştir. Kuralların farklı sözlükler üzerinde farklı başarılar elde etmesinin altında, sözlüklerde yer alan verinin nitelik ve standart bakımından farklı özelliklere sahip

olması yatmaktadır. Bu konuya Bölüm 4 altında çalışmanın sonuçlarında ayrıca değinilecektir.

Bilgisayarla doğal dil işlemede pek çok farklı yöntem ve teknik kullanılmaktadır. Biçimsel diller ve özdevinirler kuramı (*automata theory*) üzerine inşa edilen düzenli ifadeler (*regular expressions*) de bu amaçla kullanılan tekniklerden biridir [12]. Birden fazla benzer kalıbı tek bir kuralda toplayabilmek ve uygulamada esnekliği sağlamak amacıyla bu çalışmada düzenli ifadelerden faydalanılmıştır.

Düzenli ifadelerin yazım biçimi dilden dile az da olsa farklılık gösterebilmektedir, bu belgede yer verilen kuralların tamamında Perl Uyumlu Düzenli İfadeler (*PCRE - Perl-compatible Regular Expressions*) [13] kullanılmıştır. Ayrıca kullanılan düzenli ifade kalıplarının tamamı büyük küçük harf ayrımına duysızdır.

Çıkarılan kurallar her bir sözcüğün sözlükteki tanımına uygulanmadan önce sözcük tanımı, [10]'da da başvurulduğu üzere ayırıcı olarak virgül karakteri kullanılarak parçalara bölünmüş ve her bir parçanın başında ya da sonunda bulunabilecek beyaz boşluklar temizlenmiştir. Kuralın uygulandığı bu parçalar bundan böyle “sözcük tanımı parçası” olarak anılacaktır.

Sözcük türü ve ilişkili sözcük türü bilgileri kuralların eşleşme doğruluklarını arttırmak amacıyla ve tek başına düzenli ifade kuralıyla yapılan bir ilişki tespitinin yeterli güvenilirlikte olmadığı durumlarda kullanılmıştır. Tablo 2 ile Tablo 21 arasında yer alan tablolar üzerindeki “Sözcük türü” ve “İlişkili sözcük türü” alanlarında sıfat, isim, isim fiil gibi sözcük türlerinin varlığı; incelenen sözcüğün veya aranan ilişkili sözcüğün türüne kısıt getirmekte, kısıda uyulmaması durumunda çok sayıda hatalı sonuç alınabileceğine işaret etmektedir. Örneğin Tablo 12’de karşıt anlamlılık ilişkisinin tespitine yönelik önerilen kuralda sözcük türünün sıfat, ilişkili sözcük türünün ise isim ya da sıfat olması istenmektedir. Sözlük verisinde “Zambiya” sözcüğünün tanımında yer alan sözcük tanımı parçalarından birisi şudur: “denize kıyısı olmayan ülke”. Tablo 12’deki kuralın sözcük türü kısıtları olmaksızın bu sözcük tanımı parçasına uygulanması “Zambiya” ile “kıyı” ya da “kıyısı” sözcüğünün zıt anlamlı olarak tespit edilmesine neden olacaktır ki bu da istenmeyen hatalı bir durumdur. Sözcük türünün sıfat, ilişkili sözcük türünü ise isim ya da sıfat olarak kısıtlamak aynı sebepten kaynaklanabilecek hataları büyük ölçüde ortadan kaldırmaktadır.

Öte yandan sözcük türünün veya ilişkili sözcük türünün “Tümü” olarak belirtilmesi ilgili kuralın çok sayıda farklı sözcük türü üzerinde uygulanabileceğini ya da kurala bir sözcük türü kısıtı getirmenin eşleşme üzerinde yeterli bir olumlu etki sağlamayacağını ifade etmektedir. Örneğin Tablo 15’te yer alan yapılmış olma ilişkisinin saptanmasına yönelik önerilen kuralda -den veya -dan ile biten bir sözcüğün ardından “yapılan” veya “yapılmış” sözcüklerinin gelmesi gerektiği belirtilmiştir. Bu düzenli ifade kuralıyla saptanabilecek ilişkilerin tamamına yakınında sözcük türü ve ilişkili sözcük türü isim olmasına rağmen bu bilgi kural için bir kısıt olarak gösterilmemiştir. Zira sözcük türü ve ilişkili sözcük türü için isim türünü kısıt olarak göstermenin kuralın başarımına katkısı oldukça azdır. Öte yandan yeterli katkıyı

sağlıyor olmaksızın böyle bir kısıt kurala getirmenin bazı olumsuz etkileri vardır. Bunların en önemlisi, sözcük türlerinin kesin bir doğrulukla saptanamayışından ötürü doğru olarak tespit edilen bir ilişkinin, kuralın sözcük türü kısıtı sağlanamamasından dolayı sonuca dahil edilmeyişidir. Yazımı aynı olduğu halde türü farklı olan sözcüklerin varlığı bu durum üzerindeki olumsuz etkiyi daha da arttırmaktadır.

3.1.1. Alt Kavram – Üst Kavram İlişkisi

Alt kavram – üst kavram ilişkisi bir varlığın hangi üst varlık kümesine ait olduğu bilgisini içerir. Bu ilişkinin saptanmasında varlığın sözlük anlamının yanında sözlükteki kategorisinden de faydalanılabilir.

İncelenen sözcük tanımı parçası “tümü” sözcüğü ile sonlanıyorsa “tümü” sözcüğünden önce gelen sözcüğün yalnız halde incelenen sözcüğün üst kavramı olarak alınabilir.

Tablo 2: Alt kavram – üst kavram ilişkisi - Kural 1

Düzenli ifade	(?P<w>[^\s]*)? tümü\$	
Sözcük türü	İsim	
İlişkili sözcük türü	İsim	
Örnek	Sakal: Bazı hayvanlarda çene altında bulunan kılların tümü	
	Sözcük	Sakal
	İlişkili sözcük	Kıl
	İlişki türü	Alt kavram – üst kavram
Eşleşme sayısı	TDK Sözlük	124
	Vikisözlük	68

İncelenen sözcük tanımı parçası içerisinde “bir çeşit” ya da “bir tür” sözcük öbeklerinden sonra isim türü bir sözcük geliyorsa bu sözcük incelenen sözcüğün üst kavramı olarak alınabilir.

Tablo 3: Alt kavram – üst kavram ilişkisi - Kural 2

Düzenli ifade	bir (çeşit tür) (?P<w>[^\s]*)\$	
Sözcük türü	İsim	
İlişkili sözcük türü	İsim	
Örnek	Barbunya: Taneleri yuvarlak, oval veya yassı, kırmızı benekli, bir tür fasulye	
	Sözcük	Barbunya
	İlişkili sözcük	Fasulye
	İlişki türü	Alt kavram – üst kavram
Eşleşme sayısı	TDK Sözlük	473
	Vikisözlük	268

İncelenen sözcük tanımı parçası “iş” sözcüğü ile sonlanıyorsa incelenen sözcüğün üst kavramı “iş” olarak alınabilir.

Tablo 4: Alt kavram – üst kavram ilişkisi - Kural 3

Düzenli ifade	(?P<w>iş)i\$	
Sözcük türü	Tümü	
İlişkili sözcük türü	Tümü	
Örnek	Ağartma: Kuyumculukta gümüşü temizleme işi	
	Sözcük	Ağartma
	İlişkili sözcük	İş
	İlişki türü	Alt kavram – üst kavram
Eşleşme sayısı	TDK Sözlük	5633
	Vikisözlük	1119

İncelenen sözcük tanımı parçası “bilimi” ile sonlanıyorsa bu sözcüğün üst kavramı “bilim” olarak belirlenir.

Tablo 5: Alt kavram – üst kavram ilişkisi - Kural 4

Düzenli ifade	(?P<w>bilim)i\$	
Sözcük türü	İsim	
İlişkili sözcük türü	İsim	
Örnek	Semiyojoloji: Gösterge bilimi	
	Sözcük	Semiyojoloji
	İlişkili sözcük	Bilim
	İlişki türü	Alt kavram – üst kavram
Eşleşme sayısı	TDK Sözlük	130
	Vikisözlük	48

İncelenen sözcük tanımı parçası “yer” sözcüğü ile sonlanıyorsa incelenen sözcüğün üst kavramı “yer” olarak alınabilir.

