

# Türkçe Zamansal İfadelerin Yakalanması ve Tanımlanması

## Araştırma Makalesi/Research Article

 Hatice Kübra ÇAMCI,  Gülşen ERYİĞİT\*

Bilgisayar ve Bilişim Fakültesi, İstanbul Teknik Üniversitesi, İstanbul, Türkiye  
[camcih15@itu.edu.tr](mailto:camcih15@itu.edu.tr), [gulsen.cebiroglu@itu.edu.tr](mailto:gulsen.cebiroglu@itu.edu.tr)  
(Geliş/Received:04.01.2021; Kabul/Accepted:05.07.2021)  
DOI: 10.17671/gazibtd.853145

**Özet**— Zamansal ifadelerin yakalanması ve tanımlanması, anlamsal çıkarım gerektiren durumlar için kritik öneme sahip bir doğal dil işleme görevidir. Günümüzde yapay zeka alanında öne çıkan dil işleme teknolojileri pek çok araştırma ve uygulama evresinde zamansal ifadelerin çözümlenmesine ihtiyaç duymaktadır. Bu makalede, Türkçe için literatürde yer alan ilk zamansal ifade yakalama ve tanımlama sistemi tanıtılmaktadır. Yine literatürde ilk kez bu konuda takip eden çalışmalarda kullanılabilir bir sınıma veri kümesi oluşturulmuş ve araştırmacıların hizmetine sunulmuştur. Açık kaynak olarak geliştirilen sistem diğer diller için sıklıkla kullanılan HeidelTime mimarisi üzerine kurulmuş ve oluşturulan veri kümesi üzerinde dört farklı tür (tarih, saat, süre, tekrar belirten zaman ifadeleri) için ortalamada yüzde 90 civarında başarımlar elde edilmiştir. Sistem farklı literatürde yer alan farklı değerlendirme ölçütleri ile değerlendirilmiş ve temel bir Türkçe varlık ismi tanıma ve otomatik oluşturulmuş HeidelTime dil kaynakları ile karşılaştırılmıştır. Geliştirilen sistemin Türkçe doğal dil araştırmalarında eksik kalan önemli bir yapı taşını tamamladığı ve ileriki çalışmalara fayda sağlayacağı öngörülmektedir.

**Anahtar Kelimeler**— TimeML, TIMEX3, HeidelTime, doğal dil işleme, zamansal ifadeler

## Turkish Temporal Expression Extraction and Identification

**Abstract**— Temporal expression recognition and disambiguation is a critical stage for natural language processing tasks that require semantic inference. Nowadays, language processing technologies (a subfield of artificial intelligence) require the analysis of temporal expressions in many phases. In this article, a temporal expression recognition and disambiguation system for Turkish is presented for the first time in the literature. Additionally, again for the first time, a Turkish temporal expression test data set has been created and made publicly available for researchers. The introduced system was developed on HeidelTime architecture which is frequently used for other languages. On average, 90 percent performance was achieved for four different types of temporal expressions (date, time, duration, and set) on the introduced data set. The system performance is evaluated under different evaluation criteria and compared with a baseline named entity recognizer and HeidelTime automatically created language resources. It is anticipated that the system fills an important gap in the Turkish natural language studies so far and will benefit future studies.

**Keywords**— TimeML, TIMEX3, HeidelTime, natural language processing, temporal expressions

### 1. GİRİŞ (INTRODUCTION)

Zaman kavramı istikrar, düzen ve anlaşılabilirlik noktalarında hayatımız ve dünyayla iletişimimiz için büyük önem arz etmektedir. Bu durum zamanla ilgili her türlü ifadeyi, sözlü ve yazılı iletişimimizin ayrılmaz bir parçası yapmaktadır. Bu bağlamda, zamansal ifadelerin analizi doğal dil işlemenin önem arz eden bir çalışma alanı olmaktadır. Zamansal ifadeleri öncelikle yakalamak ve

ardından TimeML standartlarına [1] uygun olarak değer ve tür ataması yapmak bu analizin iki ana aşamasıdır. Bu alanın günlük yaşamımızdaki karşılığı da somut olarak görülmektedir. Örneğin cep telefonlarımızdaki sanal asistanların takvimlerimiz üzerinde yapabildiği işlemler (konuşma tanıma yoluyla toplantı ekleme, silme vb.) bu analizin sonuçlarından biridir.

