

# Yöntem ve Uygulama Açısından “Türkiye Türkçesi Söz Varlığının Derlem Tabanlı Sözlüğü”\*

Bülent Özkan\*\*

## Öz

Sözlükbilim yöntem ve uygulamaları geçmişten bugüne bilişim teknolojilerine paralel olarak bir takım yenilikleri beraberinde getirmiştir. Geçmişte fişleme, sınıflandırma, yeniden yazma vb. yöntemlerle dil nesnesini kendine konu edinen sözlükbilim, günümüzde derlem (corpus) adı verilen dilsel bütünceler üzerinden araştırma nesnesine yönelmiştir. Alanyazına baktığımızda *Oxford, Collins, Longman, Cambridge* benzeri sözlüklerin derlem tabanlı (corpus-based) uygulamalar sonucu oluşturdukları görülür. Öte yandan genel hatlarıyla Türkiye Türkçesinin Söz Varlığını (TTSV) barındıran sözlüklerin derlem dilbilimin ve çağdaş sözlükbilim çalışmalarının paralelinde olmadığı gözlemlenir. Bu çalışmada, ‘Türkiye Türkçesi Söz Varlığının Derlem Tabanlı Sözlüğü’nü oluşturmanın bir yöntem olarak tanıtımı amaçlanmıştır. Bu çerçevede derlem dilbilimin ilke ve yöntemlerinden yararlanılarak 50 milyon (+/-) sözcük içeren bir *model derlem* oluşturulmuştur. Bu derlem aracılığıyla ‘Türkiye Türkçesi Söz Varlığının Derlem Tabanlı Sözlüğü’nün nasıl oluşturabileceğini ortaya koymak açısından, uygulanan yöntemin bir yönergesini verilmiştir. Çalışmada öncelikle sözlükbilim, sözlükbilimin dünü ve bugünü üzerinde kısaca durulmuş, ardından yöntem ve *model derlem* tanıtılmış ve son olarak da örnek bir uygulamayla deneyimlenen yöntemin sonuçları aktarılmıştır. Çalışma, Türkçenin sözlükbirimsel topografyasının oluşturulması, Türkçenin ana dili ve

\* Bu çalışma TÜBİTAK-SOBAG-11K516 nolu Ulusal Araştırma Projesi’nin alt yapısına dayanmakta ve TÜBİTAK tarafından desteklenmektedir.

\*\* Yrd. Doç. Dr., Mersin Üniversitesi, Eğitim Fakültesi, Türkçe Eğitimi Bölümü – Mersin / Türkiye  
ozkanbulent@mersin.edu.tr

yabancı dil olarak öğretilmesinde söz varlığının güncel ve geçerli veriler doğrultusunda sözlükbirimselleştirilmesi anlamında alana önemli katkılar sağlayacaktır.

### Anahtar Kelimeler

Türkiye Türkçesi, söz varlığı, derlem dilbilim, sözlükbilim, Türkçe öğretimi

## 1. Giriş

*Sözcükbirimler* (lexeme) *sözlükbilimin* (lexicology) araştırma nesnesini oluşturur. Bu anlamda sözcükbirimin kendisi ve onun dildeki *örüntüsü* (patterns) doğrudan sözlükbilimin ilgi alanına girer. Dolayısıyla sözlükbilimin ne olup ne olmadığını anlamak için sözcükbirimin kendisinin ve onun dildeki örüntüsünün iyi ayrıştırılması gerekir.

Bilindiği gibi sözcükbirimler dilde anlam taşıyıcı birimler olarak yer alırlar. Genel kabul gören anlayışla, bir sözcükbirimin taşıdığı yalın anlam bir yana bırakılırsa, sözcükbirimsel anlamın asıl belirleyicisi sözcükbirimlerin diğer sözcükbirimlerle dizgelerde *birliktelik kullanımları* (co-occurrence) ve *eşdizimlilikleridir* (collocation). Söz konusu birliktelik kullanımları ve eşdizimlilikler sözcükbirimlerin anlamsal ve biçimsel görünümelerini dizgesel ve dizimsel ilişkilerle dilsel örüntüde belirginleştirirler.

Sözcükler her şeyden önce birer dilsel göstergedir. Dil ise, yine genel kabul gören biçimiyle, 'göstergeler bütünü' olarak kavramlaştırılabilir. Bu bütün, *sıradüzen* (hierarchy) bir örüntünün yanında üst ve/ya alt kavramlılık, kaplam ve işlem oluşturma, eş/yakın ve karşıt anlamlılık, anlamsal değişimler/dönüşümler, tek ve çok anlamlılık, türetimler, düzenleyimler (birleşik sözcükler ve deyimler) vb. görünümlemlerle bir var oluş sergiler (Günay 2007).

Öte yandan kültürel etkileşimler, dilin birincil işlevi olan bildirişimin de zorlamasıyla, diller arasında tüm bu sözcükbirimsel var oluşu karşılıklı olarak aktarma çabasını doğurmuştur. Sözlükbilimin tarihine göz attığımızda itici güç olarak bu çabaları açıkça gözlemleyebiliriz.<sup>1</sup> Bugün ana dili ve yabancı dil öğretiminde bu karşılıklı aktarım çabalarının önemli bir işleve sahip olduğunu söyleyebiliriz.

Sözlükler (lexicon) yukarıda kısaca değerlendirdiğimiz bu sözcüksel var oluş çerçevesinde belirli amaçlar doğrultusunda hazırlanmış koleksiyonlar olarak değerlendirilir. Bu anlamda sözlükbilimde sözlükler, içerik ve oluşturulma yöntemlerine göre:

- Betimlemeli/kuralcı sözlükler,
- Eşzamanlı/artzamanlı sözlükler,
- Genel/teknik sözlükler,
- Genel kullanım/öğrenim amaçlı sözlükler,
- Tekdilli/ikidilli sözlükler olarak sınıflandırılır (Kocaman 1998:111).

Öte yandan, içerik ve oluşturulma yöntemlerine göre sözlükler üzerine başkaca bir sınıflandırma ise:

- Bir ya da birden çok dilin sözvarlığını işlemesine [tekdilli/ikidilli sözlükler],
- Abece sırasının temel alınıp alınmamış olmasına [abecesel sözlükler, kavram (ya da kavram alanı) sözlükleri],
- Ele alınan sözvarlığının niteliğine [genel sözlükler (ortak dil, yazı dili, ansiklopedik sözlükler), lehçebilim sözlükleri, eşanlımlı, eşadlı, ters anlamlı öğeler sözlüğü, ... argo sözlükleri, ... deyim ve atasözü sözlükleri, anlatımbilim sözlükleri, sanatçı ve metin sözlükleri] göre yapılan sınıflandırmadır (Aksan 2000:75).

İçerik ve oluşturulma yöntemi ne olursa olsun sözlükbilim araştırma nesnesi olarak sözcükbirimleri ve onların örüntülerini konu edinir. Bu noktada üzerinde durulması gereken sözlükbilimin söz konusu sözcükleri nasıl ve hangi yöntemlerle kendisine konu edindiğidir.

### 1.1. Sözlükbilimin Dünü: Fişleme, Sınıflandırma, Yeniden Yazma

Sözlükbilimin tarihine göz attığımızda dilbilimciler ve sözlükbilimcilerin çalışmalarında kullandıkları dilsel malzemenin genellikle ana dili kullanıcılarının sezgilerine dayandığı ve söz konusu çalışmaların çoğunlukla yanlış veya en azından dilsel gerçeği tam olarak yansıtamayan, dilsel geçerlilikten uzak sonuçlar ortaya koyduğu gözlemlenir. Ayrıca dilbilimcilerin ve sözlükbilimcilerin kişisel tercihleri ve sezgisel yaklaşımları zorunlu olarak her dilci için farklılaşan bir takım sonuçlarla karşımıza çıkmaktadır (McEnery 2006:145).

Günümüzde tanık toplamak, kişisel bilgilere dayanmak, fiş indeksi oluşturmak vb. uygulamaların sözlük oluşturmada işlevi olmayan yöntemler olduğu kabul edilmektedir (Atkins vd. 1995: 85). Geçmişten bugüne sözlükbilim çalışmalarına baktığımızda, bilişim teknolojilerine koşut olarak, bu çalışma alanında yeni birtakım açılımların gerçekleştiği gözlemlenir. Tarihsel süreçte dilsel verileri sözlükleştirmede kullanılan *fişleme, sınıflandırma, yeniden yazma* vb. yöntemler bugün için geçerliliğini yitirmiştir. Bu yöntemler yerlerini artık sözlük çalışmalarının amaçlılığına uygun olan gerçek zamanlı dilsel verilerin bütüncül olarak incelenip yordanabildiği sayısal (dijital) platformlara bırakmıştır.

Son zamanlarda bilişim teknolojilerindeki gelişimden dil üzerine yürütülen çalışmalar da kendine düşen payı almıştır. Bugün bilgisayarlar aracılığıyla yürütülen dil çalışmalarında var olan yöntem ve uygulamalar, daha hacimli verileri çeşitli amaçlar doğrultusunda işleme ve yordama olanağını dil araştırmacılarına fazlasıyla sunmaktadır.

*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, anlam-bilim, kullanımbilim, sosyo-dilbilim, söylem çözümlemesi, biçembilim ve yazınbilim çalışmalarında yoğunlukla kullanılmaktadır (McEnery vd. 2006: 80-122, Kennedy 1998: 208-310).

Diğer taraftan DDİ paralelinde yürütülen dilbilim uygulamalarını sözlükbilim ve sözlük çalışmaları çerçevesinde değerlendirdiğimizde Türkçenin söz varlığını barındıran sözlüklerin bu çalışma alanının uzağında yer aldığı görülür.

## 1.2. Sözlükbilimin Bugünü: Derlem Dilbilim ve Sözlükbilim

Bilgisayar bilimlerindeki gelişmeler, geniş hacimli derlemlerin oluşturulmuş olması ve buna bağlı olarak gelişen yeni teknolojiler, kişisel tercihlere dayalı ve sezgisel nitelikli çalışmalarla kıyaslandığında, dilin betimlenmesinde veriye dayalı yaklaşımları bugün için oldukça geçerli kılmıştır.<sup>2</sup>

Günümüzde derlemler ve bu derlemler için geliştirilmiş yazılımlar, oldukça güvenilir dilsel verileri sağlamada ve oluşturulacak sözlük için gerekli örüntüyü elde etmede oldukça işlevseldir (McEnery ve Wilson 2004). Ancak söz konusu süreçler otomatik olarak bir sözlüğü tamamlayabilecek yeterlilikte değildir. Bu noktada dilbilimsel bir sürecin sözlükbilim çalışmalarında yerini alması gerekir. Gerçekte sözlükbilim çalışmaları ve bilgisayar kullanımı arasındaki ilişkiyi *bilgisayar destekli* (computer-aided) olarak adlandırmak da mümkündür (Jackson 2002: 169-171).

