



Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü  
Bilgi ve Belge Yönetimi Anabilim Dalı

## **Türkçe Akademik Yayınlar İçin Yapısal Öz Çıkarım Sistemi**

Ayşe Esra ÖZKAN ÇELİK

Doktora Tezi

Ankara, 2021



# TÜRKÇE AKADEMİK YAYINLAR İÇİN YAPISAL ÖZ ÇIKARIM SİSTEMİ

Ayşe Esra ÖZKAN ÇELİK

Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü

Bilgi ve Belge Yönetimi Anabilim Dalı

Doktora Tezi

Ankara, 2021

## KABUL VE ONAY

Ayşe Esra Özkan Çelik tarafından hazırlanan “Türkçe Akademik Yayınlar İçin Yapısal Öz Çıkarım Sistemi” başlıklı bu çalışma, 21.06.2021 tarihinde yapılan savunma sınavı sonucunda başarılı bulunarak jürimiz tarafından doktora tezi olarak kabul edilmiştir.

---

Prof. Dr. Aydın Erar (Başkan)

---

Prof. Dr. Umut Al (Danışman)

---

Doç. Dr. Harun Artuner (Üye)

---

Doç. Dr. Umut Sezen (Üye)

---

Dr. Öğretim Üyesi Aysun Güran (Üye)

Yukarıdaki imzaların adı geçen öğretim üyelerine ait olduğunu onaylarım.

Prof. Dr. Uğur Ömürgönülşen

Enstitü Müdürü

## YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin tamamını veya herhangi bir kısmını, basılı (kağıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinleri yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan “**Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge**” kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- Enstitü / Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren 2 yıl ertelenmiştir. <sup>(1)</sup>
- Enstitü / Fakülte yönetim kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren ..... ay ertelenmiştir. <sup>(2)</sup>
- Tezimle ilgili gizlilik kararı verilmiştir. <sup>(3)</sup>

19/07/2020

**Ayşe Esra ÖZKAN ÇELİK**

<sup>1</sup> “Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge”

- (1) Madde 6. 1. Lisansüstü teze ilgili patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda, tez danışmanının önerisi ve **enstitü anabilim dalının** uygun görüşü üzerine **enstitü** veya **fakülte yönetim kurulu** iki yıl süre ile tezin erişime açılmasının ertelenmesine karar verebilir.
- (2) Madde 6. 2. Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internetten paylaşılması durumunda 3. şahıslara veya kurumlara haksız kazanç imkanı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez danışmanının önerisi ve **enstitü anabilim dalının** uygun görüşü üzerine **enstitü** veya **fakülte yönetim kurulunun** gerekçeli kararı ile altı ayı aşmamak üzere tezin erişime açılması engellenebilir.
- (3) Madde 7. 1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, **tezin yapıldığı kurum** tarafından verilir \*. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlere ilişkin gizlilik kararı ise, **ilgili kurum ve kuruluşun önerisi** ile **enstitü** veya **fakültenin** uygun görüşü üzerine **üniversite yönetim kurulu** tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.  
Madde 7.2. Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir.

\* Tez danışmanının önerisi ve **enstitü anabilim dalının** uygun görüşü üzerine **enstitü** veya **fakülte yönetim kurulu** tarafından karar verilir.

## ETİK BEYAN

Bu alıřmadaki bütn bilgi ve belgeleri akademik kurallar erevesinde elde ettiđimi, grsel, iřitsel ve yazılı tm bilgi ve sonuları bilimsel ahlak kurallarına uygun olarak sunduđumu, kullandıđım verilerde herhangi bir tahrifat yapmadıđımı, yararlandıđım kaynaklara bilimsel normlara uygun olarak atıfta bulunduđumu, tezimin kaynak gsterilen durumlar dıřında zgn olduđunu, Prof. Dr. Umut Al danıřmanlıđında tarafımdan retildiđini ve Hacettepe niversitesi Sosyal Bilimler Enstits Tez Yazım Ynergesine gre yazıldıđını beyan ederim.

Ayře Esra ZKAN ELİK



## SUMMARY

ÖZKAN ÇELİK, Ayşe Esra. *Structured Abstract Extraction System for Turkish Academic Publications*. Ph.D. Dissertation, Ankara, 2021.

With the widespread use of information technologies in all areas of life, the long-lasting problem of researchers' inability to read the full texts of articles has gained a different dimension. Researchers are also tackling new challenges while studying the massive amount of information they access in the electronic media due to the nature of electronic versions of texts, the online media, and the digital screens. Abstracts are the most important tools that will enable potential readers to select the relevant full texts from this huge stack of electronic information. However, whereas their primary aim is to facilitate the reading of the full texts, in most cases the abstracts are produced with more readability issues and structural deficiencies in their contents when compared the actual texts.

In this dissertation, ways of contributing to the readers' experience of meeting scientific articles in today's conditions, where technology constantly transforms people's interaction with information, were investigated. Accordingly, the study focuses on the contribution of the use of structural methods in obtaining the abstracts of scientific articles through automatic methods. The system developed in the study was designed to operate with the principle of choosing representative sentences from all structural parts of the full texts of the articles. The automatically structured abstracts obtained are superior to the classical version of outputs of the system. The calculated legibility values show that all the outputs are significantly easier to read than those produced by authors. The summarization system, which was developed within the scope of the study, is independent from the data set. The summarization system can also be used for other Turkish document groups. It is believed that the results of this study and the data set, which was extracted from a total of 421 articles published in the field of Turkish librarianship and information science, can be used in other studies as well. It is considered that this structured abstract preparation system can be used to produce better readable automatic structural abstracts for the corpus produced with classical abstracts, and can also contribute to the scientific communication process as an auxiliary tool for authors and editors.

**Keywords** Scholarly communication, information retrieval, text summarization.



## İÇİNDEKİLER

KABUL VE ONAY .....	i
YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI.....	ii
ETİK BEYAN .....	iii
ÖZET.....	iv
SUMMARY .....	v
İÇİNDEKİLER .....	vi
ŞEKİLLER DİZİNİ.....	ix
TABLOLAR DİZİNİ .....	xi
TEŞEKKÜR.....	xii
1. BÖLÜM: GİRİŞ .....	1
1.1. KONUNUN ÖNEMİ.....	1
1.2. ARAŞTIRMANIN AMACI VE HİPOTEZİ.....	13
1.3. ARAŞTIRMANIN SORULARI .....	16
1.4. ARAŞTIRMADA KULLANILAN YÖNTEM VE TEKNİKLER.....	16
1.5. ARAŞTIRMANIN KAPSAMI .....	19
1.6. ARAŞTIRMANIN DÜZENİ .....	19
2. BÖLÜM: LİTERATÜR DEĞERLENDİRMESİ.....	21
2.1. BİLİMSEL YAYINLARIN EVRİMİ VE İMRAD .....	21
2.2. BİLİMSEL YAYINLARIN ÖZLERİ .....	24
2.3. ÖZLERLE İLGİLİ AKADEMİK ÇALIŞMALAR.....	30
2.4. GENEL HATLARIYLA OTOMATİK METİN ÖZETLEME .....	34
2.5. ÇIKARIMA DAYALI OTOMATİK METİN ÖZETLEME YAKLAŞIMI.....	36
2.6. OTOMATİK ÖZETLEME SİSTEMLERİNİN DEĞERLENDİRİLMESİ.....	38
2.6.1. Metin Kalitesinin Değerlendirilmesi .....	40