Başta İngilizce olmak üzere birçok dil için zamansal ifadelerin analizi üzerine çalışmalar yapılmış ve yapılmaktadır [2-10]. Bunun yanında, özellikle İngilizce odaklı olmak üzere birçok zamansal ifade etiketleyicisi bulunmaktadır; DANTE [11], Trips/Trios [12], Heidelberg [13], SUTime [14] bunlardan bazılarıdır. Maalesef ki Türkçe için bu konuda sonuca ulaşmış bir akademik çalışma ve dolayısıyla araştırmacıların kullanımına açık bir etiketleyici bulunmamaktadır. Fiil zamanlarının tespiti [15] ve olay sıralaması [16] metindeki zamanların belirlenme konusu ile ilgili tamamlanmış nadir çalışmalardandır. Bu makaledeki ana mesele olan zamansal ifadelerin analizi ile ilgili de tamamlanmamış olsa da bazı çalışmalar bulunmaktadır. Bunlardan en önemlisi Küçük ve Küçük (2015)'ün çalışmasıdır [17]. Bu çalışmada, zamansal ifadelerle ilgili sözlük ve örüntü örnekleri verilmiş fakat çalışma sadece ön analiz aşamasında kalmıştır. Şeker ve Eryiğit (2017) [18] Türkçe varlık ismi tanıma (VİT) çalışmalarında kapsadıkları varlık ismi türlerini zamansal ifadeler ile genişlettiklerini ve bu konuda veri kümesi [18,19] işaretlediklerini belirtmişlerdir. Bu çalışmada hem tanımlama kısmının olmayışı hem de kullandıkları metinlerin kısıtlı doğası ve boyutunun (gazete haberleri, romanlar vb.) sonucu olarak az sayıda ve türde (tarih ve saat) ifade etiketlenmesi nedeniyle dünya literatüründe bu konuda yapılan ayrıntılı çalışmaların çok gerisinde kaldığı görülmektedir.

Bu çalışmada, Türkçe zamansal ifade analizi alanındaki yukarıda aktarılan eksiklikten yola çıkılarak Türkçe için zamansal ifade yakalayıcı ve tanımlayıcı oluşturmak amaçlanmaktadır. Bu doğrultuda açık kaynaklı, kural tabanlı zamansal ifade etiketleyici olan Heidelberg mimarisi üzerinde Türkçe kaynak geliştirilmesi yöntemi benimsenmiştir. Makalede, Türkçe için literatürde yer alan ilk zamansal ifade yakalama ve tanımlama sistemi tanıtılmaktadır. Yine literatürde ilk kez bu konuda takip eden çalışmalarda kullanılacak bir sınıma veri kümesi oluşturulmuş ve araştırmacıların hizmetine sunulmuştur. Açık kaynak olarak geliştirilen sistem<sup>1</sup> ile oluşturulan veri kümesi üzerinde dört farklı tür (tarih, saat, süre, tekrar belirten zaman ifadeleri) için ortalamada %90'ın üzerinde başarımlar elde edilmiştir. Geliştirilen sistemin Türkçe doğal dil araştırmalarında eksik kalan önemli bir yapı taşını tamamladığı ve ileriki çalışmalara fayda sağlayacağı öngörülmektedir.

Makalenin aşağıda yer alan bölümlerinde, ilk olarak Heidelberg ve çalışma yapısı açıklanacak, ardından Metod ve Geliştirme Süreci bölümünde Türkçe zamansal ifade yakalayıcısı ve tanımlayıcısından, oluşum aşamalarından bahsedilecek, son olarak Deneyler ve Analiz bölümünde değerlendirme ve sonuç analizi verilecektir.

## 2. HEIDELTIME (HEIDELTIME)

Heidelberg [13], Heidelberg Üniversitesi Databases System Research grubu tarafından geliştirilmiş, farklı

dillere uyarlanabilen, açık kaynaklı, kural tabanlı bir zamansal ifade etiketleyicisidir. Herhangi bir dilden bağımsız, Java dilinde geliştirilen Heidelberg, dile özgü kaynak dosyaları ile desteklenerek o dilde yazılan metinlerdeki zamansal ifadeleri yakalayıp TimeML-TIMEX3 standartlarına uygun olarak ifadelere değer ve tür ataması yapmaktadır. Şu ana kadar 12 dil (İngilizce, Almanca, Flemenkçe, İspanyolca, İtalyanca, Arapça, Vietnamca [20], Fransızca [21], Hırvatça [22], Çince [23], Portekizce ve Estonca) için kaynak dosyaları insanlar tarafından geliştirilmiş, 200'ün üzerinde dil için ise kaynak dosyaları otomatik olarak oluşturulmuştur. Otomatik oluşturma İngilizceden çeviri yolu ile gerçekleştirilmiş olduğundan bu şekilde kapsanan diller üzerinde zamansal ifade yakalama ve tanımlama başarımları çok düşük ölçülmektedir.

Heidelberg'in dil kaynakları 3 bölümden oluşmaktadır. Etiketlenmesi istenen zamansal ifadelerin listelendiği *repattern* (LİSTE), TIMEX3 standartlarına uygun olarak değer tanımlarının yapıldığı *normalization* (NORMALİZASYON), LİSTE sınıfları ve yardımcı öğeler kullanılarak kalıplaştırılıp bu kalıplara NORMALİZASYON sınıfları yardımıyla değer atamalarının yapıldığı kuralları içeren *rules* (KURALLAR). LİSTE bölümünde zamansal ifadeleri düzenli ifade şeklinde kalıplaştırabilmek için uygun sınıflar oluşturulur ve tanımlanması istenen ifadeler listelenir, örneğin mevsim veya haftanın günleri sınıfı gibi. NORMALİZASYON da ise LİSTE bölümünde tanımlanan zamansal ifadelerin veya KURALLAR'daki kalıpların değer atamasına yardımcı olacak ifadelerin TIMEX3 standardına uygun olarak değerleri listelenir, örneğin Mayıs ayı için "05" gibi.

