



İlkokul Türkçe Dersi Okuma Metnindeki Kelime, Harf ve Hece Sıklıklarına Karşılaştırmalı Bir Bakış

*Mehmet Fatih KARACA**

Öz

Sıklık, metinlerin ölçülebilir özelliklerindedir. Sıklık analiziyle elde edilen veriler ilk okuma-yazma ve Türkçe dil bilgisi öğretimi, Türkçe ders kitaplarının yazımı ve değerlendirilmesi, müfredatının oluşturulması ve yenilenmesi gibi birçok konuda önemli bilgiler içerebilir. Bu çalışmada ilkökul 1, 2, 3 ve 4. sınıf Türkçe ders kitaplarında bulunan bütün okuma metninin kelime, harf ve hece sıklıkları saptanmış, sonuçlar sınıf ve yayınevlerine göre kategorik olarak sunulmuştur. 1. sınıf metinleri daha az sayıda kelimedenden oluşmuştur. Sınıf seviyesi arttıkça ve ilk okuma metninden son okuma metnine doğru ilerledikçe kelime sayısı da artmıştır. Kelimelerin ağırlıklı olarak 5 harfli ve 2-3 heceli olduğu, ünlü ve ünsüz uyumuna uydukları belirlenmiştir. Farklılıklar olsa da sınıf seviyeleri ve yayınevlerine göre en çok ve en az kullanılan harflerin yakın oranda seyrettiği saptanmıştır. Hecelerin en çok CV tipinde bulunduğu, ortalama hece sayısının sınıf seviyesi ve yayınevine göre benzerlik gösterdiği, bu değerlerin ilk okuma metninde düşükken son metinlere doğru yükseldiği tespit edilmiştir. Seslerin öğretim sırasıyla harflerin kullanım sıklıklarının örtüşmediği fakat 2019 Türkçe Dersi Öğretim Programı'nın 1. grubunda yer alan harflerin sıklık oranının %47'ye ulaştığı görülmüştür. Çalışma kapsamında analiz edilenler itibarıyla aynı sınıf seviyesindeki farklı yayınevlerinde birbirine yakın sonuçların alınması olağan bir durumken farklı sınıf seviyelerinde de birçok benzerlik görülmesi, sonuçların aynı veya çok yakın olması dikkat çekici ve vurgulanması gereken bir nokta olarak ortaya çıkmıştır.

Anahtar Kelimeler: Sıklık, sıklık analizi, ilkökul, Türkçe dersi, öğretim programı, doğal dil işleme

A Comparative View on Frequencies of Word, Letter and Syllables in Reading Texts of Elementary School Turkish Language Arts Course

Abstract

Frequency is one of the measurable characteristics of texts. Data obtained via frequency analysis may contain important information on various topics such as first reading-writing and Turkish language grammar instruction, writing and evaluating Turkish language arts course textbooks, developing and revising curriculums, etc. In this study, frequencies of word, letter and syllables in all reading texts included in Turkish language arts course textbooks of elementary school 1st, 2nd, 3rd and 4th classes were determined, results were presented as categorical data considering classes and publishers. Texts of 1st classes consisted of fewer words. The number of the word increases as the level of class increases and also considering the progress from the first reading text to last reading text. It was found out that the words predominantly consist of 5 letters and 2-3 syllables, obey vowel and consonant harmony rules. It was determined that the ratio of the letters mostly and least used are almost equal according to the level of the classes and publishers in spite of the fact that there are differences. It was found out that the syllables are mostly in CV type, the average syllable number displays similarity according to the level of the classes and publishers, and this value increases considering the progress from the first reading text to last reading text. It was seen that the instructional sequence of speech sounds does not overlap with the usage letter

* Dr. Öğr. Üyesi, Tokat Gaziosmanpaşa Üniversitesi, Erbaa Meslek Yüksekokulu, Bilgisayar Teknolojileri Bölümü, Erbaa, Tokat, mehmetfatih.karaca@gop.edu.tr, ORCID: orcid.org/0000-0002-7612-1437

frequencies but the frequency ratio of letters included in 1st group of 2019 Turkish Course Curriculum reaches to 47%. Considering the texts analysed within this study, obtaining similar results among different publishers within the same level of classes can be regarded as acceptable, on the other hand, obtaining similar results within the different level of classes can be regarded as noteworthy and should be emphasized.

Keywords: Frequency, frequency analysis, elementary school, Turkish lesson, curriculum, natural language processing

Giriş

Sosyal varlık olan insanlar, diğer kişilerle çeşitli yöntemler kullanarak iletişim kurmaktadır. Sözlü iletişime dayanan konuşma dilleri bu yöntemlerden biridir. Aralarında görülen benzerlik veya yakınlıktan kaynaklı olarak diller ailesine, kökenine ve yapısına göre sınıflandırılır. Türkçe kaynak bakımından Ural-Altay dil grubunun Altay koluna bağlı, yapı bakımındansa eklemeli diller kategorisindedir. Dünyanın farklı coğrafyalarında Türkçe gibi birçok dil konuşma ve yazma dili olarak kullanılmaktadır.

Sayıları tam olarak henüz bilinmeyen dillerin kendine özgü kuralları vardır. Zamanla oluşan, dilin oluşmasında, korunmasında ve gelişmesinde önemli bir yere sahip olan kurallar bütününe dilbilgisi denir. Dilbilgisinde olduğu gibi dillerin birtakım özellikleri de dilden dile değişkenlik gösterebilir. Örneğin bir dilde bulunan sözcük sayısı, alfabesindeki ünlü ve ünsüz harfler, harflerin kullanım sıklıkları ve konumları, hece şekilleri gibi birçok özellik dillerde başka şekillerde görülmekte; ses uyum kuralları ve hece oluşturma vb. işlemler de dilin kendi dinamikleri içerisinde meydana gelmektedir.

Türkçe gerek dilbilgisi kuralları gerekse de kullanım özellikleri itibarıyla diğer dillerden farklılık sergilemektedir. Örneğin Türkçede 29 harften oluşan alfabe kullanılmaktayken, İngilizce ve Fransızcada 26, İtalyancada ise 21 harfli alfabe kullanılmaktadır. İngilizce ve İtalyancada “a, e, i, o, u”, Fransızcada “a, e, i, o, u, y”, Türkçede ise daha fazla sayıda olmak üzere “a, e, ı, i, o, ö, u, ü” ünlüleri bulunmaktadır. Diğer dillerle kıyaslandığında Türkçe ünlü harf sayısı bakımından zengin bir dildir.

Bir dilde yer alan ünlü harf sayısı ne kadar fazla ise dil o denli ahenkli olur. Bunun bir sonucu olarak Türkçe dünyada konuşulan diller arasında en ahenkli üç dört dil arasında yer almaktadır (Hengirmen, 2007: 33, 35). Ayrıca, Türkçede bunu sağlayan kelimedeki ünlü harfleri incelik-kalınlık (önlük-artlık) bakımından inceleyen büyük ünlü uyumu ve düzlük-yuvarlaklık bakımından inceleyen küçük ünlü uyumu kuralları vardır (Demir ve Yılmaz, 2011: 57-61). Ünlü harflerin yanı sıra kelimedeki ünsüzlerin seda bakımından uyumunu inceleyen ünsüz uyumu kuralı da bulunmaktadır. Bu kurallar, belirli seslerin belirli seslerle bir arada bulunabileceğine ve birbirlerini takip edeceklerine ilişkin kuralları ihtiva eder. Örneğin büyük ünlü uyumu, kelimenin ilk hecesinde ince ünlü harf varsa sonrasında ince, kalın ünlü harf varsa sonrasında kalın ünlü harflerin gelmesini ünlü harfler açısından incelemektedir.

Türkçede yer alan hecelerle ilgili de birtakım kurallar vardır. Hece özetle bir hamlede söylenebilen, yalnız bir ünlü harf barındırabilen, en az bir ve en fazla dört harfli olabilen, sadece ünsüzlerden oluşamayan ve genelde bir ünlü ile birkaç ünsüzün birleşiminden meydana gelen ses gruplarıdır. Her ne kadar bazı çalışmalarda Türkçede 17 tip hece olduğu ifade edilmiş olsa da (Gürsoy-Naskali, 1997) genel görüş 6 tip hece bulunduğuudur. Bu hece tiplerinde hecenin çekirdeğini oluşturan ünlü harfler heceyi kurarlar ve tek başlarına hece olarak bulunabilirler (Banguoğlu, 1974: 54; Demir ve Yılmaz, 2011: 62; Ergin, 1980: 104-105; Gencan, 1979: 36-37).

Yukarıda Türkçeye dair bahsedilen bazı kurallar başka dillerde görülmeyebilir, farklı şekillerde görülebilir veya bunlardan hariç farklı kurallar da görülebilir. Bu sebeple dillerin dil bilgisi kuralları veya kelime, harf ve hece gibi dile ait bazı unsurların kullanım özelliklerinin ve sıklıklarının dilden dile farklılık gösterdiğini ifade etmek gerekir. Örneğin bir dildeki ünlü veya ünsüz harflerin kullanım sıklıkları farklı bir dilde aynı sırada seyretmeyebilir. Türkçede ünlü harflerin kullanım sıklıkları sırasıyla “a, e, i, ı, u, o, ü, ö” iken İngilizcede “e, o, a, i, u” şeklinde ortaya çıkmaktadır. Türkçede en çok kullanılan ünsüz harfler sırasıyla “n, r, l, d, k, m, y, t, s, b, ş, z, g, h, ç, ğ, c, v, p, f, j” (Serengil ve Akin, 2011), İngilizcede ise farklı sıralamada ve şu şekildedir “t, n, s, r, h, l, d, c, m, f, p, g, w, y, b, v, k, x, j, q, z” (Solso ve King, 1976). Harf ve hecelerin sıklıkları yalnızca istatistiksel bir bilgi olmanın ötesinde farklı amaçlar için önemli bir

veri olma özelliğini de barındırmaktadır. Örneğin bu verilerden okuma ve yazma öğretiminde istifade edilebilir.

İlk okuma-yazma öğretiminde çeşitli yöntemler kullanılmaktadır. 2019 Türkçe Dersi Öğretim Programı'nda "Tematik Yaklaşım" esas alınmış ve ilk okuma-yazmada "Ses Temelli İlk Okuma Yazma Öğretimi" benimsenmiştir (Millî Eğitim Bakanlığı [MEB], 2019). Fakat konuyla alakalı tartışmalardan birisi ilk okuma-yazma öğretimindeki harf öğretim sıralamasının Türkçe harf kullanım sıklığıyla aynı düzende gitmediğidir (Ferah, 2011). Buna karşın yapılan alan yazın incelemesinde Türkçe ders kitaplarında bulunan okuma metinlerindeki kelime, hece ve harf sıklıklarının birlikte analiz edildiği çalışmaya rastlanamamıştır.

Sıklık analizi, veride bulunan belirli unsurların kullanım sıklıklarının belirlenmesi ve analiz edilmesi işlemidir. Dil çalışmaları açısından sıklık analizi ise metin içindeki deyim, atasözü, kelime, hece, harf vb. özelliklerin frekans olarak sayılması ve kategorik olarak sunulması şeklinde tanımlanabilir. Sıklık analiziyle metinlerin veya bir dilin özellik ve var olan durumları net bir şekilde nicel olarak ortaya konulabilmektedir.

Sıklık, metinlerin ölçülebilir özelliklerinden biridir (Karadağ, 2020). Halen daha sıklık belirleme işleminin araştırmacılar tarafından elle yapıldığı görülmektedir. Fakat günümüz bilgisayar teknolojilerinin gelişmesiyle birlikte eldeki büyük boyuttaki verilerin sıklık analizlerinin bilgisayar yazılımlarıyla yapılmasının zaman tasarrufu sağlaması, insan faktöründen kaynaklı hataların azaltılması ve dolayısıyla daha tutarlı sonuçlar elde edilmesi gibi birçok avantajı bulunmaktadır. Bu avantajları sebebiyle yalnızca sıklık analizlerinin belirlenmesinde değil doğal dillerde yazılmış neredeyse tüm metinlerin analizinde bilgisayar teknolojilerinde yaşanan gelişmelerden faydalanmak daha doğru olacaktır. Doğal Dil İşleme (DDİ), bilgisayar bilimindeki bu gelişmelerden biridir.

