



## TERİM BİLİM ÇALIŞMALARINDA DERLEM KULLANIMI ve “KENDİ KENDİNE DERLEM OLUŞTURMA PLATFORMU” TERİM BİLİM MODÜLÜ\*

Bülent Özkan<sup>1</sup>

<sup>1)</sup> Mersin Üniversitesi, Eğitim Fakültesi, Türkçe Eğitimi Bölümü, Mersin,  
ORCID ID: 0000-0002-4319-7582, İletişim: ozkanbulent@gmail.com

### Özet

Bu çalışmanın amacı, Türkçe üzerine çalışan dil/dilbilim araştırmacıları için araştırma sorularına bağlı olarak biçimlendirilebilen, araştırmacıya özel, esnek, kolay ulaşılabilir, *veri tabanı destekli* ve bu çerçevede araştırma sonuçlarının sağlıklı bir şekilde raporlanabildiği bir derlem platformu olan “Kendi Kendine Derlem Oluşturma Platformu”nun (KKDP) Terim Bilim Modülü’nü (TBM) tanıtmaktır. Söz konusu TBM, KKDP’nin alt modüllerinden biri olarak terim bilim çalışmalarında kullanılmak üzere geliştirilmiş bir derlem modülüdür.

**Anahtar Kelimeler:** Derlem Dilbilim, Kendi Kendine Derlem Oluşturma, Terim Bilim Modülü

### USE OF CORPUS IN TERMINOLOGY STUDIES AND TERMINOLOGY MODULE OF “DO IT YOURSELF CORPUS PLATFORM”

### Abstract

The aim of this paper is to introduce the Terminology Module (TM) of the Do It Yourself Corpus Platform (DIYCP) which can be individualized in line with the research questions of the scholars, and which is exclusive to the scholar, user friendly, and is *supported by the database*. Considering this perspective, this corpus is going to provide a corpus platform in which the results of the research can be reported in a functional way. The mentioned TM is the one of the DIYCP module which can be used in the field of terminology studies.

**Key words:** Corpus Linguistics, Do It Yourself Corpora, Terminology Module

\* Bu çalışma, TÜBİTAK 1005 Ulusal Yeni Fikirler ve Ürünler Araştırma Destek Programı çerçevesinde 111E791 numaralı “Türkçe için Kendi Kendine Derlem Platformu Oluşturma” başlıklı projeden üretilmiştir. Katkıları için TÜBİTAK’a teşekkür ederim. Bu metin ilgili projenin sonuç raporunda üretilmiştir.

## 1. Giriş

*Derlem*, (corpus) dilbilimde bir dili belirgin dilbilimsel kriterlere göre örneklemek amaçlı olarak o dile ait yazılı ve sözlü metinlerden seçilerek belirli standartlara göre düzenlemiş metin parçalarının bütünü (McEnery ve ark., 2006) olarak tanımlanabilir.<sup>1</sup> *Doğal Dil İşleme* (DDİ) olarak da adlandırılan bilgisayarlı dilbilim, *derlem* adı verilen dil örnekçelerini temel alarak, *uygulamalı dilbilimin* (applied linguistics) paralelinde, başta sözlükbilim olmak üzere dilbilgisi, ağız, çeviri bilim, tarihsel dilbilgisi ve dilsel değişim, dil öğrenimi ve öğretimi, anlambilim, kullanımbilim, toplum dilbilim, söylem çözümlemesi, biçembilim ve yazınbilim çalışmalarında yoğunlukla kullanılmaktadır (McEnery, ve ark., 2006: 80-122; Kennedy 1998: 208-310).

Bugün dilin doğal ortamlarından yazılı ve sözlü olarak derlenen verilerle kendine özgü kuram ve uygulamalarla ortaya koyulan derlem, özellikle sezgisel örnekçelerden ve ikincil verilerden yalıtılmış 'gerçek zamanlı' dilsel verileri içermesi bakımından alana büyük katkılar sağlamaktadır (Özkan, 2010 ve 2013). Bu anlamda *uygulamalı dilbilimin* yukarıda bir kısmı sıralanan alt dallarında araştırmacıların yürüttükleri derleme dayalı çalışmalar günümüzde oldukça yaygınlaşmıştır.

Türkçe için derlem oluşturma çalışmaları bugün için literatürde yerini almıştır.<sup>2</sup> Ancak mevcut derlem, araştırma soruları genel dile değil de daha özelleşmiş, belirgin konu ve alanlara odaklanıldığında araştırmacıların bütün amaçlarına hizmet etmekte yetersiz kalmaktadır. Günümüzde dil/dilbilim çalışmaları daha ayrıntılı araştırma sorularına cevap arama eğilimindedir ve bu çerçevede dilbilimciler var olan standart derlem çıktıları yerine *araştırma sorularına bağlı olarak biçimlendirilebilen, araştırmacıya özel, esnek, kolay ulaşılabilir, veri tabanı destekli ve araştırma sonuçlarının sağlıklı bir şekilde raporlanabildiği* bir derlem altyapısına sıklıkla ihtiyaç duymaktadır.

