

Emlak Alanına Özgü Kelime Ağı

Domain Dependent Wordnet for Real Estate

Selen Parlar^{1,2}, Bilge Nas Arıcan², Mehmet Erkek³, Kamil Çayırılı³, Olcay Taner Yıldız⁴

¹Boğaziçi Üniversitesi Bilgisayar Mühendisliği Bölümü, İstanbul, Türkiye

²Starlang Yazılım Danışmanlık, İstanbul, Türkiye

³Zingat Gayrimenkul Bilgi Sistemleri, İstanbul, Türkiye

⁴Işık Üniversitesi Bilgisayar Mühendisliği Bölümü, İstanbul, Türkiye

selen.parlar@boun.edu.tr, bnarican@gmail.com, {mehmet.erkek, kamil.cayirli}@zingat.com, olcaytaner@isikun.edu.tr

Özetçe—Kelime ağı, anlamlarına göre organize edilmiş kelimeleri barındıran bir veritabanıdır. Bir kelime ağı, sahip olduğu kelimelerin anlamlarını, bilişsel eş anlamlılarını, türlerini, diğer anlamlar ile arasındaki ilişkilerini ve bu anlamların tanımlarını temsil eder. Bu çalışma ile, emlak alanına özgü bir sözlük oluşturmak ve bu yeni sözlüğü kullanarak daha küçük bir kelime ağı tasarlamak yoluyla biçimbilimsel çözümleme ve anlam belirsizliği giderme gibi Doğal Dil İşleme görevlerini kolaylaştıracak bir yöntem öneriyoruz. Ön çalışma olarak, emlak alanına özgü 7.000 kelime içeren bir sözlük ve yaklaşık 11.000 eş anlam kümesinden oluşan bir kelime ağı oluşturuldu ve bunlar çeşitli görevlerle doğrulandı.

Anahtar Kelimeler—Kelime ağı, Emlak

Abstract—WordNet is a thesaurus which represents word senses, meanings of lemmas, relations between senses, and definitions of these senses. In this paper, we propose a method to simplify the morphological analysis and disambiguation tasks performed in a specific domain by creating a subset of original Turkish dictionary and WordNet. As a preliminary study, we create a dictionary for the estate domain with 7,000 words and a WordNet with 11,000 synsets, then validate them on simple tasks.

Keywords—WordNet, Real Estate

I. GİRİŞ

Kelime ağı (WordNet), düğüm (node) ve bu düğümler arasındaki bağlantılardan (edge) oluşan çizge (graph) yapısına sahip, farklı kavramlar arasındaki anlamsal ilişkileri temsil eden bir ağdır (network). Bir düğüm, bir kavramı, yani, ortak özelliklere veya özelliklere sahip bazı üyelerin bir sınıfını temsil ederken, düğümler arasındaki bir bağlantı, kavramlar arasındaki ilişkileri temsil eder. Bir anlamsal ağ olan kelime ağı, kelime anlamı ve kök kelime belirsizliği giderme, bilgi geri döndürümü, makine çevirisi ve duygu analizi gibi çeşitli amaçlar için kullanılabilir, ancak kelime ağının kullanmanın temel amacı, kelimeler arasındaki benzerlikler ve ilişkileri elde etmektir. Örneğin, bir programın *Antalya'da kiralık daireler* cümlesini anlayabilmesi için *Antalya* kelimesinin Türkiye'nin Akdeniz Bölgesi'nde yer alan illerinden biri olduğu, *kiralık* kelimesinin kiraya verilecek olan anlamına gelen bir sıfat olduğu, ve *daire* kelimesinin de *kat* kelimesinin eş anlamlısı olduğu gibi bilgileri edinmesi için kelime ağı kullanımını kaçınılmazdır.

