

Türkçe Çok Sözcüklü Fiil İfadeleri Derlemi

Turkish Verbal Multiword Expressions Corpus

Gözde Berk, Berna Erden, Tunga Güngör
Bilgisayar Mühendisliği Bölümü
Boğaziçi Üniversitesi
İstanbul, Türkiye
{gozde.berk, berna.erden, gungort}@boun.edu.tr

Özetçe —Bu çalışmada, çok sözcüklü fiil ifadeleri etiketlenmiş Türkçe metin derlemi oluşturulmuştur. Derlemdeki çok sözcüklü fiil ifadeleri alt gruplarına göre etiketlenmiştir. PARSEME Ortak Çalışma 1.0'da yayınlanmış Türkçe eğitim ve test derlemleri PARSEME tarafından belirlenen etiketleme kurallarına göre düzenlenerek eğitim ve geliştirme derlemleri olarak güncellenmiştir. Aynı zamanda, bu etiketleme kuralları izlenerek yeni bir Türkçe test derlemi oluşturulmuştur. Metin derlemi gazetelerde yer alan siyaset, dünya, yaşam ve sanat konulu makaleler ve köşe yazılarından oluşmaktadır. Metin derlemi PARSEME Ortak Çalışma 1.1'de yayınlanacaktır. Derlem, sözdizimsel ayrıştırma, makine tercümesi ve n-gram temelli dil modelleme gibi Türkçe doğal dil işleme çalışmaları için önemli bir kaynak olacaktır.

Anahtar Kelimeler—Türkçe, Derlem, Doğal Dil İşleme, Çok Sözcüklü İfade, Derlem Etiketleme

Abstract—In this study, a Turkish corpus with labeled verbal multiword expressions was built. The verbal multiword expressions in the corpus were annotated according to their subcategories. The Turkish train and test corpora that was published in PARSEME Shared Task 1.0 were updated as train and development corpora based on PARSEME Annotation Guidelines. Additionally, a new Turkish test corpus was created by following the guidelines. The corpus consists of newspaper articles on politics, world, life, art and columns. The corpus will be released in PARSEME Shared Task 1.1. The corpus will be an important source to be used in many Turkish natural languages processing applications such as syntactic parsing, machine translation and n-gram language modeling.

Keywords—Turkish, Corpus, Natural Language Processing, Multiword Expression, Corpus Annotation

I. GİRİŞ

Çok sözcüklü ifadeler birden fazla kelimeden oluşan ve kendini oluşturan kelimelerden sözlüksel, sözdizimsel, anlamsal, edimsel ve istatistiksel kullanım olarak farklılık gösteren kelime gruplarıdır [1].

Çok sözcüklü ifadeler doğal dil işlemede önemli bir problemdir. Bu ifadeler konuşma dili sözlüğünün yaklaşık yarısını oluşturmaktadır [2]. İfadelerin kendini oluşturan kelimelerden farklı anlama sahip olabilmesi, alt kelimelerin cümlede farklı sırada bulunabilmesi problemin farklı dilbilimsel metotlar ile çözülmesini gerektirmektedir [3].

PARSEME (PARSING and Multiword Expressions) çok sözcüklü ifadeler üzerine çalışmalar düzenleyen ve bu tür çalışmaları destekleyen bilimsel bir gruptur. PARSEME her yıl çok sözcüklü ifadeler ile ilgili ortak çalışmalar düzenler. Bu çalışmalar sonucunda katılımcılar tarafından altın standartta derlem oluşturulmakta ve çok sözcüklü ifadeleri tanıyan otomatik sistemler geliştirilmektedir [4].

PARSEME tarafından 2017 yılında çok sözcüklü fiil ifadelerinin otomatik olarak tanınması amacıyla ortak bir çalışma düzenlenmiştir [5]. PARSEME Ortak Çalışma (Shared Task) 1.0 olarak adlandırılan bu etkinlikte, katılımcılar tarafından 18 farklı dilde derlem hazırlanmıştır. Her dile ait eğitim ve test olmak üzere iki derlem bulunmaktadır. Bu derlemler sadece baş sözcüğü fiil olan çok sözcüklü fiil ifadelerini içermektedir. Etkinliğin devamı PARSEME Ortak Çalışma 1.1 olarak 2018 Ağustos ayında düzenlenecektir. PARSEME Ortak Çalışma 1.1 kapsamında, PARSEME Ortak Çalışma 1.0'da yayınlanmış eğitim ve test derlemi PARSEME tarafından belirlenen etiketleme kuralları [6] ışığında düzenlenecektir. PARSEME Ortak Çalışma 1.0 eğitim ve test derlemi, PARSEME Ortak Çalışma 1.1 eğitim ve geliştirme derlemi adlarını alacaktır. Ek olarak, etiketleme kurallarını [6] izleyerek yeni test derlemi oluşturulacaktır.

