



METİNLERİN OKUNABİLİRLİĞİNİN ÖLÇÜLMESİ ÜZERİNE BİR YAZILIM KÜTÜPHANESİ VE TÜRKÇE İÇİN YENİ BİR OKUNABİLİRLİK ÖLÇÜTÜ

*(A SOFTWARE LIBRARY FOR MEASUREMENT OF READABILITY OF
TEXTS AND A NEW READABILITY METRIC FOR TURKISH)*

Burak BEZİRCİ*, Asım Egemen YILMAZ*

ÖZET/ABSTRACT

Bu makalede, Doğal Dil İşleme (DDİ)'nin bir konusu olan okunabilirlik kavramı ve kullanım alanları ele alınmıştır. İngilizce için tanımlanmış olan ve günümüzde farklı alanlarda kullanılan, birbirinden farklı niceliklerle oluşturulmuş okunabilirlik ölçütleri incelenmiştir. Söz konusu formüllerin Türkçe için uygunlukları ve uyarlanabilirlikleri, geliştirilen bir yazılım kütüphanesi ile farklı türlerde ve seviyelerdeki Türkçe kitap metinleri kullanılarak incelenmiştir. Ayrıca, Türkçe'nin bir takım istatistiksel özellikleri belirlenerek Türkçe için yeni bir okunabilirlik ölçütü formülü önerilmiştir. C programlama dilinde geliştirilen ve platform bağımsız olarak kullanılabilen söz konusu kütüphanenin, bu konuda çalışacak araştırmacılara bir kaynak olacağı; önerilen okunabilirlik ölçütünün de Türkçe metinlerin seviyesi hakkında fikir vermesi açısından yaygın kullanım alanı bulabileceği öngörülmektedir.

In this article, as a matter of Natural Language Processing, the concept of readability and its possible areas of application have been considered. Readability metrics, which have so far been defined for English by considering different aspects and used in different areas, are investigated quantitatively. Eligibility and applicability of the existing metrics for Turkish have been analyzed via Turkish books of various types and levels by means of a software library developed by ourselves. In addition, statistical characteristics of Turkish were determined and a new readability formula for Turkish is proposed. Being developed with the C programming language with the main feature of platform independency, the software library is expected to serve as a resource for the upcoming researches in this topic; meanwhile the proposed readability metric is expected to find wide application area for prediction of the readability levels of Turkish texts.

ANAHTAR KELİMELELER/KEYWORDS

Okunabilirlik; Doğal dil işleme; İstatistiksel dil özellikleri; Yazılım kütüphanesi

Readability; Natural language processing; Statistical language characteristics; Software library

*Ankara Ün., Tandoğan Kampüsü, Elektronik Müh. Bölümü, 06100 Tandoğan, ANKARA

1. GİRİŞ

DDİ, ana işlevi bir doğal dili çözümlene, anlama, yorumlama ve üretme olan bilgisayar sistemlerinin tasarımını ve gerçekleştirilmesini konu alan bir mühendislik alanıdır (Bozşahin vd., 1998).

DDİ'de asıl amaç bilgisayara bir dili en doğru şekilde öğretmektir. Bu da, ancak o dilin karakteristik özelliklerini iyi bilmekten geçer. DDİ ile ilgili günümüze kadar, farklı birçok çalışma yapılmış ve bu çalışmalar çeşitli yazılımlarla desteklenmiştir. Bu çalışmalardan bir kısmı da okunabilirlik (*readability*) üzerinedir.

Okunabilirlik, 19. yüzyılın başlarında ABD'de ortaya çıkmış bir kavramdır (Dubay, 2004). Genellikle okunaklılık (*legibility*) kavramıyla ile karıştırılmaktadır. Okunaklılık, metnin yazı karakteri, sayfa şekli gibi özelliklere göre belirlenmektedir. Öte yandan okunabilirlik, herhangi bir dildeki bir metnin okuyucu(lar) tarafından kolay takip edilebilir olup olmadığı bilgisidir. Bu bilgi, söz konusu dilde hece, kelime ve cümle sayılarının birbirleri arasındaki ilişkileri temel alan bir takım karakteristik özelliklerin göz önünde bulundurulması ile elde edilmektedir. Özellikle İngilizce, İspanyolca, Fransızca, Almanca, İsveççe, Rusça, Japonca ve Çince dilleri için metinlerin okunabilirliği hakkında birçok çalışma yapılmıştır (Dubay, 2004; Al-Ajlan vd., 2008). Bu çalışmalar özellikle eğitim kitaplarında, askeri uygulamalarda ve sağlık alanında kullanılmıştır (Hedman, 2008; Ley, 1996). Bu çalışmalarda, dillerin özelliklerine göre okunabilirliğe etki eden faktörler tanımlanmıştır. Bu faktörlerden bazıları; ortalama kelime uzunluğu, kelime frekansı, çok heceli (*polysyllable*) kelime sayısı, ortalama cümle uzunluğu, çok anlamlı kelime sayısı, ortalama hece sayısıdır (Al-Ajlan vd., 2008).

Bu makalenin ana çerçevesi şu şekilde yapılandırılmıştır: Konuyu özetleyen giriş bölümünden sonra, 2. bölümde okunabilirlik üzerine yapılmış çalışmalar ve bir takım okunabilirlik değerleri incelenmektedir. 3. bölümde, söz konusu okunabilirlik formüllerinin Türkçe'ye uygunlukları ele alınmaktadır. 4. bölümde, okunabilirlik üzerine bu çalışma kapsamında oluşturulmuş olan yazılım kütüphanesi ile Türkçe'de okunabilirliğe etki eden nicelikler, istatistiksel olarak değerlendirilmektedir. 5. bölümde elde edilen niceliklere göre Türkçe için yeni bir okunabilirlik formülü önerilmekte; söz konusu formül diğer okunabilirlik formülleri ile karşılaştırılmaktadır. 6. bölüm ise, elde edilen sonuçlara ilişkin tartışma ve yorumların yanı sıra bu konudaki olası yeni çalışmalara ayrılmıştır.