2.6.2. İçerik Değerlendirilmesi.....	41
2.6.2.1. Birlikte Seçim Ölçümleri.....	41
2.6.2.2. İçerik Temelli Ölçümler.....	41
2.6.2.2.1. Kosinüs Benzerliği.....	41
2.6.2.2.2. Birim Çakışması .....	42
2.6.2.2.3. En Uzun Ortak Alt Dizi .....	42
2.6.2.2.4. ROUGE-N: N-gram Birliktelik İstatistikleri .....	43
2.7. TÜRKÇE OTOMATİK METİN ÖZETLEME ÇALIŞMALARI .....	44
3. BÖLÜM: YÖNTEM .....	47
3.1. DERLEM VE ÖZELLİKLERİ .....	47
3.2. METİN ÖN İŞLEMLERİ.....	49
3.3. DOKÜMAN VEKTÖRLERİ VE AĞIRLIKLANDIRMA.....	54
3.4. OTOMATİK CÜMLE SEÇİM İŞLEMLERİ .....	55
3.4.1. Cümle Puanlamada Kullanılan Özellikler .....	56
3.4.2. Cümle Sıralama ve Seçim İşlemi .....	58
3.5. OKUNURLUK HESAPLAMALARI.....	62
3.6. OTOMATİK ÖZLERİN DEĞERLENDİRİLMESİ .....	64
4. BÖLÜM: BULGULAR.....	65
4.1. BD VE TK YAZIM KURALLARI .....	65
4.2. TÜRKÇE KÜTÜPHANECİLİK VE BİLGİBİLİM ALANINDA YAYINLANMIŞ MAKALELERİN YAPISAL BİÇİMLERİ .....	67
4.3. TÜRKÇE KÜTÜPHANECİLİK VE BİLGİBİLİM ALANINDA YAYINLANMIŞ MAKALELERİN ÖZ-TAM METİN OKUNURLUKLARI .....	70
4.4. OTOMATİK YAPISAL ÖZLERİN DEĞERLENDİRİLMESİ .....	80
4.4.1. OYÖ ve OKÖ Cümlelerinin Tam Metin İçerisindeki Dağılımları.....	81
4.4.2. Otomatik Yapısal Özlerin Okunurluğuna Dair Bulgular .....	88

4.4.3. Sistem Çıktılarının Yazar Tarafından Üretilen Orijinal Öz Üzerinden ROUGE Değerlendirmesi .....	91
5. BÖLÜM: SONUÇ VE ÖNERİLER.....	99
KAYNAKÇA .....	108
EK 1. BD200023, Yazar tarafından üretilmiş orijinal öz, OYÖ ve OKÖ .....	114
EK 2. BD201213, Yazar tarafından üretilmiş orijinal öz, OYÖ ve OKÖ .....	116
EK 3. TK200221, Yazar tarafından üretilmiş orijinal öz, OYÖ ve OKÖ.....	119

## ŞEKİLLER DİZİNİ

Şekil 1. 1951-1987 yılları arasında dünyada tahmini dergi sayıları .....	22
Şekil 2. Makalelerin yapısal bölümleri diyagramı .....	24
Şekil 3. Yapısal öz örneği .....	28
Şekil 4. Önemli cümle öz örneği.....	29
Şekil 5. Özet değerlendirme ölçümlerinin taksonomisi .....	39
Şekil 6. Otomatik yapısal özetleme sistemi mimarisi .....	48
Şekil 7. Makale seçim arayüzü.....	50
Şekil 8. Makale üstveri ayrıştırma arayüzü.....	50
Şekil 9. Makale gövde metni paragraf ayrıştırma/IMRAD işaretleme arayüzü.....	51
Şekil 10. Kelime sıklık vektörlerinin üstverileri .....	54
Şekil 11. Kelime sıklık vektörlerinin temsili .....	54
Şekil 12. OYÖ ve OKÖ için tam metinden cümle seçimleri .....	60
Şekil 13. OKÖ cümlelerinin seçime hazırlanmış görüntüsü .....	61
Şekil 14. OYÖ cümlelerinin seçime hazırlanmış görüntüsü.....	61
Şekil 15. Tam metinlerde rastlanan yapısal düzenlerin makale sayılarına göre ağırlıkları.....	68
Şekil 16. Makalelerdeki yapısal bölüm düzeninin yıllara göre değişimi .....	69
Şekil 17. Öz okunurluklarının karakter sayısına göre saçılımı .....	72
Şekil 18. Tam metin okunurluk (a), orijinal öz okunurluk (b), tam metin kelime sayısı (c), orijinal öz kelime sayısı (d) yığılımları.....	77
Şekil 19. Tam metin (a), orijinal öz (b) okunurluk dağılımları.....	78
Şekil 20. Tam metin, orijinal öz karşılaştırmalı okunurluk değerleri .....	79
Şekil 21. Tam metin, orijinal öz AOD ilişkisi .....	80
Şekil 22. Tam metin içeriklerinin IMRAD'a göre sınıflandırılması.....	82

Şekil 23. OKÖ (kırmızı), OYÖ (mavi) IMRAD düzenlerinin, tam metin IMRAD düzenlerine göre sınıflandırılması.....	84
Şekil 24. Derlem cümlelerinin IMRAD sayısı ve düzenine göre ağırlıkları.....	85
Şekil 25. OYÖ (a), OKÖ (b) cümlelerinin derlem IMRAD desenine göre IMRAD bölümü ağırlıkları.....	86
Şekil 26. OYÖ, OKÖ IMRAD bölümü sayıları .....	87
Şekil 27. Sistem çıktısı OYÖ (a) ve OKÖ (b)'lerin okunurluk dağılımları .....	88
Şekil 28. Derlem okunurluk değerleri .....	89
Şekil 29. Tam metin (a), orijinal öz (b), OYÖ (c) ve OKÖ (d) okunurluk dağılımları ..	90
Şekil 30. IMRAD'a göre ROUGE-1, ROUGE-2, ROUGE-L ve ROUGE-SU4 ölçümlerinin anma (a), duyarlık (b) ve f-ölçüm (c) değerleri.....	93
Şekil 31. Orijinal öz, OYÖ, OKÖ kelime sayıları.....	94
Şekil 32. IMRAD'a göre ROUGE-1 (a), ROUGE-2 (b), ROUGE-SU4 (c) ölçümlerinin f-ölçüm değerleri.....	98

**TABLolar DİZİNİ**

Tablo 1. Tez çalışması kapsamında kullanılan yöntemler ve araçlar.....	18
Tablo 2. Öz yazımında yapılması ve kaçınılması gerekenler .....	26
Tablo 3. Çalışmada kullanılan yapısal bölümler .....	52
Tablo 4. Ateşman okunabilirlik ölçeği.....	62
Tablo 5. IMRAD'a uyumlu bölüm sayıları ve yapısal düzenlerine göre makale sayıları.....	67
Tablo 6. Makale tam metin ve orijinal özlerinin okunurluk özet tablosu .....	71
Tablo 7. OYÖ ve OKÖ'lerin IMRAD'a göre ROUGE ölçümlerinin ortalama f-ölçüm değerleri .....	96

## TEŞEKKÜR

Çoğunlukla boyumu aşan denizlerde bol su yutarak yüzmeyi öğrendiğim süreçlerin bir araya gelmiş halidir bu tez. Bu süreçlerde, beni her zaman daha iyisi için yönlendiren çok değerli danışmanım Prof. Dr. Umut Al'a emekleri için çok teşekkür ederek söze başlamalıyım. Çalışma boyunca kendisine her ihtiyacım olduğunda tezime vakit ayırdı ve değerli görüşlerini paylaştı. Sadece tezin ortaya çıkmasında değil benim araştırmacı olarak gelişimimde de katkıları büyüktür. Kendisiyle çalışma şansım olduğu için çok mutluyum. Tekrar teşekkür ederim.

