

BİLGİSAYAR DESTEKLİ METİN OKUNABİLİRLİĞİ ANALİZİ

Tolga GÜYER* Turan TEMUR** Ebru SOLMAZ***

Öz

Okunabilirlik kavramı dile özgü bir kavramdır ve genel olarak bir metnin belirli bir düzeydeki okuyucular tarafından anlaşılma derecesi şeklinde tanımlanmaktadır. Kullanılan dile özgü olması itibarı ile bu alanda yapılan çalışmaların daha çok ulusal nitelikte olması kaçınılmazdır. Amacımız, disiplinler arası katılımı gerektiren bu türdeki çalışmaların süreç aşamalarını ortaya koyarken, benzer araştırmalarda kullanılacak ve Türkçe için eksikliğini hissettiğimiz bir metin analizi yazılımının geliştirilmesi sürecini de konuyla ilgilenen araştırmacılarla paylaşmaktır.

Anahtar Sözcükler: *Okunabilirlik, anlaşılabilirlik, okunabilirlik formülleri*

Abstract

Readability concept is a concept unique to language and is generally defined as the level of comprehensibility by readers of a certain level. Considering that it is unique to the language used, it is inevitable that studies conducted in this field are mostly national in quality. Our aim is to share the process of the development of text analysis software for Turkish which we felt its deficiency and which can be used in similar researches by the researchers interested in the subject, while putting forward the process stages of these kinds of studies requiring interdisciplinary participation.

Keywords: *Readability, comprehensibility, readability formulas*

Yazışma adresi: *Yrd. Doç. Dr., Gazi Üniversitesi, Gazi Eğitim Fakültesi, Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü, tguyer@gazi.edu.tr; **Yrd. Doç. Dr., Dumlupınar Üniversitesi, İlköğretim Bölümü, Sınıf Öğretmenliği Anabilim Dalı; ***Bilgisayar Öğretmeni, Ankara Kazan İlköğretim Okulu.

En genel şekliyle, bir metnin belirli bir düzeydeki okuyucular tarafından anlaşılma derecesi şeklinde tanımlanan okunabilirlik kavramının ortaya çıkışının Aristo ve Plato'ya kadar uzandığı söylenebilir. Bununla birlikte, bu alanda yapılan ilk akademik çalışmalar 1920'lere rastlamaktadır. Dolayısıyla okunabilirliğin en azından felsefi anlamda, uzun bir geçmişe sahip olduğunu söylemek mümkündür. Eğitim açısından bakıldığında, okunan eğitsel bir metnin öğrenen tarafından anlaşılmasının öneminin fark edilmesinden itibaren okunabilirliğin ölçülmesi üzerine çalışmalar da yoğunlaşmıştır (Chall, 1988). Okunabilirlik üzerine yapılan araştırmaların ve bu araştırmalarda okunabilirlik kavramının ele alınma biçiminin genel çerçeveleri itibariyle evrensel oldukları görülmektedir. Ancak detaya inildiğinde ve bu konunun dile bağımlı olduğu düşünüldüğünde, gerçekleştirilen çalışmaların ve ilgili kavramın kullanılan dilin dil bilimsel özelliklerine, o dili konuşan toplumun sosyal ve kültürel yapısına, dilin zenginliğine ve çeşitli seviyelerdeki yaş ve eğitim grupları tarafından kullanım farklılıklarına göre şekillendiği ortaya çıkmaktadır. Başka bir deyişle, okunabilirlik uluslararası anlamda kabul gören bir kavram olmasının yanında, kullanılan dile özgü bir yapı göstermektedir. İlk olarak İngilizce dilinde başlayan okunabilirlik çalışmaları, diğer bir yaygın dil olan İspanyolcada devam etmiştir. Daha sonra Çince, Almanca, Fransızca, Fince ve Rusça gibi dillerde de çeşitli araştırmalar ve formül geliştirme çalışmaları yapılmıştır (Klare, 1988).

Okunabilirlik Nedir?

Okunabilirlik konusunda yerli ve yabancı kaynaklar incelendiğinde, bu kavramın tasvir edildiği birçok tanımla karşılaşmıştır. Analiz edildiğinde her tanımın, birbirlerini tamamlayan bir yapıda okunabilirliğin çeşitli yönlerine değindiği görülmektedir. Aşağıda, alanyazıda en çok kabul görmüş olan tanımlara değinilmiştir:

Ateşman (1997) okunabilirliği, okuyucu için okuduğu metnin kolay ya da zor anlaşılır olma durumu şeklinde tanımlamıştır. Okunabilirlik kavramı metinlerin güçlük seviyesi ve okuyucu düzeyine uygunluk seviyesi ile ilgilidir.

Klare (1963)'e göre okunabilirlik, yazma stilinden dolayı, anlama ve kavramanın kolaylığıdır. Bu tanım biçim, yapı özellikleri ve içerik gibi faktörlerden çok yazma stili üzerine odaklanmıştır (Dubay, 2004).