Bir derlemden geçerli ve güvenilir dilsel verilerin elde edilebilmesi, o derlemin belirli nitelikleri karşılıyor olmasıyla gerçekleşebilmektedir. Amaç, dili betimlemek ve bu betimlemelerden işe yarar bilgiler üretmekse, her şeyden önce, betimlenecek dilin sağlam kuruluşlu bir derlemle örneklenmesi gerekir.

Alanyazında bir derlemin dili betimler nitelikte olduğunun kabulü, o derlemin söz konusu dili ne kadarıyla örnekleyebildiği ile paralel bir değerliktir. Bu anlamda, bir derlem oluşturulurken üzerinde karar verilmesi gere-

ken en önemli nokta söz konusu derlemin tasarımıdır. Bir dilin tipik ve karakteristik kullanımlarını ortaya çıkarmak amaçlı bir derlem tasarımı yapılırken bu derlemin *temsil gücü* (representativeness) ve buna bağlı olarak *türsel dağılımı/dağılımsallığı* (balance) ve *örnekleme seçimi/örnekleme* (sampling) önemli sorunlar olarak karşımıza çıkar (Sampson vd. 2005: 174). Bu noktada belirleyici olan diğer bir ölçüt ise araştırmacının araştırma sorusudur.

Elbette ki bir derlem farklı amaçlar için oluşturulabilir. Alanyazında söz konusu bu farklı amaçlılıkları karşılayan derlemlerin varlığını görebilmekteyiz. İçlemleriyle ve amaçlılıklarıyla koşut olarak alanyazında yazılı ve/ya sözlü içeriğe sahip ve belirli bir dilde dilsel çeşitliliği yansıtabilen *genel derlemler* (general corpora); ‘petrokimya derlemi’ ya da ‘bilgisayar bilimleri derlemi’ benzeri *özel alan derlemleri* (specialized corpora); belirli bir yazı dilini, belirli dönem ve örneklemlerle içeren *yazılı derlemler* (written corpora); genel derlemlerin içleminde ya da bağımsız olarak sözlü dili temsil için tasarımılanan *sözlü derlemler* (spoken corpora); eş süremliler olarak dilsel veriler içeren *eş süremliler derlemler* (synchronic corpora); belirli bir dilin farklı zaman aralıklarıyla içlemlendiği *art süremliler derlemler* (diachronic veya historical corpora); yine ana dili ve yabancı dil olarak belirli bir dilin öğretiminde temel verilerin derlenmesi amaçlı tasarımılanan *öğrenici derlemleri* (learner corpora) ve diğer derlem türlerinden daha hacimli ve ayrınıltandırılmış, kapsamlı derlemlerden olan *izlem derlemleri* (monitor corpora) ile karşılaşırız (McEnery vd. 2006: 59-70).

Sözünü ettiğimiz bir dilin doğal ortamlarından yazılı ve sözlü olarak derlenen verileri temel alan ve kendine özgü yöntem ve uygulamalarla ortaya koyulan sağlam kuruluşlu derlemler, özellikle dil öğrenimi ve öğretimi alanında sezgisel örnekçelerden ve ikincil verilerden yalıtılmış ‘gerçek zamanlı’ dilsel verileri içermesi bakımından alana büyük katkılar sağlamaktadır (Özkan ve Gündoğdu 2011).

## 2. Yöntem

### 2.1. İçerik ve Nitelik Olarak **TÜRKÇE DERLEM**

*Derlem* (corpus), *dilbilimde* bir dili belirgin dilbilimsel ölçütlere göre örnekleme amaçlı olarak o dile ait yazılı ve sözlü metinlerden seçilerek belirli standartlara göre düzenlemiş metin parçalarının bütünü (McEnery vd. 2006: 4) olarak tanımlanabilir. Bir derlem için gerekli ölçütlerin başında ise oluşturulduğu dili *örnekleme* ve o dili *temsil gücü* gelir. Bunun yanı sıra *belirli bir sınıra* (finite size) sahip olması ya da olmaması (dynamic size), *bilgisayarlarca okunabilir* (machine-readable) bir yapıda hedef

dili betimlemede ölçünlü bir *başvuru kaynağı* (standart reference) niteliği taşıması bir derlemin oluşturulmasında önemli ölçütlerdendir (McEnery ve Wilson 2004: 29-32).

Bir dilin, kullanıldığı karakteristik ortamlarından biri de *yazın dilidir*. Yazın dili özelliği gereği o dilin genel yapısını betimlemek için önemli veriler sağlar. Öte yandan, yazı dili araştırma nesnesi olarak ölçünlü ve araştırmacılar için kolay erişilebilirdir. Bir derlemden bu anlamda derlenen verilerin güvenilirliği yapılacak betimlemenin de güvenilirliği anlamına gelmektedir. Derlem dilbilimde, derlemin içeriği ve niteliği kadar derlem üzerinde yapılacak uygulamaların da önemi büyüktür.

Derlemin tasarımında, örneğin, içerikte yer alan metinlerin türü, metin sayısı, belirli metin parçalarının seçimi, metin parçaları arasından örneklem seçimi, örneklemin genişliği vb. tasarımın bilinçli yapılp yapılmadığının ölçüsüdür (Sampson vd. 2005: 174). Bunun yanı sıra bir derlemin tasarımında metin türleri, tabakalar, tabaka oranları ve zaman aralığı gibi unsurlar önemli ölçütler olarak karşımıza çıkar.

Derlemlerde yer alan metinlerin seçiminde birtakım ölçütler gözetilmektedir. Bunlardan en bilinen ve deneyimlenenleri, birbiri arasında çeşitli avantaj ve dezavantajları da barındıran, *uzman görüşlerine dayanma* (literary/academic merit), *tesadüfi seçim* (random selection), *okunurluk* (currency), *ulaşılabilirlik* (availability), *demografik örnekleme* (demographic sampling), *deneysellik* (empirical) ve *daha geniş bir kaynaktan seçim* (a broad range of sources)'dir (Summers 1993:191). Bu seçimlerden biri ya da birkaçı bir derlem oluşturulurken kullanılabilir.

Bir derlem oluşturulurken alınılan metin büyüklükleri (text size) üzerine ortaklaşan yaklaşım ise her metin kaydından 40.000 sözcüklük bir seçim yapmaktır.<sup>4</sup> (Summers 1993:199). Oluşturulan *TÜRKÇE DERLEM* de söz konusu yaklaşımlar göz önünde bulundurulmuştur. Araştırma sorusu gözetilerek derleme alınılan metinlerde 40.000 sözcük sınırı korunmuş, derlemin oluşturulmasında uzman görüşleri, okunurluk, ulaşılabilirlik ve demografik örnekleme yöntemleri kullanılmıştır.

Bu anlamda 35 milyon sözcük (+/-) içeren *Türkçe Derlem-2* (TD-2)'den evrilerek oluşturulan *TÜRKÇE DERLEM* metin türleri açısından zenginleştirilmiş, biçimbirimsel olarak *işaretlenmiş* (tagging) ve *etiketlenmiştir* (annotation). Bu açıdan *TÜRKÇE DERLEM*, *Türk Yazın Dili*'nde yer alan metin türlerinin tamamını, yazın dilinde yer aldıkları oranlar göz önünde bulundurularak, örnekleyebilecek bir yapıda tasarlanmıştır.

Bu çerçevede *TÜRKÇE DERLEM* temel olarak üç tabakadan oluşmaktadır.

Bu tabakalardan **birincisinde** basılı kitaplar yer almaktadır. Kitaplardan metin seçiminde sözcük sayısı sınırı ortalaması 40.000'dir [Longman/Lancaster English Language Corpus ve TNC (<http://www.tnc.org.tr/>)]. Bu çalışmada, birinci tabakada yer alan metinler aynı zamanda evreni (Türk Yazın Dili) en iyi şekilde temsil edecek tabaka olarak öngörülmüştür. Zaman aralığı olarak ise metinlerin %85'i 1970-2009 dönemine aittir.

Derlemin **ikinci** tabakasında 2000-2009 yıllarına ait internet metinleri<sup>5</sup> konu duyarlı olarak (haber, yaşam, kültür-sanat, makale) yer alırken; **üçüncü** tabakada ise (aşağıdaki tabloda yer aldığı biçimiyle) çeşitli türlere ait sözcük sınırı ortalama sayısı gözetilerek sayısallaştırılan basılı kitaplar yer almaktadır. Bu anlamda *TÜRKÇE DERLEM* toplam 50 milyon (+/-) söz içerir bir yapıda tasarlanmıştır (*bk.* Tablo 1).

**Tablo 1:** *TÜRKÇE DERLEM*'de Metin Türleri, Tabakalanma ve Zaman Aralığı

Tabakalar		Tabaka Oranı	Zaman Aralığı		
<b>Türk Yazın Dili</b>					
1	<ol style="list-style-type: none"> <li>1. Roman</li> <li>2. Şiir</li> <li>3. Deneme-Eleştiri</li> <li>4. Öykü</li> <li>5. Tiyatro</li> <li>6. Anı</li> <li>7. İnceleme-Araştırma</li> <li>8. Mizah</li> <li>9. Sohbet-Söyleşi-Makale</li> <li>10. Gezi</li> <li>11. Mektup</li> <li>12. Biyografi</li> <li>13. Günlük...</li> </ol>	%40	Cumhuriyet Dönemi Türk Yazını (1923-2006)		
<b>İnternet Metinleri</b>					
2	<b>Haber vb.</b> 1-Siyaset 2-Ekonomi-Finans 3-Dünya-Güncel 4-Hava Durumu 5-Spor ...	%20	<b>Kültür-Sanat vb.</b> 1-Sağlık 2-Kitap 3-Sinema 4-Tiyatro ...	%5      %40	İnternet Metinleri (2000-2009)

	<b>Yaşam</b> 1-Teknoloji 2-Eğitim 3-Magazin ...	<b>%5</b>	<b>Makale</b> (Köşe Yazıları) ...	<b>%10</b>	
<b>Çeşitli Metinler</b>					
3	beslenme, bilgisayar bilimleri, biyoloji, davranış bilimleri, dilbilim, göstergebilim, efsane, eğitim bilimleri, ekoloji, enformasyon, estetik, evrim, felsefe, fıkra, fizik, geometri, hukuk, iktisadi bilimler, istatistik, kimya, kişisel gelişim, mantık, masal, matematik, mimarlık, mitoloji, mühendislik (çeşitli), müzik, nutuk, psikoloji, resim, sanat, sinema, siyaset, sosyoloji, strateji, tarih (çeşitli), tıp ...			<b>%20</b>	(1940-2009)
<b>Toplam</b>				<b>% 100</b>	

### 2.1.1. TÜRKÇE DERLEM'in Sözdizimsel, Sözcükbirimsel ve Biçimbirimsel Yapısı

Derlemde yer alan toplam sözcük sayısı 50 milyon (+/-)'dur. Derlem tümcesel olarak bölümlenmiştir (parsing), her tümce için ait oldukları ana tabakalar ve alt tabakalar [*Yazın Türleri* (roman, şiir, öykü, anı vb.), *İnternet Metinleri* (haber, yaşam, kültür-sanat vb.), *Çeşitli Türler* (bilgisayar bilimleri, biyoloji, davranış bilimleri vb.) *yazar adı, eser adı, yayın yılı* vb. metin dışı unsurlar açısından etiketlenmiştir. Derlem üzerinde bu aşamadan sonra *sözbirimselendirme* (lemmatizing/lemmatization) işlemi yapılmış ve sözbirimselendirme *tekilleştirilmiştir* (deduplication).

