



Araştırma Makalesi

Journal of Innovative Engineering
and Natural Science

(Yenilikçi Mühendislik ve Doğa Bilimleri Dergisi)

<https://dergipark.org.tr/en/pub/jieng>

Türkçe doğal dil işleme: ses bilgisi ve morfolojik analiz

ID Mehmet Fatih Karaca^{a,*} ID Şafak Bayır^b^aTokat Gaziosmanpaşa Üniversitesi, Erbaa Sosyal ve Beşeri Bilimler Fakültesi, Yönetim Bilişim Sistemleri Bölümü, 60500, Erbaa, Tokat, Türkiye.^bKarabük Üniversitesi, Edebiyat Fakültesi, Eğitim Bilimleri Bölümü, 78050 Karabük, Türkiye.

MAKALE BİLGİSİ

Makale Geçmişi:

Geliş 23 Nisan 2024

Düzeltilme 14 Haziran 2024

Kabul 3 Temmuz 2024

Çevrimiçi mevcut

Anahtar Kelimeler:

Türkçe

Dil Bilgisi

Ses Bilgisi Analizi

Morfolojik Analiz

Doğal Dil İşleme

ÖZET

Bu çalışmada Türkçe Doğal Dil İşleme teknikleriyle 230 cümlelerin ses bilgisi ve morfolojik analizi gerçekleştirilmiştir. İşlemlerde Zemberek'le entegre şekilde çalışan *Visual Studio* ortamında ve *C#* dilinde geliştirilen yazılım kullanılmıştır. Zemberek'te bulunan 130 ekten 120'sinin Türkçe dil bilgisindeki ek karşılıkları belirlenmiştir. Zemberek eklerinden 13'ü kök adı, 46'sı çekim eki, 39'u yapım eki ve 22 tanesi diğer eklerdir (birleşik fiil, fiilden sıfat fiil ve fiilden zarf fiil türeten). Veri kümesindeki 230 cümlelerin 744 kelimededen oluştuğu, kelimelerin ortalama 5.30 harf ve 2.26 heceli olduğu, hecelerin yaklaşık yarısının ünsüz-ünlü birleşiminden oluşan CV hece türünde bulunduğu ve kelimelerin yarısından fazlası için Zemberek tarafından 1 çözümlene üretildiği belirlenmiştir. Ayrıca, kelimelerin yarısının ek almadığı, çekim eklerinin diğer ek türlerinden daha çok kullanıldığı, yapım eklerinden sonra yapım veya çekim eklerinin, diğer eklerden (birleşik fiil ve fiilden sıfat fiil türeten) veya çekim eklerinden sonra ise yalnızca çekim eklerinin geldiği tespit edilmiştir. İsim türündeki kelime sayısının diğer türlerdeki kelime sayısından, ünlü ve ünsüz uyum kurallarına uyan kelime sayısının uymayan kelime sayısından daha yüksek olduğu görülmüştür.

Turkish natural language processing: phonetic and morphological analysis

ARTICLE INFO

Article history:

Received 23 Apr 2024

Received in revised form 14 Jun 2024

Accepted 3 July 2024

Available online

Keywords:

Turkish

Grammar

Phonetic Analysis

Morphological Analysis

Natural Language Processing

ABSTRACT

In this study, phonological and morphological analysis of 230 sentences was carried out using Turkish Natural Language Processing techniques. Software developed in the *Visual Studio* environment and *C#* language, which works integrated with Zemberek, was used in the analyses. The Turkish grammar equivalents of 120 out of the 130 suffixes in Zemberek were determined. 13 suffixes are root nouns, 46 are inflectional suffixes, 39 are derivational suffixes, and 22 are other suffixes (compound verbs, deriving verbal adjective from verbs and verbal adverb from verbs) of the Zemberek suffixes. It was determined that the 230 sentences in the dataset consisted of 744 words, the words had an average of 5.30 letters and 2.26 syllables, approximately half of the syllables were in the CV syllable type consisting of consonant-vowel combinations, and 1 analysis was produced by Zemberek for more than half of the words. In addition, it was determined that half of the words do not take suffixes, inflectional suffixes are used more than other types of suffixes, derivational or inflectional suffixes come after derivational suffixes, and only inflectional suffixes come after other suffixes (compound verbs and deriving verbal adjective from verbs) or inflectional suffixes. It was observed that the number of words in the noun type is higher than the words in other types, and the number of words that comply with the vowel and consonant harmony rules is higher than those that do not.

I. GİRİŞ

Bilişim teknolojilerinde yaşanan gelişmeler ve bu teknolojilerin kullanımının yaygınlaşmasına paralel olarak dijital ortamdaki veri miktarı da artmaktadır. Buna karşın verilerin tamamı, yapısal forma sahip değildir. Yapısal olmayan doğal dillerde yazılmış metinsel verilerin bilgisayarlar tarafından işlenebilmesi için metinlerin DDİ teknikleri uygulanarak yapısal forma dönüştürülmesi gerekmektedir.

Yapay zekâ ve dil bilimin alt birimi olan Doğal Dil İşleme (DDİ), doğal dillerde yazılmış metinlerin yazıldığı dilin kuralları çerçevesinde bilgisayarlar tarafından çözümlenmesi, anlaşılması ve yeniden üretilmesi süreçlerini kapsar. İnsan bilgisayar etkileşiminin örneği olan ve günümüzde oldukça popüler hale gelen ChatGPT [1] veya Gemini [2] benzeri sohbet robotları DDİ'nin önemini ve görünürliğini daha da arttırmıştır. Fakat sohbet robotları dünya üzerinde konuşulan bütün dillerde hizmet vermemektedir. Her doğal dilin dil bilgisi denilen kendine has kuralları olması, birtakım benzerlikler olsa da bu kuralların dilden dile farklılık göstermesi, bir dil için uygulanan DDİ adımlarının başka dillerde farklı şekillerde uygulanamaması, kelime türetiminden söz dizimine kadar bütün işlemlerin dilin kuralları içerisinde gerçekleşmesi ve DDİ sürecinin zorluğu, bunun sebepleri olarak gösterilebilir.

DDİ, her ortamdaki metinsel içerikteki verilere uygulanacak yöntemleri ifade etmektedir. Yapılmak istenilen işe, veriye ve verinin kaynağına göre DDİ adımları farklılık göstermektedir. Örneğin, doküman sınıflandırmadaki işlemler ile yazar tanımadaki işlemler veya web ortamındaki metinsel veriye uygulanan DDİ adımları ile basılı bir dokümana uygulanan adımlar aynı değildir.

İnsan-bilgisayar etkileşiminin seviyesini belirleyen DDİ, yapay zekanın en komplike konularından biridir. Akıllı cihazlar veya web ortamındaki sanal asistan kavramıyla insanların hayatlarında yer almaya başlayan DDİ, bugün birçok uygulamada kendini göstermektedir. Metin sınıflandırma [3], yazar tanıma [4], duygu analizi [5], metin özetleme [6], metin üretme [7], çeviri sistemleri [8, 9], insan-bilgisayar etkileşimi [10], akıllı asistanlar [11] ve sohbet robotları [1, 2] DDİ'nin kullanım alanlarına örnek olarak gösterilebilir.

II. TÜRKÇENİN ÖZELLİKLERİ

Dünyada konuşulan 7000'i aşkın dil [12], kaynağı ve yapılarına göre sınıflandırılmaktadır. Türkçe, Ural-Altay dil grubunun Altay koluna bağlı eklemeli diller kategorisindedir [13]. Diğer dillerde olduğu gibi Türkçede gerçekleşen bütün işlemler dil bilgisi kurallarına göre icra edilmektedir. Konuşma dilleri arasında benzerlikler görülse de her dilin dinamikleri farklıdır. Doğal dillerle ilgili çalışmalarda çalışma yapılan dilin dil bilgisi kuralları bilinmeli, dilin kurallarına veya özellikle dilin çalışma yapılan kısmıyla ilgili kurallara hâkim olunmalıdır. Bu çalışmada Türkçenin ses bilgisi ve morfolojik analiz konularına odaklanılmıştır.

2.1 Ses Bilgisi Analizi

Bir dilin en küçük parçası olan ses tek başına anlamı olmayan, hece veya kelimeye dönüştüğünde anlam kazanan unsurlardır. Türkçede sesler, harfler ve ekler birleştirilirken ses olayları meydana gelmektedir. Bu olaylar fonetik veya ses bilimi olarak da ifade edilen DDİ'nin ses bilgisi alanında incelenmektedir. Seslerin ünlü/ünsüz uyumundan ek birleşimindeki ses düşmesine kadar olan bütün ses özellik ve olayları ses bilgisi analizinde incelenir.

Harfler, seslerin yazılı olarak temsili olan sembollerdir. Farklı dillerde farklı harflerin bulunduğu bilinmektedir. Örneğin; Türkçede İngilizceden farklı olarak ç, ğ, ı, ö, ş ve ü harfleri vardır. İngilizcede ise Türkçeden farklı olarak q, w ve x harfleri yer almaktadır. Türkçede 8 ünlü ve 21 ünsüz harften oluşan alfabe kullanılır. Esasında 8 ünlü harf bulunsada konuşmada bazı ara seslerin kullanıldığı da görülmektedir. Bununla birlikte bir dilde ünlü harf sayısına bağlı olarak dilin ahengi artar [14] ve Türkçe ünlü harf sayısı bakımından zengin bir dildir [14, 15].

