

**YAZININ DOĞAL GELİŞİMİNE BİLGİ KURAMI TEMELİNDE
FARKLI BİR BAKIŞ (TÜRKÇE ÖRNEĞİNDE)
A DIFFERENT VIEW ON THE NATURAL DEVELOPMENT OF
WRITING ON THE BASIS OF INFORMATION THEORY
(IN THE CASE OF TURKISH)**

Çağlayan YILMAZ*
Yasin SANCAR**

Öz

İletilerdeki karakterlerin (harflerin veya rakamların) yazı dizgesindeki işlevlerini onların kullanım sıklıklarıyla ve buna bağlı olarak bilgi değerleriyle ilişkilendiren bilgi kuramı, bir yandan ölçülebilir (ve karşılaştırılabilir) veriler sunarken diğer yandan yazı dizgesinin gelişimi üzerine öngöründe bulunma imkânı sağlıyor. Ancak bunun için öncelikle standart bir alfabedeki karakterlerin iletilerdeki kullanım sıklıklarını gösteren veriler gerekiyor. Bu sebeple standart Türk alfabesindeki karakterlerin ("w", "x", "q" ve boşluk karakteri (-) ile birlikte) ve karakter permütasyonlarının (2'li, 3'lü, 4'lü, 5'li, 6'lı, 7'li ve 8'li permütasyonlar) son 10 yılda yayımlanmış 100 adet metinden seçilen pasajlarla oluşturulan derlemdaki kullanım sıklıkları tespit edildi. Sonrasında bu sıklık değerleri üzerinden iletilerde kullanılan her karakter ve karakter bloğunun bilgi değeri hesaplandı. Bilgi değerlerine ilişkin elde edilen verilerle, söz konusu karakter ve karakter bloklarının işaretlenme aşamaları açıklanmaya çalışıldı. Ulaşılan sonuçlar, Türkçedeki sözcüklerinin günümüz konuşma ve yazı dilindeki kullanımıyla ve tarihî metinlerdeki şekilleriyle ilişkilendirildi.

Anahtar Kelimeler

alfabe, harf, karakter, işaret, işaretleme, ses bilimi, bilgi kuramı, yazı dizgesi

Abstract

The information theory, which associates the functions of the characters (letters or numbers) in the messages in the writing system with their frequency of use and accordingly with the information values, provides the opportunity to make predictions on the development of the writing system on the one hand. However, this requires data that shows how often characters in a standard alphabet are used in messages. For this reason, the frequency of use of characters and character permutations (2s, 3s, 4s, 5s, 6s, 7s and 8s permutations) in the standard Turkish alphabet (with "w", "x", "q" and the space character (-)) in the compilation created by passages selected from 100 texts published in the last 10 years have been determined. Then, the information values of each character and character block used in the messages have been calculated over these frequency values. With the data obtained from

* Dr. Öğr. Üyesi, Artvin Çoruh Üniversitesi, cagom49@yahoo.com, <https://orcid.org/0000-0002-1995-212X>
Makalenin birinci (ve sorumlu) yazarıdır. Makaledeki verilerin çözümlenmesi, yorumlanması ve makalenin metin kurgusu birinci yazara aittir. Ayrıca verilerin elde edildiği derlem, birinci yazar tarafından hazırlanmıştır.

** Öğretim Görevlisi, Atatürk Üniversitesi, sancaryasin@gmail.com, <https://orcid.org/0000-0002-4200-1293>
Makalenin ikinci yazarıdır. Derlemdaki istatistikî veriler, ikinci yazar tarafından hazırlanmıştır.

Gönderim Tarihi: 06.01.2020
Kabul Tarihi: 27.02.2020

the information values, it has been tried to explain the stages of the character and character blocks in question. The results reached have been associated with the use of Turkish words in today's spoken and written language and their shapes in historical texts.

•

Keywords

alphabet, letter, character, code, coding, phonology, information theory, writing system



GİRİŞ

Tarih boyunca tasarlanan yazı dizgeleri ve buna bağlı olarak geliştirilen alfabeler, aslında gitgide karmaşıklaşan iletilerin (mesajların) eklemenebilir parçalarına (öğelerine) ayrıştırılma çabasıdır. Resimlerden piktogramlara, nihayetinde seçil yazılardaki fonem karşılıkları grafemlere ulaşıncaya kadar yazının gelişimini, metinlerin nicelik ve niteliğinde aramak gerekir. Benzer bir süreci, aritmetikte de gözlemlemek mümkündür. Üretim ve mülkiyetle birlikte ortaya çıkan ticarî ilişkiler ve buna bağlı olarak gerek metinlerde gerekse sayılarla yapılan işlemlerde, günden güne derinleşen düşünsel ilişkileri yazıda gösterme zorunluluğu, insanları “gösterge işaretleme” konusunda farklı yollara itmiştir.

Bir kültürel dizge olarak yazıdaki göstergeler de başlangıçta gerçeklikle kurgusal bir ilişki içindedir. Ancak kullanılmaya başlandığı andan itibaren göstergeler kontrolden çıkar ve kendi gelişim sürecini izler. Bu süreç içinde, yazı dizgesindeki göstergelerin, gerçeklikle başlangıçta kurulmuş yapay bağı gittikçe zayıflar. Bu sebeple ürün veren bir dizgedeki göstergelerin anlamını, artık onların gerçeklikle olan ilişkisi üzerinden değil, dizgedeki değeri üzerinden belirleyebiliriz (Saussure, 1998, s. 178).

Burada, çalışmanın amacına bağlı olarak, “gösterge”, dildeki “sözcük”leri ve aritmetikteki “sayı”ları karşılamak için kullanılmıştır. Benzer yaklaşımla, hedeflenen ürün olması münasebetiyle, “metin”ler ve “aritmetik işlem”ler birer iletidir. Bu durumda “harf”ler ve “rakam”lar, göstergeleri (sözcükleri ve sayıları) yazıda göstermeye yarayan karakterler olacaktır.

Söz konusu benzerliklerin ardından artık, “alfabe” kabaca “göstergeleri işaretlemekte kullanılan karakterlerin oluşturduğu küme” gibi düşünülmelidir. Sözcükleri işaretlemek üzere kullanılan alfabenin yanında evrensel olarak kullanılan, 10 rakamdan oluşan küme de bir alfabe olarak kabul edilmelidir.

Burada, alfabenin sınırlı sayıda karakterden oluşmuş bir küme olduğunun altı çizilmelidir. Bu sınırlandırma, göstergelerin (sözcüklerin veya sayıların) miktarındaki önlenemez yükselişin sonucu olarak her göstergenin ayrı bir karakterle gösterilmesinin insan hafızasında ortaya çıkardığı ağır yükü, mümkün olduğunca hafifletmek üzere gerekli görülmüştür.

Sayıların doğal dillerdeki ifadelerine bakıldığında, bu açıkça görülebilir. İnsanın iki elinde toplam, 10 adet parmak olduğundan, doğal dillerde genellikle ilk 10 sayma sayısının (ve bazı dillerde “0”ın) karşılığı olan göstergeleri görüyoruz. Daha büyük sayılar için mevcut göstergelerin eklemesi gerekiyor. Ama tarih boyunca ilk 12 sayma sayısının, belki de ayak parmaklarını ekleyerek ilk 20 sayma sayısının, hatta ilk 60 sayma sayısının, ayrı ayrı göstergelerle (ve ayrı ayrı işaretlerle) temsil edildiği dillerde (saatin 60 dakikaya, dakikanın 60 saniyeye bölünmesi) sayı ve rakam arasındaki (gösterge sayısı ile alfabedeki karakter sayısı arasındaki) ilişkinin, ihtiyaçlar nispetinde yaygın kullanılan iletilere bağlı olarak kurulduğunu gösteriyor (Ifrah, 1998, s. 11).

