

Hüseyin ÖZTÜRK

Çağdaş Türk Romanında Bir Sözvarlığı İncelemesi

- *Yazının bilimsel içeriğinden yazarı sorumludur.*
- *Bu çalışma referans gösterilmeden kullanılamaz. Referans için aşağıdaki adresi kullanabilirsiniz:*

Öztürk, H. (2013). Çağdaş Türk Romanında Bir Sözvarlığı İncelemesi. *E-Dil Dergisi*, Sayı 1, 157-173.

www.e-dil.net

bilgi@e-dil.net

Çağdaş Türk Romanında Bir Sözcük İncelemesi¹

A Vocabulary Analysis on Contemporary Turkish Novel

Özet

1980 ile 2010 yılları arasında yayımlanmış olan Türk romanlarının sözcük açısından incelendiği bu çalışmada, öncelikle Türkçenin yabancı dil olarak öğretimi konusunda eksikliği hissedilen okuma becerisi ve sözcük öğretimi konusunda kaynak olarak yararlanılacak romanlar seçilmiştir. Sonrasında öğrenci gereksinimlerini inceleyen çalışmalar taranarak söz konusu yıllarda basılmış; çok satan, ödül alan, eleştirmenlerce önemli görülmüş romanlardan her yıl için birer örnek seçilerek yaklaşık 2,5 milyon sözcüklük bir bütüncü oluşturulmuştur. Bu bütünceden elde edilen sonuçlar, ilk aşamada her roman için tür ve türce sayılarıdır; sonrasında Türk romanının ortalama kaç sözcükle yazıldığı ve tür/türce oranının yabancı dilde okuma yaparken bize nasıl bilgiler verdiği.

Çalışmamızda, ayrıca bu bütüncü üzerinden sıklık incelemesi yapılmış ve bütüncede yer alan en sık sözcükler, başsözcükler (gövde) ve kökler sıralanmıştır. Son olarak yabancı dil olarak Türkçe derslerinde öğretilmesi gereken sözcüklerin belirlenirken bu sıklık listelerinden ne ölçüde yararlanılması ve öğrencinin Türkçe roman okumak için kaç sözcük bilmesi gerektiği gibi konular tartışılmıştır.

Anahtar Sözcükler : Türk romanı, bütüncü/derlem, sıklık, tür/türce oranı, ortak sözcükçe

Abstract

In this study on vocabulary analysis of Turkish novels published between 1980 and 2010, primarily Turkish novels were selected in order to provide for the need in the areas of reading skills and vocabulary teaching in Turkish as a foreign language classes. After that, studies that investigate learner's needs by using a corpus were reviewed and a corpus, consisting 30 best-selling, award winning novels, was built. First a quantitative research was conducted and the numbers of types and tokens were defined. Also type/token ratio and its usage while reading in a foreign language was discussed.

¹ Bu çalışma, Dokuz Eylül Üniversitesi Eğitim Bilimleri Enstitüsü Yabancı Dil olarak Türkçe Öğretimi Anabilim Dalında yürütülmüş olan "1980'den 2010 Yılına Okunurluğu Yüksek Türk Romanlarının Sözcük Açısından İncelenmesi" adlı yüksek lisans tezinden yapılmış bir çalışmadır.

Moreover a frequency study was carried out on the corpus and the most frequent words, lemmas and word stems of Turkish were arrayed. Also the novels were separated into the groups of printing dates; four distinctive groups of three decades and the corpus as a whole were analysed in terms of common lexicon.

Finally upon the arrangement of the words that are to be taught in a Turkish as a foreign language classroom, the use of these frequency lists and which words to master prior to reading Turkish novels for foreigners were discussed in the last part.

Key words : Turkish novel, corpus, word frequency, word count, type/token ratio, common lexicon

Giriş

Türkçe, gerek konuşulduğu coğrafya gerekse konuşan sayısı bakımından dünya dilleri arasında önemli bir yere sahiptir. Her geçen gün önemi daha da artan Türkçenin, yabancı bir dil olarak algılanması ve öğretilmesi ise henüz yeni bir bilim alanıdır. Bu konu günümüzde önem kazanmaya devam etse de alanda yapılan ve yürütülen çalışmalar henüz maalesef istenilen düzeyde değildir.

Bu çalışmamızda, sözcükbilimsel çalışmalardaki eksikliğin giderilmesi adına 1980 ile 2010 yılları arasında yayımlanan; okunurluklarıyla, satış rakamlarıyla, aldıkları ödüllerle benzerleri arasından ayırt edilen önemli Türk romanları sözcük varlığı açısından incelenecektir. Çalışmamızın sonucu farklı alanlarda kullanılabileceği gibi yabancı dil olarak Türkçe öğretiminde de etkin bir biçimde kullanılabilecektir.

Türk romanını sözcük varlığı açısından inceleyen bu çalışma; sözcük sıklığı, geleneksel dilbilgisinin bakış açısıyla köklerin ve gövdelerin sıklığı, ortak sözcükçe gibi içerisinde barındırdığı birçok veriyi yorumlayacak, yabancı dil olarak Türkçe derslerinde özellikle okuma becerileri gibi derslerde hangi sözcüklerin hedef sözcük varlığına alınacağı konusunda aydınlatıcı olacaktır. Bununla beraber çalışmada, Türk romanı dönemsel özelliklerine göre, hangi dönemde hangi romanları okumak için ne kadar sözcük varlığına sahip olmak gerektiği sonuçlarına ulaşılabilecektir.

Sinclair'e göre bütüncü çalışmalarının bilgisayar destekli ilk örnekleri önce Amerika ve İngiltere'de altmışlı yıllarda başlamıştır. 1961 yılında Brown Üniversitesi'nde, Nelson Francis ve Henry Kuçera Amerikan İngilizcesinin yazılı dildeki bir milyon sözcükle ilk bütüncesini oluşturmuştur. Bu sırada Edinburgh Üniversitesi İngiliz İngilizcesinin sözlü bütüncesini oluşturmaya başlamıştır. Sözlü bütüncü oluşturmak daha yoğun bir emek gerektirdiğinden bu bütüncü 300 bin sözcükle sınırlı kalmıştır (Sinclair, 2001:VIII).