Çalışmamda jüri olarak yer alan değerli Hocalarım Prof. Dr. Aydın Erar, Doç. Dr. Harun Artuner, Doç. Dr. Umut Sezen, Dr. Öğretim Üyesi Aysun Güran ile tanışmaktan ve çalışmamda isimlerinin yer almasından büyük gurur duymaktayım. Hepsine değerli katkıları için çok teşekkür ederim. Özellikle tez izleme komitelerimden itibaren bu yolculukta araştırmam için emek sarf eden değerli hocalarım Prof. Dr. Aydın Erar ve Doç. Dr. Umut Sezen'in tezin gelişimine katkıları büyüktür. Değerli Hocam Doç. Dr. Umut Sezen'e araştırma sürecinde cesaretimin kırıldığı dönemlerde yol gösterici görüşleri ile devam etmem için verdiği güç için ayrıca teşekkür etmeliyim. Prof. Dr. Aydın Erar da yapıcı eleştirileri, düzeltmeleri ve gülen yüzü ile yoluma devam etmemde ve tezimin tutarlılığında pay sahibidir. Kendisine ayrıca çok teşekkür ederim.

Alanı bana tanıştıran değerli Hocam Prof. Dr. Bülent Yılmaz'a, bilgiye erişim konusunu tanıyıp sevmemi sağlayan ve ilk tez izleme komitelerimde bulunan değerli Hocam Prof. Dr. Yaşar Tonta'ya da çok teşekkür ederim.

TÜBİTAK Projesi'nde birlikte çalıştığım değerli Doç. Dr. Zehra Taşkın, Dr. Öğretim Üyesi Güleda Doğan, Dr. İpek Şencan'a umudum kırıldığı anlarda yüzümü güldürdükleri, işler yolundayken birlikte mutluluğumu paylaştıkları için teşekkür ederim. Doç. Dr. Zehra Taşkın'a bu çalışmaya başlamamdaki büyük katkısı için ayrıca teşekkür ederim.

Tezimin son okumasını yapan ve detaylı geri dönüşü ile tezime katkı sağlayan değerli arkadaşım ve Hocam Dr. Gürhan Bulu'ya çok teşekkür ederim. Sevgili arkadaşım

Dr. Öğretim Üyesi Sevda Çetin'e süreç boyunca verdiği öneriler, destek ve cesaret için teşekkür ederim.

Bu uzun süreçte birlikte çalıştığım eski mesai arkadaşlarıma anlayışları ve destekleri için teşekkür ederim. Murat Tezgider'e ayrıca Netbeans IDE'de bildiklerini paylaşarak işimi kolaylaştırdığı için de teşekkür ederim.

Değerli Daire Başkanım Pınar Al'a da tezin ortaya çıkması ve savunma sürecinde gösterdiği anlayış için teşekkür ederim. Son süreçte iş hayatımda kendisiyle "alanda" çalışıyor olmamın tezime katkısı büyüktür.

En büyük teşekkür, "Tezin ne zaman bitecek?" sorusunu hunharca sormaktan kaçınmamalarıyla göz dolduran anne, babama; yaramazlıklarıyla evdeki çalışma ortamımı kafama geçiren iki oğluma ve ben hafta sonları çalışırken büyük bir zevkle evle ilgilenen eşime gitmeli! Şaka bir yana başarılar sevdiklerimle kutlayınca anlamlı. Sevgili eşim Orkun'un desteği sayesinde; oğullarımız Doğu ve Tan, anne (Cihan Özkan, Nursen Çelik) ve babalarımızla (Rıfki Özkan, Süleyman Çelik) bu zor süreçten çıkıp, tezimin girişinden aileme teşekkür yazmak, benim için mutlulukların en büyüklerinden biri oldu. Sevgili ailem, bu çalışma sizlerin de eseridir. Canım oğullarım Doğu ve Tan, çalışmayı ve araştırmayı sizin de çok sevmeniz dileğiyle, hepinize tekrar çok teşekkür ederim.



## 1. BÖLÜM: GİRİŞ

Bu bölümde insanoğlunun bilgi ile iletişimini etkileyen teknolojik gelişmelerin 21. yüzyılda ulaştığı seviye, değişen kullanıcı alışkanlıkları bağlamında ele alınmaktadır. Yaşanan değişimlerin bilimsel iletişime izdüşümlerine tarihsel açıdan kısaca değinilirken, bilimsel yayınların okuyucuya ulaşması sürecinde yazar ve okur alışkanlıkları üzerinde durulmaktadır. Bu bağlamda tez çalışmasının konusunu oluşturan bilimsel iletişim sürecinde daha fazla okurun ilgisini çekmek için yapısal öz yazımının faydalarına dikkat çekilmektedir. Çıkarıma dayalı otomatik özetleme sistemlerinin bilimsel iletişim sürecine sağlayacağı yararların özetlendiği bu bölümde, ayrıca araştırmanın önemi, hipotezi ve araştırma soruları ile yöntem, kapsam ve araştırmanın düzeni hakkında bilgi sunulmaktadır.

### 1.1. KONUNUN ÖNEMİ

Araştırmacılar için ilgi alanlarına giren makaleleri güncel olarak takip etmek bir zorunluluktur. Bu zorunluluğun artan yayın hacmiyle giderek daha fazla zaman alıcı bir iş haline gelmesi ise her zaman popülerliğini korumuş bir sorundur. Örneğin, 1826 yılında Faraday, zamanının bir bölümünü deneylerine ayırmak isteyen bir araştırmacı için, ilgi alanına giren tüm kitapları ve makaleleri okuyup işine yarayacakları diğerlerinden ayırmayı da hakkıyla yapabilecek vakti bulmanın kesinlikle mümkün olmadığını yakınmıştır (Crowther, 2013, s. 96). Faraday'ın bu şikayetinin sebebi, sonuçlarını henüz yayımladığı bir çalışmasında ele aldığı olgunun 1813 yılında bir başka bilim insanı tarafından zaten çalışılmış olduğunu fark etmiş olmasıdır. Şüphesiz günümüzde, gelişmiş bilgiye erişim teknolojileri ve İnternet kullanım imkânına sahip olabilen bilim insanları için ilgi alanlarında yayımlanmış çalışmalardan haberdar olmak sorun olmaktan çıkmıştır. Artık bir konuda araştırma yaparken, araştırma konusuyla ilgili görünen yığınlarca makaleye çok kısa sürelerde erişilebilmektedir. Günümüzde yaşanan sorun, kısa sürede erişim sağlanabilen yığınların içinden araştırma konusu ile gerçekten ilgili olan dokümanları seçebilmek ve seçilenleri gerçekten okumaya zaman ayırabilmektir.