SMOG adı verilen okunabilirlik formülünün yaratıcısı McLaughlin (1969) ise okunabilirliği şu şekilde tanımlamıştır; belirli bir sınıftaki insanların okuduğu

konuyu anlama ve zorlanma derecesidir. Bu tanım okuma becerileri, ön bilgiler ve motivasyon gibi özellikleri bilinen bir sınıf okuyucu ile metin arasındaki etkileşimi vurgulamaktadır (Dubay, 2004). Klare ve McLaughlin tarafından yapılan tanımlar karşılaştırıldığında McLaughlin'in Klare'in tanımına zıt olarak, okuma becerisi, motivasyon, konu ile ilgili bilgi seviyesi gibi belirli okuyucu özelliklerinin önemini ve bunların metin ile birbirlerini nasıl etkilediklerini dikkate aldığı görülmektedir (Anagnostou ve Weir, 2006).

Okunabilirliğin en kapsamlı tanımı Dale ve Chall (1948, 1949)'ın tanımı olarak kabul görmektedir. Dale ve Chall okunabilirliği şu şekilde tanımlamışlardır; okunabilirlik okuyucu grubunun materyali kullanarak sahip olduğu başarıyı etkileyen, basılı materyal ile ilgili bütün unsurların toplamıdır. Başarı, okuyucuların materyali anlaması, uygun bir seviyede okuması ve onu ilginç bulmasının bir sonucudur (Akt. Dubay, 2004; Anagnostou ve Weir, 2006).

Johnson (2000); bir yazarın bir kitap, bir çalışma yaprağı veya sınav kâğıdı yazarken, okuyucuya bilgiyi aktarmak amacıyla olduğunu ve yazarın bunu ne kadar iyi başardığının metnin okunabilirliğine bağlı olduğunu ifade etmiştir. Okunabilirlik, okuyucu ve metni arasındaki uyum problemi ile ilgilidir. Başarılı bir okuyucu basit metinlerden sıkılırken, zayıf bir okuyucu ise akıcı okuyamadığı metinlerden vazgeçer. Johnson okuyucunun, aşağıdaki gibi bir takım özelliklere sahip bir metni okumayı istemeyeceğini belirtmiştir;

- Kötü basılmış,
- Karışık cümle yapıları içeren,
- Uzun kelimeler veya
- Tamamen yeni düşünceleri içeren metinler.

Okunabilirlik konusunda genel varsayım, kısa cümle ve kelimelerden oluşan bir metnin ortalama bir okuyucu tarafından kolayca okunduğu ve anlaşıldığıdır. Bu görüşten hareketle, okunabilirlik formülleri genellikle ortalama kelime uzunluğu ve ortalama cümle uzunluğu gibi hesaplanabilir metin özelliklerine odaklanır. Bu değişkenleri kullanan ve en popüler formül olarak kabul edilen okunabilirlik formülü, 1948'de Rudolf Flesch tarafından yayınlanmıştır.

RE : Okuma kolaylığı (1- 100 arası ölçek)

ASL : Ortalama cümle uzunluğu (kelime sayısının cümle sayısına bölünmesi)

ASW : Her kelimedeki ortalama hece sayısı (hece sayısının kelime sayısına bölünmesi)

olmak üzere

$$RE = 206.835 - 1.015 ASL - 84.6 ASW$$

biçiminde tanımlanan bu formül, daha sonra Amerika Deniz Kuvvetleri tarafından yürütülen bir çalışmada diğer okunabilirlik araştırmacıları tarafından gözden geçirilmiştir. Bu çalışmalar sonucunda uyarlanan formül, Flesh – Kincaid veya Flesch Seviye Ölçeği olarak bilinmektedir (Anagnostou ve Weir, 2006).

Kısa cümle ve kelimelerden oluşan bir metnin ortalama bir okuyucu tarafından kolayca okunduğu ve anlaşıldığı varsayımı birçok metin için doğru olsa da bu değişkenler açısından uygun gözüken ancak hedef okuyucu kitlesi tarafından anlaşılmayan kelimelerden oluşan metinler için geçerli olmayacaktır. Bu türdeki metinler bilinen okunabilirlik ölçümlerinde iyi sonuçlar verse de okuyucuların önemli bir kısmı bunların zor anlaşıldığını düşünecektir. Bununla birlikte kısa olup yaygın kullanılmayan birçok kelime olabileceği gibi çok yaygın olarak kullanılan uzun kelimeler de bulunmaktadır (Weir ve Calum, 2006). Diğer bir deyişle hece ve kelime sayısı okunabilirliğin tek belirleyeni değildir. Bir metindeki terimlerin, yabancı kelimelerin, zor sözcüklerin, somut veya soyut kelimelerin vb. de metnin okunabilirliğini etkileyeceği açıktır ve bunlar formüllerde yer alması gereken değişkenler arasında değerlendirilir (Ateşman, 1997). Örneğin kelime grubu belirli seviyede bulunan bir okuyucu grubunun en çok kullandığı kelimelerden oluşmaktadır (Temur, 2006). Verilen bir metnin bu kelimelerden oluşması o metnin o seviyedeki grup tarafından daha kolay okunacağını göstermektedir ya da her hangi bir ders kitabında o ders alanına uygun kelime veya terimlerden oluşan kelime grupları yer almaktadır. Konu alanına ait bu kelime grupları alana özgü metin kapsamında olacaktır. Alana özgü kelime ve terimler metnin okunabilirliğini etkileyen etkenlerden olduklarına göre okuyucunun alana özgü kelimeleri bilip bilmemesi, o metni anlama seviyesini doğrudan etkileyecektir.