DDİ, en genel tanımıyla doğal dillerle oluşturulmuş metinsel verilerin dilin kendi kuralları çerçevesinde bilgisayar aracılığıyla çözümlenmesi, anlaşılması veya yeniden üretilmesidir. Yapay zekâ ile dilbilimin bir alt birimi olan DDİ, insan bilgisayar etkileşiminin seviyesini belirlemesi açısından yüksek bir öneme sahiptir. DDİ, doğal dillerin kullanıldığı ve doğal dillerin işlenmesini gerektiren bütün çalışmalarda yer alabilmektedir. Yazım hatası düzeltimi (Polat vd., 2019), soru-cevap sistemleri (Çetiner vd., 2021), metin özetleme (Uzundere vd., 2008), metin sınıflandırma (Yıldırım ve Yıldız, 2018) ve daha farklı konulardaki çalışmalarda (Öztürk vd., 2020; Kul, 2020) DDİ'nin kullanıldığına rastlamak mümkündür.

Özellikle çalışmalarda kullanılan veri kümelerinin büyüklüğü göz önünde bulundurulduğunda işlemlerin elle yapılması yerine DDİ'yle yapılması daha makul ve gerekli bir yöntem gibi görülmektedir. Zira birkaç sayfalık bir metinde belirli bir harf veya kelimenin kullanım sıklığını tespit etmek basit bir işlem olarak kabul edilebilir. Fakat daha fazla sayfadan oluşan bir metinde alfabedeki tüm harflerin kullanım sıklıklarının, konumlarının ve birbirlerini takip etme durumlarının belirlenmesi, buna ek olarak metni oluşturan kelimelerin hecelenmesi, hecelerin türlerinin ve bu türlerin kelime içindeki konumlarının saptanması, hecelerde geçen harflerin tespit edilmesi ve dahası birçok işlemin gerçekleştirilmesi için DDİ'nin kullanılması bir zorunluluk halini almıştır.

Sıklık analizi çalışmalarında harf ve hecelerin tüm yönleriyle ele alınmamış olması sebebiyle DDİ'ye çok ihtiyaç duyulmadığı düşünülebilir. Ayrıca, dil bilimin nicel dalını oluşturan sıklık çalışmalarında (Dolunay, 2012) genelde kelime, kısmen de deyim ve atasözlerinin frekans sayımlarının yapıldığı görülmektedir (Karadağ, 2020). Fakat Türkçe ve dil bilgisi öğretimi ile dil bilimi açısından ele alınması gereken özellikler bunlardan ibaret değildir. Örneğin metinlerde harflerin geçiş frekansları ve birbirlerini takip etme durumları, hece tiplerinin frekansları ve metinlerin kelime uzunlukları gibi özellikleri konu edinen çalışmalardan Türkçe anlama ve anlatma becerilerinin geliştirilmesinin yanı sıra dil bilgisi öğretiminde faydalanılabilir (Sezgin, 1993: 100-101).

Ferah (2011), Türkçe Dersi Öğretim Programı'nı ses ve hece sıklığı açısından değerlendirmiştir. Çalışmasında Türkçede ünlü harflerin kullanım sıklığının "a, i, e, ı, u, ü, ö" sırasında olmasına karşın öğretim programına göre öncelikle "e" harfinin öğretildiği konusunda eleştirel bir bakış ortaya koyarak bunun ilk okuma-yazma öğretiminde zorlaştırıcı etkiye sahip olduğunu ifade etmiştir. Ayrıca, ilk okuma-yazma öğretimine "a" harfiyle başlanmasının, hece oluşturmanın revize edilerek Türkçenin yapısına uygun olacak şekilde dönüştürülmesinin ve ses temelli ilk okuma-yazma öğretiminde harflerin yazım

kolaylığının yanı sıra Türkçe harf-sözcük frekanslarının temel alınmasının en doğru yol olacağı ifade edilmiştir. Özcan ve Özcan (2014) tarafından, çocukların erken ses edinimi ile ilgili yapılan araştırmaları incelemeleri neticesinde “k, t, m, p, b, ç, c” seslerinin en erken edinilen sesler oldukları, söyleyiş açısından kolay olan bu seslerin öğrenilmesiyle Türkçedeki kullanım sıklıklarının ilişkisi olduğu belirtilmiştir. Bununla beraber, çocukların büyük çoğunluğunun ses çözümlemesini yapabilmeleri ve harfleri çizebilecek durumda olmaları sebebiyle ilk okuma-yazmadaki esas problemin harf ve seslerin örtüşürülerek zihinsel eşleştirmesinin yapılması sorunu olduğu, seslerin çıkarılma kolaylığıyla kullanım sıklıklarının ilişkili olması nedeniyle sık kullanılan sözcüklere yer vermenin kolay ve kalıcı öğrenmenin sağlanmasına fayda sağlayacağı vurgulanmıştır. Balantekin (2020), okuma-yazma sürecinde zorluk yaşayan öğrencilerin özellikleri ve süreçte yaşadıkları zorlukları belirlemek amacıyla bir çalışma yürütmüştür. Okuma-yazma sürecinde zorlanan öğrencilerin %23.03’ünün sesleri öğrenme aşamasında ve yaklaşık yarısının seslerden hece oluşturma sürecinde zorlandıkları, yaklaşık %61.99’unun 2. sınıfta okuma-yazmayı öğrendikleri, %71.73’ünün Eylül ve Ekim aylarında zorlandıklarının fark edildiği, %81.37’sinin diğer öğrencilere göre daha küçük yaşta okula başladıkları ve okul öncesi eğitim alma oranının düşük olduğu tespit edilmiştir. Okuma-yazma sürecinde zorlanan öğrencilerin yaşadıkları zorlukları azaltabilmek adına okul öncesi eğitimi almış öğrencilerin okullarında, almamışların ise ilkokula kayıt döneminde rehber öğretmenler tarafından okul olgunluğu tespitlerinin yapılması ve buna göre ilkokula kayıt yaptırılıp/yaptırılmayacağıyla ilgili karar alınması tavsiye edilmiştir. Bununla beraber, öğrencilerin ses ve seslerden hece oluşturmada daha fazla zorlanmaları sebebiyle tüm öğrencilerin okul öncesi eğitimden faydalandırılarak sesbilgisel farkındalığın artırılabilmesini ve böylece yaşanması muhtemel zorlukların en aza indirilebileceği önerisinde bulunulmuştur.

Türkçedeki kullanım özelliklerini ortaya koymak adına yürütülen araştırmalarda da sıklık verilerinin kullanıldığı ve metinlerin sıklık analizlerinin yapıldığı görülmüştür. Örneğin; Karaca ve Bayır (2020), Türkçenin genel kullanım durumunu sunmak adına ilkokul öğrencilerine yönelik hazırlanmış 5 yayınevinden 5 Türk ve 5 dünya edebiyatından eserin sıklık analizini gerçekleştirmişlerdir. Harflerden en sık kullanılan 10 tanesi “a, e, i, n, r, l, ı, k, d, m”, en az kullanılan 5 tanesi ise “j, f, p, ö, v” sıralamasındadır. Yayınevlerinde en sık kullanılan ilk yedi harf ve en az kullanılan ilk 3 harf aynıdır. En uzun kelimenin 24 harfli, kelimelerin ortalama 5.97 harfli ve 2.55 heceli oldukları, hecelerin ortalama 2.34 harfli ve en sık kullanılan hece tipinin ünsüz-ünlü birleşimli olduğu ortaya konmuştur. Türkçe metinler kullanılarak yapılan çalışmalardan bir diğerinde ise Karadağ (2020), iki farklı veri kümesine ait kelimelerdeki harflerin frekanslarını sayma işlemi gerçekleştirmiştir. Harf sıklıklarının belirlenmesi, Microsoft Word programında bul-değiştir yöntemiyle sağlanmıştır. İki farklı veri kümesinden elde edilen sonuçlar farklılık göstermiştir. En sık kullanılan 10 harf sırasıyla 1. veri kümesinde “a, e, k, l, i, m, r, ı, n, t” iken 2. veri kümesinde “a, e, i, n, r, l, ı, k, d, t”; en az kullanılan 5 harf sırasıyla 1. veri kümesinde “j, f, ö, ğ, h”, 2. veri kümesinde “j, f, ö, h, p”dir. Özmen (2011), 10 eserle gerçekleştirdiği sıklık analizi çalışmasında ünlü harflerin “a, e, i, ı, u, o, ü, ö”, ünsüzlerin ise “r, n, l, d, k, m, t, y, s, b, ş, z, h, g, ç, v, ğ, c, p, f, j” sırasında kullanıldığını ortaya koymuştur.

Özetle; Türkçe metinleri bazı özellikleri bakımından inceleyen ve ilk okuma-yazma öğretiminde dikkat edilmesi gerekenleri araştıran çalışmalar olsa dahi ilkokul Türkçe ders kitaplarının bu çalışma kapsamında ele alınan boyutlarıyla incelendiği ve kitapların mevcut durumuna ait sıklıkların analiz edildiği bir çalışmaya rastlanamamıştır.

Bazı çalışmalarda Türkçenin harf sıklığına dikkat edilmeden öğretim programı oluşturulduğu ve bunun pek de doğru olmadığı ifade edilmektedir (Ferah, 2011; Karadağ, 2020). Eleştirel bakış ortaya koyan çalışmalarda özellikle harflerin ilk öğretiminde kullanılan sırasal düzenin Türkçedeki genel harf kullanım sıklığı düzeniyle eşleşmediği yönünde değerlendirmeler vardır. Var olan durumu betimlemek üzere gerçekleştirilen bu çalışmada ilkokul kademesinde ve özellikle 1. sınıftaki Türkçe ders kitaplarında yer alan okuma metinlerinin kelime, harf, hece ve bunlara ilişkin sıklıkların analiz edilmesi amaçlanmıştır. Bu yolla özelde Türkçe ders kitaplarına genelde ise Türkçenin genel kullanım durumuna ilişkin bilgilerin sunulması, öğretim programlarının düzenlenmesine kaynak oluşturması, Türkçeyle ilgili fonetik düzeyde çalışma yapacak araştırmacılara açıklayıcı bilgiler sunması amaçlanmaktadır.

Harflerin öğretim sırası özellikle 1. sınıfla alakalı bir konuyken diğer sınıf seviyelerinde durumun ne olduğu ve 1. sınıf ile diğer sınıf seviyelerindeki Türkçe kitaplarında farklılık var mıdır soruları yanıt bulmuş değildir. Bu sebeple ilkokul 1. sınıf seviyesinde 3 yayınevine, 2, 3 ve 4. sınıf seviyesinde ise 1'er yayınevine ait Türkçe ders kitapları karşılaştırmalı olarak ele alınmıştır. Bu çalışmayla sınıf ve yayınevleri ayrı ayrı değerlendirilip sunulmuş, aşağıdaki konularda sıklıkla birlikte oranları da incelenerek sınıf seviyeleri ve yayınevleri bakımından farklılık ve benzerliklerin belirlenmesine çalışılmıştır;

- Metinler kaç kelimededen oluşmaktadır?
- Kelimeler kaç harflidir?
- En çok kullanılan kelimeler nelerdir?
- Kelimelerin ses uyum kurallarına uyma/uymama durumları nedir?
- Harflerin kullanım sayıları nedir?
- Özellikleri bakımından ünlü ve ünsüz harflere ilişkin veriler nelerdir?
- Harflerin kelime içinde kaçınıcı sırada kullanıldıklarını ilişkin veriler nelerdir?
- Harflerin kelime içinde birbirlerini takip etme durumları nedir?
- Kelimelerin hece sayılarına ait veriler nelerdir?
- Hece tiplerinin kullanım sayıları nedir?
- Hece tiplerinin kelime içindeki sıraları nedir?
- En çok kullanılan heceler nelerdir?

Yöntem

Okullarda kullanılacak Türkçe ders kitapları, Talim Terbiye Kurulu kararlarıyla belirlenmektedir. Bu çalışma, kurul tarafından ders kitabı olarak kabul edilen ilkokul 1, 2, 3 ve 4. sınıf Türkçe ders kitaplarıyla gerçekleştirilmiştir. Temel nitel araştırma deseni ile yürütülen bu çalışmada, kitaplardaki okuma metinlerinin kelime, hece ve harf sıklıkları belirlenmiş, elde edilen sonuçlar değerlendirilmiş ve mevcut durum betimlenmeye çalışılmıştır. Betimsel çalışmalarda araştırma konusuyla ilgili var olan durum (Kaptan, 1989: 34), sıklık verileri aracılığıyla ortaya konulabilmektedir.

