



Türkçe Eğitimi Açısından Ses-Harf Sıklığı Üzerine Bir Değerlendirme*

Özay KARADAĞ**

Öz

Bir dil üzerine yapılan sıklık çalışmaları o dilin eğitimi açısından önemli bilgileri ortaya çıkartmaktadır. Yaygın olarak yapılan sıklık çalışmaları, bir dilde en sık kullanılan kelimeleri belirlemeye yöneliktir. Bu çalışmaların çıkış mantığı, hemen her dilde bazı kelimelerin diğerlerine oranla çok daha sık kullanılmasıdır. Dildeki sık kullanılan kelimelerin belirlenerek eğitim öğretim sürecinin planlanmasında bu kelimelerin öğretimine öncelik verilmesi dilin işlevsel olarak öğretimini kolaylaştırmaktadır. Türkçede sıklık üzerine yapılan çalışmaların çok büyük bölümü kelime, sınırlı olarak da deyim ve atasözü sıklıklarını belirlemeye yöneliktir, ses-harf sıklığı üzerine yapılmış kapsamlı bir çalışma bulunmamaktadır. Bu çalışmada Türkçenin ses-harf sıklığı iki temel bakış açısıyla belirlenmiştir. İlk olarak Türkçe Sözlük'te yer alan madde başları ses-harf sıklığı açısından analiz edilmiştir. İkinci olarak derlem çalışmaları için oluşturulan bir havuzdan örnek alınarak dilin genel yazılı kullanımındaki ses-harf sıklığı belirlenmiştir. Elde edilen ses-harf sıklığına ilişkin veriler karşılaştırılarak değerlendirilmiştir. Araştırmanın tartışma ve sonuç bölümünde ses-harf sıklığına dair elde edilen verilerle ilk okuma yazma öğretiminde kullanılan ses kümeleri değerlendirilmiş, ilk okuma yazma ses kümelerinin Türkçenin ses özellikleriyle ne oranda örtüştüğü tartışılmıştır. Son olarak ses-harf sıklığına ilişkin verilerin yabancılara Türkçe öğretiminde nasıl kullanılabileceği değerlendirilmiştir.

Anahtar Kelimeler: ses sıklığı, harf sıklığı, derlem, Türkçe öğretimi, ilk okuma yazma öğretimi

An Evaluation of Phoneme-Letter Frequency Regarding to Turkish Education

Abstract

Frequency studies on a language reveal important information for the education of that language. Common frequency studies are aimed at identifying the most commonly used words in a language. The mainspring of these studies is that in almost every language, some words are used more frequently than others. Determining the frequently used words in the language and giving priority to teaching these words in the planning of the educational process facilitates the functional teaching of the language. The majority of the studies on frequency in Turkish are aimed at determining the frequency of words and limited idioms and proverbs. There is no comprehensive study on phoneme-letter frequency. In this study, the phonetic frequency of Turkish was determined from two basic perspectives. Firstly, the headings in the Turkish Dictionary were analyzed in terms of phoneme-letter frequency. Secondly, the phoneme-letter frequency in general written use of the language was determined by taking a sample from a pool created for corpus studies. The obtained data related to the phoneme-letter frequency were evaluated by comparing. In the discussion and conclusion part of the study, the data obtained from the phoneme-letter frequency were evaluated and the clusters used in the first literacy teaching were evaluated, and the extent to which the first literacy phoneme sets overlap with the Turkish

* Bu çalışma XI. Uluslararası Büyük Türk Dili Kurultayı'nda (25-28 Eylül 2016 Budapeşte-Macaristan) aynı isimle sunulan sözlü bildirin gözden geçirilmiş ve genişletilmiş hâlidir.

** Doç. Dr, Hacettepe Üniversitesi Eğitim Fakültesi Türkçe ve Sosyal Bilimler Eğitimi Bölümü, Ankara, ozaykaradag@hacettepe.edu.tr, ORCID: 0000-0003-4596-1203

phonetic features was discussed. Finally, it was evaluated how the data related to the phoneme-letter frequency can be used in teaching Turkish to foreigners.