Türkçede Okunabilirlik Çalışmaları

Ateşman'ın 1997 yılında yapmış olduğu Türkçe için bir okunabilirlik formülü geliştirme çalışması Türkçedeki ilk ciddi çalışma olarak nitelendirilebilir. Ateşman araştırmasında Flesch'in okunabilirlik formülünü Türkçeye uyarlamıştır. Geliştirilen formül Flesch'in formülünde yer alan değişkenlerden oluşmaktadır ve Türkçe için geliştirilen ilk okunabilirlik formülü olarak ayrı bir önem taşımaktadır.

Tekbıyık (2006) ise, ders kitaplarının okunabilirliği ile ilgili bir çalışma yapmıştır. Bir okunabilirlik formülü ile bir okunabilirlik belirleme tekniğinden yararlanarak bir lise 1 fizik kitabının okunabilirlik seviyesini ve hedef yaş düzeyine uygunluğunu incelemiştir. Araştırmada okunabilirlik için Cloze, hedef yaş düzeyine uygunluk için FOG testi kullanılmıştır. Çalışma sonunda ders kitabının okuma yaşının düşük seviyedeki öğrenciler için yüksek olabileceği ve öğrencilerin kitabı okurken bazı güçlükler ile karşılaşabilecekleri belirlenmiştir.

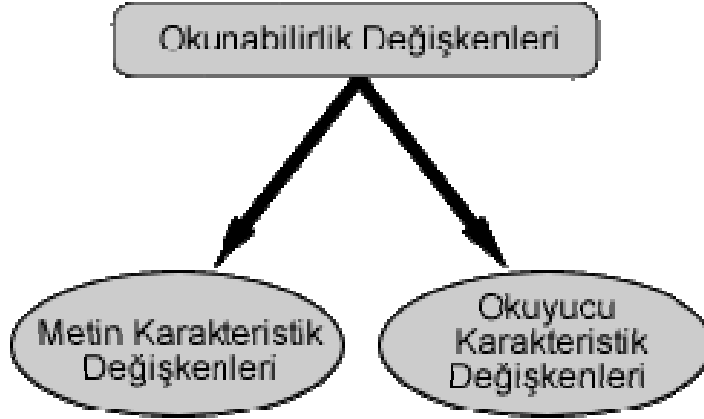
Zorbaz (2007) araştırmasında Türkçe ders kitaplarındaki masalların kelime ve cümle uzunluklarının sınıflara göre değişimini ve metinlerin okunabilirlik düzeylerini incelemiştir. Metinlerin okunabilirlik düzeylerinin belirlenmesinde, Ateşman'ın Türkçe için Flesch'ten uyarladığı, kelime ve cümle uzunluğunu temel alan formül kullanılmıştır. Araştırma sonucunda, metinlerin cümle uzunluğu ve okunabilirlik seviyesi açısından sınıflar arasında belirli bir değişimin olmadığı görülmüş, metinlerin okunabilirlik yönünden kolay oldukları tespit edilmiştir. Aynı zamanda ders kitaplarında bulunan masalların kelime ve cümle uzunlukları öğrencilerin sınıf seviyesine göre belirli bir artış göstermemektedir. Zorbaz bu çalışmada okunabilirlik testlerinin kesin sonuçlar veremeyeceğini, bu ölçümler ile metnin düzeyi ile ilgili bazı fikir ve tahminlerden öteye geçilemeyeceğini ifade etmektedir. Zorbaz'a göre okunabilirlik ölçümleri ancak nicel özelliklerin yanında nitel özelliklerin de ele alınması ile kesin bir sonuca ulaşabilir.

Okunabilirlik üzerinde yapılan yeni çalışmalardan birisi de Çiftçi, Çeçen ve Melanlıoğlu (2007) tarafından gerçekleştirilmiştir. Bu çalışmada 6.sınıf kitaplarında bulunan kırk altı düz yazının kelime ve cümle uzunlukları kullanılarak okunabilirlik seviyeleri ölçülmüştür. Bu ölçüm Ateşman'ın Flesch'den uyarladığı formül ile gerçekleştirilmiştir. Araştırmada kullanılan düz yazılar öyküleyici ve bilgilendirici olarak ikiye ayrılmıştır. Öyküleyici metinlerin cümle uzunlukları bilgilendirici öykülere göre daha kısa olduğu için okunabilirlik düzeyleri daha yüksek çıkmıştır.

Okunabilirlik Formülünün Gerekliliği

Konuya en geniş çerçeveden baktığımızda, okunabilirlik ölçümlerine ilgi duyan bütün araştırmacıların amaçladıkları ortak hedefin, yaşadıkları toplumun konuştuğu dile ait en genel okunabilirlik formülünü elde etmek olduğunu görürüz. Böyle bir formül, yazılan ders kitaplarından roman ve hikâye kitaplarına, hatta günlük gazetelerde yayınlanan makalelere kadar bütün basılı materyalin, hedefi durumundaki okuyucular için en uygun içeriğe sahip olacağını garantiyecektir.

