

E-ISSN: 3023-820X
Kış 2025 • Winter 2025
Cilt 2 • Volume 2
Sayı 2 • Issue 2

Klasik Divan

Uluslararası Dil, Kültür ve Edebiyat Araştırmaları Dergisi (KD-UDKEAD)
International Journal of Language, Culture and Literature Researches (KD-IJLCLR)



Editör

Editor-in-Chief

Öğr. Gör. Esra AKA
(Muş Alparslan Üniversitesi)

Danışma Kurulu

Advisory Board

Prof. Dr. Ahmet Atilla ŞENTÜRK
(İstinye Üniversitesi)

Prof. Dr. Bahir SELÇUK
(Fırat Üniversitesi)

Prof. Dr. Bülent BAYRAM
(Kırklareli Üniversitesi)

Prof. Dr. Bünyamin TAŞ
(Aksaray Üniversitesi)

Prof. Dr. Ebru Burcu YILMAZ
(İnönü Üniversitesi)

Prof. Dr. Fettah KUZU
(Gaziantep Üniversitesi)

Prof. Dr. Gençay ZAVOTÇU
(Kocaeli Üniversitesi)

Prof. Dr. Halil İbrahim YAKAR
(Gaziantep Üniversitesi)

Prof. Dr. İbrahim Halil TUĞLUK
(Adıyaman Üniversitesi)

Prof. Dr. Kemal YAVUZ
(Fatih Sultan Mehmet Vakıf Üniversitesi)

Prof. Dr. Mehmet Celal VARIŞOĞLU
(Tokat Gaziosmanpaşa Üniversitesi)

Prof. Dr. Mehmet Fatih KÖKSAL
(İstanbul Kültür Üniversitesi)

Prof. Dr. Mehmet Sait ÇALKA
(Recep Tayyip Erdoğan Üniversitesi)

Prof. Dr. Neslihan İlknur KOÇ KESKİN
(Ankara Hacı Bayram Veli Üniversitesi)

Prof. Dr. Nihat ÖZTOPRAK
(Fatih Sultan Mehmet Vakıf Üniversitesi)

Prof. Dr. Nimet YILDIRIM
(Atatürk Üniversitesi)

Prof. Dr. Oğuzhan DURMUŞ
(Ankara Sosyal Bilimler Üniversitesi)

Prof. Dr. Orhan Kemâl TAVUKÇUOĞLU
(Recep Tayyip Erdoğan Üniversitesi)

Prof. Dr. Ozan YILMAZ
(Sakarya Üniversitesi)

Prof. Dr. Sadık ARMUTLU
(Ağrı İbrahim Çeçen Üniversitesi)

Prof. Dr. Serkan TÜRKOĞLU
(Tokat Gaziosmanpaşa Üniversitesi)

Prof. Dr. Şerife YALÇINKAYA
(Ege Üniversitesi)

Prof. Dr. Zeki TAŞTAN
(Van Yüzüncü Yıl Üniversitesi)

Doç. Dr. Bünyamin AYÇİÇEĞİ
(İstanbul Üniversitesi)

Doç. Dr. Cemalettin YAVUZ
(Trakya Üniversitesi)

Doç. Dr. Gülşah Gaye FİDAN
(Gaziantep Üniversitesi)

Doç. Dr. Hasan KAĞLAN
(Hatay Mustafa Kemal Üniversitesi)

Doç. Dr. Mehmet Akif GÖZİTOK
(Erzurum Teknik Üniversitesi)

Sayı Hakemleri
Issue Reviewers

Prof. Dr. Erdem SARIKAYA
(Bozok Üniversitesi)

Doç. Dr. Gülşah Gaye FİDAN
(Gaziantep Üniversitesi)

Prof. Dr. Fatma Sabiha KUTLAR OĞUZ
(Hacettepe Üniversitesi)

Doç. Dr. Hasan KAPLAN
(Hatay Mustafa Kemal Üniversitesi)

Prof. Dr. İbrahim Halil TUĞLUK
(Adıyaman Üniversitesi)

Doç. Dr. İsa IŞIK
(Muş Alparslan Üniversitesi)

Prof. Dr. Mehmet Sait ÇALKA
(Recep Tayyip Erdoğan Üniversitesi)

Doç. Dr. Mustafa DERE
(Ordu Üniversitesi)

Prof. Dr. Mücahit KAÇAR
(İstanbul Üniversitesi)

Doç. Dr. Mustafa Sefa ÇAKIR
(Sivas Cumhuriyet Üniversitesi)

Prof. Dr. Orhan Kemâl TAVUKÇUOĞLU
(Recep Tayyip Erdoğan Üniversitesi)

Dr. Öğr. Üyesi Ayşe KAŞ
(Hatay Mustafa Kemal Üniversitesi)

Prof. Dr. Özer ŞENÖDEYİCİ
(Hitit Üniversitesi)

Dr. Öğr. Üyesi
Damlanur KÜÇÜKYILDIZ GÖZELCE
(Recep Tayyip Erdoğan Üniversitesi)

Prof. Dr. Sadık ARMUTLU
(Ağrı İbrahim Çeçen Üniversitesi)

Dr. Öğr. Üyesi Elif PALIÇKO
(Hacı Bayram Veli Üniversitesi)

Prof. Dr. Sadık YAZAR
(İstanbul Medeniyet Üniversitesi)

Dr. Öğr. Üyesi Gülşah ŞİŞMAN
(Recep Tayyip Erdoğan Üniversitesi)

Prof. Dr. Şerife YALÇINKAYA
(Ege Üniversitesi)

Dr. Öğr. Üyesi Hatice ÖZDİL
(Bitlis Eren Üniversitesi)

Prof. Dr. Ziya AVŞAR
(Niğde Ömer Halisdemir Üniversitesi)

Dr. Öğr. Üyesi Kadriye HOCAOĞLU ALAGÖZ
(Bursa Uludağ Üniversitesi)

Doç. Dr. Ahmet KAVAKLIYAZI
(Konya Selçuk Üniversitesi)

Dr. Öğr. Üyesi Şebnem Şerife ŞAHİNKAYA
(Nevşehir Hacı Bektaş Veli Üniversitesi)

Doç. Dr. Eyüp KUL
(Recep Tayyip Erdoğan Üniversitesi)

Arş. Gör. Dr. Arzu ACAR
(Ege Üniversitesi)

Doç. Dr. Fatih ALTUĞ
(Boğaziçi Üniversitesi)



Klasik Divan
Uluslararası Dil, Kültür ve Edebiyat Araştırmaları Dergisi (KD-UDKEAD)
International Journal of Language, Culture and Literature Researches (KD-IJLCLR)
Cilt 2 / Sayı 2 / Kış 2025