Ünlü ve ünsüz harfler kendi içlerinde sınıflandırılmaktadır. Türkçede ünlü harfler meydana geliş noktasına göre kalın veya ince, açıklık/kapalılık durumuna göre geniş veya dar, dudakların durumuna göre düz veya yuvarlak, uzunluk/kısalık durumuna göre normal veya kısa ünlü olarak tasnif edilmektedir [15, 16]. Ünsüz harfler ise ses tellerinin durumuna göre sert veya yumuşak ünsüzler, boğumlanma noktasına göre dudak, diş, damak veya gırtlak ünsüzleri, ses yolunun durumuna göre sürekli veya süreksiz ünsüzler olarak sınıflandırılmaktadır [17].

Türkçede gerek köklerdeki gerekse kök ile ekler arasındaki seslerin bir araya gelmesi gelişigüzel olmamakta; bu işlemler kurallar çerçevesinde meydana gelmektedir. Ses uyum kuralları diye ifade edilen bu kurallar Türkçenin kolay konuşulmasını ve ahenkli olmasını sağlar. Türkçe, dünyadaki en ahenkli 4 dil arasında gösterilmektedir [14]. Ural-Altay dil grubunda görülen büyük ünlü uyumu, kelimedeki bulunan ünlü harfleri incelik-kalınlık açısından inceler [14]. Büyük ünlü uyumunda kelimedeki ilk ünlü harf inceyse kelimenin devamındaki diğer ünlüler de ince, kalınsa kalın olmalıdır. Bir diğer uyum kuralı olan küçük ünlü uyumu ise günümüz Türkçesinde kendini belirgin olarak göstermektedir [15, 16]. Büyük ünlü uyumunun aksine bu kuralda ünlüler kelimedeki bütün ünlülerle değil kendinden sonra gelen ünlüyle ilgilenir; kelimedeki düz ünlüden sonra düz ünlü, yuvarlak ünlüden sonra ise dar-yuvarlak veya düz-geniş ünlü gelebilir [13, 15, 16]. Bir diğer kural olan ünsüz uyumunda kelimelerde yan yana bulunan ünsüzlerin seda bakımından birbirine uyma durumu incelenir.

Türk Dil Kurumu sözlüğüne göre hece, bir solukta söylenebilen ses veya ses birlikleri olarak tanımlanmaktadır [18] ve Türkçede hecelerle ilgili kurallar mevcuttur. Örneğin; heceler en az 1 en fazla 4 harfli olabilir, hecelerde yalnızca 1 ünlü harf bulunabilir, sadece ünsüz harflerden oluşamaz, genelde bir ünlü ile birkaç ünsüzün birleşiminden meydana gelir [14, 15]. Ayrıca, Türkçede kelime ve heceler iki veya daha fazla ünsüz harfle başlayamayacağı; kelime ve hece sonunda yan yana -lç, -lk, -lp, -lt, -nç, -nk, -nt, -rç, -rk, -rp, -rd, -rt, -st ve -şt ünsüzlerinin bulunabileceği, 3 veya daha fazla sayıda ünsüzünse bulunamayacağı; kelimedeki yan yana bulunan 3 ünsüzün ancak kelime ortasında iki heceye ait olmak üzere bulunabileceği bilinmektedir [15]. Bunlarla beraber V ünlü harfi ve C ünsüz harfi temsil etmek üzere Türkçede kullanılan hece çeşitleri şunlardır; V, VC, CV, CVC, VCC, CVCC.

2.2 Morfolojik Analiz

Biçim birim diye de ifade edilen morfem; anlamlı, görevli, anlam değiştirme veya birleştirme görevi olan, bir dilin en küçük birimini oluşturan ses veya ses topluluklarıdır. Morfolojik analizde kelimenin kök/gövdeleri ve türleri, aldıkları ekler, eklerin türleri, kelime içindeki küçük yapılar, kısacası kelimelerdeki morfemler incelenmektedir [14, 15]. Bu işlem ise morfolojik analiz diye isimlendirilmektedir.

Morfem olan ekler tek başlarına anlamı olmayan, kök/gövdelere gelerek kullanılan anlamsal vazifeli seslerdir [16]. Eklemeli diller kategorisindeki Türkçe son ekli bir dildir [15]. Farklı amaçları gerçekleştirmek üzere Türkçede yapım ve çekim ekleri vardır. Sözcüğün anlamlı, parçalanamayan en küçük birimi köktür [15, 16] ve Türkçede

kökler değişmemektedir [14, 15]. Fakat kök/gövdeler yapım eki olarak yeni anlama sahip gövdeler oluşturabilir [16]. Bununla birlikte anlamı değiştirmeyen çekim ekleri kök/gövdelerin cümle içinde kullanılmasını sağlayarak onlara işlerlik kazandırmakta, cümle içinde dil bilgisine uygun şekilde kelimelerin yer almasını sağlamaktadır [16]. Ayrıca, Türkçedeki birleşik fiil, fiilden sıfat ve zarf fiil üreten ekler bazı kaynaklarda yapım eki; bazı kaynaklarda ise ayrı bir ek olarak kabul edilmektedir.

Fiilden isim ve fiil, isimden isim ve fiil olmak üzere 4 tür yapım eki vardır [15]. Yapım ekine kıyasla daha çok kullanım alanına sahip olan, isim ve fiil çekim ekleri olmak üzere 2 sınıfa ayrılan çekim eklerinde daha fazla sayıda ek bulunmaktadır. Türkçede eklerin birbiri ardına gelmesi belirli kurallar dahilinde gerçekleşir [19]. Örneğin; kelimenin en başında kökler bulunur [13] ve Türkçede yapım ekleri kök/gövdenin hemen bitişiğinde yer alır ve kök/gövde ile yapım eki arasına çekim eki girmez, bir kelimedede birden fazla yapım veya çekim eki bulunabilir, bunların sırasında da belirli bir kural vardır, kökten sonra önce yapım ekleri, sonrasında yapım ekleri bittikten sonra çekim ekleri gelir, yapım ve çekim eklerinin de kendi içinde sıralaması vardır [15].

III. ZEMBEREK

Açık kaynak kodlu DDİ kütüphanesi olan Zemberek [20], Türk dilleri için oluşturulmuş DDİ sistemidir [21]. Kelimelerin denetlenmesi, kök/gövdelerinin ve türlerinin belirlenmesi, eklerinin tespiti, ayrıştırılması ve hecelenmesinde Zemberek kullanılabilir. Kelime denetleme, analiz öncesi bir ön işlemdir. Denetlemede başarılı olan kelimelerin çözümlenmesi yapılabilir. Fakat morfolojik analizin başlatılması, analiz öncesi ön işlemler ve çözümlenme sonrası elde edilenlerin işlenmesi için başka yazılımlara ihtiyaç vardır. Örneğin; *Visual Studio C#, Visual Basic* veya *Java*, Zemberek'le entegre şekilde çalıştırılarak bu işlemlerin gerçekleştirilmesi sağlanabilir.

Öte yandan Zemberek'in bazı eksiklikleri bulunmaktadır; rakamlarla yazılmış sayıların veya küçük harflerle yazılmış özel isimlerin çözümlenmesi yapılamamaktadır. Ayrıca Zemberek kelime bazlı çalışmaktadır. Zemberek'e metinler parçalanarak kelime kelime verilmelidir. Zemberek analiz etmek üzere kelimeyi almakta ve muhtemel bütün çözümlenmeleri sunmaktadır. Hangi çözümlenmenin seçileceğine Zemberek değil, araştırmacı karar vermelidir. Ayrıca, bazı kelimelerin çözümlenmesinde yapım eki almış kelime ayrı gövde olarak sunulmaktayken bazı kelimelerin eklerine ayrılmış şekilde birden çok çözümlenmesi sunulmaktadır. Fakat çözümlenme sonuçlarındaki en üstte bulunan çözüm genellikle aranan kök/gövdeyi işaret etmektedir. Bu ve benzeri eksik işlemlerde Zemberek'le birlikte çalışacak yazılımlar devreye girmeli, eksik işlemler yazılım yoluyla tamamlanmalıdır. Buradan hareketle Zemberek'in morfolojik analiz yapmak üzere tasarlandığı, tek başına yeterli olamayacağı, eksikliklerinin yazılımlar yoluyla tamamlanması gerektiği söylenebilir.

Alan yazında Zemberek kullanıldığı çalışmalar mevcuttur [10, 22-26]. Çalışmalarda ağırlıklı olarak Zemberek'in kelimelerin kök/gövdesinin belirlenmesi özelliğinden faydalanılmaktadır. Ayrıca, yapılan incelemede Zemberek'in ek isimleriyle bu eklerin Türkçedeki karşılıklarının ortaya konduğu bir çalışmaya rastlanmamıştır. Özellikle kelimenin aldığı çekim eklerinin belirlenmesi, Zemberek'in söz dizimi analizinde de kullanılabilmesini mümkün kılacaktır.