Keywords: phoneme frequency, letter frequency, corpus, Turkish teaching, primary literacy teaching

Giriş

Dil öğretimi alanında yapılan araştırmaların bir bölümünü sıklık çalışmaları oluşturmaktadır. Bilgisayar teknolojisinin gelişmesi ve yaygınlaşması ile birlikte sıklık üzerine yapılan çalışmalar artmakta ve farklı amaçlarla gerçekleştirilen sıklık çalışmaları, dilbilim alanında olduğu kadar dil öğretimi alanına da önemli veriler sunmaktadır.

Türkçe Sözlük'te sıklık (1998: 1966) "*sıkça geçen, kullanımı sık olma*" anlamına gelmektedir. Korkmaz (1992: 134) sıklık terimini "*bir konuşmada, bir yazıda veya bir yazarın eserinde bazı kelimelerin öteki kelimelere oranla daha çok ya da daha az kullanılması durumu*" olarak açıklamaktadır. Vardar (1998: 183) sıklık kavramını daha genel bir çerçeve ile "*Belli bir uzunlukta bir konuşma ya da yazıda aynı dilsel olgu ya da birimin gerçekleşme sayısı.*" şeklinde tanımlamış ve yabancı dil öğretiminde kullanılabilirlik ile birlikte, bu kavramdan öğretilecek öğelerin belirlenmesi ve seviyelendirilmesinde çok yararlandığını belirtmiştir.

Sıklık, metinlerin ölçülebilir özellikleri arasında yer almaktadır. Sezgin (1993: 100) metinlerdeki harfleri, harflerin birbirini takip ediş oranlarını, Türkçedeki hece tiplerini ve kelimelerin kullanım sıklıklarını metinlerin ölçülebilir özellikleri arasında sıralamıştır. Dolunay (2012: 84) metinlerin bu özelliklerinden, anlama ve anlatma becerilerinin geliştirilmesinde ve dil bilgisi öğretiminde yararlanılabileceğini belirtmektedir.

Bir dil üzerine yapılan sıklık çalışmaları o dilin eğitimi/öğretimi açısından önemli bilgileri ortaya çıkartmaktadır. Yaygın olarak yapılan sıklık çalışmaları, bir dilde en sık kullanılan kelimeleri ve veya diğer söz varlığı unsurlarını belirlemeye yöneliktir (Baş ve Karadağ, 2012). Bu çalışmaların çıkış mantığı, hemen her dilde bazı kelimelerin diğerlerine oranla çok daha sık kullanıldığıdır. Dilde sık kullanılan söz varlığı unsurlarının belirlenmesi ve eğitim öğretim sürecinin planlanmasında bu unsurlara öncelik verilmesi dilin işlevsel olarak öğretimini kolaylaştırmaktadır. Sıklığa bağlı olmak üzere, öğrencilerin yaş ve hazırbulunuşluk özelliklerini dikkate alarak okuma içeriklerinin seviyelendirilmesi de konunun pratik diğer yönünü oluşturmaktadır.

Türkçede konu üzerinde yapılan çalışmaların çok büyük bölümü kelime, sınırlı olarak da deyim ve atasözü sıklıklarını belirlemeye yöneliktir. Hece ve ses sıklıklarını belirlemeye dönük çalışmalar son derece sınırlıdır (Çimen vd. 2007; Dolunay, 2009; Ferah, 2011; Özmen, 2012). Bu çalışmalar, yazılı metinlerden oluşturulan özel amaçlı küçük hacimli derlemlerdeki harf sıklıklarının belirlenmesine yöneliktir.

Türkçedeki ses-harf sıklığı üzerine en kapsamlı ve dikkat çekici çalışma *Şifrelerin Matematiği: Kriptografi* (Çimen vd., 2007) isimli kitaptır. Bütün dillerde harflerin kullanım sıklıklarının farklılık gösterdiği gerçekliğinden hareketle, gerek şifrelemede gerekse şifre çözmede harf sıklıklarından hareket edilmesi dolayısıyla bu çalışmada Türkçedeki harf sıklıkları üzerine bir analiz yapılmıştır. Çalışmada (Çimen vd., 2007) 6.421.357 harften oluşan metinler üzerinde hesaplamalar yapılarak harflerin kullanım sıklıkları bindelik oranlarla belirlenmiştir:

Tablo 1.