Yazı İşleri Sorumlusu
Managing Editor

Fatma TÜRK

Dil Editörleri
Language Editors

İlknur CEVHER (Türkçe)

Öğr. Gör. Bestami BİLGE (Türkçe)

Öğr. Gör. Hakan SÖNMEZ (İngilizce)

Dr. Öğr. Üyesi Ebubekir Sıddık ŞAHİN (Farsça)

Musahhihler
Proofreaders

Ecemnur TOPCU
Haydar ÖZÇELİK

Kapak Tasarımı ve Mizanpaj
Cover Design and Layout

Beyza ABDUKARIMOV

Yazışma Adresi
Contact

dergiklasikdivan@gmail.com
<https://klasikdivan.com>

Dibace

Kıymetli Okurlarımız,

Bu sayı ile ikinci yılını tamamlayan *Klasik Divan: Uluslararası Dil, Kültür ve Edebiyat Araştırmaları Dergisi Cilt 2-Sayı 2*, 31 Aralık 2025 tarihinde yayımlanmıştır. Dergi, çift kör hakem uygulaması ile incelenip reddedilen yazılar elendikten sonra geriye kalan toplam sekiz yazı ile yayıma hazırlanmıştır. Bunlardan; beşi *Araştırma Makalesi*, biri *Dijital Beşeri Bilimler Yazısı* ve ikisi *Kitap Tanıtım Yazısı*'ndan oluşmaktadır. Öncelikle tamamen gönüllülük esasına bağlı olmasına rağmen derginin yayımlanma aşamasında büyük bir titizlikle çalışan Klasik Divan Çevrim içi Edebiyat Topluluğunun kıymetli üyelerine, dergimize yazılarıyla katkıda bulunan değerli araştırmacılara ve tüm yoğunluklarına rağmen hakemlik teklifimizi geri çevirmeyip yazıları inceleyen saygıdeğer hakem kuruluna teşekkürü borç biliriz. İlk sayıdan itibaren hep daha iyisini yapmanın peşinde olan ve bunun için çabalayan ekibimiz, hakemlerden ret alan makaleler sebebiyle ciddi baskılara maruz kalsa da hakemlerden geçmeyen hiçbir yazıyı dergiye almamayı ilke edinmiştir. Büyük bir emek ve titizlikle yayıma hazırlanan *Klasik Divan: Uluslararası Dil, Kültür ve Edebiyat Araştırmaları Dergisi (KD-UDKEAD)*'nin alanında özgün makalelerle uzun soluklu olmasını temenni ediyoruz. Saygılarımızla...

Baş Editör
Öğr. Gör. Esra AKA

Makale göndermek için başlangıç tarihi: 17 Temmuz 2025
Makale göndermek için son tarih: 17 Ekim 2025
Derginin yayımlanma tarihi: 31 Aralık 2025

İÇİNDEKİLER CONTENT

Dergi Künyesi:I

Dibace:V

İçindekiler:VI

Araştırma Makalesi

Research Article

Mevlânâ'da Şem' Mefhumu yahut Mesnevî'yi Mum Işığında Okumak
The Concept of Sham in Rumi or Reading the Masnavi by Candlelight

Öğr. Gör. Dr. Bestami BİLGE.....72

Fatih Divanı'nda Saç Mefhumu

The Concept of Hair in Fatih Divan

EsraKARAKUŞ.....92

18. Yüzyılda Askerî Islahatların Şiirdeki İzleri: Hamdullah ve Dönem Şairlerinin Tarih Manzumeleri

Traces of Military Reforms in Eighteenth-Century Poetry: Historical Poems by Hamdullah and The Period's Poets

HasanERGÜLEÇ.....111

Karacaoğlan'ın "Elif Elif Diye" Redifli Şiirinin Düşünce Alanı Merkezli Metin Çözümleme Yöntemine (Dam) Göre Şerhi

Commentary of Karacaoğlan's Poem With The Redif "Elif Elif Diye" According to The Thought-Centered Text Analysis Method (Dam)

NagehanKIROĞLAN.....130

Hamdullah Hamdî'nin Yûsuf u Züleyhâ Mesnevisinde Tipler ve Motifler

Characters and Motifs in Hamdullah Hamdî's Mesnevi Yusuf u Züleyha

SenaERDEM.....141

Dijital Beşerî Bilimler Araştırma Makalesi
Digital Humanities Research Article

Dijital Osmanlı Derlem Projesinde Metin Temizleme İşlemi
Text Cleansing Processes in the Digital Ottoman Corpus Project
Doç. Dr. Ayşe TARHAN-Meyase Bahar RAMANLI.....165

Kitap Tanıtım Yazısı
Book Review

Kuzey Kıbrıs Türk Edebiyatının Duayeni Harid Fedai Kıbrıs Türk Kültürü
Bildirileri -III-
Orkun ÖZAŞIK.....181

Divan Edebiyatında Türler
Hafize Tuğba ER.....186

Dijital Osmanlı Derlem Projesinde Metin Temizleme İşlemi

Text Cleansing Processes in the Digital Ottoman Corpus Project

Doç. Dr. Ayşe Tarhan

Uluslararası Kıbrıs Üniversitesi

Türkçe Öğretmenliği Bölümü

aysedarhan@gmail.com

<https://orcid.org/0000-0003-0493-7928>

Meyase Bahar Ramanlı

Dijital Osmanlı Projeleri Araştırmacısı

mbramanli@gmail.com

<https://orcid.org/0009-0003-9584-5347>

Makale Türü/Article Type: Dijital Beşerî Bilimler Araştırma Makalesi

Geliş Tarihi/Received Date: 2 Kasım 2025

Kabul Tarihi/Accepted Date: 7 Kasım 2025

Sayfa/Page: 165-180

Atıf/Citation: Tarhan, Ayşe- Ramanlı, Meyase Bahar. “Dijital Osmanlı Derlem Projesinde Metin Temizleme İşlemi”. *Klasik Divan Uluslararası Dil, Kültür ve Edebiyat Araştırmaları Dergisi* 2/2 (2025), 165-180.

İntihal/Plagiarism: Bu makale turnitin programında taranmıştır. /This article was checked by turnitin.