Tablo 6: Alt kavram – üst kavram ilişkisi - Kural 5

Düzenli ifade	(?P<w>yer)\$	
Sözcük türü	İsim	
İlişkili sözcük türü	İsim	
Örnek	Arpalık: Arpa konulan yer	
	Sözcük	Arpalık
	İlişkili sözcük	Yer
	İlişki türü	Alt kavram – üst kavram
Eşleşme sayısı	TDK Sözlük	979
	Vikisözlük	399

İncelenen sözcük tanımı parçası “kimse” sözcüğü ile sonlanıyorsa incelenen sözcüğün üst kavramı “kişi” olarak alınabilir.

Tablo 7: Alt kavram – üst kavram ilişkisi - Kural 6

Düzenli ifade	kimse\$	
Sözcük türü	İsim	
İlişkili sözcük türü	Tümü	
Örnek	Aşçı: Yemek pişirip satan kimse	
	Sözcük	Aşçı
	İlişkili sözcük	Kişi
	İlişki türü	Alt kavram – üst kavram
Eşleşme sayısı	TDK Sözlük	3535
	Vikisözlük	1405

İncelenen sözcük tanımı içerisinde “bir isim türü” ya da “bir isim çeşidi” şeklinde bir yapı mevcutsa “isim” sözcüğü incelenmekte olan sözcüğün üst kavramı olarak alınabilir.

Tablo 8: Alt kavram – üst kavram ilişkisi - Kural 7

Düzenli ifade	bir (?P<w>[^]+?) (türü çeşidi)\$	
Sözcük türü	İsim	
İlişkili sözcük türü	İsim	
Örnek	Hortum: Hava veya suyun hızla dönüp sütun biçiminde yükselmesiyle oluşan, alanı dar bir siklon çeşidi	
	Sözcük	Hortum
	İlişkili sözcük	Siklon
	İlişki türü	Alt kavram – üst kavram
Eşleşme sayısı	TDK Sözlük	233
	Vikisözlük	1000

Sözcük tanımında yer alan “-giller” ekiyle biten sözcükler, genellikle hayvan ya da bitkilerin bilimsel sınıflandırmadaki ailesini (familya) göstermekte olup bu eki alan sözcüklerin varlığı, incelenen sözcük ile bu eki alan sözcük arasında bir alt kavram – üst kavram ilişkisi olduğuna işaret eder.

Tablo 9: Alt kavram – üst kavram ilişkisi - Kural 8

Düzenli ifade	(?P<w>[^]+?)giller(den familyasından ailesinden)	
Sözcük türü	İsim	
İlişkili sözcük türü	Tümü	
Örnek	Puhu: Baykuşgillerden, orman, dağ ve kayalıklarda yaşayan, uzunluğu 65 cm, sırtı koyu kahverengi bir kuş türü	
	Sözcük	Puhu
	İlişkili sözcük	Baykuş
	İlişki türü	Alt kavram – üst kavram
Eşleşme sayısı	TDK Sözlük	833
	Vikisözlük	1475

Alt kavram – üst kavram ilişkisi WordNet ve WordNet’ten dallanarak oluşturulan diğer projelerde yer verilen ilişkiler arasındadır. Bu ilişki türünün tespiti başta metin sınıflandırma olmak üzere makine çevirisi ve makine öğrenmesi gibi çalışma alanlarında kolaylık sağlayacaktır. 1998 yılında makine öğrenmesi üzerine yapılan deneysel bir çalışma [14],

sadece alt kavram – üst kavram ilişkisinden faydalanılarak metin sınıflandırma işlemine örnek olarak gösterilebilir.

3.1.2. Eş Anlamlılık İlişkisi

Eş anlamlılık ya da anlamdaşlık ilişkisi bir dilde bulunan ve birbirleriyle aynı ya da çok yakın anlama gelen farklı sözcükler arasındaki ilişkiyi tanımlar. Eş anlamlı sözcükler, sözcüğün sözlük tanımında sıklıkla virgül karakteriyle birbirinden ayrılmış vaziyette bulunur. Virgül karakteri, sözcük tanımı parçalarını birbirinden ayıran karakter olduğundan, incelenen sözcük tanımı parçasının tek bir sözcükten oluşması, incelenen sözcük ile bu sözcük arasında bir eş anlamlılık ilişkisi olabileceği sonucunu doğurur. Tablo 10’da görülen ve Balkanet projesi dahilinde otomatik Türkçe WordNet oluşturma çalışmasında da kullanılan bu kural [8], sözcük tanımındaki her virgülle ayrılan sözcüğün incelenen sözcüğün eş anlamlısı olmaması nedeniyle çok sayıda hatalı çıktı (*false positive*) üretmeye elverişlidir. Tabloda yer alan örneklerin ilki kuralın beklenen biçimde çalışması durumunda ürettiği çıktıları gösterirken, ikinci örnek söz konusu nedenden ötürü kuralın ürettiği hatalı çıktıları göstermektedir. Birinci örnekte yer alan ilişkili sözcükler verilmek istenen anlama göre acı sözcüğü yerine kullanılabilirken ikinci örnekte yer alan ilişkili sözcükler ağlamak eyleminin nedenleri ya da etkenleri olup ağlamak eyleminin anlamdaşı sözcükler değildir.

Tablo 10: Eş anlamlılık ilişkisi - Kural 1

Düzenli ifade	^(?P<w>[^]+)\$		
Sözcük türü	Tümü		
İlişkili sözcük türü	Tümü		
Örnek	Acı: Kırıcı, üzücü, incitici, dokunaklı, korkunç		
	Sözcük	Acı	
	İlişkili sözcük	Kırıcı	
	İlişkili sözcük	Üzücü	
	İlişkili sözcük	İncitici	
	İlişkili sözcük	Dokunaklı	
	İlişkili sözcük	Korkunç	
	İlişki türü	Eş anlamlılık	
Örnek	Ağlamak: Üzüntü, acı, sevinç, pişmanlık, aldanma vb.'nin etkisiyle gözyaşı dökmek		
	Sözcük	Ağlamak	
	İlişkili sözcük	Üzüntü	
	İlişkili sözcük	Acı	
	İlişkili sözcük	Sevinç	
	İlişkili sözcük	Pişmanlık	
	İlişki türü	Eş anlamlılık	
	Eşleşme sayısı	TDK Sözlük	46926
Vikisözlük		28593	

Sözcük tanımında yer alan, sözlükteki diğer kavramlara referanslar, çoğunlukla bu kavramın daha yaygın kullanılan şekline yönlendirmeler olup kavramlar arasındaki bir eş anlamlılık ilişkisine işaret eder.

Tablo 11: Eş anlamlılık ilişkisi - Kural 2

Düzenli ifade	^bk.(?P<w>.+)	
Sözcük türü	Tümü	
İlişkili sözcük türü	Tümü	
Örnek	Tümsekli: bk. dışbükey	
	Sözcük	Tümsekli
	İlişkili sözcük	Dışbükey
	İlişki türü	Eş anlamlılık
Eşleşme sayısı	TDK Sözlük	1846
	Vikisözlük	64

Tablo 11’de görülen kuralın Vikisözlük’teki eşleşme sayısının TDK Sözlük ile karşılaştırıldığında düşük olması, Vikisözlük’ün çevrimiçi olması nedeniyle çoğunlukla “bk.” şeklinde referanslar kullanmak yerine doğrudan referans sözcüğe köprü içermesinden kaynaklanmaktadır.

Eş anlamlılık ilişkisi doğal dil işleme çalışmalarında sıkça faydalanılan bir ilişki türüdür. Otomatik çeviri ve sınıflandırmanın dışında sistem başarımını arttırmak amacıyla başvuru sorgu genişletme (*Query Expansion*) işleminde sözcükler arasındaki eş anlamlılık ilişkisinin tespiti özel bir önem arz etmektedir [15]. Ayrıca WordNet projesinde yer alan ilişkiler de eş anlam kümeleri (*synset*) olarak isimlendirilen mantıksal kümelerin etrafında organize edilmiştir [4].