Veri Kümesi

İlkokul 1. sınıf ders kitaplarındaki her 3 yayınevinde de dinleme-okuma metinlerinin yer aldığı 4; 2, 3 ve 4. sınıf ders kitaplarında ise 8 tema yer almaktadır. Bütün sınıf seviyelerindeki ders kitaplarında bulunan her temada dinleme, okuma ve serbest okuma metinleri bulunmaktadır. Çalışmanın veri kümesini 1. sınıflar için 3 yayınevi, 2, 3 ve 4. sınıflar için 1 yayınevine ait hikâye edici ve bilgilendirici tüm metinler ile şiirlerden oluşan okuma ve serbest okuma metinleri oluşturmaktadır. Kullanılan metin sayılarının sınıf ve yayınevlerine göre dağılımları Tablo 1'de sunulmuştur. Çalışma kapsamında farklı sınıf seviyelerinden toplamda 144 metin analiz edilmiştir.

Bütün sınıf seviyelerinde hikâye edici ve bilgilendirici metinlerin şiirlerden daha fazla olduğu, hikâye edici ve bilgilendirici metinlerin tüm metinlere oranının yaklaşık %70 civarında olduğu görülmüştür. Tablo 1'de verilen 1. sınıf 1. yayınevi olan Ada Yayınları S1(Y1), 1. sınıf 2. yayınevi olan Cem Yayınları S1(Y2), 1. sınıf 3. yayınevi olan MEB Yayınları S1(Y3), 2. sınıf yayınevi olan Koza Yayınları S2, 3. sınıf yayınevi olan Sonuç Yayınları S3 ve 4. sınıf yayınevi olan Koza Yayınları ise S4 olarak etiketlenmiştir.

Tablo 1.

Sınıf, Yayınevleri ve Okuma Metni Sayılarına İlişkin Bilgiler

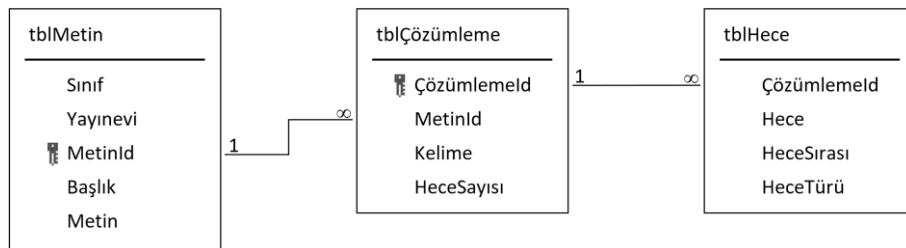
	S1(Y1)	S1(Y2)	S1(Y3)	S2	S3	S4	
1. Sınıf	16	16	16	-	-	-	48
2. Sınıf	-	-	-	32	-	-	32
3. Sınıf	-	-	-	-	32	-	32
4. Sınıf	-	-	-	-	-	32	32
	16	16	16	32	32	32	144

Verilerin İşlenmesi

Eldeki veri miktarının büyüklüğü ve çalışmada yapılması planlanan işlemleri gerçekleştirmek için gereken süre ve iş yükü göz önünde bulundurulduğunda çalışma kapsamında metinlerin dijital ortama aktarılmasından işlenmesine ve sonuçların elde edilmesine kadar olan yürütülecek tüm işlemlerin elle yapılması pek de mümkün görünmemektedir. Ayrıca, işlemleri elle yapmanın zaman alacağı, bazı olası hatalar ortaya çıkmasına ve birtakım işlemlerin gözden kaçmasına veya yapılmamasına sebebiyet verebileceği de muhtemeldir. Bu sebeple çalışmanın günümüz bilgisayar teknolojilerini de içine alacak şekilde yürütülmesi, belirtilen aksaklıkları gidermesi bakımından önem arz etmekte, gerekli olmaktadır. Bu hata ve aksaklıkları en aza indirmek veya ortadan kaldırmak adına işlemler, *Visual Studio* ortamında ve *C#* dilinde geliştirilen yazılımla gerçekleştirilmiştir. Geliştirilen yazılımda DDİ süreçlerinin bir kısmını gerçekleştirmek üzere Zemberek'ten (Zemberek, 2015) faydalanılmıştır. Zemberek, açık kaynak kodlu Türkçe DDİ Kütüphanesi olarak tanımlanabilir. Her ne kadar bu çalışmada yalnızca kelimelerin hecelenmesi işlemi Zemberek'ten faydalanılmış olsa da Zemberek'in yapabildikleri bununla sınırlı değildir. Örneğin kelimenin kök/gövdesinin, türünün ve aldığı eklerin belirlenmesi, bu eklerle ayrıştırmanın yapılması gibi daha birçok özelliği bulunmaktadır.

Çalışma genel anlamda 4 aşamadan oluşmaktadır. 1. aşamada çalışmada kullanılacak yayınevlerine ait Türkçe ders kitapları belirlenmiş ve kitapların dijital formlarına ulaşılmıştır. 2. aşamada okuma metinlerinin bulunduğu sayfalar OCR (Optical Character Recognition-Optik Karakter Tanıma) programı ile dijital ortama aktarılmıştır. Metinsel verilerdeki sayfa numarası, üst bilgi ve yazar bilgisi gibi birtakım veriler içeriklerden kaldırılmış, başlık ve metinlerden oluşan salt metinler elde edilmiştir. 3. aşamada salt metinler, veri tabanına kaydedilmiş ve veri tabanında kayıtlı metinler, işlemek üzere yazılım tarafından kelimelere parçalanmıştır. Kelimelerin denetlenmesi gerçekleştirilmiş ve denetleme sonucu başarılı olanlar Zemberek yardımıyla hecelerine ayrılmıştır. 4. aşamada ise geliştirilen yazılımla okuma metinlerinde en çok kullanılan kelimeler, kelimelerin ses uyum kurallarına göre durumları, kelimelerin harf sayısı, harflerin kullanım sayıları, harfin kaçınıcı harf olduğu, harflerin birbirlerini takip etme durumları, hece sayısı, hecenin kaçınıcı hece olduğu, hece tipi, hece tiplerinin kaçınıcı hece olduğuna ilişkin bilgiler elde edilmiştir. Bu işlemler çalışmaya dahil edilen farklı sınıf seviyesi ve yayınevlerine ait ders kitapları için uygulanmıştır.

Çalışmada kullanılan ilişkisel veri tabanının yapısı, veri tabanında kullanılan tablolar, tablolardaki alanlar ve tablolar arası ilişkiler Şekil 1'de verilmiştir. Şekle göre sınıf, yayınevi, başlık ve metin gibi okuma metinlerine ait genel veriler "tblMetin" tablosunda tutulmaktadır. "tblMetin" tablosunda "Metin" alanında tutulan okuma metinleri parçalara ayrılarak kelime kelime işleme alınmakta ve sonuçlar "tblÇözümleme" tablosuna kaydedilmektedir. "tblÇözümleme" tablosu, çözümleme numarasını, metin numarasını, kelimeyi ve kelimenin hece sayısını tutar. "tblHece" tablosunda ise kelimenin heceleri, hecenin sırası ve hece türüne ait veriler yer almaktadır. Bir metne ait kelimeler "tblÇözümleme" tablosunda, bir kelimeye ait heceler ise "tblHece" tablosunda ayrı kayıtlar şeklinde tutulmaktadır. "tblMetin" tablosundaki "MetinId" ve "tblÇözümleme" tablosundaki "Çözümlemeld" alanları birincil anahtar (PK) olarak tanımlanmış ve tablolar arası ilişki bu alanlar üzerinden sağlanmıştır. "tblMetin" tablosu ile "tblÇözümleme" tablosu "MetinId" alanı üzerinden, "tblÇözümleme" tablosuyla "tblHece" tablosu ise "Çözümlemeld" alanı üzerinden 1-∞ (1'e sonsuz) olarak ilişkilendirilmiştir. Çalışma kapsamında elde edilen tüm sonuçlar, ilişkisel veri tabanı üzerinde çalıştırılan SQL (Structured Query Language) ifadeleriyle sağlanmıştır.



Şekil 1. Veri Tabanında Bulunan Tablolar ve Alanları ile Tablolar Arası İlişkiler

İlkokul Türkçe Dersi Okuma Metinlerindeki Kelime, Harf ve Hece Sıklıklarına Karşılaştırmalı Bir Bakış

İşlemlere ait sonuçların veri tabanında nasıl temsil edildiği, 1 numaralı yayınevinde olduğu kabul edilen 1. sınıftaki “Ben ödevimi yaptım.” ve 2. sınıftaki “Akşam abinle bize gelebilirsin.” metinleri üzerinden Tablo 2’de örneklendirilmiştir. Metinler, “tblMetin” tablosuna sınıf, yayınevi, metnin kayıt numarası, başlık ve metin şeklinde tutulmaktadır. Bundan sonraki kısımlarda metin başlık ve metnin birleşimini ifade edecektir. Metinler kelimelere parçalanarak “tblÇözümle” tablosunda çözümleme kayıt numarası olan “Çözümlemeld”, hangi metne ait olduğunu belirten “MetinId”, kelime ve hece sayısı şeklinde tutulmaktayken “tblHece” tablosunda hecenin hangi çözümlemeye ait olduğunu gösteren “Çözümlemeld”, hecesi, hecenin sırası ve hece türü bilgisi yer almaktadır.

Tablo 2.

Verilerin Veri Tabanı Tablolarında Tutulma Şekli

tblMetin Tablosu					tblHece Tablosu			
Sınıf	Yayınevi	MetinId	Başlık	Metin	Çözümlemeld	Hece	HeceSırası	HeceTürü
1	1	1	Okul	Ben ödevimi yaptım.	1	o	1	V
2	1	2	Ziyaret	Akşam abinle bize gelebilirsin.	1	kul	2	CVC
					2	ben	1	CVC
					3	ö	1	V
					3	de	2	CV
					3	vi	3	CV
					3	mi	4	CV
					4	yap	1	CVC
					4	tım	2	CVC
					5	zi	1	CV
					5	ya	2	CV
					5	ret	3	CVC
					6	ak	1	VC
					6	şam	2	CVC
					7	a	1	V
					7	bin	2	CVC
					7	le	3	CV
					8	bi	1	CV
					8	ze	2	CV
					9	ge	1	CV
					9	le	2	CV
					9	bi	3	CV
					9	lir	4	CVC
					9	sin	5	CVC

tblÇözümle Tablosu			
Çözümlemeld	MetinId	Kelime	HeceSayısı
1	1	okul	2
2	1	ben	1
3	1	ödevimi	4
4	1	yaptım	2
5	2	ziyaret	3
6	2	akşam	2
7	2	abinle	3
8	2	bize	2
9	2	gelebilirsin	5

Araştırma ve Yayın Etiği

Bu çalışmada “Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesi” kapsamında uyulması belirtilen tüm kurallara uyulmuştur. Yönergenin ikinci bölümü olan “Bilimsel Araştırma ve Yayın Etiğine Aykırı Eylemler” başlığı altında belirtilen eylemlerden hiçbiri gerçekleştirilmemiştir.

Bulgular

Kelimelerin Zemberek’le çözümlenmesi durumlarına göre elde edilen sonuçlar Tablo 3’te verilmiştir. Ayrık olarak belirlenen kök/gövde sayısında 1. sınıf yayınevlerinde fark fazla olmasa dahi sınıf seviyesi arttıkça farkın dikkate değer şekilde arttığı gözlenmiştir. 1. sınıflarda ortalama 708 olan ayrık (benzersiz) kök/gövde sayısı 2. sınıfta 1274’e, 3. sınıfta 1473’e ve en yüksek olarak da 4. sınıfta 1899’a ulaşmıştır. Buradan da anlaşılacağı üzere sınıf seviyesi arttıkça buna paralel olarak metinlerin

farklı kelimelerle oluşturulma sayısı da artmıştır. Sonuçlar toplam kelime sayısı bakımından ele alındığında da ayırık kelime kök/gövde sayısındaki benzer sonuçlar elde edilmiştir. Toplam kelime sayısı 1. sınıflarda ortalama 2088 iken sınıf seviyesi arttıkça bu değer de artış göstermiştir. 189.07 olan okuma metinlerinin ortalama kelime sayısı sınıf ve yayınevlerine göre şu şekildedir. S1(Y1)'de 138.00, S1(Y2)'de 120.44, S1(Y3)'te 127.81, S2'de 153.41, S3'te 195.00 ve S4'te 309.28'dir.