Dijital Osmanlı Derlem Projesinde Metin Temizleme İşlemi¹

Text Cleansing Processes in the Digital Ottoman Corpus Project

Doç. Dr. Ayşe Tarhan

Meyase Bahar Ramanlı

Özet

Dijital Beşerî Bilimler, edebiyat gibi kültür değerlerinin korunması ve ortaya çıkarılması gibi amaçlarla araştırmacılara bilgisayar teknolojileriyle beşerî bilimleri birleştiren yeni yöntemler önermektedir. Edebî metinler birer kültür hazineleridir ve bunlar üzerinde derinlemesine incelemeler yapılması için benzer yapıların birbiriyle bağlantılarının kurulması gerekmektedir. Bunlar yer adları olabildiği gibi tematik unsurlar da olabilmektedir. Bu gibi işlemler bilgisayarların sunduğu imkânların öncesinde yılları alan fişlemeler aracılığıyla yapılmaktaydı. Şimdi bilgisayar teknolojileriyle bu işlemleri yapmak kolaylaşsa da araştırmacıyı inceleme öncesinde bazı ön işlemler beklemektedir. Bu çalışmada Dijital Beşerî Bilimlerin inceleme yöntemlerinin uygulanabilmesi için Dijital Osmanlı Derlem Projesi'nde araştırmacılar tarafından metinler üzerinde uygulanan metin madenciliğinin metin temizleme önişlemi anlatılacaktır. Dijital Osmanlı Derlem Projesi'nin amacı ve alana yapacağı katkı üzerinde durulacaktır. Çalışmada derlem tanımlanacak ve derlemler üzerinde oluşturulmuş mevcut projeler hakkında kısaca bilgi verilecektir. Ardından derlemlerde uygulanan metin temizleme işlemlerinin aşamaları sırasıyla Stopwords Listesinin Silinmesi ve Normalizasyon vb. başlıklarıyla açıklanacaktır. Her bir işlemin de kendi içinde alt başlıkları oluşturulacak ve örnekler açıklanarak görselleştirilecektir.

Anahtar Kelimeler: Dijital Osmanlı Derlem Projesi, Dijital Beşerî Bilimler, Edebî Metinler, Metin Madenciliği, Metin Temizleme Önişlemi

Abstract

Digital Humanities offers researchers new methods for the preservation and discovery of cultural values such as literature. Literary texts are cultural treasures, and for in-depth analysis, it is necessary to establish connections between similar structures, such as place names or thematic elements. Before the advancements in computers, such processes were carried out through years of indexing. While these processes are now easier with computer technology, certain processes await researchers before analysis. This study will explain the text cleansing pre-processes used by researchers in the Digital Ottoman Corpus Project to apply Digital Humanities analysis methods to texts. The purpose of the Digital Ottoman Corpus Project and its potential contribution to the field will be discussed. This study will define corpus structure and briefly discuss existing projects based on corpora. The steps of text cleansing processes applied to corpora will then be explained, starting with the Stopwords List Deletion and Normalisation. Each process will be further subheaded, and examples will be explained and illustrated.

Keywords: Digital Ottoman Corpus Project, Digital Humanities, Literary Texts, Text Mining, Text Cleansing Preprocesses..

¹ Bu makale; 30-31 Ekim 2025 tarihleri arasında Muş Alparslan Üniversitesinde düzenlenen *Uluslararası Dil, Tarih ve Medeniyet Araştırmaları Sempozyumu*'nda sunulan aynı isimli özet bildirisinin genişletilmiş tam metin hâlidir.

Giriş

Derlem, dil bilimi alanında kullanılmak üzere aracı olarak görülen ve bilgisayarlar vasıtasıyla okunabilen konuşma ve yazımdaki metinlerin oluşturduğu yığın² veya bilgisayar (sayısal) ortamlarında düzenli bir şekilde toplanıp saklanan, çözümlemeye uygun duruma getirilmiş metinler toplamı³ olarak tanımlanmaktadır.

Geçmiş; kutsal kitapların kelime listelerinin ve kelime sıklıklarının belirlenmesine kadar giden derlem çalışmalarının, din dışı metinlere de uygulanmasıyla yaygınlaşmaya başlamıştır.⁴ Derlemler, dil bilim ve edebiyat çalışmaları başta olmak üzere sosyal bilimler çerçevesindeki konularda; çeşitli akademik disiplin ve bilim alanlarında kullanılmaktadır.⁵

Dil bilim çalışmalarında kullanılan derlem, dil bilim metotları dil bilgisi araştırmalarında ele alınan dile genel bir perspektiften bakabilmek için kullanılırken, edebiyat ve/veya sosyal bilimler çalışmalarında kullanılan edebî metinlerin nicelik ve niteliklerinin incelenmesine olanak sağlayan derlem biçem bilimi metotları ise bir yazarın dil kullanımına ait verileri istatistiksel olarak çözümlemek amacıyla işlenir. Derlem dil bilimi, elektronik ortamda dil analizi yaparken biçem bilim de edebî metnin dilsel analizini yapması dolayısıyla dilsel kalıpları ve anlam kodlarını araştırıp bulma gibi ortak hedefleri vardır.⁶

Derlem çalışmalarında kullanılan yöntemler; Sıklık Analizi, Eşdizimlilik Analizi, Sözdizim Ağacı Analizi ve Duygu Analizidir. Sıklık Analizi, bir derlemde herhangi bir kelimenin ne sıklıkla ortaya çıktığını ifade eder. Eşdizimlilik yan yana kullanılan bir veya birden fazla kelimenin birbiri ile olan bağlantısını ifade eder böylece verilerin niceliksel ve niteliksel analizi yapılabilir. Cümlelerin bir ağacın dalları tarzında çözümlenmesi ve söz dizimsel yani dil bilgisel yapıyla açıklama eklenen Sözdizim Ağacı Analizinde matematiksel terimler kullanılır.⁷ Duygu Analizi ise önceleri pazar araştırmaları veya siyasi söylem çözümleme aracı olarak kullanılırken sonraları İngilizce edebî eserlere uygulanan ve çeşitli duygu durumlarının negatif, pozitif ve nötr veya -1, 1 ve 0 ile temsil edildiği bir yöntemdir.⁸

Ulusal ve Uluslararası pek çok derlem tabanlı proje bulunmaktadır. Burada; Türkçe ve Osmanlıca derlemler üzerine yapılan bir kısım çalışma örnek olarak verilecektir. Bunlar kısaca; OTAP

2 Selma Akol Göktaş, *Yöntem Olarak Derlem Araştırmaları: Derlem Dilbiliminin Dil İncelemelerinde Kullanımı* (Çanakkale: Paradigma Akademi, 2021), 198-199.

3 Mustafa Aksan, “Derlem Dilbilim ve Dil Eğitimi”, *Türkçenin Eğitimi-Öğretiminde Kuramsal ve Uygulamalı Araştırmalar*; ed. Leylâ Uzun-Ümit Bozkurt (Almanya: Die Blaue Eule, 2011), 339.