Şifrelerin Matematiği: Kriptografi Kitabındaki Harf Sıklıkları

SIRA	HARF	Binde
1.	A	121
2.	İ	107
3.	E	95
4.	N	71
5.	R	68

6.	L	62
7.	I	48
8	K	47
9.	D	41
10.	M	37
11.	Y	31
12.	T	31
13.	U	29
14.	S	28
15.	B	25
16.	O	22
17.	Ş	16
18.	Ü	15
19.	Z	14
20.	G	13
21.	H	11
22.	Ğ	11
23.	Ç	10
24.	C	9
25.	V	8
26.	P	8
27.	Ö	7
28.	F	5
29.	J	0,5

Türkiye Türkçesinde Ses Sıklığı (Özmen, 2012) isimli çalışmada Cumhuriyet dönemini temsil eden 10 eserden bir havuz oluşturulmuş ve bu havuz içinde yer alan harfler sayılmıştır. Çalışmada ünlüler yüzdelik olarak, ünsüzler ise havuzdaki sıklık değerlerine göre sıralanmıştır. Ünlülerin dağılımı şu şekilde verilmiştir: a, e ünlüleri %48.83, i, ı ünlüleri %31.77, u, ü ünlüleri %11.91 ve o, ö ünlüleri %7.46 oranındadır. Özmen'in çalışmasında (2012: 735-736) ünsüzlerin dağılımına ilişkin verdiği değerler aşağıdaki şekilde tablolştırılmıştır:

Tablo 2.

Türkiye Türkçesinde Ses Sıklığı adlı Çalışmadaki Ünsüzlere İlişkin Sıklık Listesi

SIKLIĞA GÖRE ÜNSÜZLER		
1.	R	119.097
2.	N	115.138
3.	L	96.651
4.	D	76.513
5.	K	64.754
6.	M	59.486
7.	T	52.023
8	Y	52.10
9.	S	51.108
10.	B	50.295
11.	Ş	28.958

Türkçe Eğitimi Açısından Ses-Harf Sıklığı Üzerine Bir Değerlendirme

12.	Z	28.673
13.	H	21.167
14.	G	20.510
15.	Ç	18.928
16.	V	17.778
17.	Ğ	16.108
18.	C	13.935
19.	P	12.668
20.	F	8.023
21.	J	314
Toplam:		924.137

Ses temelli cümle yönteminin uygulanmasına ilişkin değerlendirmeler yaptığı çalışmasında Ferah (2011) Türkçedeki ses sıklığının ve Türkçe dil bilgisinin ilk okuma yazma öğretiminde temel alınmadığını belirtmektedir. Çalışmada Ferah (2011: 37) *Türkçe İlk Okuma Yazmayı Öğrenme* (Ferah, 2007) isimli kendi kitabında Türk alfabesindeki seslerin sıklığını, Türkçe sözlüğü tarayarak kelimelerin sıklığına dayalı olarak verdiğini belirtmektedir. Ferah (2011) Çimen ve arkadaşlarının (2007) çalışmasında yer alan ses sıralamalarını aktarmaktadır.

Yazılı metinlerden oluşturulan metin havuzlarının harf sıklığı çözümlenmelerine dayanan bu çalışmalar dışında Türkçede ses-harf sıklığına dair çalışmaya rastlanmamıştır. Türkçede ses sıklığının nesnel yöntemlerle belirlenmesi, ana dili eğitiminde ses temelli ilk okuma yazma sürecinin planlanmasına, yabancı dil öğretiminde ses eğitiminin önceliklerinin belirlenmesine olanak sağlayacaktır. Okuma yazma öğretimine hangi seslerle başlanıp hangi seslerle devam edileceğini gösterir en somut kanıt, dildeki ses sıklığıdır.

Ses sıklığının belirlenebilmesinde ideal yol, konuşma dilini esas alan bir derlemin oluşturulması ve sıklığın bu derlemden çıkarılmasıdır. Ancak tamamı konuşma diline dayalı ve kapsamlı bir derlem oluşturmak son derece güçtür ve Türkçe için oluşturulmuş bu nitelikte bir derlem bulunmamaktadır. Bu durumda dildeki ses sıklığının belirlenmesinin diğer yolu ağırlıklı olarak yazı dilinden oluşturulan bir derlem içinde harf sıklıklarının hesaplanmasıdır. Bu çalışmanın amacı, Türkçe ses-harf sıklığının güvenilir bir sonucuna ulaşmaktır. Bu amaçla Türkçe Sözlük'teki sözlük birimlerin ve güncel, çevrimiçi, geniş kapsamlı bir Türkçe derlemin ses-harf sıklığı analizi yapılmıştır.