Bilindiği üzere batı dilleri için geliştirilmiş olan metin işleme/derlem yazılımları o dillere özgü standart birtakım veri işleme özelliklerine sahiptir. Türkçe için kısmen kullanılabilen söz konusu yazılımlar, Türkçe karakter desteği sunabilmelerine karşın bu yazılımların en büyük eksikliği, oluşturuldukları diller de dâhil olmak üzere, araştırma sorularına bağlı olarak işlenen veriler üzerinden ortaya çıkan raporların saklanamaması ve amaca uygun olarak düzenlenememesidir. Bu yazılımla-

rın belirli alanlarda özelleşmiş olmaları da söz konusu yazılımların eksiklikleri olarak değerlendirilebilir. Örneğin, *AntConc*, *WordSmith Tools*, *MonoConc Pro*, *TextStat*, *Nooj* benzeri sık kullanılan paket yazılımlar *sıklık*, *bağımlı dizin*, *anabtar sözcük gösterimi*, *n-gram* ve *eşdizimlilik hesaplama* gibi konularda geliştirildikleri diller için çıktılar verebilmektedir. Kısmen de olsa bu çıktılar metin belgeleri olarak dışa aktarılabilmeyle birlikte çoğu kez araştırmacı metin yığınlarından elde edilen ayrıştırılmamış, filtrelenmemiş başka bir metin yığını ile karşı karşıya kalmaktadır. Söz konusu paket yazılımların bir diğer olumsuz yönü ise farklı işletim sistemlerine özgü olmasıdır. Her ne kadar paket yazılımların UNIX, Mac, Windows gibi farklı işletim sistemleri için sürümleri mevcutsa da bu durum programlara erişimi ve söz konusu programların kullanımını genellikle zorlaştırabilmektedir.

Oysa dil/dilbilim çalışmaları bugün daha ayrıntılı araştırma sorularına cevap arama eğilimindedir. Türkçe üzerine yürütülen derlem çalışmalarının sınırlılığı ve gün geçtikçe derlem oluşturma ve derlem dilbilim yöntemleriyle Türkçeyi değerlendirme eğiliminin de arttığı bir gerçektir. Ayrıca söz konusu metin işleme/derlem yazılımlarını kullanmayı öğrenmek bile başlı başına bir iş olarak araştırmacıların karşısında durmaktadır. Bu anlamda *kullanıcı dostu, kolay erişilebilir, araştırma sorularına bağlı olarak geliştirilebilen esnek* platformlara ihtiyaç vardır. **Kendi Kendine Derlem Oluşturma Platformu (KKDP)** söz konusu tüm olumsuzlukları kullanıcının yararına dönüştürebilecek bir yapıdadır. Diğer taraftan her dilbilimciden derlem oluşturma/işleme gibi özel bir alana dair yazılım bilgisi edinmesini beklemek de pratik bir yaklaşım gibi görünmemektedir.

**Kendi Kendine Derlem Oluşturma Platformu** her geçen gün geçerliliği artan bir uygulama alanı olan derlem dilbilimin ilke ve yöntemlerini kuramsal ve uygulamalı açıdan temel alarak oluşturulmuştur. Bu çerçevede dilbilimciler var olan standart derlem çıktıları vermek yerine *araştırmacıların araştırma sorularına bağlı olarak biçimlendirilebilen, araştırmacıya özel, esnek, kolay ulaşılabilir, veri tabanı destekli ve araştırma sonuçlarının sağlıklı bir şekilde raporlanabildiği* bir derlem altyapısı sunmayı amaçlamaktadır.

## 2. Amaç ve Kapsam

### 2.1. Amaç

Türkçe yapısı gereği eklemeli bir dil özelliğine sahiptir. Bu durum var olan metin işleme/derlem yazılımlarıyla Türkçeyi işlemede zorluklar yaratmaktadır. Öte

<sup>1</sup> Başka tanımlar için bk. Say, 2002 ve 2003.

<sup>2</sup> Türkçe üzerine yürütülen önemli derlem çalışmaları için bk. Özkan 2010.

yandan her araştırma kendine özgü birtakım veri işleme ve raporlama süreçlerini de araştırmanın amacına uygun olarak kendisinde barındırmaktadır. Bu çerçevede araştırmacılara kendi araştırma sorularına bağlı olarak veri işleme ve raporlama imkânı sunabilen KKDP'nin alt modülü olan Terim Bilim Modülü'nü araştırmacılara tanıtmak bu çalışmanın ana amacıdır.

## 2.2. Kapsam

Literatürde yer alan bilinen metin işleme/derlem yazılımları Türkçe dışında diller için geliştirilmiş yazılımlardır. Söz konusu yazılımların Türkçeye özgü niteliklerden dolayı Türkçenin etiketlenmesi ve etiketlenen verilerin raporlanmasında yetersiz kaldığı görülür.

Bugün için kullanılan yazılımların bir diğer eksikliği ise yalnızca standart birtakım raporlama araçlarına sahip olmasıdır. Oysa her araştırma, araştırma sorularına bağlı olarak kendine özgü birtakım özellikler taşır ve söz konusu özellikler derlem dilbilimin araştırma mantığına göre farklı raporlamalar ve etiketlemeler/işaretlemler gerektirir.

Bir yazılım kümesi olarak KKDP, derlem dilbilimin yönetsel yaklaşımlarına paralel olarak kullanıcıların ihtiyaçlarını en üst seviyede karşılayabilmekte ve doğrudan Türkçenin kendine özgü yapısının araştırma sorularına göre biçimlendirilebileceği bir platformu araştırmacılara sunabilmektedir.

Bu çerçevede KKDP'nin Türkçe için özelleştirilmiş küçük ve orta ölçekli derlem projelerinde etkili bir şekilde kullanılacak bir derlem platformu olduğu söylenebilir. Bu anlamda çalışmanın konusu ve kapsamı KKDP'nin tanıtımı ve bunun yanı sıra bir alt modül olan Terim Bilim Modülü'nün terim çalışmalarında ne şekilde kullanılabileceğini ortaya koymaktır (Özkan, 2016).