İnsanoğlunun bilgi ile iletişimi meselesini başka bir boyuta taşıyan teknolojik gelişmeler uygarlık tarihinde de yeni bir sayfa açmıştır. 20. yüzyılın son çeyreğinde bilginin ve bilgi teknolojilerinin tarım, sanayi, hizmet sektörleri, eğitim, sağlık, iletişim gibi her alanda yaygın olarak kullanılır hale gelmesiyle bilgi toplumundan söz edilmeye başlanmıştır (Aktan ve Vural, 2005, s. 32). Bilgi toplumu, tarım ve sanayi toplumlarının ardından üçüncü dalga olarak kabul edilmektedir (Toffler, 1980). Üçüncü dalgayı tarihte yeni bir dönem başlatan daha önceki dönüşümlerden ayıran köklü gelişimlerin ve dönüşümlerin yoğunluğuna önceki dönemlerde rastlamak mümkün değildir (Drucker, 2012, s. 185). Bu dönemde basılı bilgi miktarının okuma zamanı karşısındaki orantısız artışının boyutları da geçmiş dönemlere göre abartılı bir hal almıştır.

21. yüzyılda Internetin gündelik hayata girmesi ile gündeme gelen sosyal ağların erişim, üyelik ve kullanımları gün geçtikçe kolaylaşmıştır. Çok az ön bilgi ile kullanıcısı olunan bu sosyal medya hesapları ve akıllı telefonlar, her yaştan Internet kullanıcısını (doğru ya da yanlış) bilgiyi oturduğu sandalyeden tüm dünyaya zahmetsizce yayabilir ve yine aynı kolaylıkla ona erişebilir hale getirmiştir. Artık önceki birkaç bin yıl süresince üretilmiş bilgiler bile her yerden bir tıkla erişilebilecek şekilde saklanabilmektedir. Internet üzerinden mesajlaşma araçları ve sosyal ağlar ise özellikle gençleri neredeyse tüm gün çevrimiçi tutmaktadır (Athreya ve Mouza, 2017, s. 2). Tabii ki bilgiye ve iletişim teknolojilerine erişim imkânlarının günden güne artmasının demokratik etkileri son derece olumludur (Athreya ve Mouza, 2017, s. 7). Ancak günün her saati, dünyanın her yerinden, her çeşit amaç için üretilmiş, pek çok formatta ve büyük miktarlardaki bilginin, toplumun her kesiminden bireyinin gözünü, kulağını ve zihnini meşgul eder hale gelmesinin yan etkileri de mevcuttur. Örneğin, araştırma ve okuma alışkanlığı basılıdan elektronik tarafa doğru hızla kaymaktadır (Mangen, Walgermo ve Brønneck, 2013, s. 61). Bireylerin elektronik veri ormanındaki avcılara ve toplayıcılara dönüştüğü bu süreçte büyüyen nesiller, teknolojinin baskın olduğu bu dünyaya doğup, önceki kuşakların büyük çoğunluğunun yabancı olduğu, onlar için yaratılmış yepyeni öğrenme, sosyalleşme ve iletişim araçlarına sahip olmuşlardır (Carr, 2010). Kuşkusuz bundan sonraki nesillerin bilgi ile iletişimde de henüz öngöremediğimiz teknolojik yenilikler var olacaktır.

Kişisel bilgisayarlarımız aracılığı ile elektronik ortamda ürettiğimiz ve kişisel Internet kotalarımız ile elektronik ortamda eriştiğimiz bilgiyi yine elektronik aygıtlar aracılığı ile okuyup işleyerek tüketmekteyiz. Üstelik bu döngüyü sadece akademik araştırmalarımız için değil, sosyal medya hesaplarımızı gün içinde rutin olarak kontrol ederken de yaşamaktayız. Bu haliyle, 21. yüzyıl insanı eriştiği büyük miktarlardaki bilgiyi işleyebilmek için temel okuryazarlık becerilerine ek olarak görsel, işitsel, elektronik okuryazarlık, medya ve bilgisayar okuryazarlığı vb. çoklu becerilere gün boyu ihtiyaç duymaktadır (Kurbanoglu, 2013, s. 85). Analiz, sentez ve değerlendirme gibi üst düzey bilişsel işlemler (Bloom, 1956, s. 204) gündelik hayat deneyiminin bir parçası haline gelmiştir. Bu sürekli etkileşim sebebiyle insanoğlunun okuma veya yazmaya yoğunlaşma, araştırma, odaklanarak okuma, anlama, derin düşünme gibi akademik becerilerinin zarar gördüğünü öne süren çalışmalar da mevcuttur (Carr, 2008; Issa ve Isaias, 2016; Merzenich, 2008). Yoğun etkileşim altında kimsenin ilgilendiği konuya yönelik olarak erişebildiği tüm yayınları okumaya ayıracak vakti ya da dikkati yoktur. Kısa sürede aradığı bilgiyi bulabilmek her kullanıcı için çok önemlidir. Bu yüzden, istediği anda istediğinden fazla bilgiye erişebilen 21. yüzyıl insanı için ilgilendiği konuyla ilgili dokümanlara hızlıca eriştikten sonra göz gezdirerek kısa zamanda tarayabilmek önemli bir meseledir. Ancak bu yüksek hızın bedelini “derin okuma” yeteneğinden kaybederek ödeyen kullanıcıları (Carr, 2008; Merzenich, 2008) eriştiği bir metnin içeriğinin konusuyla ilgili olduğuna karar verdikten sonra dahi tam metni okumaya ikna etmek gün geçtikçe zorlaşmaktadır. Kullanıcılar Internette eriştikleri ve hızlıca taradıktan sonra araştırma konuları ile ilgili olduğuna karar verdikleri dokümanları bile çoğu kez derin okuma yapmadan geçebilmektedirler. Faraday’a günümüz perspektifinden bakarak bilimsel iletişime geri dönecek olursak, bilimsel makalelerin okuyucuları da taradıkları dokümanlar içinde başlığını okudukları makalelerin sadece onda birini, özünü ayrıntılı olarak okudukları makalelerin ise yarısını gerçekten okuyabilmektedirler (Mabe ve Amin, 2002, s. 154).