Başlangıçta sayı ile rakam arasında bir fark yokmuş gibi görünüyor. Günümüzde kullanılan 10 adet rakam, aynı zamanda sayıdır. İhtiyaç duyulan her sayı için bir rakam seçiliyor. Bu durum, her anlam birimi (sözcükleri oluşturan “kök” veya “ek”ler) için bir karakter kullanıldığı “kavram yazıları”na benzemektedir. Yani dildeki gösterge sayısı ile alfabedeki karakter sayısı eşit olduğundan, bir iletideki gösterge sayısı ile karakter sayısı da birbirine eşit olacaktır.

Göstergelerin parçalanarak gösterilmesi, bir dildeki gösterge sayısı ile alfabe karakter sayısı arasındaki eşitliğin bozulmasının sonucudur. Bu eşitsizlik, herhangi bir iletideki gösterge sayısı ile karakter sayısı arasındaki eşitliğin de bozulmasını zorunlu kılacaktır. Bu eşitsizlikler ters orantılıdır. Gösterge sayısının alfabe karakter sayısından çok daha fazla olması, her bir göstergenin birden fazla karakterle gösterilmesini gerektireceğinden, neticede bir iletideki karakter sayısı, söz konusu iletinin gösterge sayısından daha fazla olacaktır. Yani alfabe karakter sayısının dildeki gösterge sayısına oranı azaldıkça, iletideki karakter sayısının iletideki gösterge sayısına oranı artar.

Taban aritmetiği, aslında sayıların işaretlenmesi sürecinde, “taban”ı gösteren sayıyla, sayının rakam (karakter) sayısı arasındaki bu ters orantıyı açıkça ortaya koyar. Burada tabanı gösteren sayının, “alfabe karakter sayısı” anlamına geldiği de hatırlanmalıdır.

Tablo 1.

1. taban (alfabe karakter sayısı)	2. sayı (gösterge)	3. sayının rakam adedi (göstergedeki karakter sayısı)
2	10101000	8
5	1133	4
10	168	3
20	88	2

“Tablo 1’e bakıldığında 1. sütundaki değerler arttıkça 3. sütundaki değerler azalmaktadır. Hâlbuki 2. sütundaki sayılar farklı tabanlarda, birbirine eşit sayılardır.

Aynı ters orantı, sözcükler için de söz konusudur. Günümüz Türk alfabesinde kapanmalı sesleri karşılayan harflerdeki titreşimli-titreşimsiz karşıtlığının (“p-b, t-d, ç-c, k-g” harflerinin temsil ettiği sesler arasındaki karşıtlığın) titreşimsiz olanları kaldırıldığında harf sayısı 4 adet eksilir.

Tablo 2.

1. alfabenin harf (karakter) sayısı	2. sözcük	3. sözcüğün harf (karakter) sayısı
29	çam (bir ağaç türü)	3
25	cham (bir ağaç türü)	4

“Tablo 2”den anlaşılacağı üzere alfabe ve iletinin harf sayıları arasındaki bu ters orantı sözcükler için de geçerlidir.

İnsan, hafızasını rahatlatmak üzere alfabe karakter sayısını mümkün olduğunca sınırlandırmak ister. Bununla birlikte, artan gösterge sayısına bağlı olarak iletideki göstergelerin karakter sayısının büyümesini de önleyerek zaman ve enerji tasarrufu sağlamaya çalışır (Gemalmaz, 1982, s. 28). İşte bu yaklaşım, tüm kültürel dizgelerin olduğu gibi yazının doğal gelişiminin de anahtarıdır.

Bu süreci, basit bir örnekle somutlaştırmak mümkündür:

Örnek 1.

0’dan 10 kadar (0, 1, 2, 3, 4, 5, 6, 7, 8, 9) sayıları (sayı göstergelerini) işaretlemek üzere 2 “elemanter işaretten” (“Elemanter işaret” terimi, çalışmamızda karakter ve karakter bloklarını işaretlemek için kullanılan 0 ve 1’lerden oluşan “eklemlenebilir işaret” anlamında kullanılmıştır.) oluşan bir alfabe kullanıldığında,

$\log_2 10 = 3,32$ bit sonucu elde edilir.

Bu sayının, kendisinden büyük ve kendisine en yakın tam sayıya eşitlenmesi zorunlu

olduğundan her sayı için 4 elemanter işaret kullanılması gerekecektir.

Bu logaritmik fonksiyonun tabanı (burada 2) alfabedeki elemanter işaret sayısını, genliği (burada 10) işaretlenmesi gereken gösterge sayısını, sonuç ise (burada 3,32'den 4'e tamamlanmış olarak) her bir göstergeyi karşılamak üzere gerekli elemanter işaret sayısını göstermektedir.

Tablo 3.

sayı (gösterge)	elemanter işaretlerle gösterimi
0	0000
1	0001
2	0010
3	0011
4	0100
5	0101
6	0110
7	0111
8	1000
9	1001

"Tablo 3" incelendiğinde, 2 elemanter işaretten oluşan bir kümenin (alfabenin) 4'lü tekrarlı permütasyonu $2^4 = 16$ 'dır. Yani bu alfabedeki elemanter işaretler 4 kez kullanılarak en fazla 16 gösterge işaretlenebilir.

Anlaşılabacağı üzere 16 mümkün işaretlemeden 6'sı (1010, 1011, 1100, 1101, 1111) herhangi bir göstergeyi karşılamamaktadır. Anlamli işaretlemelerin sayısı mümkün işaretlemelerin sayısına oranlandığında,

$$\frac{\text{anlamli işaretlemeler}}{\text{mümkün işaretlemeler}} = \frac{10}{16} = 0,83 \text{ sonucuna ulaşılır.}$$

Gösterge sayısının artmasıyla birlikte zamanla azalacak olan bu oran, yazı dizgesindeki verimliliğin de düşmesine (daha açık bir ifadeyle, işaretleme sürecinde zaman ve enerji artışına) sebep olacaktır. Bu nedenle yazı dizgeleri, işaretlemek zorunda oldukları göstergelerin sayısındaki bu önlenemez artış karşısında, verimliliğini arttırmanın yollarını arayacaktır.

Öyleyse, öncelikle verimlilikten neyin kastedildiği daha açık bir ifadeyle ortaya konmalı, buna bir ölçüt belirlenmelidir. Bilgi kuramı açısından bir alfabedeki elemanter işaretlerin (ve karakterlerin) iletilerde birbirine mümkün olduğunca yakın oranlarda kullanılması, işaretleme dizgesindeki verimliliği arttıran bir ölçüttür. Hartley tarafından, alfabedeki elemanter işaretlerin eşit oranda kullanıldığı bir yazı dizgesinde bir işaretin "bilgi değeri" (enformasyonu) olarak ortaya konulan "log k" fonksiyonu, verimliliğin ulaşmayı hedeflediği değer olarak kabul edilebilir (Hartley, 1928, s. 536-540). Bu fonksiyonda "k", alfabedeki elemanter işaret sayısını ifade eder. Logaritmik fonksiyonun tabanı alfabedeki elemanter işaret sayısına eşitlendiğinden, bu fonksiyon "log_k k" olarak düşünülmelidir (Alfabedeki her bir elemanter işaretin kullanım oranının $\frac{1}{k}$ olduğu unutulmamalıdır.).

Söz konusu fonksiyondan yola çıkarak 2 elemanter işaretten oluşan bir alfabenin kullanıldığı bir yazı dizgesindeki verimliliğin, mümkün olduğunca, 1'e yaklaşmayı hedeflediği söylenebilir.

Ancak 10 adet sayının her birinin 4 elemanter işaretle gösterildiği "Tablo 3"e bakıldığında, 25 kez "0"ın, 15 kez "1"in kullanıldığı görülür. Alfabedeki elemanter işaretlerin eşit oranda kullanılmadığı böyle durumlarda, her bir işaretin bilgi değeri, onun kullanım oranı (Buna "kullanım sıklığı" veya "kullanım olasılığı" da denilebilir.) üzerinden hesaplanır (Abramson, 1963, s. 12).