Etiketlenen zamansal ifadeye atanacak bilgiler kural oluştururken belirlenir. Her zamansal ifadenin taşıması gereken iki bilgi vardır: değer ve tür. TimeML-TIMEX3 standartlarına göre zamansal ifadeler 4 türe ayrılmaktadır; tarih (*date*), saat (*time*), süre (*duration*), tekrar belirtenler (*set*) [1]. KURALLAR bölümünde yer alan örüntü kuralları bu sınıflandırmaya uygun olarak 4 sınıfta listelenir ve ifadenin türü kuralın bulunduğu sınıfa göre atanır.

Kural oluşturmak için;

- Kural ismi verilir.
- İfadenin kalıbı düzenli ifade biçiminde, LİSTE sınıfları, sembol, noktalama işareti gibi yardımcı öğeler, gerektiğinde ifadenin kendisi ya da belli bir parçası kullanılarak oluşturulur.
- İfadenin değeri NORMALİZASYON sınıfları, ifadenin kendisi ya da sabit bir değer kullanılarak belirlenir. İfadenin kalıbındaki LİSTE sınıfları ve düzenli ifade karakter sınıfları sırasına göre group1, group2 vb. şeklinde adlandırılır. Değer

<sup>1</sup>Şu anda özel kod deposunda <https://bitbucket.org/camcihkubra/heideltime/src/master> bulunan sistem ve veri kümesi, makalenin kabulünü

takiben Heidelberg kod deposuna makale referansı ile birlikte yerleştirilecek ve bu bölüm ilgili adres ile güncellenecektir.

ataması yaparken ifadenin kendisi kullanılacaksa ya da NORMALİZASYON sınıfına parametre olarak gönderilecekse bu adlandırmalar kullanılır.

Şekil 1'de tarih sınıfına ait bir kural verilmiştir. Bu kural örneğinde "gün-ay-yıl" kalıbına ve "." veya "/" sembollerine sahip ifadeler için düzenli ifade oluşturulmuş; bunun için ayın günlerinin numaralarının listelendiği 'reDayNumber',

```
RULENAME="date_r0h",
EXTRACTION="%reDayNumber[./]%reMonthNumber[./]%reYear4Digit",
NORM_VALUE="group(3)-%normMonth(group(2))-%normDay(group(1))"
```

Şekil 1. Tarih sınıfına ait kural örneği  
(A rule example of date type)

```
The Wakefield gallery opened on the banks of the River Calder in <TIMEX3 tid="t4" type="DATE" value="2011-05">May 2011</TIMEX3> to replace the city's former art gallery which opened in <TIMEX3 tid="t2" type="DATE" value="1934">1934</TIMEX3>. [24]
```

Wakefield galerisi, şehrin 1934'te açılan eski sanat galerisinin yerini almak üzere Mayıs 2011'de Calder Nehri kıyısında açıldı.

Şekil 2. İngilizce metin örneğinde HeidelTime çıktısı  
(Example of HeidelTime output on English text)

### 3. METOT VE GELİŞTİRME SÜRECİ (METHOD AND DEVELOPMENT PROCESS)

Öncelikle, HeidelTime'da Türkçe için yukarıda söz edilen çeviri yöntemi ile otomatik olarak oluşturulmuş kaynak dosyaları incelenmiştir. Türkçe için otomatik oluşturulmuş 26 LİSTE sınıfı, 25 NORMALİZASYON sınıfı ve 25'i tarihler, 6'sı saatler, 2'si süreler ve 4'ü tekrar belirtenler için olmak üzere 37 kural bulunmaktadır. Yapılan incelemelerde, otomatik oluşturulan dil kaynaklarının genel olarak aşağıdaki konularda eksik veya yanlış sonuç verdikleri belirlenmiştir:

- Ek alan ifadelerde sonuç üretilememesi,
- Günün bölümleri dışında saat ifadelerine dair herhangi bir sınıf veya kural bulunmaması,
- LİSTE ve dolayısıyla NORMALİZASYON sınıflarındaki ifade eksiklikleri,
- Özel gün ve bayramlarla ilgili kaynak bulunmaması,
- Karmaşık ifadeleri çözümleyebilecek kalıplara yer verilmemiş olması.

Yukarıda listelenen eksiklerin tamamlanmasının ardından, Türkçeye özgü çalışılması gereken aşağıdaki konular belirlenmiştir:

- Türkçe sondan eklemeli dil grubuna dahil olduğundan ekler dilimiz için dolayısıyla etiketleme için de büyük önem taşımaktadır [25]. Ekler;
  - Kelime kökünde değişime sebep olmakta, "ayın dördü"
  - Varlığına göre zamansal ifadenin türünde değişime sebep olmakta; "780 yıl" ve "780 yılında" örneklerinde biri süre ifade ederken ek alan tarih belirtmektedir.

ay numaralarının listelendiği 'reMonthNumber' ve basamaklı yılların listelendiği 'reYear4Digit' LİSTE sınıfları kullanılmış, '%' sembolüyle bu sınıfların kullanıldığı belirtilmiştir. İfadenin değerini belirlemek için kalıptaki 3. grup yani yıl ifadesi direkt olarak, sırasıyla 2. ve 1. grubu parametre alan 'normMonth' ve 'normDay' NORMALİZASYON sınıfları kullanılmıştır.

