

# METİN MADENCİLİĞİ VE STİLOMETRİ YAKLAŞIMIYLA TANZİMAT'TAN CUMHURİYET'E TÜRK ROMANINDA DİL DEĞİŞİMİ

Hülya BAYRAK AKYILDIZ\*

Murat AKYILDIZ\*\*

**Öz:** Bu çalışma, Tanzimat döneminden 1950'ye kadar Türk romanlarında dil değişimlerini nicel yöntemlerle incelemeyi amaçlamaktadır. Çalışma, döneme özgü dil özelliklerinin analizini yaparak bu değişimlerin kelime zenginliği, kelime çeşitliliği, cümle uzunluğu ve tamlama kullanımına etkisini araştırmıştır. Çalışmada Tanzimat'tan 1950'ye kadar yazılmış 29 roman analiz edilmiştir. Metinler, dijitalleştirilerek metin madenciliği ve stilometrik yöntemlerle değerlendirilmiş, kelime frekansları, türleri ve tamlamalar arasındaki farklılıklar incelenmiştir. Tanzimat dönemine özgü 23237 kelimenin %56'sı sonraki dönemlerde de kullanılmış, özellikle Cumhuriyet dönemine aktarılan kelimeler dikkati çekmiştir. Cumhuriyet dönemi, kelime çeşitliliği açısından en zengin dönem olmuştur (44454 özgün kelime). Buna karşın Servet-i Fünûn döneminde ortalamada eser uzunluğundan dolayı daha düşük çeşitlilik gözlenmiştir. Osmanlı Türkçesine özgü nazarı, hüsn-i, suret-i gibi tamlama yapımında kullanılan kelimelerin kullanımı Cumhuriyet döneminde belirgin şekilde azalmıştır. Ayrıca, Cumhuriyet döneminde Türkçe kökenli duygu, evlenmek, yaşam gibi kelimelerin kullanımı artarken, Arapça ve Farsça kökenli kelimelerin kullanımı azalmıştır. Bulgular, Tanzimat ve Cumhuriyet dönemleri arasında dilde sadeleşme ve modernleşme yönünde bir eğilim olduğunu göstermektedir. Bu hareketlere muhalefetin en önemli sebeplerinden olan dilin yoksullaşması ise söz konusu değildir. Cumhuriyet dönemindeki söz varlığının yüksek çeşitlilik oranı bunun göstergesidir. Bu, Yeni Lisan hareketi ve Dil Devrimi'nin etkilerini yansıtmaktadır. Türkçenin söz varlığında meydana gelen değişimlere rağmen, kelime türlerinin dağılımının dönemler boyunca büyük ölçüde tutarlı olduğu tespit edilmiştir.

**Anahtar Sözcükler:** Stilometri, Metin madenciliği, Türk romanı, Yeni Lisan, Dil Devrimi, Dil değişimi

## Language Change in Turkish Novel From Tanzimat to Republic Era with Text Mining and Stylometric Approach

**Abstract:** This study aims to examine language changes in Turkish novels from the Tanzimat period to 1950 using quantitative methods. By analyzing the linguistic features specific to each period, the study investigates the effects of these changes on lexical richness, word diversity, sentence length, and the use of nominal phrases. A total of 29 novels written between the Tanzimat period and 1950 were analyzed. The texts were digitized and evaluated using text mining and stylometric methods, examining differences in word frequencies, types, and nominal phrases. 56% of the words of Tanzimat period continued

\* Prof. Dr., Anadolu Üniversitesi, Edebiyat Fakültesi, Türk Dili ve Edebiyatı Bölümü, Eskişehir/TÜRKİYE, E-posta: hbayrak@anadolu.edu.tr, ORCID: 0000-0002-1382-4047.

\*\* Prof. Dr., Anadolu Üniversitesi, Açıköğretim Fakültesi, Eskişehir/TÜRKİYE, E-posta: muratakyildiz@anadolu.edu.tr, ORCID: 0000-0001-5069-0132.

*to be used in later periods, with a notable transfer of words into the Republican era. The usage of Ottoman Turkish words that are used to form nominal phrases such as nazar-ı, hüsn-i, and suret-i significantly declined during the Republican era. Additionally, while the use of Turkish-origin words like duygu (emotion), evlenmek (to marry), and yaşam (life) increased in the Republican period, the usage of Arabic and Persian-origin words decreased. The findings indicate a trend toward simplification and modernization of the language between the Tanzimat and Republican periods. However, concerns about the impoverishment of the language, which formed one of the main oppositions to these movements, proved unfounded. The vocabulary richness of the Republican period attests to this. These findings reflect the impact of the New Language movement and the Language Reform. Despite the changes in the Turkish lexicon, the distribution of word types remained largely consistent across periods.*

**Keywords:** Stylometry, Text mining, Turkish novel, Yeni Lisan, Language Revolution, Language evolution

## Giriş

Bu çalışmada Tanzimat döneminden 1950 yılına kadar Türk romanlarında döneme özgü dil özelliklerinin belirlenmesi ve dönemler arasında bir değişimin olup olmadığının nicel yöntemlerle ortaya konması amaçlanmıştır. Bu doğrultuda Tanzimat'tan 1950'ye kadar yazılmış yirmi dokuz roman kelime çeşitliliği, kelime zenginliği, cümle uzunluğu, sık kullanılan kelimeler, kelime türleri, bakımından incelenmiş, dönemler arası karşılaştırmalar yapılmıştır.

Türk edebiyatında 1839 Tanzimat Fermanı ile başlayan modernleşme dönemi edebi meyvelerini 1860lardan itibaren vermeye başlar ve bu dönem 1896'ya kadar sürer. 1896-1901 arası Servet-i Fünûn Devri olarak kabul edilir (Akyüz, 1995, s. 87). II. Meşrutiyet dönemi Fecr-i Âti ve Milli edebiyat hareketlerini kapsayan dönemi ifade etmek üzere kullanılmıştır. 1923'te Cumhuriyet'in kuruluşundan bugüne uzanan döneme ise Cumhuriyet dönemi denir. Bugüne kadar yapılan çalışmalar her bir dönemin özellikle dilin kullanımı bakımından farklı karakteristiklere sahip olduğunu bildirmektedir (Önertoy, 1984; Akyüz, 1995; Kurdakul, 1994; Gökner, 2008; Stone, 2010). Bu bildirim çoğunlukla nicel verilere dayanmayan, öznel gözlemlere ve okumalara dayalı bildirimlerdir. Bu çalışmayla dönemler arasındaki benzerlik ve farklılıklar nicel olarak tespit edilmiştir.

## 1. Tanzimat'tan Cumhuriyet'e Dil Tartışmalarına Genel Bir Bakış

Yaklaşık yüz yıllık bir süreyi kapsayan bu dönemde elbette çok sayıda dil tartışması ve farklı görüşler vardır. Burada bu çalışmanın odağı olan dil değişiminin neden söz konusu dönem ve eserler üzerinden incelendiğinin kısaca ortaya konmasıyla yetinilecektir.

### 1.1. Tanzimat Dönemi ve Dil Arayışları

Tanzimat Fermanı, Türk modernleşmesinin en önemli adımlarından biri ve devletin kurumlarını, hukuki temelini, eğitim sistemini, kültürel yönelimini ve bağlarını yeniden tanımladığı bir dönüm noktasıydı (Akyüz, 1995, s. 5, 13). 1860'tan itibaren ilk örneklerini vermeye başlayan Tanzimat edebiyatıyla birlikte, kültürel değişimin bir parçası ve sonucu olarak dil alanında değişimler başladı.

1839 Tanzimat Fermanı Osmanlı Devleti'nin yüzünü Batı'ya dönmesinin, medeniyet değişiminin dönüm noktasıdır. Bu noktadan itibaren edebiyat da geleneksel olarak üzerinde temellendiği Doğulu köklerini büyük oranda terk ederek Batılı kaynaklara

yönelmiş ve yeniden şekillenmiştir. Halman (2011, s. 2), bunu Osmanlı İmparatorluğu'nun gücünün azalmasına bağlar, ona göre aydın ve yazarların Batı estetiği, türleri ve teknikleri için dinamik bir arayışa girmesinin sebebi budur. Sonuç olarak İmparatorluğa yeni fikir, ilke ve kavramların girdiği ve uygulama alanı bulduğu bu dönem arayış ve değişimlerle karakterize olur.

Türkçe, Türk tarihindeki kültürel yönelim, inanç, yönetim sistemi dönüşümlerine rağmen halk dilinde hemen hemen değişmeden kalmayı başarmıştır. Buna karşılık aydınlar ve edebiyatçılar onuncu yüzyıldan yirminci yüzyıla kadar Farsça ve Arapça kelime dağarcığının yanı sıra bu iki dilin bazı dilbilgisi araçlarını da benimsemişti (Halman, 2011). Dolayısıyla yazı dili konuşma dili olan Türkçeden çok uzaktı (Namık Kemal, 1866). Tanzimat döneminde ilk sadeleşme çabaları başlamıştır (Akyüz, 1995). Tanzimat yıllarının üç büyük edebiyatçısı Şinasi, Namık Kemal ve Ziya Paşa, Türkçenin o günkü durumundan memnun değillerdi. Dönemin ünlü yazarları Ahmet Mithat Efendi, Şemsettin Sami, Ali Suavi, Ahmet Vefik Paşa, Basiretçi Ali, Süleyman Paşa; Osmanlı sahasındaki Türkçe ile Doğu Türkçesinin, yazı dili ile konuşma dilinin, imlanın farklılıkları ve bunların doğurduğu zorluklar üzerinde durmuşlardı (Polat, 2020).

Namık Kemal (1866), “Lisan-ı Osmani'nin Edebiyatı Hakkında Bazı Mülahazatı Şamildir” başlıklı yazısında yazı dili ile konuşma dilinin birleştirilmesi gerektiğini bildirmiş ve ilk sadeleşme hareketinin önünü açmıştır. Osmanlıca kurallarının düzenlenmesini, Türkçenin söz varlığını ortaya koyan bir sözlük hazırlanmasını, var olan eserler arasından doğal söyleyişin lefasetine sahip olanlardan bir seçki hazırlanarak okullarda okutulmasını, Türkçeye özgü bir belagat kitabı yazılarak okullarda okutulmasını önerir. Ona göre sadeleşme kaçınılmazdır, edebiyatın “en iyi söz, halkın sadeliği dolayısıyla anladığı, seçkinlerin de hakikate uyduğu için benimsediğidir” ilkesine göre düzenlenmesi gerekir. Gereksiz süslemeye karşıdır.

Tanzimat öncesi Türk nesrinin insandan kopukluğuna ve süslülüğüne işaret eden Tanpınar, “Türkçede nesrin teşekkülü için insanın ve cemiyet müesseselerinin değişmesi, tahsil sisteminin Türkçeye dönmesi lâzımdı, İşte Tanzimat bunu yaptı” der (Tanpınar, 2013, s. 46).