2. OKUNABİLİRLİK ÜZERİNE YAPILAN ÇALIŞMALAR

Metinler üzerine ilk niceliksel çalışmalar, 9. yüzyılda din adamları tarafından kutsal metinlerdeki önemli sözcükleri önemsiz olanlardan ayırt etmek için yapılmıştır (Ateşman, 1997). Bunun yanında 19. yüzyılda farklı metinlerden yaklaşık 11 milyon sözcüğü inceleyerek sık kullanılan kelimeler sözlüğü hazırlayan Keading, okunabilirlik için önemli bir çalışma hazırlamıştır (Ateşman, 1997). Okunabilirliğe etki eden faktörlerin dilin yapısıyla çeşitlenmesinden ötürü, okunabilirlik hakkında özellikle İngilizce için 200 adetten fazla formül ve binlerce araştırma yayımlanmıştır (Dubay, 2004).

Bu formüllerden, İngilizce için en çok kullanılanları; SMOG, Gunning Fog, ARI, FRES, Flesch-Kincaid Grade Level'dir (McLaughlin, 1969; Gunning, 1952; Senter ve Smith, 1967; Flesch, 1948; Ateşman, 1997). Türkçe için ise (yazarların bilgisi dahilinde) bugüne kadar tek okunabilirlik formülü, Ateşman tarafından tanımlanmıştır (Ateşman, 1997).

2.1. Gobbledygook Ölçütünün Smog-Simple Ölçümü

McLaughlin tarafından 1969 yılında tanımlanan *Gobbledygook'un basit ölçüm* (SMOG) Değeri, okunabilirlik için basit bir formül olmasına rağmen ABD'de eğitim ve sağlık alanında yazılan metinlerde uzun yıllar boyunca kullanılmamıştır (Hedman, 2008; Ley, 1996).

SMOG değerinin hesaplanmasında aşağıdaki yöntem kullanılmaktadır (McLaughlin, 1969).

- Metnin başından, ortasından ve sonundan en az 10 cümlelik bölümler alınır.
- Ele alınan bu cümlelerdeki çok heceli kelime sayıları (3 hece veya daha fazlası) hesaplanır.
- Çıkan değer, metnin ABD eğitim sistemine göre tanımlanmış olan seviyesini (*US grade level*) belirtir (American Education System, 2010).

$$SMOG = 1,043 \times \sqrt{30 \times \frac{\text{Çok Heceli Kelime Sayısı}}{\text{Cümle Sayısı}}} + 3,1291 \quad (1)$$

2.2. Gunning-Fog Değeri

Gunning'in 1952 yılında kelime uzunluğu ve cümle uzunluğuna göre tanımlamış olduğu bu formül, sadece iki nicelikle metnin hangi yaş grubuna (*US grade level*) hitap ettiği, buna göre de metnin ne derece kolay veya zor olduğu hakkında bilgi vermektedir (Gunning, 1952). Formülün basit ve kolay hesaplanabilmesinden dolayı birçok ünlü dergi ve gazete bu formülü kendi yayınlarında kullanıp yayınlamışlardır.

$$Gunning - Fog = 0,4 \times \left(\frac{\text{Kelime Sayısı}}{\text{Cümle Sayısı}} + 100 \times \frac{\text{Çok Heceli Kelime Sayısı}}{\text{Kelime Sayısı}} \right) \quad (2)$$

2.3. ARI-Otomatik Okunabilirlik İndeksi

ABD ordusu tarafından kullanılmakta olan teknik dökümanlarının bir standardizasyona tabi tutulması amacıyla 1967 yılında Senter ve Smith tarafından tanımlanan bu formül, diğer okunabilirlik formüllerinden farklı olarak; metindeki ortalama kelime uzunluğunun kaç harften oluştuğu niceliğini de göz önünde bulundurarak, metnin hangi yaş grubuna hitap ettiği hakkında bilgi vermektedir (Senter ve Smith, 1967).

$$ARI = 4,71 \times \left(\frac{\text{Harf Sayısı}}{\text{Kelime Sayısı}} \right) + 0,6 \times \left(\frac{\text{Kelime Sayısı}}{\text{Cümle Sayısı}} \right) - 21,43 \quad (3)$$

2.4. Fresch Reading Ease Score (FRES)

İlk ciddi okunabilirlik formülü olan FRES 1948 yılında Flesch tarafından yayınlanmıştır.

$$FRES = 206,835 - 1,015 \times \left(\frac{\text{Kelime Sayısı}}{\text{Cümle Sayısı}} \right) - 84,6 \times \left(\frac{\text{Hece Sayısı}}{\text{Kelime Sayısı}} \right) \quad (4)$$

Bu formülden 0 ile 100 arasında çıkan değer, Çizelge 1'e göre metnin okunabilirliğinin kolaylığı hakkında bilgi vermektedir.