IV. ÇALIŞMANIN AMACI

Bu çalışma, DDİ kütüphanesi olan Zemberek'teki eklerin Türkçe dil bilgisi karşılıklarının belirlenmesi, doğal bir dil olan Türkçede yazılmış cümlelerin DDİ teknikleriyle incelenerek ses bilgisi ve morfolojik analizlerinin yapılması ve işlemler sonucu elde edilen nicel verilerin ortaya konulması amacıyla gerçekleştirilmiştir. Çalışma kapsamında aşağıdaki sorulara yanıt aranmıştır;

- Zemberek ek isimlerinin Türkçe dil bilgisindeki karşılıkları nedir?
- Nicel veriler nelerdir?
 - Kelimeler kaç harflidir?
 - Kelimeler kaç hecelidir
 - Hece türlerinin kullanım sayıları nedir?
 - Heceler kaç harflidir?
 - Kelimelerin çözümlene sayıları nedir?
 - Kelimeler kaç ek almıştır?
 - Kelimelerin aldıkları eklerin türlerine göre sayıları nedir?
 - Kelimedeki eklerin birbirlerini takip etme sayıları nedir?
 - Türlerine göre kelime sayıları nedir?
 - Ses uyum kurallarına uyma/uyumama durumuna göre kelime sayıları nedir

Çalışmanın en önemli özelliklerinden birisi Zemberek eklerinin Türkçe dilbilgisi karşılıklarının belirlenmesidir. Ve bu, birçok alanda kullanılacak veriler ihtiva eder. Örneğin; Türkçedeki yapım ve çekim ekleri ile bu eklerden hangilerinin daha çok kullanıldığının tespitinde, eklerin sıralanışının ortaya konulmasında, Büyük Dil Modellerindeki (Large Language Model) işlemlerde, anlam bilimsel (semantic) ve söz dizimsel (syntactic) analizlerde buradan elde edilecek veriler anlamlı bilgiler içerebilir. Bununla birlikte metinlerdeki heceler, hece tiplerinin ve heceler harf sayılarının belirlenmesi hesaplamalı dil bilimin yanı sıra Türkçe öğretiminde (örneğin 1.sınıf metinlerinde) önemli bir konu olan metinlerin öğrenci grubuna uygun olup olmadığının ortaya konulmasına yönelik araştırmalar için de yardımcı bilgiler barındırır.

V. MATERYAL ve METOT

Bu çalışmada, cümle düzenindeki metinsel verilerin analizi gerçekleştirilmiştir. Zemberek'le *Visual Studio* ortamında C# dilinde geliştirilen yazılım birlikte kullanılarak metinler analiz edilmiştir. Kelimelerin çözümlenmesi işlemleri için Zemberek, bunlar dışındaki tüm işlemler içinse geliştirilen yazılım kullanılmıştır.

5.1 Veri Kümesi

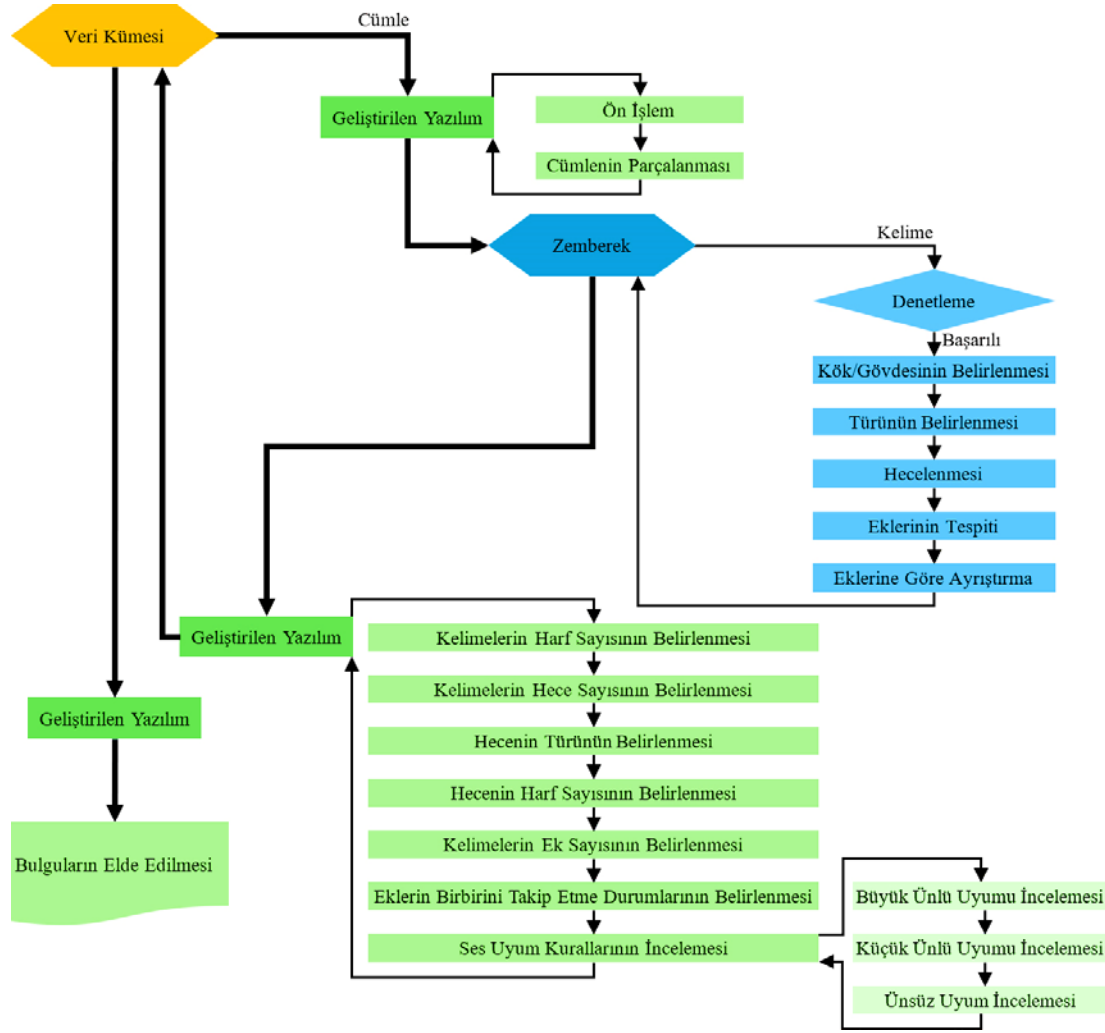
Çalışmanın veri kümesini 230 cümle oluşturmaktadır. Nicel verilerin yanı sıra Türkçede bulunan yapım eki, çekim eki ve diğer eklerin kullanımlarının DDİ yöntemleriyle ortaya konulmasının amaçlanması nedeniyle cümle düzenindeki metinler tercih edilmiştir.

5.2 Veri İşleme

Verilerin DDİ ile işlenmesinde Zemberek'le entegre şekilde çalışan, *Visual Studio* ortamında ve *C#* dilinde geliştirilen yazılımdan faydalanılmıştır. Cümle düzenindeki verilerin ön işlemden geçirilmesi, cümlelerin kelimelere parçalanması, kelimelerin harf ve hece sayılarının, hecelerın türlerinin ve harf sayılarının, kelimelerin aldıkları ek sayılarının, türlerine göre eklerin kullanım sayılarının, eklerin birbirini takip etme durumlarının, türlerine göre kelime sayılarının, kelimelerin ses uyum kuralına uyma/uymama durumlarının kontrolü ve sayıları, geliştirilen yazılımla gerçekleştirilmiştir. Kelimelerin denetlenmesi, kök/gövdelerin belirlenmesi, türlerinin saptanması, hecelenmesi, aldığı eklerin tespiti ve eklerine göre ayrıştırılması işlemlerinde ise Zemberek kullanılmıştır.

Çalışmanın işlem akış şeması Şekil 1'de gösterilmiştir;

- Veri kümesindeki 230 cümle döngüyle sisteme girdi olarak sırayla alınmaktadır.
- Her bir cümle sırasıyla işlenmek üzere geliştirilen yazılıma verilmektedir.
- Geliştirilen yazılım işlemleri;
 - Ön işlem: Cümle ön işlem sürecinden geçirilmektedir. Ön işlemde cümle içinde bulunan rakamla belirtilen sayıların yazıyla ifadesi gerçekleşmekte, ayrıca kesme işareti (') dışındaki noktalama işaretleri içerikten kaldırılmaktadır.
 - Cümlelerin parçalanması: Zemberek'e sunulmak üzere cümle içinde bulunan kelimeler boşluk karakterine göre parçalanmakta ve kelimeler elde edilmektedir.
- Zemberek işlemleri;
 - Cümleyi oluşturan kelimeler döngüyle Zemberek'e girdi olarak sırayla alınmaktadır.
 - Zemberek'le öncelikle kelimenin denetlenmesi yapılmaktadır.
 - Denetleme sonucu başarılı olan kelimenin çözümlenmesi gerçekleştirilmektedir.
- Geliştirilen yazılım işlemleri;
 - Zemberek çözümlenmesi sonrasında kelimenin harf ve hece sayıları, hecelerın türü, kelimenin ek sayıları, eklerin birbirini takip etme durumları ve ses uyum kurallarına uyma/uymama durumları belirlenmektedir.
- Geliştirilen yazılım işlemleri;
 - Veri işleme süreci ve cümleye ait döngü tamamlandıktan sonra çalışma kapsamında elde edilen bulgular elde edilmektedir.