## 1.2. Servet-i Fünûn Döneminde Dil Tartışmaları

Servet-i Fünûn döneminde dilde sadeleşme konusunda iki ana görüş olduğu söylenebilir. Bunlardan birincisine göre, tasfiye zorunludur. Arapça ve Farsça kelime ve kurallar tasfiye edilmedikçe ilerleme olamaz ve Türkçe gelişemez. İkinci görüşe göre ise Arapça ve Farsça kelimeler dilimizi zenginlik ve ahenk vermektedir. Bunlar artık bize mal olmuştur ve dil bu haliyle yeterli ve mükemmeldir (Rıza Tevfik, 1896). Servet-i Fünûn Döneminde özellikle şiir alanında sadeleşme hareketi ikinci görüşün daha çok benimsenmesi nedeniyle tersine dönmüş, dil daha da ağırlaşmıştır (Akyüz, 1995; Tanpınar, 2013). Bu dönemde gerek Arapça ve Farsçadan gerekse Fransızcadan alınan kelimeler en yüksek seviyeye çıkar ve bu durum dilin yapay ve anlaşılmaz hale getirildiği eleştirilerini beraberinde getirir (Ahmet Mithat, 1897).

Dönemin ünlü yazarlarından ve ikinci görüşe daha yakın olan Halit Ziya; yazarların dile müdahale etmesini doğru bulmaz. Ona göre de yazı dilini mümkün olduğu kadar konuşma diline yaklaştırmak gerekir. Zaten yarım asırdan beri dilde yapılan yenilik bu yöndedir. Ancak konuşma dilinin ne olduğu kararlaştırılmış değildir. İstanbul'daki seçkin irfan sahiplerinin kullandığı dil, konuşma dili sayılırsa mesele yok demektir. Bugünkü

lâtif ve güzel Osmanlıca'yı öğretilirse millet de “İstanbul Osmanlıcası”nı anlar, söyler ve yazar. Ancak lisanın tasfiyesini isteyenler, Osmanlıca'yı öğretmek için okutmayı kâfi görmemekte, onu okutmadan öğrenilecek bir hâle getirmek istemektedir. Karşılığı bulunan Arapça ve Farsça kelimeleri atmak mümkün değildir, “sema” yerine “gök” sözcüğü kullanılırsa “sahâri-semâvât, sümuvv-i cenâb, nazâr-ı semâvî” yerine de “göklere kırılışı, öz ululuğu, gök bakışlı” demek zorunda kalınır. Yapılacak en güzel şey, dili kendi hâline bırakmaktır (Uşaklıgil, 1899, s. 179-182; Aydemir, 2016).

### 1.3. II. Meşrutiyet Dönemi: Yeni Lisan ve Millileşme

1900lerden itibaren yükselen Türkçülük hareketleriyle Türk dil ve kültür tarihine duyulan ilgi artar. Bu dönemin en önemli dil hareketi Yeni Lisan'dır.

II. Meşrutiyet döneminde Selanik'te çıkan Genç Kalemler dergisinde Ali Canip, Ziya Gökalp ve Ömer Seyfettin tarafından başlatılan Yeni Lisan hareketi sadeleşmenin Tanzimat'tan sonraki ikinci ayağını oluşturur (Seyfettin, 1911). Yeni Lisan millileşme arayışını vurgulayan ilkeler içerir ve Milli edebiyatın doğuşunu hazırlar. Bu ilkeler, Genç Kalemler dergisinin 11 Nisan 1911 tarihli sayısında yer alan Yeni Lisan makalesinde ortaya konur. Hareketi etraflıca anlatmak, tenkitlere cevap vermek ve giderek sistemleştirmek amacıyla diğer sayılarda da Yeni Lisan başlıklı makalelerin yayımı sürdürülür. Yazılarda dilde sadeleşme özendirilmeye, sadeleşmenin ilkeleri belirlenmeye ve Türkçenin bilim ve sanat dili olarak yeterliliği ve güzelliği ispatlanmaya, bu dilin etrafında bir edebiyat anlayışı oluşturulmaya çalışılmıştır.

İlk Yeni Lisan makalesinde, eski lisanın hasta olduğu, hastalıklarının da içindeki lüzumsuz ve ecnebi kaideler olduğu belirtilir. Yazı diline esas olarak İstanbul Türkçesi önerilir ve klişe olmuşlar hariç yabancı tamlamaların kullanılmaması gerektiği belirtilir. Yazı diliyle konuşma dili birleştirilmelidir. Yabancı kural ve tamlamalar atılmalı, Arapça ve Farsça edatlar kullanılmamalıdır. Tamlamalar Türkçenin kurallarıyla yapılmalıdır.

Sadeleşme çabaları bir muhalefete de yol açmıştır. Yeni Lisan makalesinde bu çabanın tasfiyecilik olarak görülerek eleştirilerek uğradığından bahsedilir (Ömer Seyfettin, 1911). Yeni Lisan'a yapılan itirazlar şöyle sıralanabilir: Terkiplerin dilden atılması dilde bir yoksullaşmaya sebep olacaktır. Yeni Lisan'ın savunduğu dil, ancak bir bilim dili olabilir, bu dille sanat eserleri üretilemez. Dildeki sadeleşmeler ancak tabii bir şekilde gerçekleşebilir. Dile dışarıdan müdahale edilemez, oysa Yeni Lisancılar dile müdahale etmişlerdir. Yeni Lisancı gençlerin yaptığı sınırlı bir grubun dile müdahalesidir ve siyasal kaynaklıdır. Yeni Lisancılar tasfiyecidirler ve Türkçeyi tarihî dönemlerine döndürmek istemektedirler (Argunşah, 2011, s. 110).

Yeni Lisan hareketinin liderlerinden Ömer Seyfettin (1914a). “Milliyetimiz nasıl Türklük, vatanımız nasıl Türkiye ise, lisanımız da Türkçedir. Türkçe bizim manevî ve mukaddes vatanımızdır. Bu manevî vatanın istiklâli, kuvveti, resmî ve millî vatanımızın istiklâlinden daha mühimdir,” der. Dilin sadeleştirilmesine karşı çıkanların Türk olmayan Osmanlılar olduğunu, bunların maksatlarının ise öncelikle Türk dilini, sonra da Türk milletini inkâr etmek olduğunu söyler. Ne kadar Arapça, Acemce terkipler yapsanız yapın, “tabii, güzel ve terkipsiz Türkçe” galebe çalacak ve Osmanlıca denilen Enderun dili yok olacaktır, der (Ömer Seyfettin, 1914b; Aydemir, 2016).

Yeni Lisan hareketine karşı çıkanlardan biri olan Halit Ziya, Yeni Lisan, Servet-i Lehçe ve Kelimatta Hayal isimli yazılarında, Arapça ve Farsça kelimeleri atmakla Türkçeleşmenin sağlanamayacağını, böyle bir tutumun zengin ve nüanslı bir sanat dilini zayıflatacağını, sanatta ve hayatta değişme ve tekâmülün esas olduğunu, dolayısıyla doğal bir işleyişin dikkate alınması gerektiğini söyler (Balcı, 2012).

#### 1.4. Cumhuriyet Dönemi: Ulus Devletin Doğuşu ve Dil Devrimi

Cumhuriyet dönemindeki Dil Devrimi sadeleşmenin üçüncü ve en önemli ayağıdır. Dili daha çok yerli öğelerin egemen olduğu bir kültür dili durumuna getirmek amacıyla yapılan ve devletin desteğini kazanmış olan ulus çapındaki dili geliştirme eylemine “Dil Devrimi” adı verilmektedir (İmer, 1976, s. 31).

Yeni Lisan hareketinin yol açtığı Milli Edebiyat gerek dil gerekse içerik açısından Cumhuriyet dönemin öncülü olmuştur. Hatta bu iki dönemi birleştiren ve tek bir dönem olarak gören araştırmacılar vardır. Milli Edebiyat hareketinin genellikle 1922 ya da 1923'te tamamlandığı kabul edilir. Bunu ileri sürenler 1922 İstiklal Savaşı'nın tamamlanmasını ya da 1923 Türkiye Cumhuriyeti'nin kuruluşunu esas almaktadırlar (Argunşah, 2011, s. 90). Fakat bazı araştırmacılara göre Milli Edebiyat 1940'a kadar uzanmaktadır (Ercilasun, 1997, s. 453; Çetişli, 2007, s. 129; akt. Argunşah, 2011, s. 90). Bütün bir dönemi Milli Edebiyat hareketine dahil etmesek de aradaki süreklilik açıktır. Türkçülük akımından bağımsız düşünülemez olan Yeni Lisan ve Milli Edebiyat hareketlerinin dayandığı ilkeler ulus devletle birlikte tartışmasız bir şekilde egemen olan edebiyat anlayışına dönüşmüştür. Millileşme, toplumcu sanat anlayışı, sade dil, Türkçeleşme Cumhuriyet döneminde de artarak sürmüştür, örneğin “Memleket Edebiyatı” Milli Edebiyatın Cumhuriyet dönemindeki tezahürlerinden biridir.

Dil, ulus devlet ve ulus kimliği için son derece önemli bir bileşendir. Cumhuriyet de dilini sadeleştirmek, kendi söz varlığını kullanmak, yabancı dillerin etkisini mümkün mertebe azaltmak, Türkçeyi bir bilim ve sanat dili yapmak için çalışmıştır. 1928'de Latin alfabesine geçilmesi, 1932'de Türk dilini geliştirmek amacıyla Türk Dil Kurumu'nun kuruluşu, tarama ve derleme sözlükleri yapılması, sadeleşme çabaları Cumhuriyet dönemindeki dil devriminin aşamalarını oluşturur. Latin alfabesine geçiş Türkçeleşmeyi de hızlandırmıştır. Falih Rıfkı Atay yeni alfabe tartışmalarını aktarırken “Yeni yazı Türkçeleşme hareketine hız vermiştir. Osmanlıcanın devam etmesine imkân yoktu,” der. Velidedeoğlu “Harf Devrimi kesinlikle Dil Devrimini, yani karma bir dil olan Osmanlıcadan ulusal bir dil Türkçeye geçişi getirecekti ve nitekim getirdi de” der (Yücel, 1982, s. 33).

Dil devriminin gerçekleşmesini sağlayan etkenler, aynı zamanda onun amaçlarını ortaya koymaktadır. Uluslaşma etkeni dili yabancı öğelerden temizleme amacını, öteki de kültür dili durumuna getirmeyi amaçlamaktadır (İmer, 1976, s. 32).

Dil devrimine eleştiriler de yöneltmiş, özellikle Atatürk'ün ölümünden sonra bu eleştiriler yoğunlaşmıştır. Dilimizin fakirleştiği endişesi, yeni türetilen kelimelerin uydurma olduğu, bu dil tercihinin siyasi olduğu eleştirileri dile getirilmiş, hatta Dil Devrimi taraftarlarına hakaretler yapılmıştır (Yücel, 1982, s. 45). Dil devriminin en büyük savunucularından Nurullah Ataç bu eleştirilere karşı “Dili değiştirmeye kalkan biz değiliz ki! Bu dil, en aşağı yüzyıldan beri boyuna değişiyor. Niçin değişiyor. Bir kişi öyle dilemiş de buyurmuş, onun için mi değişiyor? Olur mu öyle şey? Yüzyıldan beri boyuna değişiyorsa demek ki bir sıkıntısı var, kendi kendine yetmiyor, kendini beğenmiyor; sınırları dar geliyor” der (Ataç, 1962, s. 113).