Bu çalışmada, PARSEME Ortak Çalışma 1.1 tarafından belirlenen çok sözcüklü fiil ifadeleri etiketleme kurallarının [6] Türkçe'ye uyarlanması ile 284043 kelime ve 589 cümleden oluşan, 509 tane çok sözcüklü ifade içeren yeni PARSEME Ortak Çalışma 1.1 test derlemi oluşturulmuştur. Ayrıca, PARSEME Ortak Çalışma 1.0'da hazırlanmış olan Türkçe eğitim ve test derlemi yeni kurallar [6] çerçevesinde düzeltilerek PARSEME Ortak Çalışma 1.1 eğitim ve geliştirme derlemi haline getirilmiştir. Düzenleme yapılarak oluşturulan Türkçe eğitim ve geliştirme derlemleri ve yeni sunulacak olan Türkçe test derlemi, Türkçe dilinde çok sözcüklü ifadelerle ilgili yapılacak çalışmalarda önemli bir kaynak olacaktır.

II. ALTYAPI

Türkçe çok sözcüklü fiil içeren ifadeler, [7] tarafından belirlenen çok sözcüklü ifade gruplarından aşağıdaki beş gruba dahil olmaktadır:

- **Fiilimsi Çok Sözcüklü İfadeler:** Bu tür ifadelerde bir araya gelen kelimeler ifadeyi oluştururken anlamsal özelliklerini kaybetmezler.

- **Katkısız Eylemden Oluşan İfadeler:** Katkısız eylemler *olmak, etmek, yapmak, kılmak, eylemek, ve buyurmak* olarak dilbilimciler tarafından belirlenmiştir.
- **Deyimler:** Deyimleri oluşturan kelimeler bir araya geldiğinde düz anlamlarını kaybederek ifadeye yeni bir anlam yüklerler.
- **Formüsel İfadeler:** Fiili dilek kipi halinde bulunan iyilik, hoşnutluk dileme amaçlı kullanılan kelime grupları da çok sözcüklü ifadeye dahil edilmiştir.
- **Atasözleri:** Atasözleri kalıplaşmış bir biçime sahiptir. Bu yüzden bir bütün halinde bir anlamlıdır.

PARSEME Ortak Çalışma 1.0 etiketleme kılavuzuna [8] göre çok sözcüklü fiil ifadeleri beş ana gruba ayrılmıştır: **Katkısız Eylemden Oluşan Çok Sözcüklü İfadeler** (Light verb constructions-LVC), **Deyimler** (Idioms-ID), **Doğası Gereği Dönüşlü Fiiller** (Inherently reflexive verbs-IRV), **Fiil-İlgeç İfadeleri** (Verb-particle combinations-VPC) ve **Diğer Çok Sözcüklü Fiil İfadeleri** (Other verbal MWES-OTH).

Türkçe çok sözcüklü ifadeler sadece katkısız eylemden oluşan çok sözcüklü ifadeler, deyimler ve diğer çok sözcüklü fiil ifadeleri olarak sınıflandırılabilir.

PARSEME Ortak Çalışma 1.1 etiketleme kılavuzuna [6] göre ise çok sözcüklü fiil ifadeleri beş ana gruba ayrılmıştır:

- **Katkısız Eylemden Oluşan Çok Sözcüklü İfadeler** (Light verb constructions-LVC)
- **Fiil İçeren Deyimler** (Verbal idioms-VID)
- **Doğası Gereği Dönüşlü Fiiller** (Inherently reflexive verbs-IRV)
- **Fiil-İlgeç İfadeleri** (Verb-particle constructions-VPC)
- **Birden Fazla Fiilden Oluşan İfadeler** (multi-verb constructions-MVC)