Ülkemizdeki bir milyon sözcüklük ilk derlem, Göz tarafından 1997 yılında hazırlanmaya başlamıştır. İki yıl boyunca hazırlanan bu derlem, sözcük sıklığı incelemesi alanında ilk olmasından kaynaklanan bazı problemlerden dolayı yayınlanmamış ve yeniden hazırlanarak 2003 yılında bir sözlük olarak yayımlanmıştır (Göz, 2003:I).

Bütünce üzerine yapılan çalışmalar McEnery ve Hardie'ye (2012:6) göre *bütünce temelli* (ing. corpus-based) ve *bütünce odaklı* (ing. corpus-driven) olmak üzere ikiye ayrılmaktadır. Bütünce temelli çalışmalar bir savın doğruluğunu ya da bir araştırma sorusunun yanıtını bütünce verilerini kullanarak arar. Bütünce odaklı çalışmalar ise bütünce dilbiliminin bütünce temelli çalışmalar önermesini reddedip bütüncenin kendisinin dil hakkında varsayımların kaynağı olduğunu niteler.

Sözlü dilde hazırlanmış olan bütünce üzerinden sıklık çalışmalarına, Ufuk Aşık'ın 2007 yılında hazırladığı *Sözlü Türkçenin Temel Sözcük Varlığı* çalışması örnek olarak verilebilir. Bu çalışmada Aşık, 3000 dakikalık bir ses havuzundan sözlü dilde kullanılan sözcüklerin sıklıklarını vermektedir (Aşık, 2007:6). Bir başka örnek de yine sözlü dilde sıklık hesaplanması konusunda, Türkiye'nin 7 farklı coğrafi bölgesinden 14 farklı şehirde 20 farklı bağlamda konuşmacıların doğal ortamlarındaki sözcüklerindeki biçimbirim sıklıklarını çalışan Yıldız(2010:13) olabilir.

Belirli bir bütünceyi toplayarak, bunların üzerinde sıklık, etken ve edilgen sözvarlığı, öğrenici gereksinimleri vb. çalışmaları yapmak Nation'a göre temelde bazı sözcüklerin diğerlerinden daha önemli olduğu varsayımıyla yola çıkar (Nation, 2001:32). Örneğin anlamaya yönelik okunan bir metinde anlamı bilinmeyen sözcüğün sıklığı ne kadar yüksekse, okuma sürecinde bu tür sözcüklere o kadar çok rastlanılacak ve okuduğunu anlamayı o derecede güçleştirecek demektir. Sözcük sıklığı, yabancı bir dilde işlenecek bir metnin basitleştirilmesi gerekliliğini, öğretilmeye başlamadan önce hangi sözcüklerin öğretilmesi gerektiğini ya da öğretilmemesi seçeneklerini öğreticinin ve izlenç hazırlayanların tercihine sunar.

Çalışmada geçen türce (örnekçe, *ing.token*) kavramıyla ilgili Yeşim Aksan ve arkadaşları "Her iki tarafında boşluk olan ya da bir noktalama imi ile sınırlanmış bulunan dilsel birim" (2011: 397) tanımını yapmıştır. Yani türce bir metindeki sözcüklerin her biridir. Tür (çeşit, *ing.type*) için ise derlemi oluşturan ve birbirinden farklı olan her bir sözcük (Aksan ve diğer., 2011: 397) tanımını yapmıştır. Başka bir deyişle tür, birbirinden farklı sözcüklerin her biridir.

Tür / türce oranına geldiğimizde ise bu terim tür sayısının türce sayısına bölünmesi ile elde edilir. Elde edilen rakam 100 ile çarpılarak oran yüzdelik kesirle gösterilir. Hirsh ve

Nation'a göre gerçek anlamda okuduğundan keyif alabilmek (*Ing. reading for pleasure*) okunan her 50 sözcükte en fazla 1 anlamı bilinmeyen sözcükle karşılaşılması şartıyla olanaklıdır (Nation, 2001;42). Dolayısıyla bir metinde ne kadar anlamı bilinmeyen sözcük varsa okuduğunu anlama oranı da o kadar düşecektir. O halde okuyucunun bir bağlamda yer alan sözcükleri bilmemesi ve tanımaması olasılığı konusunda bu metni okumadan önce metnin tür/türce oranını ele alarak çıkarım yapmanın imkanı var mıdır? Bu konuda bir çalışma yapan Youmans; temel İngilizce metinleri ile Shakespeare'in sonelerini tür/türce oranı olarak karşılaştıran bir çalışma yapmıştır. Bu çalışmada Shakespeare'in *Macbeth* isimli oyununun temel İngilizceye basitleştirilmiş halinden 2000 türce bir kesit ile yine Shakespeare'in sonelerinden 2000 türce bir kesit alınmış ve incelenmiştir. İlk metin için 2000 sözcük içerisinde 460 tür bulunmuştur. İkinci metin için ise 780 tür bulunmuştur (Youmans, 1990:586). Aşağıdaki tabloda bu iki metin için hesaplanmış tür/türce oranlarını görebiliriz:

Tablo 1 : Örnek Shakespeare eserlerinde tür/türce oranları

Metin açıklaması	Türce sayısı	Tür Sayısı	Tür/Türce Oranı
Macbeth (T.Takata tarafından basitleştirilmiş)	2000	460	23%
Shakespeare Soneleri	2000	780	39%

Çalışmasındaki oranların farklı çıkması karşısında kontrol amaçlı başka bir çalışma daha yapan Youmans ilk yapıttaki oranın düşük çıkması karşısında bu oranın eşitlenebilmesi için *Macbeth* eserindeki türce sayısını 4800'e çıkarmış ve sonuç olarak tür sayısının 697'ye yükseldiğini görmüştür. Bu durumda dahi tür/türce oranı artmamış aksine %14,5'e düşmüştür. Sonuç olarak Youmans "Açıkça görünen –ve kesin- çıkarım şudur ki Shakespeare'in sözvarlığı temel İngilizceninkinden çok daha geniştir (1990:588)" demiştir. İncelemeye konu olan bir metin ya da romandaki bu geniş sözvarlığı, tür/türce oranı arttıkça okuyucunun karşısında sürekli olarak daha önce metinde karşılaşmadığı yeni sözcükleri göreceği anlamına gelmektedir. Başka bir deyişle okuduğunu anlamak için gerekli olan sözvarlığı ihtiyacı tür/türce oranının yükselmesiyle artmaktadır. Bu oran düştüğünde ise okuyucu daha sınırlı bir sözvarlığı ile metindeki okuduğunu anlama oranını artırabilir.