– Türüne göre ifadenin türünü de değiştirmektedir. "6'da görüşürüz", "6'sında görüşürüz" örneklerinde ilk cümledeki ifade saat belirtirken diğeri tarih belirtmektedir.

- Kelime dizilişleri kelime grubunun anlamı dolayısıyla türü üzerinde etkili olmaktadır; "yıl 905" tarih, "905 yıl" süre anlamı taşımaktadır.
- Aynı kelime grubu diziliş açısından birçok kombinasyona sahip olabilmektedir. Örneğin; "2 Mayıs 1998", "1998 2 Mayıs" "Mayıs'ın 2'si 1998", "1998 Mayıs'ının 2'si" ifadeleri aynı kelime gruplarıyla oluşturulmuştur ve hepsi aynı değere sahiptir.

Belirtilen noktalar göz önünde bulundurularak sınıflar ve KURALLAR oluşturulmuştur. Bunların dışında, diğer HeidelTime dil kaynaklarından farklı olarak, İngilizce alfabe kullanılarak yazılan Türkçe ifadeler (örneğin "Bugün" yerine "Bugun" yazılması) ve kelimeleri arasında birden fazla boşluk bulunan ya da hiç boşluk bulunmayan ifadeler için de KURALLAR ekleyerek çalışmanın kapsamı genişletilmiştir.

Yapılan ekleme ve düzenlemelerin sonucunda 60 LİSTE, 94 NORMALİZASYON sınıfına ve 211'i tarih, 198'i saat, 11'i süre ve 18'i tekrar belirten ifadeler için olmak üzere 438 kurala sahip Türkçe kaynak dosyaları oluşturulmuş ve Türkçe zamansal ifade yakalama ve tanımlama sistemi tamamlanmıştır.

Oluşturulan sistemi değerlendirmek üzere haber metinleri, yerli/çeviri edebiyat, blog yazıları gibi farklı tür kaynaklardan toplanan ve Küçük ve Küçük (2015) [17], Sauri vd. (2006) [26] çalışmalarında verilen örüntülere göre oluşturduğumuz örneklerle bir sınama kümesi oluşturulmuş ve bu küme ISO-TimeML standartlarını baz alarak hazırladığımız Türkçe için ISO-TimeML işaretleme kılavuzuna göre elle işaretlenmiştir. Bu sınama veri

kümesinde 159 tarih, 72 saat, 53 süre, 24 tekrar belirten toplamda 308 ifade<sup>2</sup> bulunmaktadır. Bu sınama kümesinin yanı sıra, kurallar yazılırken kuralları doğrulamak için oluşturulmuş 426 ifadeli bir doğrulama veri kümesi de bu çalışmayla beraber araştırmacıların hizmetine sunulmuştur. Bir sonraki bölümde verilen değerlendirmeler sınama veri kümesi üzerindeki sistem başarımlarını sunmaktadır.

Şekil 3'de Türkçe metin için HeidelTime çıktı örneği, Şekil 4'te ise bu çıktı örneğini oluşturmak için kullanılan kurallardan bazıları verilmiştir. Sol üstte verilen doküman oluşum tarihi yani DOT, metindeki ifadeler için referans olacak tarihtir. İlk cümlede geçen "İki gün sonra" ifadesi tarih belirten zamansal ifade olduğu için yakalanmış ve tür

(*type*) bilgisi tarih (*date*), referans tarihe göre "İki gün sonra" 27.11.2020 tarihini işaret ettiği için değer (*value*) bilgisi de TimeML standartlarına uygun olarak "2020-11-27" (kural *date\_r29e*, Şekil 4) şeklinde atanmış ve tanımlama tamamlanmıştır. Bu değer ataması ile birlikte metnin başında "2020-11-25" olan referans tarihi "2020-11-27" olarak değişmiştir. Diğer cümle grubunda ise "saat dokuza bir mi vardı", "saat tam dokuz" ve "saat dokuzda" ifadeleri saat belirten zamansal ifade oldukları için yakalanmış ve tür bilgisi saat (*time*) olarak atanmıştır. TimeML standartlarına göre saat belirten zamansal ifadelerle değer ataması yaparken önce tarih ardından "T" harfiyle saat bilgisi verilir. Buna uygun olarak ifadelerin değer bilgisi ilk ifade için "2020-11-27T08:59", diğer ifadeler için "2020-11-27T09:00" (kural *time\_r16a*, Şekil 4) olarak atanmış ve tanımlama tamamlanmıştır.

```
DOT: 25.11.2020

<TIMEX3 tid="t1" type="DATE" value="2020-11-27">İki gün sonra</TIMEX3> yirmi kişiye: "O benim bileceğim şey" ne manaya gelir diye sordum. Hiçbiri doğru dürüst bir mana veremedi.

...

Siz ne fikirdesiniz? Sağdan şu kadar, faraza doksan sekiz adım yürürsem Taksim sinemasının önünde olacağım. Evden çıkarken <TIMEX3 tid="t2" type="TIME" value="2020-11-27T08:59">saat dokuza bir mi vardı</TIMEX3>? Ağır yürüyorum. Şimdi <TIMEX3 tid="t3" type="TIME" value="2020-11-27T09:00">saat tam dokuz</TIMEX3> olmalı. Börekçi Mahmut'un <TIMEX3 tid="t4" type="TIME" value="2020-11-27T09:00">saat dokuzda</TIMEX3> dükkândan ayrıldığı görülmemiştir. [27]
```