Türkçe çok sözcüklü ifadeler bu beş ana gruptan sadece üç grupta sınıflandırılabilir. Bu üç grup, katkısız eylemden oluşan çok sözcüklü ifadeler, fiil içeren deyimler ve birden fazla fiilden oluşan çok sözcüklü ifadelerdir. Sırasıyla açıklarsak;

- **Katkısız eylemden oluşan ifadeler,** bir fiil ve bir veya birden fazla isimden oluşur. İsim aslında yüklemcil özellik taşıyor ve bir etkinliği tanımlar. Bu tür ifadeler iki alt gruba ayrılır. İlk grup olan tam katkısız eylemli ifadelerde (LVC.full), eylem isme sadece biçimsel özellik kazandırır ve anlamsal bir nitelik katmaz. İkinci alt grup ise, eylemin sebebi olan isimlerden oluşan nedensel katkısız eylemli ifadelerdir (LVC.cause). Türkçede katkısız eylemler sadece ilk gruba dahil olur.
- **Deyimler** ise bir fiil ve bu fiile bağlı en az bir kelimedenden oluşur. Fiile bağlı kelimenin türü isim veya sıfat olabilir.
- **Birden fazla fiilden oluşan ifadeler** ise ardışık olarak birden fazla fiil içerir.

PARSEME 1.0 eğitim ve geliştirme derlemi, PARSEME Ortak Çalışma 1.0 etiketleme kılavuzu [8] referans alınarak oluşturulduğu için, PARSEME Ortak Çalışma 1.1’de yer almayan diğer çok sözcüklü fiil ifadeleri (OTH) olarak etiketlenmiş ifadeler bulunmaktadır.

III. ETIKETLEME SÜRECİ

PARSEME 1.1 test derleminin oluşturulması ve PARSEME 1.0 eğitim ve geliştirme derlemlerinin düzenlenmesi dört ay sürmüştür. Etiketleyiciler makalenin iki yazarı anadili Türkçe olan kişilerdir. Öncelikle, PARSEME 1.1 tarafından sunulan karar ağacı testleri [6] Türkçe’ye uyarlanmıştır. Daha sonra, PARSEME 1.0 derlemlerindeki diğer çok sözcüklü fiil ifadeleri (OTH) olarak işaretlenmiş kelime grupları incelenmiştir. PARSEME Ortak Çalışma 1.1 etiketleme kılavuzuna göre böyle bir grup bulunmamaktadır. Güncellenen kılavuza göre, katkısız eylemden oluşan çok sözcüklü ifadelerin ve deyimlerin kapsamı değişmiştir. Bununla birlikte Türkçeye birden fazla fiilden oluşan ifadeler grubu eklenmiştir. Diğer çok sözcüklü fiil ifadesi olarak işaretlenmiş kelime gruplarından bazıları ise aslında çok sözcüklü fiil ifadesi değildir, hatalı olarak etiketlenmiştir.

Yukarıdaki değişiklikler göz önünde bulundurularak, önce diğer çok sözcüklü fiil ifadelerinin (OTH) çok sözcüklü fiil ifadesi olup olmadığı tartışılmıştır. Çok sözcüklü fiil ifadesi olmayanlar, herhangi bir sınıfa dahil edilmemiştir ve etiketlenmemiştir. Çok sözcüklü fiil ifadesi olduğuna karar verilenlerin hangi gruba ait olduğu tartışılmış ve PARSEME 1.1 etiketleme kılavuzuna göre katkısız eylemden oluşan çok sözcüklü ifadeler, fiil içeren deyimler veya birden fazla fiilden oluşan ifadeler olarak sınıflandırılmıştır.

PARSEME 1.0 derlemlerindeki LVC olarak işaretlenmiş kelime grupları LVC.full, ID olarak işaretlenmiş kelime grupları VID olarak otomatik güncellenmiştir.

Yapılan düzenlemeler sonrasında PARSEME Ortak Çalışma 1.0’da yayınlanmış eğitim derlemi ve test derlemi, PARSEME Ortak Çalışma 1.1 eğitim derlemi ve geliştirme derlemi olarak hazırlanmıştır.