Özetle modernleşme hareketleri, İmparatorluk'un çözülme süreci, Cumhuriyet'in kuruluşu ve uluslaşma süreci dil değişimlerini beraberinde getirmiş, bu dönemlerde verilen eserleri dil değişimi açısından ilginç bir inceleme alanı haline getirmiştir.

Bu çalışma dildeki değişimlerin romandaki görüntüsünü nicel olarak tespit etmeyi amaçlamıştır. Bu bağlamda romancıların kullandıkları kelime ve kelime gruplarının çeşitliliği, en çok tercih edilen kelime grupları, dönemler arasında kaybolan ya da yeni ortaya çıkan kelimeler, tamlamalar ve cümle içinde öge kullanım sıklıkları gibi değişkenler açısından yapılan incelemeyle sınıflandırmalar yapılmıştır. Böylece bu eserlere ya da dönemlere ilişkin gözlem ve izlenimlerden oluşan nitel incelemelerin nicel ve nesnel ölçülerle karşılaştırılabilmesine olanak sağlanmak istenmiştir.

## 2. İlgili Araştırmalar

Türk edebiyatı alanındaki araştırmalardan pek azı istatistiksel veya metin madenciliğine dayalıdır. Bu çalışmada da kullanılan metin madenciliği analizi ve stilometrik tespit yöntemlerini Türk edebiyatı metinleri üzerinde deneyen az sayıda çalışmadan biri Can ve Patton'a aittir. Can ve Patton (2010) 20. Yüzyıl Türk romanlarından 40 tanesinde stilometrik analiz gerçekleştirerek romanlarda en sık kullanılan kelimelerden yararlanarak romanları yazıldıkları tarihlerde görece yüksek doğrulukla sınıflandırmışlardır. Alkan vd. (2023), Ömer Seyfettin'in 165 öyküsüne duygu analizi uygulamak için genel ve kendilerinin geliştirdikleri spesifik bir sözlük kullanmışlardır. Çalışmalarında kelime sıklıklarını kelime bulutu ile görselleştirmişler ve duygularda en sık rastlanan kelimeleri tespit etmişlerdir. Akgün vd. (2022), Cumhuriyet'in ilanından hemen önceki şiirler ile Cumhuriyet ilan edildikten sonraki şiirlerin duygu analizini yaparak karşılaştırmışlardır. Çalışkan (2020) çalışmasında Orhan Pamuk'un *Benim Adım Kırmızı* romanının iki farklı dildeki çevirilerini orijinal dildeki metinle metin madenciliği ve stilometrik analizler yaparak karşılaştırmıştır. Yıldırım ve Can (2024), Türk romanının ünlü dört yazarının romanlarında yazım stillerini metin madenciliği, stilometrik ve stilokronometrik analizlerle birbirleriyle karşılaştırmışlardır. Zenkov vd. (2021), Orhan Pamuk ve Yaşar Kemal'in romanlarında sayıların kullanılma sıklıklarını stilometrik ve metin madenciliğine dayanan analizlerle bulmuşlar ve iki yazarın birbirinden ayırt edilebileceğini göstermişlerdir. Patton ve Can (2004) Yaşar Kemal'in dörtlemesinde stilometrik analizler yaparak yazarın sıklıkla kullandığı kelimeleri, cümle uzunluklarını ve dörtlemedeki stilometrik benzerlik ve farklılıkları göstermişlerdir. Türkçe olmayan romanların metin madenciliği ve stilometrik yöntemlerle incelendiği çalışmalar daha yaygın olarak karşımıza çıkmaktadır. Bu çalışmaların bir kısmı romanların yazım biçimine dayalı olarak çeşitli kategorilerde doğru sınıflandırılmasına odaklanmaktadır (Dilai ve Dilai, 2024; Sun, 2024; Ullah ve Mahmoud, 2020). Bir kısmı ise romanlarda yer alan duyguların tespit edilmesini konu edinmektedir (Alkan vd., 2023).

## 3. Araştırmaya konu olan romanlar

### Tanzimat Dönemi

1. Felatun Bey ile Rakım Efendi (1875), Ahmet Mithat Efendi
2. Müşahedat (1891), Ahmet Mithat Efendi
3. İntibah (1876), Namık Kemal
4. Cezmi (1880), Namık Kemal
5. Sergüzeşt (1888), Samipaşazade Sezai
6. Araba Sevdası (1897), Recaizade Mahmut Ekrem

### Servet-i Fünun Dönemi

7. Mai ve Siyah (1899), Halit Ziya Uşaklıgil

8. Aşk-ı Memnu (1900), Halit Ziya Uşaklıgil
9. Eylül (1901), Mehmet Rauf

## II. Meşrutiyet Dönemi

10. Siyah Gözler (1911), Cemil Süleyman Alyanakoğlu
11. Şiřpsevdi (1911), Hüseyin Rahmi Gürpınar
12. Kuyruklu Yıldız Altında Bir İzdivaç (1912), Hüseyin Rahmi Gürpınar
13. Enin (1912), Fatma Aliye
14. Handan (1912), Halide Edip Adıvar
15. Yeni Turan (1913), Halide Edip Adıvar
16. İstanbul'un İç Yüzü (1920), Refik Halit Karay
17. Ateşten Gömlek (1922) Halide Edip Adıvar

## 1950'ye kadar Cumhuriyet Dönemi

18. Sözde Kızlar (1923), Peyami Safa
19. Yeşil Gece (1928), Reşat Nuri Güntekin
20. Yaprak Dökümü (1930), Reşat Nuri Güntekin
21. Dokuzuncu Hariciye Koğuşu (1930), Peyami Safa
22. Yaban (1932), Yakup Kadri Karaosmanoğlu
23. Ankara (1934), Yakup Kadri Karaosmanoğlu
24. Ayaşlı ile Kiracıları (1934), Memduh Şevket Esendal
25. Kuyucaklı Yusuf (1937), Sabahattin Ali
26. Üç İstanbul (1938), Mithat Cemal Kuntay
27. İçimizdeki Şeytan (1940), Sabahattin Ali
28. Mahur Beste (1944), Ahmet Hamdi Tanpınar
29. Huzur (1949), Ahmet Hamdi Tanpınar

## 4. Araştırma Soruları

Artık yeterince gelişmiş olan metin madenciliği araçlarını kullanarak dilde meydana gelen değişimleri test etmek mümkün hale gelmiştir. Örneğin sadeleşme çabaları ve Dil Devrimi sonucunda Türkçenin yoksullaşıp yoksullaşmadığı nicel olarak ortaya konabilir. Bu inceleme kapsamında aşağıdaki alt soruların cevapları aranmıştır.

- Dönemler arasında kelime sayıları/oranları değişiklik göstermiş midir?
- Dönemler arasında ortalama farklı (tekil) kelime kullanma oranları değişiklik göstermiş midir?
- Farklı dönemlerde kullanımı artan ya da azalan kelimeler nelerdir?
- İlk defa kullanılmaya başlanan ya da kullanımdan düşen kelimeler nelerdir?
- Farklı dönemlerdeki kelime türlerinin oranları değişiklik göstermiş midir?
- Her dönemde en çok kullanılan kelimeler nelerdir?

## 5. Yöntem

### 5.1. Verilerin düzenlenmesi ve analizi

Çalışmada Tanzimat döneminden Cumhuriyet Döneminde 1950'ye kadar yazılmış olan kendi dönemlerinde başarıya ulaşmış, okur ve araştırmacıların ilgisine mazhar olmuş, bir başka deyişle kanonik olmuş 29 roman metin madenciliği ve stilometri yöntemleri ile incelenmiştir.

Metin madenciliği, büyük hacimli metin verilerinden anlamlı bilgiler ve desenler çıkarmak için kullanılan disiplinler arası bir süreçtir. Doğal dil işleme, makine öğrenmesi ve istatistiksel yöntemleri bir araya getirerek, metinlerden içgörüler elde etmeyi amaçlar. Metin madenciliği; duygu analizi, konu modelleme, kelime frekans analizi, isim varlık tanıma ve anlamsal ağ oluşturma gibi teknikleri kapsar.

Stilometri, yazılı metinlerin istatistiksel ve dilbilimsel özelliklerine dayanarak yazar kimliği, metin türü veya dönem gibi faktörleri analiz eden bir araştırma alanıdır. Kelime frekansları, cümle uzunlukları, sözdizimi kalıpları ve karakter dizilimleri gibi dilsel göstergeler kullanılarak, bir metnin kime ait olduğu veya hangi üslup özelliklerini taşıdığı belirlenebilir. Özellikle edebiyat, adli dilbilim ve tarih araştırmalarında kullanılan stilometri, makine öğrenmesi ve doğal dil işleme teknikleriyle birleşerek modern veri bilimi araçlarıyla daha güçlü hale gelmiştir.

Çalışmada orijinal olarak Arap harfleriyle basılmış olan metinlerin Latin harfli baskıları kullanılmıştır. Halihazırda Latin harfleriyle basılı olanlar oldukları haliyle kullanılmıştır.

1. Adım: Dijital kopyası olmayan/bulunamayan eserler yüksek çözünürlüklü olarak taranmış ve eserlerin her birinin dijital kopyası oluşturulmuştur. Elde edilen txt uzantılı düz metin dosyalarından analize alınmayacak olan harf dışındaki karakterler R (R Core Team, 2024) ortamında temizlenmiştir. Temizleme esnasında sadece metnin ve nokta işaretinin kalması sağlanmıştır. Bunlar dışında parantezler, soru işareti, eşittir işareti, iki nokta gibi noktalama işaretleri metinden çıkarılmıştır.

2. Adım: Bu dijital kopyalardaki yazarlara ait metin kısımları bir Optical Character Recognition (OCR) yazılımı ile karakterlere çevrilmiştir. Optical Character Recognition (OCR) metin analizinde basılı materyallerin harflerinin/rakamlarının resim halinden tanınarak bilgisayarların anlayabileceği dijital öğelere dönüştürülmesi sürecidir (Mori vd., 1992). Bu süreçte örneğin basılı materyal üzerindeki a harfi bilgisayar tarafından da a harfi olarak anlaşılacaktır. OCR süreci metin analizinin en önemli kısmıdır (Jääskeläinen vd., 2023). Bu aşamada harflerin ve sayıların doğru tanınması yapılacak analizin doğruluğunu etkilemektedir. Kullanılan kitapların baskılarının bazılarının eski olması nedeniyle OCR sürecinde doğruluğu yüksek yazılım olarak literatürde bildirilmiş olan (Tafti vd., 2016; Jain vd., 2021) Fine Reader (ABBY, 2024) yazılımı kullanılmıştır. OCR yapılmadan önce eserlerin her birisi için tarama sırasında oluşmuş olabilecek sayfa eğriliklerinin düzeltilmesi (deskew) işlemi uygulanmıştır. Harflerin bazılarının sayfa içinde eğri durması ihtimaline karşı yamukluk ve bozulmalar tespit edilip düzeltme işlemi uygulanmıştır.