3.1.3. Karşıt Anlamlılık İlişkisi

Sözcüğün sözlük tanımında yer alan isim ya da sıfat türü bir sözcükten sonra gelen “olmayan” anahtar sözcüğü aracılığıyla karşıt anlamlılık ilişkisi saptanabilir.

Tablo 12: Karşıt anlamlılık ilişkisi - Kural 1

Düzenli ifade	(?P<w>[^]+ olmayan)	
Sözcük türü	Sıfat	
İlişkili sözcük türü	İsim ya da sıfat	
Örnek	Denetimsiz: Denetlenmiş olmayan	
	Sözcük	Denetimsiz
	İlişkili sözcük	Denetlenmiş
	İlişki türü	Karşıt anlamlılık
Eşleşme sayısı	TDK Sözlük	698
	Vikisözlük	69

Sözcük tanımında yer alan “karşıtı” sözcüğünün varlığı incelenen sözcük ile “karşıtı” sözcüğünden önce gelen sözcük arasında bir karşıt anlamlılık ilişkisi olduğuna işaret eder.

Tablo 13: Karşıt anlamlılık ilişkisi - Kural 2

Düzenli ifade	(P<w>[^]+ karşıtı)	
Sözcük türü	Sıfat	
İlişkili sözcük türü	İsim ya da sıfat	
Örnek	Alçak: Yerden uzaklığı az olan, yüksek karşıtı	
	Sözcük	Alçak
	İlişkili sözcük	Yüksek
	İlişki türü	Karşıt anlamlılık
Eşleşme sayısı	TDK Sözlük	100
	Vikisözlük	34

Karşıt anlamlılık ilişkisi eş anlamlılık ilişkisiyle birlikte metin sınıflama ve makine çevirisi gibi Yapay Zeka uygulamalarının yanında belge erişim sistemlerindeki arama hatalarını en aza indirmek amacıyla da kullanılabilir. Yapay Zekadaki çerçeve probleminin bir alt türü olarak nitelendirilen anlamsal karşıtılık probleminin çözümünde de karşıt anlamlı sözcüklerden faydalanılabilmektedir [16].

3.1.4. Parça - Bütün İlişkisi

Parça – bütün ilişkisi bir varlığın bütününe hangi parça varlıklardan oluştuğu bilgisini barındırır. Parça – bütün ilişkisinin literatürde farklı tanım ve yorumları mevcuttur. Burada adı geçen parça – bütün ilişkisi Cruse'nin genişletilmiş tanımı [17] uyarınca üye – bütün ilişkisini de içine almaktadır.

İsmin ayrılma hal ekini (-den, -dan) barındıran isim türü bir sözcükten sonra gelen “oluşan” ve “oluşmuş” anahtar sözcükleri bu ilişkinin saptanmasında kullanılabilir. Aynı tanım içerisinde bir varlığı oluşturan birden fazla varlık “ve” ya da “veya” bağlaçlarıyla birbirinden ayrılmış vaziyette bulunabilir.

Tablo 14: Parça – bütün ilişkisi kuralı

Düzenli ifade	^(([^]+)(lar ler)?(den dan tan ten)(ve veya))?([^\]+)?(lar ler)?(den dan tan ten)(oluşan oluşmuş)	
Sözcük türü	İsim	
İlişkili sözcük türü	İsim	
Örnek	Bölük: Takımlardan oluşan, üçü veya dördü bir tabur oluşturan ve öbür birliklerin temeli sayılan birlik	
	Sözcük	Bölük
	İlişkili sözcük	Takım
	İlişki türü	Parça - bütün
Eşleşme sayısı	TDK Sözlük	33
	Vikisözlük	10

Literatürde meronimi olarak da bilinen parça – bütün ilişkisi, yapılan anlamsal bir aramanın daha sığ bir sorguyla gerçekleşmesinde yardımcı olabilir. “A, B’yi içerir mi?” ya da “A nelerden oluşur?” gibi ontolojik sorulara cevaplar bulmaya yarayan parça – bütün ilişkisi; otomatik soru cevaplama ve bilgi çıkarımı sistemlerinde de kullanılabilir.

3.1.5. Yapılmış Olma İlişkisi

Yapılmış olma ilişkisi bir varlığın hangi varlık ya da varlıklardan yapıldığı bilgisini içerir. Bu ilişki parça – bütün ilişkisinin özel bir hali olup bir bütününe hangi öğelerden oluştuğunun yanında öğenin ya da öğelerin belli bir işleme tabi tutularak bütünü oluşturduğu bilgisine de gönderme yapar.

İsmin ayrılma hal ekini (-den, -dan) barındıran isim türü bir sözcükten sonra gelen “yapılan” ve “yapılmış” anahtar sözcükleri bu ilişkinin saptanmasında kullanılabilir. Aynı tanım içerisinde bir varlığın yapılmış olduğu birden fazla varlık “ve” ya da “veya” bağlaçlarıyla birbirinden ayrılmış vaziyette bulunabilir.

Tablo 15: Yapılmış olma ilişkisi kuralı

Düzenli ifade	^((([^\]+)(lar ler)?(den dan tan ten)(ve veya))?([^\]+)?(lar ler)?(den dan tan ten)(yapılan yapılmış)	
Sözcük türü	Tümü	
İlişkili sözcük türü	Tümü	
Örnek	Halat: Kenevirden yapılmış çok kalın ip	
	Sözcük	Halat
	İlişkili sözcük	Kenevir
	İlişki türü	Yapılmış olma
Örnek	Katma: Kıldan veya yünden yapılmış ip	
	Sözcük	Katma
	İlişkili sözcük	Kıl
	İlişkili sözcük	Yün
	İlişki türü	Yapılmış olma
Eşleşme sayısı	TDK Sözlük	163
	Vikisözlük	57

ConceptNet [7], yapılmış olma (*madeOf*) ilişkisini burada olduğu gibi parça – bütün ilişkisinden bağımsız olarak tanımlamaktadır. Uygulamada parça – bütün ilişkisiyle aynı amaca yönelik faydalanılabilecek bu ilişki türü daha fazla bilgiyi daha yüksek bir kesinlikte barındırması nedeniyle ayrı olarak değerlendirilmiştir.

3.1.6. Yapan – Yapılan İlişkisi

Bu ilişki, yapılan bir eylemin kimin tarafından yapıldığı ya da bir varlığın veya soyut bir kavramın kimin tarafından oluşturulduğu bilgisini içerir. Sözcüğün sözlük tanımında yer alan “tarafından” anahtar sözcüğü bu ilişkinin saptanmasında kullanılabilir.

Tablo 16: Yapan – yapılan ilişkisi kuralı

Düzenli ifade	(?P<w>[^]+)(lerce larca tarafından) (yapılan oluşturulan)	
Sözcük türü	Tümü	
İlişkili sözcük türü	Tümü	
Örnek	Ekspertiz raporu: Ekspertler tarafından yapılan inceleme sonunda hazırlanan rapor	
	Sözcük	Ekspertiz raporu
	İlişkili sözcük	Ekspert
	İlişki türü	Yapan – yapılan
Eşleşme sayısı	TDK Sözlük	5
	Vikisözlük	3

Yapan – yapılan ilişkisi ConceptNet'te *CreatedBy* adıyla yer almaktadır.

3.1.7. Araç - Amaç İlişkisi

Araç – amaç ilişkisi bir varlığın hangi amaçla kullanıldığı bilgisini içerir. Sözlük tanımında bulunan “için” ve “amacıyla” sözcükleri bu ilişkinin tespit edilmesinde anahtar sözcük olarak kullanılabilir. Amaç rolünü üstlenen sözcüğün çoğunlukla isim fiil olması bu ilişkinin saptanmasında kullanılan bir başka veridir. Aynı tanım içerisinde bir varlığın birden fazla kullanım amacı “ve” ya da “veya” bağlaçlarıyla birbirinden ayrılmış vaziyette bulunabilir.