Tüm bilimsel makalelerin üretilme amacı yapılan bilimsel çalışmayı diğer araştırmacılara duyurmaktır. Bunun yolu da yayımlanan makaleyi okunur kılmaktan geçer. Ancak bir bilim insanı çalışmalarını sadece duyurmakla da yetinmek istemez. Yayımlanan bir araştırma için, tartışılabilir olsun ya da olmasın, dergi etki faktörleri ve makale atıfları

günümüz akademik dünyasındaki önemli kavramlardır (Wu, 2011, s. 1345). Bu kavramlar, araştırma projelerinin, dergilerin, bilim insanlarının ve kurumların performansını değerlendirmek için ölçüt olarak gün geçtikçe daha fazla kullanılmaktadır (Taşkın ve Al, 2018, s. 335). Bilimsel iletişim süreci kapsamında yazar, editör ve hakem arasında gelişen sürecin sonunda yayımlanan bir bilimsel araştırmanın meslektaşlar tarafından incelenip kabul edilerek bir gerçeklik kazanması beklenir. Kaldı ki, araştırmanın sonuçları uygun kitlelere sunulmadıkça, fonlar da boşa harcanmış olacaktır. Bu açıdan bakınca, verimli ve etkili iletişim, araştırma sürecinin önemli bir parçasıdır (Meadows, 1997, s. 10). Ancak bilimsel iletişim sürecinin her iki yönünde de yer alan araştırmacılar yazar kimlikleriyle daha fazla kişiye ulaşip, daha fazla okunmayı hedeflerlerken, okur kimlikleriyle de her gün daha fazla makale okumayı istemekte midirler? 9000 araştırmacı ile yapılan görüşmeler sonunda elde edilen sonuçlara göre, araştırmacıların yazar olarak beklentileri daha fazla yayın yapmak ve daha fazla okunmak iken, okur olarak beklentileri daha az okumak ve daha az çaba sarf ederek tarama yapabilecekleri ortamlara sahip olmaktır (Mabe ve Amin, 2002, s. 150). Gerçek hayatta ise araştırmacılar her zaman beklentilerinin ötesinde okumak durumunda kalmaktadırlar (Mabe ve Amin, 2002, s. 156).

Bilim insanlarının okumaya yetişememeyi sorun olarak görüp, akademik makaleler için ilk otomatik metin özetleme çalışmalarına başlamaları bilgi toplumunun öncesine dayanır. İnternetin ve bilgi teknolojilerinin getirdiği kolay erişim imkânlarının henüz gündemde olmadığı yıllarda Luhn (1958) okuyucuya bir makale veya rapordaki faydalı bilgileri bulmada zaman ve çabadan tasarruf ettirme amacıyla otomatik metin özetleme alanındaki öncü çalışmasına başlamıştır. Artan yayın hacmine yetişememe kaygısı tarihi varlığını korumaya hâlâ devam ederken, araştırmacıların karşı karşıya kaldıkları yayın miktarı günümüze değin keskin bir şekilde artarak gelmiştir. 1980 yılından 2012 yılına kadar yapılmış tüm yayınların *WoS* verilerine göre, küresel bilimsel yayın sayılarındaki artışın üstel olarak gerçekleştiği gösterilmiştir (Bornmann ve Mutz, 2015, s. 2217).

Akademik yayın hacmi, 21. yüzyılın başlarına kadar dergilerin her bir sayıda yayımladıkları makale sayısı ve bu makalelerin uzunluklarına göre de artarken, pek çok yeni başlık da araştırmacıların ilgisine sunulmuştur. (Tenopir ve King, 2014, s. 171). Bu

süreç devam ederken teknolojik gelişmelerle birlikte e-dokümanların kullanım alanları büyük bir hızla artmış, akademik makaleler çoğunlukla elektronik ortamda erişilen ve okunan dokümanlar haline gelmişlerdir (Tenopir ve King, 2014, s. 174). Halen bilimsel yayınların basılı ve elektronik formları birçok başlık için bir arada bulunmaya devam edebiliyor olsa da, elektronik formatlarda daha fazla başlık yer almaktadır ve erişilebilirlik bu formatlarda daha ön plandadır. Bu bağlamda erişim ve yayınlama süreçlerini hızlandıran gelişmiş işlevsellikleriyle elektronik formatlar, bilimsel yayıncılıkta hızla kabul görerek, elektronik dönüşüm süreçlerini tamamlanmıştır denilebilir (Tenopir ve King, 2014. s. 159).

Bilimsel makaleler, e-kitaplardan sonra, okuyucuların elektronik ortamda karşı karşıya kaldıkları en uzun metinlerdir. Kullanıcılar kitap okumada çoğunlukla basılı sürümleri kullanmaya devam ediyor olsalar da, bilimsel yayınların elektronik sürümleri basılı olanlara göre daha fazla tercih edilir olmayı kısa sürede başarmıştır (Dewan, 2012, s. 29). Bu hızlı gelişmeler elektronik formatlara özgü kullanılabilirlik ve erişilebilirlik seçenekleri üzerine kütüphanelerde, yayıncılıkta ve bilimsel topluluklarda canlı tartışmalara yol açmıştır (Tenopir ve King, 2014. s. 159). Bilimsel yayınların elektronik bir aygıt üzerinden, “her zaman”, “her yerden” erişim konforları diğer yayınlara göre büyük bir hızla artarken, elektronik ortamın okuyucunun metinle ilişkisine etkisi çoğunlukla göz ardı edilmiştir. Yüksek oranda ve uzun düz metinler içeren yapıları ile bilimsel yayınların elektronik formatları basılı formatlarının neredeyse bire bir aynısıdır. Bu halleriyle bilimsel yayınların “elektronik okuma” konforları erişim konforları ile birlikte artmamıştır. Aksine, elektronik formatlar, akademik yayınların günden güne artan zor okunurluklarına (Meadows, 1985, s. 29) bir de dijital okumanın doğası gereği eklenen yeni zorluklar getirmişlerdir. Bu zorlukların başlıcalarını aşağıdaki gibi sıralayabiliriz:

- Çevrimiçi olmanın yarattığı dikkat dağınıklığı (her an mesaj alışverişi ve sosyal medya hesaplarındaki hareketler).
- Arama sonuçlarında erişilen yığınlarca ilgili makale içinden eleme yapma zorunluluğu.
- Bağlantılı web sayfaları içinde tıklayarak ilerlerken yolunu kaybetmeden araştırmayı sürdürmenin zorluğu.

- Elektronik ortamda uzun bir metni sayfada kaydırma yaparak, ileri geri gezinerek okuyup bitirmek ve anlamanın basılı sürüme göre daha fazla zihinsel çaba gerektirmesi.

Yukarıdaki ilk üç madde web üzerinden herhangi bir okuma yaparken kullanıcının baş etmesi gereken, derin okuma ve anlama yeteneklerini körelten başlıca unsurlar olarak sıralanmaktadır (Carr, 2008; Issa ve Isaias, 2016, s. 594; Merzenich, 2008). Elektronik ortamda uzun bir dokümanı kaydırarak okumanın aynı dokümanın basılı sürümünü sayfa sayfa okumaya göre daha fazla zihinsel çaba gerektirdiği ve sonuç olarak elde edilen anlama düzeyinin elektronik ortamda yapılan okumalarda daha düşük olduğu ise laboratuvar ortamında yapılan deneylerle doğrulanmıştır (Singer ve Alexander, 2017, s. 1034; Wästlund, 2007, s. 35). Akademik makaleler de bu kaydırarak okunan dokümanların başlıcalarındandır. Gazete makaleleri, ya da dijital ortamda doğmuş pek çok diğer yayın türü için dijital ortamda anlamayı kolaylaştıracak çeşitli düzenlemeler sürekli geliştirilmekte iken, akademik makalelerin dijital sunumları basılı sürümlerinden pek de farklı bir yapıda üretilmemektedir. Uzun düz metin içerikleri ve günden güne artan sayfa sayılarına rağmen akademik makalelerin elektronik ortamda okunmasını kolaylaştıracak, yaygın kabul görmüş farklı sunum türleri üzerinde uzlaşmaya varıldığı söylenemez.