Derlemin biçimbirimsel anlamda öznitelişsel çözümlemesinde Kemal Oflazer'in geliştirdiği biçimbirimsel çözümleme aracı<sup>6</sup> kullanılmış, tekilleştirilmiş sözcükbirimler öznitelişsel olarak çözümlenmiştir. Bu aşamadan sonra *TÜRKÇE DERLEM*'de veri tabanında işaretli tümceler biçimbirimsel çözümlemelerle eşleştirilmiş ve sözcükbirimler tümcesel olarak sorgulanabilir hale getirilmiştir.

#### 2.1.1.1. Tümce ve Tümce Etiketleme Örneği

<<Bütün malzemeleri derin bir kaba koyup, pütür kalmayacak şekilde doğrayın.>>

<3.tabaka> **tabaka**

<beslenme> **metin türü**

< leman cılızoğlu eryılmaz> **yazar adı**

< türk mutfağından seçme yemekler> **eser adı**

<1997>... **yayın yılı**



### 2.1.1.2. Özniteliksel Çözümleme Örneği

<<Bütün malzemeleri derin bir kaba koyup, pütür kalmayacak şekilde doğrayın.>>

**bütün**

bütün+Adj

**malzemeleri**

malzeme+Noun+A3pl+Pnon+Acc

**derin**

derin+Adj

**bir**

*bir+Adverb*

bir+Det

bir+Num+Card

**kaba**

kap+Noun+A3sg+Pnon+Dat

**koyup**

koy+Verb+Pos^DB+Adverb+AfterDoingSo

**pütür**

pütür+Noun+A3sg+Pnon+Nom

**kalmayacak**

kal+Verb+Neg^DB+Adj+FutPart+Pnon

**şekilde**

şekil+Noun+A3sg+Pnon+Loc

**doğrayın**

doğra+Verb+Pos+Imp+A2pl

Bu biçimbirimsel ve sözcükbirimsel çözümlemelerin yanı sıra özel yazımı olan eski dile ait sözcükbirimlerin sorgulanmaları 'serbest sorgulama' (derlem sorgu arayüzünden) ile ve ilgili sözcükbirim olası yazımsallıkları göz önünde bulundurularak yapılmaktadır. Biçimbirimsel çözümleme her ne kadar %99 başarımla çalışsa da çözümleyemediği yapılar (<cigaraları ???> | <yontmağa ???> | <selâmü-naleyküm ???> | <lövantenlik ???> vb. ile özel isimler ve OCR hataları nedeniyle yanlış okumalar) ??? işaretiyle sistem tarafından tutulmakta ve kontrolden geçirilmektedir.<sup>7</sup>

### 2.2. Bir Derlemden Sözlükbirimsel Veri Çıkarımı

Sözlükbilim uygulamalarında belirli bir dilde oluşturulmuş derlemlerden veri çıkarımında standartlaşmış birtakım araç ve yöntemler kullanılır. Bu araç ve yöntemlerden *sözcük sıklığı çıkarımı* (frequency) ve *bağımlı dizin oluşturma* (concordance) en yaygın kullanılanlardandır. Öte yandan, daha

karmaşık ve üst düzey araç ve yöntemler de derlem araştırmalarında yerini almıştır. Bunlar ise: *sözbirimleştirme* (lemmatizing/lemmatization), *sözcük türü etiketleme* (part-of-speech veya tagging), *cümle ayırlama* (parsing) ve *eşdizim çıkarımı* (collocation) gibi uygulamaları içermektedir (Atkins ve Clear 1992:3-4). Sözlükbilim uygulamalarında bu sıralanan derlem araçları amaca uygun olarak etkin bir şekilde kullanılmaktadır.

Sözlükbirimsel yaklaşımla, anlambirimsel ve sözdizimsel ilişkiler çerçevesinde, bir derlemden sözcükbirim araştırmasında diğer ilişkisellikler bir yana bırakılırsa kaynak metnin alıntılanma nitelikleri (yazarı, yayın tarihi, türü vb.), anahtar sözcüğün dizisel (paradigmatic) ilişkiselliği (eş ve yakın anlamlılıkları vb.), anahtar sözcüğün bir derlemde bağıl sıklık istatistikleri ve tekil olarak anlamları ile sözcüğün bir derlemde anlamlı eşdizimsel görünüşleri oldukça önemlidir (Atkins 2003:252).

Sözlükbilim araştırmalarında dilsel bileşenlerin tanınması öncelikli bir uygulamadır. Sıra-düzen ve ilişkisel olarak sözlük madde başları belirli aşamalarla gerçekleştirilmektedir. Bu aşamalardan ilki sözlükbilim çalışmalarında madde başlarının seçimi ve oluşturulmasına, *sözbirimleştirmeye*, karşılık gelmektedir. Bu aşama, aynı sözcük birimden türetmelerin yığılanarak sözcük listelerinin oluşturulmasını ve biçimbirimsel çözümleme uygulamalarını kapsar. Tüm bunlar, sözlükbilim uygulamaları için geliştirilmiş bir takım özel yazılımlarla gerçekleştirilmektedir. Her ne kadar bugün var olan Türkçe sözlükler söz varlığının belirlenmesinde yeterli veri sağlayabilecek güçteyse de dilde var olan yeni türetme sözcüklerin [*türenti* (neologism)] belirlenebilmesinde, derlem üzerinde nispeten ikincil bir uygulama olarak düşünülmesine karşın *sözbirimleştirme* çalışması yapılması da gerekir.

İkinci aşama, araştırma konusu olan derlemde edinilen madde başlarının söz varlığında (lexicon/vocabulary) *gerçekleşme sıklığını* (frequency) belirleme aşamasıdır. Bu aşama *sözbirimleştirme* aşamasıyla doğrudan ilgilidir ve bize hiyerarşik söz varlığı değerlendirme yapma olanağı verir. Ayrıca sıklık listeleri nesnel bir dilsel gerçekliği de dil araştırmacılarına sunabilmektedir.<sup>8</sup> Bir derlemde sıklık temelli veri çıkarımının oldukça farklı boyutları vardır. Özellikle sözlükbirimin yazınsal ve anlamsal kullanımının sözlük girdilerinin oluşturulma aşamasında göz önünde bulundurulması, sözlükbirimlerin sözlükte sunumunda özellikle anlamsal kullanım sıklıklarını dikkate almak, oluşturulacak sözlüğün işlevselliğini de amaca uygun olarak arttırmaktadır. Bu tip bir sıklık uygulaması, ana dili ve yabancı dil olarak Türkçenin öğretiminde önceliklerin belirlenmesi açısından ayrı bir önem taşımaktadır.

Üçüncü aşama ise *bağımlı dizin* oluşturmaktır. Bu aşama belirli bir derlemede seçimlik sözcükbirimlerin, -bağlam içinde inceleme konusu yapılacak olan anahtar sözcüklerin (keywords-in-context) (Atkins 2003:253)-belirlenmesine ve özellikle sıklığı yüksek olanlarının birlikte kullanıldıkları yapıları bütün olarak görebilmeye karşılık gelir. Öte yandan, bağlam içi seçimlik sözcükbirimler üzerinde yürütülen 'bağımlı dizin oluşturma'nın yanında, bir diğer bağımlı dizin türü ise sözcük türü temelinde yapılan bağımlı dizinlerdir. Bu dizinlerde sözcük türlerinin birliktelikleri araştırma konusudur. Bu tip bağımlı dizinler ayrıntılı sözcük türü etiketlenmesi yapılmış derlemlerde uygulanabilmektedir.<sup>9</sup>

Diğer bir aşama ise derlemden *birliktelik kullanımları* ve *eşdizimli* yapıların çıkarımıdır. Bu yapılar çeşitli istatistiksel uygulamalarla bir derlemden tespit edilebilmekte ve genellikle dil öğretiminde önemli sözlükbirimsel materyallerin derlendiği bir uygulama olarak varlık göstermektedir. Öte yandan, eşdizimlilik kavramı bugüne kadar yapılan çalışmalarda araştırma amacına uygun olarak farklı bakış açılarıyla ele alınmıştır. Bu ele alış biçimi, kavram olarak eşdizimliliği ve onun ne olduğu da çeşitlendirmiştir. Bu çeşitlenme iki ana anlayışı ortaya çıkarmıştır. Bunlardan birincisi, belli bir dizimde belli bir aralıkta birlikte kullanımlar arasındaki sıklık dağılımının ortaya çıkarıldığı *istatistik temelli yaklaşım* (statistically oriented approach) ya da *sıklık temelli yaklaşım* (frequency-based approach); ikincisi ise, sözcük birlikteliklerinin tamamıyla değil de belli birtakım ölçüler dahilinde bir araya gelerek ortaya çıktığının düşünüldüğü *anlam temelli yaklaşım* (significance oriented approach)'dır. Yine, bazı dil birliklerinin ise dizge içerisinde gerçekleşme biçimleri eşdizim incelemelerine konu olur. Örneğin *ad+ad*, *sıfat+ad*, *belirteç+fiil* vb. (Özkan 2011).

Öte yandan, bir sözlüğün hazırlanmasında sözlük girdileri düzenlenirken sözcükbirimle ilgili sözlükte yer alması gereken ana yapılandırmalar bugün için belirlidir. Sözcükbirimle ilgili bu yapılandırma doğrudan sözcükbirimle ilintili olarak gerçekleşse de genel olarak sözlüklerde aşağıda sıralanan madde başı yapılandırmalar yer alır.

Bunlar:

**yazımsallık** (spelling): Madde başı sözlükbirimin genel kullanımındaki yazımı ve bunu takip eden farklı yazımlar.

**sesletim** (pronunciation): Farklı sesletim özellikleri.<sup>10</sup>

**çekimlenme biçimleri** (inflections): Sözcüğün farklı çekimleniş biçimleri.

**sözcük türü** (word class): Genel-uzlaşım sal sözcük türü etiketlemeleri (isim, fiil...).

**anlam/lar** (senses): Farklı sözcük ya da sözcük gruplarına ve alt gruplara ait anlam ya da anlamları kapsayan sözlükbirimlerin birden çok anlama sahip olduklarında numaralandırılan anlamlar.