27646 kelimedenden 420 tanesi Zemberek tarafından çözümlenememiş, çözümlenmesi gerçekleştirilen toplamda 27226 kelimeyle çalışma gerçekleştirilmiştir. “edison, jackson, yaşasın, gideceğüz, serunluğunda, cırrrt, hımmm, eee, ooo, şeey, şişşşt, yooo” gibi çözümlenemeyen kelimelerin ortak özellikleri özel isim, şive veya değişik formlarda metinlerde yazıma uygun olacak şekilde kelimelerin değiştirilmesi ve seslenme ifadesi olmaları şeklinde özetlenebilir.

Çözümlenemeyen kelimeler kapsam dışında tutulmuş ve çalışmanın bundan sonraki kısımlarına dahil edilmemiştir. S1(Y1)'de 2208, S1(Y2)'de 1927, S1(Y3)'te 2045, S2'de 4909, S3'te 6240 ve S4'te 9897 olmak üzere toplamda 27226 kelime ile çalışma yürütülmüştür.

Tablo 3.

Kelimelere İlişkin Sayılar

	S1(Y1)	S1(Y2)	S1(Y3)	S2	S3	S4	
Ayrık Kök veya Gövde Sayısı	741	659	723	1274	1473	1899	6769
Toplam Kelime Sayısı	2222	1944	2098	4976	6292	10114	27646
Çözümlenemeyen Kelime Sayısı	14	17	53	67	52	217	420
Çalışmanın Gerçekleştirildiği Kelime Sayısı	2208	1927	2045	4909	6240	9897	27226

Okuma metinlerini oluşturan kelimelerin harf sayısı bakımından incelenmesi neticesinde elde edilen sonuçlar Tablo 4'te verilmiştir. Tüm kelimelerin ortalama harf sayısı 6.14'tür ve yayınevleri arasında küçük farklılıklar vardır. Kelimelerin ortalama harf sayısının yayınevlerindeki değerleri S1(Y1)'de 6.13; S1(Y2)'de 6.06; S1(Y3)'de 6.02; S2'de 6.17; S3'te 6.27; S4'te 6.09 şeklinde seyretmiştir. Tabloya göre kelimelerin yaklaşık %63.50'si 3-7 harflidir. Tüm sınıf seviyeleri ve yayınevlerinde en fazla kelime 5 harfli gruptadır. 1 ve 12-21 harfli kelimelerin oranı yaklaşık %4.54'tür ve bu oran en düşük 1. sınıf, en fazla ise 3. sınıf okuma metinlerindedir. “o” 1 harfli, “kahramanlarımızdandır” ise 21 harfli kelimelere örnektir. Sonuçlar sınıf ve yayınevleri açısından değerlendirildiğinde sırasıyla en fazla ve en az kelime olan harf grupları şu şekildedir; S1(Y1)'de en fazla 5, 3 ve 4 harfli, en az 18, 20 ve 17 harfli; S1(Y2)'de 5, 7 ve 3 harfli, en az 17, 21, 1 ve 15 harfli, S1(Y3)'te 5, 3 ve 6 harfli, en az 1617 ve 15 harfli, S2'de 5, 6 ve 7 harfli, en az 18, 17 ve 16 harfli, S3'te 5, 7 ve 6 harfli, en az 19, 18 ve 17 harfli, S4'te 5, 4 ve 3 harfli, en az 18, 19 ve 20 harfli gruptadadır.

Sınıf ve yayınevlerine göre en çok geçen 10 kelime sırasıyla şu şekildedir; S1(Y1) bir (76), bu (30), çok (29), yavaş (21), zaman (20), daha (18), ve (18), ben (15), için (15), küçük (14); S1(Y2) bir (56), ve (28), çok (22), bu (18), gün (12), hacivat (12), karagöz (12), diye (11), küçük (11), minik (11); S1(Y3) bir (64), çok (39), ve (20), bu (17), ne (16), atatürk (14), daha (13), dedi (12), ama (9), güzel (9); S2 bir (159), ve (65), de (55), bu (49), çok (41), da (37), ne (28), için (25), diye (23), sonra (23); S3 bir (141), ve (75), bu (72), için (39), ne (39), sonra (39), çok (36), daha (31), her (31), karagöz (28); S4 bir (304), ve (123), bu (109), çok (79), o (76), ne (66), için (60), kadar (56), ama (42), ben (42).

Okuma metnini oluşturan kelimelerin büyük ünlü, küçük ünlü ve ünsüz uyumu bakımından ses uyum kontrolleri de yapılmıştır. Tüm sınıf seviyeleri ve yayınevleri birlikte değerlendirildiğinde kelimelerin yaklaşık %88.94'ünün büyük ünlü, %94.86'sinin küçük ünlü, %99.54'ünün ünsüz uyumuna uydukları görülmüştür. Sınıf ve yayınevleri arasında küçük farklılıklar olsa da kelimelerin %86.43'ünün hem büyük hem de küçük ünlü uyumuna uydukları, %5.14'ünün ne büyük ne de küçük ünlü uyumuna uydukları belirlenmiştir. Ses uyum kurallarının sınıf ve yayınevlerine göre incelenmesi neticesinde S1(Y1)'de kelimelerin yaklaşık %88.41'inin büyük ünlü, %94.84'ünün küçük ünlü, %99.37'sinin ünsüz uyumuna; S1(Y2)'de kelimelerin yaklaşık %88.69'unun büyük ünlü, %93.88'inin küçük ünlü, %99.53'ünün ünsüz uyumuna; S1(Y3)'de kelimelerin yaklaşık %86.45'inin büyük ünlü, %93.59'unun küçük ünlü, %99.66'sinin ünsüz uyumuna; S2'de kelimelerin yaklaşık %90.10'unun büyük ünlü,

İlkokul Türkçe Dersi Okuma Metinlerindeki Kelime, Harf ve Hece Sıklıklarına Karşılaştırmalı Bir Bakış

%94.36'sının küçük ünlü, %99.78'inin ünsüz uyumuna; S3'te kelimelerin yaklaşık %87.77'sinin büyük ünlü, %94.76'sının küçük ünlü, %99.42'sinin ünsüz uyumuna, S4'te kelimelerin yaklaşık %89.78'inin büyük ünlü, %95.62'sinin küçük ünlü, %99.52'sinin ünsüz uyumuna uydukları gözlenmiştir.

Tablo 4.

Harf Sayılarına Göre Kelime Sayı ve Oranları

	S1(Y1)		S1(Y2)		S1(Y3)		S2		S3		S4		f	%
	f	%	f	%	f	%	f	%	f	%	f	%		
1 harfli	10	0.45	5	0.26	7	0.34	16	0.33	23	0.37	77	0.78	138	0.51
2 harfli	130	5.89	113	5.86	136	6.65	328	6.68	448	7.18	731	7.39	1886	6.93
3 harfli	275	12.45	230	11.94	267	13.06	552	11.24	564	9.04	1120	11.32	3008	11.05
4 harfli	245	11.10	197	10.22	194	9.49	516	10.51	684	10.96	1146	11.58	2982	10.95
5 harfli	439	19.88	382	19.82	407	19.90	842	17.15	1073	17.20	1676	16.93	4819	17.70
6 harfli	227	10.28	227	11.78	244	11.93	627	12.77	780	12.50	1124	11.36	3229	11.86
7 harfli	229	10.37	271	14.06	236	11.54	588	11.98	815	13.06	1111	11.23	3250	11.94
8 harfli	213	9.65	173	8.98	180	8.80	495	10.08	633	10.14	961	9.71	2655	9.75
9 harfli	144	6.52	128	6.64	158	7.73	364	7.41	407	6.52	762	7.70	1963	7.21
10 harfli	131	5.93	80	4.15	86	4.21	241	4.91	312	5.00	532	5.38	1382	5.08
11 harfli	86	3.89	55	2.85	51	2.49	145	2.95	202	3.24	276	2.79	815	2.99
12 harfli	36	1.63	33	1.71	43	2.10	89	1.81	132	2.12	172	1.74	505	1.85
13 harfli	18	0.82	15	0.78	12	0.59	49	1.00	81	1.30	96	0.97	271	1.00
14 harfli	12	0.54	11	0.57	16	0.78	28	0.57	47	0.75	53	0.54	167	0.61
15 harfli	5	0.23	5	0.26	5	0.24	12	0.24	13	0.21	38	0.38	78	0.29
16 harfli	4	0.18	-	-	1	0.05	8	0.16	11	0.18	10	0.10	34	0.12
17 harfli	2	0.09	1	0.05	2	0.10	7	0.14	6	0.10	5	0.05	23	0.08
18 harfli	1	0.05	-	-	-	-	2	0.04	5	0.08	2	0.02	10	0.04
19 harfli	-	-	-	-	-	-	-	-	4	0.06	2	0.02	6	0.02
20 harfli	1	0.05	-	-	-	-	-	-	-	-	3	0.03	4	0.01
21 harfli	-	-	1	0.05	-	-	-	-	-	-	-	-	1	0.00

Harflerin kullanım sayısı ve oranları Tablo 5'te verilmiştir. Buna göre tüm sınıf ve yayınevlerinde "a" en çok, "j" ise en az kullanılan harflerdir. Tüm okuma metinleri birlikte değerlendirildiğinde sırasıyla "a, e, i, n, r" en çok, "j, f, ö, p, c" ise en az kullanılan harfler oldukları tespit edilmiştir. Sınıf ve yayınevlerine göre en çok ve az kullanılan harfler şunlardır; S1(Y1) "a, i, e, n, r" ve "j, f, ö, h, c"; S1(Y2) "a, e, i, n, r" ve "j, f, p, ğ, ö, v"; S1(Y3) "a, e, i, r, n" ve "j, f, v, ö, p, ğ"; S2 "a, e, i, r, n" ve "j, f, ö, h, c"; S3 "a, e, i, r, n" ve "j, f, ö, v, c"; S4 "a, e, i, n, r" ve "j, f, p, c, h".

Ünlü harflerin kullanım oranı yaklaşık %42.63, ünsüz harflerin ise %57.37'dir. Ünlü harfler kendi içinde incelendiğinde kalın ünlü harflerin kullanım oranı %52.14 iken ince ünlülerin oranı %47.86'dır. Düz ünlülerin oranı %79.43 iken yuvarlak ünlülerin oranı %20.57'dir. Geniş ünlülerin oranı %56.92, dar ünlülerin oranı ise %43.08'dir. Ünsüz harfler kendi içinde değerlendirildiğinde yumuşak ünsüzlerin oranı %72.03, sert ünsüzlerin oranı %27.97'dir. Sürekli ünsüzlerin oranı %66.15 iken süreksiz ünsüzlerin oranı %33.85'tir. Dudak ünsüzlerin oranı %15.64, diş ünsüzlerin oranı %64.01, damak ünsüzlerin oranı %18.64 ve gırtlak ünsüzünün oranı %1.71'dir. 1. sınıf yayınevlerindeki okuma metinlerinde ünlü harflerden ince, düz ve dar; ünsüz harflerden ise yumuşak, sürekli ve diş ünsüzlerin kullanım oranınının 2, 3 ve 4. sınıf okuma metinlerine göre nispeten daha düşük kullanımda olduğu gözlenmiş ve bunlar dışındaki kullanım oranlarında sınıf seviyeleri ve yayınevleri arasında büyük ölçüde benzerlik görülmüştür.

Her ne kadar sonuçlar ayrı bir tablo olarak sunulmamış olsa da harflerin kelime içinde kaçınıcı olarak yer aldıklarına ait bilgiler çalışma kapsamında ele alınmıştır. Tüm kelimelerin yaklaşık %13.10'u "b", 8.81'i "d" ve %8.58'i "k" harfiyle başlamış; %14.90'ı "n", %13.02'si "r" ve %11.73'ü "e" harfiyle

sonlanmıştır. Hiçbir kelime “ğ, j” harfleriyle başlamamış, “c, g, ö” harfleriyle sonlanmamış ve “b, d” harfleriyle 1 kez, “j” harfiyleyse 2 kez sonlanmıştır. 1-10 harfli olan kelimelerin tüm kelime sayısına oranı yaklaşık %92.84’tür. Bu sebeple ağırlığı oluşturan kelimelerin ilk 10 harfine odaklanılmıştır. Yapılan incelemeler neticesinde ortaya çıkan bulgulara göre kelimelerin ilk 10 harfinde yer alan harflere ilişkin olarak sırasında en çok geçen 1. harfin %13.01’inin “b”, 2. harfin %20.42’sinin “a”, 3. harfin %15.83’ünün “r”, 4. harfin %16.18’inin “a”, 5. harfin %11.64’ünün “n”, 6. harfin %11.30’unun “l”, 7. harfin %13.72’sinin “a”, 8. harfin %15.47’sinin “r”, 9. harfin %13.13’ünün “i”, 10. harfin %18.14’ünün ise “n” olduğu belirlenmiştir.