Şekil 3. Türkçe metin örneğinde sistem çıktısı  
(Example of the system output on Turkish text)

```
RULENAME="date_r29e",
EXTRACTION="%reNumber(\s)*%reUnit(\s)*[Ss]onra%reSuffix*",
NORM_VALUE="UNDEF-REF-%normUnit(group(3))-PLUS-%normNumber(group(1))"

RULENAME="time_r16a",
EXTRACTION="[Ss]aat(%rePartOfDayforTime|%reAdverb|(\s)|%reSuffix*)*%reTimeHour"?%reSuffix*",
NORM_VALUE="UNDEF-REF-day-PLUS-0T%normTimeMinute(group(6)):00"
```

Şekil 4. Tarih ve saat sınıfına ait kurallar  
(Rules of date and time types)

#### 4. DENEYLER VE ANALİZ (EXPERIMENTS AND ANALYSIS)

Bu bölümde sistemi değerlendirmek amacıyla oluşturulan sınama veri kümesi üzerinde 2 farklı değerlendirme sonucu verilmiştir. Bunlardan ilki Semeval 2013 çalıştayında [28] kullanılan ölçütler temel alınarak hesaplanan varlık bazlı kesinlik (*precision*), hassasiyet (*recall*), F1 skorları, tür bazlı F1 ve değer bazlı F1 skorlarıdır. Varlık bazlı değerler esnek (yakalanan ifadeyle asıl ifade arasında sınırları birebir aynı olmasa da çakışma olması ve yakalanan ifadenin TIMEX olarak etiketlenmesi, örneğin yakalanan "2 Temmuz" zamansal ifadesi asıl ifade olan "2 Temmuz sabahı" ile karşılaştırıldığında iki ifadenin sınırları aynı olmasa da "2 Temmuz" grubunu ortak olarak barındırmaları yakalanan ifadenin esnek eşleşme altında

değerlendirilmesini gerektirmektedir) ve tam (yakalanan ifadenin asıl ifadeyle birebir aynı olması ve yakalanan ifadenin TIMEX olarak etiketlenmesi) eşleşme durumları için ayrı ayrı hesaplanmıştır. Tablo 1'de ilk iki sütun bloğu altında bu değerler sunulmaktadır. Tür bazlı F1 skorları (Tablo 1 3. sütun bloğunda "Tür F1"), esnek değerlendirme altında bulunan TIMEX varlıklarının aynı zamanda doğru türe atanmış olmalarını göz önüne alarak değerlendirme yapar. Değer bazlı F1 skorları (Tablo 1 3. sütun bloğunda "Değer F1"), esnek değerlendirme altında bulunan TIMEX varlıklarının aynı zamanda doğru değer ataması yapılmış olmalarını göz önüne alarak değerlendirme yapar. İkinci değerlendirmede ise Moriceau ve Tannier (2013)'deki detaylı sonuç bilgisine benzer olarak [21], zamansal ifadelerin türlerine göre yakalama ve tanımlama başarı yüzdeleri verilmiştir (Tablo 2). Yakalama başarı ölçütü

<sup>2</sup> Buradaki ifade kavramı hem kısa zamansal ifadeleri hem de varsa bunların içerisinde bulunduğu cümleleri kapsamaktadır.

ifadenin tam yakalanması; tanımlama başarı ölçütü ise esnek yakalama durumunda tür ve değer atamalarının aynı anda doğru yapılmasıdır. Örneğin, "Mayıs ayında" ifadesi

"Mayıs" olarak yakalanmış ve tür olarak "tarih", değer olarak da "05" atanmışsa etiketlemenin sonucu hatalı yakalama-doğru tanımlama olarak değerlendirilmektedir.