Kapsamlı bir kelime ağı oluşturabilmek için öncesinde oluşturulmuş bir sözlüğe ihtiyaç vardır. Ancak, kelime türetiminde herhangi bir üst sınır bulunmayan eklemeli diller grubuna dahil olan Türkçe için sözlükte bulunan kelimelerin sayısında da bir üst sınır olmayacaktır. Bunun yanı sıra, kelimelerin çok anlamlılığa sahip olması çok sayıda anlamsal tutarsızlığa sebebiyet verir. Bu nedenle, bir dildeki tüm kelimeleri ve anlamları bir araya getirmek sanıldığı kadar basit bir iş değildir. Örneğin, bir Türkçe sözlük 50.000'den fazla kelime içerir ancak bu fazlasıyla geniş sözlüğü sadece belirli ve dar bir alan için kullanmak oldukça belirsiz ve alan ile alakası bulunmayan sonuçlar verebilirken, genel kullanımda bulunmayan ve sadece o alana özgü bir anlama sahip olan kelimeler de dışarıda kalabilir.

Bu çalışmada, yukarıda belirtilen hususları ele almak amacıyla *alana özgü kelime ağı ve Türkçe sözlüğü oluşturma* yaklaşımı önerilmiştir ve çalışma alanı olarak da **emlak** alanı seçilmiştir. Çalışmada, bu alanla ilgili kullanıcı değerlendirmelerinden ve tercihlerinden derlenen veriler kullanılarak, yalnızca yaygın olarak kullanılan kelimeleri değil, aynı zamanda sadece o alana özgü anlamları barındıran kelimeleri de içeren bir Türkçe sözlük oluşturulmuştur. Emlak alanı için oluşturulan sözlük 7.000 kelimedenden oluşmakta ve halihazırda kullanılan Türkçe sözlüklerle karşılaştırıldığında, yeni oluşturulan sözlüğün diğerlerine oranla yaklaşık 10 katı kadar daha az kelime barındırdığı görülmektedir. Çalışma sırasında da, alana özgü kelime ağı oluşturulurken bu küçük boyutlu sözlükten yararlanılmıştır. Makale şu şekilde düzenlenmiştir: Öncelikle, Bölüm II'de Türkçe hakkında kısa bir inceleme yapılmış ve Bölüm III'de kelime ağı literatüründen bahsedilmiştir. Bölüm IV'de, alana bağlı kelime ağının nasıl oluşturulduğu açıklanmış, ve Bölüm V'de yeni oluşturulan kelime ağı hakkındaki istatistikler ve deney sonuçları verilmiştir. Son olarak, çalışma Bölüm VI'da sonuçlandırılmıştır.

II. TÜRKÇE

A. Dil

Türkçe, yapısı bakımından sondan eklemeli diller grubuna aittir ve bu dillerde, yeni kelimeler ve terimler türetmek oldukça kolaydır. Dilde en yaygın kullanılan cümle yapısı, özne-nesne-fil biçimindedir. Ancak, Türkçenin esnek bir dil

olmasından dolayı devrik cümlelere günlük yaşamda sık rastlanır. Türkçede, çekim ekleriyle yapılan değişiklikler kök üzerinde yapılır ve bu sırada kök herhangi bir değişikliğe uğramaz. Yapım ekleriyle yapılan değişiklikler, tek bir kökten düzinelere yeni kelime türetilmesine izin verir. Cümleyi oluşturan unsurların ilgisine ilişkin olarak, cümle, düşünce geliştirmek için değil, tamamlanmış düşüncenin doğal bir hiyerarşisi olarak kurulur.

B. Biçimbilimsel Çözümleme

Sona veya başa ek ekleme, yazım veya fonetikte herhangi bir değişiklik yapmadan bileşik sözcükler oluşturmak için biçimribimlerin (morpheme) birleştirilmesiyle yapılan dilbilimsel bir işlemdir. Ural-Altay dil ailesine ait olan Türkçe, sondan eklemeli bir dildir; tek bir kökten sınırsız sayıda kelime formu elde edilebilir ve çok zengin bir biçimbilime (morphology) sahiptir. Türkçe gibi türetimsel ve çekimsel değişikliklerle yeni bir kelime üretmenin kolay olduğu üretken dillerde, kelimenin gerçek kökünü bulmak için biçimbilimsel ayrıştırma (morphological parsing) ve biçimbilimsel çözümleme (morphological analysis) işlemlerinin yapılması önem arz eder. Biçimbilimsel ayrıştırma, bir cümleyi veya bir sözcüğü bileşenlerine ayırmayı ve sözdizimsel (syntactic) rollerini tanımlamayı amaçlar. Örneğin, *kitaplar* kelimesi, biçimbirimlerine *kitap* ve *-lar* olarak ayrıştırılır ve de çoğul eki almış bir sıfat olduğu çıktı olarak verilir.