**tanımlama** (definition): Her bir anlamın açıklaması.

**tanıklama** (examples): Farklı font karakterleri ile belirginleştirilmiş açıklayıcı birer söz öbeği ya da tümce anlamının açıklanması.

**kullanım** (usage): Ağız, resmi-resmi olmayan, argo, sözlü ya da yazılı dil, kullanım alanı (bilim, sanat, felsefe) gibi sözlükbirimin anlamıyla sınırlı olarak bağlamını aktaran etiketler.

**diğer türetimler** (run-ons): Deyim, kalıp ifade vb. yapılar.

**köken bilgisi** (etymology): Sözcükbirimlerin kökeni sözlük girdisi için ayrıca belirtilen yapılandırma ölçütleridir (Jackson 2002:26-27, Hanks 2003:56-57).

Bunlardan hangilerinin sözlük girdisinde yapılandırılacağı sözlükbilimcinin amacı doğrultusunda şekillenir. Bu şekillenmenin temelinde ise derlemlerden elde edilen veriler yer alır.

Sözlükbilim araştırmalarında anahtar bir sözcükbirimle ilgili olarak o sözcükbirimin kullanım örnekleri, ilgili örneklerden sezilen anlam farklılaşmaları, dizgesel birlikteliklerin anlam farklılaşmalarına etkisi ve bunların belirlenmesi ölçülenmiş uygulamalar arasında yerini almıştır (Atkins vd. 2003:254).

Yukarıda kısaca değindiğimiz dilsel *bileşenlerin tanılanması, tanılanan bileşenlerin söz varlığında gerçekleşme sıklıklarının tespiti, bağımlı dizinlerin oluşturulması ve eşdizimli yapıların çıkarımı* aşamalarının toplamı, sözlükbilim çalışmalarında sözlük girdilerinin seçiminde nitelikli ve geçerli verileri araştırmacılara sunmaktadır. Bu noktada iki ayrı yöntemle karşılaşılır. Birincisi, bizim tercih ettiğimiz daha işlevsel ve sonuçları bakımından verimli bir yöntem olan: sözvarlığında var olan sözcükbirimleri nitelikli ve temsil gücü yüksek bir derlemden denetlemektir. İkincisi ise *sözbirimleştirme* sunucunda elde edilecek olan sözcükbirimlerin ve dile yeni girmiş yapıların (neology) yine bir derlemden çıkarımıdır. İkinci yöntem birincisine nazaran daha verimli sonuçlar doğurabilecek gibi görülse de bu yöntem zaman alıcı ve maliyeti yüksek bir altyapıyı gerektirmektedir.

Türkçe için henüz Türkçenin karakteristik ve tipik kullanım ortamlarından derlenmiş, Türkçeyi temsil niteliği olan yazılı ve/ya sözlü bir derlem oluşturulamamış olsa da araştırmacılar Türkçe için derlem oluşturma çalışmalarını sürdürmektedir.<sup>11</sup> Böyle bir derlemin henüz oluşturulamamasının nedenleri bir derlemin bilgisayar tarafından okunabilir-işlenebilir bir yapıda işaretlenmiş olma zorunluluğu ve söz konusu uygulama alanında altyapı olanaklarının yetersizliğidir. İşaretlenmiş derlem uygulamaları alanyazında sıklıkla karşılaştığımız uygulamalar arasında yerini almıştır. Bu tip derlemler yukarıda bahsettiğimiz ikinci yöntem ile veri çıkarımı yapılabilecek derlemlerdir (bk. BNC: British National Corpus, <http://info.ox.ac.uk/bnc>).

Öte yandan sözlükbilim çalışmalarında bilgisayar destekli bir uygulama olarak tasarımlanan yapının işlevsel bir takım nitelikleri de karşılaması gerekir. Bu noktada tercih edilecek yöntem, sözcükbirimleri sözlükbirimselleştirirken onları etiketleri, birimsel ve anlamsal sıklıkları ile bunların dizimsel özelliklerini derleyebilecek bir yapıda olmalıdır.

### 3. Uygulama

Sözlükbilim, sözcüklerin kullanım ve anlamlarıyla ilgilenir. Geleneksel bakış açısıyla sözlükbilimsel araştırmalar sözcüklerin anlamlarını ve eşanlamlılıklarını vb. kendine konu edinmiştir. Son zamanlarda ise benzer çalışmalar derlem-tabanlı yöntemlerle sözcüklerin dağılımları, bir sözcüğün yüklendiği farklı anlamları, diğer sözcüklerle olan dizgesel bağlılıkları vb. kendine konu edinmekte, tüm bu çalışmalar ise sözlük oluşturma merkezinde gerçekleştirilmektedir. Öte yandan, dilbilgisel ve sözlükbirimsel bileşenler dizgesel yollarla gerçeklik bulmaktadır. Bu anlamda sözlükbilim belirli bir sözcüğün farklı bağlamlarda ne şekilde kullanıldığıyla yakından ilgilenmektedir (Biber vd. 2002: 21-24).

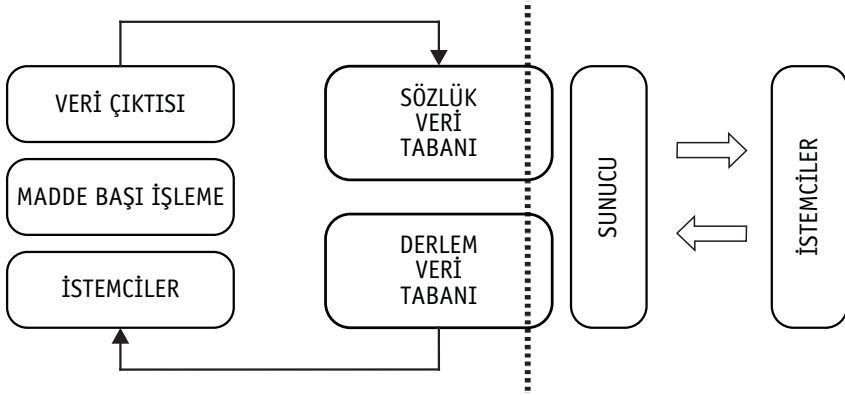
Bu çerçevede oluşturulan *TÜRKÇE DERLEM*'in sözdizimsel, sözcükbirimsel ve biçimbirimsel olarak çözümlenmesi ve "Türkiye Türkçesinin Derlem Tabanlı Sözlüğü'nün oluşturulmasındaki veri işleme aşamaları bu bölümde konu edinecektir.

TTSV'nin derlem tabanlı sözlüğünü oluşturmak amaçlı tasarımlanan derlem üzerinden yürütülecek çalışma, bütünlük ve katılımlı bir uygulamayı gerekli kılmaktadır. Bilişim teknolojilerinin sunduğu imkânları böyle bir çalışmada kullanmak işlenecek verinin yoğunluğu düşünüldüğünde bir zorunluluk olarak karşımıza çıkar.

Böyle bir amaç için bilişim teknolojilerinin paralelinde kurulacak olan bütünlük ve katılımlı sistemde veri işleme diyagramı, bir *sunucu* (server) üzerinde yer alan iki ayrı veri tabanı kullanıcıların veri işlemesi denetleme-

sinde kesintisiz olarak çalışmaktadır. Kullanıcılar *Derlem Veri Tabanı*'ndan yapacakları sorgulamaları *Veri İşleme* aşamasından geçirip *Sözlük Çıktısı* olarak kontrol ettikten sonra *Sözlük Veri Tabanı*'na kaydedilmektedir. Bu yöntemin uygulandığı ve sözcük türü olarak Türkiye Türkçesinde fiillerin ve sıfatların konu edinildiği iki ulusal araştırma projesinde, sistem sözlük girdisi olarak sözcükbirimlerin işlenmesinde aktif olarak kullanılmaktadır. Sözcükbirimler söz konusu derlemde elde edilen tümcesel sorgulamaların yordamlanması aşamasından sonra, kurulan sistem üzerinden 'sözlük veri tabanı'na kaydedilmektedir (bk. Şekil 1)<sup>12</sup>

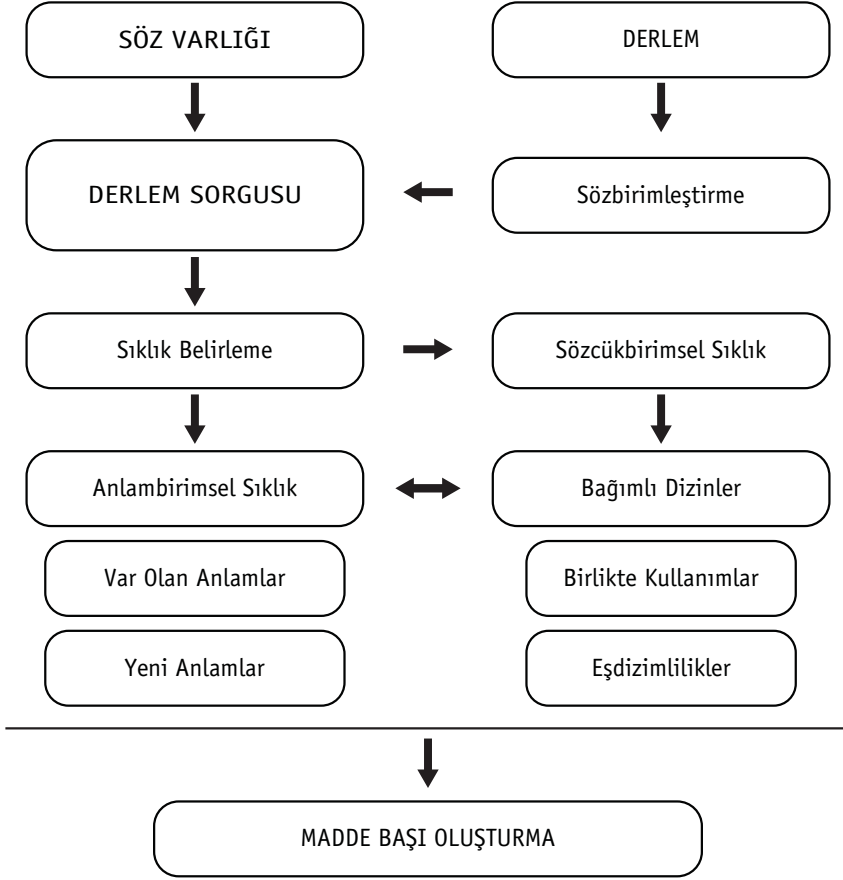
**Şekil 1:** Veri İşleme ve Derlem Sorgulama Diyagramı



Bu sözcükbirim işleme süreci, amaca uygun olarak hazırlanmış 'sözlük veri tabanı' işaretlemelerini kapsamaktadır. Söz konusu sistem üzerinden sözcükbirimler var olan yazımsallık, sözcükbirimsel ve anlambilimsel sıklıklar, anlamlar, tanıklar, birliktelik kullanımları ve eşdizimli yapılar için ayrı ayrı işaretlemelerle 'sözlük veri tabanı'na aktarılmaktadır.