Tablo 1. Sistem başarımları  
(System performances)

	Esnek			Tam			Esnek	
	P	R	F1	P	R	F1	Tür F1	Değer F1
HeidelTime Temel Sistemi [13]	0,95	0,51	0,67	0,33	0,18	0,23	0,45	0,28
Türkçe VİT [18]	1,00	0,10	0,19	0,24	0,02	0,04	0,13	-
Türkçe Zamansal İfade Yakalayıcı-Tanımlayıcı	0,99	0,97	0,98	0,95	0,93	0,95	0,98	0,92

Tablo 2. Türe bağlı detaylı sonuçlar  
(Detailed results by type)

	HeidelTime Temel Sistemi [13]		Türkçe VİT [18]	Türkçe Zamansal İfade Yakalayıcı-Tanımlayıcı	
	Yakalama	Tanımlama	Yakalama	Yakalama	Tanımlama
Tarih	0,18	0,24	0,06	0,93	0,92
Saat	0,05	0,05	0,10	0,92	0,92
Süre	0,37	0,37	-	1,00	0,88
Tekrar Belirten İf.	0,25	0,25	-	1,00	0,95

#### 4.1. Deneyler (Experiments)

Geliştirilen sistemin performansı iki temel sistem ile karşılaştırılmıştır. Bunlardan ilki HeidelTime otomatik oluşturulmuş Türkçe sistemi, ikincisi ise zamansal ifadeleri yakalayan bir Türkçe VİT [18] aracıdır. Daha önceki bölümlerde belirtildiği üzere, bu ikinci temel sistemin tanımlama özelliği olmayıp sadece yakalama özelliği bulunmaktadır

Tablo 1’de oluşturulan sınama veri kümesi üzerinde bu iki temel sistemin ve bu makalede sunduğumuz yeni Türkçe zamansal ifade yakalayıcı ve tanımlayıcısının esnek ve tam eşleşmelere göre hassasiyet, kesinlik ve F1 skorları, esnek eşleşme durumuna ait tür ve değer F1 skorları, Tablo 2’de ise bu sistemlerin türe dayalı detaylı başarımlar oranları verilmiştir.

HeidelTime temel sistemi çeviri yoluyla oluşturulması sebebiyle sadece basit ve yaygın ifadeleri tanımakta, eğer bu ifadeler ek aldysa neredeyse hiç sonuç üretmemektedir. Bunun neticesinde, bu sistemin ürettiği sonuçlar içerisinde, en yüksek başarı elde edilen sınıflar daha basit ve tekdüze ifadelerle sahip olan süre (0,37) ve tekrar belirten ifade (0,25) sınıflarıdır. Saat ifadelerine neredeyse hiç yer verilmemesi dolayısıyla saat ifadelerinde elde edilen başarımlar oranı (0,05) oldukça düşüktür. Tarih kategorisinde ise normalden farklı olarak tanımlama başarı oranı (0,24) yakalama başarı oranından (0,18) daha yüksek elde edilmiştir. Bu durum ise belirtildiği gibi basit yapıları kapsamasından dolayı bazı örnekleri yakalamada eksik kalması ama yakaladığı kısım değer ve tür ataması konusunda yeterli geldiği için doğru tanımlama yapabilmemesinden kaynaklıdır. Genele bakıldığında HeidelTime, esnek eşleşmede kendi içinde başarılı bir sonuç elde ederken (P = 0,95) tam eşleşmede başarılı bir sonuç elde edememiştir. Tam eşleşmedeki yetersizlik tür ve değer F1 skorlarının düşük elde edilmesine sebep olmuştur.

Şeker ve Eryiğit (2017)’in Türkçe varlık ismi tanıma sistemleri [18] için kapsama dahil olan tarih ve saat ifadelerinin yakalama sonuçları verilmiştir. Bu kategorilerde ise sadece daha açık ve belirli tip ifadelerle yanıt vermesi, bununla birlikte bu ifadeleri tam olarak yakalamada eksik kalması ortalamada %10’un altında başarımlar oranı elde edilmesine sebep olmuştur. Genelde ise HeidelTime’a benzer olarak esnek eşleşmede kendi içinde başarılı (P = 1,00) ama tam eşleşmede ve dolayısıyla Tür F1 skorunda başarısız bir sonuç elde edilmiştir.

Bu makalede sunulan sistem ise diğer ikisine kıyasla gözle görülür ölçüde daha iyi bir başarı oranına sahiptir. HeidelTime yapısı içinde çözümlenemeyen bazı ifadelerin kurallarının karışıklığına sebep olmaması sebebiyle kapsam dışı bırakılması ve bazı durumlarda Türkçenin dil bilgisel yapısının HeidelTime sistemi ile birleştirilememesi nedenleriyle sistem belirli noktalarda eksik kalsa da bu dört kategorinin ortalamasında %90’ı aşkın başarı oranı ve genelde ise dikkate değer sonuçlar (Tür F1 = 0,98, Değer F1 = 0,92) elde edilmiştir.

#### 4.2. Hata Analizi (Error Analysis)

Örnek 1: Yanlış hatırlamıyorsam <TIMEX3 tid="t1" type="DURATION" value="P3Y">iki ya da üç yıl</TIMEX3> Avustralya’da çalışmıştı.

Bir önceki bölümde bahsedildiği üzere HeidelTime’ın yapısı bazı ifadelerin çözümüne olanak vermemektedir. Örnek 1’de görülen ifade "iki" kelimesi "iki yıl" anlamını taşımakta, dolayısıyla yakalanması ve tanımlanması gerekmektedir. "iki" ifadesinin yıl anlamı taşıdığından sistem tarafından anlaşılabilmesi için ya kendinden sonra gelen sözcükler kontrol edilip ona göre işaretlenmeli ya da "iki ya da 3 yıl" ibaresi bir bütün olarak işaretlenmelidir. İlk seçenek HeidelTime çalışma yapısına uymadığı, ikinci seçenekte ise bir yakalamadan iki farklı değer ataması yapılamayacağı için ifadenin tamamı yakalanıp ilk değer

göz ardı edilmektedir, bu durumda ilk değerin yakalanması başarılı olsa da değer atama kısmı eksik kalmaktadır.