4 Şahru Pilten Ufuk, “Derlem Dilbilim ve Edebiyat Çalışmalarının Kesişim Noktası: Derlem Biçem Bilimi”, *Disiplinlerarası Edebiyat Çalışmaları*, ed. Ömer Solak-Samet Doykun (Çanakkale: Paradigma Akademi, 2021), 145.

5 Akol Göktaş, *Yöntem Olarak Derlem Araştırmaları: Derlem Dilbiliminin Dil İncelemelerinde Kullanımı*, 199.

6 Pilten Ufuk, “Derlem Dil bilim ve Edebiyat Çalışmalarının Kesişim Noktası: Derlem Biçem Bilim”, 145-146.

7 Akol Göktaş, *Yöntem Olarak Derlem Araştırmaları: Derlem Dilbiliminin Dil İncelemelerinde Kullanımı*, 207-210.

8 Ayşe Tarhan, *Dijital ve Beşerî Bilimler Araçlarının Türkçe ve Osmanlıca Metin Analizinde ve Görselleştirmede Kullanılabilirliği* (İstanbul: DBY Yayınları, 2025), 28.

(Osmanlıca Metin Arşivi Projesi)⁹, Metinbankası Projesi, TEBDİZ (Tarih ve Edebiyat Metinlerinin bağlamli dizini ve İşlevsel Sözlüğü)¹⁰-LEHÇEDİZ¹¹ Projeleri, LexiQamus (Lexicon Dijitalleştirme Projesi)¹², Müteferrika (Osmanlı ve Türkiye Çalışmaları Keşif Portalı)¹³, Wikilala¹⁴, BAKİ Projesidir. DOP (Dijital Osmanlı Projeleri)¹⁵ çatısı altında yer alan: Latifi Projesi-Bin Kelime Projesi-Dijital Osmanlı Derlem Projesi, TUD (Türkçe Ulusal Derlem Projesi)¹⁶ ve TS Corpus¹⁷ projeleridir.

Dijital Osmanlı Derlem Projesi

Yukarıda örnekleri verilen projeler etrafında oluşturulan mevcut derlemlerin 700 yılı aşkın zamandır oluşan eserleri kapsayamayışı (böyle bir kapsam bir projeden beklenilmemektedir) nedeniyle hem modern Türkçe hem de Osmanlıca üzerine derlemlerin üretilmesinin elzem olduğunu kanıtlamaktadır. *Dijital Osmanlı Derlem Projesi* Osmanlıca edebî eserler açısından alanında bu boşluğu doldurmak amacıyla atılan bir adımdır. Proje açık erişime sahip olacaktır. Araştırmacılar, Osmanlı dönemi edebî eserlerinin derlemlerine *Dijital Osmanlı Projeleri* web sitesinden ücretsiz ulaşabileceklerdir. Projenin logosu aşağıda yer almaktadır.



Şekil 1: Dijital Osmanlı Derlem Projesi logosu.

Proje kişileri Türkiye'nin farklı üniversitelerinden Yüksek Lisans ve Doktora öğrencileridir.¹⁸ Her bir öğrenci, aylık planlarını kendisinin çalışma gücüne göre ayarlamakta ve divan temizleme işlemini bu plana göre tamamlamaktadır.

Derlemlerin oluşturulmasında veri kaynağı olarak Kültür Bakanlığının e-kitap sayfası merkeze alınmıştır. Osmanlı dönemi edebî metinleri Türkiye Kültür Bakanlığı tarafından dijital ortamda paylaşılmıştır. Dijital ortamda paylaşılan bu metinler proje kapsamında araştırmacılar tarafından metin temizleme gibi işlemlere tabi tutulmaktadır. Dolayısıyla açık erişimle paylaşılacak olan der-

9 <http://otap.bilkent.edu.tr/> Erişim 8 Eylül 2025.

10 <https://www.tebdiz.com/> Erişim 8 Eylül 2025.

11 <https://lehcediz.com/> Erişim 8 Eylül 2025.

12 <https://www.lexiqamus.com/tr/lexicon-digitization-project> Erişim 8 Eylül 2025.

13 <https://www.muteferrika.com/> Erişim 8 Eylül 2025.

14 <https://www.wikilala.com/> Erişim 8 Eylül 2025.

15 <https://digitalottomanprojects.org/language> Erişim 8 Eylül 2025.

16 <https://www.tnc.org.tr/tr/> Erişim 8 Eylül 2025.

17 <https://nlp.tscorpus.com/frequency/> Erişim 8 Eylül 2025.

18 Projeye katılmak isteyen araştırmacıların, Proje yürütücüsü olan Doç. Dr. Ayşe Tarhan'a aysed_dalyan@hotmail.com üzerinden ulaşmaları beklenmektedir.

lemler, Doğal Dil İşlemesine (Natural Language Processing) dayalı araştırmalar açısından bilgisayar tarafından üzerinde çalışmalar yapılmasına olanak sağlayacaktır.

Proje kapsamında ilk eser seçimi Osmanlı edebiyatına adını vermiş olan şairlerin toplu şiirlerinin yer aldığı divanlardır. Osmanlı'nın kuruluşundan itibaren ulaşılan divanların yeniden dijitalleşmiş derlemleri oluşturulacaktır. Bu bağlamda ilk etapta Kültür Bakanlığı tarafından e-kitap olarak paylaşılmış 100 divan yeniden dijitalleştirilecektir.

Bir Bilişim Teknolojisi terimi olan yeniden dijitalleştirme (re-digitalisation) çevrim içi paylaşılan ancak Word ya da txt dosya biçimlerine dönüşmemiş çoğunlukla HTML, XML, PDF veya JPEG gibi dosyaların dönüştürülmesine denilmektedir. Bunlar yazmalar.org'daki el yazmalarını, elektronik kitapları, taranmış gazeteleri vb. materyalleri de içerebilmektedir. Kültür Bakanlığında PDF dosya formatıyla paylaşılan elektronik kitaplar da bu kapsama dâhil edilebilmektedir. Bu kitaplar, araştırmacılar tarafından Osmanlıca transkripsiyon harfleri için oluşturulmuş *Oktay New Transkripsiyon* gibi fontları kullandığından bu fontları tanımayan bilgisayar araçları tarafından okunamamasıyla sonuçlanmaktadır. Bu nedenle bu dosyaların hepsinde bilgisayarlar tarafından tanınan fontlara dönüştürmek ve Doğal Dil İşlemede Gürültü Giderme (noise removal) adı verilen eser dışı unsurların silinmesi gerekmektedir. Bahsi edilen dönüştürme işlemine kısaca Derlemlerde Metin Temizleme ve Normalizasyon adı verilmektedir.