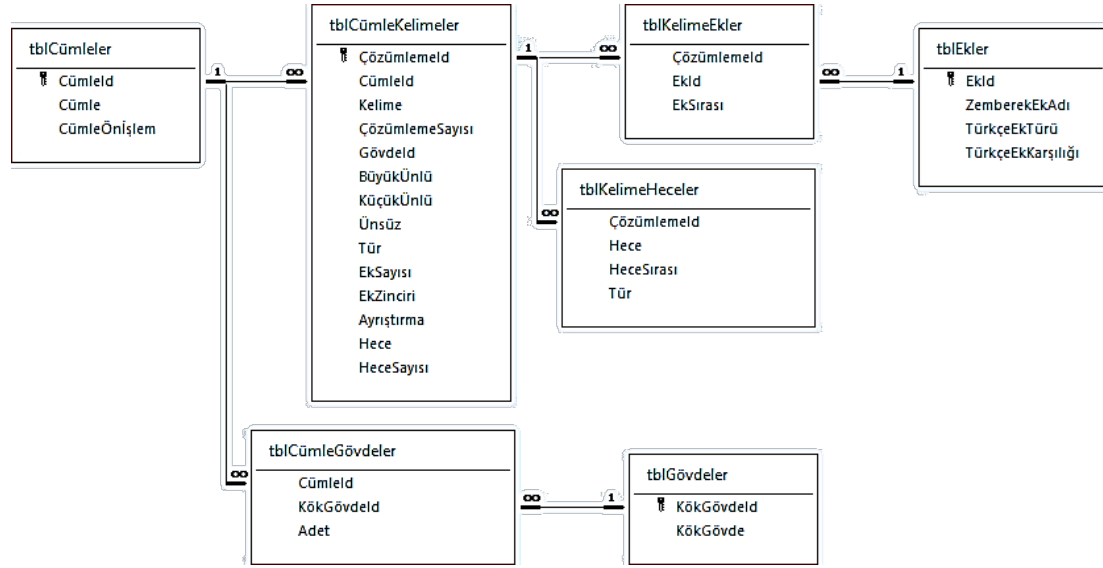


Şekil 1. İşlem akış şeması (Yeşiller geliştirilen yazılımla, maviler Zemberek'le yapılan işlemleri belirlemektedir)

Çalışmada kullanılan veri tabanına ait şema Şekil 2'de görülmektedir. Normalizasyon kurallarına ve ilişkisel veri modeline uygun olarak oluşturulmuş veri tabanında 7 tablo bulunmaktadır. Tablolar arası ilişkilerden faydalanılarak veri tabanı üzerinde çalıştırılan SQL (Structured Query Language / Yapılandırılmış Sorgu Dili) sorgularıyla çalışmaya ait verilerin elde edilmesi gerçekleştirilmiştir.

Tablolara ait açıklamalar şu şekildedir;

- tblCümleler: Veri kümesini oluşturan 230 cümlelerin ve ön işlem sonrası elde edilen formlarının tutulduğu tablodur.
- tblCümleKelimeler: Cümleleri meydana getiren kelimelerin işlenmesi sonucu elde edilen verilerinin tutulduğu tablodur.
- tblKelimeEkler: Kelimelerin aldıkları eklerle ilişkin verilerin tutulduğu tablodur.
- tblEkler: Zemberek eklerinin Türkçe dil bilgisindeki karşılıklarının, eşleşmelerinin tutulduğu tablodur.
- tblKelimeHeceler: Kelimelerin hecelerine ilişkin verilerin tutulduğu tablodur.
- tblCümleGövdelere: Kelimelerin kök/gövdelere ile onların cümle içindeki kullanım sayılarının tutulduğu tablodur.
- tblGövdelere: Ayrık (unique-benzersiz) kök/gövdelere tutulduğu tablodur.



Şekil 2. Veri tabanı şeması

Tablolara ait açıklamaları ve Şekil 2'de şeması verilen veri tabanına girilmiş verilerin örneği Şekil 3'te görülmektedir. 230 cümleden temsili olarak yalnızca 3 tanesiyle işlemler gerçekleştirilmiş ve veriler veri tabanına aşağıdaki şekilde kaydedilmiştir. Diğer cümlelere ait işlemler de benzer şekilde veri tabanında tutulmuştur. Bunlarla birlikte Zemberek eklerinin yer aldığı tblEkler tablosunda, 130 Zemberek ekinden örnek olarak 50 tanesi gösterilmiştir.

CümleId	Cümle	CümleÖnişlem
1	Benim gözüm rengim yeşildir.	Benim gözüm rengim yeşildir.
2	0 çocuk kim?	0 çocuk kim
3	Pazartesi okul başlıyor.	Pazartesi okul başlıyor

CümleId	KökGövdeld	Adet
1	1	1
1	2	1
1	3	1
1	4	1
1	5	1
1	6	1
1	7	1
1	8	1
1	9	1
1	10	1

KökGövdeld	KökGövdde
1	ben
2	göz
3	renk
4	yeşil
5	0
6	çocuk
7	kim
8	pazartesi
9	okul
10	başlamak

CümleId	EkId	EkSırası	Tür
1	be	1	CV
1	mm	2	CV
1	göz	1	CV
1	ren	1	CV
1	gim	2	CV
1	ye	1	CV
1	gil	2	CV
1	dir	3	CV
1	o	1	V
1	go	1	CV
1	çuk	2	CV
1	7	kim	1
1	pa	1	CV
1	8	paz	2
1	8	pa	3
1	8	si	4
1	9	o	1
1	9	kul	2
1	10	baş	1
1	10	2	CV
1	10	yor	3

CümleId	EkId	EkSırası	Tür	EkZinciri	Ayrıştırma	Hece	HeceSayısı
1	1	1	1	ZAMIR	2	ZAMIR_KOK, ZAMIR_SAHPLIK_IM	ben im
1	2	1	1	ISIM	1	ISIM_KOK	göz
1	3	1	1	ISIM	2	ISIM_KOK, ISIM_SAHPLIK_BEN_IM	renk im
1	4	1	1	SIFAT	2	ISIM_KOK, ISIM_TANIMLAMA_DIR	yeşil dir
1	5	1	1	ZAMIR	1	ZAMIR_KOK	o
1	6	1	1	ISIM	1	ISIM_KOK	çocuk
1	7	1	1	ISIM	1	ISIM_KOK	kim
1	8	1	1	ISIM	1	ISIM_KOK	pazartesi
1	9	1	1	ISIM	1	ISIM_KOK	okul
1	10	1	1	FIL	2	FIL_KOK, FIL_SIMDIKIZAMAN_YOR	başla yor

EkId	ZemberekEkAdı	TürkçeEkTürü	TürkçeEkKarşılığı
0	BAGLAC_KOK		
1	EDAT_KOK		
2	FIL_KOK		
3	GENL_KOK		
4	INEX_KOK		
5	ISIM_KOK		
6	OZEL_KOK		
7	SAYI_KOK		
8	SORU_KOK	Çekim Eki	Soru Eki
9	UNLEM_KOK		
10	YANKI_KOK		
11	ZAMAN_KOK		
12	ZAMIR_KOK		
13	BOS_EK		
14	FIL_BELIRTEME_DIK	Diğer Ekler	Fill→Şıfat Fill
15	FIL_BERABERLIK_IS	Yapım Eki	Fill→Fill
16	FIL_DEK_IL	Diğer Ekler	Fill→Zarf Fill
17	FIL_DEVAMLILIK_DIKCE	Diğer Ekler	Fill→Zarf Fill
18	FIL_DONUSUM_ECKE	Diğer Ekler	Fill→Şıfat Fill
19	FIL_DONUSUM_EN	Diğer Ekler	Fill→Şıfat Fill
20	FIL_DONUSUM_ESI	Diğer Ekler	Fill→Şıfat Fill
21	FIL_DONUSUM_ECKE	Diğer Ekler	Fill→Şıfat Fill
22	FIL_DONUSUM_ESIYE	Diğer Ekler	Fill→Zarf Fill
23	FIL_DONUSUM_IK	Yapım Eki	Fill→İsim
24	FIL_DONUSUM_ILU		
25	FIL_DONUSUM_ILM	Yapım Eki	Fill→İsim
26	FIL_DONUSUM_INTI	Yapım Eki	Fill→İsim
27	FIL_DONUSUM_IS	Yapım Eki	Fill→İsim
28	FIL_DONUSUM_ME	Yapım Eki	Fill→İsim
29	FIL_DONUSUM_MEZ	Yapım Eki	Fill→İsim
30	FIL_DONUSUM_MIS	Yapım Eki	Fill→İsim
31	FIL_EDUGEN_IL	Yapım Eki	Fill→Fill
32	FIL_EDUGENESKI_N	Yapım Eki	Fill→Fill
33	FIL_EMIR_O_SIN	Çekim Eki	Tasarımlama Kipi (Emir)
34	FIL_EMIR_ONKAD_SINLER	Çekim Eki	Tasarımlama Kipi (Emir)
35	FIL_EMIR_SIZ_IN	Çekim Eki	Tasarımlama Kipi (Emir)
36	FIL_EMIR_SIZRESMI_NIZ	Çekim Eki	Tasarımlama Kipi (Emir)
37	FIL_ETTIRGEN_TEKRAR_T	Yapım Eki	Fill→Fill
38	FIL_ETTIRGEN_TIS	Yapım Eki	Fill→Fill
39	FIL_GEÇMISZAMAN_DI	Çekim Eki	Zaman (Bildirme) Kipi (Geçmiş)
40	FIL_GEÇMISZAMAN_MIS	Çekim Eki	Zaman (Bildirme) Kipi (Geçmiş)
41	FIL_GELECEKZAMAN_ECKE	Çekim Eki	Zaman (Bildirme) Kipi (Gelecek)
42	FIL_GELECEKZAMAN_IR	Çekim Eki	Zaman (Bildirme) Kipi (Gelecek)
43	FIL_GILB_CESINE	Diğer Ekler	Fill→Zarf Fill
44	FIL_IMSI_IP	Diğer Ekler	Fill→Zarf Fill
45	FIL_ISTEK_E	Çekim Eki	Tasarımlama Kipi (İstek)
46	FIL_ISTEK_SENE	Çekim Eki	Tasarımlama Kipi (Dilek-Şart)
47	FIL_ISTEK_SENZE	Çekim Eki	Tasarımlama Kipi (Dilek-Şart)
48	FIL_KISI_BEN	Çekim Eki	Şahıs Eki
49	FIL_KISI_BIZ	Çekim Eki	Şahıs Eki
50	FIL_KISI_O		

Şekil 3. Veri tabanında kayıtlı veriler

VI. BULGULAR VE TARTIŞMA

Bu bölümde sunulan veriler 230 cümle ve cümleleri oluşturan 744 kelimenin analizinden elde edilmiştir. Öncelikle, Zemberek ek adlarının Türkçe dil bilgisi karşılıkları örneklerle verilmiş, sonrasında ise nicel verilere ait bulgular tablolarda sunulmuştur.