Çizelge 1. Flesch okunabilirlik değerleri

FRES	Metnin Okunabilirliği
0-30	Çok Zor
30-40	Zor
50-60	Biraz Zor
60-70	Normal
70-80	Biraz Kolay
80-90	Kolay
90-100	Çok Kolay

2.5. Flesch-Kincaid Değeri

Flesch'in tanımladığı FRES ve Kincaid'in tanımladığı NRI (*Navy Readability Index*) okunabilirlik formüllerinden yola çıkarak tanımlanan Flesch-Kincaid Değeri, FRES'den farklı olarak metnin eğitim seviyesine paralel olarak hangi yaş grubuna hitap ettiğini (*US grade level*) belirtir.

$$Flesch - Kincaid = 0,39 \times \left(\frac{\text{Kelime Sayısı}}{\text{Cümle Sayısı}} \right) + 1,18 \times \left(\frac{\text{Hece Sayısı}}{\text{Kelime Sayısı}} \right) - 15,59 \quad (5)$$

2.6. Ateşman Okunabilirlik Değeri

Şimdiye kadar Türkçe için bilinen tek okunabilirlik formülü, Ateşman tarafından 1997 yılında tanımlanmıştır. Ateşman'a göre Türkçe'nin ortalama cümle uzunluğu 9-10 sözcük, ortalama kelime uzunluğu ise 2,6 hecedir (Ateşman, 1997). Bu niceliklerden yola çıkarak kolay ve zor metinlerin incelenmesiyle elde edilen değerlerden Eşitlik 6'da verilmiş olan formül ortaya çıkmıştır. Ateşman, FRES için hazırlanan Çizelge 1'de gösterilen skala, benzer bir skala ile Çizelge 2'deki gibi incelediği metinlerin okunabilirlik değerlerini 0 ile 100 arasında vermektedir.

$$Atesman = 8,825 - 40,175 \times \left(\frac{\text{Hece Sayısı}}{\text{Kelime Sayısı}} \right) - 2,610 \times \left(\frac{\text{Kelime Sayısı}}{\text{Cümle Sayısı}} \right) \quad (6)$$

3. OKUNABİLİRLİK FORMÜLLERİNİN TÜRKÇE'YE UYGUNLUĞU

Türkçe ve İngilizce yapı olarak birbirinden tamamen farklı iki dil olmasına rağmen okunabilirlik formüllerinde ele alınan nicelikler hemen hemen aynıdır. Bu noktada akla gelen sorunlar ve hususlar; söz konusu niceliklerin Türkçe için uygun olup olmadığı, söz konusu formüllerin Türkçe için uygun katsayılar tanımlanarak geliştirilebilir olup olmadığı şeklinde özetlenebilir. Bu çalışma kapsamında incelenen metinlerin okunabilirlik değerleri Çizelge 3'te verilmiştir.

Çizelge 2. Ateşman'ın okunabilirlik ölçütü değerlerine ilişkin olarak tanımlanmış olduğu okunabilirlik skalası

Ateşman Değeri	Metnin Okunabilirliği
1-29	Çok Zor
30-49	Zor
50-69	Orta Güçlükte
70-89	Kolay
90-100	Çok Kolay

SMOG değeri, metnin çok heceli kelime sayısına bağlı olduğu için metindeki kelimelerin sadece hece sayısına göre yapısı hakkında bilgi vermektedir (Eşitlik 1). Türk dili sondan eklemeli bir dildir; köklerin sonuna eklerin gelmesiyle, bir sözcük ile ifade edilen kavram bazen başka bir dilde ancak bir cümle ile ifade edilebilir (Bozşahin vd, 1998). Bu nedenden dolayı, SMOG'nin ele almış olduğu nicelik Türkçe için yeterli değildir ve incelenen metinlerden çıkan değerler Çizelge 3'te gösterilmektedir. Çizelge 3'teki sonuçlara göre SMOG değeri, İngilizce'deki gibi kararlılık göstermemektedir.

Eşitlik 2, Eşitlik 3 ve Eşitlik 5'de ele alınan nicelikler hemen hemen aynıdır. Çıkan sonuçlar da ABD eğitim sistemindeki seviyeleri belirtmektedir. Türkçe metinlerin okunabilirliği için bu eşitliklerin incelenmesinin nedeni; aynı niceliklerin (ortalama kelime ve cümle uzunlukları) Türkçe'nin karakteristik özelliklerini gösterip göstermediğini anlamak üzere farklı katsayılar kullanan bu eşitliklerden çıkan değerlerin sonucunun önemli olmasıdır. Çizelge 3'e göre sonuçlar karşılaştırıldığında (*US grade level*) seviyeleri oldukça yüksek çıkmıştır. Bu çalışmada, Türkçenin yapısına göre eşitliklerin katsayıları değiştirilmiş ve değerler küçültülmüştür. Ancak değerlerin belli yerlerde yoğunluk göstermesi, ölçütlerin İngilizce'deki gibi tutarlı olmadığını göstermektedir.

Eşitlik 4 ise, yine aynı nicelikleri (ortalama kelime ve cümle uzunluklarını) ele alarak çıkan sonucu 0-100 arasında derecelendirmiştir. Çizelge 3'te taranan Türkçe metinlerin FRES değerlerinin negatif çıkması, bu formülün Türkçe metinlerin okunabilirliği için kullanılamaz olduğunu göstermektedir.

Ateşman tarafından FRES formülü Türkçe'ye uyarlanmıştır (Eşitlik 6). Çizelge 3'teki sonuçlara göre incelenen kitaplarda Ateşman'ın değerleri, en anlamlı çıkan sonuçlardır. Ancak tanımlanan eşitliğe göre değerlerin 40-80 arasında yoğunlaşmakta olduğu gözlenmektedir. Bu değerler de Çizelge 2'ye göre, metinlerin orta zorlukta ve kolay arasında olduğunu göstermektedir.

4. OKUNABİLİRLİK YAZILIM KÜTÜPHANESİ

Okunabilirlik üzerine oluşturulan yazılım kütüphanesi, C programlama dilinde geliştirilmiştir. Okunabilirlik kütüphanesi içinde farklı işlevleri olan, metin düzeltme ve incelenen kitapların orijinal bölümlerine ayıran, yazılım kütüphaneleri vardır. Okunabilirlik kütüphanesi, Windows işletim sistemi üzerinde derlenen bütün yazılım dillerinde 'dll' biçiminde kullanılabilir. Ayrıca diğer işletim sistemlerinde de platform bağımsız olarak derlenebilmektedir.