Derlemlerde Metin Temizleme

Yazılı ve/veya sözlü dillerin tanınması, anlaşılması, üretilmesi adına dil bilgisel yapıları ve sözcüklerin özel anlamlarını incelemek, bunlardan çeşitli anlamlar üretip sonuçlar çıkarmak için Doğal Dil İşleme Sistemleri kurulmuştur. Bu hedef doğrultusunda da işlenmemiş doğal dil metinlerinden bilgisayarlı yöntem ve tekniklerle metin madenciliği¹⁹ aracılığıyla araştırma, inceleme ve keşifler yapılmaktadır. Doğal dillerin sahip olduğu metinlerde keşif öncesinde yapılması gereken bazı işlemler bulunmaktadır. Derlem oluşturma bu işin ilk aşamasıdır. Sonrasında yapılması gereken işlem metin temizleme işlemidir.

Metin temizleme ön işlemi, derlemlerde depolanan veya saklanan yapılandırılmamış veri ve/veya metinlerin yapılandırılması/işlenebilmesi için analiz hedeflerine uymayan, asıl metinde olmayan kimi gürültülerin azaltılması gibi birtakım işlemleri kapsayan bir adımdır.²⁰ Öyle ki bu, metin ve/veya veri için en önemli adımdır; çünkü sonrasında yapılacak olan analizin başarısı ve kalitesi, yapılan temizleme ve standartlaştırma ile doğru orantılıdır.

Metin temizlemede yapılan işlemler projeden projeye değişmekle birlikte teknik eksiklikler ve metinlerin Word/txt dosyasına aktarımıyla ortaya çıkan kimi problemler dolayısıyla kelime ve

19 Tarhan, *Dijital ve Beşerî Bilimler Araçlarının Türkçe ve Osmanlıca Metin Analizinde ve Görselleştirmede Kullanılabilirliği*, 26.

20 Neri Van Otten, "Top 20 Essential Text Cleaning Techniques [Practical How To Guide In Python]", *Spot Intelligence*.

harflerde oluşan farklılaşma sebebiyle standart olarak gerçekleştirilen; sayıların temizlenmesi, analiz sonuçlarına katkısı/etkisi olmayan kelimelerin temizlenmesi, ilgisiz sözcüklerin temizlenmesi, noktalama işaretleri ve karakterlerin temizlenmesi²¹ diğer bir ifadeyle de Stopwords Listesi Belirleme ve Silme işlemi yapılmaktadır. Bunun yanında Normalizasyon olarak adlandırılan araştırmacı tarafından yanlış transkribe edilen veya bilgisayarın tanımadığı harf/kelimelerin transliterasyon problemlerinin giderilmesi²² de yer almaktadır.

Dijital Osmanlı Derlem Projesi kapsamında divan metinleri üzerinde metinlerde Gürültü Giderme, Stopwords Listesi Belirleme-Silme ve Kelime veya Harflerde Standartlaştırma/Normalizasyon işlemleri yapılmaktadır. Bu işlemler word dosyasına aktarılan her bir divan için aynı olmamaktadır. Örneğin; *Şeyh Gâlib Divanı*'nda divanın parçası olmayan çok fazla kelime ve sayı silinmiş; bilgisayarın okuyamadığı ve Word dosyasına aktarımla oluşan kelimenin anlamını/bağlamını değiştirecek boyutta harf hataları bulunmuş ve gerekli düzeltmeler yapılmıştır. *Şeref Hanım Divanı*'nda ise harf ve kelime hataları daha az olmakla beraber aruz kalıplarının, dipnotların, başlık ve sayfa numaralarının temizliği yapılmıştır. Proje kapsamında incelenen *Zâtî Divanı*'nda ise diğer divanlarda yapılan standart temizleme işleminin yanı sıra standardizasyon işlemine daha fazla ihtiyaç duyulduğu görülmüştür.

Proje çerçevesinde pek çok divan temizleme işlemine tabi tutulmuştur. Burada sadece üç divan üzerinden örnekler verilecektir. Şeyh Gâlib, Şeref Hanım ve Zâtî Divanlarında Gürültü Giderme ve Stopwords Listesi Belirleme-Silme, Normalizasyon işlemi şu şekilde yapılmıştır:

A-Gürültü Giderme ve Stopwords Listesi Belirleme-Silme İşlemi

Proje kapsamında derlemlerde yapılan bu aşamada sayılar, dipnotlar, analiz sonuçlarına etkisi/katkısı olmayan kelime ve cümleler, noktalama işaretleri, aruz kalıpları ve Arapça ve/veya Farsçadan geçen bazı harflerin yerine kullanılan semboller silinmiştir.

1-Sayıların Silinmesi

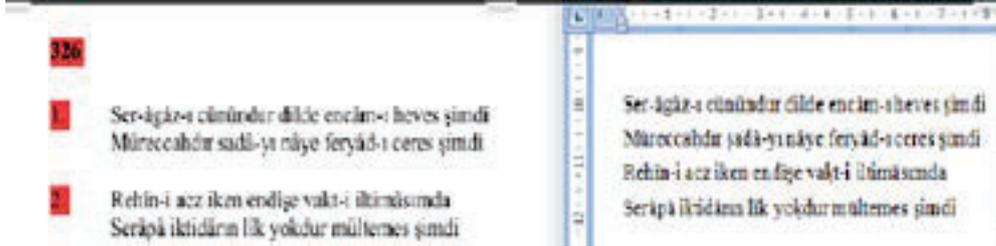
Metinleri transkribe eden araştırmacı tarafından metnin daha iyi anlaşılması için eklediği sayılar divan metinlerinin bir parçası olmadığından derlemlerden silinmektedir. Bu sayılar Romen rakamları, günümüzde kullanılan Desimal sayı sistemine ait sayılar veya herhangi bir dile ait rakam dizgesi olabilmektedir. Bunlar metinde sayfa, şiir, beyit, dipnot numarası veya tarih manzumelerinin tarih beyitlerinin sonuna yerleştirilmiş olan tarihler olabilmektedir. Proje kapsamında PDF formatlı metinden (1 ile işaretli), Word'e aktarılan metinde (2 ile işaretli) sayıların temizlenme işlemi örnekleri aşağıdaki gibidir.²³

21 Suat Atan, "Metin Madenciliği İmkânlar, Yöntemler ve Kısıtlar", *Mehmet Akif Ersoy University Journal of Social Sciences Institute* 31 (2020), 231.