Eklemeli diller için biçimbilimsel ayrıştırmanın çözümlemesi, kelime başına düşen biçimbirim oranı çok düşük olan tek heceli dillerden daha karmaşıktır. Kelime oluşturma işlemi, zengin sonekle türetme süreci nedeniyle oldukça verimlidir ve verimlilik bu diller için büyük sözlüklerin ortaya çıkmasına neden olur. Örneğin; internette bulunan gazetelerden derlenen bir Türkçe derlemde yaklaşık 10 milyon kelime bulunurken bir İngilizce derlemde bunun yaklaşık dörtte biri oranında kelime bulunur [5].

Sözlüğün büyüklüğü arttıkça, olası biçimbilimsel çözümlemelerin sayısı da artar. Aynı kelime için birden fazla biçimbilimsel çözümlemenin olması belirsizlikler meydana getirir ve bu belirsizlik olası sözcük etiketleri arasından en doğru ayrıştırmanın seçilmesini amaçlayan biçimbilimsel belirsizlik giderme (morphological disambiguation) işlemini olumsuz yönde etkiler. Örnek olarak *Turistler bu semte akın ettiler.* cümlesinde bulunan *akın* kelimesinin birden fazla biçimbilimsel çözümlemeye sahip olduğu Tablo I'de gösterilmiştir. Geniş kapsamlı sözlükleri sınırlı ve dar bir alanda yapılacak olan biçimbilimsel çözümlemeler için kullanmak biçimbilimsel belirsizlik giderme problemini zorlaştırır. Tablo I'de de görüldüğü gibi *akın* kelimesinin kökü isim, fiil ve sıfat olabilir, lakin kullanıldığı cümle emlak alanına göre değerlendirildiğinde kökün fiil olarak kullanılma ihtimali isim olarak kullanılma ihtimalinden daha yüksektir. **Sonuç olarak, bir sözlükteki kullanılma ihtimalleri düşük olan anlamları ortadan kaldırmak veya hedef etki alanındaki ilişkilerle dayanarak yeni bir sözlük oluşturmak biçimbilimsel belirsizlik gidermek için gereken işlem süresini azaltır ve DDİ sürecini olumlu yönde etkiler.**

TABLO I: *akın* kelimesinin olası biçimbilimsel çözümlemesi

Kök	Ekler
<i>ak</i>	+ADJ^DB+NOUN+ZERO+A3SG+P2SG+NOM
<i>ak</i>	+ADJ^DB+NOUN+ZERO+A3SG+PNON+GEN
<i>ak</i>	+VERB+POS+IMP+A2PL
<i>akı</i>	+NOUN+A3SG+P2SG+NOM
<i>akın</i>	+NOUN+A3SG+PNON+NOM

III. LİTERATÜR ÖZETİ

Kelime ağı, düğümleri, kelime manaları ve bu manalara karşılık gelen literallerden oluşan, bağıntıları da mana çiftleri arasındaki anlambilim (semantic) ilişkilerinden meydana gelen bir çizge yapısıdır. Bir kelime ağı oluşturmak yoğun emek gerektiren bir iştir. Kelime ağına açıklama eklemek, farklı alanlardaki kelimelerin kullanım modlarına aşına olmanın yanı sıra, sözlük yetkinliğini de gerektirir.

İlk kelime ağı projesi 1995 yılında George Miller tarafından başlatılan Princeton WordNet (PWN) projesidir [7]. Şu anda, PWN'nin son sürümü olan 3.1 sürümü 117.000 adet eş anlam kümesi (synset), 206.941 adet ise kelime-anlam çiftine sahiptir. PWN'nin daha ayrıntılı bir tarihi ve açıklaması [4]'de verilmiştir. PWN'nin piyasaya sürülmesinden kısa bir süre sonra, diğer diller için kelime ağı sistemleri oluşturulmuştur. Bu kelime ağı sistemlerinin kapsamı PWN'nin kapsamı kadar büyük değildir [8], [2].