Yöntem

Çalışmada ilk olarak Türkçe Sözlük'te (2011) yer alan sözlük birimlerin harf sıklığı hesaplanmıştır. Sözlükteki sözlük birimler bir havuz haline getirilmiş ve Microsoft Word 2010'da yer alan "*Bul ve Değiştir*" özelliği kullanılarak Türkçe alfabedeki her harf tek tek "*Bul ve Değiştir*" işlemine tabi tutulmuştur. Microsoft Word bu işlemi gerçekleştirdiğinde işlem sayısını vermektedir. İşlem alfabedeki 29 harfin tamamı için tekrarlanmıştır. Elde edilen değerler sözlük birimlerden oluşturulan havuz içindeki harflerin kullanım sıklığı olarak kaydedilmiştir.

Çalışmanın derlem tabanlı verileri Türkçe derlemler, veri setleri, doğal dil işleme araçları üretmeyi hedefleyen ve bağımsız bir proje olarak yürütülen TS Corpus'tan (<https://tscorpus.com/>) elde edilmiştir. TS Corpus (<https://tscorpus.com/>) Mart 2012 tarihinde, Türkiye Türkçesi için sözcük türü ve biçimbirimsel olarak etiketlenmiş ilk genel amaçlı çevrimiçi Türkçe derlemi olarak kullanıcılara sunulmuştur. Bu derlem TS Corpus V2 (<http://tscorpus.com/corpora/ts-corpus-v2/>) olarak anılmaktadır.

TS Corpus V2 derleminden veri kaynağı olarak kullanılan toplam havuz 491.360.398 birimden (token) oluşmaktadır. Elde edilen bu veri, Linux kabuğu (shell) üstünde çalışan *grep*, *uniq*, *sort* vb. komutlar ve *vim* metin editörü yardımıyla işlenmiştir. Harf sıklığı bilgisine ulaşmak için *grep* ve *grep'in* sayısal sonuç döndüren *-c parametresi* kullanılmıştır. Komut şu şekilde çalıştırılmaktadır: *grep -c "a" [girdi_veri]*. Böylece, Türkçe alfabede bulunan 29 harfin, derlem içindeki ham sıklığı elde edilmiştir.

Bulgular

Bu çalışmada Türkçe Sözlük'ten (2011) ve TS Corpus V2 elde edilen bulgular iki alt başlıkta verilmiştir.

Türkçe Sözlük'teki Sözlük Birimlerin Ses-Harf Sıklıklarına İlişkin Bulgular

İncelenen Türkçe Sözlük'te (2011) toplam 110.785 sözlük birim yer almaktadır. Sözlük birimlerde uzun ünlüler olan â 1219, î 264 ve û 82 kez kullanılmıştır. Bu harflerin kullanım sıklıkları toplamı sıralamaya etki etmediği için değerlendirme dışında bırakılmıştır. Türkçe Sözlük'te (2011) yer alan sözlük birimlerinin toplamda 1.051.332 harften oluştuğu tespit edilmiştir. Harflerin sıklık sıralaması aşağıdaki şekildedir:

Tablo 3.

Türkçe Sözlük'teki Biçim Birimlerin Harf Sıklıkları

TÜRKÇE SÖZLÜK HARF SIKLIKLARI		
SIRA	HARF	SIKLIK
1.	A	137.712
2.	E	96.044
3.	K	87.184
4.	L	71.120
5.	İ	69.210
6.	M	66.634
7.	R	56.282
8	I	48.245
9.	N	46.948
10.	T	46.467
11.	S	34.319
12.	U	28.692
13.	Y	26.896
14.	O	25.559
15.	D	25.512
16.	B	22.974
17.	Ş	20.179
18.	Ü	19.616
19.	Z	18.737
20.	Ç	14.829
21.	G	13.880
22.	P	12.659
23.	V	12.287
24.	C	11.960
25.	H	11.243
26.	Ğ	8.833
27.	Ö	8.195
28.	F	7.998
29.	J	1.118
Toplam Harf: 1.051.332		