$I(x)$: "x" işaretinin bilgi değeri,

$P(x)$: "x" işaretinin kullanım sıklığı,

k: alfabedeki elemanter işaret sayısı olmak üzere,

$I(x) = \log_k \frac{1}{P(x)}$ fonksiyonu kullanılır.

Bu fonksiyon kullanılarak "Tablo 3"ten,

$I(0) = 0,678$ bit ve $I(1) = 1,415$ bit sonuçlarına ulaşılır.

Ortalama bilgi değerini hesaplamak için, Shannon tarafından geliştirilen bir fonksiyonda (Shannon, 1948, s. 11), bu elemanter işaretlerin bilgi değeri kullanılarak, yazı dizgesinin verimliliği (alfabedeki her bir elemanter işaretin ortalama bilgi değeri) ölçülebilir.

$H = -\sum_{i=1}^k P_i \times \log_k P_i$ fonksiyonunda,

"H" alfabedeki her bir elemanter işaretin ortalama bilgi değeri (entropisi) olarak kabul edildiğinden, "Tablo 3"teki kullanım sıklıkları üzerinden bu değer,

$H = 0,954$ bit olarak belirlenir.

Artık yazı dizgesinin verimliliği için bir oran belirlenebilir. Hartley tarafından tüm elemanter işaretlerin eşit oranda kullanımıyla ulaşılan bilgi değerinin, $\log_k k$ fonksiyonundan, 1 olduğu biliniyor.

Bu durumda, yazı dizgesinin verimlilik oranı "V" ile gösterilirse,

$V = \frac{1-H}{1}$ oranı 0'a yaklaştıkça yazı dizgesinin verimliliğinin arttığı sonucuna varılacaktır.

Öyleyse "Tablo 3"teki işaretlemeyle ilgili ulaşılan sonuçlar ve yazı dizgesinin verimliliği bir tabloda gösterilebilir:

Tablo 4.

gösterge sayısı	:	10
göstergelerin elemanter işaret sayısı	:	4
anlamli işaretlemelemlerin mümkün işaretlemelemlere oranı	:	0,83
alfabedeki elemanter işaret sayısı	:	2
alfabedeki elemanter işaretlerin kullanım sıklığı	:	0→0,625 ve 1→0,375
alfabedeki elemanter işaretlerin bilgi değeri	:	0→0,678 bit ve 1→1,415 bit
alfabedeki her bir elemanter işaretin ortalama bilgi değeri	:	0,954 bit
yazı dizgesinin verimlilik oranı	:	0,046

Örnek 2.

Şimdi yazının, iletilere bağlı olarak, doğal gelişimine bakılabilir. İşaretlenmesi zorunlu sayı adedinin 10'dan 100'e ulaştığı düşünülerek iki farklı işaretleme yaklaşımı geliştirilebilir.

Birinci yaklaşım: Burada iki elemanter işaretlemeden oluşan alfabeyle (0 ve 1) 0'dan 100'e kadar sayıların 10 tabanında basamaklarına ayrılmış bir hâldeki gösterimleri işaretlenecektir. Çünkü "Örnek 1"deki yazı 10 adet sayıyı işaretleme üzere tasarlanmıştı. Bu durumda her sayı iki basamaktan oluştuğu için mevcut alfabeyle her sayı (iki sayı gibi düşünülüp) 8 elemanter işaretlemeyle gösterilecektir. (Bu durumda $99 (9 \times 10 + 9 \times 1 = 90 + 9 = 99)$ '10011001 (1001_1001)'

olarak gösterilmelidir.)

2 elemanter işaretten oluşan bir alfabe 8 kez kullanılarak en fazla $2^8 = 256$ gösterge işaretlenebilir. Ancak burada 100 adet gösterge bulunduğu için anlamlı işaretlemelerin mümkün işaretlemelere oranı $\frac{100}{256} = 0,39$ olacaktır.

“Tablo 4”teki “anlamlı işaretlemelerin mümkün işaretlemelere oranı”yla (0,83) karşılaştırıldığında, gösterge sayısındaki artışa bağlı olarak bu oranda belirgin bir düşüş (0,39) görülüyor. Alfabedeki elemanter işaret sayısını arttırarak bu düşüşün önüne geçmek mümkündür. Ancak bu yöntemle, gösterge sayısının sürekli artışına bağlı olarak alfabedeki elemanter işaret sayısı da sürekli artacaktır. Alfabenin işaret sayısındaki sürekli artış da zamanla hafızayı çok yoracak, alfabe var oluş amacından uzaklaşacaktır. Bu nedenle doğal dillerde artan gösterge sayısına rağmen alfabeler, karakter sayılarını sabit tutma eğilimindedirler.

İkinci yaklaşım: Öte yandan alfabedeki işaret sayısını sabit tutarak anlamlı işaretlemelerin mümkün işaretlemelere oranındaki düşüşün önüne geçme imkânının olmadığı açıktır. Ancak iki elemanter işaretten oluşan alfabedeki işaretlerinin ortalama bilgi değeri, 1’e yaklaştırılıp yazı dizgesinin verimliliği arttırılarak bu düşüş yavaşlatılabilir.

“Örnek 1”deki yazı 10 adet sayıya göre tasarlandığı için, “Örnek 2”deki “Birinci yaklaşım”da 0’dan 100’e kadar sayıların her basamağı, ayrı bir sayı (gösterge) gibi düşünülüp (10 tabanında basamaklarına ayrılmış bir hâlde) işaretlenmişti. Ancak yazı dizgesi, doğal seyri içinde gösterge sayısındaki artışa uyum sağlamaya çalışır. Sayı (gösterge) adedi 100 olduğuna göre bu kez sayılar basamaklarına ayrılmak yerine bütün olarak (iki basamaklı değil, bir basamaklı) işaretlenmelidir. Bu durumda 100 adet sayı, 100 tabanında görülüp her sayı tek bir rakamla gösterildiğinde,

$$\log_2 100 = 6,64 \text{ bit (tam sayı olarak 7) sonucu elde edilir.}$$

Her sayı için (aynı zamanda her rakam için) 7 elemanter işaret kullanılacağından, bu yaklaşımla $2^7 = 128$ mümkün işaretleme yapılabilir. İşaretlenmesi gereken 100 adet gösterge olduğuna göre,

$$\frac{\text{anlamlı işaretlemelerin sayısı}}{\text{mümkün işaretlemelerin sayısı}} = \frac{100}{128} = 0,78 \text{ oranı elde edilir.}$$

Bu oran (0,78) “Birinci yaklaşım”daki “anlamlı işaretlemelerin mümkün işaretlemelere oranı”yla (0,39) karşılaştırıldığında, düşüşün önemli ölçüde yavaşladığı görülüyor.

Söz konusu oranın düşüşündeki yavaşlama, yazı dizgesinin verimliliğindeki artışla (daha açık bir ifadeyle, iki elemanter işaretten oluşan alfabedeki elemanter işaretlerin ortalama bilgi değerinin 1’e yaklaşmasıyla) ilişkilendirilmelidir.

100 adet sayının her biri 7 elemanter işaretle gösterildiğinde 384 kez “0” ve 316 kez “1” işareti kullanılır. Öyleyse her bir elemanter işaretin kullanım sıklığı (iletideki her bir elemanter işaretin kullanım miktarının iletideki tüm elemanter işaretlerin kullanım miktarına oranı) kullanılarak

$$I(x) = \log_k \frac{1}{P(x)} \text{ fonksiyonundan,}$$

$$I(0) = 0,866 \text{ bit ve } I(1) = 1,147 \text{ bit sonuçlarına ulaşılır.}$$

Bu durumda,

$$H = -\sum_{i=1}^k P_i \times \log_k P_i \text{ fonksiyonundan,}$$

H (alfabedeki her bir işaretin ortalama bilgi değeri) 0,993 olurken H’ye bağlı olarak da

$$V = \frac{1-H}{1} \text{ oranı } 0,007 \text{ olur.}$$

Görüldüğü üzere “İkinci yaklaşım”da “alfabedeki işaretlerin ortalama bilgi değeri”, 1’e daha fazla yaklaşırken, ters orantılı olarak “yazı dizgesinin verimliliğinin oranı”, 0’a daha çok yaklaşmıştır. Yani “İkinci yaklaşım”da daha verimli bir işaretleme yapılmıştır.