Yöntem

Çalışmada öncelikle bütüncüyi oluşturmak için *Radikal*, *Cumhuriyet*, *Hürriyet*, *Milliyet* ve *Zaman* gibi çeşitli günlük gazetelerin güncel ve Ahmet Pıřtına İzmir Kent Arşivi Müzesinin arşivindeki örnekleri ile bunların düzenli olarak çıkardıkları kitap ekleri taranmıştır. Bunun yanısıra, çeşitli yayınevi ve topluluklarca çıkarılmış olan ‘*Yapı Kredi kitap-luk*’, ‘*Hece Türk Romanı Özel Sayısı*’ vb. yazın dergileri incelenmiştir. Türkiye’de 1986’dan günümüze kadar yayın hayatını sürdürmüş ve edebiyat dünyasıyla ilgili verileri derleyen ‘Türkiye kültür ve sanat yıllığı’, yine romanlar ve diğer edebiyat yapıtlarının eleştirileri ile basım yayım bilgilerini toplayan ‘Türkiye istatistik yıllığı’, 1984 yılına kadar çıkarılmış olan ve romanlar için ayrı bir bölümü olan ‘Varlık yıllığı’, son olarak da bir kurul ile yılın edebi yapıtlarını değerlendiren ‘Nesin vakfı edebiyat yıllığı’ gibi nicel ve nitel verileri bir arada toplayan yayınlardan yararlanılmıştır.

Bunlarla beraber Fethi Naci’nin ‘*Yüzyılın Yüz Türk Romanı*’ adlı kitabı, Alemdar Yalçın’ın ‘*Çağdaş Türk Romanı*’ incelemesi, TDK Yayınlarının ‘*Güzel Yazılar: Romanlar*’ derlemesi çalışmalarımıza ışık tutan basılı kaynaklar olmuştur. O yıl basılan, çeşitli kurumlarca yılın romanı seçilen yapıtlar incelenip bu çalışmamızın sonunda iki aşamalı seçim sistemimizin ön bir listesi elde edilmiştir. Son olarak yayınevlerince yılın en çok satan romanları ile ‘Orhan Kemal Roman Ödülü²’, ‘Yunus Nadi Ödülü’, ‘Sait Faik Armağanı³’ ve ‘Cevdet Kudret Edebiyat Ödülleri⁴’ gibi edebiyat alanında çeşitli kurum, vakıf ve kuruluşlarca o sene ödül alan romanlar çalışmamızda var olan romanların yanlarında açıklama olarak eklenmiş ve 363 romandan oluşan son bir karar şeması elde edilmiştir.

Bu sentezin sonunda 1980 yılından 2010 yılına kadar yayınlanmış olan Türk Romanlarından, sözcük seçiminde kişilerin tercihleri ve sözvarlıklarının etkisinde kalmamak için aynı yazardan sadece bir kitap almak koşuluyla ve yayınlanmış romanların olabildiğince her yıl için bir tane olmasına özen gösterilerek 30 yıl için seçilmiş 30 adet Türk romanı belirlenmiştir. Bu seçim sırasında nesnel kriterler ön planda tutulmaya çalışılmıştır. Örneğin 1982 yılı için seçilen Orhan Pamuk’un ‘Cevdet Bey ve Oğulları’ adlı romanı hem Fethi Naci’nin 100 yılın 100 romanı derlemesinde yer bulmuş, hem de Orhan Kemal Roman ödülüyle onurlandırılmıştır. Aynı şekilde ‘Babil’de Ölüm İstanbul’da Aşk’ satış rakamlarıyla; ‘Aşk’ aday gösterildiği uluslararası ve aldığı ulusal ödülleriyle, ‘Mutluluk’ Barnes&Noble

² < <http://orhankemal.org/odul.htm> > (S.e.t. 6 Şubat 2012)

³ < <http://www.darussafaka.org/tr-TR/darussafaka/Darussafaka-cemiyeti/Guncel/Sayfalar/Raporlar> > (S.e.t 3 Mart 2012)

⁴ < <http://www.cevdetkudretodulleri.com/> > (S.e.t. 3 Mart 2012)

yeni büyük yazar keşif ödülüyle, ‘Bu İşte Bir Yalnızlık Var’ Radikal Kitap ekinde editörlerin yılın kitabı seçmesiyle bu listemize girmek açısından ön plana çıkan romanlardan bazılarıdır.

Belirlenen bu kitaplar öncelikle taranarak sayısal ortama aktarılmıştır. Sonra *Optik Karakter Tanıma* programlarıyla işlenen veriler sözcüklerin incelenmesine hazır hale getirilerek üzerinde *Metin Analizcisi* yazılımıyla yaklaşık 2,5 milyon sözcüklük bütüncü üzerinde tür ve türce incelemesi yapılmıştır. Türkçede biçimbirimler eklendiği sözcüklere göre eşyazımlı olabildiği için (bkz. Tablo 2) halihazırda bu ekleri sorunsuz olarak tanıyabilecek bir yazılım yoktur. Bu sözcüklerin çekim eklerinden ayrıştırılmış başsözcükleri (ing. lemma) ancak metinler taranarak el ile yapılabilmektedir. Nitekim 50 milyon sözcükle oluşturulmuş Türkçe Ulusal Derlemi için söz konusu işlem öğretim elemanları ile yüksek lisans ve lisans öğrencilerinden oluşan bir grup tarafından el ile yapılmıştır (Aksan ve Mersinli, 2011:31).