Harflerin konumlarının yanı sıra kelime içinde birbirlerini takip durumları da analiz edilmiştir. Sonuçlar için 841 elemanlı 29x29’luk bir matris oluşturulmuştur. Harf ikilileri takip eden ve edilen olarak nitelendirilmiştir. Örneğin; “ab” ikilisinde “a” takip eden, “b” takip edilendir. Okuma metinlerinde birbirini en çok takip eden 10 harf ikilisi sırasıyla “a, r”, “e, r”, “l, a”, “i, n”, “d, e”, “b, i”, “i, n”, “y, a”, “m, a” ve “k, a” olarak saptanmıştır. En çok takip eden harflerin sırasıyla “a, e, l, i, d”; en az takip edenlerin “j, f, ö, p, ş” oldukları belirlenmiştir. En çok takip edilen harfler sırasıyla “r, a, e, n, i” iken en az takip edilen harflerin “j, g, f, c, h” oldukları tespit edilmiştir.

Tablo 5.

Harflerin Kullanım Sayı ve Oranları

	S1(Y1)		S1(Y2)		S1(Y3)		S2		S3		S4		f	%
	f	%	f	%	f	%	f	%	f	%	f	%		
a	1597	11.80	1394	11.94	1504	12.22	3474	11.47	4426	11.32	6826	11.32	19221	11.50
b	391	2.89	355	3.04	375	3.05	753	2.49	946	2.42	1672	2.77	4492	2.69
c	130	0.96	132	1.13	132	1.07	287	0.95	381	0.97	566	0.94	1628	0.97
ç	237	1.75	163	1.40	161	1.31	393	1.30	450	1.15	790	1.31	2194	1.31
d	576	4.26	493	4.22	578	4.69	1346	4.45	1666	4.26	2735	4.54	7394	4.42
e	1170	8.64	1083	9.27	1047	8.50	2847	9.40	3789	9.69	5514	9.15	15450	9.24
f	44	0.33	55	0.47	33	0.27	136	0.45	154	0.39	212	0.35	634	0.38
g	187	1.38	213	1.82	176	1.43	434	1.43	557	1.42	790	1.31	2357	1.41
ğ	131	0.97	87	0.75	121	0.98	321	1.06	402	1.03	671	1.11	1733	1.04
h	117	0.86	120	1.03	131	1.06	265	0.88	428	1.09	582	0.97	1643	0.98
ı	660	4.88	478	4.09	601	4.88	1492	4.93	1875	4.80	2995	4.97	8101	4.85
i	1181	8.72	946	8.10	961	7.81	2400	7.93	3327	8.51	5024	8.33	13839	8.28
j	15	0.11	1	0.01	1	0.01	9	0.03	5	0.01	8	0.01	39	0.02
k	643	4.75	526	4.50	632	5.13	1427	4.71	1799	4.60	2731	4.53	7758	4.64
l	771	5.70	647	5.54	688	5.59	1814	5.99	2328	5.95	3700	6.14	9948	5.95
m	579	4.28	539	4.62	538	4.37	1266	4.18	1526	3.90	2217	3.68	6665	3.99
n	943	6.97	843	7.22	819	6.65	2161	7.14	2751	7.04	4425	7.34	11942	7.14
o	398	2.94	323	2.77	384	3.12	791	2.61	962	2.46	1560	2.59	4418	2.64
ö	101	0.75	103	0.88	99	0.80	251	0.83	319	0.82	604	1.00	1477	0.88
p	154	1.14	74	0.63	113	0.92	326	1.08	408	1.04	469	0.78	1544	0.92
r	905	6.69	814	6.97	850	6.90	2262	7.47	2841	7.27	4249	7.05	11921	7.13
s	351	2.59	358	3.07	303	2.46	790	2.61	1157	2.96	1762	2.92	4721	2.82
ş	314	2.32	256	2.19	227	1.84	570	1.88	689	1.76	1173	1.95	3229	1.93
t	390	2.88	333	2.85	402	3.27	957	3.16	1257	3.21	1766	2.93	5105	3.05
u	389	2.87	347	2.97	459	3.73	1010	3.34	1291	3.30	1925	3.19	5421	3.24
ü	282	2.08	304	2.60	212	1.72	566	1.87	702	1.80	1278	2.12	3344	2.00
v	156	1.15	121	1.04	98	0.80	320	1.06	371	0.95	605	1.00	1671	1.00
y	495	3.66	371	3.18	459	3.73	1023	3.38	1398	3.58	2282	3.79	6028	3.61
z	229	1.69	198	1.70	207	1.68	590	1.95	895	2.29	1158	1.92	3277	1.96

İlkokul Türkçe Dersi Okuma Metinlerindeki Kelime, Harf ve Hece Sıklıklarına Karşılaştırmalı Bir Bakış

Hece sayısına göre kelime sayı ve oranları Tablo 6'da sunulmuştur. Tabloya göre tüm kelimelerin yaklaşık %93.46'sı, S1(Y1)'de %92.71'i, S1(Y2)'de %94.66'sı, S1(Y3)'te 93.88'i, S2'de %94.00'i, S3'te %92.45'i ve S4'te %93.69'u 1-4 hecelidir. Tüm sınıf seviyeleri ve yayınevlerinde en fazla kelime 2 heceli, ardından 3 heceli, ardından 1 heceli ve ardından 4 heceli gruptadır. 6-9 heceli kelimelerin oranı oldukça düşük seviyede olup yaklaşık %1.71'dir ve bu oran en düşük 1. sınıf, en fazla ise 3. sınıf okuma metinlerindedir. 9 heceli kelime yalnızca S4 yayınevinde görülmüştür. Tüm sınıf seviyeleri ve yayınevleri birlikte değerlendirildiğinde toplam hece sayısının 71271, kelimelerin ortalama hece sayısının 2.62 olduğu belirlenmiştir. Bu veriler yayınevlerinde şu şekildedir; S1(Y1)'de 5778'e 2.62; S1(Y2)'de 4978'e 2.58; S1(Y3)'de 5267'ye 2.58; S2'de 12831'e 2.61; S3'te 16691'e 2.67; S4'te ise 25726'ya 2.60'tır. "bir, yıl, fark" 1 heceli; "a, kaç, lan, dı, ra, cak, la, rı, na, dü, zel, ti, le, me, ye, ce, ği, ni" ise okuma metinlerinde geçen 9 heceli kelimelere örneklerdir. "o, bu, çok, ilk, renk" gibi toplamda 4632 kelimenin hecesi ile aynı olduğu tespit edilmiştir.

Tablo 6.

Hece Sayısına Göre Kelime Sayı ve Oranları

	S1(Y1)		S1(Y2)		S1(Y3)		S2		S3		S4		f	%
	f	%	f	%	f	%	f	%	f	%	f	%		
1 heceli	387	17.53	302	15.67	369	18.04	826	16.83	949	15.21	1791	18.10	4624	16.98
2 heceli	778	35.24	685	35.55	717	35.06	1674	34.10	2126	34.07	3362	33.97	9342	34.31
3 heceli	549	24.86	584	30.31	535	26.16	1367	27.85	1814	29.07	2592	26.19	7441	27.33
4 heceli	333	15.08	253	13.13	299	14.62	747	15.22	880	14.10	1527	15.43	4039	14.84
5 heceli	120	5.43	79	4.10	96	4.69	223	4.54	333	5.34	463	4.68	1314	4.83
6 heceli	33	1.49	22	1.14	20	0.98	56	1.14	109	1.75	130	1.31	370	1.36
7 heceli	6	0.27	1	0.05	9	0.44	11	0.22	23	0.37	26	0.26	76	0.28
8 heceli	2	0.09	1	0.05	-	-	5	0.10	6	0.10	4	0.04	18	0.07
9 heceli	-	-	-	-	-	-	-	-	-	-	2	0.02	2	0.01

Hece gösteriminde genellikle ünlü harf için V, ünsüz harf için C kullanılmaktadır. Hece tipleri, hecede bulunan ünlü ve ünsüz harflerin kullanım durumlarına bağlı olarak V ve C ile gösterilmektedir. Örneğin; "bir" hecesi, ünsüz-ünlü-ünsüz birleşiminden oluşmakta ve tip olarak CVC ile temsil edilmektedir.

Türkçede 6 çeşit hece tipi bulunmaktadır ve bu hece tiplerine uymayan hece barındıran kelimeler Türkçe değildir. Örneğin; 6 hece tipi dışında metinlerde yer alan ve yabancı dillerden geçen kelimelerde CCV (gri), CCVC (tren), CCVCC (prens) ve CCCVC (stres) gibi heceleme türleri de kullanılmıştır. Esasında bunlar söyleyişte araya bir ünlü alırlar ve Türkçe değildirler; g(i)ri, t(i)ren, p(i)rens, s(i)tres gibi.

Hece tiplerine göre hece sayısı ve oranlarının verildiği Tablo 7'ye göre tüm hecelerin yaklaşık %51.86'sı 2 harfli hece olan CV, %37.93'ü 3 harfli hece olan CVC tipinde olmakla birlikte sınıf seviyeleri ve yayınevleri arasında farklılık düşük seviyededir. Öte yandan VCC ve CVCC tiplerindeki hece sayılarının kullanımı oldukça azdır. Açık hecelerin oranı ortalama yaklaşık %56.63, kapalı hecelerin oranı %42.32, tabloda diğer olarak ifade edilen Türkçe hece tipine uymayanların oranı ise %0.10 civarındadır. Türkçe hece tipine uymayan hecelere 1. sınıftaki S1(Y2) ve S1(Y3) yayınevlerinde hiç yer verilmemişken 2, 3 ve 4. sınıf yayınevlerinde az da olsa yer verilmiştir. Hece tiplerine okuma metinlerinden örnek olarak şunlar verilebilir; V tipinde "a, e, o", VC tipinde "ab, el, ok", CV tipinde "ba, la, te", CVC tipinde "cek, fen, tan", VCC tipinde "ant, ilk, ört", CVCC tipinde "dost, harf, taht", diğer tipinde ise "gru, kra, plas". Heceler harf sayısı bakımından incelendiğinde tüm sınıf ve yayınevlerinde 2.35 olan hecelerin ortalama harf sayısının yayınevlerindeki değerleri şu şekildedir; S1(Y1)'de 2.34; S1(Y2)'de 2.35; S1(Y3)'de 2.34; S2'de 2.36; S3'te 2.34; S4'te ise 2.34'tür. Bu değerler sınıf seviyesi ve yayınevine göre çok küçük oranda değişiklik göstermiştir.

Tablo 7.

Hece Tiplerine Göre Hece Sayı ve Oranları (V: Ünlü Harf, C: Ünsüz Harf)

	S1(Y1)		S1(Y2)		S1(Y3)		S2		S3		S4		f	%
	f	%	f	%	f	%	f	%	f	%	f	%		
V	275	4.76	249	5.00	276	5.24	611	4.76	767	4.60	1217	4.73	3395	4.76
VC	241	4.17	239	4.80	259	4.92	600	4.68	702	4.21	1265	4.92	3306	4.64
CV	3037	52.56	2543	51.08	2710	51.45	6480	50.50	8851	53.03	13344	51.87	36965	51.86
CVC	2175	37.64	1909	38.35	1988	37.74	5024	39.16	6219	37.26	9721	37.79	27036	37.93
VCC	4	0.07	15	0.30	3	0.06	25	0.19	31	0.19	18	0.07	96	0.13
CVCC	25	0.43	23	0.46	31	0.59	76	0.59	101	0.61	146	0.57	402	0.56
Diğer	21	0.36	-	-	-	-	15	0.12	20	0.12	15	0.06	71	0.10

Hece tiplerinin kaçınıcı hece olarak konumlandığına ilişkin bilgilerin sunulduğu Tablo 8-Tablo 11 incelendiğinde 1. sıradaki hecelerde V, VC, CV ve CVC hece tiplerinin toplam bulunma oranının %98.66, 2-8. sıradaki hecelerde CV ve CVC hece tiplerinin bulunma oranının (ağırlıklı olarak CV tipinde) %98.98 ile %100.00 aralığında, 9. hecede CV tipinin bulunma oranı %100.00 olduğu görülmüştür. Ayrıca, Türkçe kelimelerde bulunmayan diğer hece tipinin tamamının ilk hecede yer aldığı belirlenmiştir.