Çizelge 3. Metinlerin okunabilirlik sonuçları

Kitap Türü	Kitap Adı - Yazarı	SMOG	Gunning Fog	ARI	Flesch-Kincaid	FRES	Ateşman
<i>Felsefe</i>	Totem ve Tabu–Sigmund Freud	21,02	28,86	18,20	24,20	-46,03	40,06
	Denemeler–Montaigne	18,02	25,01	14,63	21,01	-30,67	56,25
<i>Roman</i>	Semerkant–Amin Maalouf	14,58	22,34	11,46	18,57	-23,00	71,64
	Yeşil Yol–Stephen King	14,22	21,00	10,52	17,44	-14,67	75,39
	Kar–Orhan Pamuk	17,37	24,21	13,49	20,12	-25,82	60,42
<i>Türk</i>	Memleket Hikayeleri–R. H. Karay	17,87	25,58	15,05	21,62	-36,58	55,36
<i>Klasikleri</i>	Yaprak Dökümü–R. N. Güntekin	14,96	22,52	12,84	20,15	-33,24	65,49
<i>Dünya</i>	Oliver Twist–Charles Dickens	12,68	19,48	9,43	16,65	-12,49	80,52
<i>Klasikleri</i>	Ivan Ilyiç'in Ölümü–Lev Tolstoy	16,23	22,98	11,89	18,94	-20,41	66,60
	Vahşetin Çağırısı–Jack London	17,89	24,60	13,64	19,92	-22,69	59,86
	Yol Arkadaşım Öyküleri–M. Gorki	12,44	21,11	9,83	16,89	-16,02	81,01

4.1. Kütüphanenin İşlevsel Yapısı

Okunabilirlik eşitlikleri, incelenen Türkçe metinlerin düzenlenmesinin ve üç farklı düzeyde analizlerinin yapılmasının ardından hesaplanmaktadır. Birinci düzey, tüm okunabilirlik eşitliklerinin ele aldıkları nicelikleri hesaplarken, ikinci ve üçüncü düzey Türkçe'nin karakteristik verilerini hesaplamaktadır. Elde edilen değerlere göre okunabilirlik eşitliklerinin değerleri hesaplanmıştır ve metinlerin istatistiksel analizleri yapılmıştır.

4.1.1. Metin Düzeltme Kütüphanesi

Bu yazılımda, ele alınan bir Türkçe metin niceliksel verilerinin doğru hesaplanabilmesi için, analizleri yapacak olan fonksiyonlara uygun bir şekilde yapısal olarak düzeltilmektedir. Fonksiyonlar sadece harflerin, hecelerin, kelimelerin ve cümlelerin istatistiksel verilerini inceleyeceği için metni düzeltirken numaralar ve harf dışındaki karakterler (boşluk ve cümle sonlandırıcı karakterler hariç) metnin dışında tutulmuştur.

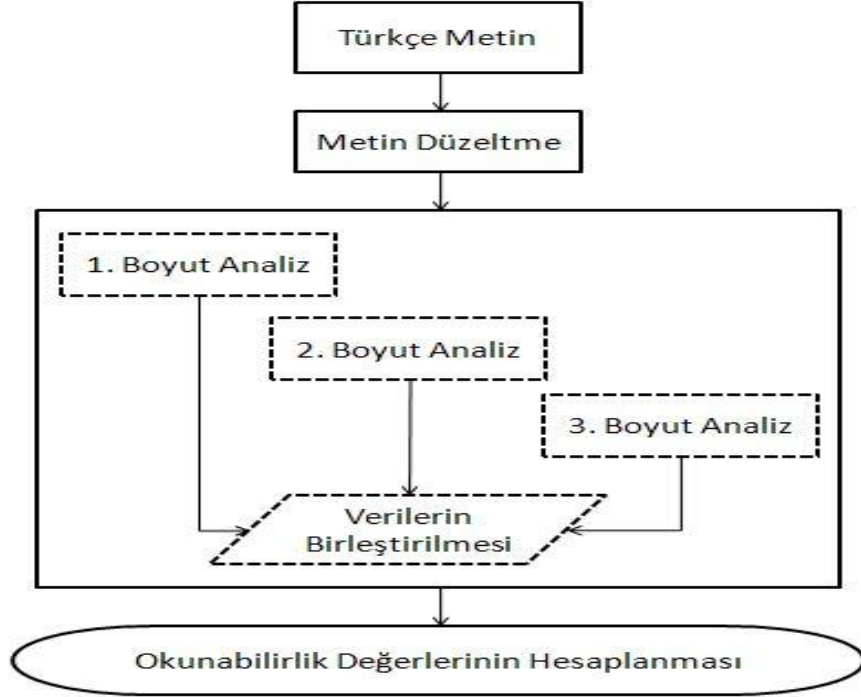
Kelimeler arası bir karakterden daha fazla boşluk olabileceği için fazla boşluklar, tek karaktere düşürmek yerine, tek karaktermiş gibi hesaplanmaktadır. Böylece metinler düzeltilerek, yazılımla yeniden yazılmadığı için sadece nicelikleri düzeltilmiş şekilde hesaplandığından birçok metin düzeltme algoritmasına göre çok daha hızlı çalışmaktadır (Şekil 1).

4.1.2. Metinlerin Niceliksel Analizi

İncelenen metinlerdeki nicelikler düzeylerine göre Şekil 2'deki gibi sınıflandırılmıştır.

4.2. Türkçe'de Okunabilirliğe Etki Eden Nicelikler

Yazılım sayesinde elde edilen sonuçlar, Türkçe'de okunabilirliğe etki eden niceliklerin ağırlıkla, ortalama cümle uzunluğu ve hecelerın frekans değeri olduğunu göstermektedir.