Ancak çok geniş bir örnekleme ve detaylı istatistiksel analizlerle gerçekleştirilmesi gereken böyle bir çalışmanın yapılması hiç de kolay değildir. Bu türdeki çalışmalarda yaşanan güçlükler sadece örneklemin geniş olmasından kaynaklanmamakta, okunabilirliği etkilediği düşünülen değişkenlerin belirlenmesi ve bunların kontrol altına alınması da önemli bir problem olarak ortaya çıkmaktadır. Bu değişkenleri iki ana grupta inceleyebiliriz. Bu değişken grupları Şekil 1’de verilmiştir.



Şekil 1

Okunabilirlik Analizinde Değişken Grupları

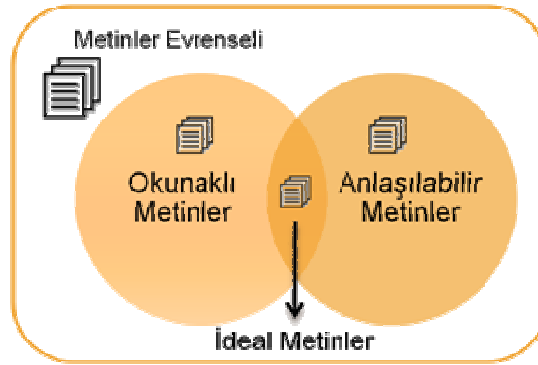
Metin karakteristiğini belirleyen değişkenlerin ölçülmesi genellikle daha kolay olmakta, ancak bu değişkenlerin sayısındaki artış, yapılacak regresyon analizleri için gerekecek örnek sayısını da aynı oranda artırmaktadır. Dolayısıyla bu bağımsız değişkenlerin sayısına sınırlama getirilmesi kaçınılmaz olacaktır. Bu durumda örneklemimizi oluşturan örnekler sadece metinlerden oluşacaktır, okuyucuların nicelik ve niteliklerinin önemi yoktur.

Diğer yandan okuyucuların karakteristik değişkenlerini de analize dâhil etmek istediğimiz takdirde, motivasyon durumları, okudukları metnin içeriğine olan ilgileri ya da bu konudaki ön bilgi düzeyleri gibi kontrol altına alınması oldukça zor olan durumlarla karşılaşmaktadır. Ayrıca bu durumda örneklemimizi oluşturan sadece metinler olmayacak, metinlerin yanı sıra okuyucular da doğrudan ölçümlere dâhil olacaktır.

Okunabilirlik ve Anlaşılabilirlik Kavramları

Okunabilirlik kavramının önem kazandığı 1950’li yıllar düşünüldüğünde, bu kavramın dönem özelliklerinden etkilenecek şekilde şekillendiği anlaşılabilir. Metinlerin tamamının daktilo kullanılarak yazıldığı, dolayısıyla renk, metin yerleşimi, yazı tipi ya da yazı tipi boyutu gibi teknoloji ile gelişen tasarım özelliklerinden mahrum oldukları bir dönemde, okunabilirlik ve anlaşılabilirlik kavramlarının çakışması da gayet doğaldır. Dolayısıyla o dönemlerde okunabilirlik olarak isimlendirilen kavram, biçim ve yapı özelliklerinin etkisinden yoksun, o metni okuyan her okuyucu için aynı şeyi ifade eden bir kavram olarak ele alınmaktadır. Sonuç olarak dönem araştırmacılarının, okunabilirliği metnin okuyucu tarafından anlaşılma seviyesi olarak ele almış oldukları söylenebilir. Başka bir deyişle okunabilirlik genel bir kavram olarak kullanılmış, fakat bu sözcük ile sadece bir metnin anlaşılabilirliği ifade edilmiştir. Metne ait karakteristik değerlerini içeren, hatta okuyucunun bireysel özelliklerini yansıtan parametrelere dahi bağlı olabilen ve ortaya çıktığı yıllarda “okunabilirlik” olarak adlandırılan kavramın, günümüzde “anlaşılabilirlik” olarak isimlendirilmesi daha doğru olacaktır.

Diğer yandan günümüzde herhangi bir basılı materyaldeki metni sadece dilsel özellikleri ile ele almak imkânsızdır. Yazı; teknolojinin kullanımı ile yazı tipi, yazı tipi büyüklüğü, renk gibi özellikler kazanmakta ve resim, tablo, harita, çizelge gibi yazının yanında bulunan çeşitli nesnelere bir kompozisyon oluşturmaktadır. Bu biçimsel ve görsel metin özellikleri ise günümüzde “okunaklılık” başlığı altında incelenmeye başlanmıştır. Sonuç olarak okunabilirlik kavramı şu anki şartlar göz önüne alınarak değerlendirildiğinde, Şekil-2’deki gibi sınıflandırılabilir.



Şekil 2

Metinlerin Okunabilirlik Açısından Sınıflandırılması

Dolayısıyla okunabilirliğin temel olarak iki ana kavramdan oluştuğunu söyleyebiliriz: Okunaklılık ve anlaşılabilirlik.

Okunaklılık: Bu kavram yazı tipi, yazı tipi boyutu, metin yerleşimi, renk ve metin/resim/tablo vb. kompozisyonu etkenlerinden oluşur (Johnson, 2000; Dubai, 2004; Anagnostou ve Weir, 2006). Okunaklılık konusunda alanyazı incelendiğinde, yapılan araştırmalar ile geçerlilikleri kanıtlanmış bir takım tasarım ilkelerine rastlanmaktadır.