6.1 Zemberek Ek İsimlerin Türkçe Dil Bilgisi Karşılıkları

Zemberek'te toplam 130 ek adı bulunmaktadır. Bunlardan 1 tanesi BOS_EK, 13 tanesiye FIIL_KOK, ISIM_KOK, EDAT_KOK gibi kelime türünü de ifade eden kök adlarıdır. Bunlarla birlikte eklerin 39 tanesi yapım eki, 8 tanesi birleşik fiil, 5 tanesi fiilden sıfat fiil türeten, 9 tanesi fiilden zarf fiil türeten ek iken 46 tanesi çekim ekidir. Zemberek çözümlenmeleri incelendiğinde 13 kökten biri olan SORU_KOK, kök olmasına karşın çekim eklerinden soru ekinde yer alması gerektiği belirlenmiştir. Öte yandan Zemberek ek listesinde yer alan 10 ekin kullanımına yapılan analizde rastlanmamıştır.

Zemberek yapım eki almış kelimeleri her zaman ayrı bir gövde olarak sunmayabilmektedir. Örneğin; yapım eki almış *öğretmen* kelimesi *öğretmen* olarak doğru şekilde çözümlenirken; *bayramlaştı* kelimesinin kök/gövdesi *bayram* olarak çözümlenmiştir. Oysaki *bayramlaştı* kelimesinde yapım ekleri bulunmaktadır ve bu kelimenin gövdesi *bayramlaş(mak)* olmalıdır. Kelimenin Zemberek çözümlenmesindeki ek zincirinde "ISIM_KOK, ISIM_DONUSUM_LE, FIIL_BERABERLIK_IS, FIIL_GECMISZAMAN_DI" eklerinin bulunduğu, bunlardan ISIM_DONUSUM_LE ve FIIL_BERABERLIK_IS eklerinin yapım eki olduğu, kelimenin *bayram-la-ş-tı* olarak ayrıştırıldığı, buna göre *bayram* köküne *la* ve *ş* eklenerek *bayramlaş(mak)* olarak kullanılması gerektiği sonucuna ulaşılmıştır. Bu sebeple eğer kelimenin yapım eki almış gövdesine ulaşılacak isteniyorsa Zemberek ek zinciri de incelenmeli, ek zincirinde yapım eki olduğu yere kadar ayrıştırmadaki ekler kök/gövdeye eklenmelidir. Böylece kelimelerin kök/gövdelerinin doğru tespit edilmesi sağlanmış olunacaktır.

Zemberek'teki yapım eklerinin Türkçe dil bilgisindeki karşılıkları, ekin hangi türden yapım eki olduğu ve ek kullanımına ilişkin örnek kelime ayrıştırmasıyla birlikte Tablo 1'de verilmiştir. Fiilden fiil türeten 7, isim türeden 11 ek varken isimden isim türeten 18, fiil türeten 3 ek vardır.

Tablo 1. Zemberek ek adlarının Türkçe dil bilgisindeki karşılıkları (yapım ekleri)

Zemberek Ek Adı	Türkçe Dil Bilgisi Karşılığı	Kelime	Ayrıştırma
FIIL_BERABERLIK_IS		bakışmak	bak <i>ış</i> mak
FIIL_EDILGEN_IL		bakılır	bak <i>ıl</i> ır
FIIL_EDILGENSESLI_N		seslenmiş	sesle <i>n</i> miş
FIIL_ETTIRGEN_TIR	Fiilden Fiil	baktırmak	bak <i>tır</i> mak
FIIL_ETTIRGEN_TEKRAR_T		baktırtmak	bak <i>tır t</i> mak
FIIL_OLDURGAN_T		yükseltmiş	yüksel <i>t</i> miş
FIIL_OLUMSUZLUK_ME		bakmadı	bak <i>ma</i> dı
FIIL_DONUSUM_IK		ayrılık	ayrıl <i>ık</i>
FIIL_DONUSUM_IM		doyum	doy <i>um</i>
FIIL_DONUSUM_INTI		bozuntu	boz <i>untu</i>
FIIL_DONUSUM_IS		kurtuluş	kurtul <i>uş</i>
FIIL_DONUSUM_ME		sevmenin	sev <i>me</i> nin
FIIL_DONUSUM_MEZ	Fiilden İsim	sevilmeyiz	sev <i>il me</i> z
FIIL_DONUSUM_MIS		sevilmiş	sev <i>il miş</i>
FIIL_MASTAR_CE		kovalamaca	kovala <i>ma ca</i>
FIIL_MASTAR_MEK		sevilmek	sev <i>il mek</i>
FIIL_TANIMLAMA_ICI		üzücü	üz <i>ücü</i>
FIIL_ZAMAN_INCE		sevince	sev <i>ince</i>
ISIM_ANDIRMA_IMSI		mavimsi	mavi <i>msi</i>
ISIM_ANDIRMA_SI		adamsı	adam <i>sı</i>
ISIM_BULUNMA_LI		köylüleri	köy <i>lü</i> ler i
ISIM_BULUNMA_LIK		yolculuk	yolcu <i>luk</i>
ISIM_DURUM_LIK		sabırsızlık	sabır <i>sızlık</i>
ISIM_GIBI_CE		irice	iri <i>ce</i>
ISIM_ILGI_CI		yardımcı	yardım <i>cı</i>
ISIM_ILISKILI_SEL		bedensel	beden <i>sel</i>
ISIM_KUCULTME_CEGIZ	İsimden İsim	adamcağız	adam <i>cağız</i>
ISIM_KUCULTME_CIK		küçükük	küçük <i>çük</i>
ISIM_TARAFINDAN_CE		bence	ben <i>ce</i>
ISIM_YOKLUK_SIZ		kaygısız	kaygı <i>sız</i>
SAYI_KOSE_GEN		üçgen	üç <i>gen</i>
SAYI_SIRA_INCI		ikinci	iki <i>nci</i>
SAYI_TOPLULUK_IZ		beşiz	beş <i>iz</i>
SAYI_ULESTIRME_ER		ikişer	iki <i>şer</i>
YANKI_DONUSUM_TI		tıkırttı	tıkır <i>tı</i> yı
ZAMAN_BELIRTME_KI		sonraki	sonra <i>ki</i>
ISIM_DONUSUM_LE		adımla	adım <i>la</i>
ISIM_DONUSUM_LES	İsimden Fiil	silikleştirdi	sil <i>ik leş</i> ti
YANKI_DONUSUM_DA		fısıldadığını	fısıll <i>da</i> dık ın ı

* Ayrıştırma sütunundaki koyu olan kısımlar, ilgili ekin kelime içindeki yerini ifade etmektedir.

Tablo 2’de Türkçede yapım eki olduğu veya olmadığıyla kesin bir yargı bulunmayan ekler görülmektedir. Bu eklerden 8 tanesi birleşik fiil iken 5 tanesi fiilden sıfat fiil, 9 tanesiye fiilden zarf fiil türetmek amacıyla kullanılmaktadır.

Tablo 2. Zemberek ek adlarının Türkçe dil bilgisindeki karşılıkları (diğer ekler)

Zemberek Ek Adı	Türkçe Dil Bilgisi Karşılığı	Kelime	Ayrıştırma	
FIIL_SURERLIK_EDUR	Birleşik Fiil	sevinedursun	sevin <i>edur</i> sun	
FIIL_SURERLIK_EGEL		alışılagelmiş	alış ıl <i>agel</i> miş	
FIIL_SURERLIK_EGOR		başlamayagörsün	başla ma <i>yagör</i> sün	
FIIL_SURERLIK_EKAL		uyuyakaldı	uyu <i>yakal</i> dı	
FIIL_TEZLIK_IVER		yapıver	yap <i>iver</i>	
FIIL_YAKLASMA_AYAZ		düşeyazdı	düş <i>eyaz</i> dı	
FIIL_YETENEK_EBİL		ölebilir	öl <i>ebil</i> ir	
FIIL_YETERSİZLIK_E		alamaz	al <i>a</i> ma z	
FIIL_BELIRTME_DİK		Fiilden Sıfat Fiil	yaptığı	yap <i>tık</i> ı
FIIL_DONUSUM_ECEK			vereceği	ver <i>ecek</i> i
FIIL_DONUSUM_EN	olan		ol <i>an</i>	
FIIL_DONUSUM_ESİ	öpülesi		öp ül <i>esi</i>	
FIIL_DONUSUM_ESİCE	kopasıca		kop <i>asca</i>	
FIIL_BERİ_ELI	Fiilden Zarf Fiil	ayrılalı	ayrıl <i>alı</i>	
FIIL_DEVAMLILIK_DİKCE		doldukça	dol <i>dukça</i>	
FIIL_DONUSUM_ESİYE		öldüresiye	öldür <i>esiye</i>	
FIIL_GIBI_CESİNE		ağlarcasına	ağla r <i>casına</i>	
FIIL_IMSI_IP		aşıp	aş <i>ıp</i>	
FIIL_OLUMSUZLUK_DEN		durmadan	dur ma <i>dan</i>	
FIIL_OLUMSUZLUK_SİZİN		aldırmaksızın	aldır mak <i>sızın</i>	
FIIL_SUREKLİLİK_EREK		giderek	git <i>erek</i>	
IMEK_ZAMAN_KEN		dönerken	dön er <i>ken</i>	

* Ayrıştırma sütunundaki koyu olan kısımlar, ilgili ekin kelime içindeki yerini ifade etmektedir.