Türkçe Sözlük'te (2011) yer alan sözlük birimlerde/madde başı sözcüklerde en sık geçen ilk beş harf A, E, K, L ve İ'dir. Bu harflerin kullanım sıklıklarının toplamı 461.270'tir. Bu değer toplam harf sayısının yaklaşık %44'üne karşılık gelmektedir. Buna karşılık listenin son beş sırasında yer alan H, Ğ, Ö, F ve J harflerinin kullanım sıklığı toplamı 37.387'dir. Bu değer toplam sıklığın yaklaşık %0,3'üne karşılık gelmektedir. Harf sıklığına ilişkin sözlük birimlerden/madde başı sözcüklerden elde edilen bu değerler Türkçenin karakteristik özelliklerinden biri olarak görünmektedir.

Türkçe Eğitimi Açısından Ses-Harf Sıklığı Üzerine Bir Değerlendirme

TS Corpus V2 Derlem Havuzundaki Ses-Harf Sıklıklarına İlişkin Bulgular

Çalışmada incelenen derlem havuzu toplam 491.360.398 birimden (token) oluşmaktadır. Derlem havuzunda tespit edilen toplam karakter sayısı 3.589.428.506, toplam harf sayısı 2.594.542.434'tür. Veriden, rakamlar, noktalama işaretleri ve benzeri harf dışı öğeler (@, ®, ™, vb.) ile düzeltme işareti içeren harfler ve Türkçe alfabede yer almayan harfler ayıklanmıştır. Ayrıca küçük/büyük harf ayrımının işleme ve elde edilecek sayısal sonuçlar üstünde etkisini ortadan kaldırmak için tüm veri küçük harfe dönüştürülmüştür. Veri içinde yer alan alfabe dışı karakter sayısı 994.886.072'dir. Türkçe Alfabede bulunan karakterlerin toplam sayısı 2.593.114.688, Türkçe olmayan karakterlerin toplam sayısı 1.427.746'dır. Türkçe olmayan karakterlerin derlem içindeki sıklığı şu şekildedir: X - 561.061, W - 811.025, Q - 55.660.

TS Corpus V2 Türkçe ses-harf sıklığı şu şekildedir:

Tablo 4.

TS Corpus V2 Harf Sıklıkları

TS CORPUS V2 SIKLIK		
SIRA	HARF	SIKLIK
1.	A	314.559.952
2.	E	243.772.539
3.	İ	223.614.150
4.	N	200.333.050
5.	R	179.602.237
6.	L	175.330.895
7.	I	129.750.538
8	K	121.380.140
9.	D	112.663.708
10.	T	89.048.257
11.	M	87.490.725
12.	Y	81.411.137
13.	S	80.285.662
14.	U	78.887.121
15.	O	66.439.499
16.	B	62.988.409
17.	Ü	48.846.539
18.	Ş	40.465.209
19.	Z	37.410.299
20.	G	34.338.246
21.	V	26.697.361
22.	Ğ	26.465.893
23.	C	25.066.137
24.	Ç	24.958.954
25.	P	24.437.375
26.	H	23.223.919
27.	Ö	19.515.662
28.	F	12.293.485
29.	J	1.837.590
Toplam Harf: 2.593.114.688		

Derlem havuzundan elde edilen sıklık sıralamasında ilk beş sırada A, E, İ, N ve R harfleri yer almaktadır. Bu harflerin sıklık değerlerinin toplamı 540.237.160'tır. Bu değer derlem havuzundaki toplam harf sayısının yaklaşık %21'ine karşılık gelmektedir. Buna karşılık listenin son beş sırasında yer alan P, H, Ö, F ve J harflerinin kullanım sıklığı toplamı 81.308.031'dir. Bu değer derlem havuzundaki toplam sıklığın yaklaşık %0,3'üne karşılık gelmektedir.