Harflerin hecelerde bulunduğu konumlar da irdelenmiştir. Tüm sınıf seviyeleri ve yayınevleri birlikte değerlendirildiğinde hecelerin yaklaşık %6.94'ünde "k", %6.51'inde "s", %6.41'inde "d", %6.12'sinde "r" ve %6.10'unda "t" 1. harf; %19.30'unda "a", %15.66'sında "e", %15.40'ında "i", %13.40'ında "ı" ve %10.71'inde "u" 2. harf; %15.80'inde "n", %15.49'unda "r", %9.34'ünde "k", %8.70'inde "m" ve %8.35'inde "l" 3. harf; %36.00'unda "t", %18.29'unda "k", %11.43'ünde "s", %9.71'inde "p" ve %8.00'unda "ç" harfi 4. harf olarak yer almıştır. Hecelerde bütün harfler 1. harf olarak kullanılmışken "g, j" 2. harf, "g, ı, ö, ü" 3. harf ve "a, c, d, e, ğ, ı, i, j, l, o, ö, u, ü, y, z" ise 4. harf olarak kullanılmamıştır. S1(Y1) yayınevine ait okuma metinlerindeki hecelerde 1. harf olarak en çok bulunanların sırasıyla "d, k, r, t, s" ve bütün harflerin hecenin ilk harfi olarak kullanıldığı, 2. harf olarak en çok bulunanların "a, e, i, ı, u" ve hiç bulunmayanların "b, c, d, f, g, j, p", 3. harf olarak en çok bulunanların "r, n, k, l, m" ve hiç bulunmayanların "c, g, ı, j, ö, u, ü", 4. harf olarak en çok bulunanların "t, p, k, m, f" ve hiç bulunmayanların "a, b, c, ç, d, e, g, ğ, h, ı, i, j, l, n, o, ö, r, ş, u, ü, v, y, z" oldukları gözlenmiştir. S1(Y2) yayınevi okuma metinlerindeki hecelerde 1. harf olarak en çok bulunanların sırasıyla "d, l, s, k, r" ve bütün harflerin hecenin ilk harfi olarak kullanıldığı, 2. harf olarak en çok bulunanların "a, i, e, ı, u" ve hiç bulunmayanların "c, d, f, g, h, j", 3. harf olarak en çok bulunanların "n, r, k, m, l" ve hiç bulunmayanların "a, c, d, e, g, ı, i, j, o, ö, u, ü", 4. harf olarak en çok bulunanların "s, ç, k, t, f" ve hiç bulunmayanların "a, b, c, d, e, g, ğ, h, ı, i, j, l, n, o, ö, r, ş, u, ü, v, y, z" oldukları gözlenmiştir. S1(Y3) okuma metinlerindeki hecelerde 1. harf olarak en çok bulunanların sırasıyla "l, s, t, k, y" ve bütün harflerin hecenin ilk harfi olarak kullanıldığı, 2. harf olarak en çok bulunanların "a, e, i, ı, u" ve hiç bulunmayanların "b, c, g, h, j", 3. harf olarak en çok bulunanların "n, r, k, m, l" ve hiç bulunmayanların "a, e, g, ı, i, j, o, ö, u, ü", 4. harf olarak en çok bulunanların "k, t, ç, p, s" ve hiç bulunmayanların "a, b, c, d, e, f, g, ğ, h, ı, i, j, l, m, n, o, ö, r, u, ü, v, y, z" oldukları gözlenmiştir. S2 yayınevine ait okuma metinlerindeki hecelerde 1. harf olarak en çok bulunanların sırasıyla "k, t, d, r, s" ve bütün harflerin hecenin ilk harfi olarak kullanıldığı, 2. harf olarak en çok bulunanların "a, e, i, ı, u" ve hiç bulunmayanların "g, j", 3. harf olarak en çok bulunanların "r, n, k, m, l" ve hiç bulunmayanların "c, g, ı, i, j, ö, ü", 4. harf olarak en çok bulunanların "t, k, n, s, p" ve hiç bulunmayanların "a, b, c, d, e, f, g, ğ, ı, i, j, l, o, ö, ş, u, ü, v, y, z" oldukları gözlenmiştir. S3 yayınevi okuma metinlerindeki hecelerde 1. harf olarak en çok bulunanların sırasıyla "s, k, r, d, n" ve bütün harflerin hecenin ilk harfi olarak kullanıldığı, 2. harf olarak en çok bulunanların "a, e, i, ı, u" ve hiç bulunmayanların "b, g, j", 3. harf olarak en çok bulunanların "n, r, k, m, l" ve hiç bulunmayanların "g, ı, i, j, ö, ü", 4. harf olarak en çok bulunanların "t, p, k, ç, s" ve hiç bulunmayanların "a, c, d, e, ğ, h, ı, i, j, l, o, ö, u, ü, v, y, z" oldukları gözlenmiştir. S4 okuma metinlerindeki hecelerde 1. harf olarak en çok bulunanların sırasıyla "k, s, d, r, l" ve bütün harflerin hecenin ilk harfi olarak kullanıldığı, 2. harf olarak en çok bulunanların "a, i, e, ı, u" ve hiç bulunmayanların "g, j", 3. harf olarak en çok bulunanların "r, n, k, l, y" ve hiç bulunmayanların "g, ı, ö,

İlkokul Türkçe Dersi Okuma Metinlerindeki Kelime, Harf ve Hece Sıklıklarına Karşılaştırmalı Bir Bakış

u, ü”, 4. harf olarak en çok bulunanların “t, p, k, ç, s” ve hiç bulunmayanların “a, c, d, e, ğ, h, ı, i, j, l, o, ö, u, ü, v, y, z” oldukları gözlenmiştir.

Tablo 8.

1. Sınıf (S1) Okuma Metinlerindeki Hece Tiplerinin Kelime İçindeki Konumuna Göre Sayı ve Oranları

	1. sırada		2. sırada		3. sırada		4. sırada		5. sırada		6. sırada		7. sırada		8. sırada	
	f	%	f	%	f	%	f	%	f	%	f	%	f	%	f	%
V	789	12.77	7	0.14	4	0.14	-	-	-	-	-	-	-	-	-	-
VC	726	11.75	12	0.23	1	0.03	-	-	-	-	-	-	-	-	-	-
CV	2619	42.38	2830	55.25	1757	59.72	762	59.81	253	65.04	57	60.64	11	57.89	1	33.33
CVC	1960	31.72	2259	44.10	1158	39.36	512	40.19	136	34.96	37	39.36	8	42.11	2	66.67
VCC	22	0.36	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CVCC	43	0.70	14	0.27	22	0.75	-	-	-	-	-	-	-	-	-	
Diğer	21	0.34	-	-	-	-	-	-	-	-	-	-	-	-	-	

Tablo 9.

2. Sınıf (S2) Okuma Metinlerindeki Hece Tiplerinin Kelime İçindeki Konumuna Göre Sayı ve Oranları

	1. sırada		2. sırada		3. sırada		4. sırada		5. sırada		6. sırada		7. sırada		8. sırada	
	f	%	f	%	f	%	f	%	f	%	f	%	f	%	f	%
V	603	12.28	4	0.10	4	0.17	-	-	-	-	-	-	-	-	-	-
VC	583	11.88	4	0.10	13	0.54	-	-	-	-	-	-	-	-	-	-
CV	2027	41.29	2222	54.42	1389	57.66	611	58.64	179	60.68	39	54.17	11	68.75	2	40.00
CVC	1595	32.49	1845	45.19	996	41.34	431	41.36	116	39.32	33	45.83	5	31.25	3	60.00
VCC	25	0.51	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CVCC	61	1.24	8	0.20	7	0.29	-	-	-	-	-	-	-	-	-	
Diğer	15	0.31	-	-	-	-	-	-	-	-	-	-	-	-	-	

Tablo 10.

3. Sınıf (S3) Okuma Metinlerindeki Hece Tiplerinin Kelime İçindeki Konumuna Göre Sayı ve Oranları

	1. sırada		2. sırada		3. sırada		4. sırada		5. sırada		6. sırada		7. sırada		8. sırada	
	f	%	f	%	f	%	f	%	f	%	f	%	f	%	f	%
V	758	12.15	5	0.09	4	0.13	-	-	-	-	-	-	-	-	-	-
VC	688	11.03	9	0.17	5	0.16	-	-	-	-	-	-	-	-	-	-
CV	2768	44.36	3045	57.55	1842	58.20	805	59.59	300	63.69	76	55.07	15	51.72	-	-
CVC	1901	30.46	2209	41.75	1310	41.39	546	40.41	171	36.31	62	44.93	14	48.28	6	100.00
VCC	31	0.50	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CVCC	74	1.19	23	0.43	4	0.13	-	-	-	-	-	-	-	-	-	
Diğer	20	0.32	-	-	-	-	-	-	-	-	-	-	-	-	-	

Tablo 11.

4. Sınıf (S4) Okuma Metinlerindeki Hece Tiplerinin Kelime İçindeki Konumuna Göre Sayı ve Oranları

	1. sırada		2. sırada		3. sırada		4. sırada		5. sırada		6. sırada		7. sırada		8. sırada		9. sırada	
	f	%	f	%	f	%	f	%	f	%	f	%	f	%	f	%	f	%
V	1195	12.07	16	0.20	6	0.13	-	-	-	-	-	-	-	-	-	-	-	-
VC	1215	12.28	40	0.49	10	0.21	-	-	-	-	-	-	-	-	-	-	-	-
CV	4258	43.02	4509	55.63	2745	57.86	1331	61.88	376	60.16	99	61.11	18	56.25	5	83.33	2	100.00
CVC	3097	31.29	3508	43.28	1969	41.51	820	38.12	249	39.84	63	38.89	14	43.75	1	16.67	-	-
VCC	17	0.17	1	0.01	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CVCC	100	1.01	32	0.39	14	0.30	-	-	-	-	-	-	-	-	-	-	-	-
Diğer	15	0.15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Sınıf ve yayınevlerinde kullanılan ilk 10 hecenin sayı ve oranları Tablo 12’de verilmiştir. “la, le, de, da, ya, ma” heceleri tüm sınıf ve yayınevlerindeki okuma metinlerinde en çok geçen heceler iken “di” S1(Y1), “bir” ise S3 yayınevi dışında en çok geçen 10 heceden içindedir. İlk 10 hece itibariyle farklı sıralamalar olsa da genel itibariyle sınıf ve yayınevleri arasında benzerlik gözlenmiştir.

Tablo 12.

En Çok Kullanılan 10 Hecenin Kullanım Sayı ve Oranları

S1(Y1)			S1(Y2)			S1(Y3)			S2			S3			S4					
Hece	f	%	Hece	f	%	Hece	f	%	Hece	f	%	Hece	f	%	Hece	f	%			
la	149	2.58	la	115	2.31	de	127	2.41	la	290	2.26	la	410	2.46	la	594	2.31	la	1685	2.36
le	130	2.25	de	106	2.13	la	127	2.41	le	287	2.24	le	349	2.09	le	567	2.20	le	1517	2.13
de	105	1.82	da	77	1.55	le	112	2.13	de	232	1.81	ya	296	1.77	de	505	1.96	de	1370	1.92
ya	103	1.78	a	75	1.51	bir	85	1.61	da	205	1.60	de	295	1.77	da	470	1.83	da	1148	1.61
da	85	1.47	di	72	1.45	ma	83	1.58	di	203	1.58	di	255	1.53	ya	390	1.52	ya	1106	1.55
bir	82	1.42	le	72	1.45	ya	83	1.58	bir	185	1.44	da	235	1.41	di	364	1.41	di	1048	1.47
ka	82	1.42	ma	68	1.37	di	82	1.56	ka	176	1.37	ma	226	1.35	bir	352	1.37	ma	951	1.33
a	80	1.38	ra	65	1.31	da	76	1.44	ya	175	1.36	ka	220	1.32	o	351	1.36	bir	939	1.32
ma	80	1.38	bir	64	1.29	o	74	1.40	ma	173	1.35	ra	219	1.31	ma	321	1.25	o	865	1.21
ne	80	1.38	ya	59	1.19	ne	65	1.23	a	169	1.32	ye	217	1.30	lar	302	1.17	a	861	1.21

Tartışma ve Sonuç

Alan yazın incelemesinde Türkçe metinlerle yapılan sıklık analizi çalışmalarında genellikle kelime, kısmen de deyim ve atasözleri incelenmiştir. Buna karşın Türkçe ders kitaplarının sıklık analizlerine ilişkin bir çalışmaya rastlanmamıştır. Oysaki kelime, harf ve hecelerle ilgili yapılacak kapsamlı bir sıklık analizi çalışması sonucu elde edilecek veriler, ilk okuma-yazma ve Türkçe dil bilgisi öğretimine, Türkçe kitaplarının yazımına ve değerlendirilmesine, müfredat programlarının oluşturulmasına ve revize edilmesine kaynak olacak nitelikte bilgiler barındırıyor olabilir. Bununla birlikte dilbilim ve dil öğretimi alanında çalışma yapacak ve konuyu bütüncül bir yaklaşımla ele almak isteyen araştırmacılara da yol gösterici bilgiler sunabilir.