Tablo 17: Araç – amaç ilişkisi kuralı

Düzenli ifade	(([^]+(mek mak me ma)) (ve veya) ?)([^]+(mek mak me ma)) (için amacıyla) yapılan	
Sözcük türü	Tümü	
İlişkili sözcük türü	İsim fiil	
Örnek	Anma: Ölmüş bir insanı hatırlamak için yapılan tören	
	Sözcük	Anma
	İlişkili sözcük	Hatırlamak
	İlişki türü	Araç - amaç
Örnek	Rampa: Özellikle istasyonlarda, vagonlara eşya yüklemek veya boşaltmak için yapılan, ambarın önünde bulunan set	
	Sözcük	Rampa
	İlişkili sözcük	Yüklemek
	İlişkili sözcük	Boşaltmak
	İlişki türü	Araç - amaç
Eşleşme sayısı	TDK Sözlük	137
	Vikisözlük	60

3.1.8. Bulunma İlişkisi

Bulunma ilişkisi, bir varlığın nerede bulunduğu konusunda bilgi verir. İsmi bulunma hal ekini (-de, -da) barındıran isim türü bir sözcükten sonra gelen “bulunan” anahtar sözcüğü, bulunma ilişkisinin sözcük anlamından tespit edilebilmesinde kullanılabilir.

Tablo 18: Bulunma ilişkisi kuralı

Düzenli ifade	^(?P<w>[^]+)(ler lar)?(de da te ta) bulunan	
Sözcük türü	İsim	
İlişkili sözcük türü	İsim	
Örnek	Granül: Stoplâzma bulunan küçük tanecikler	
	Sözcük	Granül
	İlişkili sözcük	Stoplâzma
	İlişki türü	Bulunma
Örnek	Derbent: Sınırlarda bulunan küçük kale	
	Sözcük	Derbent
	İlişkili sözcük	Sınır
	İlişki türü	Bulunma
Eşleşme sayısı	TDK Sözlük	79
	Vikisözlük	53

Burada bahsi geçen bulunma ilişkisi WordNet'te parça – bütün ilişkisi dahilinde değerlendirilmektedir. ConceptNet ise bulunma ilişkisine *LocationOf* adı altında parça – bütün ilişkisinden bağımsız olarak yer vermiştir. Bu ilişki 3.1.4 altında incelenen parça – bütün ilişkisine göre daha zayıf olup bir varlığa ilişkin “Nerede bulunur?” sorusuna cevap bilgisini taşır.

3.1.9. Kullanılma İlişkisi

Kullanılma ilişkisi bir varlığın hangi amaçla kullanıldığı konusunda bilgi verir. Sözcüğün sözlük tanımında yer alan “için kullanılan” ve “amacıyla kullanılan” sözcük öbekleri bu ilişkinin saptanmasında kullanılabilir. Bu ilişkinin araç – amaç ilişkisindeki farkı, araç – amaç ilişkisinde bir oluş söz konusuken bu ilişkide bir amaç uğruna başka bir varlıktan faydalanma söz konusudur. Araç – amaç ilişkisindeki benzer biçimde bu ilişkide de faydalanılan varlığın tanım içerisinde çoğunlukla isim fiil olarak yer alması bu ilişkinin tespitinde kullanılan bir veridir.

Tablo 19: Kullanılma ilişkisi - Kural 1

Düzenli ifade	(?P<w>[^]+?(mek mak me ma)) (için amacıyla) kullanılan	
Sözcük türü	İsim	
İlişkili sözcük türü	İsim fiil	
Örnek	Çıkış: Bir yerden çıkmak için kullanılan yer	
	Sözcük	Çıkış
	İlişkili sözcük	Çıkmak
	İlişki türü	Kullanılma
Örnek	Sprey: Püskürtme amacıyla kullanılan araç	
	Sözcük	Sprey
	İlişkili sözcük	Püskürtme
	İlişki türü	Kullanılma
Eşleşme sayısı	TDK Sözlük	367
	Vikisözlük	233

İsmi yönelme hal ekini (-e, -a) barındıran isim türü bir sözcükten sonra gelen “yarayan” sözcüğü, kullanıma ilişkisinin saptanmasında faydalanılan bir başka kuraldır.

Tablo 20: Kullanılma ilişkisi - Kural 2

Düzenli ifade	(P<w>[^]+?(me ma))(ye ya) yarayan	
Sözcük türü	Tümü	
İlişkili sözcük türü	Tümü	
Örnek	Kimlik: Herhangi bir nesneyi belirlemeye yarayan özelliklerin bütünü	
	Sözcük	Kimlik
	İlişkili sözcük	Belirleme
	İlişki türü	Kullanılma
Eşleşme sayısı	TDK Sözlük	639
	Vikisözlük	304

Kullanılma ilişkisi ConceptNet’in yer verdiği ilişkiler arasındadır.

3.1.10. Benzerlik İlişkisi

Benzerlik ilişkisi bir varlığın başka hangi varlıklara benzediği konusunda bilgi verir. Buradaki benzerlikten kasıt çoğunlukla biçimsel benzerliktir.

İsmi yönelme hal ekini (-e, -a) barındıran isim türü bir sözcükten sonra gelen “benzeyen” anahtar sözcüğü, benzerlik ilişkisinin çıkarılmasında kullanılabilir.

Tablo 21: Benzerlik ilişkisi kuralı

Düzenli ifade	(P<w>[^]+?)y?(a)e benzeyen	
Sözcük türü	İsim	
İlişkili sözcük türü	Tümü	
Örnek	Halka: Su gibi sıvıların içine katı bir nesnenin düşmesiyle oluşan, gittikçe büyüyen açılan çembere benzeyen biçim	
	Sözcük	Halka
	İlişkili sözcük	Çember
	İlişki türü	Benzerlik
Eşleşme sayısı	TDK Sözlük	185
	Vikisözlük	123

Benzerlik ilişkisinin ilk aklı gelen kullanım alanı kavram sınıflandırma ve kümelemedir. Herhangi bir amaca yönelik sınıflama ya da kümeleme yaparken aralarında fiziksel benzerlik bulunan varlıkları aynı kümeye ya da sınıfa yerleştirmek çoğunlukla istenilen bir durumdur. Böyle bir durum arzu edildiğinde benzerlik ölçütü olarak burada geçen benzerlik ilişkisine başvurulabilir. Benzerlik ilişkisi WordNet’te de yer almakla birlikte WordNet’teki tanımı daha ziyade anlamsal benzerlik üzerinedir.

3.2. Sözlük Kategorisine Dayalı Anlamsal İlişki Çıkartımı

Sözcüğün sözlük tanımı ve türü dışında sözcüğün kategori bilgisi de o sözcüğe ait alt kavram – üst kavram ilişkilerinin çıkartımında kullanılabilir. Türk Dil Kurumu’na ait Türkçe sözlükte yer alan kategoriler “zooloji”, “botanik” ve “tarih” gibi sözcüğün terim olarak kullanıldığı alanı (*context*) ifade ederken; Wikisözlük’te bu tür bir bölümlendirmenin yanında “kuş”, “balık” ve “organ” gibi sözcüğün doğrudan üst kavramını ifade eden kategorilerin yer alması yeni üst kavramların çıkartılabilmesine olanak sağlamaktadır. Tablo 22’de Wikisözlük’te yer alan bu veriden faydalanılarak çıkartılan üst kavramlar görülmektedir. Tablodaki verilere sadece 10 ve daha fazla sözcük içeren kategoriler dahil edilmiştir.