Bu iki farklı yaklaşımla ulaşılan sonuçlar da bir tabloda karşılaştırılabilir:

Tablo 5.

	Birinci yaklaşım	İkinci yaklaşım
gösterge sayısı	: 100	100
göstergelerin elemanter işaret sayısı	: 8	7
anlamli işaretlemelerin		
mümkün işaretlemelere oranı	: 0,39	0,78
alfabedeki elemanter işaret sayısı	: 2	2
alfabedeki elemanter	: 0→0,625	0→0,548
işaretlerin kullanım sıklığı	: 1→0,375	1→0,452
alfabedeki elemanter	: 0→0,678 bit	0→0,866 bit
işaretlerin bilgi değeri	: 1→1,415 bit	1→1,147 bit
alfabedeki her bir elemanter işaretin ortalama bilgi değeri	: 0,954 bit	0,993 bit
yazı dizgesinin verimlilik oranı	: 0,046	0,007

“Tablo 5”ten, 100 adet göstergenin 2 elemanter işarettten oluşan bir alfabeyle işaretlendiği iki farklı yaklaşımdan ikincisinin daha verimli olduğu açıkça görülüyor. “İkinci yaklaşım”da “alfabedeki elemanter işaretlerin ortalama bilgi değeri”, 1’e daha fazla yaklaşırken, ters orantılı olarak, “yazı dizgesinin verimliliğinin oranı”, 0’a daha çok yaklaşmıştır. Yani “İkinci yaklaşım”da daha verimli bir işaretleme yapılmıştır. Çünkü “İkinci yaklaşım”da “alfabedeki elemanter işaretlerin kullanım sıklığı” ve “alfabedeki elemanter işaretlerin bilgi değeri” birinciye göre birbirlerine daha yakındır.

İNCELEME

Her yazı dizgesi “Birinci yaklaşım”la tasarlanır, ancak “İkinci yaklaşım”a göre gelişimini sürdürür. Tanımlamalarda, gösterge olması bakımından sözcüklerle sayılar aynı kategoride değerlendirilirken sayılardan yola çıkılarak ifade edilmeye çalışılan bu iki yaklaşım farkının, sözcükler için de geçerli olduğunun gösterilmesi amaçlanmıştır.

Bu çalışmada, daha önce Türkçe metinlerden oluşturulan derlemdeki 33 karakterin (Türk alfabesindeki 29 harf ile “q, x, w” harfleri ve boşluk karakteri (“-” ile gösterilmiştir.)) 1’li, 2’li, 3’lü, 4’lü, 5’li, 6’lı, 7’li, 8’li karakter permütasyonlarının (kendi kümeleri içinde) kullanım sıklığı belirlendi (“Karakter” terimi, “göstergenin basamaklarında karşılaştırılması muhtemel işaret” olarak “elemanter işaret”ten farklı bir anlamda kullanılmıştır.). Bu kullanım

sıklıklarından hareketle her karakter permütasyonunun bilgi değeri, 2 tabanında (“bit” cinsinden) hesaplandı (Söz konusu derlem “Türkçede Anlam Birimlerinin Bilgi Kuramı Temelinde İşaretlenmesi” başlıklı çalışma için, özellikle son 10 yılda yayımlanmış konuşma diline yakın 100 adet metinden seçilmiş parçalardan oluşturulmuştu (söz konusu çalışma için hazırlanmış olan derlemin niteliği için bk. Yılmaz, 2018, s. VI).).

Öyleyse şimdi metinlerden rastgele seçilen sözcükler önce 1’li, sonra 2’li, daha sonra 3’lü vb. (uzunluğuna göre) basamak bloklarına ayrılıp her bloğun bilgi değeri (ilgili kümeden) bulunabilir ve daha sonra bu blokların bilgi değerlerinin toplamı karşılaştırılabilir.

Örnek 3.

Burada, derlemde en sık kullanılan sözcük olan “bir” sözcüğü, basamak bloklarına ayrılarak işaretlenmiştir. Ancak boşluk karakteriyle birlikte (Yazıda ‘boşluk’ (-) da, “.”, “,”, “;”, “?”, “!”, ..., gibi bir ‘noktalama işareti’dir. Örn.: “hanımeli” ≠ “hanım eli”) bu sözcüğün 4 karakterden (bir-) oluştuğuna dikkat edilmeli. Bu durumda 4 karakterden (sayılarla kıyasarsak 4 basamaktan) oluşan bir gösterge ancak 1’li, 2’li ve 4’lü bloklar hâlinde işaretlenebilir. Bu blokların, kullanım sıklıklarına göre bilgi değeri aşağıdaki gibidir:

1’li bloklar hâlinde işaretlemede:

“b”	→	5,05’ten 6 bit	blokların toplam bilgi değeri	:	15,78 bit
“i”	→	3,74’ten 4 bit	bloktaki elemanter işaretlerin		
“r”	→	4,12’den 5 bit	toplam bilgi değeri	:	18 bit
“-”	→	2,87’den 3 bit	aradaki fark	:	2,22 bit

2’li bloklar hâlinde işaretlemede:

“bi”	→	7,06’dan 8 bit	blokların toplam bilgi değeri	:	13,08 bit
“r-”	→	6,02’den 7 bit	bloktaki elemanter işaretlerin		
			toplam bilgi değeri	:	15 bit
			aradaki fark	:	1,92 bit

4’lü blok hâlinde işaretlemede:

“bir-”	→	8,32’den 9 bit	blokların toplam bilgi değeri	:	8,32 bit
			bloktaki elemanter işaretlerin		
			toplam bilgi değeri	:	9 bit
			aradaki fark	:	0,68 bit

“Örnek 3”te bloklardaki basamak sayısı arttıkça gösterge (burada “bir” sözcüğü) daha az işaretle karşılanabiliyor. Nihayetinde sözcük tek bir blok (4’lü blok) olarak işaretlendiğinde, 1’li bloklar hâlindeki işaretleme göre $18 - 9 = 9$ bit tasarruf ediliyor.

Öte yandan “blokların toplam bilgi değeri” ile “bloktaki elemanter işaretlerin toplam bilgi değeri” arasındaki farkın bloktaki basamak sayısı artışına bağlı olarak azaldığına da dikkat edilmelidir. (1’li bloklarda 2,22 bit; 2’li bloklarda 1,92 bit; 3’lü bloklarda 0,68 bit)

Ancak blokların basamak sayısındaki artış, her zaman verimli bir işaretleme sağlamayabilir. “Örnek 2”de 100 adet sayının iki ayrı basamakta işaretlendiği “Birinci yaklaşım”da her basamağın tabanı aynıydı. Ancak çoğu zaman yazı dizgelerinde göstergelerdeki basamakların tabanı aynı olmaz. Yani göstergenin her basamağına gelmesi muhtemel karakter sayısı eşit olmayabilir. Saatlerdeki, takvimlerdeki, elektrik sayaçlarındaki ve bilgisayar belleklerinin veri

dökümündeki göstergeler gibi daha pek çok gösterge böyledir.

İşaretlenmesi amaçlanan bloktaki basamak sayısına uygun karakter permütasyonlarından kullanım sıklığı yüksek olanlar, işaretlemede verimliliği arttırırken kullanım sıklığı düşük olanlar ise verimliliği azaltacaktır. Çünkü kullanım sıklığı ile bilgi değeri ters orantılıdır. Yani bir gösterge, kullanım sıklığı düşük karakter permütasyonu (veya karakter permütasyonları) içeriyorsa göstergeyi oluşturan bloklardaki basamak sayısı artmış olsa bile verimli bir işaretleme mümkün olmaz.