Tablo 2 : Eklerin eşyazımlılığını gösteren tablo

<i>Kök/Gövde</i>	<i>İşlevi</i>
süs Ø	Kök
süsleme	Yapım Eki (Eylem → Ad)
süsleme	Çekim Eki (Olumsuzluk)
süslemeli	Çekim Eki (Gereklilik Kipliği)
süslemeli	Yapım Eki (Ad → Önad)

Dolayısıyla çalışmamızın ikinci aşaması olan başsözcük ve kök sayımı için bütüncenin tamamını tek kişi el ile tarayamayacağı için ikinci bir yazılıma gereksinim duyulmuştur. Bu adımda Fatih Üniversitesi Bilgisayar Mühendisliği Bölümü Doğal Dil İşleme Grubu’na hazırlanmış ve üniversitenin genel ağ sitesinde⁵ kullanıma sunulmuş olan ‘*Türkçe Metin Sıklık Çözümleyicisi*’ yazılımı kullanılmıştır. Hazırlanan yaklaşık 1 milyon sözcüklük bütüncenin altderleminde bu program, tarama ve sayım işlevini gerçekleştirmiş; tanıma ve bir araya getirme işlemleri el ile yapılmıştır.

⁵ < <http://nlp.ceng.fatih.edu.tr/metinanaliz/> > (S.e.t 31 Mayıs 2013)

Bulgular ve Sonular

Türce Sayıları

1980 ile 2010 yılları arasında yayınlanmış 30 Türk romanını ele aldığımız çalışmamızda araştırma sorularımızı yanıtlamak için 2.418.372 sözcükle oluşturduğumuz bütüncemizde elde ettiğimiz ilk veri, romanların kaç sözcükle yani kaç türce ile yazıldığı olmuştur.

TDK Türkçe Sözlük'e göre roman: “İnsanın ya da çevrenin karakterlerini, göreneklerini inceleyen, serüvenlerini anlatan, duygu ve tutkularını çözümleyen, kurmaca veya gerçek olaylara dayanan uzun edebi tür”dür. Görüldüğü üzere bu tanımda bir romanın ortalama kaç sözcükle yazıldığına dair bir bilgi yoktur. Aynı şekilde Ali Püsküllüoğlu'nun Türkçe Sözlüğünde, Mehmet Doğan'ın Büyük Türkçe Sözlüğünde; ve Türkçe dışında İngilizce *Oxford Advanced Learner's* ve Fransızca *Le Petit Robert* sözlüğünde bir romanın ortalama olarak kaç sözcükle yazıldığı konusunda bir rakam verilmemektedir.

Konuyla ilgili olarak taranan çalışmalarda, romanların kaç sözcük kullanılarak yazıldığı konusunda; Türkçe için Bayram Baş'ın 2006'da yaptığı çalışmada çocuk romanlarının ortalama 9700 sözcükle yazıldığını belirtmiştir (2006:162). Başka bir örnekte İngilizce için Migdale'in (2012:5) çalışmasında yüksek lisans derecesi için hazırlanmış olduğu eseri 40.000 sözcüğün altında olduğu için kısa roman (*ing. novella*) olarak değerlendirmektedir. Buna kaynak olarak da McBein'in 2007'de yaptığı tanıma işaret ederek; Kısa romanların, 5000 sözcükle yazılan hikayelerden, ve en az 60.000 sözcükle yazılan romanlardan farklı olarak 10.000 ile 40.000 arasında sözcük içerdiğini söylemiştir (*Aktaran Migdale, 2012:5*). Bu tanımda ise 40.000 sözcük ile 60.000 sözcük arasındaki yayımların tanımı açık bırakılmıştır.

Romanların kaç sözcükle yazıldığı konusunda yazarların yorumlarını ve edebiyat dergilerini taradığımızda ise Zülfü Livaneli'nin “Aslında zor bir konu bu. ... konu, dönem ve tanıtılan karakterlerin çokluğu bir romanda sayfa sayısını belirliyor” (Livaneli, 2012) dediğini görüyoruz. Romanların sözcük sayısı konusunda en kesin rakamları bir Amerikan edebiyat dergisinde Chuck Sambuchino şöyle veriyor: Çocuk romanları: 20-45 bin, gençlik romanları için 55-70 bin, yetişkin romanları için 70-100 bin ve son olarak da günümüzde okur kitlesi bir

hayli artan bilim kurgu ya da fantastik konulu romanlarda ise 100-115 bin sözcük olmalıdır (Sambuchino, 2012).

Tablo 3'e türce sayısı açısından baktığımızda, bütüncemizde en düşük türce sayısına sahip roman 27.419 sözcükle yazılmış Metin Kaçan'ın *Ağır Roman* adlı yapıtı; en yüksek türce sayısına sahip roman ise 167.344 sözcük ile yazılmış Yaşar Kemal'in *İnce Memed* adlı yapıtıdır. 30 romanla oluşturduğumuz bütüncemizde bir roman ortalama 80.612 sözcüğe sahiptir. Buradan da yola çıkarak diyebiliriz ki 1980-2010 yılları arası Türk romanı ortalama 80 bin sözcükle yazılmıştır. Bütüncemizde yer alan romanlardan 13 tanesi ortalama türce sayısının üzerinde sözcük içermekte 17 tanesi ise bu ortalamanın altına düşmektedir.

Tür Sayıları

Çalışmamızın ikinci aşaması, hazırlanan romanların tek tek *Metin Analizcisi* yazılımında incelenmesidir. Bunun sonucu olarak elde ettiğimiz bir diğer veri ise romanlardaki tür sayılarıdır. Buna göre romanlarda kaç farklı sözcük kullanıldığı bilgisini veren tür sayısı Tablo 3'deki gibidir. En düşük tür sayısı ile yazılmış yapıt Nazlı Eray'ın *Aşkî Giyinen Adam* adlı romanıdır. En yüksek tür sayısı ile yazılmış olan yapıt ise 39,209 tür içeren Alev Alatlî'nin *Schrödinger'in Kedisi*'dir. Toplam 543,051 sayısı romanların sahip olduğu tür sayılarının toplamıdır. Bu rakama göre sözü geçen yıllar arasında yazılmış bir romandaki ortalama tür sayısı 18.102'dir. Buradan yola çıkarak 1980 ile 2010 yılları arasında yazılmış olan Türk Romanı ortalama 18 bin tür (farklı sözcük) içermektedir denilebilir. Fakat yaklaşık 2,5 milyon sözcüğe sahip bütüncemizdeki tür sayısı 189155'tir. Bu fark her türün yalnızca bir kez sayılmasından kaynaklanmaktadır. Bütüncede yer alan 30 romanın 16 tanesi bu ortalamanın altında tür sayısına sahiptir 14 roman ise ortalamanın üzerinde tür sayısı taşımaktadır.