Şekil 1. Kütüphanenin işlevsel yapısı

4.2.1. Ortalama Cümle Uzunluğu

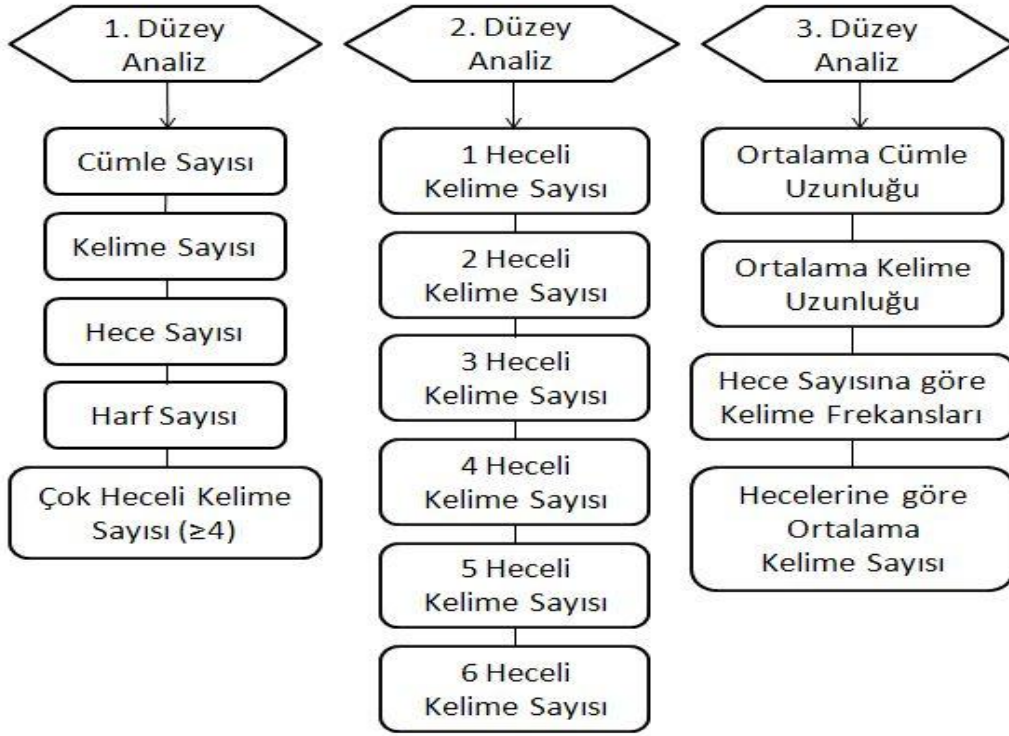
Cümle uzunluğu sadece Türkçe için değil, diğer diller için de önemli bir husustur. Bir cümlede kelime sayısı arttıkça, söz konusu cümlenin okunabilirliği ve anlaşılabilirliği azalmaktadır. Türkçe bir metinde, bu çalışma kapsamında Çizelge 4'te belirtilen, taranan eserlerden elde edilen sonuçlara göre; ortalama cümle uzunluğu 10-11 sözcük olarak tespit edilmiştir. Bu nicelik Ateşman'a göre ise 9-10 sözcüktür (Ateşman, 1997). Değerlerin bu kadar birbirine yakın olması, cümle uzunluğunun okunabilirlik açısından Türkçe için bir nicelik olabileceğini göstermektedir.

4.2.2. Hecelerin Frekansları

Türkçe bir metinde, bu çalışma kapsamında taranan eserlerden elde edilen sonuçlara göre; ortalama kelime uzunluğu Çizelge 4'te gösterildiği üzere 2,60 hece'dir. Bu sayı, Ateşman'a göre de 2,60 hece'dir (Ateşman, 1997). Ortalama kelime uzunluğunun bire bir aynı çıkması Türkçe'nin bu anlamda karakteristik bir özelliğini açıkça ortaya koymaktadır. Bu nicelikten yola çıkarak hecelerin frekansları hesaplanmıştır. Çizelge 5'te taranan eserlerde 3, 4, 5, 6 ve daha fazla heceli kelimelerin toplam kelimelere oranı verilmiştir. İlginç bir şekilde, bu değerler farklı eserler için kararlı bir şekilde yaklaşık aynı çıkmıştır.

Frekans değerleri, metinlerde bu derece kararlı bir dağılım göstererek Türkçe'nin karakteristik özelliğini ortaya koymaktadır. Bu değerlerin Şekil 3'teki dağılımlarına bakılırsa; bir, iki ve üç heceli kelimelerin frekans değerlerinin oldukça yüksek olduğu görülmektedir.

Özellikle bir ve iki heceli kelimeler genellikle “bu, şu, o” gibi sıfatlar, “de, da, ve, ile, için” gibi edat ve bağlaçlar olmakla birlikte metinlerde oldukça fazladır. Bundan dolayı bir metindeki ortalama kelime uzunluğunu, sık kullanılan bir ve iki heceli kelimeler azaltmaktadır. Bu da okunabilirliğin zorluk derecesini düşürmektedir. Ateşman’ın ve diğer formüllerin değerlerinin belli bir aralıkta yoğunlaşmasının sebebi bu niceliği göz ardı etmelerindedir.



Şekil 2. Metinlerin niceliksel analizi

5. TÜRKÇE İÇİN YENİ OKUNABİLİRLİK FORMÜLÜ

Özellikle Hint-Avrupa dil ailesinde bulunan dillere ait metinlerin okunabilirliği üzerine yapılan çalışmaların oldukça eskilere dayanıyor olmasına; bu dillere dair ölçütlerin sayısının çok olmasına rağmen Türkçe’de (yazarların bilgisi dâhilinde) bugüne değin sadece Ateşman tarafından bir eşitlik geliştirilmiştir. Bu nedenle, Türkçe için yeni bir okunabilirlik eşitliği tanımlamak adına, bu çalışmada oluşturulan yazılım kütüphanesi sayesinde, 20 farklı türde kitap ve dergi ayrı ayrı aşağıdaki paragraflarda detaylandırılmış olan adımlar izlenerek incelenmiştir.