Bu durum bir örnekle somutlaştırılabilir.

Örnek 4.

Aşağıda “sonra” sözcüğü (“sonra-” şeklinde altı basamak kabul edilerek) 1’li, 2’li, 3’lü ve 6’lı basamak bloklarına ayrılarak, derlemdeki kullanım sıklıklarına göre her karakter bloğunun bilgi değeri, ilgili kümeden bulunmuş ve bu değerler toplanarak işaretlemenin verimliliği karşılaştırılmıştır:

1’li bloklar hâlinde işaretlemede:

“s”	→ 5 bit	blokların toplam bilgi değeri	: 24,77 bit
“o”	→ 5,56’dan 6 bit	bloktaki elemanter işaretlerin	
“n”	→ 3,96’dan 4 bit	toplam bilgi değeri	: 27 bit
“r”	→ 4,12’dan 5 bit	aradaki fark	: 2,23 bit
“a”	→ 3,26’dan 4 bit		
“-”	→ 2,87’dan 3 bit		

2’li bloklar hâlinde işaretlemede:

“so”	→ 9,19’dan 10 bit	blokların toplam bilgi değeri	: 26,17 bit
“nr”	→ 11,03’ten 12 bit	bloktaki elemanter işaretlerin	
“a-”	→ 5,95’ten 6 bit	toplam bilgi değeri	: 28 bit
		aradaki fark	: 1,83 bit

3’lü blok hâlinde işaretlemede:

“son”	→ 10,16’dan 11 bit	blokların toplam bilgi değeri	: 19,98 bit
“ra-”	→ 9,82’dan 10 bit	bloktaki elemanter işaretlerin	
		toplam bilgi değeri	: 21 bit
		aradaki fark	: 1,02 bit

6’lı blok hâlinde işaretlemede:

“sonra-”	→ 11,48’dan 12 bit	blokların toplam bilgi değeri	: 11,48 bit
		bloktaki elemanter işaretlerin	
		toplam bilgi değeri	: 12 bit
		aradaki fark	: 0,52 bit

Ancak “Örnek 4”e dikkat edilecek olursa “sonra” sözcüğünü “1’li blok hâlinde işaretleme”de 27 bitken “2’li blok hâlinde işaretleme”de 28 bittir. Yani bloktaki basamak sayısının artışı işaretlemeyle daha verimli kılınmıştır. Çünkü “2’li blok hâlinde işaretleme”de kullanım sıklığı düşük (buna bağlı olarak bilgi değeri yüksek) bir blok (“nr” bloğu) vardır.

$I(nr) < I(n) + I(r)$ olması beklenirken,

11,06 bit > 3,969 bit + 4,127 bit eşitsizliği "nr" bloğunun bir bütün hâlinde işaretlenmemesi gerektiğini gösteriyor.

Aynı tabana sahip olmayan basamaklardan oluşan göstergelerde (Doğal dillerin göstergeleri yazıda böyle işaretlenir.) karakter bloklarının çoğu zaman kullanım sıklığı düşük permütasyonlardan oluşması sık karşılaşılan bir durumdur. Zaten yazı dizgesi, doğal gelişiminde kullanım sıklığı yüksek karakter permütasyonlarını bir bütün olarak görür ve bir bütün olarak (karakter bloğunu birleştirip tek bir karakter gibi) işaretler. Yani bir karakter sürekli aynı karakter (veya karakter bloğu) ile birlikte kullanılıyorsa, sürekli birlikte kullanılan karakterlerin (veya karakter bloklarının) birleşerek tek bir karaktere (ligature) dönüşmesi kaçınılmazdır.

Ancak yazı dizgesi, bir göstergeyi yukarıda ("Örnek 3" ve "Örnek 4"te) olduğu gibi 1'li veya 2'li ya da 3'lü vb. eşit sayıda basamaktan oluşmuş bloklara ayırarak işaretlemeyebilir. Yani kullanım sıklığı yüksek 1'li, 2'li, 3'lü vb. karakter blokları aynı gösterge içinde en verimli işaretlemeyi sağlamak amacıyla bir arada bulunabilir.

Bu durum bir örnekle daha anlaşılır kılınabilir.

Örnek 5.

Burada "söyle" sözcüğü (emir kipi ikinci tekil kişi çekimiyle "söyle-" şeklinde altı basamaklı), bloklarına ayrılıp işaretlenmiştir. Ancak bu kez eşit sayıda bloklara ayrılmamıştır. En verimli işaretlemenin bulunabilmesi için öncelikle mümkün tüm blokların bilgi değerinin ilgili kümeden bulunması gerekmektedir.

1'li karakter bloklarının bilgi değeri ve bloğu işaretlemek için kullanılan elemanter işaretlerin bilgi değeri:

"s"	→	5,167'den 6 bit
"ö"	→	7,163'ten 8 bit
"y"	→	5,073'ten 6 bit
"l"	→	4,194'ten 5 bit
"e"	→	3,615'ten 4 bit
"-"	→	2,876'dan 3 bit

2'li karakter bloklarının bilgi değeri ve bloğu işaretlemek için kullanılan elemanter işaretlerin bilgi değeri:

"sö"	→	10,643'ten 11 bit
"öy"	→	10,153'ten 11 bit
"yl"	→	8,917'den 9 bit
"le"	→	6,343'ten 7 bit
"e-"	→	5,749'dan 6 bit

3'lü karakter bloklarının bilgi değeri ve bloğu işaretlemek için kullanılan elemanter işaretlerin bilgi değeri:

"söy"	→	11,335'ten 12 bit
"öyl"	→	10,228'den 11 bit
"yle"	→	9,568'den 10 bit
"le-"	→	8,656'dan 9 bit

4'lü karakter bloklarının bilgi değeri ve bloğu işaretlemek için kullanılan elemanter işaretlerin bilgi değeri:

- “söyl” → 11,335’den 12 bit
 “öyle” → 10,254’ten 11 bit
 “yle-” → 10,267’den 11 bit

5’li karakter bloklarının bilgi değeri ve bloğu işaretlemek için kullanılan elemanter işaretlerin bilgi değeri:

- “söyle” → 11,392’den 12 bit
 “öyle-” → 11,203’ten 12 bit

6’lı karakter bloğunun bilgi değeri ve bloğu işaretlemek için kullanılan elemanter işaretlerin bilgi değeri:

- “söyle-” → 15,451’den 16 bit

Başlangıçta her karakter ayrı ayrı işaretlenirken, zamanla bazı karakterlerin birleştirilmesiyle karakter bloklarının (ligature) oluştuğu ve söz konusu blokların bir karakter gibi düşünülerek işaretlendiği ve yazı dizgesinin doğal gelişiminin bu şekilde ilerlediği daha önce belirtilmişti. Örneğin 6 basamaktan oluşan “söyle-” sözcüğünün zamanla iki karakteri birleştirilerek bir blok oluşturulur. Söz konusu bloktaki birleştirilmiş karakterler bir karakter (ligature) gibi işaretlenir. Böylece basamak sayısı 6’dan 5’e iner. Sonraki adımlarda aynı işlem tekrarlanır. Yani her adımda yeni bir blok oluşturularak basamak sayısı sırasıyla 4’e, 3’e, 2’ye ve nihayet 1’e indirilir.

Göstergenin (daha doğrusu göstergedeki karakter ve karakter bloklarının) kullanım sıklığına bağlı olarak, aşamalı ve kendiliğinden gerçekleşen bu süreçte, yazı dizgesi her zaman verimliliğini en fazla arttıracak karakter bloklarını seçer. Yani kullanım sıklığı en fazla olan (buna bağlı olarak bilgi değeri en düşük olan) karakter blokları işaretlemede öncelikli tercih edilir.