Bu çalışmada Talim Terbiye Kurulu kararlarıyla Türkçe ders kitabı olarak kullanılabilen kitaplardaki 144 okuma metninin sıklık analizleri gerçekleştirilmiştir. Veri kümesi olarak ilkökul 1. sınıftan 3; 2, 3 ve 4. sınıfların her birinden 1 yayınevine ait Türkçe ders kitapları kullanılmış ve sonuçlar karşılaştırmalı olarak ele alınmıştır. Çalışmayla metni oluşturan kelimeler, kelimelerin harf sayıları, en sık kullanılan kelimeler, kelimelerin ses uyum kurallarına uyup uymama durumları, harflerin kullanım sayıları, konumları (sıraları) ve birbirlerini takip etme durumları, kelimelerin hece sayıları, hecelerin harf sayıları, tipleri, konumları ve en çok kullanılan hecelere ait sıklıklar ortaya konmuş ve sonuçlar analiz edilmiştir. İşlemler, *Visual Studio* ortamında ve *C#* dilinde geliştirilen yazılımla gerçekleştirilmiştir. Geliştirilen yazılım DDİ süreci de barındırmaktadır. DDİ sürecinin kelimelerin hecelenmesi adımıyla Türkçe DDİ kütüphanesi olan Zemberek kullanılmıştır.

Çalışma, 144 okuma metni ve metinleri oluşturan toplam 27226 kelime ile gerçekleştirilmiştir. 189.07 olan okuma metinlerinin ortalama kelime sayısında sınıf seviyesindeki artışa paralel olarak kelime sayısında da artış gözlenmiştir. Türkçe ders kitaplarının ilk okuma metinleri daha az sayıda kelimedenden, kelimeler daha az sayıda harften oluşmaktayken okuma metinlerinde ilerledikçe bu değerlerde, özellikle 1. sınıf okuma metinlerinde ilerledikçe 11 ve üstü harfli kelime sayısında artış gözlenmiştir. Kelimeler ortalama 6.14 harflidir. Kelimeler ortalama harf sayısı itibariyle en az S1(Y3)’te, en fazla ise S3’tedir. Tüm sınıf seviyeleri ve yayınevlerinde en fazla kelimenin 5 harfli grupta yer aldığı, en çok kullanılan kelimelerin genelde aynı olduğu tespit edilmiştir. Metni oluşturan kelimelerin küçük ünlü ve ünsüz uyumuna büyük ünlü uyumundan daha fazla uydukları belirlenmiştir. Karaca ve Bayır’a (2020) göre kelimelerin ortalama harf sayısı 5.97, ağırlıklı olarak kelimeler 5 harfli grupta ve küçük ünlü ve ünsüz uyumuna uyan kelime sayısı büyük ünlü uyumuna uyan kelime sayısından fazladır.

Değişik nedenlerden kaynaklı Türkçede harfler aynı frekansta kullanılmazlar (Özmen, 2011). Bununla beraber harflerin Türkçedeki kullanım sayılarıyla başka bir dildeki kullanım sayıları da farklılık gösterebilir. Örneğin; Türkçede en sık kullanılan 5 harf “a, e, i, n, r” (Serengil ve Akin, 2011) iken İngilizcede “e, t, o, a, i”dir (Solso ve King, 1976). Tüm sınıf seviyeleri ve yayınevleri birlikte değerlendirildiğinde bu çalışmadaki harflerin kullanım sıklıklarına göre sıralaması “a, e, i, n, r, l, ı, k, d, m, y, u, t, s, b, o, ü, z, ş, g, ç, ğ, v, h, c, p, ö, f, j” şeklinde oluşmuştur. Sınıf seviyelerine ve yayınevlerine göre kullanım sıklıklarında küçük farklılaşmalar tespit edilmiştir. Ünlü harflerin kullanım oranı %42.58, ünsüz harflerin ise %57.42’dir. Kelimelerin ağırlıklı olarak “b, d, k” ile başladığı, “n, r, e” ile bittiği; “ğ, j” ile başlayan ve “c, g, ö” ile biten kelime olmadığı belirlenmiştir. Harflerin kelime içinde birbirini takip etme durumları bakımından birbirini en çok takip eden ikililerin “a, r”, “e, r”, “l, a”, “i, n”, “d, e”, “b, i”, “ı, n”, “y, a”, “m, a” ve “k, a” olduğu saptanmıştır. Karaca ve Bayır (2020) harflerin kullanım sıklıklarının sırasıyla “a, e, i, n, r, l, ı, k, d, m, y, u, b, s, t, o, ş, ü, z, g, ç, ğ, h, c, v, ö, p, f, j”, ünlü harflerin kullanım oranını %42.65, ünsüz harflerin kullanım oranını ise %57.35 olduğunu, kelimelerin en çok “b, k, d” ile başladığını ve “n, e, a” ile bittiğini, “ğ” ile başlayan ve “g” ile biten kelime olmadığını, birbirini en çok takip eden ikililerin “a, r”, “l, a” ve “e, r” olduklarını saptamışlardır.

Kelimelerin ortalama hece sayısı 2.62, hecelerinin ortalama harf sayısı 2.35’tir. Ortalama hece sayısı en az olan yayınevleri S1(Y2) ve S1(Y3)’de, en fazla olan ise S3’tür. Sınıf seviyeleri ve yayınevleri itibarıyla hecelerinin ortalama harf sayısı çok yakın, neredeyse aynı değerdedir. Bununla birlikte sınıf seviyeleri ve yayınevlerindeki ortalama hece sayısı farklılığı da düşük seviyededir. Sınıf ve yayınevlerinin başlangıcında yer alan okuma metinlerindeki kelimeler daha az sayıda heceden oluşmaktadır. 1. sınıf yayınevlerindeki ilerleyen metinlerde hece sayısı ve CVCC hece tipinin kullanımı artış göstermiştir. Kelimeler ağırlıklı olarak 2 ve 3 heceli olmakla birlikte 5 ve üstünde heceye sahip kelime sayılarının oranı düşüktür. Tüm sınıf ve yayınevlerinde hecelerinin yaklaşık %90’ı 2 harfli ünsüz-ünlü harf birleşimli CV ve 3 harfli ünsüz-ünlü-ünsüz harf birleşimli CVC tipindedir. S1(Y1), S2, S3 ve S4 yayınevlerinin okuma metinlerinde çok az da olsa “gri, kra, plas, tren” gibi Türkçe hece yapısına uymayan hecelere rastlanmıştır. Yalnızca 1 ünlüden oluşan V, ünlü-ünsüzden oluşan VC ve ünlü-ünsüz-ünsüzden oluşan VCC hece tiplerinin yaklaşık %98.21’inin, Türkçe kelimelerde bulunmayan diğer hece tipinin ise %100.00’ünün kelimelerin ilk hecesi olarak buldukları tespit edilmiştir. Hecelerin ilk harfinde en çok sırasıyla “k, s, d, r, t”, ikinci harfinde “a, e, i, ı, u”, üçüncü harfinde “n, r, k, m, l” ve dördüncü harfinde “t, k, s, p, ç” bulunmaktadır. Hecelerin ilk, ikinci, üçüncü ve dördüncü olarak en çok bulunan harflerinde farklılık gözlenmiştir. “la, le, de, da, ya, ma” tüm sınıf ve yayınevlerindeki okuma metinlerinde en çok kullanılan heceler olarak ortaya çıkmıştır. Bu çalışmada elde edilen sonuçlarla Karaca ve Bayır (2020) tarafından yapılan çalışmada elde edilen sonuçlar birbirlerine yakın değerlerde seyretmiştir; kelimelerin hece ortalamasının 2.55, hecelerinin ortalama harf sayısının 2.34 ve heceler genellikle 2 ve 3 heceli olduğu, CV ve CVC hece tiplerinin ağırlıklı olarak kullanıldıkları görülmüştür. Çalışma kapsamında ele alınanlar itibarıyla aynı sınıf seviyesindeki farklı yayınevlerinde birbirine yakın sonuçların alınması olağan bir durumdur. Fakat sınıf seviyeleri arasında da birçok benzerlik görülmesi, sonuçların aynı veya çok yakın seyretmesi dikkat çekici ve vurgulanması gereken bir noktadır. Türkçe ders kitaplarındaki okuma metinleri oluşturulurken metin uzunluğunda (kelime sayısı) sınıf seviyelerine dikkat edildiği görülmektedir. Buna karşın gerek kelimelerin harf ve hece sayısı gerekse hecelerinin harf sayısında sınıf seviyelerine dikkat edildiği söylenemez.

Bu çalışma kapsamında elde edilen veriler, ilk okuma-yazma öğretimi açısından da irdelenmiştir. Bazı araştırmacılar tarafından ilk okuma-yazma öğretiminde Türkçedeki harf sıklığına dikkat edilmediği ifade edilmiştir. Yapılan değerlendirmelere bakıldığında özellikle harflerin ilk öğretimindeki sırasal farklılığın ele alındığı; Türkçedeki harflerin genel kullanım sıklıklarıyla bunların öğretim sıralamasının eleştirel bir bakış açısıyla karşılaştırıldığı görülmektedir. Örneğin; ses/harf öğretimine “e” ile başlandığı fakat Türkçede daha fazla kullanılan “a” harfiyle başlamanın daha doğru olacağı, “e” harfinin en kolay el yazısı harfi olmasından kaynaklı olarak öğretime bu harfle başlandığı ifade edilmektedir (Ferah, 2011). Bu çalışmada elde edilen sıklıklara bakıldığında Türkçede “a” kullanımının “e” kullanımından 1. sınıflarda %35.33, 2. sınıflarda %21.07, 3. sınıflarda %15.81, 4. sınıflarda %23.10 ve genelde %23.55 daha fazla olduğu görülmüştür. Karaca ve Bayır (2020) tarafından farklı yayınevlerinde farklı değerler elde edildiği, “a” ile “e” kullanım sıklıkları arası farkın en düşük

%30.56, en yüksek %38.93 ve ortalama %33.76 olduğu ortaya konmuştur. Karadağ (2020) ise “a” harfinin 1. veri kümesinde %43.38, 2. veri kümesinde ise %29.04 olarak “e” harfinden daha fazla kullanıldığını belirtmiştir. Bu çalışmayla ortaya çıkarılan sonuçlarda Türkçede “a” harfinin kullanım sıklığının “e” harfinden daha fazla olduğunu doğrulamaktadır.