Tür/Türce Oranları

Tür ve türce verileri birbiri ile oranlandığında daha önceki bölümlerde üzerinde durduğumuz gibi bize tür/türce oranını verir. Bir roman için tür/türce oranı, tür sayısının türce sayısına bölünmesi ile çıkan sonucun yüzdeler olarak gösterilebilmesi için 100 ile çarpılmasından oluşur. Aşağıda bütüncemizi oluşturan romanların tür/türce oranına göre düşükten yükseğe sıralanmış listesi verilmiştir.

Tablo 3 : Tür/Türce Oranları

Yazarı	Adı	Tür Sayısı	Türce Say.	TTO (%)
Orhan Pamuk	Cevdet Bey Ve Oğulları	20257	161389	12,55
Yaşar Kemal	İnce Memed 4	22341	167344	13,35
İpek Ongun	Şimdi Düğün Zamanı	13760	102972	13,36
Nazlı Eray	Aşkı Giyinen Adam	5723	39937	14,33
Osman Aysu	İktidar Merdiveni	11352	62366	18,20
Turgut Özakman	Şu Çılgın Türkler	30101	145991	20,62
Ayşe Kulin	Adı: Aylın	18098	85185	21,25
Hıfzı Topuz	Hatice Sultan	12416	57425	21,62
Burak Turna	Metal Fırtına 2	9366	42896	21,83
Pınar Kür	Bir Cinayet Romanı	17117	77328	22,14
Canan Tan	Piraye	15380	69147	22,24
Latife Tekin	Sevgili Arsız Ölüm	13702	59500	23,03
Ahmet Ümit	Sis Ve Gece	13883	60092	23,10
Oğuz Atay	Tutunamayanlar	38252	162822	23,49
Vedat Türkali	Mavi Karanlık	24019	101830	23,59
Zülfü Livaneli	Mutluluk	20256	84450	23,99
Buket Uzuner	Kumral Ada Mavi Tuna	26561	107659	24,67
Celil Öker	Kramponlu Ceset	11220	44402	25,27
İnci Aral	Mor	20260	78979	25,65
Hakan Yel	Sultana Dokunmak	19469	74314	26,20
Hasan Ali Toptaş	Gölgesizler	11304	42811	26,40
Adalet Ağaoğlu	Bir Düğün Gecesi	22313	84352	26,45
Elif Şafak	Aşk	24769	92790	26,69
İskender Pala	Babil'de Ölüm İstanbul'da Aşk	24854	90719	27,40
Ahmet Altan	Sudaki İz	13055	46397	28,14
Alev Alatlı	Schrödinger'in Kedisini: Kabus	39209	138710	28,27
Tuna Kiremitçi	Bu İşte Bir Yalnızlık Var	10093	35698	28,27
Ahmet Karcılılar	Yağmur Hüznü	10908	34239	31,86
Erdal Öz	Gülün Solduğu Akşam	13272	39209	33,85

Metin Kaçan	Ağır Roman	9741	27419	35,53
Toplam		543051	2418372	22,46

Tablo 3 incelendiğinde ilk dikkatimizi çeken bir önceki bölümde en düşük türce sayısına sahip *Ağır Roman* adlı yapıtın tür adedine göre oranlandığında en yüksek tür/türce oranına sahip olduğunu görüyoruz. Yine aynı şekilde en yüksek türce adedine sahip *İnce Memed 4* adlı roman ise tür/türce oranında en düşük ikincidir. Buradan çıkartılabilecek yargı –oluşabilecek genel kanının aksine- çok sözcüğe sahip, sayfa sayısı yüksek ve uzun bir romanda, çok daha kısa yazılan bir romana göre bilinmeyen sözcüklerle karşılaşma olasılığı daha yüksek değildir.

Önceki bölümlerde bir romanın ortalama 18102 tür ve 80612 türce ile yazıldığını belirtmiştik. Bu rakamlara göre tür/türce oranı için de bir ortalama çıkarttığımızda; bir roman için tür/türce oranı %22,46 olmalıdır. Bütüncemizde yer alan toplam türlerin sayılarının toplam türcelere oranlaması yukarıda verdiğimiz ortalama ile aynı doğrultuda %22,46 olmuştur. Buna göre bütüncemizdeki 11 roman bu orandan daha düşük, 19 roman ise bu orandan daha yüksektir. Tablomuzda en sık tekrarlanan değer, %23 olmuştur. Bu değere göre, tür/türce oranı >%23 olan romanların, okuyucuyu anlamı bilinmeyen ve yeni sözcüklerle karşılaşma olasılığı açısından zorlayacağını; bu oranın altında kalan romanların ise okuyucu için görece daha kolay olacağını söyleyebiliriz.

Bütüncenin Tür Sıklığı

Bütüncemizde yer alan 30 yıl içerisinde yazılmış 30 Türk romanı toplamda 2,418,372 türce ile yazılmıştır. Bu nicelik 189155 türün farklı sıklıklarda geçmesiyle oluşmuştur. Bütüncemizde yer alan yaklaşık 2,5 milyon sözcüğün en sık olanları yoğunlukla önad (bir, bu, iki, vb.), bağlaç (ve, ama vb.), ilgeç (gibi, için, diye, vb.) ve adıllardan (o, ben, vb.) oluşmaktadır. Bunlardan farklı olarak ‘de-‘ eyleminin geçmiş zaman çekimi 10109 defa tekrar etmiştir. En sık 2000 sözcüğün toplamda 981,082 defa tekrar ettiğini görüyoruz. Bu nicelik bütüncemizde yer alan 2,418,372 sözcüğün %41’ine denk gelmektedir. En sık sözcük olan ‘bir’in bütüncedeki türcelere oranı binde 25, yine en sık 2000’inci sözcük olan ‘kısmı’ sözcüğünün ise bütüncedeki türcelere oranı yüzbinde 5’dir.