Anlaşılabilirlik: Bu kavram hece sayısı, kelime sayısı, kelime grupları, zor kelimeler, yabancı sözcükler, terimler, soyut ve somut kelimeler (Ateşman, 1997) gibi anlamsal etkenleri içermektedir. Kabaca, bir metni okumaya devam etmeyi isteyen biri onu anlayabiliyor demektir. Diğer taraftan birinin okuduğu bir materyali anlaşılır bulması onu okunabilir bulduğu anlamına gelmez (McLaughlin, 1974). Bugünkü okunabilirlik formülleri, gerçekte anlaşılabilirliği tahmin etmektedir.

Hem anlaşılabilir, hem de okunaklı metinler ise, metinler evreninde ideal metinleri temsil ederler (Şekil-2).

Bilgisayar Destekli Metin Analizi

Dilin bilgisayar destekli olarak analiz edilmesini teknik anlamda iki kategoride değerlendirebiliriz: Birinci kategori, “doğal dil işleme” olarak da adlandırılan ve diller arası çeviri algoritmalarını da kapsayan araştırma alanıdır ve doğrudan bilgisayar bilimlerinin yapay zekâ dalıyla ilgilidir.

İkinci yaklaşım ise, dilin dilbilgisi yapısından çok, okunabilirlik araştırmalarında da kullanılan, kelime ve cümlelerin nicelik verilerinin incelenmesi ilkesi üzerine kurulmuştur. Bu yaklaşımda yazılım geliştirmek ilkinde göre daha kolay olsa da ele alınan dilin yapısına bağlı olarak, özellikle de eklemeli dillerde bazı problemler yaşanabilmektedir.

Okunabilirlik alanında gerçekleştirilen çalışmaların ürünlerinden olan okunabilirlik formülleri, her ne kadar pratik kullanım amacıyla geliştirilmiş olsalar da uzun metinlerde elle hesaplamayı olanaksız kılan değişkenlere sahip olabilirler. Bu durumda kullanılacak kelime-işlem yazılımları da oldukça kısıtlı ve sınırlayıcı olacaktır. Özelleştirilebilir bir yapıda olan ve istenilen formül ve değişkenlerle analiz yapmaya olanak tanıyan bir yazılım, bu konuda gerçekleştirilecek her türde araştırma için önemli bir materyal olacaktır. Dolayısıyla okunabilirlik analizi için geliştirilecek bir yazılımın iki ana amaç doğrultusunda kullanılması söz konusudur. Bunlar:

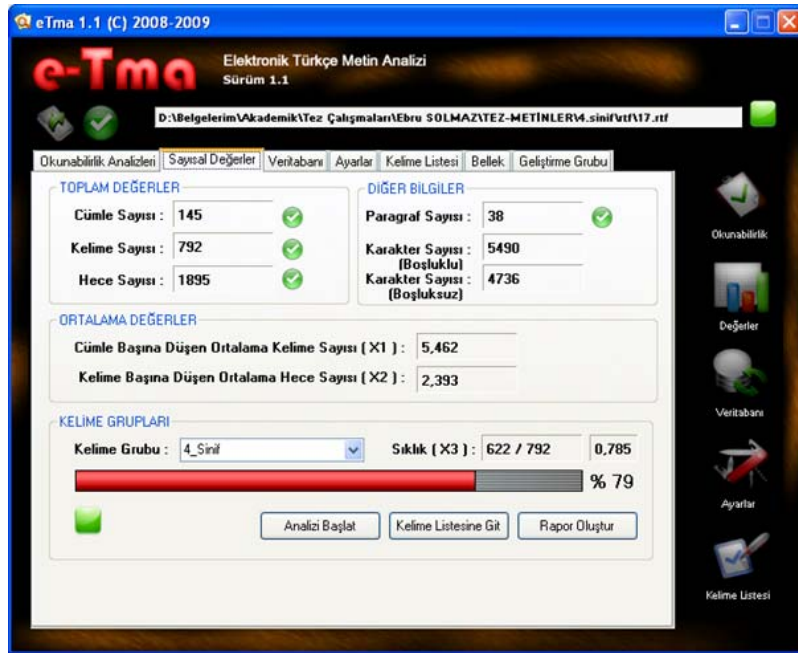
1.Çalışmaların sonucunda elde edilen formüllerin metinlere uygulanması işleminin bilgisayar destekli olarak gerçekleştirilmesi:

Okunabilirlik formülleri, içerdikleri değişken grubuna bağlı olarak, çok sayıda değişkenden oluşabilir. Ayrıca bu formüller regresyon analizinden elde edilen katsayıları kullandıklarından, ondalık hassasiyet dereceleri yüksek olacaktır. Bütün bu sebeplerden, özellikle metin sayısının arttığı durumda hesaplamaların elle yapılması yanlış sonuçlara ulaşılması olasılığını yükseltecektir. Bu durumda bir yazılımın desteğine gereksinim duyulması kaçınılmaz olacaktır.