PARSEME 1.1 test derlemi, PARSEME 1.0 eğitim ve geliştirme derlemleri ile uyumlu olabilmesi için gazetelerde yer alan siyaset, dünya, yaşam, sanat konulu makaleler ve köşe yazılarından oluşturulmuştur. Toplanan makale ve köşe yazıları öncelikle İTÜ NLP Tool [9] aracılığıyla CoNLL-U formatına çevrilmiş ve her bir etiketleyici tarafından bireysel olarak işaretlenmiştir. Bir sonraki adımda, bulunan ifadeler karşılaştırılmıştır. Farklı olarak etiketlenen ya da sadece bir etiketleyici tarafından etiketlenmiş ifadeler karşılıklı tartışılarak nihai bir karara bağlanmıştır.

A. Etiketleme Kılavuzu

Derlemin etiketlenmesi için PARSEME 1.1 Ortak Çalışma tarafından sunulan karar ağacı metodolojisi [6] kullanılmıştır. Buna ek olarak Türkçeye özel bazı çözüm yöntemleri geliştirilmiştir. Daha önce de değindiğimiz üzere Türkçe, PARSEME 1.1’in sunduğu altı gruptan ağırlıklı olarak katkısız eylemden oluşan çok sözcüklü ifadeler, deyimler ve birden fazla fiilden oluşan çok sözcüklü ifadeleri barındırır.

1) *Katkısız eylemden oluşan çok sözcüklü ifadeler:* PARSEME 1.1 etiketleme kılavuzuna göre, katkısız eylemden oluşan çok sözcüklü ifadeler bir fiil ve tek ya da bileşik isimden oluşur. İsim, fiile doğrudan ya da edat ile bağlı olabilir. İsim yüklemcil olmalıdır ve olay ya da durum ifade etmelidir. Yukarıda da belirtildiği gibi, katkısız eylemden oluşan çok sözcüklü ifadeler, tam katkısız eylemli ifadeler ve nedensel

katkısız eylemli ifadeler olmak üzere ikiye ayrılır. Türkçede yalnızca tam katkısız eylemli ifadeler bulunur.

Daha önce belirtildiği üzere, Türkçede altı tane katkısız eylem tespit edilmiştir. Bunlar *olmak*, *etmek*, *yapmak*, *kılmak*, *eylemek* ve *buyurmak* olarak belirlenmiştir. Katkısız eylemlerin kullandığı her kelime grubu tam katkısız eylemli ifade olmadığı gibi tam katkısız eylemli ifadeler belirtilen altı eylemle de sınırlı değildir. Örneğin, *maskara olmak* bir deyimdir ama *olmak* katkısız eylemiyle oluşmuştur. Tam katkısız eylemli ifade oluşturan yardımcı eylemler de bulunmaktadır. Katkısız veya yardımcı fiillerle oluşan kelime gruplarının katkısız eylemli ifadeler olup olmadıklarına PARSEME tarafından sunulan karar ağacı takip edilerek karar verilmiştir.

Karar ağacı [6] çeşitli testlerden oluşmaktadır. Çok sözcüklü ifadenin tam katkısız eylemli ifade olması için tüm testleri geçmesi gerekmektedir. Karar ağacının kullanımını örnekleyebilmek açısından, tam katkısız eylemli ifade olan *karar vermek* kelime grubunu inceleyebiliriz:

- LVC.0 - [N-ABS]: İsim soyut olmalıdır. Örnekte de *karar*, soyut bir isimdir.
- LVC.1 - [N-PRED]: İsim yüklemci olmalıdır. Diğer bir deyişle, en az bir tane anlamsal argümanı olmalıdır. Örnekte *karar vermek* bir olaydır ve *karar veren* anlamsal argümanına sahiptir.
- LVC.2 - [N-SUBJ-N-ARG]: Eylemin öznesi, ismin anlamsal argümanı olmalıdır. "*Ali ders çalışmaya karar verdi.*" cümlesini ele alacak olursak, Ali eylemin öznesidir ve karar veren yani anlamsal argümandır.
- LVC.3 - [V-LIGHT]: Bir diğer teste göre ise, isim orijinal anlamını korumalıdır. Bu örnekte *karar* kelimesi anlamını korumaktadır. Ek olarak, eylem isme sadece biçimsel bir özellik kazandırabilir ama anlamsal bir nitelik katmamalıdır. Örnekte de *vermek* hiç bir anlamsal katkı sağlamaz, sadece karar eylemini gerçekleştirme anlamı verir.
- LVC.4 - [V-REDUC]: Son olarak, eylem düşürmesi ile elde edilen isim tamlaması aynı olay ya da durumu ifade etmelidir. *Ali'nin kararı* isim tamlaması, aynı karar verme olayına değinmektedir.