5.1. İzlenen Adımlar

- Eserlerdeki toplam cümle sayısı, kelime sayısı, harf sayısı, karakter sayısı, hece sayısı, çok heceli kelime sayısı (4 heceli veya daha fazlası için) hesaplanmıştır.
- İngilizce için tanımlanmış olan 5 önemli okunabilirlik formülü (SMOG, Gunning-Fog, ARI, FRES ve Flesch-Kincaid) Türkçe için geliştirilmeye çalışılmış ve çıkan değerler karşılaştırılmıştır.

Çizelge 4. Ortalama cümle ve kelime uzunlukları

Kitap Adı - Yazarı	Bir Cümledeki Ortalama Kelime Sayısı	Bir Kelimedeki Ortalama Hece Sayısı	Cümle Sayısı	Kelime Sayısı
Totem ve Tabu–Sigmund Freud	18,18	2,77	2132	38758
Denemeler–Montaigne	13,77	2,64	3434	47291
Semerkant–Amin Maalouf	8,47	2,62	6914	58594
Yeşil Yol–Stephen King	8,39	2,52	10776	90458
Kar–Orhan Pamuk	13,12	2,59	8736	114624
Memleket Hikayeleri–Refik Halit Karay	11,47	2,72	3015	34577
Yaprak Dökümü–Reşat Nuri Güntekin	8,58	2,73	3081	26426
Faust–Goethe	5,46	2,42	1971	10768
Oliver Twist–Charles Dickens	5,73	2,51	3391	19442
Ivan Ilyiç'in Ölümü–Lev Tolstoy	10,49	2,55	1610	16895
Vahşetin Çağrısı–Jack London	14,08	2,54	1776	25011
Yol Arkadaşım Öyküler–Maksim Gorki	5,63	2,57	4086	23009
Ortalama	10,28	2,60	4243,5	42154,4

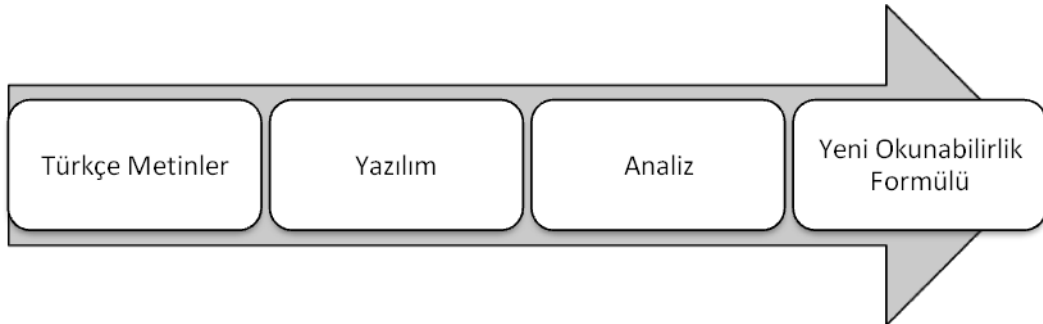


Şekil 3. Hece frekans dağılımları

- Tüm metinler göz önüne alınarak Türkçe için ortalama değerler hesaplanmıştır.
- Metinlerdeki 1 heceli, 2 heceli, 3 heceli, 4 heceli, 5 heceli, 6 veya daha fazla heceli kelime sayıları hesaplanmıştır.
- Bir cümlede hece sayısına göre ortalama kelime sayısı ve hece sayısına göre kelimelerin frekansları (rastlanma sıklıkları) hesaplanmış ve dağılım grafikleri çıkarılmıştır.

Çizelge 5. Hecelerin frekansları

Frekans	1 Heceli	2 Heceli	3 Heceli	4 Heceli	5 Heceli	≥6 Heceli
Totem ve Tabu–Sigmund Freud	0,18	0,28	0,26	0,18	0,07	0,006
Denemeler–Montaigne	0,18	0,33	0,25	0,15	0,06	0,004
Semerkant–Amin Maalouf	0,19	0,33	0,24	0,15	0,06	0,005
Yeşil Yol–Stephen King	0,22	0,33	0,23	0,14	0,05	0,007
Kar–Orhan Pamuk	0,21	0,31	0,24	0,15	0,06	0,005
Memleket Hikayeleri–R. H. Karay	0,17	0,32	0,26	0,15	0,06	0,014
Yaprak Dökümü–R. N. Güntekin	0,18	0,35	0,23	0,14	0,06	0,023
Faust–Goethe	0,20	0,40	0,25	0,10	0,04	0,005
Oliver Twist–Charles Dickens	0,22	0,36	0,24	0,11	0,04	0,014
Ivan Ilyiç'in Ölümü–Lev Tolstoy	0,20	0,34	0,25	0,15	0,05	0,002
Vahşetin Çağırısı–Jack London	0,22	0,30	0,26	0,15	0,05	0,003
Yol Arkadaşım Öyküleri–M. Gorki	0,19	0,34	0,26	0,15	0,05	0,002
Ortalama	0,20	0,33	0,25	0,14	0,06	0,008



Şekil 4. Okunabilirlik formülünün oluşturulmasındaki çalışma adımları

5.2. Okunabilirlik Formülünün Oluşturulması

Türkçe metinler için okunabilirlik 3 ve daha fazla heceli kelimelerin metinde buldukları orana göre değişmektedir. Ortalama kelime uzunluğunun da 2,60 bulunması bu önerilen kuramı doğrulamaktadır. Bu oranların farklı türdeki ve uzunluktaki eserlerde hemen hemen aynı çıkması; 3, 4, 5, 6 ve daha fazla heceli kelimeler için bir katsayı bulunmasını sağlamaktadır. Çizelge 5'teki ortalama değerlerin en küçük ortak kat (EKOK)'ları alınıp kendilerine bölünerek Eşitlik 7'deki katsayılar bulunmuştur.

$$(H3 \times 0,84) + (H4 \times 1,5) + (H5 \times 3,5) + (H6 \times 26,25) \quad (7)$$

Eşitlikteki değerlerin tanımları aşağıdaki gibidir.