Elektronik ya da basılı, her iki formatta makale bileşenleri bellidir ve hemen hemen aynıdır. Bu bileşenler, başlık, öz, anahtar kelimeler, makale gövde metni, teşekkür, kaynakça ve ekler olarak sıralanabilir. Bu bileşenlerin her birinin işlevleri belirlidir. Başlık, öz ve anahtar kelimeler, bir akademik makalenin potansiyel okurlarının, ilgili dokümanın tamamına göz gezdirmesi gerekmeden tam metni okuyup okumayacağına karar vermesini sağlarlar. Bunlardan başlık ve anahtar kelimelerin görevi makalenin arama motorlarında bulunması ve okuyucu ilgisinin makaleye çekilmesi iken; öz içeriğe dair detaylı bilgi de vermektedir. Öz, okurun tam metni incelemeye geçmeden önce içerik hakkında bilgi edinmesini sağlayan son durağıdır. Bu yönüyle çoğu okuyucu için bir makale, özü kadar ilgi çekicidir.

Bir bilimsel makalenin özü genellikle 150 ile 300 kelime arasında bir uzunluktadır. İyi bir özden beklenen okuyucunun bilgi ihtiyacı ile dokümanın ilgisini sağlıklı bir şekilde karşılaştırma imkânı sağlamasıdır (Atanassova, Bertin ve Larivière, 2016, s. 636). Özler makalenin içeriği hakkında başlığa ve anahtar kelimelere kıyasla daha üst düzey ayrıntı verirler (Hartley, 2008, s. 23). Anahtar kelimeler ve başlık yayının bulunmasına ve okurun ilgisini çekmesine yardımcı olduktan sonra, okuyucuların ilgisini artırmak ve meraklarını korumak işi öze kalmıştır (Hartley, 2008 s. 23; Rodrigues, 2013). Öz bu üç unsur arasında en etkili mekanizmaları tetikleyebilen işlevlere sahip olması ile göze çarpmaktadır. İyi hazırlanmış bir öz bir bilimsel makalenin en değerli paragrafıdır (APA, 2010, s. 26). Tam metin içeriğini iyi yansıtan bir öz daha iyi bir dizinlemeye sahip olarak arama motorlarında, ilgi alanına girdiği okurların önüne üst sıralardan düşecektir. Okunurluğu yüksek ve elektronik ortamın okuma alışkanlıklarına uygun bir düzende sunulmuş, ilgi çekici bir içeriğe sahip özlerin, makalelerin tam metninin okunma şansını artırmada payları büyüktür. Özü iyi planlanmamış bir makalenin erişilebilirliği ve tam metninin okunma ihtimali iyi bir öze sahip olan makalelerden düşük olacaktır.

Özlerin konuyla ilgili ön bilgisi olmayan okurları ya da akademi dışından uzmanları da konu hakkında bilgilendirebilecek güçte ve sadelikte tasarlanması önemlidir. Kullanılan dil, mesleğe yeni girmekte olan öğrenciler, kendi konuları dışında olup o makaleyi okuyabilecek araştırmacılar ve ana dili makalenin dili olmayan okur kitlesinin de anlayabileceği özellikte olmalıdır. Öz aynı zamanda içerik olarak tam metni kapsamalıdır.

Bilimsel yayınların, okurları ile buluşmalarında ilk ve çoğunlukla son şanslarını özleri ile kullanmalarına rağmen (APA, 2010, s. 26), araştırmacılar bir makale yazmaya karar verdiklerinde, zamanlarının büyük bölümünü makalenin tam metninin yapısal bütünlüğü üzerinde harcarlar (Rodrigues, 2013). Uygulamada bir bilimsel yayından beklenen, araştırmanın motivasyonunu, yöntemini, bulgularını ve sonuçlarını evrensel olarak kabul görmüş bir yapısal düzeni takip ederek aktarmasıdır. Makalelerin yapısal düzenleri, gönderileceği derginin yayın politikasına uygun olacak şekilde tasarlanır. Derginin yayın politikasında olsun ya da olmasın makalelerin tam metnlerinde araştırmanın kapsamı, amaçları, yöntemi, bulguları ve sonuçları hakkındaki bilgilerin ayrı bölümler halinde organize edilmiş şekilde verilmesi alışkanlığı yaygındır. Benzer şekilde, yazarın öz

yazımında da takip edebileceği çeşitli formatlar mevcuttur. Dergilerin kullandıkları öz formatları çeşitlilik gösterse de sırasıyla klasik ve yapısal öz yazımı en baskın talep edilebilen türlerdir. Yazar yapısal öz yazımında tam metinde kullandığı yapısal bütünlüğü özde de sergilemek zorunluluğunda iken, en yaygın kullanımda olan klasik özlerde ise özün içeriği yazar inisiyatifindedir. Yazar da her zaman tam metnin içeriğini yansıtmada titiz davranmayabilmektedir. Özler üzerinde yapılmış araştırmalara göre, klasik özlerde çoğunlukla giriş ve sonuç hakkında bilgilere yer verilip, araştırma problemi, yöntem ya da sonuçlara dair bilgilerin atlandığı gözlemlenmektedir (Bitri ve Keseroğlu, 2015, s. 251; Hartley, 2008, s. 32). Oysa bir makalenin tam metni bittikten sonra her bir bölümünden önemli cümleleri olduğu gibi alarak bu cümleler üzerinde düzeltmeye gitmek (Labaree, 2018), ya da hazırlanmış özü tam metnin başlıklarının ana hatları ile karşılaştırarak doğrulamak (APA, 2010, s. 26) iyi bir öz yazımı için tavsiye edilen yöntemlerdir.

Tıp başta olmak üzere, psikoloji, eğitim gibi başka bazı deneysel çalışma alanlarında yayımlanan dergilerde yapısal öz kullanımı 1980'lerden itibaren kademeli olarak zorunlu olmaya başlamıştır. Yine de bilim camiasında yaygın eğilim halen klasik öz kullanımınıdır. Yapısal öz talep eden yayınlar, yapısal başlıkları önceden belirleyip, yazardan ilgili başlıkların altını doldurarak öz hazırlanmasını ister. Klasik öz ise tek bir blok paragrafta, hiç alt başlık kullanmadan hazırlanmalıdır. Bu da yazar için hazırlaması, editör için tam metnin içeriğiyle karşılaştırarak kontrol etmesi, okur için ise okuması ve anlaması zor olan bir düzenleme biçimidir. Yapısal öz isteyen dergilerin özle ilgili talep ettikleri başlıklar yazar ve editör için bir kontrol listesi yerine de geçmektedir. Yapılandırılmış özler, yazarların kilit ayrıntıları ihmal etmelerini önler (Hartley ve Betts, 2009, s. 2016). Okuyucuların da yapısal başlıklara bölünmüş özler hakkındaki fikirleri aynı özün klasik sunumuna göre daha pozitifdir (Hartley ve Betts, 2007, s. 2338). Dergiler ise bu şekilde öz yazımının insani hataların azaltılmasında da işe yarayacağını umarak kullanıma aldıklarını bildirmektedirler. Sonuç olarak klasik öz ile karşılaştırıldığında yapısal öz; hem yazar, hem editör, hem de okuyucu cephesinde uygulama kolaylıkları olan bir formattır.