2.Okunabilirlik araştırmalarında, süreç içerisinde bir araştırma materyali olarak kullanılması amacıyla:

Okunabilirlik formüllerinin geçerliliği, özellikle okuyucuların karakteristik değişkenlerini içerdikleri durumda, büyük ölçüde katılımcıların niceliğine ve niteliğine bağlı olacaktır. Dolayısıyla, okuyuculardan okunabilirlik puanlarının, yani okuyucuların metinleri ne kadar anladıklarını belirleyen puanların toplanması ve bu puanların metinlerin karakteristik özellikleri ile birlikte değerlendirilmeleri işlemlerinde, yine yazılım tarafından sağlanacak destek önem kazanacaktır.

Mevcut okunabilirlik formüllerini metinlere uygulamanın yanı sıra bu alanda gerçekleştirilecek çalışmalarda da kullanılacak bir yazılım olan *elektronik Türkçe metin analizi* (eTma)'nin yapımına başlanmıştır ve yazılım hâlihazırda geliştirilme aşamasındadır (Şekil-3).



Şekil 3
eTma'dan Bir Ekran Görüntüsü

Yazılım, metinleri genel olarak kabul gören bir kelime işlem formatı olan zengin metin biçiminde kabul etmektedir. Okunabilirlik analizlerinin yanı sıra, metin hakkında çeşitli sayısal değerlerin de hesaplandığı yazılım, ayrıca veritabanı bağlantılı olarak kelime gruplarının tanımlanmasına ve bu grupların metin içersindeki kullanım sıklıklarının hesaplanmasına da olanak vermektedir.

Türkçenin de üyesi olduğu eklemeli dillerin bilgisayar destekli olarak analiz edilmesinde karşılaşılan zorluklar, eklerle kelime kökünün birbirinden ayrılması sırasında başlar. Eğer kök durumunda olan bir fiil ise, bu durumda ilk ek tarafından gerçekleştirilen harf yumuşamaları (git-gidiyor gibi) kök durumunda olan fiilin ayrılmasını daha da güçleştirmektedir. eTma bünyesinde, bu gibi durumlar için *akıllı eşleştirme sistemi* geliştirilmiştir. Sistem, fiilde yer alan mastar ekini elimine ederek, kelimenin analiz edilen metin içersinde yer alan bütün biçimlerini eşleştirebilmektedir. Örneğin “gelmek” fiilinin metin içersinde yer alması muhtemel olan “geliyordu” ya da “gelecek” gibi biçimleri, sistem tarafından yakalanabilmektedir.

Bunların yanı sıra, kelimelerin metin içersinde hangi anlamları ile yer aldıklarının belirlenmesi problemi de vardır ki, bu problemin çözümü için bilgisayar destekli analizin mutlak surette bir insan tarafından desteklenmesi gerekmektedir. Bir bilgisayar için bazı durumların birbirlerinden farkının ayırt edilmesi olanaksız olabilmektedir. Örneğin, “Lokantada sunulan yemeklerin hepsi de kırmızı etti.” cümlesi ile “Neden böyle etti, anlamadım.” cümlesinde yer alan “etti” kelimesi, anlamsal olarak farklı olsalar da bilgisayar için kullanılan yöntemle bağlı olarak ya etmek fiilinin geçmiş hâli olarak ya da “et idi” olarak algılanacaktır.

Benzer şekilde “yer” ve “yerli” kelimeleri arasındaki ilişki ile “ceket” ve “ceketli” kelimeleri arasındaki ilişki, bir yazılım için tam olarak aynı şey olabilir. Dolayısıyla bu gibi istisnai durumlara karşı, analiz edilen metinler için yazılım tarafından gerçekleştirilen eşleşmelerin insan gözüyle yeniden taranması ve kontrol edilmesi hata oranını düşürecektir.

eTma yazılımında bu gibi durumlar için bir tür “öğrenme” modülü de geliştirilmektedir. Bu modülün temel işlevi, yapılan kelime eşleşmelerini araştırmacıya bir liste olarak sunmak ve bu liste üzerinde araştırmacıya hatalı bulunduğu eşleşmeleri işaretleme şansı vermektir. Daha sonraki analizlerde bu eşleşmeler yazılım tarafından dikkate alınmamakta, dolayısıyla bu konuda yazılıma “geçmiş tecrübelerden yararlanma” özelliği kazandırılmış olmaktadır.

Okunabilirlik analizinde Türkçe için bir ilk konumunda olan eTma tarzında bir yazılımın geliştirilme süreci, deđişkenlerin belirlenmesi ile başlar. Yazılım tarafından kabul edilecek metin format(lar)ı, bu metinler için belirlenen deđişkenlerin deđerlerinin nasıl hesaplanacağı ya da veritabanı bağlantıları gibi teknik problemlerin çözümü, bilgisayar programlama konusunda uzmanlık gerektiren konulardır. Yazılımın kullanımı, okunabilirlik formüllerinde yer alan metin karakteristik deđişkenlerinin farklı metinler için aldıkları deđerleri ölçmede araştırmacıya önemli bir hız kazandıracaktır. Bilgisayar destekli metin analizinin ilk aşamasını oluşturan bu sürecin ardından, yazılım, araştırmanın bir ürünü olarak geliştirilecek okunabilirlik formülünün tanımlanmasına olanak tanyacaktır. Bu şekildeki özelleştirilebilir yapısı sayesinde yazılımın, formül geliştirme sürecinin ardından, geliştirilen formüllerin kullanıldığı araştırmalarda da kullanım alanı bulması öngörülmektedir.