3. Adım: Bazı eserlerde eski Türkçe kelimelerin günümüz Türkçe karşılıkları sayfaların altında özel bir bölümde dipnot olarak verilmiştir. Bu dipnotların her sayfada kapladıkları alan birbirinden farklı olduğundan dipnotların otomatik bir yöntemle temizlenmesi mümkün değildir. Sayfaların içlerinde farklı alanları kaplayan bu metinle ilgisiz sözlük kısımlarının temizlenmesi için PdfMasher 0.7.4 (hardcoded, 2013) yazılımdan yararlanılmıştır. Bu yazılım her bir sayfanın içindeki metinlerin x ve y ekseninde sıraya dizilerek görüntülenmesine imkân tanımakta ve her sayfadan metne dahil olmayan alt ve üst bölümlerin kolayca çıkarılmasını sağlamaktadır. Böylesi bir problemi bulunan sayfalardaki dipnot sözlükler tespit edilerek sayfalardan çıkarılmıştır.

4. Adım: OCR yazılımları işlem bittiğinde düzenlenebilir metin dosyası çıktısı vermektedir. Bu dosyalar veri madenciliği yazılımları tarafından işlenebilmektedir. Oluşan metin dosyaları kendi döneminin klasörüne kaydedilmiştir.

5. Adım: Elde edilen metin dosyaları her bir eser bir metin dosyası olacak şekilde dönem klasöründe saklanmış ve bu dosyalar eserlerin her birindeki kelime sayısı, farklı kelime sayısı, cümle sayısı gibi özellikler bulunurken kullanılmıştır. Ayrıca bu dosyalar başka bir klasöre taşınarak her bir eserin metin dosyası kendi dönemindeki diğer eserlerin metin dosyaları ile birleştirilmiştir. Böylece her bir dönemin tüm eserleri tek bir dosya içinde toplanmıştır. Bu işlem yapılırken eserlerin kronolojik olarak birbirlerinden sonra gelmesi sağlanmıştır.

Araştırma sorularıyla ilgili kelime sayılarının dönemler boyunca değişip değişmediğini bulmak için her döneme ait metin dosyasındaki kelimelerin toplam sayısı bulunmuştur. Kelimeler ekleri ile analize alınmıştır. Böylece yazarların kelimeleri ve ekleri kullanma durumlarının kelime çeşitliliği ile etkileşmesi sağlanmıştır. Kelimelerin çeşitlilik oranlarını belirlemek amacıyla her bir dönem içinde, her kelimedenden kaç adet kullanıldığı sayılmıştır. Bu işlem tokenizasyon (belirteçleme) olarak isimlendirilmektedir. Tokenizasyon sırasında kelimelerin aldıkları eklerle birlikte değerlendirilmesi sağlanmıştır. Tokenize edilen kelimeler her bir dönemde toplam kaç farklı kelimenin kullanıldığının görülmesini sağlamıştır. Kelime çeşitliliği oranı hesaplanırken aşağıda formül kullanılmıştır.

$$\text{Farklı kelimelerin oranı} = \frac{\text{Tokenize edilen kelime sayısı}}{\text{Toplam kelime sayısı}} \times 100$$

Her bir dönemdeki farklı kelime sayısı toplam kelime sayısına bölünerek kelime çeşitliliği oranları hesaplanmıştır. Her bir dönemdeki eserler için bu işlem tek tek gerçekleştirilerek eserlerin kendi içlerindeki çeşitlilikleri bulunmuş, ardından bu değerlerin ortalamaları alınarak dönemlerin ortalama farklı kelime kullanma oranları bulunmuştur.

Dönemler arasında kronolojik olarak ilerledikçe kullanımı azalan, tamamen ortadan kalkan, artan ve ilk defa kullanılmaya başlanılan kelimelerin olup olmadığını belirlemek amacıyla dönemlerde yer alan kelimelerin frekanslarının ve bu frekansların toplam kelime sayısı içindeki oranları belirlenmiştir. Bir dönemden diğerine %0.1 (binde bir) ve daha fazla oransal kullanım değişimi geçirmiş ilk on kelime tespit edilmiştir. Bu aşamada %0.1'in (binde bir) seçilmesinin sebebi yüksek oranda gerçekleşen değişimlerinin fark edilerek düşük frekansta değişim geçiren kelimelerin elenmesinin amaçlanmasıdır. Metinlerdeki tamlamaların bulunması aşamasında Osmanlıca tamlamalara odaklanılmıştır. Medâr-ı iftihâr, cenâb-ı hak gibi iki kelime arasında ı, i, u ve ü barındıran veya – işareti ile yapılmış tamlamalara odaklanılmıştır. Bu doğrultuda metinlerde tek başına bulunan ı, i, u, ü gibi harfler ve – işaretleri tespit edilmiş ve bunların öncesindeki ve sonrasındaki kelimeler ile birlikte ayrı bir metne dönüştürülmüş ve tamlama olup olmadıkları kontrol edilmiştir. Yapılan kontrollerde metinlerdeki ayrı yazılmış bu harflerin hepsinin tamlama olduğu görüldüğünden bu harflerin tekil olarak kullanıldığı durumlarda önceki ve sonraki kelime ile birlikte alınarak bu üçlü ifade bir tamlama kullanımı olarak işaretlenmiştir. Tamlamaların tespitinde kullanılan – işaretlerinin hepsinin de tamlama olduğu tespit edilmiştir. Konuşma belirten çizgi simgelerin veya başka amaçlarla kullanılan ve – işareti ile karıştırılabilecek tüm kullanımların tamlamalardan farklı simgelerle gösterildiği tespit edildiğinden – işareti ile yapılmış olan kelime bağlamaların tamamı tamlama olarak kabul edilmiştir. Tamlamaların sayısının belirlenmesinde de frekans analizi kullanılmıştır.

Metin dosyalarındaki kelimelerin morfolojik analizini yaparak isim, fiil, sıfat gibi hangi tür öge olarak kullanıldığını bulmak amacıyla Java ortamında Zemberek paketi kullanılmıştır (Akın ve Akın, 2007). Zemberek her bir kelimenin içinde bulunduğu cümlenin bağlamı içinde kelimenin hangi tür öge olarak (sıfat, zarf, isim, fiil, vb.) kullanıldığını morfolojik analizle bulabilen Türkçe dili için en başarılı (Tunalı ve Bilgin, 2012; Doğuç vd., 2020) pakettir. Zemberek paketi kelimelerin morfolojik analizini cümle bazında değerlendirebilmektedir. Buna bağlam analizi (context analysis) denilmektedir. Zemberek morfolojik analiz yaparken ilk aşamada cümleleri kelimelerine ayırır. Kelimeler ekleriyle birlikte metinde geçtikleri haliyle ele alınır. İlk aşamada analiz edilecek kelime, sistem tarafından giriş olarak alınır. İkinci aşamada kelimenin olası kökleri, dahili sözlük ve kök ağacı kullanılarak tespit edilir. Bu aşamada, kelimenin başlangıç kısmı potansiyel köklerle eşleştirilir. Üçüncü aşamada belirlenen kök adaylarına, Türkçe dilbilgisi kurallarına uygun ekler sırasıyla eklenir. Bu süreçte, eklerin geçerliliği ve kökle uyumu kontrol edilir. Eğer kelime sözlükte bulunmuyorsa, sistem büyük harfle başlayan özel isimler veya sayılar için geçici sözlük öğeleri oluşturur ve analiz etmeyi dener. Dördüncü aşamada her bir geçerli kök ve ek kombinasyonu için olası morfolojik analiz sonuçları oluşturulur. Bu sonuçlar, kelimenin kökü (kelimenin temel anlamını veren bölüm), ekleri (kelimeye eklenen ve anlam veya gramer işlevini değiştiren parçalar), dilbilgisel özellikleri (kelimenin türü (isim, fiil, sıfat vb.), zaman, şahıs, çoğulluk gibi) bilgileri içerir. Eğer birden fazla geçerli çözümleme varsa, bağlam analizi yapılarak en uygun çözümleme seçilir. Bu adım, özellikle cümle içindeki kelimelerin doğru anlam ve işlevlerini belirlemek için kritiktir. Zemberek bu aşamada tercihe bağlı olarak köklendirme (stemming) ve lemmatizasyon olarak iki seçenek sunmaktadır. Lemmatizasyon köklendirmeye kıyaslandığında bağlam bilgisini de içermesi ve kelimelerin sözlük hallerini dikkate alması (Khyani vd., 2021; Balakrishan ve Lloyd-Yemoh, 2014) dolayısıyla bu çalışmada tercih edilmiştir. Aşağıda bir cümle için yapılan analize ait örnek Zemberek'in ne kadar başarılı olduğu hakkında bilgi verecektir.

“Her tesadüfte onu bir parça daha sürükleyip götüren bu nazarlar nihayet vusulünden ihtiraz olunan bu netice karşısında bütün azimlerini kırmıştı”, [her:Det] her:Det | [tesadüf:Noun] tesadüf:Noun+A3sg+te:Loc | [o:Pron,Demons] o:Pron+A3sg+nu:Acc | [bir:Det] bir:Det | [parça:Noun] parça:Noun+A3sg | [daha:Adv] daha:Adv | [sürüklemek:Verb] sürükle:Verb|yip:AfterDoingSo→Adv | [götürmek:Verb] götür:Verb|en:PresPart→Adj | [bu:Det] bu:Det | [nazar:Noun] nazar:Noun+lar:A3pl | [nihayet:Noun] nihayet:Noun+A3sg | [vusul:Noun] vusul:Noun+A3sg+ü:P3sg+nden:Abl | [ihtiraz:Noun] ihtiraz:Noun+A3sg | [olmak:Verb] ol:Verb|un:Pass→Verb|an:PresPart→Adj | [bu:Det] bu:Det | [netice:Noun] netice:Noun+A3sg | [karşı:Adj] karşı:Adj|Zero→Noun+A3sg+sı:P3sg+nda:Loc | [bütün:Adj] bütün:Adj | [azim:Noun] azim:Noun+ler:A3pl+i:P3sg+ni:Acc | [kırmak:Verb] kır:Verb+mış:Narr+tı:Past+A3sg |

Yukarıdaki örnek analizde sırasıyla Det (zamir), Noun (isim), Adv (zarf fiil), Verb (fiil), AfterDoingSo (sonraki zamanı bildirir ek), PresPart (sıfat fiil eki), A3pl (üçüncü çoğul şahıs eki), A3sg (üçüncü tekil şahıs eki), P3sg (üçüncü tekil şahıs iyelik eki), Abl (çıkma/ayrılma (den) eki), Pass (edilgenlik eki), Adj(sıfat), Loc (bulunma (de) eki), Acc (belirtme (i) eki), Narr (rivayet geçmiş zaman eki) ifadeleri her bir kelimenin kökünün, ekinin ve kök+ek bütünüün metin anlamı içinde cümle öğelerinden hangisinin görevini yerine getirdiğini göstermektedir. Görüleceği gibi Zemberek cümleyi anlamını da

dikkate alarak bütün öğeleriyle analiz edebilen başarılı bir pakettir. Yukarıdaki cümle meşrutiyet dönemindeki eserlerden birisinden alınmış eski Türkçe ifadelerin yoğun bulunduğu cümlelerden birisidir. Zemberek cümlenin bağlamını dikkate alarak her bir kelimeyi doğru bir şekilde morfolojik olarak analiz etmiştir. Örnek olarak “vusûlünden” kelimesini alırsak zemberek [vusûl:Noun]vusûl:Noun+A3sg+ü:P3sg+nden:Abl şeklinde analiz etmiştir. Buna göre vusul isim olarak sınıflandırılmıştır. +ü eki ile 3. tekil şahıs kipindedir (A3sg). +n kaynaştırma harfidir. Abl (ablative) kararı ise çıkma durumunu (-den) göstermektedir. Bu sınıflandırma doğrudur ve oldukça ayrıntılıdır. Bu çalışmada zemberek analizi sonucunda elde edilen kelimelerin öge olarak hangi görevi yerine getirdiğine bakılmış diğer sınıflandırmalar dikkate alınmamıştır. Örneğin vusûl kelimesinin isim olduğuna karar verildiyse sadece bu bilgi alınmış, diğer bilgiler analize dahil edilmemiştir.