Tablo 22: Wikisözlük’teki kategorilerden faydalanılarak çıkartılan üst kavramlar

Üst Kavram	Adet
köy	17831
belediye	2687
belde	1932
kuş	1290
ilçe	880
ülke	361
memeli	187
yemek	149
renk	119
balık	110
ağaç	84
il	81
organ	81
hastalık	58
silah	56
meyve	55
böcek	54
içecek	52
fırtına	47
akrabalık	42
sebze	37
çiçek	36
şehir	33
ada	33
siyasi parti	31
element	30
deniz	27
oyun	26
ölçü birimi	24
baharat	23
mahalle	21
bilim dalı	18
sürüngen	14
dağ	12
semt	10
makam	10
Toplam	26630

Tabloda üst kavram olarak yer verilen kavramlardan bazıları diğerlerinden farklı olarak alt kavram olarak özel isimleri barındırmaktadır. Köy, ilçe, belde ve mahalle gibi yerleşim

birimi türleri; ada, deniz ve dağ gibi coğrafi oluşumlar ile siyasi parti üst kavram türleri alt kavram olarak özel isimleri barındırmaktadır.

Kategorilerden faydalanılarak alt kavram – üst kavram ilişkisi çıkartma işlemi sadece Wikisözlük üzerinde uygulanmış olup TDK sözlüğünde sözcüğün kullanım alanı dışında bir kategori ayırımının mevcut olmamasından ötürü TDK sözlüğü üzerinde böyle bir çalışma yapılmamıştır.

3.3. Sözlük Tarafından Sağlanan Anlamsal İlişkiler

Vikisözlük, sözcük tanımı ve sözcük türü gibi pek çok faydalı bilginin yanında bu belgenin konusu olan kavramlar arası anlamsal ilişkilere de yer vermektedir. Wikisözlük'te yer alan girdilerin çözümlenmesi (*parsing*) sırasında edinilen kavramlar arası anlamsal ilişkilere ait veri Tablo 23'te gösterilmiştir.

Tablo 23: Wikisözlük'te yer alan kavramlar arası anlamsal ilişkiler

İlişki Türü	Adet
Eş anlamlılık	7271
Bilimsel adı	1885
Sözcük birliktelikleri	1868
Alt kavram	1306
Yan kavram	1165
Üst kavram	1025
Karşıt anlamlılık	259
Benzer sözcük	120
Toplam	14899

Vikisözlük, sözlük verisini ilişkisel yapıda saklamıyor olmasından ötürü ilişkiler arası simetriklik ve geçişlilik gibi özelliklerden faydalanamamaktadır. Bunun bir sonucu olarak alt kavram – üst kavram ilişkisinde olduğu gibi bir varlığın hangi varlığın alt kavramı olduğu bilinirken, üst kavram olarak bilinen varlığa ilişkin sözlük verisinde bu varlığın hangi varlık ya da varlıkların üst kavramı olduğu bilgisinin yer almadığı durumlar olabilmektedir. Tablo 23'te verilen ilişki türü sayıları sözlükte doğrudan bulunan ilişki sayılarıdır. Anlamsal ilişki türleri arasındaki geçişlilik ve yansıma özelliklerinden faydalanılarak ilişki sayılarını arttırmak mümkündür. Yazarlar, bu işlemin uygulama ihtiyacına göre çalışma anında yapılabılır olduğu kanaatini taşıdıklarından elde edilen ilişkileri bu işleme tabi tutmamıştır.

Vikisözlük, literatürde yaygın kullanılan anlamsal ilişkilerin yanında “Bilimsel adı”, “Yan kavram” ve “Sözcük birliktelikleri” olmak üzere üç farklı ilişki türüne de yer vermektedir. “Bilimsel adı”, çoğunlukla canlı varlıklar için var olup canlının bilimsel sınıflandırmadaki adını içermektedir. Örnekleme gerekirse Wikisözlük'te *Kedi* sözcüğünün bilimsel adına bakıldığında *Felis catus* ve *Felis domesticus* olduğu görülmektedir. Yan kavram ise Wikisözlük tarafından şu şekilde tanımlanmaktadır: “Sözcüklerin anlamsal sınıflandırılmasında yan kavram (yani aynı düzeyde yer alan kavramlar) olarak görülen sözcükler burada yer alır.” [18]. *Otomobil* sözcüğünün yan kavramlarına örnek olarak *kamyon*, *TIR* ve *traktör* sözcükleri gösterilmiştir. Sözcük birliktelikleri ise Türkçede istatistiksel olarak anlamlı bir şekilde birlikte görülme sıklıkları yüksek olan sözcükleri ifade etmektedir.

Örneğin *empati* sözcüğü için *empati kurmak* sözcük öbeği sözcük birlikteliği olarak gösterilirken *beyaz* sözcüğü için *beyaz perde* sözcük birlikteliği olarak gösterilen sözcük öbekleri arasındadır. Sözcük birliktelikleri birer anlamsal ilişki olmamakla birlikte başta anlam belirsizliklerinden (*Word Sense Disambiguation*) kaynaklanan sorunları istatistiksel yöntemler kullanarak en aza indirmek üzere doğal dil işleme uygulamalarında sözcük birlikteliklerinden yararlanılması mümkündür.

4. Sonuçlar ve Tartışma

Bu çalışmada iki farklı sözlük kullanılarak bilgisayar uygulamalarının kullanımına yönelik anlamsal ilişki çıkarımıyla ilgili farklı otomatik yöntemler ve bu yöntemlerin uygulamaları ele alınmıştır. Her bir yöntemin uygulanması sonucu çıkartılan anlamsal ilişki türü ve sayıları ilgili yöntemin incelendiği bölüm altında verilmiştir. Çalışma kapsamında çıkartılan toplam ilişki sayıları ise Tablo 24'te görülmektedir. Tabloda yer alan “Toplam” sütunu iki sözlükten çıkartılan ilişkilerin birleşiminin sayısını göstermekte olup tekrarlı ilişkilerin sayısı bu sütundaki veriye dahil edilmemiştir. Tablodaki veriler Bölüm 3.1 altında incelenen ve bu belgede esas olarak üzerinde durulan kurala dayalı ilişki çıkarımı sonucu elde edilen verilerdir. Tablo 22'de ve Tablo 23'te yer alan diğer yöntemlere ait veriler bu tablodaki verilere dahil edilmemiştir, dahil edilmeyen bu verilerle birlikte toplam çıkartılan ilişki sayısı 127.203 adettir. Bu ilişkileri barındıran veritabanına http://www.kemik.yildiz.edu.tr/data/File/TKD_viki.rar adresinden erişilebilir.

Tablo 24: Çalışma kapsamında, kurala dayalı ilişki çıkartımı yöntemiyle çıkartılan anlamsal ilişki sayıları

İlişki Türü	İlişki Sayısı		
	TDK Sözlük	Vikisözlük	Toplam
Alt – üst kavram	11754	6354	16405
Eş anlamlılık	47196	28185	66531
Karşıt anlamlılık	109	106	190
Yapılmış olma	160	56	199
Parça - bütün	32	10	38
Yapan – yapılan	229	241	422
Araç - amaç	135	57	166
Bulunma	79	53	123
Kullanılma	1000	530	1331
Benzerlik	184	123	269
Toplam	60878	35715	85674

Çıkartılan ilişkilerin doğruluğu, kullanılan yöntemlere göre farklılık göstermektedir. Sözlük tarafından sağlanan anlamsal ilişkiler insan eliyle girildiğinden bu ilişkilerin güvenilirliklerinin bir hayli yüksek olduğu kabul edilmektedir. Sözlükteki kategori verisinin de insan eliyle girilmiş olması nedeniyle benzer bir çıkarım sözcüğün sözlükteki kategorisine dayanılarak çıkartılan anlamsal ilişkiler için de yapılabilir. Kurala dayalı çıkartılan anlamsal ilişkilerin güvenilirliği ise sözlük verisinin niteliği ve uygulanan kuralın niteliğine göre değişiklik göstermektedir. Burada geçen sözlük verisinin niteliğinden kasıt sözlük verisinin bilgisayarlarca otomatik olarak işlenmeye elverişli olup olmadığıdır.