Sistemde veri işleme aşamaları aşağıdaki gibidir (bk. Şekil 2):

**Şekil 2: Madde Başı Oluşturma Diyagramı**



### 3.1. Madde Başı Oluşturma Örneği

Madde başları Güncel Türkçe Sözlük'te (GTS) tanımlı madde başlarıdır. Derlem sorgusu için GTS'de yer alan madde başları ile derlemde *sözbirimleştirilmiş* yapı uyumu öncelikli olarak kontrol edilmektedir. GTS'de yer almayan sözcükbirimler ise değerlendirme dışı bırakılmaktadır (bk. Tablo 2).

**Tablo 2: “pütür” ve Türetimlerinin Sözbirimsel Görünümü**

Değerlendirmeye alınanlar		Değerlendirme dışı bırakılanlar
GTS	Türkçe Derlem	Yeni Türetimler
pütür	pütür	-
pütür pütür	pütür pütür	-
pütürlü	pütürlü	<i>pütürlülük</i>
pütürlen-	pütürlen-	-

Uygulama olarak hem GTS’de madde başı tanımlı hem de *TÜRKÇE DERLEM*’de *sözbirimleşmiş* sözlükbirimlerden olan ‘pütür’ ve ‘pütür’ün türetimleri üzerine örnek bir çözümleme yapılacaktır. Veri işleme aşamaları yukarıda Şekil 2’de de yer aldığı biçimiyle aşağıdaki gibidir.

### 3.1.1. TTSV’de (GTS) Tanımlı Sözcükbirimler: Anlam ve Tanıkları<sup>13</sup>

**Şekil 3: GTS’de Tanımlı madde başları ‘pütür’ ve türetimleri**

<b>pütür</b> <i>isim</i>
<i>isim</i> Küçük kabarcık “Ovuşturduğu ellerinden hamur pütürleri, dan taneleri dökülüyordu.” – M. N. Sepetçioğlu
<b>pütür pütür</b> <i>sıfat</i>
1. <i>sıfat</i> Üzerinde pek çok pütür bulunan
2. Sertleşip çatlamış “Pütür pütür deri.”
<b>pütürlü</b> <i>sıfat</i>
<i>sıfat</i> Pütürü olan, pürüzlü, pürtüklü “Muşambanın üstünde körpecik bir salatalık, çiçeği burnunda, pütürlü.” – N. Hikmet
<b>pütürlenmek</b> <i>nsz</i>
<i>nsz</i> Pütürlü duruma gelmek “Sırtını ve omuzlarını yoklamak kabil olsa cildinin tavuk derisi gibi pütürlendiği hissedilecektir.” – R. N. Güntekin



### 3.1.2. Derlemeden Sözcükbirimsel Sorgulama 'pütür'

Söz konusu sözcükbirimler oluşturulan derlem üzerinden sorgulama arayüzü aracılığıyla sorgulanmaktadır. Sorgulama sonucu, biçimbirimsel çözümlenmesi yapılmış sözcükbirim için aşağıdaki gibidir (*bk.* Şekil 4).

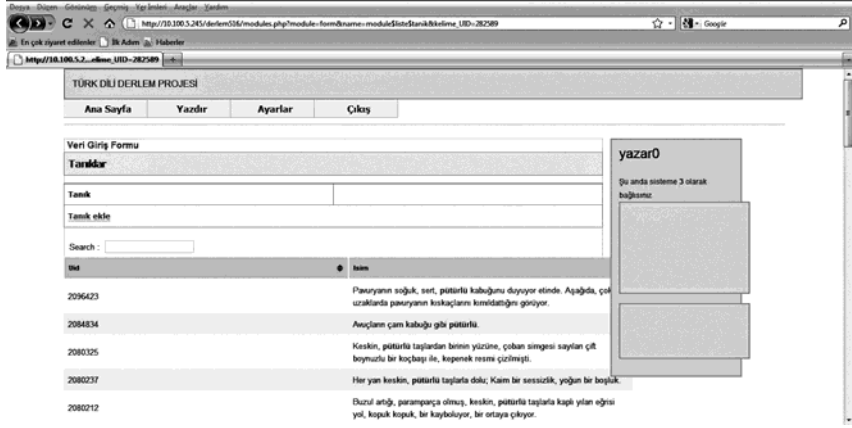
Derlem sorgusu sonucu biçimbirimsel çözümlenmesi yapılan sözcükbirimler listelenmektedir. Bu ekranda (*bk.* Şekil 4) olası yanlış çözümlenmeler kaldırılmaktadır. Ayrıca her *sözbirim* derlemde sıklık temelli olarak ayrı ayrı listelenebilmektedir.

**Şekil 4:** Sözcükbirim derlem sorgu ekranı 'pütür' ve türetilimleri

Arakam #	Çözümleme	Sil	Sıklık
pütürlerinde	pütür+Noun+A3pl+P3sg+Loc pütür+Noun+A3pl+P2sg+Loc pütür+Noun+A3sg+P3pl+Loc pütür+Noun+A3pl+P3pl+Loc	Kaldır	1
pütürler	pütür+Noun+A3pl+Pron+Nom pütür+Noun+A3sg+Pron+Nom+DB+Verb+Zero+Pres+A3pl	Kaldır	1
pütür	pütür+Noun+A3sg+Pron+Nom	Kaldır	11
pütürü	pütür+Noun+A3sg+Pron+Nom+DB+Adj+With	Kaldır	33
pütürlüğü	pütür+Noun+A3sg+Pron+Nom+DB+Adj+With+DB+Noun+Ness+A3sg+P3sg+Nom pütür+Noun+A3sg+Pron+Nom+DB+Adj+With+DB+Noun+Ness+A3sg+Pron+Acc	Kaldır	2
pütürüydü	pütür+Noun+A3sg+Pron+Nom+DB+Adj+With+DB+Verb+Zero+Past+A3sg	Kaldır	1
pütürleşen	pütür+Noun+A3sg+Pron+Nom+DB+Verb+Acquire+Pas+DB+Adj+PresPar+T	Kaldır	2
pütürdü	pütür+Noun+A3sg+Pron+Nom+DB+Verb+Zero+Past+A3sg	Kaldır	2

Sorgulama sonucunda sözcükbirimle ilgili olan uygun örnekler tümesel olarak seçilir. Yapılan seçim listesi bir metin editörüne aktarılır. Anlam/lar içinse veri girişi yapan kullanıcının var olan tanımlığı kontrol etmesi ve yeniden yazması gerekmektedir (*bk.* Şekil 5). Her bir sözcükbirim başka bir ekrana metin editörü aracılığıyla madde başı oluşturma işlemine kaynaklık etmek üzere kaydedilir.

Şekil 5: Tanık seçme ekranı ‘pütürlü’



### 3.1.3. Sözcükbirimsel-Anlambirimsel Sıklık Belirleme ve Diğer İşaretlemeler

Bu aşamada derlemden bir metin editörü aracılığıyla alımlanan veriler öncelikli olarak sözbirimleşme durumlarına göre ayrılır. Her sözbirim anlam girişleri temel alınarak kullanım sıklıkları, sözcük türü, anlam sıklıkları ve tanıklarına göre sınıflandırılır.

#### 3.1.3.1. Madde Başı Örnek Kayıt İşlemleri ve İşaretlemeler

pütür

Sözcükbirim	Sıklık	Sözcük Türü
Anlam	3	ad

Tanıklar

- *Bütün malzemeleri derin bir kaba koyup, pütür kalmayacak şekilde doğrayın.*
- *Hüseyin Feyzullah, annesi çıkınca yatağında gövdesini duvara doğru çevirerek parmağım badananın pütürlerinde gezdirdi.*
- *Beyaz, gergin derinin üstünde, iğne ucu kadar küçük, kırmızı pütürler belirmiş.*

pütür pütür

Genel Sıklık	Sözcük Türü
15	belirteç, önad, ad

1. Anlam

Sıklık	Sözcük Türü
8	belirteç

Küçük sert kabarcıklı bir biçimde.

### Tanımlar

- *Keseyi sürdükçe, kapkara, **pütür pütür** dökülüyordu üst deri tabakası.*
- *...mor menevşeler açsın, badem ağaçları **pütür pütür** olsun ha?*
- *Balığın **pütür pütür** olmaması için.*
- *Bir seferde sıcak sütü dökün ve kanşımın **pütür pütür** olmaması için iyice çırpın.*
- *Kadın kısmının belli bir yaştan sonra bel nahiyesinin altı; ki biz buna popo veya kasnak diyoruz; portakal kabuğu gibi **pütür pütür** olur...*
- *Karardı, çatladı, **pütür pütür** oldu.*
- *Sonra da mor menevşeler açsın, badem ağaçları **pütür pütür** olsun ha?*
- *Şöyle kör gözle baktın mı, ağaçlar yine kupkuru görünür, fakat can gözüyle gördün mü bir de dallara su yürüdüğünü, **pütür pütür** olduğunu anlar, sen de canlanır, cıvıldardın.*

	Sıklık	Sözcük Türü
2. Anlam	4	önad
	Küçük sert kabarcıklı, pütürlü.	

### Tanımlar

- *Dudakları taşınlarak boyanmış, kirpikleri takma, **pütür pütür** rimelle ağırlaştırılmış: bakışlarını gölgelendiren, yoksa bu mu?*
- *Fatih Peker, yaptığı yazılı açıklamada, selülitin, derinin alt tabakasında yağ dokusunun hemen çevresinde oluşuşun ve derinin üst bölümünde **pütür pütür** görüntü bırakan bir hastalık olduğunu ifade etti.*
- *Oluşuşun selülitler derinin üst bölümünde **pütür pütür** bir görüntü bırakır ve fiziksel olarak kendine güvensizliğe sebep olur.*
- *Örneğin yazın daha yağlıdır, kışın ise soğuk hava ve rüzgârlar nedeniyle kuru, **pütür pütür** bir görünüm alır.*

	Sıklık	Sözcük Türü
3. Anlam	3	ad
	Küçük sert kabarcıklı, pütür, pürüz.	

### Tanımlar

- *Elleri **pütür pütür**dü.*
- *Parmaklarının derisi sert, **pütür pütür**dü.*
- *Şu ellerime bak, **pütür pütür**.*

## pütürlü

	Genel Sıklık	Sözcük Türü
	19	önad, belirteç, ad

	Sıklık	Sözcük Türü
1. Anlam	12	önad
	Küçük sert kabarcıklı, pürüzlü bir yüzeyi olan.	