Dolayısıyla, *karar vermek* tüm testleri geçtiği için tam katkısız eylemli ifade olarak etiketlenir.

2) *Deyimler*: PARSEME 1.1 etiketleme kılavuzu fiil içeren deyimleri kapsamaktadır. Bu yüzden, bu makale boyunca baş sözcüğü fiil olan deyimler incelenmiştir. Fiil içeren deyimler bir tane fiil ve fiile bağlı bir veya birden fazla sözcükten oluşur. Fiile bağlı sözcükler isim veya sıfat olabilir. Ayrıca, atasözleri de deyimler grubuna dahil edimiştir.

Karar ağacı testlerini incelediğimizde, deyimler bir derecede esneksizlik gösteren kelime gruplarıdır. Yani, deyimler sözlüksel, biçimsel ve sözdizimsel olarak sabit bir yapıda kullanılır. Eğer aday ifade, karar ağacı testlerinden birinde başarılı olursa deyim olarak işaretlenir.

Örneğin, *ifade vermek* kelime grubu tam katkısız eylemli ifade ya da deyim olabilir. Bir yandan, *ifade* sözcüğü *etmek* katkısız eylemiyle kullanılmaktadır. *etmek* eylemi *ifade* etkinliğinin gerçekleşmesini sağlamakta fakat anlamsal olarak

etkinliğe katkı sağlamamaktadır. Diğer bir yandan, *vermek* kelimesini incelediğimizde, bu kelime *ifade* etkinliğine yasal olma anlamı katmaktadır. Bu yüzden kelime grubu deyim testi kapsamında aşağıdaki gibi incelenmiştir.

- VID.1 - [CRAN] Aday ifade tek başına anlamı olmayan ama başka kelimelerle kullanılabilen kelime içeriyorsa deyimdir. *ifade vermek* ifadesinin kelimelerinin her biri tek başına anlama sahiptir. Aday ifade, bu kurala uymamaktadır.
- VID.2 - [LEX] Deyimler sözlüksel olarak sabit yapıda olabilir. *ifade* kelimesi *açıklama* kelimesi ile eş anlamlıdır. Bu iki kelimeyi değiştirip kelime grubunun sözlüksel esnekliğini kontrol edersek; *açıklama vermek* kelime grubu Türkçe'de kullanılmamaktadır. Aday ifade sabit yapıdadır.
- VID.3 - [MORPH] Deyimler biçimsel olarak sabit yapıda olabilir. *ifade* kelimesine çoğul eki eklersek, *ifadeler vermek* kelime grubu dilbilgisi kurallarına uymamaktadır. Aday ifade sabit yapıdadır.
- VID.4 - [MORPHSYNT] Deyimler biçimsözdizimsel olarak sabit yapıda bulunabilir. *Karakolda ifade verdim* cümlesindeki *ifade* kelimesinin iyelik ekini değiştirirsek, *Karakolda ifadeni verdim* cümlesi doğru değildir. Aday ifade sabit yapıdadır.
- VID.5 - [SYNT] Deyimi oluşturan kelimeler cümle içerisinde sözdizimsel olarak sabit sırada bulunabilir. Bir önceki testteki cümleyi *Karakolda verdiğim ifade tekrar kontrol edildi* şeklinde değiştirirsek, aday ifadedeki kelimeler cümle içinde yer değiştirebilmektedir. Bu test başarısızdır.

Testler sıra ile uygulanmalıdır. *ifade vermek* kelime grubu VID.2 - [LEX] testine göre sabit yapıdadır ve deyim olarak işaretlenir. Kalan testlerin uygulanmasına gerek yoktur.

Sonuçta, *ifade vermek* kelime grubunun sabit bir yapıda bulunduğunu görmekteyiz.

3) *Birden fazla fiilden oluşan ifadeler*: Birden fazla fiilden oluşan ifadeler, aşağıdaki özelliklere sahip olan birden fazla fiilden oluşur. Örneğin, birden fazla fiilden oluşan ifade olan *gidip gelmek* kelime grubunun "*İki seçenek arasında gidip geliyorum.*" cümlesindeki kullanımı incelenerek fiillerin sahip olması gereken özellikler daha iyi anlaşılabilir.