Sonuç

Okunabilirlik araştırmalarının da içinde yer aldığı, genel olarak anlambilim üzerine yapılan deneysel çalışmalar, oldukça yoğun işgücünü ve fazla sayıda katılımcıyı gerektiren çalışmalardır. Bu çalışmalarda veriler, katılımcıların yanı sıra ele alınan metinlerin analiz edilmesi ile de sağlanır. Dolayısıyla bu aşamada, hata yapılması olasılığı oldukça yüksek olacaktır.

Bilgisayar desteğinin zorunlu hâle geldiđi aşama da bu analiz aşamasıdır. Uzun metinlerde, özellikle kelime analizlerinin gerçekleştirilmesi, Türkçenin de içinde bulunduğu eklemeli dillerde daha da güçleşmektedir. Bu gibi durumlarda, yapısal olarak Türkçeden çok farklı diller olan İngilizce ya da Almanca metinler üzerinde çalışan yazılımların Türkçe metinler üzerinde çalıştırılması sonuç vermeyecektir. Türkçe üzerine özel olarak tasarlanmış yazılımların geliştirilmesi bir zorunluluktur. Diđer bir deyişle, paket programların kullanımı yeterli olmayacaktır; araştırmaya özel, zeki ve duyarlı özelliklere sahip yazılımların geliştirilmesi gereklidir. Bu türdeki yazılımlar ise, bilimsel olarak projelendirilmiş araştırmaların ürünleri olarak ortaya çıkacaklardır. Bu anlamda, sayıları yetersiz olan eTma tarzındaki yazılımların;

1. Genel olarak dil bilimi, özellikle de anlam bilim alanlarında gerçekleştirilecek araştırmalarda araştırmacılara ölçme aracı anlamında önemli bir destek sağlamaları açısından,

2. Araştırmalar sonucunda elde edilen okunabilirlik formülleri gibi ölçme yöntemlerinin uygulanmaları aşamasında doğrudan alan uygulamacılarına sağlayacakları kolaylıklar açısından

sayılarının çoğalması ve desteklenmeleri önem taşımaktadır.

Kaynaklar

- Anagnostou, N. K., Weir, G. R. S. (2006). From corpus-based collation frequencies to readability measure. *ICT in the Analysis, Teaching and Learning of Languages*, Preprints of the ICTATLL Workshop. Glasgow, 33-46.
- Ateşman, E. (1997). Türkçe’de okunabilirliğin ölçülmesi. *A.Ü. TÖMER Dil Dergisi*, 58, 171-174.
- Chall, J. S. (1988). The beginning years. İçinde: Zakaluk, B. L., Samuels, S. J. (Ed.), *Readability: Its Past, Present ve Future*, 2-13, Newark, Delaware: IRA.
- Çiftçi, Ö., Çeçen, M. A., Melanlıoğlu, D. (2007). Altıncı sınıf ders kitaplarındaki metinlerin okunabilirlik açısından değerlendirilmesi. *Elektronik Sosyal Bilimler Dergisi*, 6(22), 206-219.
- Dale, E., Chall, J. S. (1948). A formula for predicting readability. *Educational Research Bulletin*, January 21 ve February 17, 27:1-20, 37-54.
- Dale, E., Chall, J. S. (1949). *The concept of readability*. Elementary English 26:23.
- Dubay, W. H. (2004). *The principles of readability*. U.S.: Impact Information.
- Johnson, K. (2000). *Readability*. <http://www.timetabler.com/reading.html> adresinden 28.03.2008 tarihinde indirilmiştir.
- Klare, George R. (1963). *The measurement of readability*. Ames, Iowa: Iowa State University Press.
- Klare, George R. (1988). The formative years. In B. L. Zakaluk , S. J. Samuels (Eds.), *Readability: Its Past, Present ve Future* (pp. 24-34), Newark, Delaware: IRA.
- McLaughlin, G. H. (1969). SMOG Grading - A new readability formula. *Journal of Reading*, 22, 639-646.
- McLaughlin, G. H. (1974). Temptation of the flesch. *Journal of Instructional Science*, 2(4), 367- 383.

- Tekbıyık, A. (2006). Lise Fizik I ders kitabının okunabilirliđi ve hedef yař düzeyine uygunluđu. *Kastamonu Eđitim Dergisi*, 14(2), 441-446.
- Temur, T. (2006). *İlköđretim 4 ve 5. sınıf öđrencilerinin yazı dilindeki kelime hazinelerinin bazı deđişkenler açısından incelenmesi*. Yayınlanmamıř doktora tezi, Gazi Üniversitesi, Eđitim Bilimleri Enstitüsü, Ankara.
- Weir, G., Ritchie, C. (2006). Estimating readability with the strathclyde readability measure. *ICT in the Analysis, Teaching ve Learning of Languages*, Preprints of the ICTATLL Workshop 2006. Glasgow, 25-32.
- Zorbaz, K. Z. (2007). Türkçe ders kitaplarındaki masalların kelime – cümle uzunlukları ve okunabilirlik düzeyleri üzerine bir deđerlendirme. *Eđitimde Kuram ve Uygulama*, 3(1), 87-101.