### Tanımlar

- *Buzul artığı, paramparça olmuş, keskin, **pütürlü** taşlarla kaplı yılan eğrisi yol, kopuk kopuk, bir ortaya çıkıyor.*
- *Dili sudan çıkarıp, hafif ılıyınca üstünün **pütürlü** derisini bıçakla traş edercesine soyunuz.*

- *Duyularımızın bileşimi, yüzeyi **pütürlü** ve çatlak oylumlara benziyor, ince ama derin, ta merkeze kadar inen çatlaklar...*
- *Hamuru bu süre içinde birkaç kere çıkarıp, katlar arasındaki yağ artıkları **pütürlü** bir şekilde donmasını diye, elinizle bastırıp, tekrar buzdolabına koyunuz.*
- *Her yan keskin, **pütürlü** taşlarla dolu; Kaım bir sessizlik, yoğun bir boşluk.*
- *Keskin, **pütürlü** taşlardan birinin yüzüne, çoban simgesi sayılan çift boy-nuzlu bir koçbaşı ile, kepenek resmi çizilmişti.*
- *Mızrakı, geniş, **pütürlü** bir yaprak.*
- *Pavuryanın soğuk, sert, **pütürlü** kabuğunu duyuyor etinde.*
- *Sanki falakaya yatıp sopalar indirmemişler de, o tabanı, **pütürlü** bir yüzeye sürmüşler sürmüşler, bütün çıkıntılan kabartılan giderip bir mermer yüzeyi gibi dümdüz yapmışlardı.*
- *Sonra da **pütürlü** yanaklarından damlacıklar dökülmeye başladı.*
- *Tahsin Usta, yatağında yorgun uyuyan, düşlerinde kuğuların kanatlarında uçan, yağmur kokulu oğlunun, ürpertili, yaz bulutu gibi serin, saz yaprağı gibi **pütürlü** yanağına koydu dudaklarını.*
- *Sivilce dökmüş gibi üstü **pütürlü** şekerlerden nohut şekerine, jelibonlardan vapur tutmasına iyi gelen şekerlere kadar her çeşit şeker var dükkânında.*

	Sıklık	Sözcük Türü
2. Anlam	4	Belirteç
	Yüzeyi küçük sert kabarcıklı, pürüzlü bir biçimde.	

#### Tanımlar

- *Çünkü babasının, kendisini görünce ne diyeceğini bilmiyor, sonra da yüzünün, ömrü boyunca **pütürlü** kalmasından çekiniyordu.*
- *Aksi takdirde çorba **pütürlü** olur.*
- *Elleri suyun içinde durmaktan annemin elleri gibi pembe pembe, **pütürlü pütürlü** olmuştu.*
- *Yoksa salçanız **pütürlü** olur.*

	Sıklık	Sözcük Türü
3. Anlam	3	Ad
	Yüzeyi küçük sert kabarcıklı, pürüzlü, yıpranmış.	

#### Tanımlar

- *Avuçların çam kabuğu gibi **pütürlü**.*
- *Elleri **pütürlüydü**, sertti, kocamandı.*

### pütürlen-

	Sıklık	Sözcük Türü
Anlam	2	fiil (nsz)
	Küçük kabarcıklı, pürüzlü bir duruma gelmek.	

#### Tanımlar

- *Denizin kumda **pütürlen**en, kırılıp dökülen küçük dalgacıklarına basa basa, denizin üstünde yürüyormuşçasına, kıyı boyunca ilerledim.*
- *Derindeki sızıntılı sarmıçların sularını, batık kentin gölgelerindeki çardaklı evlerinin terleyen küpleriyle sırlı testileri üstünde **pütürlen**en mavi ışıltılı su damlacıklarını, araştırmacı kökleriyle bulup emerek yıkıntılan yeşertiyorlar.*

Her bir sözbirim için oluşturulan sıklık, sözcük türü, anlam ve tanık tabloları birliktelik kullanımı ve eşdizimlilik analizinde kaynaklık etmektedir. Bu aşamada *Concordance v3.2* metin analizi yazılımı<sup>14</sup> kullanılmaktadır. Program aracılığıyla bağımlı dizinler oluşturulmakta ve birliktelik kullanımları ve eşdizimli yapılar belirlenmektedir (bk. Şekil 6).

### 3.1.4. Bağımlı Dizin Oluşturma, Birliktelik Kullanımları ve Eşdizimli Yapıları Belirleme

Şekil 6: Bağımlı dizin oluşturma aşaması

Headword	Context	W.	Context	R
parmağın	Bütün malmeleri derin bir kaba koyup,	pütür	kalmayacak şekilde doğrayın.	1
pavyarın	Keseysi sürdüğe, kapkara,	pütür	pütür doküyordu üst deri tabakası.	6
peker	...mor meneşeler açsın, badem ağaçları	pütür	doküyordu üst deri tabakası.	6
pembe	...mor meneşeler açsın, badem ağaçları	pütür	olsun ha?	7
popo	Balgın	pütür	olsun ha?	7
portakal	Balgın pütür	pütür	olmaması için.	8
pütür	Balgın pütür	pütür	olmaması için.	8
pütürü	Bir seferde sıcak sütü dokün ve karışsın	pütür	olmaması için iyice çirpin.	9
pütürülen	Bir seferde sıcak sütü dokün ve karışsın	pütür	olmaması için iyice çirpin.	9
pütürleri	Kadın kısmını belli bir yaştan sonra bel nahyesinin altı, ki biz buna ...	pütür	olar...	10
pütüründe	Kadın kısmını belli bir yaştan sonra bel nahyesinin altı, ki biz buna ...	pütür	olar...	10
pütürüdü	Karardı, çattadı,	pütür	pütür oldu.	11
resni	Karardı, çattadı, pütür	pütür	okdu.	11
rimelle	Sonra da mor meneşeler açsın, badem ağaçları	pütür	pütür olsun ha?	12
rizgârlar	Sonra da mor meneşeler açsın, badem ağaçları	pütür	olsun ha?	12
sağcaz	Şöyle kör gözde baklın mı, ağaçlar yine kapkuru görünür, fakal can g...	pütür	pütür olduğunu anlar; sen de cantanı, ovidârdın.	13
sarık	Şöyle kör gözde baklın mı, ağaçlar yine kapkuru görünür, fakal can g...	pütür	olduğunu anlar; sen de cantanı, ovidârdın.	13
sarmaçlar	Dudakları taşarak boyanmış, kirpikleri takma,	pütür	pütür rimelle ağırıştırmış; bakışların gölgelendiren, yoksa bu mu?	14
sayılan	Dudakları taşarak boyanmış, kirpikleri takma,	pütür	rimelle ağırıştırmış; bakışların gölgelendiren, yoksa bu mu?	14
saz	Fatih Peker, yaptığı yazı açıklamada, selâlin, derinin alt tabakası...	pütür	pütür görünü braker ve fiziksel olarak kendine güvenişe sebep...	15
sebep	Fatih Peker, yaptığı yazı açıklamada, selâlin, derinin alt tabakası...	pütür	pütür görünü braker ve fiziksel olarak kendine güvenişe sebep olur.	15
seferde	Fatih Peker, yaptığı yazı açıklamada, selâlin, derinin alt tabakası...	pütür	görünü braker ve fiziksel olarak kendine güvenişe sebep olur.	15
selâlin	Olğan selâtiler derinin üst bölümünde	pütür	bir görünü braker ve fiziksel olarak kendine güvenişe sebep olur.	16
selâtiler	Olğan selâtiler derinin üst bölümünde	pütür	bir görünü braker ve fiziksel olarak kendine güvenişe sebep olur.	16
sen	Örneğin yazın daha yağdır, keş işe soğuk hava ve rizgârlar neden...	pütür	pusur bir görünü alır.	17
serin	Örneğin yazın daha yağdır, keş işe soğuk hava ve rizgârlar neden...	pütür	bir görünü alır.	17
	Eleri	pütürü		18
	Darmaktanın derisi ceat	oñtür	oñtürü	19

Şekil 7: Birliktelik kullanımlarını ve eşdizimli yapıları belirleme

1 right		2 right		3 right		4 right	
No.		No.		No.		No.	
8	Pütür	1	Ağırıştırmış	1	Ağırıştırmış	1	Bakışlarını
4	Pütürü	1	Anlar	1	Anlar	1	De
1	Doküyordu	1	Doküyordu	1	Bakışlarını	1	Deri
1	Kalmayacak	1	Ha	1	Deri	1	Gölgelendiren
1	Okdu	1	Okdu	1	Doğrayın	1	Sen
1	Olduğunu	1	Olduğunu	1	Ha	1	Tabakası
1	Olmuştur	1	Olmuştur	1	Sen	1	
1	Olsun	1	Olsun	1	Üst	1	
1	Rimelle	1	Rimelle	1			
1		1	Şekilde	1			
4 left		3 left		2 left		1 left	
No.		No.		No.		No.	
1	Açsın	2	Parmaklarının	2	Derisi	8	Pütür
1	Boyanmış	1	Açsın	1	Ağaçları	2	Eleri
1	Dallara	1	Badem	1	Sert	2	Sert
1	De	1	Bir	1	Bak	1	Ağaçları
1	Derin	1	Boyanmış	1	Çattadı	1	Bak
1	Keseysi	1	Dallara	1	Elerine	1	Çattadı
1	Konjonktiva	1	Elerine	1	Kaba	1	Kapkara
1	Meneşeler	1	Karardı	1	Kapkara	1	Koyup
1	Şu	1	Keseysi	1	Karardı	1	Takma
1	Taşarak	1	Kirpikleri	1	Kirpikleri	1	Yanı
1		1	Konjonktiva	1	Su	1	Yürüdüğü

Şekil 7’de *pütür pütür olmak* eşdizimli yapısının çıkarımında bağımlı dizin analizinin yapıldığı görülür. Bu aşamada ayrıcı bir öncül olarak *sözbirimleş-tirme* işlemi de gerçekleştirilmiş olmaktadır. ‘pütür’ sözcükbiriminin tüm çekimli biçimleri listelenmekte ve bağımlı dizin içerisinde görüntülenebilmektedir. Bu anlamda ‘Türkiye Türkçesinin Derlem Tabanlı Sözlüğü’ için kurulan derlem sorgu sistemi de tüm çekimleri, sıklıklarını ve istendiğinde tanıklarını kullanıcılara sunabilmektedir (*bk.* Şekil 4 ve Şekil 5).