Balkan dilleri için oluşturulan BalkaNet [9] bugüne kadarki en kapsamlı çalışmadır. BalkaNet'in Türkçe kelime ağı kısmı için [1], araştırmacılar tek dilli bir Türkçe sözlükten eş anlamlıları (synonym), zıt anlamlıları (antonym) ve alt anlamlıları (hyponym) otomatik olarak çıkarmışlardır. En kapsamlı Türkçe kelime ağı olan KeNet 110.000 kelime-anlam çiftini kapsayan 80.000 adet eş anlam kümesine sahiptir [3].

IV. ALANA ÖZGÜ KELİME AĞI

A. Ön İşleme

Daha önce de belirtildiği gibi, çalışmada emlak alanına ait veriler kullanılmıştır. Veriler, incelemeleri veya tercihleri yoluyla doğrudan kullanıcılardan alınmıştır. Kullanıcılar yorum yaparken yalnızca cümlelerin anlamlarına odaklandığı ve bunu yaparken de genellikle dil bilgisine dikkat etmeyen günlük bir dil kullandığı için, orijinal girdilere dayanarak doğrudan doğal dil işleme (DDİ) adınımlarını gerçekleştirmek çok uygulanabilir değildir. Bu alt bölümde, veriyi DDİ işlemleri uygulanabilecek hale getirmek için yapılan ön işleme süreci detaylandırılmıştır.

Çalışmaya paragrafları cümlelere bölme işlemiyle başlanmıştır. Cümlelerdeki noktalama işaretlerinin çıkartılmasının ardından, kelimelere bölünüp, küçük harfe dönüştürülmüştür. Sonrasında, biçimbilimsel çözümleme programı kullanılarak kelimelerin kök halleri elde edilmiştir. Sonuç olarak her biri birbirinden farklı kelime birimleri elde edilmiştir.

Verilerin kullanıcılar tarafından doğru bir şekilde yazılması, sözdizimlerinin düzeltilmesi için ayrıca işlem gerektirir. Kullanıcılar genellikle yorumlarında Türkçe karakter kullanımına özen göstermemektedirler. Bunun için ilk olarak

Türkçe karaktere dönüştürme (asciify/deasciify) işlemi gerçekleştirilmiş ve hatalı yazılmış Türkçe karakterler doğru yazımlarına dönüştürülmüştür.

İşlemler sonucu, eğer bir kelime hala biçimbilimsel olarak çözümlenemez durumdaysa, kelimelerin yanlış yazıldığı varsayılarak, ayrı yazılması gerekirken bitişik yazılmış olan eklerin ayrılması, ya da bitişik yazılması gerekirken ayrı yazılmış kelimelerin bitleştirilmesi gibi kural tabanlı yöntemlerle düzeltilen kelimeler tekrar analiz edilmeye çalışılmıştır. İşlemler sonucunda elde edilen kelime hala analiz edilemiyorsa halihazırda sözlüğümüzde bulunan kelimeler ile olan Levenshtein mesafesine ve Damerau Levenshtein mesafesine bakılarak kendisine en benzer kelime olarak düzeltilmiştir.

B. Sözlük

Yeni bir kelime ağı oluşturulması adına ilk olarak yeni bir Emlak sözlüğü hazırlanmıştır. Bir emlak şirketinin internet ortamında bulunan sistemindeki kelimeler ve yorumlar kullanılarak, en çok kullanılan kelimelerle IV-A'da bahsi geçen işlemlerin ardından, alana özgü bir sözlük meydana getirilmiştir. Emlak sözlüğü genellikle ilçe isimleri, mahalle isimleri, ve proje adları içermektedir. Sözlükteki kelimeler yapılarına göre etiketlenmiştir ve bu sayede, kelimeler özel isim, isim, fiil, sıfat, ve zarf olmalarına göre ayrılmıştır.