## 3. Tartışma ve Değiştirme

### 3.1. KKDP'nin Özgün Değeri

Batıda derlem-tabanlı çalışmalar, her ne kadar daha eskiye dayansa da 1960'lı yılların başlarından bu yana edebiyat, sözlükbilim, ağız (diyalekt) çalışmaları, dil öğretimi ve dilbilgisinde bir yöntem olarak hızlı bir gelişim göstermiştir. Günümüzde ise, derlem odaklı çalışmaların özellikle sözlükbilim, dil öğretimi ve dilbilgisi alanlarında yoğunlaştığı gözlemlenmektedir (McEnery, ve ark., 2006).

Dilbilimsel çalışmalar çerçevesinde web tabanlı yazılımlara literatürde rastlanmaktadır. Bu yazılımların başında bir metin işleme/derlem oluşturma yazılımı olan ve BNCweb ile bütünleşik olarak çalışan *WordSmith* gelmektedir. Lancaster Üniversitesince ortaya koyulan bu yazılım yine standart derlem işleme aşamalarına bağlı olarak raporlar sunabilmektedir.<sup>3</sup>

Derlem işleme üzerine paket metin işleme/derlem yazılımları dışında İngilizce için web tabanlı olarak da kullanılabilen sözcük türü etiketlenmesi ve gramatikal etiketleme gibi gelişmiş etiketlemeler yapabilen CLAW (Constituent-Likelihood Automatic Word-Tagging System) benzeri<sup>4</sup> çeşitli formatlarda çıktılar üretebilen çalışmalar da literatürde yerlerini almıştır.

Bugün derlem çalışmalarında açık kaynak kodlu yazılımlar, metin işleme/derlem oluşturma yazılımları ve web tabanlı uygulamalar derlem oluşturma çalışmalarında sıklıkla kullanılmaktadır. Söz konusu yazılımlar Php, Perl, Java, Python, C vb. programlama dilleri kullanılarak oluşturulmuş yazılımlardır. Bu anlamda Steven B., Ewan K., Edward L. (2009) <http://www.nltk.org/> adresinden de erişilebilen Python programlama diliyle Unix tabanlı sürümünün yanında diğer işletim sistemleri için de üretilmiş NLTK (Natural Language Toolkit) önemli bir yer tutmaktadır. Ancak söz konusu program kullanıcı için programlama bilgisi gerektirmektedir.

ICEweb olarak bilinen küçük ve basit web tabanlı derlem uygulaması ise *International Corpus of English*'e dayanmaktadır. Bahsi geçen diğer uygulamalar gibi bu yazılım sözcük türü etiketleme, bağımlı dizi oluşturma, sıklık belirleme gibi olanakları araştırmacılara sunabilmekte; kurulum dosyası üzerinden bilgisayara kurularak çalıştırılabilmektedir.

<https://cqpweb.lancs.ac.uk/> adresinden de ulaşılabilen CQPweb ise açık kaynak kodlu yazılımların bir araya getirilmesiyle oluşturulmuş, derlem oluşturma ve işlemede kullanılan Perl ile yazılmış bir yazılım kümesidir. CQPweb<sup>5</sup> (Corpus Query Processor), var olan derlem için (BNCWeb XML -British National Corpus XML Sürümü- gibi) bütüncül sorgulama imkânını kullanıcılara sunabilmektedir.

Yine bu tip yazılımlara başka bir örnek UAM CorpusTool'dur. Yine bir kurulum dosyasıyla bilgisayar üzerinden çalışabilen özellikle batı dilleri için geliştirilmiş son kullanıcı için düşünülse de kullanıcı tarafından ayrıca öğrenilmesi gereken ve veri tabanı desteği sunamayan, bu

<sup>3</sup> [http://www.lancaster.ac.uk/fss/courses/ling/corpus/blue/diy\\_top.htm](http://www.lancaster.ac.uk/fss/courses/ling/corpus/blue/diy_top.htm)

<sup>4</sup> <http://ucrel.lancs.ac.uk/claws/>

<sup>5</sup> <https://cqpweb.lancs.ac.uk/>

anlamda da raporlama aşamasında beklentileri karşılayamayan, bir derlem yazılımıdır.<sup>6</sup>

Örnekleri çoğaltmak mümkünse de tüm bu yazılımlar batı dilleri, özellikle de İngilizce için geliştirilmiş yazılımlardır. Bu yazılımları Türkçe üzerine çalışan dil/dilbilim araştırmacıları için araştırma sorularına bağlı olarak biçimlendirilebilen, araştırmacıya özel, esnek, kolay ulaşılabilir, veri tabanı destekli ve bu çerçevede araştırma sonuçlarının sağlıklı bir şekilde raporlanabildiği derlem yazılımları olarak değerlendirmek güçtür.

Alan uzmanları derlem araştırmalarının ya oluşturulmuş derlemeler üzerinden yapılabildiği ya da amaca uygun bir derlem oluşturularak genellikle sözcük sıklığı belirleme, bağımlı dizin oluşturma gibi basit-standart derlem çıktıları alınabileceği konusunda birtakım bilgiler vermekte ve söz konusu çıktıların alınabileceği paket yazılımlar ya da web uygulamalarını önermektedir (Sinclair, 1991; Stevens, 1995).