Bu çalışmada çokluk eki diye ifade edilen ekin çoğul eki, hal ekinin durum eki, iyelik ekinin aitlik eki olarak isimlendirilmesi gibi aynı çekim eklerinin farklı kaynaklarda farklı şekillerde isimlendirilebildiği görülmektedir. Çekim eklerinin Zemberek'teki isimleri ile Türkçe dil bilgisindeki karşılıkları eşleştirilerek Tablo 3'te verilmiştir. Çokluk ekinde 2, ekeylemede 5, hal ekinde 6, iyelik ekinde 10, soru ekinde 1, şahıs ekinde 5, tasarlama kipinde 10 ve zaman kipinde 7 olmak üzere toplamda 46 çekim eki bulunmaktadır.

Anlam değiştirmeyen, eklendiği kelimenin cümle içinde yer almasını sağlayan ve cümledeki kelimelerin birbirlerinin durumlarına göre çekimlenmesini sağlayan çekim ekleri, kelimelerde en sonda yer alır. Dolayısıyla ek incelemelerinde önce kök/gövde, ardından yapım ekleri ve son olarak da çekim ekleri araştırılmalıdır. Cümle içindeki kelimeleri bağlama ve onların bir arada kullanılmasını sağlama görevi nedeniyle kelimelerin aldığı çekim eklerinin belirlenmesi özellikle söz dizimsel analiz yapılan çalışmalarda önem taşımaktadır.

Tablo 3. Zemberek ek adlarının Türkçe dil bilgisindeki karşılıkları (çekim ekleri)

Zemberek Ek Adı	Türkçe Dil Bilgisi Karşılığı	Kelime	Ayrıştırma	
ISIM_COGUL_LER	Çokluk Eki	bitkiler	bitki <i>ler</i>	
ISIM_KISI_ONLAR_LER		görenler	gör en <i>ler</i>	
ISIM_KISI_BEN_IM	Ekeylem	şanslıyım	şanslı <i>yım</i>	
ISIM_KISI_SEN_SIN		şanslısın	şanslı <i>sın</i>	
ISIM_KISI_BIZ_IZ		şanslıyız	şanslı <i>ız</i>	
ISIM_KISI_SIZ_SINIZ		şanslısınız	şanslı <i>sınız</i>	
ISIM_TANIMLAMA_DIR		şanslıdır	şanslı <i>dır</i>	
ISIM_BELIRTME_I		Belirtme	oğlu	oğul <i>u</i>
ISIM_KALMA_DE	Hal Eki	Bulunma	pazarda	
SAYI_KESIR_DE		binde	bin <i>de</i>	
ISIM_CIKMA_DEN		Uzaklaşma	oradan	ora <i>dan</i>
ISIM_BIRLIKTELİK_LE		Vasıta	kalemle	kalem <i>le</i>
ISIM_YONELME_E		Yönelme	avcıya	avcı <i>ya</i>
ISIM_SAHİPLİK_BEN_IM		İyelik Eki	evim	evi <i>m</i>
ISIM_SAHİPLİK_SEN_IN	evin		ev <i>in</i>	
ISIM_SAHİPLİK_O_I	evi		ev <i>i</i>	
ISIM_SAHİPLİK_BIZ_IMIZ	evimiz		ev <i>imiz</i>	
ISIM_SAHİPLİK_SIZ_INIZ	eviniz		ev <i>iniz</i>	
ISIM_SAHİPLİK_ONLAR_LERI	evleri		ev <i>leri</i>	
ZAMIR_SAHİPLİK_IM	benim		ben <i>im</i>	
ISIM_BULUNMA_KI	evdeki		ev de <i>ki</i>	
ISIM_TAMLAMA_I	ihtiyacı		ihtiyaç <i>ı</i>	
ISIM_TAMLAMA_IN	ihtiyacın		ihtiyaç <i>ın</i>	
SORU_KOK	Soru Eki		Ben mi?	ben <i>mi</i>
FIIL_KISI_BEN	Şahıs Eki		yaptım	yap tı <i>m</i>
FIIL_KISI_SEN		yaparsın	yap ar <i>sın</i>	
FIIL_KISI_BIZ		yaparız	yap ar <i>ız</i>	
FIIL_KISI_SIZ		yaptınız	yap tı <i>nız</i>	
FIIL_KISI_ONLAR		yaparlar	yap ar <i>lar</i>	
FIIL_EMIR_O_SIN		Emir	arasın	ara <i>sın</i>
FIIL_EMIR_ONLAR_SINLER	arasınlar		ara <i>sınlar</i>	
FIIL_EMIR_SIZ_IN	arayın		ara <i>yın</i>	
FIIL_EMIR_SIZRESMI_INIZ	aramayınız		ara ma <i>yınız</i>	
FIIL_ISTEK_E	İstek		gideyim	git <i>e yim</i>
FIIL_ISTEK_SENE	Dilek-Şart		alsana	al <i>sana</i>
FIIL_ISTEK_SENIZE		alsanıza	al <i>sanıza</i>	
FIIL_SART_SE		alsa	al <i>sa</i>	
IMEK_SART_SE		alırsa	al ır <i>sa</i>	
FIIL_ZORUNLULUK_MELI		Gereklilik	olmalı	ol <i>malı</i>
FIIL_GECMISZAMAN_DI		Zaman (Bildirme) Kipi	Geçmiş	geldim
FIIL_GECMISZAMAN_MIS	gelmiş		gel <i>miş</i>	
FIIL_GELECEKZAMAN_ECEK	Gelecek		geleceğim	gel <i>ecek im</i>
FIIL_SIMDIKIZAMAN_IYOR	Şimdiki		geliyor	gel <i>iyor</i>
FIIL_GENISZAMAN_IR	Geniş		gelir	gel <i>ir</i>
IMEK_RIVAYET_MIS	Öğrenilen Geçmiş		gelecekmış	gel ecek <i>miş</i>
IMEK_HIKAYE_DI	Görülen Zaman		gelecekti	gel ecek <i>ti</i>

* Ayrıştırma sütunundaki koyu olan kısımlar, ilgili ekin kelime içindeki yerini ifade etmektedir.

6.2 Nicel Veriler

Harf sayısına göre kelime sayıları Tablo 4'te verilmiştir. Harf sayılarının belirlenmesinde kelimelerdeki noktalama işaretleri kaldırılmıştır (' . ? !). Kelimeler toplam 3945 ve ortalama 5.30 harflidir. Kelime havuzunda en az 1, 12 ve 13 harfli; en çok ise 5 harfli kelimeler bulunmaktadır. Veri kümesindeki en uzun kelime *dikdörtgendir* kelimesidir. Kelimelerin %76'sı 3-7 harflidir.

Tablo 4. Harf sayısına göre kelime sayıları

Harf Sayısı	Kelime Sayısı
1	1
2	57
3	137
4	68
5	172
6	97
7	92
8	53
9	41
10	21
11	3
12	1
13	1

Hece sayısına göre kelime sayıları Tablo 5'te görülmektedir. Kelimeler toplam 1683 hecelidir. Kelimelerin ortalama hece sayısı ise 2.26'dır. Veri kümesini en az 0 ve 5 heceli, en çok 2 heceli kelimeler oluşturmaktadır. En çok heceye *a-ğa-be-yim-le* ve *bi-sik-le-ti-mi* kelimeleri sahiptir. Bunlarla birlikte 0 heceli kelime de tespit edilmiştir. 0 heceli olan *spor* kelimesinin Zemberek denetlemesinden geçtiği, kökünün *spor* olarak belirlendiği fakat hecelenemediği görülmüştür. Bu kelimenin Türkçenin özelliklerinden olan kelime ve hecelerin başında 2 ünsüzün yan yana bulunamayacağı kuralına aykırı olduğu saptanmıştır.

Tablo 5. Hece sayısına göre kelime sayıları

Hece Sayısı	Kelime Sayısı
0	2
1	152
2	317
3	197
4	74
5	2

V, VC, CV, CVC, VCC ve CVCC olmasına göre hece türlerinin sayıları Tablo 6'da sunulmuştur. Türkçede var olan ve ünlü-ünsüz-ünsüz harf birleşiminden meydana gelen VCC türünden bir heceye veri kümesinde rastlanmadığı; en çok ünsüz-ünlü harf birleşiminden oluşan CV hece türünün bulunduğu belirlenmiştir. Öte yandan Türkçedeki hecelerde görülmeyen ünsüz-ünsüz-ünlü harf birleşiminden oluşan CCV tipine ait 1 kelime, ünsüz-ünsüz-ünlü-ünsüz harf birleşiminden oluşan CCVC hece tipine ait 2 kelime kullanıldığı saptanmıştır. CCV tipindeki heceye sahip *sporu* kelimesi Zemberek denetlemesinden geçmiş, kök/gövdesi *spor* olarak belirlenmiş ve *spo-ru* olarak hecelenmiştir. CCVC tipindeki heceye sahip olan *spor* kelimesi de Zemberek denetlemesinden

geçmiş, kök/gövdesi ve hecesi *spor* olarak tespit edilmiştir. Esasında *spor* ve *spo* hecelerinin, Türkçedeki hecelerin 2 ünsüz harfle başlayamayacağı kuralına aykırı olmalarına karşın Zemberek tarafından hecelerin üretilebildiği gözlenmiştir.