C. Kelime Ağı

Kelime ağı, DDİ işlemlerinde önemli role sahip büyük bir sözlüksel veritabanıdır. Kelime ağında özel isimler, cins isimler, fiiller, sıfatlar ve zarflar gruplandırılarak alt-üst anlamlarıyla ilişkilendirilirler. Bu çalışma temel olarak alana özgü kelime ağı oluşturulurken ortaya çıkan sözlüğe dayanmaktadır. Emlak alanına özgü kelime ağı için üç kategori belirlenmiştir; özel isimler, cins isimler ve fiiller (Örnek kelimeler Tablo II'de gösterilmiştir). Özel isim kategorisindeki kelimelerin büyük oranda inşaat projesi, emlakçı, il, ilçe ve mahalle isimlerinden oluştuğu görülmektedir. Kelimeler, bulunduğu alana göre gruplandırılmış, insan ismi ve soyisimlerinin çalışmada kullanılmamasına karar verilmiştir.

TABLO II: Emlak kelime ağından örnekler

Kelime	Instance Hypernym
Academy	Emlakçı
Cordelia	İnşaat Projesi
Bornova	İlçe
Ankara	İl
İtalya	Ülke

İsim kategorisindeki kelimelerle çalışılmaya başlanmış ve bu çalışmanın yaklaşık yarısı tamamlanmıştır. Bu kategorideki kelimelerin alt, üst ve eş anlamları önem kazanmıştır. Kelimeler şu ilişkilere göre analiz edilmiştir: Karşıtlık (antonymy), üyelik bildiren bütün-parça ilişkisi (member holonymy), madde bildiren bütün-parça ilişkisi (substance holonymy), kısımlık bildiren bütün-parça ilişkisi (part holonymy), ilişkili konu (domain topic), ve özellik (attribute).

V. İSTATİSTİKLER

A. Veri Kümesi

Kelime ağı ve sözlük oluşturmak için çok fazla veri kullanımını gerektirmektedir, daha önce bahsedildiği gibi bu çalışmada emlak hizmetleri veren bir firmanın internet ortamında bulunan kullanıcı incelemeleri ve kullanıcı tercihlerinden veriler toplanarak 7,000 kelimelik bir sözlük oluşturulmuştur.

```

{"event_name": "FilterWithParameters",
 "event_params": [
  {"key": "listingType", "value": {
    "string_value": "Satılık"}},
  {"key": "keyword", "value": {
    "string_value": "samsun da uygun fiyatli satilik evler",
    {"key": "propertyType", "value": {
      "string_value": "Konut"}}
  ]}]

```

Şekil 1: JSON veri yapısında örnek kullanıcı incelemesi

Genellikle, kullanıcılar internet ortamındaki anketlerde metinleri yazarken dil bilgisi kurallarına uygun bir dil kullanmamaktadırlar. Bu nedenle, veri ham halde alınıp, Bölüm IV-A'da açıklanan ön işleme adımları uygulandıktan sonra iyileştirilmiş cümleler işleme alınmaktadır. Örneğin; Şekil 1'in anahtar kelime (keyword) bölümünün değeri "samsun da uygun fiyatli satilik evler" olarak bulunmaktadır ve bu hatalı cümle ön işleme sonucunda "Samsun'da uygun fiyatli satilik evler." olarak düzeltilmektedir.

Veriler toplandıktan ve ön işlem den geçirildikten sonra, biçimbilimsel çözümlenmesi için her bir verinin sözcük etiketleri tanımlanmaktadır. Örneğin, Türkiye'de bir şehri ifade eden *Samsun* kelimesi bir özel isimdir ve sözcük etiketi sözlükte "IS_OA" olarak gösterilmektedir; benzer şekilde, *ev* kelimesi bir cins isimdir ve sözlükte "CL_ISIM" olarak ifade edilmektedir. Tablo III, emlak ve Türkçe sözlüklerinde sık kullanılan dört sözcük etiketlerinin yüzdelerini göstermektedir; sırasıyla bu etiketler IS_OA (özel isim), CL_ISIM (cins isim), CL_FIIL (fiil) ve IS_ADJ (sıfat)'dir.