### 3.1.5. Madde Başlarını Oluşturma ‘pütür’ ve Türetimleri

Madde başı işleme aşamasının sonunda oluşan sözlükbirimsel yapılandırma ve açıklamaları şu şekildedir: (a) yazımsallık (*spelling*); (b) sesletim (*pronunciation*); (c) çekimlenme biçimleri (*inflections*); (d) sözcük türü (*word class*); (e) anlamlar (*senses*); (f) tanımlama (*definition*) ve tanıklama (*examples*); (g) kullanım (*usage*); (h) diğer türetimler (*run-ons*). Ayrıca madde başı oluşturma örneğinde → imi eşdizimli yapıları belirginleştirmek için kullanılmıştır. Bu yapılandırmalara sözlüğün oluşturulma amacına uygun olarak diğer örüntüler (eş yakın anlam, karşıt anlam vb.) de eklenebilir.

Madde başı işleme örneğinin sonuçları TTTSV’de tanımlı sözcükbirimlerle karşılaştırılmış ve gerekli düzenlemeler yapılmıştır. Söz konusu madde başlarının oluşturulma aşamasında veri seti olarak TÜRKÇE DERLEM kullanılmıştır.

### 3.2. Derlem Tabanlı Sözlük için Madde Başı Örnekleri: ‘pütür’ ve Türetimleri

...

**pütür**<sup>(a)</sup>: [pütür]<sup>(b)</sup> | çoğul (-ler)<sup>(c)</sup> | ad<sup>(d)</sup>

Küçük sert kabarcık.<sup>(e)</sup> “Beyaz, gergin derinin üstünde, iğne ucu kadar küçük, kırmızı **pütürler** belirmiş.”<sup>(f)</sup>

... (h)

**pütürlen-**<sup>(a)</sup>: [pütürlenmek]<sup>(b)</sup> | filimsi (-en)<sup>(c)</sup> | fiil<sup>(d)</sup>

Küçük kabarcıklı, pürüzlü bir duruma gelmek.<sup>(e)</sup> “Denizin kumda **pütürlenen**, kırılıp dökülen küçük dalgacıklarına basa basa, denizin üstünde yürüyormuşçasına, kıyı boyunca ilerledim.”<sup>(f)</sup>

... (h)

**pütür pütür**<sup>(a)</sup>: [pütür pütür]<sup>(b)</sup> | ikileme<sup>(c)</sup> | önad-ad-belirteç<sup>(d)</sup>

1. belirteç<sup>(d)</sup> Küçük sert kabarcıklı bir biçimde.<sup>(e)</sup> “Keseyi sürdükçe, kapkara, **pütür pütür** dökülüyordu üst deri tabakası.”<sup>(f)</sup>

2. *önad*<sup>(d)</sup> Küçük sert kabarcıklı, pütürlü.<sup>(e)</sup> “*Oluşan selülitler derinin üst bölümünde **pütür pütür** bir görüntü bırakır ve fiziksel olarak kendine güvensizliğe sebep olur.*”<sup>(f)</sup>

3. *ad*<sup>(d)</sup> Küçük sert kabarcıklı, pütür, pürüz.<sup>(e)</sup> “*Parmaklarının derisi sert, **pütür pütürdü.***”<sup>(f)</sup>

→ (*bir şey*) **pütür pütür olmak**

“*Karardı, çatladı, **pütür pütür oldu.***”<sup>(f)</sup>

... (h)

**pütürlü**<sup>(a)</sup> : [pütürlü]<sup>(b)</sup> | yalın<sup>(c)</sup> | *önad*-belirteç-*ad*<sup>(d)</sup>

1. *önad*<sup>(d)</sup> Küçük sert kabarcıklı, pürüzlü bir yüzeyi olan.<sup>(e)</sup>

“*Sonra da **pütürlü** yanaklarından damlacıklar dökülmeye başladı.*”<sup>(f)</sup>

2. *belirteç*<sup>(d)</sup> Yüzeyi küçük sert kabarcıklı, pürüzlü bir biçimde.<sup>(e)</sup>

“*Aksi takdirde çorba **pütürlü** olur.*”<sup>(f)</sup>

3. *ad*<sup>(d)</sup> Yüzeyi küçük sert kabarcıklı, pürüzlü, yıpranmış.<sup>(e)</sup>

“*Elleri **pütürlüydü, sertti, kocamandı.***”<sup>(f)</sup>

... (h)

#### 4. Sonuç

Bir dilin temel söz varlığını derlemek sözlükbilimin çalışmalarının ana amaçlarındanıdır. Sözlükbilim sadece ana dilin kullanıcıları için genel amaçlı bir sözlük değil belirgin amaçlar doğrultusunda hazırlanmış *özel amaçlı sözlükler* (atasözleri ve deyimler sözlüğü, terim sözlükleri, eş anlamlı sözcükler sözlüğü, zıt anlamlı sözcükler sözlüğü, eşdizim sözlüğü vb.) oluşturmakla da ilgilenir (Özkan 2008). Bu açıdan TTSV'nin yetkin bir biçimde işlenmemiş olduğu görülür.

TTSV'yi derleme çalışmaları resmi olarak 1939'da başlamıştır. Bu çalışmalar genellikle önceki sözlük çalışmalarından derlemelere dayanmaktadır. Bu anlamda ilk sözlük 1945'te yayınlanmıştır. 1945'ten bu yana çok büyük emekler verilerek TTSV'yi barındıran Türkçe Sözlük'ün 10. baskısı (2005) yapılmış, sözlük genel ağ ortamında yayınlanmıştır (GTS). Türkiye Türkçesi sözlükbilimsel anlamda tanımlı söz, terim, deyim, ek ve anlamlardan oluşan 104.481 sözlükbirimden oluşmaktadır (Türkçe Sözlük 2005).

Genel hatlarıyla bakıldığında TTSV’yi barındıran Türkçe sözlüklerin derlem dilbilimin yöntem ve uygulamaları ile çağdaş sözlükbilim çalışmalarının paralelinde olmadığı gözlemlenir. Bu çerçevede yapılması gereken asıl iş, TTSV’yi barındıran sözlüklerimizin içerik olarak çağdaş uygulamalar çerçevesinde yeniden düzenlenmesidir. TTSV’nin konu edinilen uygulamalar ışığında yeniden gözden geçilmesi sözlükbilim çalışmalarımızı bilimsel bir platforma taşıyacaktır.

Bu anlamda, TTSV’nin oluşturulacak *temsil gücü* yüksek bir *derlem*den madde başlarını belirlemek ve bu madde başlarının yazımsal özelliklerini sözlük girdisi olarak işlemek, TTSV’de bu çerçevede oluşturulacak madde başlarını anlamsal sıklık temelli olarak ortaya koymak, bu aşamada bağımlı dizinlerle sözlükbirimlerin sözlük girdisi olarak anlam ve işlevlerini tam olarak betimlemek, bunlara ek olarak TTSV’de madde başı olan sözlükbirimleri eşdizimli yapılarla zenginleştirmek, TTSV’de var olan sözlük girdilerinin tanıklarının<sup>15</sup> yetersizliklerini gerçek zamanlı dil verileriyle desteklemek gerekmektedir (Özkan ve Gündoğdu 2011).

Bugüne kadar sözlüklerde derlenen TTSV’de tanımlı yaklaşık 104 bin sözün, dilin kullanımda ortaya çıkan yeni görünümle değiştiği ve sözcükbirimler için bugüne kadar yapılmış olan tanımlarının da kullanıma dayalı bakış açısından uzak bir anlayışla ele alınmış olduğu göz önünde bulundurulmalı ve derlem dilbilimin yöntemlerinden sözlükbilimin amacına uygun olanlar kullanılarak TTSV güncellemelidir.

“Türkiye Türkçesi Söz Varlığının Derlem Tabanlı Sözlüğü”nü oluşturmanın yöntem ve uygulaması 50 milyon (+/-) sözcüklük geliştirilebilir ve güncellenebilir yapıda tasarılan bir *model derlem* (TÜRKÇE DERLEM) üzerinden örnek bir uygulamayla ortaya koyulmuştur. Örnek uygulamada da görüleceği gibi, var olan tanımlı madde başları ile derlem tabanlı uygulamanın sonuçları arasında sözcükbirimlerin anlamları, sözcük türü etiketleri, sıklıkları vb. yönlerden anlamlı derecede farklılıklar olduğu deneyimlenmiştir. Tüm söz varlığımızın bu anlamda yeniden derlenmesi Türkçenin sözlükbirimsel topografyasının oluşturulması açısından önemlidir.

Ayrıca Türkçenin ana dili ve yabancı dil olarak öğretilmesinde söz varlığının sözlükbirimselleştirilmesinde güncel ve geçerli veriler alana önemli katkılar sağlayacaktır. Türkçede yeni türetimlerin sistem üzerinden belirlenebilmesi (pütürlülük gibi.) devam çalışmalarında Türkçede yeni *türetimlerin* (neology) sözlükbirimleşmesi açısından da önemlidir.

Bu çalışmada tanıtmaya çalıştığımız sözcük türlerinden sıfatlar ve filleri konu alan iki ulusal araştırma projesinin yürütüldüğü sistem ve yöntemin



işlevselliği deneyimlenmiştir. Elbette ki derlem dilbilimin yöntemleri oldukça farklı uygulamalarla dil nesnesini kendine konu edinebilmektedir. Uygulanan yöntem, devam çalışmaları için bir model olma niteliği taşımaktadır.

## Teşekkür

*TÜRKÇE DERLEM*'in biçimbirimsel çözümlemesi için Prof. Dr. Kemal Oflazer'e teşekkürlerimi sunarım.