Summary

COMPUTER ASSISTED TEXT READABILITY ANALYSIS

Tolga GÜYER*

Turan TEMUR**

Ebru SOLMAZ***

The first academic researches related to the readability concept which can be defined as the level of comprehensibility by readers of a certain level coincides with 1920s. It is seen that the researches conducted on readability and the way the readability concept has been handled are universal in their general frameworks. But when it is taken into consideration in detail and considered to be dependent on the language, it is realized that this concept is shaped by the grammatical features of the language used, by the social and cultural structure of the society using this language, by the richness of the language and its being used by people at different ages and educational levels. In other words, besides being an internationally accepted concept, readability has a structure unique to the language used.

The structure of the language used is very important when the concepts which are proper to language are investigated as readability. If we think that this kind of studies are began for English, and Turkish is quite different language from English structurally, than we realize the fact that the methods used for these studies have to be revised for Turkish texts' readability investigations. Unlike English, most of the linguistic structures are formed by additions in Turkish. Therefore, Turkish is among languages which are difficult in terms of semantic analysis. This situation is also an uneasy process for the scientists who study on computer sciences. The assortment of the additions and difficulty in finding the roots of the words makes this process quite complex for computers as well as humans. Sometimes, when it is required to decide the root of the word, that is, to find the position where the first addition starts, it could be a necessity to use a dictionary database for computer programs. Some of the words have to be analyzed according to their usage in the

Address for correspondence: *Yrd. Doç. Dr., Gazi Üniversitesi, Gazi Eğitim Fakültesi, Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü, tguyer@gazi.edu.tr; **Yrd. Doç. Dr., Dumlupınar Üniversitesi, İlköğretim Bölümü, Sınıf Öğretmenliği Anabilim Dalı; ***Bilgisayar Öğretmeni, Ankara Kazan İlköğretim Okulu.

context. In such situations, it is inevitable to manipulate the results by the researchers. For this reason, although the texts are analyzed by a computer, contribution of the researcher to the process or to the results is inevitable.

If we consider these difficulties in terms of our subject, readability, it is necessary to determine the root of the words for calculating the ratio of the usage of the classified words. Consequently, the algorithms of the programs have to be good enough because of the abovementioned reasons.

The most detailed definition of readability is made by Dale and Chall's (1948, 1949). Dale and Chall defined readability as the sum of all elements related to the printed material that affects the success the reader group achieves using the material. Success is a result of readers' understanding, reading at an appropriate level and finding the material interesting.

General assumption in readability is that if a text is composed of short sentences and words, it is read and understood easily by the average reader. Based on this opinion, readability formulas generally focus on text features such as average word length and average sentence length. The most famous formula readability formula that uses these variables was published by Rudolf Flesch in 1948. Flesch's formula was obtained by the comprehensibility points got by making the readers read certain texts, by subjecting the texts' ASL (average sentence length) and ASW (average number of syllables per word) values into regression analysis and expressed as:

$$\text{Flesch Reading Ease Score} = 206.835 - 1.015 \text{ ASL} - 84.6 \text{ ASW}$$

Although the assumption that a text composed of short sentence and words is easily read and understood by an average reader is true for many texts, it won't be valid for the texts which seem appropriate in terms of these variables but which are composed of incomprehensible words for the target audience. Although these kinds of texts give good results in readability measurements, many of the readers will think that these are difficult to understand. In addition to this, there may be many words that are short but not used widely while there may be long words used broadly (Weir and Calum, 2006). In other words, the number of syllables and words is not the only determinant of readability. It is clear that the terms, foreign words, difficult words, tangible and intangible words, etc. will affect the readability of the text and they are evaluated among the variables that need to take part in formulas (Ateşman, 1997). For example, the text may be composed of the most used words for a reader group with certain vocabulary level (Temur, 2006). That the given text is

composed of these words shows that the text will be understood more easily by that group level. These word groups belonging to the field will be in the scope of field specific text. As field specific words and terms are among the factors that affect readability of the text, whether the reader knows the field specific words or not will directly affect his/her comprehensibility level.

On the other hand, it is impossible to handle a text on any printed material only by its linguistic features. Writing gains features such as font, font size, and color by the usage of technology and it constitutes a composition by various objects like drawings, tables, maps, charts existing besides writing. These formal and visual text features have been examined under the title of “legibility”. As a result, when the concept of readability is evaluated under current conditions, we can conclude that readability consists of two main concepts: legibility and comprehensibility.

Readability formulas are the products of the studies made in the field of readability. They were developed for practical use, may have variables that are impossible to calculate manually in long texts. In this situation, the word-process software that can be used will also be limited and restrictive. The software that can be personalized and that enables making analysis by the desired formulas and variables, and this will be an important material for every kind of research to be conducted on this subject. Therefore, the software to be developed may be used for two objectives:

1. To perform the process of applying formulas, which are obtained at the end of the studies, to texts in a computer-aided way, or
2. To be used as a research material during the process in readability researches.

The electronic Turkish text analyze software, which can be used in applying current readability formulas to texts as well as in the studies to be made on this field and the program, is currently under construction.