- *H3*: Bir cümledeki ortalama üç heceli kelime sayısı
- *H4*: Bir cümledeki ortalama dört heceli kelime sayısı
- *H5*: Bir cümledeki ortalama beş heceli kelime sayısı
- *H6*: Bir cümledeki ortalama altı veya daha fazla heceli kelime sayısı

Çizelge 6'da işaretli olan en küçük, ortalama ve en büyük değerler alınıp Eşitlik 7'de yerlerine konularak, Çizelge 7'deki işlem çizelgesindeki sonuçlar elde edilmiştir.

Cümle uzunlukları da benzer şekilde minimum-ortalama-maksimum şeklinde ele alınmış; Eşitlik 7'ye eklenerek Eşitlik 8 tanımlanmıştır.

$$OKS \times ((H3 \times 0,84) + (H4 \times 1,5) + (H5 \times 3,5) + (H6 \times 26,25)) \quad (8)$$

Eşitlikteki *OKS* değeri, bir cümledeki ortalama kelime sayısıdır. Bir metin, bu yöntemle; 3,03–kolay, 8,30–orta, 18,82–zor şeklinde sınıflandırılmıştır. Aynı şekilde cümle uzunluğu da minimum-ortalama-maksimum şeklinde ele alınıp Eşitlik 8'de yerlerine konularak, Çizelge 8'deki işlem çizelgesi ortaya çıkarılmıştır.

Çizelge 6. Bir cümlede hecelerine göre ortalama kelime sayısı

Kitap Adı–Yazarı	3 Heceli	4 Heceli	5 Heceli	6 Heceli
Totem ve Tabu–Sigmund Freud	4,75	3,21	1,36	0,11
Denemeler–Montaigne	3,46	2,13	0,82	0,06
Semerkant–Amin Maalouf	2,04	1,29	0,51	0,04
Yeşil Yol–Stephen King	1,93	1,15	0,43	0,06
Kar–Orhan Pamuk	3,19	1,97	0,79	0,06
Memleket Hikayeleri–Refik Halit Karay	2,99	1,76	0,70	0,16
Yaprak Dökümü–Reşat Nuri Güntekin	1,95	1,17	0,52	0,20
Faust–Goethe	1,36	0,52	0,24	0,02
Oliver Twist–Charles Dickens	1,37	0,62	0,24	0,08
Ivan Ilyiç'in Ölümü–Lev Tolstoy	2,59	1,54	0,55	0,02
Vahşetin Çağrısı–Jack London	3,69	2,09	0,69	0,04
Yol Arkadaşım Öyküleri–Maksim Gorki	1,48	0,82	0,29	0,01
Ortalama	2,57	1,52	0,59	0,07

Çizelge 7. Eşitlik 7'ye ilişkin işlem çizelgesi

Eşitlik 7	Minimum	Ortalama	Maksimum
3 Heceli	1,36	2,57	4,75
4 Heceli	0,52	1,52	3,21
5 Heceli	0,24	0,59	1,36
≥6 Heceli	0,01	0,07	0,20
Sonuçlar	3,03	8,30	18,82

Çizelge 8. Eşitlik 8'e ilişkin işlem çizelgesi

Çarpma İşlemi	Cümle Uzunlukları		
	7	10	14
3,03	Minimum	Ortalama	Maksimum
3,03	21,21	30,30	42,42
8,3	58,10	83	116,20
18,82	131,74	188,20	263,48

Bu çalışma kapsamında Türkçe için geliştirilen yeni okunabilirlik formülünün sistematik büyüklükleri son olarak Çizelge 8'de bulunmuştur. Bu sayılar metnin okunabilirliğini anlatmaya çok yardımcı olmadığı için bir normalizasyon yapılarak sayılar arasında bir bağ çıkarılması gerekmektedir. Çizelge 8'deki değerlerin karekökleri alındığı zaman, Çizelge 9'daki sonuçlar bulunmuştur ve değerler anlamlı hale gelmiştir. Yeni tanımlanan okunabilirlik formülünün son hali ise Eşitlik 9'da gösterilmiştir.

$$YOD = \sqrt{OKS \times ((H3 \times 0,84) + (H4 \times 1,5) + (H5 \times 3,5) + (H6 \times 26,25))} \quad (9)$$

Çizelge 9. Eşitlik 9'a ilişkin işlem çizelgesi

Karekök Alma ($\sqrt{\quad}$)	7	10	14
3,03	4,61	5,50	6,51
8,3	7,62	9,11	10,78
18,82	11,48	13,72	16,23

6. SONUÇLAR VE TARTIŞMA

Türkiye'deki eğitim sistemi;

- 1-8 arası sınıflar için ilköğretim,
- 9-12 arası sınıflar için lise,
- 2-16 arası sınıflar için lisans,
- 16 ve sonrası sınıflar için akademik

şeklinde sınıflandırılabilir. Çizelge 9 oluşturulurken, incelenen metinlerdeki niceliklerin (hece frekansları ve cümle uzunlukları) minimum-ortalama-maksimum değerleri ele alınmıştır. Buna göre Çizelge 9, en kolay metin ile en zor metin arasında değerlerin sadece dağılımını göstermektedir. Bu dağılıma göre, en kolay okunabilirlik derecesi 4,61; en zor okunabilirlik derecesi 16,23 çıkmaktadır. Bu çalışmanın sonucunda yeni okunabilirlik eşitliğinin sonucu, metnin Türkiye'deki eğitim sistemine göre hangi sınıf düzeyine hitap ettiğini açıklamaktadır. Örneğin, Stephen King'in Yeşil Yol adlı romanının okunabilirlik değeri 7,34 çıkmıştır. Bu değer, söz konusu eserin 7-8. sınıfa giden bir öğrencinin seviyesine hitap ettiğini belirtmektedir. Eşitlik 9, incelenen eserlere uygulanırsa Çizelge 10'daki değerler çıkmaktadır.