Türkçe dersi öğretim programlarında çeşitli zamanlarda güncellemeler yapıldığı görülmektedir. Örneğin; 2015 Türkçe Dersi Öğretim Programı’nda ses/harfler 6 grup, 2017 ve 2019’da ise 5 grup olarak belirlenmiştir. 2015 Türkçe Dersi Öğretim Programı’na göre “e, l, a, n” 1. grup, “i, t, o, b, u” 2. grup, “k, ı, r, ö, s, ü” 3. grup, “m, d, ş, y, c, z” 4. grup, “ç, g, p, h” 5. grup ve “f, v, ğ, j” 6. grup harfler iken 2017 ve 2019 Türkçe Dersi Öğretim Programı’na göre “e, l, a, k, i, n” 1. grup, “o, m, u, t, ü, y” 2. grup, “ö, r, ı, d, s, b” 3. grup, “z, ç, g, ş, c, p” 4. grup ve “h, v, ğ, f, j” 5. grup harfler olduğu görülmektedir (MEB, 2015; MEB, 2017; MEB, 2019). Sınıf ve yayınevleri arasında küçük farklılıklar olsa da genel itibarıyla çalışmada en sık kullanılan 10 harfin sırasıyla “a, e, i, n, r, l, ı, k, d, m” oldukları, bunlardan “a, e, i, n, l, k” harflerinin 1. grup, “m” harfinin 2. grup, “r, ı, d” harflerinin ise 3. grup harflerde yer aldığı; en az kullanılan 5 harfin “j, f, ö, p, c” olduğu ve “f, j” harflerinin 5. grup, “c, p” harflerinin 4. grup, “ö” harfinin ise 3. grup harflerde yer aldığı tespit edilmiştir. Bu çalışmayla ortaya konan kullanım sıklıklarına göre 5 harf grubu oluşturulsa 1. grupta “a, e, i, n, r, l”, 2. grupta “ı, k, d, m, y, u”, 3. grupta “t, s, b, o, ü, z”, 4. grupta “ş, g, ç, ğ, v, h”, 5. grupta “c, p, ö, f, j” olması gerekirdi. Fakat kullanılan metne göre kullanım sıklıklarının değişebildiğini de vurgulamak gerekir. Örneğin; en çok kullanılan 10 harf açısından bu çalışmada elde edilen kullanım sıklıklarıyla Karaca ve Bayır (2020) ve Karadağ’ın (2020) TS Corpus veri kümesindeki kullanım sıklıklarının aynı harfler aynı sırada olacak şekilde birebir örtüştüğü; en az kullanılan harflerin ise büyük ölçüde eşleştiği görülmüştür. Özcan ve Özcan (2014), çocukların ses edinimleriyle ilgili yapılan çalışmaları incelemeleri neticesinde “k, t, m, p, b, ç, c” seslerinin en erken edinilen ve söyleyişleri daha kolay sesler olduğunu, bu seslerin öğrenilmesiyle Türkçe kullanım sıklıklarının ilişkisinin bulunduğunu belirtmişlerdir. Fakat en erken edinilen “k, t, m, p, b, ç, c” seslerinin kullanım oranının %17.57, çalışmada elde edilen sonuçlara göre 2019 Türkçe Dersi Öğretim Programı 1. grup harflerin kullanım oranının ise %46.75 olduğunu söylemek gerekir. Türkçe Dersi Öğretim Programı oluşturulmasında harflerin öğretim sırası ile sıklık sıralarının birebir örtüşmediği doğru olsa dahi harf grupları açısından özellikle de 1. grup harfler açısından harflerin Türkçedeki kullanım sıklıklarına hiç dikkat edilmediğini söylemek doğru olmayacaktır. Harflerin kullanım sıklıkları gibi söyleyiş kolaylıkları da okuma-yazma öğretiminde göz ardı edilmemesi gereken diğer önemli bir husustur. Bu hususların dikkate alınarak harf öğretim sıralamalarının belirlenmesi, okuma-yazma başarısına olumlu etki yapabilir.

Araştırma ve Yayın Etiği

Bu çalışmada “Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesi” kapsamında uyulması belirtilen tüm kurallara uyulmuştur. Yönergenin ikinci bölümü olan “Bilimsel Araştırma ve Yayın Etiğine Aykırı Eylemler” başlığı altında belirtilen eylemlerden hiçbiri gerçekleştirilmemiştir.

Yazarların Katkı Oranı

Çalışma, yazar tarafından tek başına hazırlanmıştır.

Çıkar Çatışması

Çalışmada, çıkar çatışması teşkil edebilecek durum veya ilişki bulunmamaktadır.

Kaynaklar

- Balantekin, Y. (2020). Okuma yazma öğrenme sürecinde zorlanan öğrencilerin özelliklerinin ve süreçte yaşadıkları zorlukların belirlenmesi: Bir karma yöntem araştırması. *Eğitim ve Bilim*, 45(201), 207-230.
- Banguoğlu, T. (1974). *Türkçenin grameri*. İstanbul: Edebiyat Fakültesi Basımevi.
- Çetiner, M., Yıldırım, A., Öksüz C. ve Onay, B. (2021, Eylül). *Mevzuat verisetinde soru cevaplama uygulaması*. 6th International Conference on Computer Science and Engineering (UBMK), Ankara, 603-607.

- Demir, N. ve Yılmaz, E. (2011). Ses dizimi. H. Pilancı (Ed.), *Türkçe ses bilgisi* içinde (s. 57-62). Eskişehir: Anadolu Üniversitesi.
- Dolunay, S. K. (2012). Türkçe ve dil bilgisi öğretiminde sıklık çalışmalarının yeri. *Gazi Türkiyat*, 1(10), 81-89.
- Ergin, M. (1980). *Türk dilbilgisi*. İstanbul: Boğaziçi Yayınları.
- 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.
- Gencan, T. N. (1979). *Dilbilgisi*. Ankara: Türk Dil Kurumu Yayınları.
- Gürsoy-Naskali, E. (1997). Türkçe ve Türkçenin heceleri. *Erdem*, 9(27), 1197-1200.
- Hengirmen, M. (2007). *Türkçe dilbilgisi*. Ankara: Engin Yayınları.
- Kaptan, S. (1989). *Bilimsel araştırma ve gözlem teknikleri*. Ankara: Tekışık A.Ş. Veb Ofset Tesisleri.
- Karaca, M. F. ve Bayır, Ş. (2020). Türk ve dünya edebiyatından eserlerin doğal dil işlemeyle incelenmesi. *Selçuk Üniversitesi Edebiyat Fakültesi Dergisi*, (44), 379-404. DOI: 10.21497/sefad.845451
- Karadağ, Ö. (2020). Türkçe eğitimi açısından ses-harf sıklığı üzerine bir değerlendirme. *Ana Dili Eğitimi Dergisi*, 8(1), 1-13.
- Kul, S. (2020). Türkçe ders anlatan yapay zekâya giden yolda doğal dil işleme. *Yönetim Bilişim Sistemleri Dergisi*, 6(2), 43-56.
- Millî Eğitim Bakanlığı. (2015). Türkçe dersi öğretim programı (1-8. Sınıflar). Ankara: Devlet Kitapları Müdürlüğü.
- Millî Eğitim Bakanlığı. (2017). Türkçe dersi öğretim programı (İlkokul ve ortaokul 1, 2, 3, 4, 5, 6, 7 ve 8. Sınıflar). Ankara: Devlet Kitapları Müdürlüğü.
- Millî Eğitim Bakanlığı. (2019). Türkçe dersi öğretim programı (İlkokul ve ortaokul 1, 2, 3, 4, 5, 6, 7 ve 8. Sınıflar). Ankara: Devlet Kitapları Müdürlüğü.
- Özcan, A. O. ve Özcan, A. F. (2014). Türk çocuklarının ses gelişim özellikleri ve ilk okuma yazma öğrenme. *İstanbul Gelişim Üniversitesi Sosyal Bilimler Dergisi*, 1(2), 67-86.
- Özmen, M. (2011, Ekim). *Türkiye Türkçesinde ses sıklığı*. Prof. Dr. Mine Mengi adına Türkoloji Sempozyumu, Adana, 729-739.
- Öztürk, A., Durak, Ü. ve Badıllı, F. (2020). Twitter verilerinden doğal dil işleme ve makine öğrenmesi ile hastalık tespiti. *Konya Mühendislik Bilimleri Dergisi*, 8(4), 839-852.
- Polat, H., Sever, H., Oyucu, S. ve Tekbaş, Ş. (2019). Otomatik konuşma tanıma sistemlerinde kullanılan gerçek metin verisinde biçimbilimsel-sözdizimsel hataların tespiti ve düzeltilmesi. *Veri Bilimi*, 2(2), 18-24.
- Serengil, S. I. & Akin, M. (2011, Şubat). *Attacking Turkish texts encrypted by homophonic cipher*. 10th WSEAS International Conference on Electronics, Hardware, Wireless and Optical Communications, Cambridge, UK, 123-126.
- Sezgin, F. (1993). *Dil ve edebiyatta istatistik ve bilgisayar uygulamaları*. İstanbul: Dergah Yayınları.
- Solso, R. L. & King, J. F. (1976). Frequency and versatility of letters in the English language. *Behavior Research Methods & Instrumentation*, 8(3), 283-286.
- Uzundere, E., Dedja, E., Diri, B. ve Amasyalı, M. F. (2008, Haziran). *Türkçe haber metinleri için otomatik özetleme*. Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu (ASYU 2008), Isparta, 729-739.
- Yıldırım, S. ve Yıldız, T. (2018). Türkçe için karşılaştırmalı metin sınıflandırma analizi. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 24(5), 879-886.
- Zemberek. (2015, 12 Temmuz). Açık kaynak kodlu Türkçe doğal dil işleme kütüphanesi. Erişim adresi: <http://code.google.com/p/nzemberek>

Extended Abstract

Introduction

In Turkish, an alphabet consists of 29 letters in total, 8 of them is vowel and 21 of them is consonant, is used. If a language includes more number of vowel letters, it becomes more harmonious (Hengirmen, 2007: 33). Therefore, Turkish is listed among three of four most harmonious languages used around the world (Hengirmen, 2007: 35).

Spoken languages have various rules peculiar to themselves called as grammar. Major vowel harmony examines vowel sounds in a word in terms of hard or soft, minor vowel harmony examines vowel sounds in a word in terms of flat or round (Demir and Yılmaz, 2011: 57-61), consonant harmony examines consonant sounds in terms of speech fluency (Ergin, 1980: 75-76) and the rules related to syllables can be given as examples. Grammar rules can be different from language to language. The usage frequency of speech sounds in a language cannot be similar either.

Frequency analysis can be defined as counting features such as idiom, proverb, word, syllable and letter in a text in terms of frequency and presenting these as categorical data. Frequency is one of the measurable characteristics of texts (Karadağ, 2020). As a result of literature review, it was seen that generally word, partially idiom and proverbs are examined in frequency analysis studies carried out with Turkish texts. On the other hand, to our knowledge, no study has been found on frequency analysis of Turkish language arts course textbooks. Whereas data obtained as a result of frequency analysis study on word, letter and syllables can provide useful information for the following topics: first reading-writing and Turkish language grammar instruction, writing and evaluating Turkish books, developing and revising curriculums. Moreover, these data can provide leading lights and guiding principles for the researchers carrying out studies on linguistics and language teaching considering holistic approach.

Method

In this study, word, letter and syllable frequencies of 144 reading texts included in the textbooks used in Turkish language arts course of elementary school 1st, 2nd, 3rd and 4th were analysed. Results were discussed comparatively according to the level of classes and publishers. With this study, the following characteristics of relevant data were presented: words comprising text, number of letters in words, most frequently used words, whether the words obey the rules of speech sound harmonies, usage number of letters, positions (orders) of letters and sequences, syllable numbers of words, letter numbers of syllables, types of syllables, positions of syllables and frequencies related to most frequently used syllables.

Analyses were carried out with the software developed in *Visual Studio* setting and in *C#* language. Zemberek, Turkish Natural Language Process (NLP) library, was used within the spelling stage of the words in the NLP of developed software.

Result and Discussion

Average word number of reading texts is 189.07 and also increase in word number was observed in parallel to the increase in class level. Although the first reading texts of Turkish language arts course textbooks consist of fewer words, and also the words consist of fewer letters, an increase was determined in these values as the progress continues in reading texts. It was seen that the words predominantly consist of 5 letters and 2-3 syllables, generally obey minor vowel harmony and consonant harmony rules, the ratio of the speech sounds mostly and least used are almost equal according to the level of the classes and publishers in spite of the fact that there are differences, consonant letters were used more compared to vowel letters. The average syllable number displays similarity according to the level of the classes and publishers, and this value increases considering the progress from the first reading text to last reading text. 90% of syllables were in CV and CVC types, and it was found out that the syllables "la, le, de, da, ya, ma" are the most used ones in all classes and publishers. Words consist of 6.14 letters and 2.62 syllables on average and syllables consist of 2.35 letters on average.

With this study, it was proved that the people who criticise that the instructional sequence of speech sounds does not overlap with the usage frequencies while developing curriculum are right. But it is not right to say that the usage frequencies of letters in Turkish were never taken into consideration in terms of letter groups especially in terms of 1st group letters. Because the usage ratio of 1st group letters included in 2019 Turkish Course Curriculum is 46.75%. Furthermore, ease of pronunciation of letters beside the usage frequencies of letters is another important issue that should be taken into consideration in reading-writing instruction. Determination of letter instruction sequences by

considering these issues may affect reading-writing success positively. On the other hand, if 5-letter-group was composed as in Turkish Course Instruction Curriculum by using the usage frequencies obtained within the study, groups, letters and frequency ratios would be as follows: 1st group “a, e, i, n, r, l” with 49.24%, 2nd group “ı, k, d, m, y, u” with 24.75%, 3rd group “t, s, b, o, ü, z” with 15.16%, 4th group “ş, g, ç, ğ, v, h” with 7.67%, 5th group “c, p, ö, f, j” with 3.17%.