## Açıklamalar

- <sup>1</sup> Sözlükbilimin kısa bir tarihi için bk. Bingöl 2005.
- <sup>2</sup> Günümüz sözlük bilim ve bilgisayar uygulamaları için bk. Tahiroğlu 2006.
- <sup>3</sup> *TÜRKÇE DERLEM*'in oluşturulma aşamaları TD-1 (Türkçe Derlem-1) ve TD-2 (Türkçe Derlem-2) ile benzerdir. Oluşturulma aşamaları ile ilgili alanyazın ve işlem basamakları her üç derlem için standart niteliklere sahiptir. Bu anlamda bu bölüm diğer yayınlarda yer alan alanyazınla zorunlu olarak ortaklaşmaktadır.
- <sup>4</sup> Kitap, süreli yayın, internet vb. metin parçaları uzman görüşleri alınarak 5846 numaralı Fikir ve Sanat Eserleri Kanunu'nun Madde 35. (1-4 Fıkralar). Madde 36. Madde 37.'de belirtilen hükümler çerçevesinde sayısallaştırılmıştır. ([www.mevzuat.adalet.gov.tr/html/957.html](http://www.mevzuat.adalet.gov.tr/html/957.html)).
- <sup>5</sup> İnternet metinlerini alınmasında iki farklı yöntem kullanılmıştır. Pdf formatındaki kaynaklar OCR yazılımlarıyla, internet metinleri ise *Teleport Pro 1.50* ([www.tenmax.com/teleport/pro/home.htm](http://www.tenmax.com/teleport/pro/home.htm)) yazılımı aracılığıyla sayısallaştırılmıştır.
- <sup>6</sup> bk. [www.hlst.sabanciuniv.edu/TL/](http://www.hlst.sabanciuniv.edu/TL/) 19.05.2010
- <sup>7</sup> Biçimbirimsel çözümleme aracı ve işaretleyicileri için bk. [www.hlst.sabanciuniv.edu/TL/](http://www.hlst.sabanciuniv.edu/TL/) ve Say 2003.
- <sup>8</sup> *Sıklık* (frequency) üzerine 1.006.306 sözcüklük bir metin arşivi üzerinden yürütülen Türkçe için ilk çalışma için bk. Öz 2003.
- <sup>9</sup> Ayrıca bk. Kennedy 1998.
- <sup>10</sup> bk. International Phonetic Alphabet [en.wikipedia.org/wiki/International\\_Phonetic\\_Alphabet](http://en.wikipedia.org/wiki/International_Phonetic_Alphabet)
- <sup>11</sup> Türkçe üzerine yürütülen önemli derlem çalışmaları için bk. Özkan 2010.
- <sup>12</sup> 109K104 numaralı “*Türkiye Türkçesi Söz Varlığında Sıfatların Eşdizimliliği -Derlem Tabanlı Bir Uygulama-*” ve 109K516 numaralı “*Türkiye Türkçesi Sözvarlığında Fiillerin Derlem Denetimi ve Derlem Tabanlı Sözlüğü*” TÜBİTAK-SOBAG Ulusal Araştırma Projeleri. [turkcederlem.mersin.edu.tr/](http://turkcederlem.mersin.edu.tr/)

<sup>13</sup> Güncel Türkçe Sözlük'ten (GTS) alınmıştır [www.tdk.gov.tr/](http://www.tdk.gov.tr/) (11.05.2013).

<sup>14</sup> [www.concordancesoftware.co.uk/](http://www.concordancesoftware.co.uk/)

<sup>15</sup> TTSV'yi barındıran yoğun diskte (CD) tüm madde başlarına karşılık gelen tanık sayısının 22.707 olduğu görülür. Bu anlamda sözlükbirim başına düşen tanık sayısı ortalama 0.2'dir.

## Kaynaklar

Aksan, Doğan (2000). *Her Yönüyle Dil Ana Çizgileriyle Dilbilim*. III. Cilt. Ankara: TDK Yay.

Atkins B. T. S. and Beth Levin (1995). "Building on a Corpus: A Linguistic and Lexicographical Look at some Near-Synonyms". *International Journal of Lexicography* 8 (2): 85-114.

Atkins, Sue and Jeremy Clear (1992). "Corpus Design Criteria". *Literray and Linguistics Computing* 7 (1): 1-16.

Atkins, Sue (2003). "Lexicographic Relevance: Selecting Information from Corpus Evidence". *International Journal of Lexicography* 16 (3): 251-280.

Biber, D. vd. (2002). *Corpus Linguistics, Investigating Language Structure and Use*. UK: Cambridge University Press.

Bingöl, Zekeriya (2005). "Sözlük ve Sözlükçülük Üzerine Bir Araştırma". *Akademik Bakış Dergisi* 9 (Temmuz): 197-206.

Günay, Doğan (2007). *Sözcükbilime Giriş*. İstanbul: Multilingual.

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

[http://en.wikipedia.org/wiki/International\\_Phonetic\\_Alphabet](http://en.wikipedia.org/wiki/International_Phonetic_Alphabet) (International Phonetic Alphabet), Erişim tarihi: 23.10.2010.

<http://turkcederlem.mersin.edu.tr/derlem516/>, Erişim tarihi: 10.11.2011.

<http://www.mevzuat.adalet.gov.tr/html/957.html>, Erişim tarihi: 16.6.2010

<http://www.hlst.sabanciuniv.edu/TL/>, Erişim tarihi: 19.10.2010.

<http://info.ox.ac.uk/bnc>, Erişim tarihi: 23.10.2010.

<http://www.tnc.org.tr/>, Erişim tarihi: 16.4.2011

<http://www.concordancesoftware.co.uk/>, Erişim tarihi: 4.10.2011.

- <http://www.tdk.gov.tr>, Erişim tarihi: 10.11.2011.
- <http://www.tenmax.com/teleport/pro/home.htm>, Erişim tarihi: 10.11.2011.
- Jackson, Howard (2002). *Lexicography: An Introduction*. USA: Routledge.
- Kennedy, Graeme (1998). *An Introduction to Corpus Linguistics*. New York: Addison Wesley Longman Limited.
- Kocaman, Ahmet (1998). “Dilbilim, Sözlük, Sözlükçülük”. *Kebikeç* 3 (6): 111-113.
- McEnery, Tony and Andrew Wilson (2004). *Corpus Linguistics -An Introduction-*. Edinburg University Press.
- McEnery, Tony vd. (2006). *Corpus-Based Language Studies An Advanced Resource Book*. New York: Routledge.
- Oxford Collocation Dictionary for Students of English* (2003). Oxford University Press.
- Öz, İlyas (2003). *Yazılı Türkçenin Kelime Sıklığı Sözlüğü*. Ankara: TDK Yay.
- Özkan, B. ve A. E. Gündoğdu (2011). “Ana Dili ve Yabancı Dil Olarak Türkçenin Öğretiminde Belirteçler -Derlem Tabanlı Bir Uygulama-“. *Theoretical and Applied Researches on Turkish Language Teaching*. Ed. L. Uzun, Ü. Bozkurt). Almanya: Die Blaue Eule. Essen. 317-336.
- Özkan, Bülent (2011). *Türkiye Türkçesinde Belirteçlerle Fiillerin Birlikte Kullanılması ve Eş Dizimlilikleri-Derlem Tabanlı Bir Uygulama*-. Ankara: TDK Yay.
- \_\_\_\_\_, (2010). “Türkçenin Öğretiminde Sözlüğü: Yöntem ve Uygulama”. *E-Uluslararası Eğitim Araştırmaları Dergisi* 1 (2): 51-65.
- \_\_\_\_\_, (2008). “Güncel ‘Zarf Olarak Tanımlı Sözlük Birimlerin Derlem-Denetimi Üzerine Bir Değerlendirme”. VI. Uluslararası Türk Dili Kurultayı. Ankara. 20-26 Kasım.
- Sampson, Geoffrey and Diana McCarthy (2005). *Corpus Linguistics: Reading in a Widening Discipline*. New York: Continuum.
- Say, Bilge (2003). *Türkçe için Biçimbirimsel ve Sözdizimsel Olarak İşaretlenmiş Ağaç Yapılı Bir Derlem Oluşturma*. Proje Raporu. EK-B: Türkçe Ağaç Yapılı Derlem Kılavuzu (Proje No: 199E026). Ankara.
- Say, Bilge vd. (2002). “Bilgisayar Ortamında Derlem Geliştirme Çalışması”. Akademik Bilişim Konferansı. Konya.

Summers, D. (1993). "Longman/Lancaster English Language Corpus - Criteria and Design". *International Journal of Lexicography* 6 (3): 181-208.

Tahiroğlu, B. Tahir (2006). *Bilgisayar Destekli Dil Bilimi Çalıştayı Bildirileri*. Ankara: TDK Yay. 89-98.

*Türkçe Sözlük* (2005). Ankara: TDK Yay.

Uzun, N. Engin (2006). *Biçimbilim -Temel Kavramlar-*. İstanbul: Papatya Yay.

# Corpus Based Dictionary of Turkey Turkish's Lexicon: Method and Application

Bülent Özkan\*

## Abstract

In line with new information technologies, there have recently been a number of innovations in the methods and applications of lexicography. In the past lexicography focused on language as a research object and employed the methods of tagging, classification, and re-writing. Today, on the other hand, lexicography explores its research object through linguistic collections called “corpus”. When we review the current literature, we can see that dictionaries such as *Oxford, Collins, Longman, Cambridge* etc. have all been compiled through corpus-based research. However, the dictionaries of the Turkish Lexicon as Spoken in Turkey (TLST) do not follow the methods of corpus linguistics and contemporary lexicographic studies. This study aims to present the methodology of compiling a “Corpus-based Dictionary of the Turkish Lexicon as Spoken in Turkey”. In this context, a *prototype corpus*, which includes 50 million words (+/-), has been built according to the methods and practices of corpus linguistics. Through this corpus, we have provided instruction on the method and practice of generating a “Corpus-based Dictionary of the Turkish Lexicon as Spoken in Turkey”. This study first discusses the history of lexicography briefly and then introduces the methodology and the *prototype corpus*. Last but not least, it discusses the results of the method reached through an example application. The present study will make important contributions to the field by compiling the topography of Turkish vocabulary and by providing valid and up-to-date lexical data in the teaching of Turkish both as a mother tongue and a foreign language.

## Keywords

Turkish as spoken in Turkey, vocabulary, corpus linguistics, lexicology, Turkish teaching

---

\* Assist. Prof. Dr., Mersin University, Education Faculty, Turkish Language Teaching Department – Mersin / Turkey  
ozkanbulent@mersin.edu.tr

# Словарь лексикона турецкого языка на основе корпуса: методы и применение

Бюлент Озкан\*

## Аннотация

Вместе с развитием информационных технологий был введен ряд нововведений в методы и применение лексикографии. Если в прошлом лексикография исследовала язык как объект такими методами, как пометки, классификация, переписывание и т.п., то на сегодняшний день объектом изучения лексикографии являются языковые единицы, которые называются корпус (corpus). При обзоре литературы данной области можно увидеть, что словари, аналогичные Оксфордскому, Коллинза, Лонгмана, Кембриджскому стали результатом применения основы корпуса (corpus-based). С другой стороны, можно заметить, что словари, содержащие в себе лексикон турецкого языка, не развивались параллельно работам корпусной лингвистики и современной лексикографии. Эта работа призвана ознакомить с методом составления словаря лексикона турецкого языка на основе корпуса. В этом контексте, на основе принципов и методов корпусной лингвистики была создана модель корпуса, включающая 50 миллионов (+/-) слов. Посредством этого корпуса продемонстрировано возможное создание словаря лексикона турецкого языка на основе корпуса и дано положение используемого метода. В работе первоначально кратко показано развитие лексикографии, затем представлен метод и модель корпуса, и в заключение показаны результаты применения данного метода на конкретном примере. Работа вносит важный вклад в формирование лексикографической топографии турецкого языка и в процесс лексикографирования актуальных и действительных данных лексикона при преподавании турецкого языка как родного и как иностранного языка.

## Ключевые Слова

Турецкий язык, лексикон, корпус, лингвистика, лексикография, обучение турецкому языку

---

\* Док., университет Мерсин педагогический факультет кафедры обучения турецкому языку – Мерсин / Турция  
ozkanbulent@mersin.edu.tr