Verilerin analizi için R (R Core Team, 2024) ve RStudio (RStudio Team, 2024) ve Intellij Idea (Intellij IDEA, 2024) kullanılmıştır. Metinlerdeki kelimelerin tespit edilmesi, için R paketlerinden tidytext (Silge ve Robinson, 2016), veri dosyalarının birleştirilmesi, veri manipülasyonu ve kelimelerin frekanslarının belirlenmesi için tidyverse (Wickham vd., 2019) ve dplyr (Wickham vd, 2023) kullanılmıştır. Grafiklerin çizilmesi için ggplot2 (Wickham, 2016) ve VennDiagram (Chen ve Boutros, 2011) paketleri kullanılmıştır. Kelimelerin Türkçe morfolojik analizi için Intellij Idea ortamında Zemberek (Akın ve Akın, 2007) paketi kullanılmıştır.

## 6. Bulgular

Araştırmanın “dönemlerin farklı kelime sayıları/oranları artma ya da azalma yönünde değişiklik göstermiş midir” sorusuna yanıt verebilmek amacıyla her bir dönemdeki eserlerin tamamında kelime sayıları, farklı kelime sayıları ve farklı kelime kullanma oranları bulunmuş ve aşağıda Tablo 1’de gösterilmiştir.

Özellik	Tanzimat	Servet-i Fünûn	II. Meşrutiyet	Cumhuriyet
Döneme Özgü Kelime Frekans	23237	9445	27616	44454
Toplam Farklı Kelime Frekans	52655	33594	65227	83309
Döneme Özgü Kelime Kullanım Oranı (%)	44.13	28.11	42.34	53.36
Toplam Kelime Hazinesi İçindeki Özgünlük Oranı (%)	14.64	5.95	17.40	28.02
Toplam Kelime Hazinesinden Faydalanma Oranı (%)	33.18	21.17	41.11	52.50
Tüm Dönemlerde Farklı Kelime Frekans		158669		

**Tablo 1.** *Dönemlerin kendilerine özgü kelime frekansları ve oranları*

Yukarıdaki tablodan Tanzimat, Servet-i Fünûn, II. Meşrutiyet ve Cumhuriyet dönemlerinin eserlerinin hepsini bir araya getirdiğimizde ortaya çıkan farklı kelime sayısı 158669 olarak görülmektedir. Bir başka ifadeyle tüm dönemler bir aradayken toplamda 158669 farklı kelimeye rastlanmaktadır. Bir bakıma bu sayı Türk edebiyatının Tanzimat’tan 1950’ye dek kullandığı toplam kelime hazinesidir. Bu Türkçe kelime hazinesinin 44454 tanesi Cumhuriyet dönemine özgüdür. Yani bu dönemde ilk kez ortaya çıkmıştır. 23237 tanesi ise sadece Tanzimat döneminde kullanılmış ve sonraki dönemlerde tamamen kullanımdan kalkmıştır. 9445 tanesi sadece Servet-i Fünûn

döneminde kullanılmış diğer dönemlerde hiç kullanılmamıştır. 27616 tanesi sadece II. Meşrutiyet döneminde kullanılmış diğer dönemlerde kullanılmamıştır. Buna göre Cumhuriyet döneminin toplam Türk romanının kelime hazinesine yaptığı katkı en yüksektir (44454 kelime ile toplam kelime hazinesinin %28.02'sini üretmektedir).

Toplam farklı kelime frekansları incelendiğinde Cumhuriyet döneminin 83309 farklı kelime kullanımıyla en yüksek farklı kelime kullanım sayısına sahip olduğu görülmektedir. Servet-i Fünûn dönemi ise farklı kelime kullanma sayısının en düşük olduğu dönemdir. Bu dönemde toplamda 33594 birbirinden farklı kelime kullanılmıştır. 2. Meşrutiyet döneminde kullanılan birbirinden farklı kelime sayısı 65227 iken Tanzimat döneminde 52655'tir.

Cumhuriyet döneminin diğer dönemlerle ortak olmayan kendine özgü kelime barındırma bakımından en yüksek orana (%53.36) sahip olduğu görülmektedir. Servet-i Fünûn dönemi ise kendine özgü kelime barındırma oranının en düşük olduğu dönemdir (%28.11). Tanzimat dönemindeki kelimelerin %44.13'ü daha sonraki dönemlere aktarılmayıp sadece Tanzimat dönemi içinde kullanılmıştır. Cumhuriyet dönemindeki kelimelerin yarısından fazlası (%53.36'sı) Cumhuriyet döneminde türetilmiş ya da kullanılmaya başlanmış kelimelerdir. Servet-i Fünûn döneminin daha düşük sayı ve orana sahip olmasının sebebi bu dönemdeki eser sayısının düşük olmasıyla ilgili olabilir.

Her bir dönemin önceki ve sonraki dönemle veya dönemlerle ortak kelimelerin ve kendisinin getirdiği orijinal kelimelerin sayısını tespit edebilmek amacıyla hesaplamalar yapılmış ve aşağıda Tablo 2 ve Tablo 3'te verilmiştir. Tablo 2'de dönemlerin kendilerine özgü ve diğer bir dönemle olan ortak kelime frekanslarını gösterirken Tablo 3 üç dönemin ortak kelime frekanslarını ve sonunda da tüm dönemlerin ortak kelime frekanslarını göstermektedir.

Dönem	Tanzimat	Servet-i Fünûn	Meşrutiyet	Cumhuriyet
Tanzimat	23237	1429	4543	4406
Servet-i Fünûn	1429	9445	1481	3427
II. Meşrutiyet	4543	1481	27616	9501
Cumhuriyet	4406	3427	9501	44454

**Tablo 2.** *Dönemlerin ortak ve özgün kelime frekansları*

Dönemler	Ortak frekans
Tanzimat, Servet-i Fünûn, Meşrutiyet	1590
Tanzimat, Servet-i Fünûn, Cumhuriyet	1937
Tanzimat, Meşrutiyet, Cumhuriyet	5755
Servet-i Fünûn, Meşrutiyet, Cumhuriyet	4071
Tüm Dönemlerde Ortak	10214

**Tablo 3.** *Üç Dönemin ortak kelime frekansları ve toplam kelime frekansları*

Tablo 2 ve Tablo 3'ten görüleceği gibi dönemlerde ortak kelime frekansı 10214'tür. En çok ortaklık gösteren dönemler 9445 ortak kelime ile ikili eşleşmelerde Cumhuriyet ile Meşrutiyet dönemleridir. Üçlü eşleşmelerde ise en çok ortak kelime Tanzimat, Meşrutiyet ve Cumhuriyet dönemlerindedir. En az ortak kelime ise Tanzimat, Servet-i Fünûn ve Cumhuriyet eşleşmesindedir.

Cumhuriyet dönemine en fazla kelime mirası bırakan dönem Tanzimat'tır. Tanzimat döneminde kullanılan kelimelerin %41,51'i Cumhuriyet dönemine ulaşmıştır ve bu kelimelerin bir kısmı doğrudan, bir kısmı ise ara dönemlerden geçerek aktarılmıştır.

Tanzimat dönemine özgü kelimelerden 23237'si yalnızca bu dönemde kalmış, 4543'ü Meşrutiyet dönemine, 4406'sı ise Cumhuriyet dönemine taşınmıştır. Meşrutiyet dönemi, Tanzimat'ın kelimelerini alarak Cumhuriyet dönemine aktarımda bir köprü görevi üstlenmiştir. Tanzimat dönemine ait kelimelerin %44'ü zamanla kaybolmuş olsa da %56'sı sonraki dönemlerde varlığını sürdürmüştür. Bu durum, Tanzimat döneminin dil yapısının, sonraki dönemlerin kelime dağarcığını şekillendirmede önemli bir rol oynadığını ortaya koymaktadır.

Servet-i Fünûn ve Meşrutiyet dönemleri arasında kelime aktarımı sınırlıdır. Servet-i Fünûn dönemine özgü kelimelerin büyük bir kısmı bu dönemde kalmış ve diğer dönemlere daha az aktarılmıştır. Bu dönemden Meşrutiyet'e geçen 1481 kelime, Servet-i Fünûn dilinin daha sınırlı bir etkisinin olduğunu göstermektedir. Buna karşılık, Servet-i Fünûn'dan Cumhuriyet dönemine geçen kelimelerin sayısı (3427) azımsanmayacak düzeydedir ve bu durum, Servet-i Fünûn dilinin bazı unsurlarının doğrudan Cumhuriyet dönemine etki ettiğini göstermektedir. Servet-i Fünûn, diğer dönemlere kıyasla daha özgün bir kelime hazinesine sahiptir ve özellikle Cumhuriyet dönemine kadar ulaşan yenilikçi bir katkı sağlamıştır.

II. Meşrutiyet dönemi hem Tanzimat'tan hem de Servet-i Fünûn'dan daha fazla kelime üretmiş ve bu kelimelerin bir kısmını Cumhuriyet dönemine aktarmıştır. Meşrutiyet dönemine özgü 27616 kelime yalnızca bu dönemde kullanılmış, 9501 kelime ise Cumhuriyet dönemine taşınmıştır. Bu, Cumhuriyet döneminin dil yapısının Meşrutiyet döneminden yoğun bir şekilde beslendiğini göstermektedir.