- Genellikle aynı özneye sahiptirler. Örnekte *gitmek* ve *gelmek* aynı ben gizli öznesine sahiptir.
- Genelde birbirine bağlı ya da birbiriyle ilişkili eylemleri ifade ederler ve aynı olayın parçası olarak kabul edilebilirler. Örnekte de *gitmek* ve *gelmek* ayrı olayları değil tek bir olayı ifade eder.
- Birlikte tek bir yüklem olarak görev yaparlar. Verilen cümlelerin yüklemi *gidip geliyorum* kelime grubudur.

B. Çıkarımlar

Tam katkısız eylemli ifadeleri etiketlerken dikkat edilmesi gereken en önemli nokta, fiilin isme anlamsal olarak katkı sağlayıp sağlamadığıdır. Genellikle anlamsal katkı sağlanmaması

durumunda, çok sözcüklü fiil ifadeleri tam katkısız eylemli ifadelerdir.

Bazı isimler doğası gereği bazı belirli fiillerle bağlantılıdır. Mesela, *talimat* kelimesi, *talimat vermek* anlamı vermeye daha yatkındır. Bu yüzden, *talimat* ismi *vermek* fiiliyle bağlantılıdır. Bu da *vermek* fiilini anlamsal olarak katkısız hale getirir.

Çıkarımlardan bir diğeri ise, bir isim ve *vermek* fiili birlikte tam katkısız eylemli ifade oluşturduğunda, aynı isim ve *almak* fiili birlikte tam katkısız eylem değil, deyim oluşturur. Çünkü *vermek* fiili ismin eyleme dönüşmesini sağlarken, *almak* fiili isme yeni bir anlam katmış olur. Örneğin, *oy vermek* tam katkısız eylemli ifadedir; çünkü *vermek* herhangi bir anlam katmaz, sadece oy aktivitesini gerçekleştirme anlamı verir. Ancak, *oy almak* bir deyimdir çünkü *oy* kelimesine, o aktiviteyi gerçekleştirmek dışında tamamen farklı bir anlam katar.

Yukarıdaki yöntemlere ek olarak, anlamsal nitelik testi için Türk Dil Kurumu'nun Türkçe sözlüğü [10] de kullanılmıştır. Söz *vermek* kelime grubunu örnek olarak alalım. TDK'ya göre söz vaat, söz vermek ise vadetmek anlamına gelmektedir. Görüldüğü üzere, *vermek* fiili hiç bir anlam katmamıştır, çünkü *etmek* katkısız eylemdir ve *vadetmek*, *vaat* ile *etmek* kelimelerinin birleşmesiyle oluşmuştur.

Aynı zamanda *sağlamak*, *duymak*, *ulaşmak*, *karşılama*, *kullanmak*, *uygulamak* fiilleri ile oluşturulan kelime gruplarının çok sözcüklü fiil ifadeleri olup olmadığı konusunda karışıklık yaşanmaktadır. Ama genel olarak bu fiillerle oluşturulan kelime grupları çok sözcüklü fiil ifadeleri değildir.

IV. HAZIRLANAN DERLEM

İstatistikleri incelediğimizde, varolan PARSEME 1.0 derlemi Tablo I ve Tablo II'deki gibi güncellenmiştir ve PARSEME 1.1 eğitim ve geliştirme derlemlerini oluşturmuştur. Yeni PARSEME 1.1 test derlemi Tablo III'te gösterildiği gibi toplamda 509 çok sözcüklü fiil ifadesi içermektedir. Bu ifadelerin 273 tanesi katkısız eylemden oluşan ifade, 235 tanesi fiil içeren deyim ve 1 tanesi birden fazla fiilden oluşan ifadedir.