Türkçe Sözlük'teki (2011) sözlük birimlerde en sık geçen ilk beş harfin kullanım sıklığının toplam sıklığa oranı (%44) ile derlem havuzunda en sık geçen ilk beş harfin kullanım sıklığının toplam sıklığa oranı (%21) arasında oransal olarak iki kat fark oluşmaktadır. Bu farkı yaratan derlem içindeki kelimelerin kullanım sıklıklarıdır. Türkçe Sözlük'ten (2011) elde edilen havuzda her kelime/sözlük birim bir kez geçmektedir. Buna karşılık derlem havuzunda kelimeler kullanım sıklıklarına göre yer almaktadır. Ancak gerek sözlük gerek derlem havuzunda son beş harfin kullanım sıklıklarının toplam sıklığa oranı benzer bir sonuç ortaya çıkarmaktadır. Her iki veri setinde %0,3'lük oran tespit edilmiştir. Her iki listede son beş sırada yer alan H, Ö, F ve J harfleri ortaktır. Elde edilen sonuca bu durumun etkisi de söz konusudur.

Tartışma, Sonuç ve Öneriler

Ses-harf sıklığına ilişkin mevcut listeler, bu çalışma kapsamında belirlenen ses-harf sıklık listeleri aşağıdaki tabloda bir arada sunulmaktadır:

Tablo 5.

Mevcut Harf Sıklık Çalışmalarının Karşılaştırmalı Listesi

SIRA	Çimen vd.	Türkçe Sözlük	TS Corpus V2
1.	A	A	A
2.	İ	E	E
3.	E	K	İ
4.	N	L	N
5.	R	İ	R
6.	L	M	L
7.	I	R	I
8.	K	I	K
9.	D	N	D
10.	M	T	T
11.	Y	S	M
12.	T	U	Y
13.	U	Y	S
14.	S	O	U
15.	B	D	O
16.	O	B	B
17.	Ş	Ş	Ü
18.	Ü	Ü	Ş
19.	Z	Z	Z
20.	G	Ç	G
21.	H	G	V
22.	Ğ	P	Ğ
23.	Ç	V	C
24.	C	C	Ç
25.	V	H	P
26.	P	Ğ	H
27.	Ö	Ö	Ö
28.	F	F	F
29.	J	J	J

Bu çalışma kapsamında yararlanılan TS Corpus V2'den elde edilen ses-harf sıklıkları en nesnel veriyi oluşturmaktadır. 491.360.398 kelimelik hacmiyle Türkçe için oluşturulan en büyük derlem

Türkçe Eğitimi Açısından Ses-Harf Sıklığı Üzerine Bir Değerlendirme

özelliğine sahip olan TS Corpus V2 içinde bu çalışma için 2.593.114.688 karakter/harf sıklık açısından değerlendirilmiş ve Türkiye Türkçesi için ses-harf sıklıkları tespit edilmiştir.

Yukarıdaki tablo incelendiğinde sıklık bakımından son sırada yer alan ses-harfin tüm listelerde ortak olduğu görülmektedir. Diğer seslerde farklı listeler zaman zaman kesişse de kapsamlı bir uyumdan söz etmek mümkün gözükmemektedir. Çimen vd. (2007) tarafından şifrelemede kullanılmak üzere hazırlanan sıklık listesi ile TS Corpus V2 listesi arasında büyük oranda uyum olduğunu görülmektedir.

Türkçe Sözlük'teki sözlük birimlerin ses sıklığı ile TS Corpus V2 arasındaki uyumun ikinci sırada yer aldığı söylenebilir. Listelerin üst sıralarında daha yüksek bir uyum bulunurken ortalarında bu uyumun bozulduğu ve listenin sonunda, yani sıklığı düşük olan ses-harflerde tekrar uyumun olduğu söylenebilir.

2005 yılında yürürlüğe giren *İlköğretim Türkçe Dersi Öğretim Programı ve Kılavuzu*'nda (MEB:2005) ve 2015 yılında yürürlüğe giren *Türkçe Dersi (1-8. Sınıflar) Öğretim Programı*'nda (MEB: 2015) ilk okuma yazma öğretimi için önerilen ses sıralamaları aşağıdaki tabloda gösterilmektedir:

Tablo 6.