22 Tarhan, *Dijital ve Beşeri Bilimler Araçlarının Türkçe ve Osmanlıca Metin Analizinde ve Görselleştirmede Kullanılabilirliği*, 53.

23 Çalışma boyunca verilen örneklerde bu yol izlenmiştir.

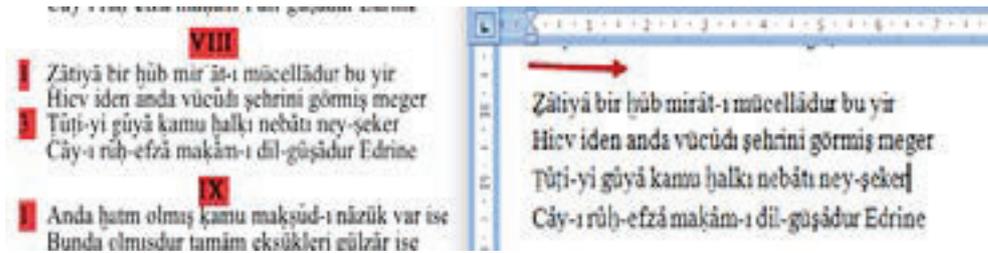
Desimal Sayıların Silinmesi;



1 2

Şekil 2: Şeyh Gâlib Divanı (Okçu, s. 412)

Romen Rakamların silinmesi;



Şekil 3: Zâtî Divanı (Kurtoğlu 2017, s. 282)



Şekil 4: Şeref Hanım Divanı (Arslan, 2018, s. 57)

2-Dipnotların Silinmesi

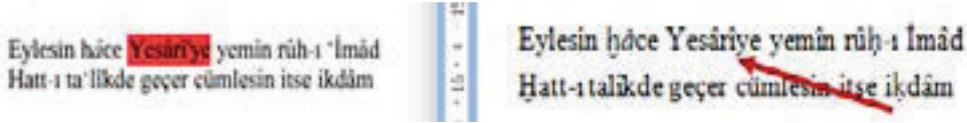
Dipnotlar asıl metinde olmayan bilgilendirici sözcük ve/veya sözcük grupları olup hem karışıklığa/gürültüye sebebiyet verdiği için hem de metnin sayısal verilerini olumsuz etkileyeceğinden silinmektedir.

Tipografi İşaretlerinin Silinmesi;



Şekil 11: Zâtî Divanı (Kurtoğlu, 2017, s. 135)

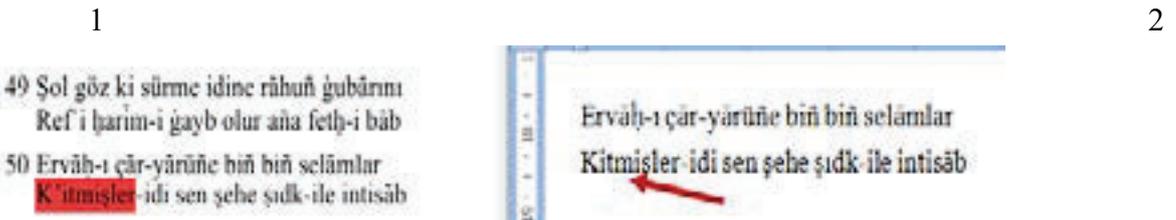
Kesme İşaretinin Silinmesi;



Şekil 12: Şeref Hanım Divanı (Arslan, 2018, s. 99)

Özel isimlere gelen kimi ekleri ayırmada kullanılmakla beraber kesme işaretinin derlemlerde kimi kalıp ifadelerde de kullanıldığı görülmektedir. Aşağıdaki örneklerde “ki”, “ne” ve “ve” ile kurulmuş olan kalıp ifadeler vardır. Derlemlerde “ki i(e)tmışler” yerine “k’itmişler”, “ne ola” (ne olur) yerine “n’ola” ve “ve ey” yerine “ve’y” kullanımı çoğunlukla şiirin veznine uyum için yapılan bir kısaltmada²⁴ kesme işareti olarak kullanılmış olduğundan bilgisayar araçları tarafından kesme işaretinden dolayı bu yapıları ikiye ayırarak bu yapılarda bulunan “n”, “v” ve “k” gibi seslerin stopwords olarak okunmasını önlemek amacıyla bu yapılardaki kesme işaretleri silinmiştir.

Kesme İşaretinin Silinmesi;



Şekil 13: Zâtî Divanı (Kurtoğlu, 2017, s. 58)

Kesme İşaretinin Silinmesi;



Şekil 14: Zati Divanı (Kurtoğlu, 2017, s. 119)

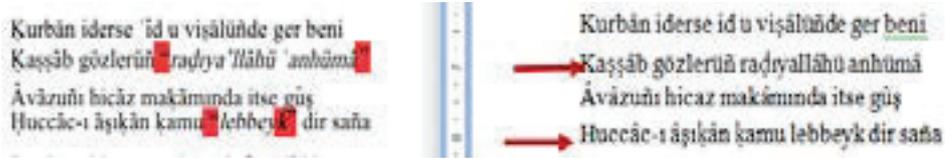
²⁴ Tarhan, *Dijital ve Beşerî Bilimler Araçlarının Türkçe ve Osmanlıca Metin Analizinde ve Görselleştirmede Kullanılabilirliği*, 52.

Kesme İşaretinin Silinmesi;



Şekil 15: Zâtî Dîvânı (Kurtoğlu, 2017, s. 76)

Tırnak İşaretinin Silinmesi;



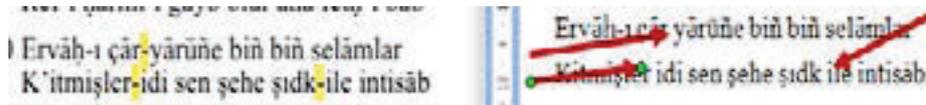
Şekil 16: Zâtî Divânı (Kurtoğlu, 2017, s. 198)

Soru İşaretinin Silinmesi;



Şekil 17: Zâtî Divânı (Kurtoğlu, 2017, s. 159)

Tire İşaretinin Silinmesi;



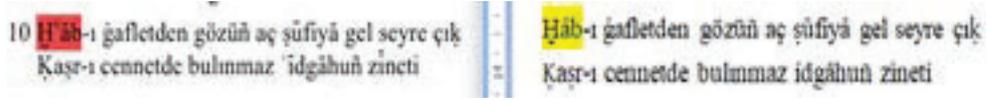
Şekil 18: Zâtî Divânı (Kurtoğlu, 2017, s. 58)