Özün bir makaledeki görevi ilgili makaleye erişen kişilerin özü kolaylıkla okuyup anlaması ve içerik hakkında bilgilenmesidir. Bu sebeple kolay okunurluk, özlerin



kalitesini belirlemede bir diğer önemli ölçüt olarak göze çarpmaktadır (APA, 2010, s. 26). Çünkü bir okuyucunun bir yazar tarafından iletilen mesajı anlayabilmesinin kolaylığı okunabilirlik olarak tanımlanmaktadır (DuBay, 2004, s. 3). Okunurluk bağlamında, okuyucu ve metin arasındaki ilişkiyi etkileyen pek çok unsur olabilir. Metnin türü, akıcılığı, anlaşılabilirliği, sözcük sayısı, imla ve noktalama kurallarına uygun yazılıp yazılmadığı, cümlelerin uzunluğu gibi metin özellikleri; metnin okunduğu ortama ait çevresel faktörler ve okuyucunun hazır bulunuşluğu, yaşı, okuma motivasyonu, fiziksel engelinin olup olmaması gibi okuyucu özellikleri okuma sürecini etkilemektedir (Çoban, 2014, s. 97). Yani okuyucu-metin ilişkisi çoğu kez metnin, ortamın ve okuyucunun özelliklerine bağlı bir şekilde göreceli olarak farklılaşabilmektedir. Bir metnin okunurluğunu düzenlerken tüm bu unsurları hesaba katmak mümkün değildir. Yaygın kullanımda olan okunurluk hesaplamaları, çoğunlukla metinde geçen kelimelerin hece uzunluğu ve cümlelerin kelime uzunluğunu temel alırlar. Bazen bu hesaplama o dilde çok kullanılan kelimeleri de katmak mümkündür ki, bunun için dilbilimcilerin titizlikle hazırladıkları çok kullanılan kelimeler sözlüklerine ihtiyaç duyulur. Bir metnin okunurluğu hesaplanırken kullanılan ölçekte dokümanın aldığı değer ile hedef okuyucu kitlesinin özelliği eşleşiyorsa o dokümanın uygun okunurluk değerinde üretildiğine karar verilir.

Akademik makalelerin özleri, okunurluk ölçekleri ile değerlendirildiklerinde çoğunlukla “çok zor” okunurluklu bulunmaktadır (Gazni, 2011, s. 279; Hartley, Pennebaker ve Fox, 2003, s. 395). Araştırmacıların makale özlerini “çok zor” okunurlukta üretiyor olmaları akademik çevrelerde sorun olarak görülmemektedir. Bu durumun sebeplerinden biri okuyucu hedef kitlesinin o alanda yoğunlaşarak çalışmış, üst düzey eğitim görmüş kişiler olmasıdır. Bir diğer sebep ise, özlerde belirgin bir şekilde zor okunurluklu dil kullanımının akademik camiada normal kabul edilmesidir. Bu zor dili okumak akademik çevrelerde “tolere” edilen, “prestijli” bir iştir (Gazni, 2011, s. 279). Az ve kolay okumayı talep eden bilim insanlarının taramaları gereken yayın hacmi göz önüne alındığında, zor okunurluklu yazma alışkanlığı bilimsel iletişim sürecini zorlaştıran bir unsur olarak göze çarpmaktadır.

Özleri yukarıda değindiğimiz unsurları çerçevesinde ele alıp; okuyucu ilgisini çekme, anlaşılır olma ve tam metnin okunma oranını artırma işlevlerine göre değerlendiren pek çok araştırma yapılmıştır. Bu çalışmalara göre klasik özlerin tespit edilmiş temel sorunları aşağıdaki gibi listelenebilir:

- **Sunum Sorunları:** En çok tercih edilen klasik özler ister basılı ister dijital ortamda olsun okuyucunun dikkatini çeken bir sunum formatına sahip değildir. Sunum konusu okunurluk bağlamından ayrı ele alınması gereken bir unsur olsa da (DuBay, 2004, s. 3), paragrafsız ve alt başlıksız tek bir blok halinde, tam metinden daha küçük fontta, bazen italik yazı karakterinde yazılmaları okuma eylemini de zorlaştırmaktadır (Ufnalska ve Hartley, 2009, s. 71, Meadows, 1985, s. 117).
- **Organizasyon Sorunları:** Özler, eğer derginin bu konuda bir politikası yoksa tam metinde takip edilen yapısal düzeni izlemez. Okuyucunun akademik bir makalenin tam metnini okumaya karar verme sürecinde en çok ilgilendiği bölümler araştırma problemi, yöntem ya da sonuçlara dair açıklayıcı bilgilerdir. Öz organizasyonunda bu bölümlere ilişkin bilgilerin atlanması sıklıkla gözlemlenen bir problemdir (Atanassova vd., 2016, Bitri ve Keseroğlu, 2015, s. 251; Hartley, 2008, s. 32; Jamar, Šauperl ve Bawden, 2014; Šauperl, Klasinc, ve Lužar, 2008).
- **Okunurlukla İlgili Sorunlar:** Bir bilimsel yayının özünde çoğunlukla bağlaçlı, uzun, devrik cümleler ve yoğun teknik terim ya da alana has jargon kullanımına rastlanır. Sade ve basit bir dille yazılması gereken bilimsel makalelerde daha üstün dil özellikleri kullanımı bilinçli olarak tercih edilmektedir. Özlerin okunurlukları makalenin diğer bölümlerine göre daha zordur (Day, 1996, s. 22; Gazni, 2011, s. 279; Hartley vd., 2003, s. 395).

Yapısal öz yazımının aşağıda verilen, belirlenmiş faydaları ise bu sorunlara çözüm olacak niteliktedir:

- Yapısal öz yazmak klasik öz yazmaktan daha kolaydır. Yazar yayının tüm bölümlerini öze aktarmış olduğundan emin olur. Öz-tam metin tutarlılığı artar.
- Yayının dizinlenme performansı, dolayısıyla da yazar için ilgili yayının arama sonuçlarında görünme olasılığı artar.

- Okuyucuya aramalarında erişim kolaylığı ve arama sonuçlarında uygunluk sağlar. Her seviyeden okuyucuya aradığı belirli bir bilgiye uygun yayını bulmasında kolaylık ve netlik sağlar.
- Özün bilgilendiriciliği artar.
- Alt başlıklarla bölünmüş cümleler okuyucu için okunma kolaylığı sağlar ve böylece okuyucunun anlama düzeyi artar.
- Okuyucular tarafından klasik sürümlere göre daha çok tercih edilir.
- Yazarlar tarafından tercih edilir (Hartley 2008, s. 368; Hartley ve Betts, 2007, s. 2338; Mulrow, Thacker ve Pugh, 1988, s. 615; Sharma ve Harrison, 2006, s. 529).

Günümüzde akademik makalelere erişim ve okumanın neredeyse tamamen elektronik ortamda yapıldığı gözlenmektedir. Bilimsel iletişim bağlamında sorunlara yol açabilen formatların ve dilin özellikleri ise başlangıcından bugüne neredeyse hiç değişmemiştir. 21. yüzyıl teknolojilerinin araştırma yapma ve okuma alışkanlıkları üzerindeki etkilerinin bilimsel iletişim bağlamında da değerlendirilmesinin önünde somut bir engel yoktur. Üstelik bilimsel iletişim bağlamında bilgi teknolojilerinden daha fazla faydalanmanın yollarını araştırmak da mümkündür. Bilimsel yayınlar belli başlı kurallara göre yazılan, belirli kurallara göre arşivlenen, oldukça yapılandırılmış (Atanassova, Bertin ve Mayr, 2015, s. 1092) sayılabilecek metinlerdir. Bu yönleriyle üzerlerinde metin madenciliği teknikleri uygulanmasına son derece müsaitlerdir. Otomatik metin özetleme çalışmalarının çoğunda kullanılan veri setlerinde bilimsel yayınlar bu yönleriyle de tercih edilir.