KKDP'nin bu anlamda temel amacı açık kaynak kodlu, internet üzerinden erişilebilen, dilbilimsel etiketlemeler yapabilen, eşdizimlilik ve sıklık listesi çıktıları verebilen, bunun yanında istatistiksel yordamalar sunan, sonuçları kaydedebilen bir derlem altyapısı ve arayüzü oluşturmaktır.

Derlem dilbilim konusunda ulusal ve uluslararası patent, faydalı model, tescil taramasında<sup>7</sup> genellikle derlemelerin oluşturulması, derlemlerde sorgu sonuçlarının verimliliği, derlemlerden bir takım standart veri çıkarımları ya da bunların iyileştirilmesi veya sözlü derlem vb. doğal dil işlemenin alt araştırma alanlarında bir takım metod ve yöntemlerin geliştirilmesi üzerine patent, faydalı model veya tescil alındığı görülür. Bu anlamda *Türkçe için Kendi Kendine Derlem Oluşturma Platformu* özgün ve alanında ulusal ve uluslararası yenilikleri barındıran bir yapıya sahiptir.

Öte yandan derlem dilbilim alanında batı literatüründe özellikle altı çizilen nokta, derlemelerin araştırma sorularına göre şekillenen araştırma evrenleri olmalarıdır (McEnery, T. ve ark., 2006, Sinclair, J. 1991, McEnery T. ve Wilson A. 1996, Barnbrook, G. 1996). Bu anlamda KKDP söz konusu ihtiyacı karşılayacak niteliktedir.

### 3.2. Kendi Kendine Derlem Oluşturma Platformu Terim Bilim Modülü ve Örnek Bir Uygulama.

"Kendi Kendine Derlem Oluşturma Platformu" araştırmacılara var olan standart derlem çıktılarının yanı sıra *araştırma sorularına bağlı olarak biçimlendirilebilen, araştırmacıya özel, esnek, kolay ulaşılabilir, veri tabanı destekli ve araştırma sonuçlarının sağlıklı bir şekilde raporlanabildiği* bir derlem altyapısı sunabilmektedir.<sup>8</sup>

Var olan metin işleme/derlem paket yazılımları yukarıda da sözünü ettiğimiz üzere batı dilleri için geliştirilmiş olan derlem yazılımlarıdır ve o dillere özgü birtakım veri işleme ve raporlama özelliklerine sahiptir. Türkçe için de kısmen kullanılabilen söz konusu yazılımların en büyük eksikliği, oluşturuldukları diller de dahil olmak üzere, işlenen verilerin saklanamaması ve araştırma sorularına bağlı olarak sağlıklı bir şekilde söz konusu verilerin raporlanamamasıdır. Ayrıca bu yazılımları Türkçe üzerine çalışan dil/dilbilim araştırmacıları için araştırma sorularına bağlı olarak biçimlendirilebilen, araştırmacıya özel, esnek, kolay ulaşılabilir, veri tabanı destekli ve bu çerçevede araştırma sonuçlarının sağlıklı bir şekilde raporlanabildiği derlem yazılımları olarak değerlendirmek güçtür.

Derlem oluşturmak bilgisayar bilimleri uzmanlarının desteği olmaksızın gerçekleşmesi güç bir iş olarak görülmektedir. Diğer taraftan batı dilleri için geliştirilen metin işleme/derlem oluşturma paket yazılımları Türkçeye özgü olmamaları bir tarafa bırakılırsa ayrıca öğrenilmesi gereken ayrıntılı ve karmaşık yazılımlardır. Bir dilbilimcinin söz konusu yazılımları kullanabilmesi için bu yazılımların ayrıntılı ve karmaşık yapılarını çözümlemesi gerekmektedir. Oysa dilbilimciler, bir tarayıcı (browser) üzerinden, kullanıcı dostu, açık ve anlaşılır arayüzlerle kendi uzmanlık alanlarına ve araştırma sorularına bağlı olarak bir derlem oluşturabilme olanağına sahip olabilmelidirler.

KKDP, araştırmacıların araştırma sorularına bağlı olarak kuracakları derlemde yer alacak tabakaları oluşturabileceği, metin girişleri yapabileceği ve bunların metadatalarını belirleyebileceği, bunlara bağlı olarak birtakım standart derlem işlemlerinin yanında yine araştırma sorularına bağlı olarak kendi etiketleme sistemini esnek olarak belirleyip oluşturacağı, derlemi bu etiket sistemine göre etiketleyebileceği ve tüm bunları esnek bir raporlama sistemi üzerinden çıktı olarak alabileceği bir platformdur.

<sup>6</sup> Bahsedilen metin işleme/derlem oluşturma yazılımlarının bir listesi ve söz konusu yazılımların linkleri için (<http://www.uow.edu.au/~dlee/software.htm>) adresi önemli bir kaynaktır.

<sup>7</sup> <http://www.tpe.gov.tr>, [tr.espacenet.com](http://www.espacenet.com) ve <http://www.epo.org/> adreslerinden "*corpus, corpora, DIY corpora, building corpus, own corpus, derlem*" gibi anahtar sözcüklerle yapılan sorgu sonuçları incelendiğinde benzer bir çalışmanın olmadığı görülür.