Türkçe Dersi Öğretim Programları Ses-Harf Sıralamaları

2005		2015	
Program		Program	
1.	E	1.	E
2.	L	2.	L
3.	A	3.	A
4.	T	4.	N
5.	İ	5.	İ
6.	N	6.	T
7.	O	7.	O
8.	R	8.	B
9.	M	9.	U
10.	U	10.	K
11.	K	11.	İ
12.	İ	12.	R
13.	Y	13.	Ö
14.	S	14.	S
15.	D	15.	Ü
16.	Ö	16.	M
17.	B	17.	D
18.	Ü	18.	Ş
19.	Ş	19.	Y
20.	Z	20.	C
21.	Ç	21.	Z
22.	G	22.	Ç
23.	C	23.	G
24.	P	24.	P
25.	H	25.	H

26.	Ğ	26.	F
27.	V	27.	V
28.	F	28.	Ğ
29.	J	29.	J

İlk okuma yazma öğretimi için 2005 ve 2015 yıllarında hazırlanan öğretim programlarında oluşturulan ses grupları ve bu yolla ortaya çıkan ses sıralamalarının diğer listelerle uyumlu olmadığını söylemek mümkündür. Seslerin öğretimi açısından ünlü ünsüz dengesinin kurulması ve Türkçenin yapısına uygun hece oluşturma seçenekleri dikkate alınarak sıralamada değişikliklerin yapılması doğal olmakla birlikte 2005 programında ortaya çıkan sıralamada böylesi bir durumun gözlemlenmediği, buna karşılık 2015 programında bu açıdan bir düzenleme yapıldığı söylenebilir.

Bu çalışmada TS Corpus V2 derleminden hareketle Türkiye Türkçesi için belirlenen ses-harf sıklıkları, yabancı dil öğretiminde hangi seslerin öğretilmesine öncelik verilmesi gerektiği konusuna da ışık tutmaktadır. Türkçeyi yabancı dil olarak öğrenecek kişilerin hem ses farkındalıklarını geliştirmek hem de söyleyiş özelliklerinin iyileştirilmesinde kullanım sıklığının esas alınması geçerli bir yol olacaktır.

Kaynaklar

- Baş, B. ve Karadağ, Ö. (2012). Söz varlığı üzerine yurt dışında ve Türkiye’de yapılan temel araştırmalar. *Millî Eğitim*, 193, 81-105.
- Çimen, C., Akleyek, S. ve Akyıldız, E. (2007). *Şifrelerin matematiği: Kriptografi*. Ankara: ODTU Yayıncılık.
- Dolunay, S.K. (2009). İlköğretim ikinci kademedeki zaman ekleri ve fonksiyonlarının öğretimi. Ankara: Gazi Üniversitesi Eğitim Bilimleri Enstitüsü (Yayımlanmamış Doktora Tezi).
- Ferah, A. (2011). 2005 Türkçe öğretim programı ve ses temelli cümle yönteminin ses sıklığı ve hece oluşturma açısından incelenmesi. *Çağdaş Eğitim*, 36 (383), 34-41.
- Komisyon. (2005). *İlköğretim Türkçe dersi öğretim programı ve kılavuzu*. Ankara: Devlet Kitapları Müdürlüğü Basım Evi.
- Komisyon. (2015). *Türkçe dersi (1-8. sınıflar) öğretim programı*. Ankara: MEB Yayınları.
- Korkmaz, Z.(1992). *Grammer terimleri sözlüğü*. Ankara: TDK.
- Sezgin, F. (1993). *Dil ve edebiyatta istatistik ve bilgisayar uygulamaları*. İstanbul: Dergâh Yayınları.
- Özmen, M. (2012). Türkiye Türkçesinde ses sıklığı. Prof. Dr. Mine Mengi Adına Türkoloji Sempozyumu (20-22 Ekim 2001) Bildiri Kitabı, 729-739.
- TDK. (1998). *Türkçe sözlük*. Ankara: TDK.
- TDK. (2011). *Türkçe sözlük*. Ankara: TDK.
- Vardar, B. (1998). *Açıklamalı dilbilim terimleri sözlüğü*. İstanbul: ABC Kitabevi.
- TS Corpus (2019). <http://ts corpus.com>, Erişim Tarihi: 20.10.2019
- TS Corpus V2 (2019). <http://ts corpus.com/corpora/ts-corpus-v2/> Erişim Tarihi: 20.10.2019.