Örnek 2:

DOT: 30.11.2020

<TIMEX3 tid="t1" type="DATE" value="2020-10-29">Cumhuriyet Bayramından</TIMEX3> <TIMEX3 tid="t1" type="DATE" value="2020-11-8">10 gün sonra</TIMEX3> dönmüş olur.

Örnek 2'de ise ifade sistem tarafından hatalı biçimde ikiye bölünerek işaretlenmiştir. Yazılan kural yapılarında, bu tür bir ifadeye doğru değer ataması yapılabilmesi için öncelikle referans tarihinin Cumhuriyet Bayramı olarak güncellenmesi gerekmektedir. Sonrasında yeni referans tarihinden 10 gün ileriye gidilerek doğru değer bulunur. Bu durum, HeidelTime'in etiketleme sırasında referans tarihin değiştirilememe kısıtlamasından kaynaklanmaktadır.

Yukarıda verilen örnekler dışında Türkçe dil bilgisi kuralları sebebiyle yanlış çözümlenen örnekler de mevcuttur. Aşağıda verilen Örnek 3'te fiil gelecek zaman kipi taşıdığı için "bu salı" ifadesi önümüzdeki salı anlamı taşımaktadır, fakat Örnek 4'te fiildeki geçmiş zaman kipi, "bu salı" ifadesine geçtiğimiz salı anlamı katmaktadır. Bu tarz bir çözümleme henüz çalışmanın kapsamı dışında olduğu için 3. örnekte doğru, 4. örnekte ise yanlış sonuç elde edilmektedir.

DOT: 02.12.2020

Örnek 3: <TIMEX3 tid="t1" type="DATE" value="2020-12-08">Bu salı</TIMEX3> önceğimiz tatilden.

Örnek 4: <TIMEX3 tid="t1" type="DATE" value="2020-12-08">Bu salı</TIMEX3> döndük tatilden.

## 5. SONUÇ (CONCLUSION)

Bu makalede, HeidelTime mimarisi üzerinde geliştirilen, Türkçe metinlerdeki zamansal ifadeleri yakalayıp tanımlanmasına olanak veren dil kaynakları ve bu kaynakların değerlendirme sonuçları sunuldu. Türkçe için bir ilk olan sistemle, zamansal ifadeleri yakalama ve tanımlama noktalarında dikkate değer başarı oranları elde edildi. Derleş yapısı itibarı ile çok farklı türde zamansal ifadelerin bir araya toplandığı ve yine Türkçe için bir ilk olma özelliği taşıyan sına veri kümesi sunuldu. Bu kümenin bu konuda yapılacak kural tabanlı veya veriye dayalı gelecek çalışmalar için faydalı bir sına kümesi olması öngörülmektedir. Gelecek çalışmalarda bu veri kümesinin gerçek dünya örnekleri ile genişletilerek daha büyük sına kümeleri oluşturulması Türkçe çalışmalar için büyük önem teşkil etmektedir.

İleriki aşamalarda, Türkçenin dil bilgil özellikleri, fiil kipleri, anlam kaymaları gibi konuların daha detaylı olarak dahil edilmesi ile sistemin kapsamı genişletilecek ve Türkçenin dil bilgil yapısına daha uygun bir sistem sunularak başarı artışı hedeflenecektir.