TABLO III: Sık kullanılan sözcük etiketleri yüzdesi

	Emlak%	Türkçe%
IS_OA	83.89	36.87
CL_ISIM	12.48	53.07
CL_FIIL	1.18	8.353
IS_ADJ	5.01	7.38

Tablo III'te görüldüğü gibi, Emlak sözlüğünde bulunan kelimeler içerisinde özel isim yüzdesi Türkçe sözlüğüne nazaran büyük çoğunluğu oluşturmaktadır. Bunun temel sebebi, Emlak sözlüğünün çok fazla miktarda şehir, ülke, mahalle, proje ismi gibi sözcükleri barındırmasıdır. Bunun yanı sıra, Emlak sözlüğünde fiillerin sayısı oldukça azdır çünkü kullanılan veri kullanıcıların bir arama yaparken kullandığı cümlelerden veya yorumlardan oluşmaktadır ve kullanıcılar da genellikle yaptıkları aramalarda fiil kullanmamakta sadece anlamlı bir sorgu yazmayı amaçlamaktadırlar. Kullanıcılar metinlerde genellikle günlük dil kullanmayı tercih ettiklerinden, Emlak sözlüğü ile Türkçe sözlüğün pek çok ortak kelimesi bulunmaktadır. Örneğin, Emlak sözlüğündeki kelimelerin %42,5'i Türkçe sözlüğüyle ortaktır. Tablo IV, Emlak-Türkçe sözlüklerinin ortak kelimelerinin sözcük etiketlerinin yüzdesini göstermektedir.

TABLO IV: Emlak ve Türkçe sözlüklerindeki ortak kelimelerin sözcük etiketlerinin yüzdeleri

	Ortak Kelime Etiketleri%
IS_OA	73.77
CL_ISIM	22.83
CL_FIIL	2.76
IS_ADJ	6.88

B. Sonuçlar

Çalışmada, Bölüm V’te açıklanan veri kümesini kullanarak bir sözlük ve kelime ağı oluşturulmuş ve yeni oluşturulan alana özgü sözlük, kelime ağı ve genel Türkçe sözlük kullanılarak birtakım analizler yapılmıştır. Başlangıç olarak, veriyi derinlemesine anlamak için yeni oluşturulan kelime ağından üst anlam (hypernym) ilişkisi çıkarılmıştır. Tablo V, Emlak alanında en sık geçen 5 üst anlamı göstermektedir. Tahmin edilebileceği gibi, Emlak sözlüğü büyük oranda inşaat projesi ve gayrimenkul isimlerinden oluşmaktadır.

TABLO V: Emlak alanında en sık geçen 5 üst anlam oranları

Üst Anlam	%
<i>Inşaat Projesi</i>	18.77
<i>Emlakçı</i>	14.85
<i>Mahalle</i>	14.55
<i>İlçe</i>	8.15
<i>Inşaat Şirketi</i>	3.69

Emlak alanı için oluşturulan sözlüğü doğrulamak amacıyla bu alana ait, ve öncesinde analizi yapılmamış kullanıcı yorumlarından derlenen bir veri kümesi üzerinde test edilmiş ve sonuçları genel Türkçe sözlük sonuçlarıyla karşılaştırılmıştır. Tablo VI, Emlak ve Türkçe sözlükleri anahtar kelimeye ve kelimeye dayalı analizlerin sonuçlarını göstermektedir. Anahtar kelime tabanlı analizde, 40.000 farklı anahtar kelime üzerinde biçimbilimsel çözümleme yapılarak, kelime temelli analizde de, her bir 116.786 sözcük için ayrı ayrı biçimbilimsel çözümleme yapılarak sözlüklerin performansları kontrol edilmiştir. Özetlemek gerekirse, Emlak alanı için alana özgü bir sözlük oluşturmak, yapılan cümle temelli analiz miktarını %23,49 ve kelime temelli analiz miktarını %8,63 oranında arttırmıştır.

TABLO VI: Emlak ve Türkçe sözlükleri ile analiz edilebilen kelimelerin ve anahtar kelimelerin oranları

Sözlük	Büyüklik	Anahtar Kelime %	Kelime %
Emlak	7,000	97.69	99.17
Türkçe	51,552	74.20	90.54

Son olarak, anahtar kelimeler, emlak veri setinde de belirtilen kullanıcı tercihi bilgileriyle, yani, kullanıcı tarafından

Herhangi bir kelime için çoklu biçimbilimsel çözümleme yapmak, belirsizlik problemini ortaya koymaktadır. Yaklaşımımızla, söz konusu belirsizlik sorununu sözlük boyutlarını azaltarak çözmeyi hedeflemekteyiz. Bu amaçla, sadece belirlenen alanla ilgili kelime anlamları kullanılmış, onların dışında kalan anlamlara sözlükte yer verilmemiştir. Performansı incelemek amacıyla, sadece bir tane biçimbilimsel çözümlemesi olan kelimelerin yüzdesi hesaplanmıştır. Tablo VII, emlak alanı için %21’lik gelişmeyi göstermektedir. Sonuç olarak, sözlük boyutunu küçültmek, belirsizlik giderme sorununa yardımcı olmaktadır.