Alt Tire İşaretinin Silinmesi;



Şekil 19: Zâtî Divânı (Kurtoğlu, 2017, s. 243)

Vav “و” Harfini Karşılaman “~” Sembolünün Silinmesi



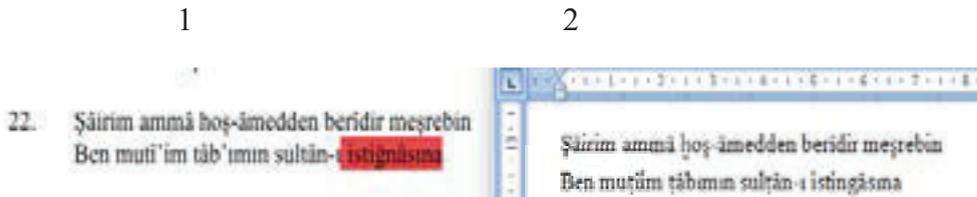
Şekil 22: Zâtî Dîvânı (Kurtoğlu, 2017, s. 66)

B-Normalizasyon (Kelime ve Harflerde Standartlaştırma) İşlemi

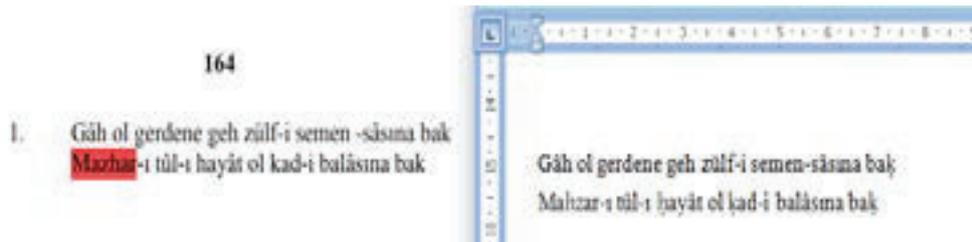
Dijital Osmanlı Derlem Projesi kapsamında incelenen metinlerde temizleme işleminin yanı sıra bir de Normalizasyon dediğimiz standartlaştırma işlemi yapılmaktadır. Bu işlem kimi zaman bir kelime içinde yerleri değişen harflere kimi zaman yanlış veya bilgisayar kaynaklı transkripsiyon-transliterasyon problemleri harflere ve dolayısıyla kelimelere uygulanmaktadır. Bunlar üç başlık altında ele alınabilir. Birincisi yerleri değişen harfler sebebiyle okunuşu bozulan sözcüklerin harflerinin düzeltilmesi, ikincisi yanlış transkripsiyon ve/veya transliterasyon problemi olan harf ve kelimelerin düzeltilmesi, üçüncüsü ise araştırmacı tarafından transkribe edildiği hâlde bilgisayar tarafından okunamayan harflerin standart hâle getirilmesi işlemleridir.

1-Yerleri Değişen Harfler Sebebiyle Okunuşu Bozulan Sözcüklerin Harflerinin Düzeltmesi

Proje kapsamında metin temizleme ön işlemi yapılan derlemlerde bir takım teknik problemler dolayısıyla bazı harfleri yer değiştiren ve anlamı bozulan kelimeler düzeltilmiş ve düzenlenmiştir. Standardizasyon İşleminde karşılaşılan bu örneklerden biri “istiğna” kelimesinin ğ ve n harfleri yer değiştirerek “istinga” şekline dönüşmesidir. Bir diğer örnek de “Mazhar” kelimesinin z ve h harflerinin yer değiştirerek “mahzar” kelimesine dönüşmesidir.



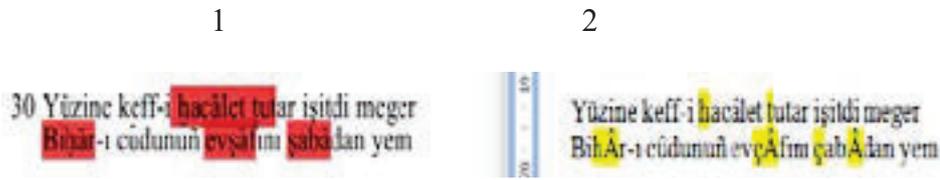
Şekil 23: Şeyh Gâlib Divanı (Okçu, s: 45)



Şekil 24: Şeyh Gâlib Divanı (Okçu, s. 304)

2-YanlıŞ Transkripsiyon ve/veya Transliterasyon Problemi Olan Harf ve Kelimelerin Düzeltilmesi

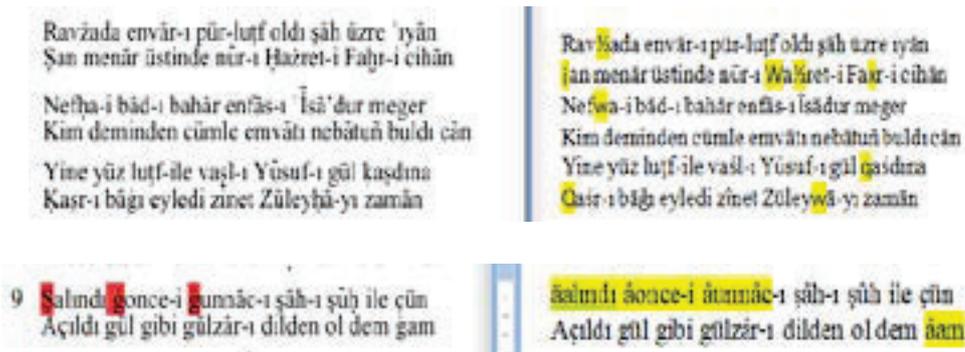
Transkripsiyon işlemine tabi tutulan metindeki harf veya kelimeler kimi zaman bilgisayar tarafından doğru okunamamakta dolayısıyla harfler olduğu gibi aktarılamamaktadır. Harfler bazen büyük (â harfi Â) bazen sayısal bir karakter (z harfi “½”) bazen de farklı bir harf (h harfi x, K harfi Q veya q) şeklinde karşımıza çıkmıştır. Proje kapsamında metin ön işleminde yapılan Standardizasyon İşlemiyle bu harfler olması gereken duruma getirilmiştir. PDF dosyasından (görseldeki 1 sayısının altındaki örnek) Word dosyasına (görseldeki 2 sayısının altındaki örnek) aktarıldıktan sonra metinde oluşan hatalı harfler aşağıda örneklendirilmiştir.