Sıklığı en yüksek 2000 tür gruplarına göre değerlendirildiğinde; en sık ilk 500 tür toplamda 451,387 kez tekrarlanmıştır, bu rakamın bütüncedeki türce sayısına oranı

%18.67'dir (Bkz. Şekil 1). İkinci 500 tür ise toplamda 210,745 defa tekrarlanmıştır; yine bu rakamın türce sayısına oranı %8.71'dir. Sıklık listesinde 1001-1500 arasında yer alan türler 169,624 defa tekrarlanmış; bunların da türce sayısına oranı %7.01'dir. Son 500 tür 169,326 defa tekrarlanmış ve türce sayısına oranları da % 6.99'da kalmıştır. Görüldüğü üzere türlerin sıklıkları ilk 500 sözcük için bütüncedeki toplam sözcük sayısının neredeyse %20'sine denk gelmekteyken sonraki 500'er tür içeren gruplar %9, %7 ve %7 oranları ile azalmaktadır.

Sıklığı en yüksek 2000 sözcük arasında yukarıda bahsettiğimiz bağlaç, önad, ilgeç gibi türlerin yanında diğer sözcük türlerinden adları ve eylemleri de göz önüne aldığımızda; akrabalık ilişkileri bildiren adların dikkat çektiğini söyleyebiliriz. *Abi* sözcüğü 161 kez, *abla* 187, *teyze* 130, *amca* 113 kez tekrar etmiştir. Bunların haricinde *annem* 641 kez, *ana* 432 kez, *annesi* 350, *anne* 298, *annesinin* 287, *annemin* 194, *annemle* 145, *anneme* 119, *annesine* 114 kez; *babası* 463, *babam* 429, *babasının* 376, *baba* 264 *babamın* 232, *babasına* sözcüğünün ise 136 kez tekrar ettiğini görüyoruz. Özellikle *anne* sözcüğünün 9 ve *baba* sözcüğünün 6 farklı biçimbirim taşıyan halleri çalışmamızda bir başsözcük (ing. *lemma*) belirleme yapılmasını gerekli hale getirmiştir.

Akrabalık bildiren sözcüklerin haricinde adlar arasında *el* sözcüğü en yüksek sıklıkta 1054 kez *elini* haliyle ve en düşük sıklıkta 112 kez *elleriyle* şeklinde toplamda 20 farklı şekilde bütüncemizin en sık sözcükleri arasında yer almaktadır. Eylemlerde ise en sık 2000 sözcükte 59 farklı çekimlenmiş hali bulunan 'ol-' eyleminin farklı sıklıklarda görüyoruz. Bunların en sık şekli, *olduğunu* 3206, *olan* 2550 ve en az sıklıkta *olabileceğini* 130 kez geçmektedir. 'Ol-' eyleminin bütüncenin tamamında 1097 kez farklı çekimli halinin farklı sıklıklarda geçmesi bütüncemizdeki sözcüklerin geleneksel dilbilgisinin kök olarak adlandırdığı biçimbirimlerden tamamen arındırılmış hallerinin de çalışmamızda belirlenmesi gerektiğini göstermiştir.

Bütüncenin Başsözcük Sıklığı

Bütüncenin tamamında yer alan sözcükler birbirinden farklı olan türler olarak ele alındığından dünya adına eklenmiş her ekle oluşturulan sözcük ayrı bir tür kabul edilmektedir, *dünya*, *dünyayı*, *dünyaya*, *dünyada* vb... Bu halde bütüncemizde yer alan dünya sözcüğü 65 ayrı tür olarak karşımıza çıkmıştır. Bu 65 türden bir önceki bölümde ele aldığımız en sık 2000 sözcük arasında sadece 5 tanesini görebiliyoruz; *dünya*, *dünyayı*, *dünyada*, *dünyaya* ve *dünyanın*. Bu türlerin sıklığı önemli bir veri sunsa da; yabancı dil öğreniminde ve okuma

becerisi öğretiminde Türkçenin hangi sözcüklerinin öğretileneğinin belirlenmesi için özellikle bu eklerden arındırılmış halleri sözcüklerin gerçek sıklıklarını verecektir.

Bu doğrultuda hazırladığımız listemizde, çekim eklerinden ayrılmış en sık 1000 başsözcüğün toplamda 594,401 defa tekrarlandığını görüyoruz. Bu nicelik yaklaşık 1 milyon sözcüklük alt derlemimizde toplam türlerin %66'sına denk gelmektedir. Bu grubu ikiye ayıracak olursak en sık 500 başsözcüğün sıklık toplamının 510,983 olduğunu bunun da alt derlemdeki sözcüklerin %57'sine denk geldiğini söyleyebiliriz.

Bütüncemizde türlerin ilk 500 tanesinin toplam sıklığının 451,387 olduğunu bir önceki bölümde vermiştik. Bu rakamla aynı sözcükleri içermese de başsözcük listesinde en sık 500 sözcük sıklığı arasındaki fark, sadece rakam olarak ele alınamaz. Bütünce ve alt derlemin içerdiği toplam sözcük sayısının farklı olmasından dolayı bu artışın daha iyi anlaşılabilmesi için oranlarıyla birlikte değerlendirilmelidir. İlk 500 tür bütüncenin %19'una denk gelmekteyken ilk 500 başsözcük altderlemin %57'sine denk gelmektedir. Bu artış sözcüklerin çekim eklerinden ayrılarak tek bir madde başı haline gelmesiyle olanaklı hale gelmiştir.