Öyleyse, en verimli işaretlemeye ulaşmak üzere 6 basamaktan oluşan “söyle-” sözcüğündeki basamak sayısı aşamalı olarak azaltılırken seçilmesi gereken karakter bloklarının tespit edilebilmesi için tüm işaretleme seçeneklerinin listelenmesi zorunludur.

Yukarıda “söyle-” sözcüğünü oluşturan 6 karakterin bilgi değerleri toplandığında,

$$\dot{I}(s) + \dot{I}(\ddot{o}) + \dot{I}(y) + \dot{I}(l) + \dot{I}(e) + \dot{I}(-) = 28,088 \text{ bit olduğu görülür.}$$

Şimdi aşamalı olarak “söyle-” sözcüğünün basamak sayısı azaltılabilir.

Birinci aşama: Bu aşamada basamak sayısını 5’e indirmek için en uygun 2’li karakter bloğu tespit edilecektir. Bunun için 5 basamaklık tüm seçeneklerin (toplam bilgi değerleriyle eşleştirilerek) listelenmesi gerekmektedir.

1. seçenek → $\dot{I}(s\ddot{o}) + \dot{I}(y) + \dot{I}(l) + \dot{I}(e) + \dot{I}(-) = 26,401 \text{ bit,}$
 2. seçenek → $\dot{I}(s) + \dot{I}(\ddot{o}y) + \dot{I}(l) + \dot{I}(e) + \dot{I}(-) = 26,005 \text{ bit,}$
 3. seçenek → $\dot{I}(s) + \dot{I}(\ddot{o}) + \dot{I}(yl) + \dot{I}(e) + \dot{I}(-) = 27,738 \text{ bit,}$
 4. seçenek → $\dot{I}(s) + \dot{I}(\ddot{o}) + \dot{I}(y) + \dot{I}(le) + \dot{I}(-) = 26,622 \text{ bit,}$
 5. seçenek → $\dot{I}(s) + \dot{I}(\ddot{o}) + \dot{I}(y) + \dot{I}(l) + \dot{I}(e-) = 27,346 \text{ bit.}$

Yukarıdaki seçeneklerden en verimli işaretleme 2. seçenektir. Çünkü 2. seçenekteki “öy” bloğu en az bilgi değerine (26,005 bit) sahiptir. Öyleyse “öy” karakter bloğundaki karakterler birleştirilerek bir karakter gibi işaretlenecektir. Bu işaretleme, sonraki aşamada kullanılacaktır.

İkinci aşama: Bu aşamada en verimli işaretlemeyle basamak sayısını 4'e indirmek için 4 basamaklık tüm seçenekler (toplam bilgi değerleriyle eşleştirilerek) listelenecektir.

1. seçenek	→	$\dot{I}(s\ddot{o}) + \dot{I}(y\dot{l}) + \dot{I}(e) + \dot{I}(-) = 26,051$ bit,
2. seçenek	→	$\dot{I}(s\ddot{o}) + \dot{I}(y) + \dot{I}(le) + \dot{I}(-) = 24,935$ bit,
3. seçenek	→	$\dot{I}(s\ddot{o}) + \dot{I}(y) + \dot{I}(l) + \dot{I}(e-) = 25,659$ bit,
4. seçenek	→	$\dot{I}(s) + \dot{I}(\ddot{o}y) + \dot{I}(le) + \dot{I}(-) = 24,539$ bit,
5. seçenek	→	$\dot{I}(s) + \dot{I}(\ddot{o}y) + \dot{I}(l) + \dot{I}(e-) = 25,263$ bit,
6. seçenek	→	$\dot{I}(s) + \dot{I}(\ddot{o}) + \dot{I}(yl) + \dot{I}(e-) = 26,996$ bit,
7. seçenek	→	$\dot{I}(s\ddot{o}y) + \dot{I}(l) + \dot{I}(e) + \dot{I}(-) = 22,02$ bit,
8. seçenek	→	$\dot{I}(s) + \dot{I}(\ddot{o}yl) + \dot{I}(e) + \dot{I}(-) = 21,886$ bit,
9. seçenek	→	$\dot{I}(s) + \dot{I}(\ddot{o}) + \dot{I}(yle) + \dot{I}(-) = 24,774$ bit,
10. seçenek	→	$\dot{I}(s) + \dot{I}(\ddot{o}) + \dot{I}(y) + \dot{I}(le-) = 26,059$ bit.

Yukarıdaki seçeneklerden 1, 2, 3, 6, 9 ve 10 numaralı seçenekler "öy" bloğunu içeren bir karakter bloğuna sahip olmadığından, öncelikle değerlendirme dışında tutulacaktır. Diğer seçeneklerin (4, 5, 7 ve 8 numaralı seçenekler) içinden 8. seçenek en az bilgi değerine sahip olduğundan (21,886 bit) en verimli işaretlemedir. Burada elde edilen "öyl" bloğu sonraki aşamada kullanılacaktır.

Üçüncü aşama: Bu aşamada en verimli işaretlemeyle basamak sayısını 3'e indirmek için 3 basamaklık tüm seçenekler (toplam bilgi değerleriyle eşleştirilerek) listelenecektir.

1. seçenek	→	$\dot{I}(s\ddot{o}) + \dot{I}(yl) + \dot{I}(e-) = 25,309$ bit,
2. seçenek	→	$\dot{I}(s\ddot{o}y) + \dot{I}(le) + \dot{I}(-) = 20,554$ bit,
3. seçenek	→	$\dot{I}(s\ddot{o}y) + \dot{I}(l) + \dot{I}(e-) = 21,278$ bit,
4. seçenek	→	$\dot{I}(s\ddot{o}) + \dot{I}(yle) + \dot{I}(-) = 23,087$ bit,
5. seçenek	→	$\dot{I}(s\ddot{o}) + \dot{I}(y) + \dot{I}(le-) = 24,372$ bit,
6. seçenek	→	$\dot{I}(s) + \dot{I}(\ddot{o}yl) + \dot{I}(e-) = 21,144$ bit,
7. seçenek	→	$\dot{I}(s) + \dot{I}(\ddot{o}y) + \dot{I}(le-) = 23,976$ bit,
8. seçenek	→	$\dot{I}(s\ddot{o}yl) + \dot{I}(e) + \dot{I}(-) = 17,826$ bit,
9. seçenek	→	$\dot{I}(s) + \dot{I}(\ddot{o}yle) + \dot{I}(-) = 18,297$ bit,
10. seçenek	→	$\dot{I}(s) + \dot{I}(\ddot{o}) + \dot{I}(yle-) = 22,597$ bit.

Yukarıda, "öyl" bloğunu içeren karakter bloklarına sahip 6, 8 ve 9 numaralı seçenekler arasında en verimli işaretleme 17,826 bitle 8. seçenektir. Yani "söyl" bloğu sonraki aşamada kullanılacaktır.

Dördüncü aşama: Bu aşamada en verimli işaretlemeyle basamak sayısını 2'ye indirmek için 2 basamaklık tüm seçenekler (toplam bilgi değerleriyle eşleştirilerek) listelenecektir.

1. seçenek	→	$\dot{I}(s\ddot{o}y) + \dot{I}(le-) = 19,991$ bit,
2. seçenek	→	$\dot{I}(s\ddot{o}yl) + \dot{I}(e-) = 17,084$ bit,
3. seçenek	→	$\dot{I}(s\ddot{o}) + \dot{I}(yle-) = 20,91$ bit,
4. seçenek	→	$\dot{I}(s\ddot{o}yle) + \dot{I}(-) = 14,268$ bit,
5. seçenek	→	$\dot{I}(s) + \dot{I}(\ddot{o}yle-) = 16,37$ bit.

Yukarıda, "söyl" bloğunu içeren karakter bloklarına sahip 2 ve 4 numaralı seçenekler arasında en verimli işaretleme 14,268 bitle 4. seçenektir. Artık "söyle" bloğu sonraki aşamada kullanılacaktır.