Okunabilirlik değerleri, metinlerin üslubu hakkında bilgi vermektedir. Dolayısıyla bir metnin okunabilir olduğunu söylemek veya o metnin hangi sınıfa hitap ettiğini adreslemek

metinlerin anlaşılabilirliği hakkında az da olsa bilgi verebilmektedir. Ancak, okunabilirlik eşitlikleri metnin anlaşılabilirliği hakkında kesin veriler içermemektedir.

Önerilmekte olan okunabilirlik eşitlikde cümleler çok uzun olsa bile kısa kelimelerin kullanılması ile metnin okunabilirlik seviyesi düşürülebilir. Örneğin, Goethe'nin Faust adlı çeviri eserinin okunabilirlik seviyesi, 4-5. sınıf düzeyinde çıkmıştır. Sonucun bu seviyede çıkması anlaşılabilirlik açısından çok doğru değildir. Ancak, Goethe'nin üslubu hakkında oldukça doğru bir yaklaşım ortaya koymaktadır. Faust adlı çeviri eserin ortalama cümle uzunluğu 5,46 sözcük, ortalama kelime uzunluğu 2,42 hece ve 1-3 arası heceli kelimelerin toplam frekansı 0,86 çıkmıştır. Bu da, Faust adlı eserin oldukça kısa cümlelerle ve basit kelimelerle tercüme edildiğini göstermekte ve doğrulamaktadır.

Bu makalenin okunabilirlik değeri ise, önerilmekte olan yeni okunabilirlik ölçütüne göre 14,45 çıkmakta olup; metnin lisans düzeyinde (akademik) olduğunu göstermektedir.

Bu çalışmada oluşturulan yazılım, özellikle web uygulamalarında ve Türkçe DDİ çalışmalarında kullanılabilir bir yazılım kütüphanesi olması amacıyla platform bağımsızlığı gözetilerek geliştirilmiştir. Bu çalışmanın devamında, okunabilirlik ve anlaşılabilirlik kavramlarının bir yazılımla ilişkilendirilmesi hedeflenmektedir.

Çizelge 10. Yeni okunabilirlik değerleri ve sınıfları

Kitap Türü	Kitap Adı - Yazarı	YOD	Sınıf Seviyeleri
<i>Felsefe</i>	Totem ve Tabu–Sigmund Freud	17,38	>16
	Denemeler–Montaigne	12,05	12
<i>Roman</i>	Semerkant–Amin Maalouf	7,41	7-8
	Yeşil Yol–Stephen King	7,34	7-8
	Kar–Orhan Pamuk	11,44	11-12
<i>Türk</i>	Memleket Hikayeleri–Refik Halit Karay	11,64	11-12
<i>Klasikleri</i>	Yaprak Dökümü–Reşat Nuri Güntekin	9,48	9-10
<i>Dünya</i>	Faust–Goethe	4,24	4-5
<i>Klasikleri</i>	Oliver Twist–Charles Dickens	5,36	5-6
	Ivan Ilyiç'in Ölümü–Lev Tolstoy	8,53	8-9
	Vahşetin Çağrısı–Jack London	11,69	11-12
	Yol Arkadaşım Öyküler–Maksim Gorki	4,59	4-5

KAYNAKLAR

- Al-Ajlan A. A., Hend S. A., AbdulMalik S. A. (2008): “Towards the Development of an Automatic Readability Measurements for Arabic Language”, 3rd Int. Conf. Digi. Inf. Mng. (ICDIM), s. 506–511.
- American Education System, Çevrimiçi: <http://www.indobase.com/study-abroad/countries/usa/usa-education-system.html> Son Erişim Tarihi: 15.07.2010.
- Ateşman E. (1997): “Türkçede Okunabilirliğin Ölçülmesi”, Dil Dergisi, Cilt 58, s. 71-74.
- Bozşahin C., Zeyrek D., Pembeci I. (1998): “Computer-aided Learning of Turkish Morphology”, Debrecen, Hungary, Proc. Joint Conf. ACH ALLC.
- DuBay W. H. (2004): “The Principles of Readability, Impact Information”, Costa Mesa, CA.

- Flesch R. (1948): "A New Readability Yardstick", *Journal of Applied Psychology*, Cilt 32, s. 221-233.
- Gunning R. (1952): "The Technique of Clear Writing", McGraw-Hill International Book Co, New York.
- Hedman A. S. (2008): "Using the SMOG Formula to Revise a Health-Related Document", *American Journal of Health Education*, Cilt 39, Sayı 1, s. 61-64.
- Ley P. T. (1996): "The Use of Readability Formulas in Health Care", *Psychology, Health & Medicine*, Cilt 1, Sayı 1, s. 7-28.
- McLaughlin G. H. (1969): "SMOG Grading-A New Readability Formula", *Journal of Reading*, Cilt 12, Sayı 8, s. 639-646.
- Senter R. J., Smith E. A. (1967): "Automated Readability Index", Technical Report, Cincinnati University, Ohio.