Şekil 25: Zâtî Divanı (Kurtoğlu, 2017, s. 69)



Şekil 26: Zâtî Divanı (Kurtoğlu, 2017, s. 323)



Şekil 27: Zâtî Divanı (Kurtoğlu, 2017, s. 68)

3-Araştırmacı Tarafından Transkribe Edildiği Hâlde Bilgisayar Tarafından Okunamayan Harflerin Standart Hâlde getirilmesi

Proje kapsamında Normalizasyon işlemine tabi tutulan derlemlerde kimi zaman araştırmacı tarafından transkripsiyonu doğru yapılan PDF dosyasındaki metin, Word dosyasına aktarılırken

bazı harfler bilgisayar tarafından doğru okunamamaktadır. Bilgisayar tarafından doğru okunamayan harfler sebebiyle değişen kelimeler, analiz sonuçlarını olumsuz etkilememesi için olması gereken hâle yani standart hâle getirilir.



Şekil 28: Zâtî Divanı (Kurtoğlu, 2017, s. 61)

Sonuç

Dijital Beşerî Bilimler alanının en zor ve zahmetli iş paketi Metin Temizleme, Stopwords listesini belirleme-silme ve Normalizasyon işlemleridir. Dijital Osmanlı Derlem Projesi bu işlemleri büyük bir ekiple gerçekleştirmektedir. Ekipte yer alan her bir araştırmacı kendisine özgü plan düzenlemekte, Metin Temizleme ve Normalizasyon işlemlerini bu doğrultuda tamamlamaktadır.

Araştırmacılar tarafından divanlara uygulanan Metin Temizleme ve Normalizasyon işlemleri bazı önemli deneyimlere ulaştırmıştır. Kültür Bakanlığı tarafından e-kitap olarak paylaşılan PDF dosya formatına sahip her bir divanın araştırmacısı tarafından uyguladığı font tiplerinin farklı olması nedeniyle metinlerin PDF dosyasından Word dosyasına aktarımı sırasında ortaya farklı problemler çıktığı gözlemlenmiştir. Bu çalışmada Şeref Hanım, Şeyh Gâlib ve Zâtî Divanları örneklem alınmış ve örneklemeler bu divanlar üzerinden yapılmıştır.

Hem bu proje hem de bu makale ileride derlem oluşturacak araştırmalar için model olma özelliğine sahip olacaktır. Tarihî dönemlerini de kapsayacak şekilde Türk Dili ve Edebiyatı alanında derlem oluşturulmasını bekleyen pek çok eser bulunmaktadır. Bu proje bu bağlamda ilklerin arasında yer almasıyla öncü projelerden sayılabilmektedir.

Kaynaklar

- Akol Göktaş, Selma. *Yöntem Olarak Derlem Araştırmaları: Derlem Dilbiliminin Dil İncelemelerinde Kullanımı*. Çanakkale: Paradigma Akademi, 2021.
- Aksan, Mustafa. “Derlem Dilbilim ve Dil Eğitimi.” *Türkçenin Eğitimi-Öğretiminde Kuramsal ve Uygulamalı Araştırmalar*. ed. Leylâ Uzun-Ümit Bozkurt. 343-349. Almanya: Die Blaue Eule, 2011.
- Arslan, Mehmet. *Şeref Hanım Divanı*, E-kitap: Kültür ve Turizm Bakanlığı, 2018. <http://ekitap.kulturturizm.gov.tr>

- Atan, Suat. “Metin Madenciliği: İmkânlar, Yöntemler ve Kısıtlar”. *Mehmet Akif Ersoy University Journal of Social Sciences Institute* 21 (2020), 220-239. <https://doi.org/10.20875/makusobed.476524>
- Dijital Osmanlı Derlem Projesi. (2024). Erişim 8 Eylül 2025. <https://digitalottomanprojects.org/language/tr/dijital-osmanli-derlem-projesi/>
- Dijital Osmanlı Derlem Projesi. (2022). Erişim 8 Eylül 2025. <https://digitalottomanprojects.org/>
- Kutoğlu, Orhan. *Zati Divanı*. E-kitap: Kültür ve Turizm Bakanlığı Yayınları, 2017. <http://ekitap.kulturturizm.gov.tr>
- Lehçediz Projesi. Erişim 8 Eylül 2025. <https://lehcediz.com/>
- Lexiqamus. Erişim 8 Eylül 2025. <https://www.lexiqamus.com/tr/lexicon-digitization-project>
- Müteferrika. Erişim 8 Eylül 2025. <https://www.muteferrika.com/>
- Okçu, Naci. *Şeyh Gâlib Dîvânı*. E-kitap: Kültür ve Turizm Bakanlığı Yayınları. <https://i.tmgrup.com.tr/fikriyat/ebooks/seyh-glib-dvni-1569856543070.pdf>.
- OTAP Project. Erişim 8 Eylül 2025. <http://otap.bilkent.edu.tr/>
- Pilten Ufuk, Şahru. “Derlem Dil bilim ve Edebiyat Çalışmalarının Kesişim Noktası: Derlem Biçem Bilim”. *Disiplinlerarası Edebiyat Çalışmaları*. ed. Ömer Solak-Samet Doykun. 145-173. Çanakkale: Paradigma Akademi, 2021.
- Tarhan, Ayşe. “Edebiyat Araştırmacıları İçin Bilgisayar Dilleri ve Metin Madenciliği”. *Littera Turca Journal of Turkish Language and Literature* 10/3 (2024), 444-457. <https://doi.org/10.20322/littera.1477535>
- Tarhan, Ayşe. *Dijital ve Beşerî Bilimler Araçlarının Türkçe ve Osmanlıca Metin Analizinde ve Görselleştirmede Kullanılabilirliği*. İstanbul: DBY Yayınları, 2025.
- Tebdiz Projesi. Erişim 8 Eylül 2025. <https://www.tebdiz.com/>
- Van Otten, Neri. “Top 20 Essential Text Cleaning Techniques [Practical How To Guide In Python]”. Spot Intelligence. https://spotintelligence-com.translate.google/2023/09/18/top-20-essential-text-cleaning-techniques-practical-how-to-guide-in-python/?_x_tr_sl=en&_x_tr_tl=tr&_x_tr_hl=tr&_x_tr_pto=wa
- Wikilala. Erişim 8 Eylül 2025. <https://www.wikilala.com/>

Klasik Divan

Uluslararası Dil, Kültür ve Edebiyat Araştırmaları Dergisi (KD-UDKEAD)
International Journal of Language, Culture and Literature Researches (KD-IJLCLR)

E-ISSN: 3023-820X
Kış 2025 *Winter 2025*
Cilt 2 *Volume 2*
Sayı 2 *Issue 2*