Tablo 6. Türüne göre hece sayıları

Hece Türü	Hece Sayısı
V	105
VC	74
CV	839
CVC	653
VCC	-
CVCC	11
Diğer (CCV)	1
Diğer (CCVC)	2

* V: ünlü harf, C: ünsüz harf

Hecelerin harf sayısına göre sayılarının gösterildiği Tablo 7 incelendiğinde hecelerin %93'ünün 2-3 harften oluştuğu, 4 harfli hece sayısının %1'in altında olduğu, hecelerin ortalama 2.34 harften meydana geldiği görülmüştür.

Tablo 7. Harf sayısına göre hece sayıları

Harf Sayısı	Hece Sayısı
1	105
2	913
3	654
4	13

Zemberek işlemek üzere aldığı kelimenin olası tüm çözümlerini kullanıcıya sunmaktadır. Özellikle ek alan kelimelerde birden fazla çözümlenme Zemberek tarafından sunulmaktadır. Fakat hangi çözümlenmenin aranan çözümlenme olduğuna araştırmacı dikkatli şekilde inceleyerek karar vermelidir. Çözümlenme sayılarına göre kelime sayıları Tablo 8'de verilmiştir. Kelimelerin %67'sinin 1 çözümlenmeli olması, kök/gövde şeklinde olduklarını ve muhtemelen yapım veya çekim almadıklarına ilişkin bilgi vermektedir. 1 çözümlenmeli kelime sayısının oranının yüksekliği, araştırmacıların hangi çözümlenmenin aranan çözümlenme olduğunu belirlemeleri açısından kolaylık sağlayacaktır. Kelimelerin %96'sının 1-3 çözümlenmeye sahip olduğu, 5-7 çözümlenmeli kelime oranı %2'den az olduğu gözlenmiştir.

Tablo 8. Çözümlenme sayısına göre kelime sayıları

Çözümlenme Sayısı	Kelime Sayısı
1	575
2	169
3	73
4	24
5	6
6	5
7	1

Türkçede çekim eki olan *mi* soru eki, cümlelerde ayrı yazılmaktadır. Fakat kelimelerin aldıkları ek sayılarının tespitinde soru eki ayrı bir kelime olarak değil; önündeki kelimenin eki olarak işleme alınmıştır. Bu bilgiler ışığında Tablo 9'a göre kelimelerin en fazla 3 ek aldığı, %50'sinin hiçbir ek almadan kullanıldığı ve 1 ekli kelime oranının %36 olduğu belirlenmiştir.

Tablo 9. Ek sayısına göre kelime sayıları

Ek Sayısı	Kelime Sayısı
0	371
1	268
2	86
3	19

Veri kümesindeki 230 cümlelerin analizi neticesinde cümlelerdeki kelimelerde kullanılan yapım eki, diğer ekler ve çekim eklerinin kullanım sayıları Tablo 10'da verilmiştir. Yapım eklerinden en çok fiilden fiil ve isimden isim türetenlerin, diğer eklerden birleşik fiillerin, çekim eklerindense zaman kipinin ve iyelik ekinin kullanıldıkları görülmüştür. Tüm ekler birlikte değerlendirildiğinde çekim eklerinin yapım ve diğer eklerden çok daha fazla kullanıldığı gözlenmiştir; çekim eklerinin toplam kullanım oranı %91, yapım eklerinin %7 ve diğer eklerin kullanım oranı %2'dir. Öte yandan yapım eklerinden isimden fiil türeten ve diğer eklerden fiilden zarf fiil dönüşümü yapan ek kullanımına rastlanmamıştır. Ayrıca, Zemberek'in yapım eki almış bazı kelimelerin çözümlemelerinde yapım eki almış kelimeyi bazen ayrı bir gövde olarak bazen de kök/gövde ve yapım eklerini vererek sunması nedeniyle, yapım eklerinin kullanım sayıları az olabilir.

Tablo 10. Türlerine göre eklerin kullanım sayıları

Ek Türü	Ek Adı	Kullanım Sayısı
Yapım eki	Fiilden fiil	17
	Fiilden isim	5
	İsimden isim	14
Diğer ekler	Birleşik fiil	7
	Fiilden sıfat fiil	3
Çekim eki	Çokluk eki	2
	Ekeylem	20
	Hal eki	85
	İyelik eki	119
	Soru eki	25
	Şahıs eki	58
	Tasarlama kipi	2
Zaman kipi	140	

Eklerin birbirlerini takip etme durumlarını ortaya koymak amacıyla, hangi ekten sonra hangi ekin geldiği belirlenerek Tablo 11 oluşturulmuştur. Buna göre fiilden fiil türeten yapım ekinden sonra yalnızca zaman kipinin, fiilden isim türeten yapım ekinden sonra isimden isim türeten yapım eki veya hal ekinin; isimden isim türeten yapım ekinden sonra ekeylem, iyelik eki veya soru ekinin; birleşik fiilden sonra yalnızca zaman kipinin, fiilden sıfat fiil türeten yapım ekinden sonra yalnızca iyelik ekinin geldiği; buna karşın yapım ekinden sonra yapım ekinin gelmediği gözlenmiştir. Çekim eklerinin durumu incelendiğinde hal ekinden sonra yalnızca çokluk ekinin; iyelik ekinden sonra ekeylem, hal, iyelik ve soru ekinin; soru ekinden sonra yalnızca ekeylemin, şahıs ekinden sonra

yalnızca soru ekinin, tasarlama kipinden sonra yalnızca şahıs ekinin, zaman kipinden sonra ise ekeylem, soru eki ve şahıs ekinin geldiği saptanmıştır.

Tablo 11. Eklerin birbirlerini takip sayıları

Ek Türü	Ek Adı	Yapım Eki			Diğer Ekler						Çekim Eki		
		-	İsimden İsim	Çokluk Eki	Ekeylem	Hal Eki	İyelik Eki	Soru Eki	Şahıs Eki	Zaman Kipi	-	-	-
Yapım Eki	Fiilden Fiil	-	-	-	-	-	-	-	-	-	-	-	11
	Fiilden İsim	-	2	-	-	1	-	-	-	-	-	-	-
	İsimden İsim	-	-	-	3	-	3	1	-	-	-	-	-
Diğer Ekler	Birleşik Fiil	-	-	-	-	-	-	-	-	-	-	-	2
	Fiilden Sıfat Fiil	-	-	-	-	-	3	-	-	-	-	-	-
Çekim Eki	Hal Eki	-	-	1	-	-	-	-	-	-	-	-	-
	İyelik Eki	-	-	-	1	18	2	1	-	-	-	-	-
	Soru Eki	-	-	-	1	-	-	-	-	-	-	-	-
	Şahıs Eki	-	-	-	-	-	-	7	-	-	-	-	-
	Tasarlama Kipi	-	-	-	-	-	-	-	-	2	-	-	-
	Zaman Kipi	-	-	-	1	-	-	8	55	-	-	-	-

Türlerine göre kelime sayılarının verildiği Tablo 12'ye göre, en az edat, en fazla isim türünden kelime bulunmaktadır. İsim türünden kelimelerin fazla olması Türkçenin bir özelliğidir. Fakat bunun başka bir sebebi olarak, Zemberek'in bazı edat, bağlaç ve soru ifade eden kelimelerin türlerini isim olarak belirlemesi gösterilebilir.

Tablo 12. Türlerine göre kelime sayıları

Tür	Kelime Sayısı
Edat	1
Fiil	192
İsim	337
Özel	10
Sayı	20
Sıfat	100
Soru	23
Zaman	21
Zamir	40

Tablo 13'te ses uyum kurallarına uyan/uymayan kelime sayıları sunulmuştur. Kelimelerin %81'inin büyük ünlü, %91'inin küçük ünlü ve %99'ununsa ünsüz uyumuna uyduğu belirlenmiştir. Öte yandan kelimelerin %76'sının büyük ve küçük ünlü uyumuna birden, %81'inin büyük ünlü ve ünsüz uyumuna birden, %90'ının küçük ünlü ve ünsüz uyumuna birden, %75'ininse her üçüne birden uyduğu tespit edilmiştir.

Tablo 13. Ses uyum kurallarına göre kelime sayıları

Ses Uyum Kuralı	Uyan/Uymayan	Kelime Sayısı
Büyük Ünlü Uyumu	Uyan	606
	Uymayan	138
Küçük Ünlü Uyumu	Uyan	675
	Uymayan	69
Ünsüz Uyumu	Uyan	738
	Uymayan	6

VII. SONUÇLAR

Bu çalışmada 230 Türkçe cümle DDİ teknikleriyle analiz edilmiştir. Cümleleri oluşturan kelimelerin ses bilgisi ve morfolojik incelemelerinin yanı sıra DDİ kütüphanesi olan Zemberek eklerinin Türkçe dil bilgisindeki karşılıkları tespit edilmiştir. İşlemler, Zemberek’le entegre şekilde çalışan *Visual Studio* ortamında ve *C#* dilinde geliştirilen yazılımla gerçekleştirilmiştir. Cümlelerin ön işlemden geçirilmesi ve kelimelere ayrılması, kelimelerin harf ve hece sayılarının, hecelerın türlerinin ve harf sayılarının, kelimelerin aldıkları ek sayılarının, türlerine göre eklerin kullanım sayılarının, eklerin birbirini takip etme durumlarının, türlerine göre kelime sayılarının, kelimelerin ses uyum kuralına uyma/uymama durumlarının kontrolü, uyan/uymayan kelime sayılarının belirlenmesi işlemlerinde geliştirilen yazılım; kelimelerin denetlenmesi, kök/gövdelerin belirlenmesi, türlerinin saptanması, hecelenmesi, aldığı eklerin tespiti ve eklerine göre ayrıştırılmasında ise Zemberek kullanılmıştır.