Tüm dönemler arasında ortak kullanılan kelime sayısı incelendiğinde, Tanzimat, II. Meşrutiyet ve Cumhuriyet dönemleri arasında daha güçlü bir bağ görülmektedir. Bu üç dönemin 5755 kelimedeki kesişmesi, Tanzimat'tan itibaren gelişen bir sürekliliği işaret eder. Servet-i Fünûn ve II. Meşrutiyet dönemlerinden Cumhuriyet dönemine taşınan kelime sayıları ise görece daha azdır. Tüm dönemlerde ortak kullanılan 10214 kelime, dilin temel unsurlarını oluşturmuş ve dönemler arası geçişlerde sürekliliği sağlamıştır. Tanzimat'tan II. Meşrutiyet'e, ardından Cumhuriyet'e geçen kelimeler, dilin modernleşme sürecinde önemli bir yapı taşı olarak karşımıza çıkmaktadır. Bu durum, her dönemin özgün katkıları kadar önceki dönemlerle olan bağlarının da dilin gelişiminde belirleyici olduğunu ortaya koymaktadır.

Her dönemde en çok kullanılmış olan kelimeler tespit edilmiş ve bunların kelime bulutları oluşturulmuştur. Kelime bulutları oluşturulurken dönem içinde gözlenen kelime frekansları dönemin toplam kelime frekansına bölünerek frekanslar normalize edilmiştir. Böylece kelime bulutlarındaki boyutlara bakılarak dönemlerin birbirleriyle karşılaştırılabilmesi mümkün hale getirilmiştir. Örneğin "bu" kelimesi Servet-i Fünûn, II. Meşrutiyet ve Cumhuriyet dönemlerinde aynı büyüklüktedir. Buna göre "bu" kelimesi bu üç dönemde aynı oranda rastlanan kelimedir. Kelime bulutları aşağıda Grafik1-Grafik 4'te gösterilmiştir.



aldığına işaret ediyor olabilir. *Ben* ve *sana* kelimelerinin sıklığı, kişisel anlatıların güçlendiğini ve bireysel duyguların önem kazandığını düşündürür.

Cumhuriyet kelime bulutunda, *bir*, *bu*, *ve*, *kadar*, *fakat*, *gibi* gibi temel bağlaç ve edatlar baskın olsa da *adam*, *şey*, *onun*, *birdenbire* gibi kelimeler dikkat çekmektedir. Bu kelimeler, anlatılarda bireyden çok genel bir bakış açısının ve toplumsal meselelerin vurgulandığını düşündürür. Bu dönemin dilinin, olay odaklı ve genel bir perspektifi benimsemiş bir yapıyı yansıttığı söylenebilir.

Bu kelime bulutları birlikte değerlendirildiğinde, her dönemin dil yapısının kendine özgü temaları ve odak noktaları olduğu görülmektedir. Bütün dönemlerin metinleri tek bir metin gibi değerlendirildiğinde oluşan kelime sıklıklarını gösteren kelime bulutu ise aşağıda Grafik 5'te verilmiştir.



Grafik 5. Bütün dönemlerin kelime bulutu

Grafik 5'ten görüleceği gibi *bir*, *bu*, *ve*, *kadar*, *sonra*, *için*, *gibi*, *o*, *de*, *da*, *ne* kelimeleri bütün dönemlerde ve eserlerde en çok kullanılan kelimeler olmuştur. *Sonra*, *şimdi*, *artık*, *olduğu*, *etti*, *dedi* gibi kelimeler, olay anlatımına odaklanıldığını ve metinlerin genelde hikâyeler etrafında şekillendiğini düşündürür.

Yukarıda verilen bulgular her bir dönemde yer alan eserlerin tamamı dikkate alınarak elde edilmiştir. Dönemlerdeki eserlerin ortalama değerleri üzerinden başka bir deyişle her bir dönemdeki ortalama bir eser açısından çeşitlilikler karşılaştırılmak istenirse neyle karşılaşılacağını belirlemek amacıyla her bir dönemdeki tüm eserlerin ortalaması üzerinden bulgular da elde edilmiş ve aşağıda sunulmuştur.

Dönem	Ortalama kelime sayısı	Ortalama Farklı kelime sayısı	Farklı kelime sayısının oranı (%)	Ortalama Cümle sayısı	Cümle sayısının oranı (%)	Ortalama cümle uzunluğu (kelime sayısı/cümle sayısı)	Osmanlıca Tamlama sayısı ortalaması	Tamlama sayısının kullanılan kelimelere oranı (%)
<b>Tanzimat</b>	49113.83	14002.67	28.51	4089.33	8.3	12.01	948.33	6.8
<b>Servet-i Fünûn</b>	73547.33	15744	21.64	5739	7.75	12.93	219	1.4
<b>II. Meşrutiyet</b>	44814.63	13632.88	30.89	4879.13	10.63	9.54	505.63	2.99
<b>Cumhuriyet</b>	55200.5	14528.75	28.27	6325.17	11.61	8.91	40.17	0.18

**Tablo 4.** *Eserlerin kelime sayısı, farklı kelime sayısı ve farklı kelime oran ortalamaları*

Tablo 4 incelendiğinde ortalama bir eserde kullanılan kelimelerin birbirinden farklı olması bakımından en çok farklı kelime kullanılan dönemin Servet-i Fünûn dönemi olduğu görülmektedir (15744 farklı kelime). Ortalamada en az farklı kelime kullanılan eserlere sahip dönem ise II. Meşrutiyet dönemidir (13632.88 kelime). Farklı kelime kullanma sayıları dönemdeki eserlerde yer alan ortalama kelime sayıları ile oranlandığında en yüksek oranlı dönemin II. Meşrutiyet dönemi olduğu, en düşük oranlı ortalama esere sahip dönemin ise Servet-i Fünûn dönemi olduğu görülmektedir. Bu bilgiler ışığında Servet-i Fünûn döneminin diğer dönemlere oranla eser başına kelime hazinesinin daha yüksek olduğu fakat ortalamada uzun eserler yazılan bir dönem olduğu için dönem içindeki ortalama çeşitlilik oranının düşük olduğu söylenebilir.

Servet-i Fünûn dönemi, eserlerin ortalama uzunluğu bakımından diğer dönemlerin önüne geçmektedir; bu dönemde eser başına düşen ortalama kelime sayısı 73547'dir. Buna karşılık, II. Meşrutiyet dönemi, ortalama 44814,63 kelime ile daha kısa eserlerin yazıldığı bir dönemdir. Servet-i Fünûn döneminde ortalama bir eserde yer alan 73547 kelimenin 15744'ü, II. Meşrutiyet döneminde ise 44814,63 kelimenin 14528,75'i birbirinden farklıdır. Bu farklı kelime sayılarının toplam kelime sayısına oranlanması, dönemlerin kendi içinde ne kadar çeşitli kelime kullanıldığını gösterir. Aynı zamanda bu oran, bir dönemden seçilecek ortalama uzunlukta bir eserde rastgele alınan 100 kelimenin kaç tanesinin farklı kelimeler olacağını ifade eder.

Ortalama farklı kelime sayısının ortalama toplam kelime sayısına olan oranı II. Meşrutiyet döneminde %30,89 ile en yüksek değere sahiptir. Servet-i Fünûn döneminde ise bu oran %21,64 olarak hesaplanmıştır. Başka bir deyişle, II. Meşrutiyet döneminden alınan ortalama bir eserden rastgele seçilen 100 kelimenin yaklaşık 31'i farklı iken, Servet-i Fünûn döneminde bu sayı 22'ye düşmektedir. Tanzimat döneminde bu oran %28,51, Cumhuriyet döneminde ise %28,27 olarak kaydedilmiştir. Servet-i Fünûn döneminde düşük bir oranın görülmesinin temel nedeni, bu dönemdeki eserlerin oldukça uzun olmasıdır; bu da kelime tekrarlarının artmasına yol açmaktadır.

Tanzimat ve Cumhuriyet dönemleri ise ortalama farklı kelime sayısı bakımından birbirine oldukça yakın değerler göstermektedir. Tanzimat döneminde bir eserde ortalama 14002,67 farklı kelime bulunurken, Cumhuriyet döneminde bu sayı 14528,75'tir. Cumhuriyet döneminde bu fark yalnızca 526 kelimedir ve bu küçük fark, belirgin bir üstünlük ya da fark olarak değerlendirilemez. Bu açıdan bakıldığında, Tanzimat ve Cumhuriyet dönemlerinin ortalama farklı kelime kullanımı ve bu kelimelerin eserlere dağılımı oldukça benzer özellikler göstermektedir.

Tanzimat ve Cumhuriyet dönemlerinin ortalama farklı kelime kullanma sayılarının değişmemesinin aynı kelimelerle mi yoksa farklı kelimelerle mi gerçekleştiğini görebilmek amacıyla dönemler arasında kaybolan veya ilk defa ortaya çıkan kelimelerin frekansları tespit edilmiştir. Elde edilen bulgular aşağıda Tablo 7'de gösterilmiştir.

Bu aşamada kullanımı biten ya da önemli derecede azalan kelimeler tespit edilmiştir. Ardından önceki dönemlerde bulunmayıp Cumhuriyet döneminde ilk kez ortaya çıkan kelimeler tespit edilmiştir.

Farklı dönemlerdeki eserlerde kullanılan kelimelerin sıklıklarının değişim gösterip göstermediğini görebilmek amacıyla her dönemde her bir kelimenin kullanım sıklığı çıkarılmış ve aşağıda kronolojik olarak kullanılmazken kullanılır hale gelen ve kullanılırken kullanımdan düşen kelimeler tablolar halinde verilmiştir.

Kelime	Tanzimat (%)	Tanzimat (frekans)	Servet-i Fünûn (%)	Servet-i Fünûn (frekans)	II. Meşrutiyet (%)	II. Meşrutiyet (frekans)	Cumhuriyet (%)	Cumhuriyet (frekans)
cihet/cihetle	0.9	484	0	18	0.3	202	0	21
eyle	3	1602	0	4	0.5	359	0	12
hâne	0.6	337	0	0	0.4	246	0	8
derece	0.9	518	0.3	117	0.4	281	0.3	289
sûret	0.7	376	0.2	67	0.5	337	0.1	98
tabip/tabib	0.2	83	0.1	25	0	22	0	1
bâdehû	0.1	60	0	0	0	12	0	0
biçâre	0.5	278	0.2	65	0.1	88	0.1	91
kıyâm	0.1	54	0	0	0	10	0	0
vâkıâ	0.3	154	0	8	0	9	0.1	49

**Tablo 5.** Kullanımı azalan veya ortadan kalkan kelimeler

Görülebileceği gibi *cihet* veya *cihetle* kelimesi Tanzimat döneminde %0.9 (binde dokuz) kullanılırken Cumhuriyet döneminde kullanımı neredeyse tamamen bitmiştir. *Eyle* ile yapılan kelimeler (eylemek, eyledi, eylemişti vb.) Tanzimat döneminde %3 (yüzde üç) düzeyinde kullanılırken Cumhuriyet döneminde neredeyse tamamen kaybolmuştur. Benzer şekilde *hâne*, *derece*, *sûret*, *biçâre* ve *vâkıâ* kelimelerinin kullanımları önemli derecede azalırken *tabip/tabib* (sadece bir defa), *bâdehû* ve *kıyâm* kelimelerinin kullanımı tamamen ortadan kalkmıştır.