<sup>8</sup> KKDP üzerinden yapılmış çalışmalar için bk. <http://kkd.mersin.edu.tr/>

### 3.3. KKDP ile İlgili Çeşitli Tanımlar<sup>9</sup>

#### 3.3.1. Derlem Oluşturma ile ilgili Tanımlar

**Tabaka:** Derlemlerde araştırma sorularına göre biçimlendirilebilen benzer nitelikli metinlerin dili örneklemek ve dili temsil gücü gibi niteliklere bağlı kalınarak oluşturulan ana metin bloklarından her biri.

Örneğin *yazın türleri* ayrı bir tabaka oluşturabileceği gibi *internet metinleri* ya da herhangi bir ana tabakaya girmeyen *çeşitli türler* (bilimsel yazılar vb.) başkaca tabakalar olarak derlemlerde yer alabilir. Ayrıca sadece araştırma yapılacak alanla ilgili tek bir tabaka oluşturulabilir.

**Alan Tanımlama:** Oluşturulan arayüzle derlemde yer alan tabaka ve metin nitelikleri ile dilbilimsel etiketleri araştırma sorularına göre biçimlendirme/belirleme.

**Metadata:** Üstveri olarak da adlandırılan *metin adı, yazar adı, yayın yılı, ilk yayın tarihi, son basım tarihi, yazının cinsiyeti, metin cümle sayısı, farklı ve aynı cümle sayısı* vb. sistem üzerinden araştırma sorularına göre biçimlendirilebilen veri girişleri (Burnard, 2004).

**İşaretleme (annotation):** Sözcük türü, cümle ve diğer nitelikler açısından derlemi etiketleme.

**Etiketleme (tagging):** Derlemde yer alan birimleri belirli nitelikler açısından işaretleme.

#### 3.3.2. Derlem Standart İşlemlerle İlgili Tanımlar

**Sözbirimleştirme (lemmatizing / lemmatization):** Derlemde yer alan her bir sözcükbirimi belirleme.

**Tekilleştirilme (deduplication):** Derlemde yer alan sözlükbiçimleri tek bir sözbüçime indirgeme.

**Biçimbirimsel / Özniteliksel Çözümleme (morphological analysis):** Derlemde yer alan her bir birimi biçimbirimsel olarak çözümleme.

**Sıklık Analizi (frequency analysis):** Derlemde yer alan her bir birimin derlemde geçme sayısını belirleme.

**Tümce Bölümleme (sentencer / parser / parsing):** Derlemde yer alan tümceleri ayrıştırma.

**n-gram Hesaplama:** Temelde bir dizide yer alan ses, hece veya sözcüğü önceleyen ya da takip eden/edecek olan dizi elemanını bulmaya ya da onu kestirmeye dayanan *n-gram* modeli söz konusu yapıların çıkarımında kullanılan yaygın bir yöntemdir. (Jurafsky ve Martin 2006). *n-gramlar* [...n-3, n-2, n-1 || n+1, n+2, n+3... (bigram, trigram...)] eşdizimli yapıların çıkarımına öncülük eden birliktelik kullanımlarının listelenbilmesi olanağını sunar.

**Eşdizimlilik Hesaplama:** “Sözcükbirimlerin anlamsal ya da dilbilimsel birlikteliklerinden kaynaklanan ve kullanım sıklığına bağlı olarak sözlükbirimsel özellik taşıyabilen söz birlikleri.” (Sterkenburg, 2003:119-120); ya da “Sözdizimsel olarak anlamsal sözcük birliktelikleri” (Malmkjaer, 2001; Hartmann ve Gregory, 1998:22, 23; Baker, ve ark., 2006) olarak tanımlanan eşdizimliliklerin çeşitli istatistiksel yöntemlerle hesaplanması (*mutual information, t-test, z test, chi-squared test, likelihood ratio*) ve listelenmesi.

#### 3.3.3. Esnek Etiket Oluşturmayla İlgili Tanımlar

**Esnek Etiket:** Yazılım arayüzüyle araştırmanın araştırma sorularına göre biçimlendirilebilen etiketler oluşturmaya karşılık gelir. Örneğin, sözcük anlambilimi ile ilgili etiketleri araştırmacı araştırma sorularına göre kendisi belirleyebilmektedir.

**Terimbilimsel Etiketleme:** Sözcükbirimlerle ilgili sözlükbirimsel yapılandırmalar: *yazımsallık* (spelling), *sesletim* (pronunciation), *çekimlenme biçimleri* (inflections), *sözcük türü* (word class), *anlam/lar* (senses), *tanımlama* (definition), *tanıklama* (examples), *kullanım* (usage), *diğer türetimler* (run-ons), *köken bilgisi* (etymology) gibi araştırma sorularına göre derlemlerde etiketlenebilen yapılandırmalardır (Jackson 2002:26-27; Hanks 2003:56-57).