Sözlük verisinin niteliğinin başlıca ölçütünün sözlüğün yazımında belli ortak standartların izlenip izlenmemesi olduğu söylenebilir. Bölüm 3.1 altında sunulan kurallara ilişkin eşleşme sayıları incelendiğinde TDK sözlüğünün toplamda Vikisözlük'ten daha az sayıda kavram barındırmasına rağmen pek çok kural için Vikisözlük'ten sayıca daha fazla eşleşme yaptığı görülür. TDK sözlüğünün belli bir kurul tarafından hazırlandığı, Vikisözlük'ün ise imce yöntemiyle hazırlandığı [3] verileri dikkate alınır, sözlük yazımında belli standartlara uyulmasının sözlük verisinin niteliğini olumlu yönde etkilediği yargısının yanlış olmadığı sonucuna ulaşılır.

Uygulanan kuralların niteliği çıkartılan ilişkilerin doğruluğuna etki eden bir başka unsurdur. Kuralların kapsadığı eşleşme sayısı arttıkça eşleşme sonuçlarındaki doğruluğun azaldığı gözlemlenmiştir. Kuralların gerçekleşmesinde kullanılan düzenli ifadeler, çıkartılan ilişki sayısı ve niteliği arasında bir denge kuracak sıklıkta seçilmiştir. Tablo 10'da verilen ve [10]'da da önerilen eş anlamlılık ilişkisi çıkarma kuralı bu duruma istisnadır. Daha önce de değinildiği üzere bu kural, sözcüğün sözlük tanımında yer alan virgülle ayrılmış her sözcüğü bu sözcüğün eş anlamlısı olarak kabul etmektedir; oysa söz konusu sözcüğün eş anlamlısı olmadığı halde sözcük tanımında virgülle ayrılmış ve yalnız başına bulunan çok sayıda sözcük olabilmektedir. Sözcüğün sözlük tanımının yanında sözcüğün ve ilişkili sözcüğün türü verileri de gerek görüldüğü durumlarda elde edilen anlamsal ilişkilerin doğruluğunu artırıcı unsurlar olarak kullanılmıştır.

Elde edilen verilerin doğruluklarını kontrol etmek için tüm veri kümesinin indirgenmiş alt kümeleri üzerinde eğitici uygulanan bir yöntemin yanında Türkçe WordNet'te yer alan eş kümeler arası ilişkiler ile çıkartılan ilişkilerin otomatik olarak karşılaştırılması da yapılmıştır.

Veri kümesinin büyüklüğü göz önüne alındığında veri kümesinin tamamı üzerinde elle doğruluk kontrolü yapmanın oldukça zahmetli bir işlem olduğu görülmektedir. Bu sebeple tüm veri kümesinin doğruluğunun kontrol edilmesi yerine her ilişki türü için veri kümesinden bir alt küme seçilmiş ve ilişkiler elle doğru ve yanlış olarak işaretlenmişlerdir. Tablo 25'te bu işlemin sonuçları görülmektedir. Bu sonuçlara bakılarak yapılan otomatik bilgi çıkarımı işleminin doğruluk oranı hakkında tam olmasa da genel bir yargıya varılabilir.

Tablo 25: Çalışma kapsamında, kurala dayalı ilişki çıkartımı yöntemiyle çıkartılan anlamsal ilişki sayıları

İlişki Türü	Toplam İlişki Sayısı	İncelenen İlişki Sayısı	Doğruluk Oranı
Alt – üst kavram	16405	246	0,94
Eş anlamlılık	66531	206	0,88
Karşıt anlamlılık	190	49	0,94
Yapılmış olma	199	61	0,48
Parça – bütün	38	20	0,55
Yapan – yapılan	422	128	0,36
Araç – amaç	166	113	0,55
Bulunma	123	47	0,34
Kullanılma	1331	92	0,73
Benzerlik	269	96	0,54

Tablo 25 incelendiğinde, genelde çok sayıda ilişkinin bulunduğu ilişki türlerindeki başarı oranının daha yüksek olduğu görülmektedir. Örneğin incelenen 246 alt kavram – üst kavram ilişkisinin %94'ü, 206 eş anlamlılık ilişkisinin ise %88'i doğru ilişkilidir.

Her ne kadar elle yapılan doğruluk kontrolü güvenilirlik bakımından tatmin edici olsa da kapsam bakımından yeterli görülmeyebilir. Bundan dolayı çıkartılan ilişkilerin doğru olduğu bilinen bir anlamsal ilişki kümesiyle değişik ölçütler kullanılarak otomatik olarak karşılaştırılması, ilişkilerin güvenilirliği hakkında okuyucuya daha fazla fikir vermesi bakımından faydalı olabilir. Türkçedeki bütün anlamsal ilişkileri barındıran eksiksiz ya da en azından kapsamlı bir anlamsal ilişki veritabanı olmasa da, Türkçe WordNet, barındırdığı ilişki sayısı ve niteliği göz önünde bulundurulduğunda karşılaştırma amacıyla kullanılabilir diğer kaynaklar arasında avantajlı sayılabilir.

Çıkartılan ilişkileri Türkçe WordNet'te yer alan ilişkilerle karşılaştırmadan önce veri kümeleri arasındaki yapısal ve anlamsal farklılardan kaynaklanabilecek hataları en aza indirmek için veri kümelerini bazı ön işlemlere tabi tutmak ihtiyacı duyulmuştur.

Bahsi geçen yapısal farklılık, Türkçe WordNet'te yer alan ilişkiler ile otomatik çıkartılan ilişkilerin farklı biçimlerde temsil ediliyor olmasıdır. 2. Bölüm altında değinildiği üzere WordNet'te anlamsal ilişkiler eş anlamlı kümeleri arasındadır. Başka bir deyişle anlamsal ilişkiler (*eş küme A, ilişki türü, eş küme B*) üçlüleriyle temsil edilmektedir. Eş kümeler ise aralarında eş anlamlılık ilişkisi bulunan sözcüklerden oluşmaktadır. Bu çalışmaya konu olan anlamsal ilişkiler ise WordNet'ten farklı olarak (*sözcük A, ilişki türü, sözcük B*) üçlüleriyle temsil edilmektedir. Bu farklılığı ortadan kaldırmak için Türkçe WordNet'te yer alan ilişkiler açılarak çıkartılan ilişkilerle aynı biçimde temsil edilir duruma getirilmiştir.

Türkçe WordNet ile çıkartılan ilişkiler arasındaki anlamsal farklılık ise ilişki türlerindeki farklılık ya da ilişkilerin tanımlarındaki yorumsal farklılıklardan kaynaklanmaktadır. Örneğin ConceptNet anlamsal veritabanı bu çalışmada konu olan yapılmış olma ilişkisine yer verirken bu ilişki türü WordNet'te yer almamaktadır. Benzer durum kullanılma ilişkisi için de geçerlidir. Bulunma ve benzerlik ilişkileri ise incelendikleri alt başlıklarda da değinildiği üzere WordNet'te benzer isimlendirilen ilişki türlerine göre farklı olarak tanımlanmaktadır. Bunların dışındaki ilişkilerden alt kavram – üst kavram, eş anlamlılık ve karşıt anlamlılık türleri haricindekiler; ya çalışma kapsamında çıkartılan ilişki türlerinde ya da Türkçe WordNet'te karşılık buldukları ilişki türlerinde sayısal olarak söz konusu üç ilişki türüne göre oldukça azdır. Tüm bu nedenlerin yanında literatürdeki yaygın kullanımları da göz önünde bulundurularak karşılaştırmada sadece bu üç ilişki türüne yer verilmiştir.

Çıkartılan ilişkiler ile Türkçe WordNet arasında yapılacak bir karşılaştırmanın isabetli sonuç vermesinin önünde olabilecek engeller bunlarla da sınırlı değildir. WordNet ile yapılacak bir karşılaştırmanın doğruluğunun önünde olabilecek bir başka engel de çıkartılan ilişkilerin barındırdıkları dolaylı bilgidir kaynaklanabilecektir. Çalışma kapsamında uygulanan diğer

yöntemlerde olduğu gibi Tablo 25’te sonuçları verilen kuralla dayalı ilişki çıkartımı yönteminde de sadece doğrudan elde edilen anlamsal ilişkiler sonuç verisine yansıtılmıştır. Anlamsal ilişki türlerinin tanımları gereği simetriklik ve geçişlilik gibi özelliklerinden faydalanılarak gerçekte sonuç verisinde saklı olan anlamsal ilişkilere dair daha fazla bilgi çıkarılabilir. Örneğin eş anlamlılık ve karşıt anlamlılık simetri özelliğine sahip birer ilişki türüken alt kavram – üst kavram ilişki türü eş anlamlı sözcük kümeleri üzerinde geçişlilik özelliğine sahiptir [19]. WordNet, yapısal organizasyonunun doğası gereği ilişki türlerinin bu özelliklerinden faydalanmaktadır. Yapılacak bir karşılaştırmalı değerlendirmenin doğru sonuç verebilmesi bakımından otomatik çıkartılan ilişkiler için de bu özelliklerden faydalanılması yerinde olacaktır.