Bunlara karşılık Cumhuriyet döneminde ilk kez ortaya çıkan ya da kullanımı sıklaşan kelimeler de tespit edilmiş ve aşağıda Tablo 6'da verilmiştir.

Kelime	Tanzimat (yüzde, frekans)	Servet-i Fünûn (yüzde, frekans)	II. Meşrutiyet (yüzde, frekans)	Cumhuriyet (yüzde, frekans)
hatırlama	0.1 (61)	0.1 (24)	0.2 (145)	0.6 (525)
evlenmek	0.0 (23)	0.2 (56)	0.2 (142)	0.6 (487)
sanmak	0.0 (26)	0.0 (1)	0.0 (25)	0.3 (290)
ömür	0.2 (124)	0.1 (37)	0.2 (118)	0.3 (287)
yaşam	0.1 (41)	0.3 (101)	0.1 (87)	0.3 (275)
kımıldamak	0.0 (18)	0.0 (9)	0.0 (31)	0.3 (225)
sinir	0.0 (9)	0.1 (22)	0.1 (40)	0.3 (223)
duygu	0.0 (1)	0.0 (13)	0.0 (7)	0.2 (157)
hâki	0.0 (2)	0.0 (1)	0.0 (2)	0.1 (95)
müderris	0.0 (0)	0.0 (0)	0.0 (0)	0.1 (54)
imkân/imkânsız	0.0 (27)	0.0 (0)	0.0 (0)	0.1 (111)
süre	0.0 (2)	0.0 (1)	0.0 (2)	0.0 (39)
İlk/ilkin	0.0 (1)	0.0 (0)	0.0 (1)	0.0 (38)
tapu	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (35)
durak/duraklamak	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (33)

**Tablo 6.** Kullanımı artan veya ilk kez ortaya çıkan kelimeler

Görülebileceği gibi kullanımı en çok artan kelimeler *hatırlama* ve *evlenmek* kelimeleridir. İki kelime de önceki dönemlerde en çok %0.2 (binde iki) oranında kullanılırken Cumhuriyet döneminde %0.6 (binde altı) oranında kullanılmaya başlamıştır. *Sanmak*, *ömür* ve *yaşam* ile yapılmış kelimelerinin Cumhuriyet dönemindeki kullanımları giderek artarak %0.3 (binde üç) düzeyine ulaşmıştır. *Duygu*, *hâki*, *müderris*, *imkân* gibi kelimelerden türetilen kelimelerin kullanımı ise yaklaşık %0 düzeyinden

%0.1 (binde bir) düzeyine çıkmıştır. *Süre, ilk, tapu, durak* ve bunlarla yapılan kelimeler hemen hemen hiç kullanılmazken Cumhuriyet dönemi eserlerinde karşımıza çıkmaya başlamıştır. Bunların oranı %0.1 (binde bir)'den düşük olmakla birlikte bu dönemde ilk kez ortaya çıkmış olmaları açısından tabloya eklenmişlerdir.

Tanzimat döneminden Cumhuriyet dönemine Osmanlıca tamlama kullanımının değişimini incelemek amacıyla iki dönemde en çok kullanılan tamlama kelimeleri belirlenmiş ve bunların frekansları da bulunmuştur. Aşağıda Tablo 7'de her dönemde en çok kullanılan dört tamlama yapan kelimelere ait bulgular sunulmuştur.

Tanzimat dönemi		Cumhuriyet dönemi	
Kelime	frekans	Kelime	frekans
Nazâr-ı	118	Kâtib-i	35
Kemâl-i	111	Harb-i	33
Hüsn-i	98	Müdür-i	33
Sûret-i	91	Zât-ı	19

**Tablo 7.** *Tanzimat ve Cumhuriyet dönemlerinde en çok kullanılan tamlama kelimeleri*

Görülebileceği gibi Tanzimat döneminde Osmanlıca tamlama kullanım sıklığı daha yüksekken Cumhuriyet döneminde sıklık azalmıştır. Örneğin Tanzimat döneminde *nazâr-ı* kelimesi ile yapılan tamlamalar 118 kez tekrar ederken Cumhuriyet döneminin en sık *kâtib-i* kelimesi ile tamlama yapılmıştır ve sıklığı 35'tir. Görüleceği gibi Osmanlıca tamlama kullanımı sıklığında önemli bir düşüş olmuştur.

Her dönemde yer alan eserlerin kelime kullanımını kelimelerin türlerine göre incelemek amacıyla kelimeler lemmatizasyon yöntemi ile köklerine ayrılmış, kök kelimelerin fiil, isim, sıfat ve bağlaç gibi kullanım amaçları tespit edilmiş ve aşağıda Tablo 8'de gösterilmiştir.

Kategori	Tanzimat (%)	Servet-i Fünûn (%)	II. Meşrutiyet (%)	Cumhuriyet (%)
Sıfat	8	10	10	10
Zarf	4	5	5	4
İsim	40	38	40	41
Fiil	21	23	20	22
Belirteç	6	9	7	7
Diğer	21	15	18	16

**Tablo 8.** *Her dönemde kelime türlerinin (lemma'ların) oranları*

Tablo 8'den görüleceği gibi Tanzimat döneminde sıfat kullanımı %8 civarındayken sonraki dönemlerde %10'a çıkmıştır. Zarf, isim ve fiil kullanımı tüm dönemlerde birbirine çok yakındır. Tüm dönemlerde isim, fiil, zarf, sıfat, belirteç kullanımları birbirine yakın oranlardadır. Tüm dönemlerde cümlelerin içinde en çok yer alan öge %40 kullanım oranıyla isimlerdir. İkinci sırada %21 kullanım oranı bulunan fiiller gelmektedir. Üçüncü sırada %10 oranında kullanımıyla sıfatlar bulunmaktadır. Belirteçler %7 oranında kullanımlarıyla kullanım sıklığı bakımından dördüncü sırada yer alırken en düşük oranda (%5) kullanılan öge zarflardır. Her dönemde bu kullanım sıklıkları birbiriyle aynı yoğunlukta görünmektedir. Bir başka deyişle bir dönemde isimler daha çok kullanılırken diğer dönemde fiillerin daha çok kullanılması gibi bir durum gözlenmemektedir.

## Sonuç ve Tartışma

Bu çalışma, Türk edebiyatının dört farklı dönemi boyunca meydana gelen dilsel değişimlere dair kapsamlı bir nicel değerlendirme sunarak, dönüştürücü bir tarihsel dönemde dilin evrimine ilişkin bulgular sağlamaktadır. Metin madenciliği ve stilometri araçlarıyla desteklenen analiz, önceki araştırmalarla uyumlu ve bu çalışmaları genişleten kalıpları ortaya koymaktadır.

Çalışma, Türkçe kullanımında devam eden sadeleşme sürecini doğrulamaktadır; bu olgu, literatürde geniş çapta ele alınmıştır (Halman, 2011; Seyfettin, 1914a; Tanpınar 2013; Yücel, 1982). Nicel sonuçlar, Tanzimat döneminin sonraki dönemlerin kelime hazinesine önemli ölçüde katkı sağladığını ve bu döneme özgü kelimelerin %41,51'inin Cumhuriyet dönemine kadar varlığını sürdürdüğünü göstermektedir. Bu bulgu, Namık Kemal ve diğerlerinin erken dönem dil reformu çabalarını yansıtarak, konuşma ve yazı dilini uyumlu hale getirme gayretlerini ve kelime zenginliğini koruma çabalarını desteklemektedir (Akyüz, 1995; Polat, 2020).

Benzer şekilde, Cumhuriyet döneminin yeni kelimeler kazandırmadaki üstünlüğü (44.454 özgün kelime), Dil Devrimi'nin dönüştürücü etkisini vurgulamaktadır. Ataç (1962) gibi yazarlar, bu değişimlerin yeni bir ulusal kimlikle dili uyumlu hale getirme gerekliliği tarafından yönlendirildiğini öne sürmüştür ve bu çalışma da gözlemlenen kelime çeşitliliği ile bu iddiayı desteklemiştir. Türkçe kökenli evlenmek, duygu, yaşam, sanmak gibi kelimelerin kullanımının arttığı; buna karşılık Arapça ve Farsça kökenli cihet, vâkiâ, bâdehû, tabip gibi kelimelerin kullanımının azaldığı görülmektedir.

Dikkat çekici bir eğilim, Tanzimat ve Cumhuriyet dönemleri arasında Osmanlı Türkçesine ait *nazâr-ı*, *hüsn-i*, *sûret-i* gibi ifadelerle yapılan tamlamaların kullanımındaki belirgin düşüştür. Önceki eserlerde öne çıkan bu ifadeler, büyük ölçüde modern yapılarla değiştirilmiştir. Bu değişimler, Yeni Lisan hareketi ve Dil Devrimi gibi dilsel sadeleştirme hedefleriyle uyumludur. Nurullah Ataç ve diğerlerinin gözlemlerine göre, bu değişimler Cumhuriyet dönemi boyunca kültürel ve siyasi motivasyonlarla şekillenmiştir (Yücel, 1982).

Söz varlığındaki değişikliklere rağmen, kelime türlerinin (isimler, fiiller, sıfatlar vb.) işlevsel dağılımı tüm dönemlerde tutarlı kalmıştır. Bu durum Türk nesrinin yapısal temellerinin, söz varlığı evrilirken bile büyük ölçüde sabit kaldığını göstermektedir.

Servet-i Fünûn dönemi, yüksek ortalama kelime sayısı ile dikkat çekerken, nispeten düşük çeşitlilik oranı, daha süslü ancak daha az çeşitli bir nesir eğilimini göstermektedir. Buna karşılık II. Meşrutiyet dönemi, Yeni Lisan hareketinin dilde sadelik çağrısındaki başarısını yansıtarak daha yüksek bir çeşitlilik oranı sergilemektedir.

Bu çalışma, dilsel evrime ilişkin nitel iddialara istatistiksel doğrulama ya da yanıtlama imkânı sağlayarak, Türk edebiyatı araştırmalarında kritik bir boşluğu doldurmaktadır. Daha önceki çalışmalar, Çalışkan (2020) ve Patton ve Can'ın (2004) araştırmalarında olduğu gibi, stilometrik yöntemlerin faydasını göstermiş olsa da bu yöntemlerin Türk edebiyatına uygulanması sınırlı kalmıştır. Yüzyılı aşkın bir süreyi kapsayan 29 kanonik romanı analiz ederek, bu çalışma yalnızca mevcut bulguları doğrulamakla kalmamış, aynı zamanda kelime kullanım kalıpları ve kelime çeşitliliği hakkında yeni bilgiler sunmuştur. Çalışma ayrıca, Türk edebiyatı araştırmalarında ilk kez dört farklı dönemi nicel yöntemlerle karşılaştırmıştır.