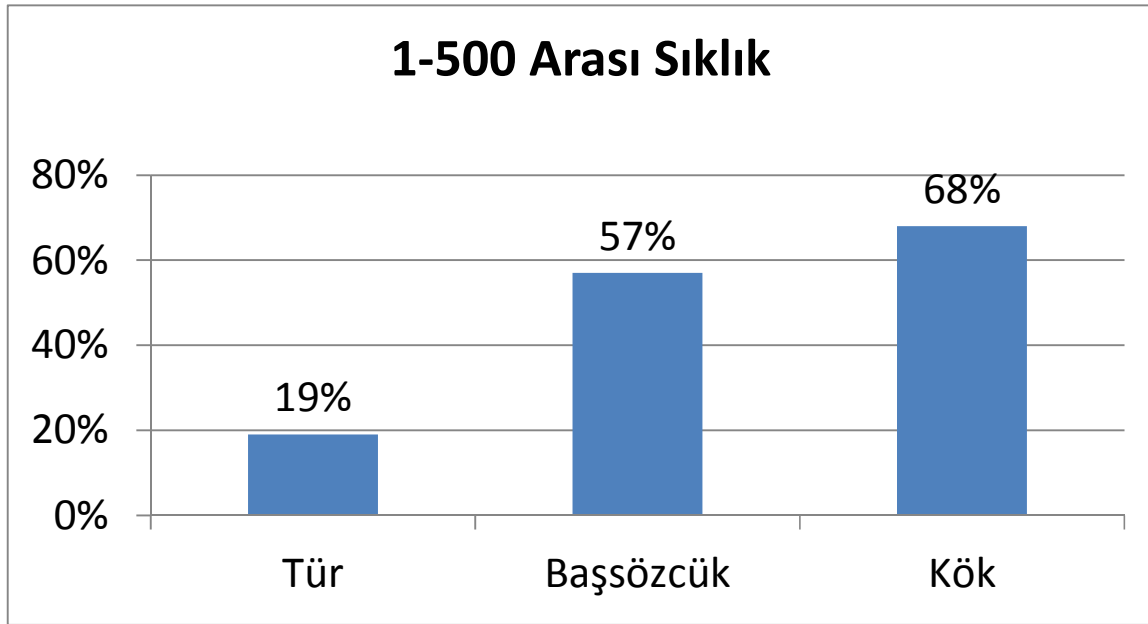
Bütüncenin Kök Sıklığı

Banguoğlu'na göre "Sözlükteki sözcüklerden bir takımı bölünmez anlam unsurlarıdır. Çoğu tek heceli olan bu ses topluluklarını daha basit parçalara bölmek mümkün olmaz. Bunlara kök denir" (2007:144). Yine Gencan "yapım eki almamış sözcükler köktür" demiştir (2007:93). Türkçenin sondan eklemeli yapısı sınırlı sayıdaki biçimbirim ile çok sayıda sözcük türetilmesini olanaklı kıldığından başsözcük listesi haricinde bir de geleneksel dilbilgisinin kök olarak nitelendirdiği, sözcüğün hiçbir ek almamış hallerini sıralamamızı gerekli kılmıştır.

Alt derlemimizde yer alan sözcük köklerinin incelendiği bu aşamada sıklığı en yüksek 500 kökün toplam sıklığının 615,938 olduğunu görüyoruz. Bu sayı 901,098 sözcük içeren alt derlemdeki sözcüklerin %68,35'ine denk gelmektedir. Açıkçası bu rakamın bu kadar yüksek olması Türkçenin ekleri ne denli etkin kullandığının da bir göstergesidir. Buradan yola çıkarak, köklerin üzerindeki eklerin işlevlerine ve anlamlarına hakim bir dil öğrencisi pek çok sözcük öğrenmeden de okuduklarını anlayabilir denilebilir. En sık kök, onbinde 314 sıklık oranıyla 'bir', en sık eylem kökü onbinde 213 sıklık oranıyla 'ol-', en sık ad kökü ise onbinde 57 sıklık oranı ile 'baş'dır. En sık 500 kök arasında sıklık oranı en düşük kök ise onbinde 2 ile 'yön'dür.

Banguoğluna göre adeylemler, sıfateylemler ve bağeylemler çekimli fiil halindedir (2007:420). Gencan'a göre ise eylemsiler hem ad, önad, bağlaç olabilir hem de eylem görevindedir (2007:431). İlk listemizde *ol, olsun, olmaz* vb... 'ol-' kökü ile başlayan 57 madde başı vardır; ikinci listemizde bu sayı, *ol, olacak, olan, olarak* olmak üzere beşe inmiştir. Yukarıdaki tanımların ışığında başsözcükler sadece çekim eklerinden arındırılmış ve adeylem, ortaç, ulaç yapan ekler yapım eki kabul edilmiştir. Bu nedenden ötürü sözcüklerin bu eklerinden de ayırıp köklerini bulmak bir zorunluluk haline gelmiştir. Bu sayede başsözcük listesinde 5 farklı girdiyle toplamda 18,334 defa tekrarlayan 'ol-' kök listesinde 19,199 defa tekrarlamıştır. Aradaki fark başsözcük listesinde ilk 1000'de *olmayan, olmuş* ya da *olmaz* gibi sözcüklerin köklere dahil edilerek sayılmasıdır.

Şekil 1 : Sıklığı en yüksek sözcüklerin bütüncüye oranı



Sıklık listelerimizde yer alan ilk 500 veriyi tür –sözcük-, başsözcük –gövde- ve kök sıklıkları açısından incelediğimizde yukarıdaki şekilde yer alan veri bize Türkçe'nin biçimbirimlerden ne kadar çok yararlandığını anlatıyor. İlk 500 sözcüğün sıklığı bütüncedeki toplam sözcüklerin %19'u kadar iken; en sık 500 kökün altderlemdeki toplam sözcüklere oranı %68'e denk geliyor. Bu veriler özellikle sözlük hazırlanırken ve Türkçe okuma yaparken sözcüklerin hangi eklerle hangi bağlamlarla kullanılabileceğinin araştırmasının yapılması gerekliliğini ortaya koyması açısından önem taşımaktadır.

Tartışma ve Öneriler

1980 ile 2010 yılları arasında yayımlanmış 30 Türk romanını incelediğimiz çalışmamızda ilk bölümde çalışmamızın amacını, araştırma sorusunu ve alt problemleri belirledikten sonra bu konuda daha önce yapılmış olan çalışmalara değinilmiştir. Sıklık, XX. yüzyılda önem kazanan ve bilgisayar desteğiyle etkin bir şekilde incelenebilen bir konu olmuştur. Sıklık çalışmalarının, ülkemizde ise özellikle son 10 yıllık süreçte çokça çalışılan bir konu olduğunu gördük. Bütünce temelli sıklık çalışmaları yapılırken, çalışmaların genellikle farklı yazın türlerindeki yapıtlardan alınan kesitler üzerinde yapıldığı görülmüştür. Bu konuda istisna olabilecek birkaç çalışmada ise bir yapıtın tamamı alınsa da belirli bir metin türüne odaklanılmamıştır. Dolayısıyla son 30 yılda yazılmış romanları ele alan çalışmamız –bilgimiz dahilinde- belirli bir yazı alanında metinlerin tamamını inceleyen tek çalışmadır.