Türkçe WordNet’te yer alan ilişkilerin (*sözcük A, ilişki türü, sözcük B*) biçimine dönüştürülmesi ve ilişki türlerinin özelliklerinden faydalanılarak dolaylı ilişkilerin açığa çıkarılması ön işlemleri sonrasında seçilen üç ilişki türü için ilişki kümelerinde yer alan ilişki sayıları Tablo 26’da verilmiştir. Tablonun birinci sütununda otomatik çıkartılan ilişki (OÇİ) sayıları, ikinci sütununda Türkçe WordNet’te (TWN) yer alan ilişki sayıları, son sütunda ise her iki kümede de ortak olarak yer alan ilişki sayıları görülmektedir.

Tablo 26: Uygulanan ön işlemler sonrasında Türkçe WordNet’te yer alan ve otomatik çıkartılan ilişkilerin güncel sayıları

İlişki Türü	İlişki Sayıları		
	OÇİ	TWN	Ortak
Alt – üst kavram	411718	21917	248
Eş anlamlılık	772621	14230	3094
Karşıt anlamlılık	15592	3035	40
Toplam	1199931	39182	3382

Tablodan anlaşılacağı üzere karşılaştırma için referans alınan veri kümesi (Türkçe WordNet) ile “hesaplanan” ya da “öngörülen” değer olarak işleme alınacak otomatik çıkartılan ilişkiler kümesi boyut itibarıyla birbirinden oldukça farklıdır. Dahası, hem Türkçe WordNet hem de otomatik olarak çıkartılan ilişkiler kümesi Türkçedeki tüm sözcükler arasındaki tüm anlamsal ilişkileri barındıran uzayın büyüklüğü göz önünde bulundurulduğunda oldukça az sayıda ilişki içermektedir. Bunun doğal bir sonucu olarak WordNet’te yer aldığı halde otomatik ilişki çıkartma işleminde kullanılan sözlüklerde yer almayan çok sayıda sözcük olduğu gibi, kaynak sözlüklerde yer aldığı halde WordNet’te yer almayan çok sayıda sözcük de mevcuttur. Bu nedenle doğruluk, isabetlilik ve geri getirim gibi ölçütler kullanarak anlamsal ilişkileri barındıran veri kümelerini değerlendirirken sadece her iki kümede de ortak olarak yer alan sözcükler arasındaki anlamsal ilişkileri değerlendirmeye almak yerinde olacaktır. Değerlendirmede referans alınacak veri kümesi olan Türkçe WordNet’te 15.590, değerlendirmeye tabi tutulacak otomatik çıkartılan ilişki kümesinde ise 57.918 adet farklı sözcük olduğu görülmüştür. Her iki kümede ortak olarak bulunan sözcüklerin sayısı ise 6.037 olarak tespit edilmiştir. Tablo 27’de, sadece bahsi geçen ortak sözcükler arasındaki anlamsal ilişkilerin kümeler bazında sayısı ile bu kümelerde ortak olarak yer alan anlamsal ilişkilerin sayısı verilmiştir.

Tablo 27: Karşılaştırmaya hazır anlamsal ilişki kümelerindeki ortak sözcükler arasındaki ilişki sayıları

İlişki Türü	İlişki Sayıları		
	OÇİ	TWN	Ortak
Alt – üst kavram	10505	9713	248
Eş anlamlılık	22869	7494	3094
Karşıt anlamlılık	769	1817	40
Toplam	34143	19024	3382

Her iki ilişki kümesinde ortak olarak yer alan sözcükler arasındaki ilişkiler üzerinden değerlendirme yapmak, bu ayrımı gözetmeksizin yapılan bir değerlendirmeye göre daha anlamlı olsa da bütünüyle sağlıklı sayılmaz. Değerlendirmede referans olarak alınan Türkçe WordNet alt kümesinin barındırdığı bütün sözcüklerin aralarında bulunabilecek bütün anlamsal ilişkilerin Türkçe WordNet’te yer aldığından söz etmek mümkün değildir. Örneğin “beceri” ve “hüner” sözcüklerinin her ikisi de karşılaştırılan her iki sözcük kümesinde ortak olarak bulunmasına rağmen, otomatik çıkartılan ilişkiler arasında (*beceri, eş anlamlılık, hüner*) anlamsal ilişkisi yer alırken Türkçe WordNet gerçekte var olan [2, 3] bu ilişkiye yer vermemiştir. Aralarında anlamsal ilişki olduğu bilinen iki sözcük için ilişki türünün doğru belirlenip belirlenmediğine dair yapılacak bir değerlendirme, bu gibi durumlardan kaynaklanabilecek yanlışları ortadan kaldıracaktır. Bu amaçla karşılaştırmaya konu olan (*sözcük A, ilişki türü, sözcük B*) biçimindeki ilişkilerden, (*sözcük A, sözcük B*) ikililerinin her iki kümede de ortak olanları tespit edilmiş ve elde edilen yeni ilişki sayıları Tablo 28’de verilmiştir.

Tablo 28: Ortak sözcük ikilileri için anlamsal ilişki sayıları

İlişki Türü	İlişki Sayıları		
	OÇİ	TWN	Ortak
Alt – üst kavram	292	796	248
Eş anlamlılık	3342	3136	3094
Karşıt anlamlılık	48	56	40
Toplam	3682	3988	3382

Tablo 28’de yer alan veriye göre otomatik çıkartılan ilişkiler (OÇİ) kümesindeki 3.682 ilişkide bulunan (*sözcük A, sözcük B*) biçimindeki sözcük ikilisi aynı zamanda Türkçe WordNet’te (TWN) de yer almaktadır. Benzer biçimde TWN’de yer alan anlamsal ilişkilerin 3.988 adetinde yer alan sözcük ikilileri OÇİ’de de ortak olarak yer almaktadır. Ortak sözcük ikililerini barındıran anlamsal ilişki sayıları arasındaki bu farkın nedeni TWN’nin aynı sözcük ikilileri arasında yer alabilecek birden fazla anlamsal ilişkiye izin vermesidir. Seçilen bu ortak sözcük ikilileri arasında yapılan karşılaştırmada ilişki türü de ortak olan ikililerin sayısı ise 3.382 olarak tespit edilmiştir. Bu sayı aynı zamanda iki kümenin (OÇİ ve TWN) bütününde ortak olarak yer alan (*sözcük A, ilişki türü, sözcük B*) üçlülerinin sayısıdır.

Türkçe WordNet’te yer alan ilişkiler (TWN) altın standardında kabul edilip referans alınarak otomatik çıkartılan ilişkiler (OÇİ) değerlendirilecek olursa, Tablo 28’de verilen ortak sözcük ikilileri arasındaki ilişkilere dair karışıklık matrisi (*confusion matrix*) Tablo 29’daki gibi olur.