#### 3.3.4. Derlem Sorgulama ile ilgili Tanımlar

**Serbest Sorgu (query):** Derlemlerde standart olarak yer alan sorgulama arayüzü.

**Çoklu Sorgu (multiple query):** Sorgu sonuçları üzerinden de arama yapabildiği arayüzü.

**Sorgu Geçmişi (query history):** Önceki derlem sorgularına kolayca ulaşılabilmesini sağlayan sorgu geçmişi belleği.

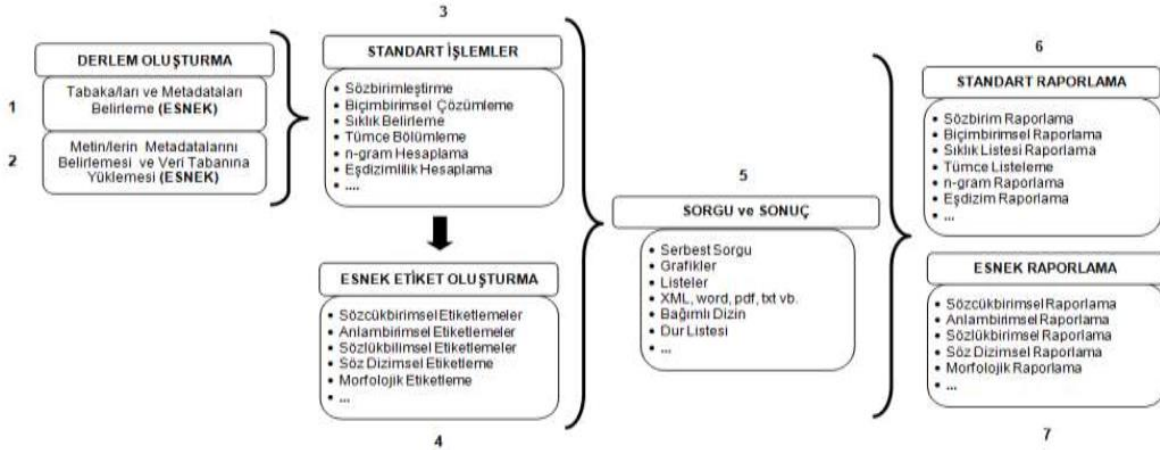
<sup>9</sup> Tanımlar, Malmkjaer, 2001 ve Baker, ve ark., 2006'dan özetlenerek alınmıştır.

**Dur Listesi (stop list):** Sadece seçilmiş belirli sözcükbirimlerin sorgulanmasına olanak veren sorgulama arayüzü.

**Bağımlı Dizin (concordance):** Aranan sözcüğün bağlamlı tümceler olarak listelenmesi.

**Raporlama:** Derlemde etiketlenmiş-ışaretlenmiş her bir birimi çeşitli raporlama arayüzleriyle (metin, grafik, harita, liste vb.) çıktı olarak dışa aktarma.

Yukarıda yapılan tanımları doğrultusunda oluşturulacak olan KKDP'nin amaca uygun araştırma sorularına bağlı olarak aşağıdaki gibi bir derlem prosedürü oluşturmaktadır (bk. Şekil 1).



Şekil 1. KKDP veri işleme algoritması

### İşlem Tanımları (Şekil 1'e göre)

İşlem	İşlem Açıklaması
(1 ve 2) Derlem Oluşturma	Derlem oluşturacak araştırmacı derlemde yer alacak tabakaları belirleyebilmekte ve bunlarla ilgili metadataların girişlerini <i>esnek alan tanılama</i> arayüzüyle oluşturabilmektedir. Aynı <i>esnek alan tanılama</i> işlemi, girişi yapılacak/yapılan metinler için de geçerlidir.
(3) Standart Derlem İşlemleri	Tüm derlemler için ortak çıktılardan olan birtakım özellikler ( <i>sözcükbirleştirme, biçimbirimsel çözümleme, sıklık belirleme, tümce bölümleme, n-gram belirleme ve eşdizimlilik hesaplama gibi</i> ) metin girişlerinin ardından otomatik olarak sistem tarafından yapılabilmekte ve veri tabanı kayıtları oluşturulabilmektedir.
(4) Esnek Etiket Oluşturma	Derlem, oluşturacak araştırmacı araştırmanın amacına göre bir önceki aşamada sistem tarafından oluşturulan yapıları esnek etiketler belirleyerek etiketleyebilmektedir.
(5) Sonuç ve Sorgu	Araştırmacı etiketlediği yapıları Sorgu-Sonuç ekranından listeleyebilmektedir.
(6) Standart Raporlama	Araştırma sorularına göre 3. aşamadaki derlem işleme sonuçlarını raporlayabilmektedir.
(7) Esnek Raporlama	Aynı şekilde araştırmacı seçimli bir raporlama ile sistem üzerinden amaca uygun olarak etiketlenmiş/ışaretlenmiş yapıları raporlayabilmektedir.

### 3.3.5. Örnek Bir Uygulama Olarak Sözlük Bilim Terimleri<sup>10</sup>

KKDP TBM kullanılarak gerçekleştirilen “Türk Sözlükbilimi Terminolojisi Üzerine Derlem Tabanlı Bir

Araştırma” başlıklı çalışmada 1003 metin, 42.831 cümle, 703.986 sözcük ve 1.616 tekil söz içeren bir özel alan derlemi oluşturulmuş ve Türk sözlükbilimi terminolojisi araştırma konusu yapılmıştır (<http://tsd.ogu.edu.tr/index.php>).