TABLO I: TÜRKÇE EĞİTİM DERLEMİ

	LVC	VID	OTH	MVC	Toplam
PARSEME 1.0	2624	2911	634	0	6169
PARSEME 1.1	2950	3169	0	1	6120

TABLO II: TÜRKÇE GELİŞTİRME DERLEMİ

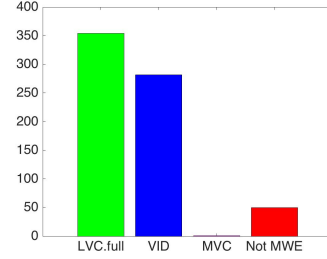
	LVC	VID	OTH	MVC	Toplam
PARSEME 1.0	199	249	53	0	501
PARSEME 1.1	227	273	0	0	500

TABLO III: TÜRKÇE TEST DERLEMİ

	LVC	VID	OTH	MVC	Toplam
PARSEME 1.1	273	235	0	1	509

Eski etiketleme kılavuzu [8] dikkate alınarak LVC ve ID olarak işaretlenmiş kelime grupları yeni kılavuz [6] ile birlikte otomatik olarak LVC.full ve VID olarak güncellendiği için, derlem güncellenen şekliyle etiketlenmiştir. PARSEME 1.1'de kaldırılan OTH olarak etiketlenmiş kelime gruplarının

Şekil 1'deki gibi 354 tanesi LVC.full, 282 tanesi VID, 1 tanesi MVC olarak sınıflandırılmıştır. 50 tanesinin çok sözcüklü fiil ifadesi olmadığına karar verilmiştir ve etiketlenmemiştir. Özet olarak, hazırlanan derlem toplamda 7129 çok sözcüklü fiil ifadesi içermektedir.



Şekil 1: OTH etiketlenmiş kelime gruplarının güncellenmesi.

V. SONUÇ VE GELECEK ÇALIŞMALAR

Etiketleme süreci boyunca, mevcut kaynaklardaki çok sözcüklü ifadeleri etiketlemek için kullanılan kuralların Türkçe için farklılık gösterdiği görülmüştür. Ayrıca, Türkçe çok sözcüklü ifade içeren veri kaynaklarının az olması sebebiyle, bu derlemler çok sözcüklü ifadeler üzerine yapılacak olan Türkçe doğal dil işleme çalışmaları için önemli bir kaynak olacaktır. Derlemin kullanım alanlarından bazıları sözdizimsel ayrıştırma, makine tercümesi ve n-gram temelli dil modelleme olarak sayılabilir. PARSEME 1.1. eğitim, geliştirme ve test derlemi Mart 2018 tarihinde kullanıma açılacaktır.

KAYNAKLAR

- [1] T. Baldwin and S. N. Kim. "Multiword expressions," in *Handbook of Natural Language Processing*, Second ed., Chapman and Hall/CRC, 2010, pp. 267–292.
- [2] R. Jackendoff. *The architecture of the language faculty*. Number 28, MIT Press, 1997.
- [3] I. A. Sag, T. Baldwin, F. Bond, A. Copestake, and D. Flickinger. "Multiword expressions: A pain in the neck for nlp," in *International Conference on Intelligent Text Processing and Computational Linguistics*, Springer, 2002, pp. 1–15.
- [4] A. Savary, M. Sailer, Y. Parmentier, M. Rosner, V. Rosén, A. Przepiórkowski, C. Krstev, V. Vincze, B. Wójciszewski, G. S. Losnegaard, C. P. Escartín, J. Waszczuk, M. Constant, P. Osenova, and F. Sangati. "PARSEME – PARSing and Multiword Expressions within a European multilingual network," in *7th Language and Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics (LTC 2015)*, Poznan, Poland, November 2015.
- [5] A. Savary, C. Ramisch, S. Cordeiro, F. Sangati, V. Vincze, B. QasemiZadeh, M. Candito, F. Cap, V. Giouli, I. Stoyanova and A. Doucet. "The parseme shared task on automatic identification of verbal multiword expressions," in *Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017)*, 2017, pp. 31–47.
- [6] <http://parsemefr.lif.univ-mrs.fr/parseme-st-guidelines/1.1/?page=home>
- [7] K. Adalı, T. Dinç, M. Gökırmak, and G. Eryiğit. "Comprehensive Annotation of Multiword Expressions in Turkish," in *The First International Conference on Turkic Computational Linguistics at CICLING*, 2016, pp. 60–66.
- [8] <http://parsemefr.lif.univ-mrs.fr/parseme-st-guidelines/1.0/>
- [9] G. Eryiğit. "Itu Turkish NLP Web Service," in *Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics*, 2014, pp. 1–4.
- [10] <http://www.tdk.gov.tr/>