Tablo 29: Türkçe WordNet'te yer alan ilişkiler referans alınarak oluşturulan karışıklık matrisi

		Öngörülen İlişki (OÇİ)			Toplam (TWN)
		Alt – üst kavram	Eş anlamlılık	Karşıt anlamlılık	
Doğru İlişki (TWN)	Alt – üst kavram	248	234	0	482
	Eş anlamlılık	42	3094	8	3144
	Karşıt anlamlılık	2	14	40	56
Toplam (OÇİ)		292	3342	48	

Tablo 29'da sunulan karışıklık matrisi oluşturulurken belli bir sözcük ikilisi arasında yalnızca bir tür ilişki bulunabileceği varsayılmıştır. Oysa karşılaştırmada referans olarak alınan TWN ilişki kümesinde aynı sözcük ikilisi arasında birden fazla ilişki yer alabilmektedir. Bu durumun karışıklık matrisine yansıyan en önemli sonucu Tablo 29'daki doğru ilişkilerin toplamını gösteren sütun ile Tablo 28'deki TWN sütununun uyuşmamasıdır. Eş anlamlılık ve karşıt anlamlılık ilişki türleri için belirgin olmasa da alt – üst kavram ilişki türü için belirgin bir sayısal fark olduğu göze çarpmaktadır. Bu durumun özünde yatan neden, TWN ilişki kümesinde eş anlamlılık ve alt – üst kavram ilişki türleri arasında sıkı bir ayrıma gidilmemiş olmasıdır. Ortak sözcük ikilileri arasındaki ilişki sayılarının gösterildiği Tablo 28'de 796 olarak verilen TWN kümesindeki alt – üst kavram ilişkilerinin 340 tanesi TWN tarafından aynı zamanda eş anlamlılık ilişkisi olarak da gösterilmiştir.

TWN ilişki kümesinde kimi ilişkiler arasında sıkı bir ayrıma gidilmemiş olmasının Tablo 29'daki karışıklık matrisi üzerindeki bir başka yanıltıcı sonucu da referans kümede (TWN) alt – üst kavram ilişkisi olarak gösterildiği halde OÇİ kümesini oluşturan modelce eş anlamlılık ilişkisi olarak öngörülen ilişki sayılarındaki fazlalıktır. Öngörülen model tarafından doğru sınıflandırılan alt – üst kavram türündeki ilişkilerin sayısı 248 iken referans kümede alt – üst kavram olarak belirtildiği halde model tarafından eş anlamlılık olarak sınıflandırılan ilişkilerin sayısı 234'tür. Durum böyle olmasının altında yatan temel nedenler yine karışıklık matrisi oluşturulurken referans kümede bir sözcük ikilisi arasında tek bir ilişki türü olabileceği varsayımı ile alt – üst kavram ve eş anlamlılık ilişki türleri arasında referans küme tarafından sıkı bir ayrıma gidilmemiş olmasıdır. Öngörülen ilişki kümesinde anlam ayrımlarının (*disambiguation*) dikkate alınmaması bu duruma katkıda bulursa da sayılan nedenlerin dışında TWN kümesinin önemli bir kısmının İngilizce dilinden çeviri olması [8] ve insan hatası gibi unsurların da bu durum üzerinde etkili olduğu düşünülmektedir. Örneğin değerlendirmede kullanılan TWN veritabanında *{besin, gıda}* eş kümesi tekrarlı olarak yer almakta olup eş kümelerden birinin diğerinin üst kavramı olduğu bilgisine yer verilmiştir. Bu durum, *{besin, eş anlamlılık, gıda}* ve *{gıda, eş anlamlılık, besin}* üçlüleriyle temsil edilen ilişkilerin referans kümede tekrarlı olarak yer almasından başka gerçekte var olmayan ilişkilerin de referans kümede yer almasına neden olmuştur. Alt kavram – üst

kavram ilişki türü δ ile ifade edilmek üzere *{besin, δ , gıda}*, *{besin, δ , besin}*, *{gıda, δ , gıda}* ve *{gıda, δ , besin}* üçlüleriyle temsil edilen ilişkiler söz konusu örnek eş küme için bu hatalı ilişkilere örnektir.

Sözlük verisinden faydalanılarak tamamen otomatik yöntemlerle belli ilişki türleri için anlamsal bilginin çıkarılabilir olduğunu göstermek, bu çalışmanın hedeflenen sonuçları arasındadır. Çalışma sonuçları farklı açılardan değerlendirilmiş ve Türkçe WordNet ilişki kümesiyle mukayese edilmiştir. Çalışmanın başlıca eksik yanı, kurallar yardımıyla anlamsal ilişki çıkarımında sözcük birlikteliklerinin ve anlam ayrımlarının dikkate alınmamasıdır. Farklı sözlükler için farklı kurallar çıkarma gerekliliği de bu yöntemin uygulanmasını zorlaştıran bir dezavantaj olarak sayılabilir.

Gelecekte hedeflenen çalışmalar arasında, kurallar yardımıyla anlamsal ilişki çıkarımında sözcük birlikteliklerinin ve anlam ayrımlarının dikkate alınması ile çıkartılan anlamsal ilişkilerin sayı ve güvenilirliklerinin artırılmasına yönelik çalışmalar sayılabilir.

Yapılan çalışma temelde şu anda sürdürülen Türkçe Hayat Bilgisi Veritabanı oluşturulması projesinin [20] destekleyici parçalarından birisidir. Bu çalışmada elde edilen ilişkiler ve güvenilirlik sayıları Hayat Bilgisi veri tabanının güncellenmesinde kullanılacaktır. Hayat Bilgisi veri tabanının arayüzüne <http://www.kemikoyun.yildiz.edu.tr/commonsense/> adresinden erişilebilir.

Kaynaklar

- [1] Chodorow, M. S. ve Byrd, R. J., "Extracting Semantic Hierarchies from a Large On-line Dictionary", *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics*, 1984, 299-304.
- [2] <http://www.tdkterim.gov.tr/bts/>
- [3] <http://tr.wiktionary.org/>
- [4] Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D. ve Miller, K., "Introduction to WordNet: An On-line Lexical Database", 1993.
- [5] Lenat, D. B., "Cyc: A Large-Scale Investment in Knowledge Infrastructure", *The Communications of the ACM*, 38(11):33-38, 1995.
- [6] Singh P., Lin, T., Mueller, E. T., Lim, G., Perkins T. ve Zhu, W. L., "Open Mind Common Sense: Knowledge acquisition from the general public", *Proceedings of the First International Conference on Ontologies, Databases, and Applications of Semantics for Large Scale Information Systems*, Irvine, CA, 2002.
- [7] Liu, H. ve Singh, P., "ConceptNet: A Practical Commonsense Reasoning Toolkit", *BT Technology Journal*, (22), Kluwer Academic Publishers, 2004.
- [8] Bilgin O., Çetinoğlu Ö. ve Oflazer K., "Building a Wordnet for Turkish", *Romanian Journal of Information Science and Technology*, cilt 7, 1-2, 2004.
- [9] Amasyalı M. F., "Türkçe Wordnet'in Otomatik Olarak Oluşturulması", *SIU 2005*, 2005.
- [10] Güngör, O. ve Güngör, T., "Türkçe için Bilgisayarla İşlenebilir Sözlük Kullanarak Kavramlar Arasındaki İlişkilerin Belirlenmesi", *Akademik Bilişim Konferansı*, 2007.

- [11] <https://zemberek.dev.java.net/>
- [12] Roche, E. ve Schabes Y., "Finite-State Language Processing", The MIT Press, Cambridge, 1997.
- [13] <http://www.pcre.org/pcre.txt>
- [14] Scott, S. ve Matwin, S., "Text Classification using WordNet Hypernyms", *Workshop on Usage of WordNet in Natural Language Processing Systems*, Canada, 1998.
- [15] Navigli, R. ve Velardi, P., "An analysis of ontology-based query expansion strategies", *Proceedings of the 14th European Conference on Machine Learning*, 2003.
- [16] Fong, S., "Semantic Opposition and WordNet", *Journal of Logic, Language and Information*, 13:159–171, 2004.
- [17] Croft, W. ve Cruse, D. A., "Cognitive linguistics", Cambridge University Press, Cambridge, 2004.
- [18] http://tr.wiktionary.org/wiki/Yardıml:Madde_şablonu
- [19] Miller, G. A., "WordNet: A Lexical Database for English", *Communications of ACM(11)*, cilt 38, 39-41, 1995.
- [20] Amasyalı, M. F., İnak, B. ve Ersen, M. Z., "Türkçe Hayat Bilgisi Veri Tabanının Oluşturulması", *Akademik Bilişim Konferansı*, 2010.