TABLO VII: Emlak ve Türkçe sözlükleri kullanıldığında oluşan 1 elemanlı biçimbilimsel çözümleme oranları

Sözlük	Biçimbilimsel Çözümleme %
Emlak	87.41
Türkçe	66.37

dan bir menüden seçilen seçeneklerle karşılaştırılmıştır. Tablo VIII’de görüldüğü gibi, Emlak sözlüğü ile kullanıcı tercihlerinin yaklaşık %30’u yakalanabilmiştir. Örneğin, Şekil 1’de kullanıcı tarafından anahtar kelime olarak *Samsun’da uygun fiyatlı satılık daire* yazılmış ve tercih olarak da *Satılık* seçeneği seçilmiş, bu sebeple *Satılık* kelimesi her iki dizide de eşleşmiştir. Ancak, *Samsun* kelimesi seçilmediğinde, sistem şehir ismi hakkında herhangi bir bilgi alamamaktadır. Gelecek çalışma planı olarak, kullanıcılar tarafından seçilmeyen seçenekleri bulmak için sisteme kelimelerin üst anlam ilişkilerini eklemeyi planlamaktayız.

TABLO VIII: Emlak sözlüğünde anahtar kelime-seçim uyum oranı

Anahtar Kelime	Emlak %
Şehir İsmi	13.05
Listeleme tipi	9.88
Bölge Adı	3.73
Emlak Tipi	1.74

VI. SONUÇ

Sonuç olarak, çalışmamızda Emlak alanındaki kullanıcı incelemeleri ve tercihleri kullanılarak alana özgü bir sözlük oluşturulmuş, bu oluşturulan sözlük ile, yeni bir kelime ağı meydana getirilmiştir. Bunu yaparak, özellikle Emlak alanı için biçimbilimsel belirsizlik giderme sorununun azaltılması sağlanmış ve son olarak, emlak alanında cümle temelli biçimbilimsel çözümlemelerin performansı %23 oranında artırılmıştır.

KAYNAKLAR

- [1] O. Bilgin, Ö. Çetinoğlu, and K. Oflazer. 2004. Building a wordnet for turkish. *Romanian Journal of Information Science*, 7:163–172.
- [2] W. Black, S. Elkateb, H. Rodriguez, M. Alkhalifa, P. Vossen, A. Pease, and C. Fellbaum. 2006. Introducing the arabic wordnet project. In *International Wordnet Conference*, pages 295–300. Masaryck University, Brno, Czech Republic.
- [3] R. Ehsani, E. Solak, and O.T. Yildiz. 2018. Constructing a wordnet for turkish using manual and automatic annotation. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 17(3):24.
- [4] C. Fellbaum. 1998. *Wordnet: an electronic lexical database*. Cambridge. MIT Press, MA, USA.
- [5] D.Z. Hakkani-Tür, K. Oflazer, and G. Tür. 2002. Statistical morphological disambiguation for agglutinative languages. *Computers and the Humanities*, 36(4):381–410.
- [6] D. Jurafsky and J.H. Martin. 2014. *Speech and language processing*, volume 3. Pearson London.
- [7] G.A. Miller. 1995. Wordnet: a lexical database for english. *ACM Communications*, 38:39–41.
- [8] V. Peik. 1997. Eurowordnet: a multilingual database for information retrieval. In *DELOS workshop on Cross-language Information Retrieval*. Vrije Universiteit, Amsterdam, Czech Republic.
- [9] D. Tufis, D. Cristea, and S. Stamou. 2004. Balkanet: Aims, methods, results and perspectives. a general overview. *Romanian Journal of Information Science*, 7:9–43.