<sup>10</sup> Erdoğan Boz, Ferdi Bozkurt ve Fatih Doğru, Corpus-Based Research on Terminology of Turkish Lexicography (CBRT-TURKLEX)

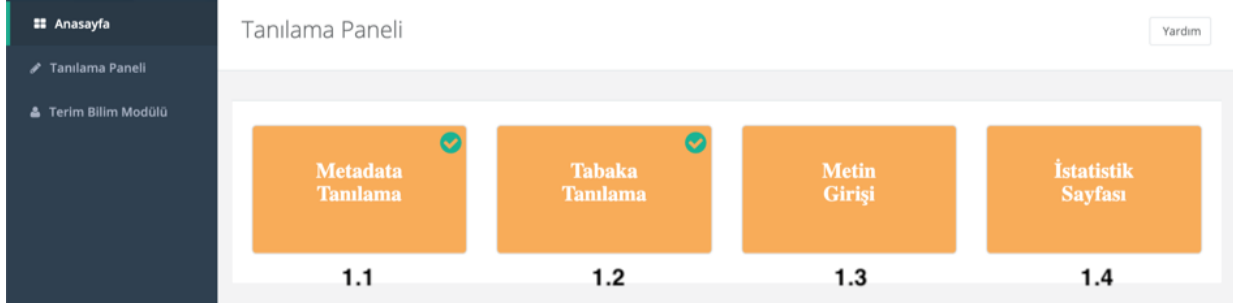
Lexikos 28 (AFRILEX-reeks/series 28: 2018): 428-439 <http://lexikos.journals.ac.za>; <https://doi.org/10.5788/28-1-147>

Söz konusu araştırmanın ana hatları örnek bir uygulama olarak aşağıdaki gibidir:

### 3.3.5.1. Derlem Oluşturma Aşamaları

Bu aşamada (Ekran 1.1) Metadata Tanılama panelinden oluşturulacak özel amaçlı derlemde yer alacak

metinlerin metadataları belirlenmekte, (Ekran 1.2) Tabaka Tanılama panelinde derlemin tabakaları belirlenmekte, (Ekran 1.3) Metin Girişi panelinde ise metin girişleri yapılmakta ve (Ekran 1.4) İstatistik Sayfası'nda basit istatistiksel/sayısal verilerle raporlanarak bir sonraki aşama olan *Terim Bilim Modülü*'nde terimler işleme hazırlanmaya hazır hâle getirilmektedir.

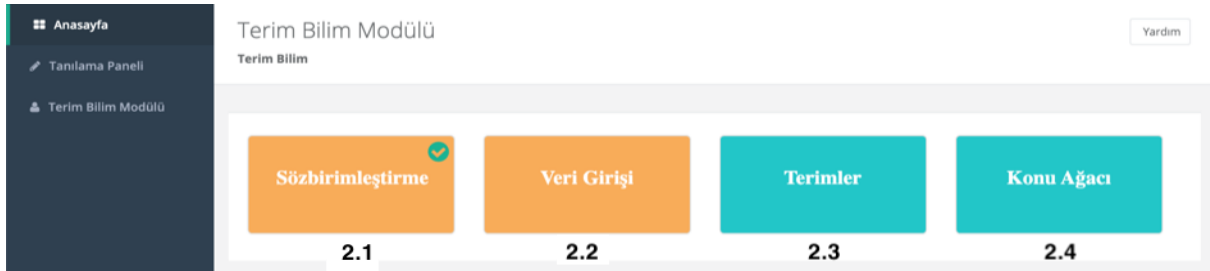


	<p><b>Ekran 1.1.</b>Bu ekranda terim bilim araştırmaları için derleme alınacak metinlerin yazar adı, soyadı, yayın yılı, metin adı, resim-görsel ya da araştırma sonucunda ulaşılması beklenenlerle ilgili diğer yapılandırmalar eklenmektedir.</p>
	<p><b>Ekran 1.2.</b>Derlemde yer alacak ana ve alt tabakalar belirlenmektedir. Bu çalışma için <i>I. Telif Eserler: makale, bildiri, kitap ... II. Çeviri Eserler: makale, bildiri, kitap vb.</i> olarak iki ana ve her ana tabaka için metin türü sayısınca alt tabakalar belirlenmiştir.</p>
	<p><b>Ekran 1.3.</b>Metadatalara bağlı olarak alt tabakalar ve bu tabakalara ait metin girişleri yapılmakta ve araştırma için oluşturulacak olan derlem hazırlanmaktadır. Metin giriş aşamasından sonra standart derlem işlemleri (söz-birimleştirme, tümcesel bölümlenme vb.) otomatik olarak sistem tarafından yapılmaktadır.</p>
	<p><b>Ekran 1.4.</b>Girişi yapılan metinlerin genel bir raporu: <i>metin sayısı, cümle sayısı, toplam kelime sayısı ve tekil kelime sayısı</i> olarak raporlanabilmektedir.</p>

### 3.3.5.2. Terim İşletme Aşamaları

Bu aşamada oluşturulan derlem üzerinden (Ekran 2.1) Sözbirimleştirme, (Ekran 2.2) Veri Girişi ekranından

terimlerin diğer yapılandırılmaları, belirlenen ve işlenmiş terimlerin raporları [(Ekran 2.3) Terimler ve (Ekran 2.4) Konu Ağacı panellerinden)] alınabilmektedir.



**Ekran 2.1.** Sözbirimleri belirlenen sözcükbirimler/terimler gövdelenecek terim listeleri oluşturulmaktadır.

**Ekran 2.2.** Terim listelerinden elde edilen sözbirimler anlam, sözcük türü, köken, eşdizim vb. yapılar olarak etiketlenerek raporlanabilmektedir.

**Ekran 2.3.** Terim olarak etiketlenip işlenen söz varlığı serbest ara-malarla listelenebilmektedir.

**Ekran 2.4.** İlişkisel söz varlığı listelerden kontrol edilerek kavram ağacı olarak düzenlenebilmektedir.

**Ekran 2.1.** Sözbirimleri belirlenen sözcükbirimler/terimler gövdelenecek terim listeleri oluşturulmaktadır.

**Ekran 2.2.** Terim listelerinden elde edilen sözbirimler anlam, sözcük türü, köken, eşdizim vb. yapılar olarak etiketlenerek raporlanabilmektedir.