Beşinci aşama: Artık basamak sayısının 1'e indirilmesi için 4. seçenekteki işaretlemenin bilgi değerinin 1 basamaktan oluşan tek işaretleme seçeneğinin bilgi değeriyle karşılaştırılması gerekmektedir.

$$1. \text{ seçenek} \rightarrow \dot{I}(\text{söyle-}) = 15,451 \text{ bit.}$$

Görüldüğü gibi, "Beşinci aşama"nın tek seçeneğinin bilgi değeri (15,451 bit) "Dördüncü aşama"nın en verimli seçeneğinin (4. seçenek) bilgi değerinden (14,268 bit) büyüktür. Yani tek basamaklı bir işaretleme verimlilik sağlamıyor. Öyleyse "söyle-" sözcüğü, en verimli "söyle" ve "-" şeklinde iki basamak olarak işaretlenebilir.

En verimli işaretlemeye ulaşmak üzere takip edilen aşamaların sonuçları aşağıdaki gibi özetlenebilir:

$$\begin{aligned} 6 \text{ basamaklık işaretleme} &\rightarrow 28,088 \text{ bit,} \\ \text{Birinci aşama (2. seçenek)} &\rightarrow \dot{I}(s) + \dot{I}(\text{öy}) + \dot{I}(l) + \dot{I}(e) + \dot{I}(-) = 26,005 \text{ bit,} \\ \text{İkinci aşama (8. seçenek)} &\rightarrow \dot{I}(s) + \dot{I}(\text{öyl}) + \dot{I}(e) + \dot{I}(-) = 21,886 \text{ bit,} \\ \text{Üçüncü aşama (8. seçenek)} &\rightarrow \dot{I}(\text{söyl}) + \dot{I}(e) + \dot{I}(-) = 17,826 \text{ bit,} \\ \text{Dördüncü aşama (4. seçenek)} &\rightarrow \dot{I}(\text{söyle}) + \dot{I}(-) = 14,268 \text{ bit.} \end{aligned}$$

Görüldüğü gibi basamak sayısı azaldıkça (buna bağlı olarak her basamakta birleşen karakter sayısı arttıkça) sözcüğü oluşturan karakterlerin toplam bilgi değeri azalmaktadır.

Artık, takip edilen aşamalarda, "söyle-" sözcüğünü oluşturan karakter bloklarının bilgi değerlerinin toplamıyla karakter bloklarının elemanter işaretlerinin bilgi değerlerinin toplamı ("Örnek 5"ten bulunup) karşılaştırılabilir.

Tablo 6.

1. işaretleme aşamaları	2. karakterlerin ve karakter bloklarının toplam bilgi değeri	3. karakter ve karakter bloklarının Elemanter işaretlerinin toplam bilgi değeri	4. "3. sütun"la "2. sütun" arasındaki fark
6 karakterlik işaretleme	28,088 bit	32 bit	3,912 bit
Birinci aşama (2. seçenek)	26,005 bit	29 bit	2,995 bit
İkinci aşama (8. seçenek)	21,886 bit	24 bit	2,114 bit
Üçüncü aşama (8. seçenek)	17,826 bit	19 bit	1,174 bit
Dördüncü aşama (4. seçenek)	14,268 bit	15 bit	0,732 bit

"Tablo 6"dan da anlaşılacağı üzere basamak sayısı azaldıkça (buna bağlı olarak ilgili basamaktaki karakter sayısı arttıkça) sözcüğü oluşturan "karakter ve karakter bloklarının elemanter işaretlerinin toplam bilgi değeri" (3. sütundaki değerler) ile sözcüğü oluşturan "karakterlerin ve karakter bloklarının toplam bilgi değeri" (2. sütundaki değerler) arasındaki fark azalmakta ve böylece işaretlemenin verimliliği artmaktadır.

Aslında yazı dizgesinde aşamalı olarak gözlenen bu süreç, konuşmada daha hızlı gerçekleşir. Zaten yazı dili de (gelişimine müdahale edilmediği sürece) konuşma dilindeki bu süreci geriden (daha yavaş) takip eder.

SONUÇ

“Örnek 5”teki “söyle-” sözcüğünün işaretlenme aşamalarına bakıldığında, ortaya çıkan bloklar yazı dilinin doğal gelişimi için olduğu gibi konuşma dilinin doğal gelişimi için de yol göstericidir.

başlangıçta	→	söyle-
Birinci aşama	→	s(öy)le-
İkinci aşama	→	s(öyl)e-
Üçüncü aşama	→	(söyl)e-
Dördüncü aşama	→	(söyle)-

Burada önemli bir ayrıntıya dikkati çekmekte fayda var. Kullanım sıklığına bağlı olarak konuşmada ortaya çıkan ses değişimleri (aslında sızıcılışma, akıcılışma, nazallaşma gibi ses zayıflamaları ve nihayet ses düşmeleri) ve ses uyumları, zamanla yazıda da kendini gösterir.

Yazılı iletilerde diğerlerine oranla daha sık karşılaşılan karakterler, kendilerinden önceki veya sonraki karakterlerle (ligature) birleşirler. Ancak bu birleşmede, bloğu oluşturan karakterlerden kullanım sıklığı yüksek karakter (bilgi değerinin az olması nedeniyle) kullanım sıklığı düşük karakterin (başka bir deyişle, bilgi değeri fazla olan karakterin) içinde kaybolur. Örneğin “söyle-” sözcüğünün işaretlenmesi sürecinde “Birinci aşama”da ortaya çıkan “öy” bloğundaki “y” karakteri, bilgi değeri kendisinden fazla olan “ö” karakterinin içinde kaybolacaktır.

$$“y” \text{ karakterinin bilgi değeri} = 5,073 \text{ bit} < “ö” \text{ karakterinin bilgi değeri} = 7,173 \text{ bit}$$

Böylece, iletilerde sıklıkla karşılaşılan karakterlerle gitgide daha az karşılaşılacaktır. Bunun sonucunda, dışarıdan bir müdahale olmadıkça, dillerin standart alfabelerindeki karakterlerinin kullanım sıklıkları ve buna bağlı olarak bilgi değerleri, ileti sayısındaki artışla birlikte, birbirine yaklaşma eğilimi gösterecektir. Yani, standart alfabedeki karakter sayısını “k” ile gösterilirse standart alfabenin ortalama bilgi değeri (entropisi) en üst sınır olan $\log k$ değerine doğru ilerleyecektir. Bunun sonucunda konuşma dilinde hissedilen ses uyumlarının yazıdaki görünümü günden güne kaybolacaktır. Aslında bu durum, her dizgenin, dıştan bir müdahale olmadığı sürece (başka bir deyişle kapalı bir dizge olduğu sürece), düzensizliğini arttırma eğiliminde olacağını söyleyen termodinamiğin ikinci yasasının sonucudur (Wiener, 1975, s. 50-74).

Yukarıdaki aşamalar takip edildiğinde “Birinci aşama”da ortaya çıkan bloğun, yazı dilinden önce, konuşma dilinde birleşme eğiliminde olduğu (“söyle-” → [sø·lɛ-] şeklinde bk. Ergenç, 2002, s. 415) görülebilir. Dilbilgisi kitaplarında tarif edilen tüm ses değişimleri ve ses uyumları da aslında dil göstergelerinin kullanım sıklığı sonucu ortaya çıkan bilgi değerleriyle, söz konusu göstergelerin “gösteren”lerinin (sesli veya yazılı) toplam bilgi değerleri arasındaki farkın azaltılması çabasından başka bir şey değildir.