Çalışmamızın sonunda ilk olarak elimize geçen veri, son 30 yılda yazılmış bir Türk romanının ortalama olarak kaç sözcük ile yazıldığıdır. Bu sorunun yanıtı ortalama 80 bin sözcüktür ki; bu rakam da yaklaşık 400 sayfaya denk gelmektedir. Daha sonra bir romanda kaç tane tür yani farklı sözcük olduğunu bulmak istediğimizde yine bu zaman aralığında yazılmış bir Türk romanında ortalama olarak 18 bin tür olduğu sonucunu elde ettik. Bu da Türkçe romanlarda bir sözcüğün ortalama olarak 5 kez tekrar ettiği anlamına gelmektedir.

Sonuç olarak bu çalışmada yer alan verilerle bir okuma sözlüğü hazırlanmasını olanaklı hatta gerekli görüyoruz. Bütünce temelli olarak hazırlanabilecek olan bu sözlük, sözlük yazarının tercihlerine göre değil bütüncedeki sözcüklere göre oluşturulduğu için ülkemizde alanında ilk olacaktır. Yabancı dil olarak Türkçe öğrenenlerin bu konudaki eksiklerine çözüm olabilecek bu sözlük aynı zamanda Avrupa Ortak Dil Çerçevesi seviyelerine göre hangi sözcüklerin hangi seviyeye ait olduğunu öğrencilere renk kodları ile de gösterebilir. Bunun yanında yabancı dil olarak Türkçe öğrenenler için okuma becerisi derslerinde hangi sözcüklerin öncelikle öğretilmesi, hangi romanların okunacağını seçme konusunda da çalışmamızın yararı olacağına inanıyoruz. İkincil bir ürün olarak bu bütüncede

yer alan sözcüklerle yabancı dil olarak Türkçe öğrenenler için başlangıç seviyesinde okuma kitapları hazırlanabilir ya da bazı romanlar belirli sözlükbaşı girdilerle sınırlı kalmak koşuluyla basitleştirilebilir. Böyle bir çalışma bu alanda yararlı olacak ve ciddi bir eksikliği giderecektir.

Bu çalışmamız "1980'den 2010 Yılına Okunurluğu Yüksek Türk Romanlarının Sözcük Varlığı Açısından İncelenmesi" adlı tezinden yapılmış bir çalışmadır. Konu ile ilgili daha ayrıntılı sayısal veriler, istatistikler ve listeler çalışmamızda alan sınırlılığında yer bulamamıştır. Söz konusu verilere YÖK ağ sayfasında yer alan tez çalışmasından ulaşılabilir.

Bunun yanında herhangi bir roman için en sık kaç kök veya gövdenin bilinmesi Türkçe bir metnin anlaşılması için gerekli olduğunun belirlenmesi başka bir araştırmanın konusudur. Yapılacak bu çalışmada, sözcük öğretimi yapıldıktan sonra sözcüklerin öğrenilip öğrenilmediği çeşitli ölçme araçları ile kontrol edilmeli ve sonrasında bu sözcükleri içeren metinler okuyuculara verildikten sonra görüşme, çoktan seçmeli ve açık uçlu sorular gibi çeşitli yöntemlerle okuduğunu anlamının yüzde kaç oranında gerçekleştiği belirlenmelidir.

Kaynakça

Aksan, M.; Mersinli, Ü. (2011) A Corpus Based Nooj Module For Turkish *Proceeding Of The Nooj 2010 International Conference And Workshop 29-39* Komotini: Democritus Uni.

Aksan, Y.; Mersinli, Ü.; Yaldır, Y. (2011). İlköğretim Türkçe Ders Kitapları Derlemi ve Türkçe Ulusal Dil Derlemi örneklemindeki sözcük sıklıkları *Türkçe Öğretimi Üzerine Çalışmalar* İzmir: Dokuz Eylül Üniversitesi Yayınları.

Aşık, U. (2007). *Yabancılar İçin Temel Türkçe Sözlük Varlığının Oluşturulması*.

Yayımlanmamış Yüksek Lisans Tezi. D.E.U. Eğitim Bilimleri Enstitüsü :İzmir

Banguoğlu, T. (2007) *Türkçe'nin Grameri* Ankara: TDK Yay.

Baş, B. (2006). *1985-2005 Yılları Arasında Çocuk Edebiyatı Sahasında Yazılmış Tahkiyeli Metinlerin Söz Varlığı Üzerine Bir Araştırma* Yayımlanmamış Doktora Tezi. Ankara: Gazi Üniversitesi, Eğitim Bilimleri Enstitüsü.

- Gencan, T. N. (2007) *Dilbilgisi* Ankara: Tek Ağaç Eylül Yay.
- Göz, İ. (2003) *Yazılı Türkçenin Kelime Sıklığı Sözlüğü*, Ankara: TDK
- Livaneli, Z. (2012, Temmuz 1) Romanlar kaç sayfa olmalıdır? *Vatan Gazetesi*
- McEnery, T. ; Hardie, A. (2012) *Corpus Linguistics: Method, Theory, Practice*. Cambridge : CUP
- Migdale, A.M.S (2012) *Beyond the snapshot: the elegant tracing of an emotional line in the writing of Anna: Three Fictions* Yayımlanmamış Yüksek Lisans Tezi. Johannesburg: Witwatersrand Üniversitesi
- Nation, P. (2001) Using Small Corpora to Investigate Learner Needs In Ghadessy et al. (Eds.) *Small Corpus Studies and ELT* Amsterdam: John Benjamins
- Sambuchino, C. (2012, Ekim) Word Count for Novels and Children's Books. *Writer's Digest*
- Sinclair, J.M. (2001) The Incredible Shrinking Corpora. In: Ghadessy et al. (Eds.) *Small Corpus Studies and ELT* Amsterdam: John Benjamins Yay.
- Yıldız, İ. (2010) *Sözlü Dildeki Bağımlı Ardılların Kullanım Sıklığı* Yayımlanmamış Yüksek Lisans Tezi. Dokuz Eylül Üniversitesi Eğitim Bilimleri Enstitüsü : İzmir
- Youmans, G. (1990) Measuring lexical style and competence : The Type-Token Vocabulary Curve. *Style*, Sayı 24. 584-599