## KAYNAKLAR (REFERENCES)

- [1] J. Pustejovsky, R. Ingria, R. Sauri, J. M. Castaño, J. Littman, R. J. Gaizauskas, A. Setzer, G. Katz, I. Mani, "The Specification Language TimeML", **The Language of Time: A Reader**, Oxford University Press, UK, 2015.
- [2] M. Verhagen, R. Gaizauskas, F. Schilder, M. Hepple, J. Moszkowicz, J. Pustejovsky, "The tempEval challenge: Identifying temporal relations in text", *Language Resources and Evaluation*, 43(2), 161-179, 2019.
- [3] B. Boguraev, J. Pustejovsky, R. Ando, M. Verhagen, "TimeBank evolution as a community resource for TimeML parsing", *Language Resources and Evaluation*, 41(1), 91-115, 2007.
- [4] A. Setzer, R. Gaizauskas, M. Hepple, "The role of inference in the temporal annotation and analysis of text", *Language Resources and Evaluation*, 39(2-3), 243-265, 2005.
- [5] H. Llorens, E. Saquete, B. Navarro-Colorado, "Applying semantic knowledge to the automatic processing of temporal expressions and events in natural language", *Information Processing and Management*, 49(1), 179-197, 2013.
- [6] M. Navas-Loro, E. Filtz, V. Rodriguez-Doncel, A. Polleres, S. Kirrane, "TempCourt: Evaluation of temporal taggers on a new corpus of court decisions", *Knowledge Engineering Review*, 2019.
- [7] J. Kocoń, M. Marcińczuk, "Supervised approach to recognise polish temporal expressions and rule-based interpretation of timexes", *Natural Language Engineering*, 23(3), 385-418, 2017.
- [8] R. Gennari, P. Vittorini, "Qualitative temporal reasoning can improve on temporal annotation quality: How and why", *Applied Artificial Intelligence*, 30(7), 690-719, 2016.
- [9] R. M. Reeves, F. R. Ong, M. E. Matheny, J. C. Denny, D. Aronsky, G. T. Gobbel, ... S. H. Brown, "Detecting temporal expressions in medical narratives", *International Journal of Medical Informatics*, 82(2), 118-127, 2013.
- [10] R. Gennari, S. Tonelli, P. Vittorini, "Challenges in quality of temporal data - starting with gold standards", *Journal of Data and Information Quality*, 6(2), 2015.
- [11] P. Mazur, R. Dale, "The DANTE temporal expression tagger", **Human Language Technology. Challenges of the Information Society: Third Language and Technology Conference, LTC 2007, Poznan, Poland, October 5-7, 2007, Revised Selected Papers**, 5603, Springer, 245-257, 2007.
- [12] N. UzZaman, J. Allen, "TRIPS and TRIOS system for TempEval-2: Extracting temporal information from text", **Proceedings of the 5th International Workshop on Semantic Evaluation**, Association for Computational Linguistics, 276-283, 2010.
- [13] J. Strötgen, M. Gertz, "Heideltime: High quality rule-based extraction and normalization of temporal expressions", **5th International Workshop on Semantic Evaluation**, Association for Computational Linguistics, 321-324, 2010.
- [14] A. X. Chang, C. D. Manning, "Sutime: A library for recognizing and normalizing time expressions", **Eight International Conference on Language Resources and Evaluation**, European Language Resources Association, 3735-3740, 2012.

- [15] Ş.E. Şeker, B. Diri, "TimeML and Turkish Temporal Logic", **2010 International Conference on Artificial Intelligence**, 10, 881-887, 2010.
- [16] Ş. E. Şeker, B. Diri, "Türkçe Metinler için Olay Sıralaması", *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 4(1), 2016.
- [17] D. Küçük, D. Küçük, "On TimeML-Compliant Temporal Expression Extraction in Turkish", *arXiv preprint arXiv:1509.00963*, 2015.
- [18] G. A. Şeker, G. Eryiğit, "Extending a CRF-based named entity recognition model for Turkish well formed text and user generated content", *Semantic Web*, 8(5), 625-642, 2017.
- [19] U. Sulubacak, G. Eryiğit, "Implementing universal dependency, morphology, and multiword expression annotation standards for Turkish language processing", *Turkish Journal of Electrical Engineering & Computer Sciences*, 26(3), 1662-1672, 2018.
- [20] J. Strötgen, A. Armiti, T. V. Canh, J. Zell, M. Gertz, "Time for more languages: Temporal tagging of Arabic, Italian, Spanish, and Vietnamese", *ACM Transactions on Asian Language Information Processing (TALIP)*, 13(1), 1-21, 2014.
- [21] V. Moriceau, X. Tannier, "French resources for extraction and normalization of temporal expressions with HeidelTime", **Ninth International Conference on Language Resources and Evaluation (LREC'14)**, European Language Resources Association (ELRA), 3239-3243, 2014.
- [22] L. Skukan, G. Glavaš, J. Šnajder, "HEIDELTIME. HR: extracting and normalizing temporal expressions in Croatian", **9th Slovenian Language Technologies Conferences (IS-LT 2014)**, 99-103, 2014.
- [23] H. Li, J. Strötgen, J. Zell, and M. Gertz, "Chinese temporal tagging with HeidelTime", **14th Conference of the European Chapter of the Association for Computational Linguistics, volume 2: Short Papers**, Association for Computational Linguistics, 133-137, 2014.
- [24] Internet: Hepworth art gallery announces plans for 10th birthday exhibition, <https://www.bbc.com/news/uk-england-leeds-55116749#:~:text=The%20Hepworth%20art%20gallery%20has,gallery%20which%20opened%20in%201934,30.11.2020>.
- [25] A. B. Ercilasun, "Türkçenin dünya dilleri arasındaki yeri", *Dil Araştırmaları*, 12(12), 17-22, 2013.
- [26] R. Sauri, J. Littman, B. Knippen, R. Gaizauskas, A. Setzer, and J. Pustejovsky, **TimeML Annotation Guidelines Version 1.2. 1**, 2006.
- [27] S. F. Abasıyanık, **Alemdağ'da Var Bir Yılan**, YKY, 2010.
- [28] N. UzZaman, H. Llorens, L. Derczynski, J. Allen, M. Verhagen, and J. Pustejovsky, "Semeval-2013 task 1: Tempeval-3: Evaluating time expressions, events, and temporal relations", **Second Joint Conference on Lexical and Computational Semantics (\*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)**, Association for Computational Linguistics, 1-9, 2013.