Ayrıca bu aşamaların, günümüzde geniş bir coğrafyada konuşulan Türk ağız ve lehçelerinin, Standart Türkiye Türkçesi’ne yakınlığını belirlemede bir ölçüt olması da söz konusudur. Türkiye’de konuşulan Türkçe ağızlarda veya Türkiye Türkçesi’ne yakın Türk yazı dillerinde “söyle-” fiilinin söyleniş ve yazılışında öncelikle “Birinci aşama”da ortaya çıkan blokta (“öy” bloğunda) farklılıklar beklenmelidir. Örneğin “söyle-” fiilinin Doğu Trakya yerli ağızında “süle-” [sy·lɛ-] (Olçay, 1995, s. 92), Kırşehir ağızında “söle-” [sø·lɛ-] (Günşen, 2000, s. 506), Bolu-Akçakoca ağızında “sövle-” [søvlɛ-] (Derleme Sözlüğü, 1978, C. 10) şekillerinde söylenişleri vardır. Türkmencede bu fiil “sözle-” (Karşılaştırmalı Türk Lehçeleri Sözlüğü, 1991,

s. 795), Hakaşçada ise “söle-” (Naskali, Butanayev, İsina, Şahin & Koç, 2007, s. 462) şeklindedir.

Hatta bu aşamalar sözcüklerin tarihî metinlerdeki görünüşleriyle de örtüşebilir. Yani “Birinci aşama”da ortaya çıkan bloğun, tarihî metinlerde farklı işaretlenme olasılığı, diğer bloklara nispetle daha fazladır. Ayrıca “Birinci aşama”da ortaya çıkan bloğu farklı işaretleyen metinler, diğer aşamalarda ortaya çıkan blokları farklı işaretleyen metinlere göre günümüze daha yakın tarihlidir (Veya daha yakın tarihlendirilmelidir.). Örneğin “söyle-” fiilinin “öy” parçası Orta Türkçe dönemi Kıpçak Türkçesi metinlerinde “sevle-” “söle-”, “sözle-” gibi farklı şekillerde işaretlenmişken (Toparlı, Vural & Karaatlı, 2007, s. 241) Divanülûgati’t-Türk’te “sözle-” (Atalay, 1985, C. 4, s. 536) şeklindedir.

Bir sonraki aşamayla (İkinci aşama) elde edilen “öyl” bloğunun farklı işaretlendiği örneklerin de daha uzaktaki Türk yazı dillerinde bulunması muhtemeldir. Örneğin “söyle-” fiili Altaycada “söstö-” (Naskali & Duranlı, 1999, s. 158), Kırgızcada “sözdö-” (Yudahin, 1998, s. 666) şeklindedir. Türkmencede ise “söyle-” fiiliyle köken ortaklığı olan “güzel ses çıkar-” anlamındaki “sayra-” fiili de (Türkmen Dilinin Düşündürüşü Sözlüğü, 2015, s. 250) dikkat çekicidir.

Kuşkusuz sonraki aşamalarda (üçüncü ve dördüncü aşamalarda) ortaya çıkan bloklardaki karakterlerin sesletimleri ve farklı yazı dizgelerindeki görünüşleri, Türkiye Türkçesi’ne gerek coğrafi gerekse tarihî açıdan daha uzak Türk konuşma ve yazı dillerinde, çok daha kapsamlı veriler ve ayrıntılı çözümlenmelerle aranmalıdır.

SUMMARY

Writing systems designed throughout history and the alphabets developed accordingly are in fact an effort to decompose elements of increasingly complex messages. It is necessary to look at the development of writing in the quantity and quality of the texts, from pictures topic to grams and ultimately to phonemes and graphs in phonetic scripts. A similar process can be observed in arithmetic. Commercial relations with production and ownership and consequently, the necessity to show the intellectual relations deepening day by day in texts and arithmetic pushed people to different ways of “marking signs”.

The signs in the writing as a cultural system are initially in a fictional relationship with reality. However, as soon as they are used, these signs are out of control and follow their own development process. In the process, the initial artificial link of the signs in the writing system with reality is gradually weakened. For this reason, the meaning of the signs can be determined any more in a over their value, not their relationship to reality.

Depending on purpose, it has been used the “sign” to meet words in linguistics and numbers in arithmetic. With a similar approach, texts and arithmetic operations are messages, as long as they are targeted products. In this case, letters and numbers are defined as the characters displaying signs in the text.

Humankind wants to limit the number of characters in the alphabet as much as possible to ease his memory. However, it tries to save time and energy by preventing the growing number of characters in the messages due to the increasing number of signs. This approach is the key to the natural development of writing as well as all cultural systems.

For this reason, the frequency of use of characters and character permutations (2s, 3s, 4s, 5s, 6s, 7s and 8s permutations) in the standard Turkish alphabet (with “w”, “x”, “q” and the space character (-)) in the compilation created by passages selected from 100 texts published in the last 10 years have been determined. Then, the information value of each character and character block used in the messages have been calculated over these frequency values. With the data

obtained from the information values, it has been tried to explain the stages of the character and character blocks.

Because characters that are more frequently seen in written messages than others, combine with the characters before or after them. However, in this combination, the character with high usage frequency and the character with low usage frequency will be lost in it. Thus, common characters will be less common over time. As a result, the frequency of use of the characters in the messages will approach each other. Because, according to the second law of thermodynamics, the entropy of the system (uncertainty of the system) will increase unless there is an external intervention. As the entropy of the writing system increases, the sound harmonies in the spoken language will be less common in writing.

KAYNAKÇA

- Abramson, N. (1963). *Information theory and coding*. New York: Mc Graww-Hill Kitap Şirketi Yay.
- Atalay, B. (1985). *Dîvânü lûgat-it-Türk*. Ankara: Türk Dil Kurumu Yay.
- Ergenç, İ. (2002). *Konuşma dili ve Türkçenin söyleyiş sözlüğü*. İstanbul: Multilingual Yay.
- Gemalmaz, E. (1982). *Standart Türkiye Türkçesi'nin formanlarının enformatif değerleri ve bu değerlerin ihtiyaç hâlinde bu dilin gelişimine muhtemel etkileri*. Erişim adresi: <http://efrasiyap.tripod.com/yazilar/STT2.pdf>
- Günşen, A. (2000). *Kırşehir ve yöresi ağızları*. Ankara: Türk Dil Kurumu Yay.
- Hartley, R. V. L. (1928). *Transmission of information*. The Bell System Tecnicl Journal, 7, 535-563.
- Ifrah, G. (1998). *Rakamların evrensel tarihi*, (K. Dinçer, Çev.). Ankara: TÜBİTAK Yay.
- Ercilasun, A. B. (Ed.) (1991). *Karşılaştırmalı Türk lehçeleri sözlüğü*. Ankara: Kültür Bakanlığı Yay.
- Naskali, E. G., Butanayev, V., İsina, A., Şahin, E., Şahin, L. & Koç, A. (2007). *Hakasça – Türkçe sözlük*. Ankara: Türk Dil Kurumu Yay.
- Naskali, E. G. & Duranlı M. (1999). *Altayca – Türkçe sözlük*. Ankara: Türk Dil Kurumu Yay.
- Olçay, S. (1995). *Doğu Trakya yerli ağızı*. Ankara: Türk Dil Kurumu Yay.
- Saussure, F. (2013). *Genel dilbilim dersleri* (B. Vardar, Çev.). İstanbul: Multilingual Yay.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Tecnicl Journal*, 27, 1-55.
- Toparlı, R., Vural, H. & Karaatlı, R. (2007). *Kıpçak Türkçesi sözlüğü*, Ankara: Türk Dil Kurumu Yay.
- Türk Dil Kurumu (1978). *Derleme sözlüğü*. Ankara.
- Türkmenistan İlimler Akademisi (2015). *Türkmen dilinin düşündürüşlü sözlüğü*. Aşkabat.
- Wiener, N. (1975). *Emek siberetik ve toplum* (İ. Keskin, Çev.). İstanbul: Özgün Yay.
- Yılmaz, Ç. (2018). *Türkçede anlam birimlerinin bilgi kuramı temelinde işaretlenmesi*. Ankara: Gazi Kitabevi Yay.
- Yudahin, K. K. (1998), *Kırgız sözlüğü* (A. Taymas, Çev.). Ankara: Türk Dil Kurumu Yay.