Sonuçlar, dil evrimi ve sosyopolitik değişim arasındaki derin ilişkiyi vurgulamaktadır. Osmanlı Türkçesi unsurlarının azalması ve modern söz varlığının ortaya çıkışı, Tanzimat reformlarından Cumhuriyet döneminin ulusal kimlik vurgusuna kadar uzanan daha geniş kültürel değişimleri yansıtmaktadır. Gelecekteki araştırmalar, aynı dönemlerden alınan edebi olmayan metinleri içeren karşılaştırmalı analizlerle bu bulguları genişletebilir ve sosyopolitik faktörlerin dil kullanımına etkisini daha derinlemesine keşfedebilir.

Gelecek çalışmalar, bu yöntemleri şiir, tiyatro, öykü gibi diğer edebi türlere uygulayabilir. Ayrıca, anlatım stillerini ve stilistik özellikleri daha ayrıntılı inceleyebilir veya çağdaş okuyucuların dilsel değişikliklere ilişkin tepkilerini analiz edebilir. Böyle bir disiplinlerarası araştırma, dilin sosyopolitik ve kültürel güçlere nasıl tepki verdiğine dair daha bütüncül bir anlayış sağlayabilir.

### Kaynakça

- ABBYY. (2024). ABBYY Fine Reader OCR Editor (Version 16.0.14). Erişim adresi: <https://www.abbyy.com/ocr-sdk/>
- Ahmet Mithat Efendi. (1897). Dekadanlar. *Sabah*. (22 Mart 1897), 2628.
- Akgün, A. N., Çalışır, Z. D., & Kaya, T. (2022). Investigation of Turkish literary poetry by unsupervised text mining methods. In Kahraman, C., Cebi, S., Çevik Onar, S., Oztaysi, B., Tolga, A. C., & Sarı, I. U. (Eds.), *Intelligent and fuzzy techniques for emerging conditions and digital transformation. Lecture Notes in Networks and Systems* (Vol. 307, pp. 105). Springer, Cham. Erişim adresi: [https://doi.org/10.1007/978-3-030-85626-7\\_105](https://doi.org/10.1007/978-3-030-85626-7_105)
- Akın, A. A., & Akın, M. D. (2007). Zemberek, an open source NLP framework for Turcic languages. *Structure*, 10(2007), 1-5. Erişim adresi: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=82b22dbb84c5bff7352fe93b80f66a1bbf4b9c80>
- Akyüz, K. (1995). *Modern Türk edebiyatının ana çizgileri*. İstanbul: İnkılâp Kitabevi.
- Alkan, B. B., Karakus, L., & Direkci, B. (2023). Generating a sentiment dictionary in R and dictionary-based sentiment analysis in Turkish texts. *Digital Scholarship in the Humanities*, 38(2), 461–476. <https://doi.org/10.1093/lc/fqac093>
- Argunşah, H. (2011). Milli Edebiyat anlayışı ve Yeni Lisan. *İkinci Meşrutiyet Dönemi Edebiyatı*. Eskişehir: Anadolu Üniversitesi Yayınları.
- Ataç, N. (1962). *Söyleşiler*. Ankara: Türk Dil Kurumu Yayınları.
- Aydemir, M. (2016). Tanzimat'tan Yeni Lisan Hareketi'ne dil sadeleştirme çalışmaları. *SOBIDER: Sosyal Bilimler Dergisi*, 9, 1–10.
- Balakrishnan, V., & Lloyd-Yemoh, E. (2014). Stemming and lemmatization: A comparison of retrieval performances. *Journal of Information Science*, 40(3), 335–345.
- Balcı, Y. (2012). II. Meşrutiyet döneminde eleştiri. R. Filizok ve M. Dayanç (Ed.), *Eleştiri Tarihi* içinde (Ünite 6). Eskişehir: Anadolu Üniversitesi Yayınları.
- Can, F., & Patton, J. M. (2010). Change of Word Characteristics in 20th-Century Turkish Literature: A Statistical Analysis. *Journal of Quantitative Linguistics*, 17(3), 167–190. <https://doi.org/10.1080/09296174.2010.485444>

- Chen, H., Boutros, P. C. (2011). VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* 12, 35. <https://doi.org/10.1186/1471-2105-12-35>
- Çalışkan, S. (2020). *Text mining analysis of translation, social communication and literary writing for Turkish* (Order No. 29046761). Erişim adresi: <https://www.proquest.com/dissertations-theses/text-mining-analysis-translation-social/docview/2652593633/se-2>
- Dilai, M., & Dilai, L. (2024). Literary Text Mining Using Verb Feature Clustering. Proceedings of the 6th International Workshop on Modern Data Science Technologies (MoDaST-2024), Lviv-Shatsk, Ukraine, May 31 - June 1, 2024. CEUR Workshop Proceedings, 3723, p. 15.
- Doğuş, Ö., Ayaç, Ö. B. ve Silahtaroglu, G. (2020). Lemmatizer: Akıllı Türkçe kök bulma yöntemi. *Turkish Studies-Information Technologies and Applied Sciences*, 15(3), 289-299.
- Gökner, E. (2008). The novel in Turkish: Narrative tradition to Nobel prize. R. Kasaba (Eds.), In *The Cambridge history of Turkey: Volume 4, Turkey in the modern world* (pp. 472–503). Cambridge University Press. <https://doi.org/10.1017/CHOL9780521620963.018>
- Halman, T. S. (2011). *A millennium of Turkish literature: A concise history*. Ed. J. L. Warner. Syracuse, New York: Syracuse University Press.
- IntelliJ IDEA. (2024). *The leading Java and Kotlin IDE* (Version 2024.3). Erişim adresi: <https://www.jetbrains.com/idea/>
- İmer, K. (1976). *Dilde değişme ve gelişme açısından Türk dil devrimi*. Ankara: TDK Yayınları.
- Jääskeläinen, A., Lipsanen, M., Föhr, A., & Räisänen, T. (2023). OCR quality: Key to enhanced data mining. In *2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)* (pp. 1–6). IEEE. <https://doi.org/10.1109/ICECCME57830.2023.10252214>
- Jain, P., Taneja, K., & Taneja, H. (2021). Which OCR toolset is good and why: A comparative study. *Kuwait Journal of Science*, 48(2).
- Khyani, D., Siddhartha, B. S., Niveditha, N. M., & Divya, B. M. (2021). An interpretation of lemmatization and stemming in natural language processing. *Journal of University of Shanghai for Science and Technology*, 22(10), 350-357.
- Kurdakul, Ş. (1994). *Çağdaş Türk edebiyatı*. İstanbul: Evrensel Basım Yayın.
- Mori, S., Suen, C. Y., & Yamamoto, K. (1992). Historical review of OCR research and development. *Proceedings of the IEEE*, 80(7), 1029–1058. <https://doi.org/10.1109/5.156468>
- Namık Kemal. (1866). Lisan-ı Osmani'nin edebiyatı hakkında bazı mülahazatı şamildir. M. Kaplan, İ. Enginün ve B. Emil (Ed.), *Yeni Türk Edebiyatı Antolojisi II* (s. 183–192). İstanbul: Marmara Üniversitesi Yayınları. (Basım: 1993).
- Ömer Seyfettin (1911). *Genç Kalemler*. İ. Parlatır ve N. Çetin (Ed.), *Genç Kalemler Dergisi* (s. 75–81). Ankara: Türk Dil Kurumu Yayınları. (1999)
- Ömer Seyfettin (1914a). Türkçeye karşı Enderunca. *Türk Sözü*, (4), 1 Mayıs 1914.

- Ömer Seyfettin (1914b). Osmanlıca değil Türkçe. *Türk Sözü*, (5), 8 Mayıs 1914.
- Önertoy, O. (1984). *Cumhuriyet Dönemi Türk Roman ve Öyküsü*. İstanbul: İş Bankası Yayınları.
- Patton, J. M., & Can, F. (2004). A Stylometric Analysis of Yaşar Kemal's "İnce Memed" Tetralogy. *Computers and the Humanities*, 38(4), 457–467. <http://www.jstor.org/stable/30204956>
- Polat, N. H. (2020). 109. yılında Yeni Lisan'a eleştirel bir bakış. *Türk Dili*, (820), 14-23.
- R Core Team. (2024). *R: A language and environment for statistical computing* (Version 4.4.2) [Software]. R Foundation for Statistical Computing. Erişim adresi: <https://www.R-project.org/>
- Rıza Tevfik. (1896). Mebahis-i lisan. *Servet-i Fünûn*, XI(266).
- RStudio Team (2024). *RStudio: Integrated Development for R*. (Version: 2024.09.1+394) [Software]. Erişim adresi: <http://www.rstudio.com/>
- Silge, J., & Robinson, D. (2016). Tidytext: Text mining and analysis using tidy data principles in R. *Journal of Open Source Software*, 1(3). <https://doi.org/10.21105/joss.00037>
- Stone, L. (2010). Inside Turkish literature: Concerns, references, and themes. *Turkish Studies*, 11(2), 235-250.
- Sun, X. (2024). Intelligent text clustering analysis of novels based on digital semantic similarity calculation. *Computer-Aided Design and Applications*, 21(S16), 199-213.
- Tafti, A. P., Baghaie, A., Assefi, M., Arabnia, H. R., Yu, Z., & Peissig, P. (2016). OCR as a service: An experimental evaluation of Google Docs OCR, Tesseract, ABBYY FineReader, and Transym. In G. Bebis, R. Boyle, B. Parvin, & D. Koracin (Eds.), *Advances in visual computing: Proceedings of the 12th international symposium (ISVC 2016)*, pp. 735–746). Springer. [https://doi.org/10.1007/978-3-319-50835-1\\_71](https://doi.org/10.1007/978-3-319-50835-1_71)
- Tanpınar, A. H. (2013). *On dokuzuncu asır Türk edebiyatı tarihi* (22. basım). İstanbul: Dergâh Yayınları.
- Tunalı, V. ve Bilgin, T. T. (2012). Türkçe metinlerin kümelenmesinde farklı kök bulma yöntemlerinin etkisinin araştırılması. *Proceedings of Elektrik Elektronik ve Bilgisayar Mühendisliği Sempozyumu (ELECO)* (ss. 598–602).
- Ullah, Z., & Mahmood, A. (2020). Stylometry of short stories through Voyant Corpus Summary Tool: A text mining study. *Kashmir Journal of Language Research*, 23(1), 1-15.
- Uşaklıgil, H. Z. (1899). Karilerime mektuplar. *Servet-i Fünûn*, (428), 179-182.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>

- Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). dplyr: A grammar of data manipulation (Version 1.1.4) [R package]. Erişim adresi: <https://github.com/tidyverse/dplyr>, <https://dplyr.tidyverse.org>
- Yıldırım, O., & Can, F. (2024). A Quantitative Style Analysis of Four Turkish Authors: Changes Over Time, and Differences. *Journal of Quantitative Linguistics*, 31(4), 353–384. <https://doi.org/10.1080/09296174.2024.2412309>
- Yücel, T. (1982). *Dil devrimi ve sonuçları*. Ankara: Türk Dil Kurumu Yayınları.
- Zenkov, A., Zenkov, E., Zenkov, M., & Sazanova, L. (2021). Numerals in authorial Turkish-language texts and the stylometric analysis. In *E3S Web of Conferences* (Vol. 270, p. 01038). EDP Sciences.