**Ekran 2.3.** Terim olarak etiketlenip işlenen söz varlığı serbest ara-malarla listelenebilmektedir.

**Ekran 2.4.** İlişkisel söz varlığı listelerden kontrol edilerek kavram ağacı olarak düzenlenebilmektedir.



#### 4. Sonuç

Türkçe üzerine çalışan dil/dilbilim araştırmacıları için araştırma sorularına bağlı olarak biçimlendirilebilen, araştırmacıya özel, esnek, kolay ulaşılabilir, veri tabanı destekli ve bu çerçevede araştırma sonuçlarının sağlıklı bir şekilde raporlanabildiği bir derlem platformu olan “Kendi Kendine Derlem Oluşturma Platformu”nun (KKDP) Terim Bilim Modülü’nü (TBM) Türkçenin terim bilim çalışmalarında hızlı güvenilir bir veri işleme platformu olarak amaca uygun bir yapılandırma sunmaktadır.

#### Kaynaklar

Baker Paul, Hardie A. and McEnery T., 2006. *A glossary of corpus linguistics*. Edinburg University Press.

Barnbrook, G., 1996. *Language and computers: a practical introduction to the computer analysis of language*. Edinburgh: Edinburgh University Press.

Boz, E., Bozkurt, F. ve Doğru F., 2017. "Türk sözlükbilimi terminolojisi üzerine derlem tabanlı bir araştırma". III. Uluslararası Sözlükbilimi Sempozyumu Bildiri Kitabı. Eskişehir Osmangazi Üniversitesi Basımevi: Eskişehir.

Boz, E., Bozkurt, F. and Doğru, F., 2018. Corpus-based research on terminology of Turkish lexicography (CBRT-TURKLEX) *Lexikos 28 (AFRILEX-revues/series 28: 2018): 428-439* <http://lexikos.journals.ac.za>; <https://doi.org/10.5788/28-1-147>

Burnard, L., 2004. *Metadata for Corpus Work*. (<http://users.ox.ac.uk/~lou/wip/metadata.html>)

Hanks, P., 2003. *Lexicography. Computational Linguistics*. (Ed. Ruslan Mitrov). Oxford University Press.

Hartmann, R. R. K. and Gregory, J., 1998. *Dictionary of lexicography*. London: Routledge.

Jackson, H., 2002. *Lexicography: an introduction*. Routledge. USA.

Jurafsky D. and Martin, J.H., 2006. *Speech and language processing: an introduction to speech recognition, computational linguistics and natural language processing*. (<http://web.mit.edu/6.863/www/fall2012/readings/ngrampages.pdf>) ET: 15.08.2019

Kennedy, G., 1998. *An introduction to corpus linguistics*. Longman.

Malmkjaer, K., 2001. *Linguistics encyclopedia*. Florence. Routledge London. United Kingdom.

McEnery T. and Wilson A., 1996. *Corpus linguistics*. Edinburg: Edinburg University Press.

McEnery, T., Xioa, R., Tono, Y., 2006. *Corpus-based language studies -an advanced resource book-*. London: Routledge.

Özkan, B., 2016. Tübitak 1005 Ulusal Yeni Fikirler ve Ürünler Araştırma Destek Programı 111E791 nolu “Türkçe için Kendi Kendine Derlem Platformu Oluşturma” başlıklı proje raporu.

Özkan, B., 2013. Yöntem ve Uygulama Açısından Türkiye Türkçesi Söz Varlığının Derlem Tabanlı Sözlüğü. *bilig. Sayı: 66 Yaz'13*.

Özkan, B., 2010. Türkçenin Öğretiminde Sıfatların Eşdizim Sözlüğü: Yöntem ve Uygulama, *e-International Journal of Educational Research*. Volume 1. No 2. Autumn 2010: 51-65.

Say, B., 2003. Türkçe İçin Biçimbirimsel ve Sözdizimsel Olarak İşaretlenmiş Ağaç Yapılı Bir Derlem Oluşturma. Proje No: 199E026. Proje Raporu-Ek-b: Türkçe Ağaç Yapılı Derlem Kılavuzu. Ankara.

Say, B., Umut, Ö., Oflazer, K., 2002. Bilgisayar Ortamında derlem Geliştirme Çalışması. *Akademik Bilişim Konferansı*. Konya.

Sinclair, J., 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sterkenburg, P., 2003. *A practical Guide to Lexicography*. Philadelphia. PA. USA: John Benjamins Publishing.

Stevens, V., 1995. Concordancing with Language Learners: Why? When? What?. *CAELL Journal vol. 6 no, 2, 2-10*.

Steven, B., Ewan, K. and Edward, L., 2009. *Natural Language Processing with Python -Analyzing Text with the Natural Language Toolkit-*. O'Reilly Media, <http://www.nltk.org/>

<http://kkd.mersin.edu.tr/>

<http://tsd.ogu.edu.tr/index.php>

<http://ucrel.lancs.ac.uk/claws/>

[http://www.lancaster.ac.uk/fss/courses/ling/corpus/blue/diy\\_top.htm](http://www.lancaster.ac.uk/fss/courses/ling/corpus/blue/diy_top.htm)

<http://www.uow.edu.au/~dlee/software.htm>

<http://www.tpe.gov.tr>, [tr.espacenet.com](http://tr.espacenet.com) ve <http://www.epo.org/>

<https://cqpweb.lancs.ac.uk/>