Zemberek ek listesinde bulunan 130 ekten 120 tanesinin Türkçe karşılığı tespit edilmiştir. Zemberek eklerinin 1 tanesinin BOŞ_EK olarak isimlendirildiği; 13 tanesinin FIIL_KOK, ISIM_KOK ve EDAT_KOK şeklinde kelimenin türünü ifade eden kök adı olarak tanımlandığı; SORU_KOK, 13 kök adından biri olarak tanımlanmasına rağmen bu ekin çekim eklerinden soru ekinde yer alması uygun görülmüştür. Öte yandan Zemberek eklerinden 39 tanesinin yapım eki, 46 tanesinin çekim eki, 8 tanesinin birleşik fiil, 5 tanesinin fiilden sıfat fiil türeten ve 9 tanesinin fiilden zarf fiil türeten ekler oldukları belirlenmiştir.

Çalışmada 230 cümle ve bu cümleleri oluşturan 744 kelime incelenmiştir. Kelimelerin ortalama 5.30 harf ve 2.26 heceden meydana geldiği, hecelerın ağırlıklı olarak ünsüz-ünlü birleşiminden oluşan CV ve ünsüz-ünlü-ünsüz birleşiminden oluşan CVC hece türünde yer aldığı, hecelerın ortalama 2.34 harfli olduğu, kelimelerin yarısından fazlasının 1 çözümlemeli ve yaklaşık yarısının ek almadığı, çekim eklerinin diğer ek türlerine kıyasla çok daha fazla kullanıldığı, yapım eklerinden sonra yapım veya çekim eklerinin geldiği, diğer eklerden (birleşik fiil ve fiilden sıfat fiil türeten) veya çekim eklerinden sonra ise yalnızca çekim eklerinin geldiği, en çok isim türünden kelimelerin bulunduğu, küçük ünlü uyumuna uyan kelime sayısının büyük ünlü uyumuna uyan kelime sayısından fazla olduğu, ünsüz uyumuna uyan kelime sayısının da ünlü uyum kurallarına uyan kelime sayısından yüksek olduğu belirlenmiştir.

Bu çalışma kapsamında yapılan işlemler ve elde edilen veriler Büyük Dil Modelleri, anlam bilimsel analiz, söz dizimsel analiz, dil bilgisi analizi, hesaplamalı dil bilim ve Türkçe öğretimi araştırmalarında kullanılabilir. İlerleyen çalışmalarda daha fazla sayıda cümleye sahip veri kümesiyle çalışma yürütülebilir. Böylece, karşılığı belirlenememiş 130 ekten 10 tanesinin Zemberek eklerinin karşılıklarının tespit edilmesi de sağlanabilir. Öte yandan Zemberek, aldığı kelimenin muhtemel tüm çözümlemelerini sunmaktadır. Çekim eklerinin kullanımlarının cümle üzerinde detaylı analizi yapılarak olası çözümlemelerden hangisinin seçileceğinin karar verilmesine yönelik bir çalışma da gerçekleştirilebilir.

BİLGİ

Bu makale, 1. yazarın “Üç Boyutlu Sanal Model ile Türk İşaret Dili Simülasyonu” başlıklı doktora tezinden üretilmiştir.

KAYNAKLAR

1. ChatGPT (2024) An artificial intelligence model. <https://chat.openai.com/> Erişim 24 Mart 2024
2. Gemini (2024) An artificial intelligence model. <https://gemini.google.com/app> Erişim 24 Mart 2024
3. Gürbüz M, Sürmeli D, Taşkın K, Cebeci Hİ (2024) Otellere için paylaşılan çevre ile alakalı yorumların metin madenciliği ile analizi: Antalya otelleri üzerine bir araştırma. *Business and Management Studies: An International Journal* 12(1):218-239. <https://doi.org/10.15295/bmij.v12i1.2369>
4. Erdoğan İ, Güllü M, Polat H (2022) Makine Öğrenmesi algoritmaları ile uçtan uca yazar tanıma uygulaması geliştirme. *El-Cezeri* 9(4):1303-1314. <https://doi.org/10.31202/ecjse.1134698>
5. Görmez Y, Arslan H, Atak B (2024) Türkçe metinlerde duygu analizi: Derin öğrenme yaklaşımlarının ve ön işlem süreçlerinin model performansına etkisi. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi* 36(1):509-520. <https://doi.org/10.35234/fumbd.1429040>
6. Pekçoşkun Güner S (2023) Çevirmen-Bilgisayar etkileşiminin kilit bileşeni: Doğal dil işleme. *Karamanoğlu Mehmetbey Üniversitesi Uluslararası Filoloji ve Çeviribilim Dergisi* 5(1):56-79. <https://doi.org/10.55036/ufced.1306746>
7. Karaca A, Aydın Ö (2024) Transformatör mimarisi tabanlı derin öğrenme yöntemi ile Türkçe haber metinlerine başlık üretme. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi* 39(1):485-496. <https://doi.org/10.17341/gazimmfd.963240>
8. Google Translate (2024) <https://translate.google.com/> Erişim 28 Mart 2024
9. DeepL Translate (2024) <https://www.deepl.com/translator> Erişim 28 Mart 2024
10. Karaca MF (2018) Üç boyutlu sanal model ile Türk İşaret Dili simülasyonu. Yayımlanmış Doktora Tezi, Karabük: Karabük Üniversitesi Fen Bilimleri Enstitüsü.
11. Harman G, Aydemir E (2022) Kolay kişiselleştirilebilir akıllı sanal asistan. *International Journal of Multidisciplinary Studies and Innovative Technologies* 6(2):143-151. <https://doi.org/10.36287/ijmsit.6.2.143>
12. Ethnologue (2024) How many languages are there in the world? <https://www.ethnologue.com/insights/how-many-languages/> Erişim 29 Mart 2024
13. Toparlı R (2000) Türk dili ve kompozisyon (1. Baskı). Sivas: Baskı Dilek Ofset Matbaacılık, ss. 20, 64, 67.
14. Hengirmen M (2007) Türkçe dilbilgisi (9. Baskı). Ankara: Engin Yayınları, ss. 33, 35, 69-71, 111, 113.
15. Ergin M (1980) Türk dilbilgisi (5. Baskı). İstanbul: Boğaziçi Yayınları, ss. 39, 40, 67, 68, 72, 73, 104-106, 115, 121-124.
16. Paçacıoğlu B (2010) Türk dili ve kompozisyon (1. Baskı). İstanbul: Hat Yayınevi, ss. 32, 34, 64, 66, 67, 79.
17. Türk Dil Kurumu (2024) Ünsüzlerin nitelikleri. <https://tdk.gov.tr/icerik/yazim-kurallari/unsuzlerin-nitelikleri/> Erişim 29 Mart 2024
18. Türk Dil Kurumu (2024) Türk Dil Kurumu sözlükleri. <https://sozluk.gov.tr/> Erişim 29 Mart 2024
19. Boz E (2012) Türkiye Türkçesi biçimsel ve anlamsal işlevli biçimbilgisi (1. Baskı). Ankara: Gazi Kitabevi, s. 5.
20. Zemberek (2007) Açık kaynak kodlu Türkçe Doğal Dil İşleme Kütüphanesi. <http://code.google.com/p/nzemberek> Erişim 12 Temmuz 2015
21. Akın MD, Akın, AA (2007) Türk dilleri için açık kaynaklı doğal dil işleme kütüphanesi: Zemberek. *Elektrik Mühendisliği* 431:38-44.
22. Demirhan T, Hacıoğlu İ (2024) 6 Şubat 2023 Türkiye'deki deprem fırtınasının X (Twitter) özelinde tanımlayıcı analizlerinin yapılması. *Fırat Üniversitesi Sosyal Bilimler Dergisi* 34(1):285-300. <https://doi.org/10.18069/firatsbed.1384400>
23. Yakar Ö, Büyüktanır B, Çil A, Altınel Girgin AB (2024) Türkçe nefret söylemi problemi analizinde farklı sınıflandırma algoritmalarının ve özellik seçimi yöntemlerinin performans karşılaştırması. *Avrupa Bilim ve Teknoloji Dergisi* (53):97-111. <https://dergipark.org.tr/en/pub/ejosat/issue/82065/1373310>
24. Kayakuş M, Yiğit Açıkgöz F (2023) Twitter'da makine öğrenmesi yöntemleriyle sahte haber tespiti. *Abant Sosyal Bilimler Dergisi* 23(2):1017-1027. <https://doi.org/10.11616/asbi.1266179>
25. Kocak S, İç YT, Sert M, Dengiz B (2023) Ar-Ge projelerinin sınıflandırılması için doğal Türkçe dil işleme tabanlı yöntem. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi* 38(3):1375-1388. <https://doi.org/10.17341/gazimmfd.889395>
26. Aram K, Erdemir G, Can B (2021) Açık kaynak kod Türkçe doğal dil işleme kütüphanelerinin robotik uygulamalarda kullanımı. *İstanbul Sabahattin Zaim Üniversitesi Fen Bilimleri Enstitüsü Dergisi* 3(2):133-137. <https://doi.org/10.47769/izufbed.880143>