Extended Abstract

Introduction

Some of the researches in the field of language teaching are frequency studies. Frequency is one of the measurable characteristics of texts. The majority of the studies on frequency in Turkish are aimed at determining the frequency of words and limited idioms and proverbs. Studies to determine syllable and phoneme frequencies are extremely limited.

Studies on the frequency of phoneme-letters (Çimen et al. 2007; Dolunay, 2009; Ferah, 2011; Özmen, 2012) are the analysis of limited text pools composed of written texts. There is no comprehensive study on this subject based on corpus. Determining the phoneme- frequency in Turkish by objective methods will enable the planning of the first phoneme-based literacy process in mother tongue education and the priorities of phoneme education in foreign language teaching.

The ideal way to determine the frequency of phoneme is to create a collection based on spoken language and to obtaining the frequency from this corpus. However, it is extremely difficult to create a comprehensive corpus based entirely on spoken language and there is no such corpus for

Turkish. In this case, the other way to determine the phoneme frequency in the language is to calculate the letter frequencies in a corpus composed mainly of written language. The aim of this study is to reach a reliable result of Turkish phonetic frequency. For this purpose, phoneme-letter frequency analysis was performed in a current, online and comprehensive corpus.

Method

In this study, firstly, the frequency of letters of dictionary tokens in Turkish Dictionary (2011) was calculated. The dictionary tokens in the dictionary were pooled and each letter in the Turkish alphabet was subjected to find and replace process by using "Find and Replace" feature in Microsoft Word 2010. When Microsoft Word performs this operation, it gives the number of transactions. The process was repeated for all 29 letters in the alphabet. The obtained values are recorded as frequency of usage of letters in the pool formed from dictionary tokens.

The corpus-based data of the study was obtained from TS Corpus (<https://tscorpus.com/>), an independent project aimed at producing Turkish corpora, datasets, and natural language processing tools. The total pool used as a data source from the TS Corpus V2 corpus consists of 491,360,398 tokens. Obtained data passed with Linux shell (shell) running on grep, uniq, sort and so on. commands and vim text editor. In order to reach the letter frequency information, the -c parameter of grep and grep, which returns numerical results, is used. The command is executed as follows: grep -c "a" [input_data]. Thus, the raw frequency of the 29 letters in the Turkish alphabet was obtained.

Result and Discussion

The corpus pool examined in the study consists of a total of 491,360,398 tokens. The total number of characters detected in the corpus pool is 3,589,428,506 and the total number of letters is 2,594,542,434. Numerals, punctuation and similar non-letter items, letters with correction marks and letters not included in the Turkish alphabet were extracted from the data. In addition, all data were converted to small letters to eliminate the effect of small letter separation on processing and numerical results. The number of non-alphabetic characters in the data is 994.886.072. Turkish total number of characters in the alphabet is 2,593,114,688 and the total number of non-Turkish characters is 1,427,746. The non-Turkish characters in the corpus are as follows: X - 561.061, W - 811.025, Q - 55.660.

The phoneme-letter frequencies obtained from TS Corpus V2 used in this study constitute the most objective data. For this study within the TS Corpus V2 having the largest Turkish TSE with a 491 360 398 words, 2593114688 characters created for the feature evaluated in terms of frequency and phoneme-letter frequency was determined to Turkey Turkish.

When the lists created in the study are examined, it is seen that the last phoneme-letter in terms of frequency is common in all lists. In other phonemes, even the different lists sometimes intersect, it is not possible to speak of a comprehensive harmony. Çimen et al. (2007) frequency list to be used in the encryption can be seen to have a harmony with TS Corpus V2 list in a large scale.

It can be said that the harmony between the phoneme frequency of the dictionary tokens in the Turkish dictionary and TS Corpus V2 takes the second place. It can be said that there is a higher fit at the top of the lists, this harmony is broken in the middle and the harmony occurs again at the end of the list, that is, the lower-frequency phoneme letters.

It is possible to say that the phoneme groups formed in the curricula prepared in 2005 and 2015 for the first literacy teaching and the resulting phoneme sequences are not compatible with the other lists.