Genel anlamda uzun bir metni özetleme, belirli kullanıcı ya da kullanıcı grubuna yönelik; belirli bir görev ya da görevleri gerçekleştirme amacıyla üretilen; ilgili tam metnin önemli bilgilerini içeren tam metnin kısaltılmış bir sürümünü elde etme işidir. Özler bu kısaltılmış sürümün özel bir halidir. Bu doğrultuda, makale tam metnini belirli bazı kurallar çerçevesinde kısaltırken içeriğin önemli noktalarını koruma işi, otomatik yöntemlerle gerçekleştirilebilmektedir. Çıkarıma dayalı ve yoruma dayalı olarak iki ana yaklaşımla geliştirilen otomatik özetleme yöntemleri sürekli üzerinde çalışılan bir alan olduğu için yeniliklere ve geliştirmeye açıktır.

Herhangi bir özet hazırlama işi oldukça sübjektif karar gerektiren bir süreç olduğu için çıktığı metinlerinin kalitesi üzerinde fikir birliğine varılması beklenmez. Bilimsel makale özlerinin üretiminde, bazı özel durumlarda profesyonel özetleyicilerin emeğine başvurulduğu olmuştur (öz dergileri gibi). Ancak, mevcut bilinen sorunlarına rağmen yazarları dışındaki kişilerce öz hazırlanması fikri kabul gören bir eğilim haline gelmemiştir. Çünkü yayının konusuna en aşina kişi yazarın kendisidir. Ayrıca, öz hazırlama sürecinde tam metnin önemli noktalarını çok daha kısa bir şekilde yeniden ifade etme işinde verilecek sübjektif kararları yazar bizzat kendisi almayı tercih eder. 2.6. Otomatik özetleme sistemlerinin değerlendirilmesi bölümünde detaylarıyla ele alınacağı gibi, profesyonel kişilerce üretilmiş olsa da bir metnin birden fazla özetleyici tarafından üretilmiş özet metinleri birbirine benzer özellikte olmak zorunda da değildir. Üretilen her türden özet metninin niteliği, tam metnin içeriğinin yanı sıra özeti hazırlayan kişinin geçmiş bilgilerinden, fikirlerinden ve eğilimlerinden her zaman etkilenir ve kalitesi de aynı şekilde göreceli bir konudur (Luhn, 1958, s. 159). Tüm bu eksilerinin yanı sıra bu şekilde öz üretimi zaman alıcı bir süreçtir, dolayısıyla ekonomik de değildir. Otomatik yöntemlerle üretilen özler kullanıcının metnin tamamını okumadan doküman hakkında temel bilgi edinmesi prensibine göre çalışır ve bahsedilen sübjektif kararların bilgisayarca, önceden belirlenmiş kurallara göre alınmasını sağlar. Bunu yaparken öz hazırlama sürecinde insan faktöründen kaynaklı göreceliliği ve maliyeti azaltarak sürece hız kazandırır (Torres-Moreno, 2014, s. 4).

Çıkarıma dayalı otomatik özetleme yöntemleri daha az karmaşık dilbilimsel yöntem kullanıyor olmaları sebebiyle uygulama kolaylığına sahiptir. Çıkarıma dayalı yöntemler, tam metinde geçen önemli cümleleri seçerek, olduğu gibi kullanır. Bu yüzden de yazarın fikirlerinden uzaklaşma ihtimalleri yoktur. Makale tam metinlerinde yapısal bölümlendirme yapılma alışkanlığı akademisyenler ve araştırmacılar arasında yaygın olarak gözlenmektedir. Bu doğrultuda, tam metin cümleleri yapısal bölümlerine göre etiketlenerek otomatik cümle seçme işlemleri uygulandığı takdirde makalelerin her bir yapısal bölümünün en önemli cümlelerinden oluşan bir çıktı elde etmek mümkündür. Bu çıktılar, her bir cümlenin tam metinden kazandığı yapısal özelliklere göre düzenlendiğinde yapısal öz formatına sahip olurlar.

Tez araştırmasında kullanılan veri setine kaynaklık eden kütüphanecilik ve bilginbilim alanı çok disiplinli çalışmalara açık bir alandır. Özellikle bibliyometrik araştırmalar bağlamında diğler alanlara örnek olacak yenilikçi yaklaşımların test edilmesi bilimsel iletişimin gelişimi için mecburidir. Yapısal düzenin başlı başına özlerin okunurluklarını artıran bir unsur olduğu değlerlendirilmektedir (Mulrow vd., 1988, s. 615; Sharma ve Harrison, 2006, s. 529; Ufnalska ve Hartley, 2009, s. 71). Otomatik öz hazırlama araçlarından elde edilen özler ile, özellikle zaman kısıtı olan, odaklanma sorunu yaşayan, konuya aşına olmayan, “zor” metinleri okuyup anlamayı “tolere” edemeyenler de dâhil olmak üzere her seviyeden ve her anadilden daha fazla okuyucu için kolaylık sağlanacağı düşünölmektedir. Bu tür öz hazırlama araçları klasik öz hazırlama süreci için de yazar ve editörlere üretim ve değlerlendirme süreçlerinde yardımcı olarak kullanılabilir.

## **1.2. ARAŞTIRMANIN AMACI VE HİPOTEZİ**

20. yüzyılın son çeyreğinden itibaren bilgi teknolojilerinin yaşamın her alanında yaygın olarak kullanılır hale gelmesi insanın bilgi ile etkileşiminde köklü değışikliklere yol açmıştır. Günümüzde de süregelen bu teknolojik gelişmelere paralel olarak insanın bilgi ile ilişkisinde farklılaşmalara şahit olmaya devam etmekteyiz. Bu ilişkiyi bilimsel iletişim bağlamında ele aldığımızda, bilimsel dergilerin bu süreçte en hızlı şekilde elektronik sürümlerini hayata geçiren yayın gruplarından biri olduğu göze çarpmaktadır. Diğler yandan, basılı bilgi miktarının bireylerin okuma zamanlarına oranla artışı da geçmiş dönemlere göre abartılı bir hal almıştır. Bu bağlamda, bilimsel yayın hacmindeki artış karşısında kullanıcıların tam metinleri okumaya yetişememeleri sorunu da eskiye göre farklı bir boyuttur.

Bilimsel makaleler, hem karmaşık bilimsel bilgiler içermelerinin gereğı hem de akademik camianın yazım alışkanlıklarının bir sonucu olarak zor okunurluk düzeyinde dokümanlardır (Meadows, 1985, s. 29). Elektronik sürümler, bilimsel makalelerin erişimine büyük kolaylıklar getirirken, okunurluklarına çevrimiçi ortamın ve dijital ekranların doğası gereğı yeni zorluklar eklemiştir. Oysa bilimsel iletişimde bilgi teknolojilerinden daha fazla faydalanmanın yollarını araştırmak çağın gereğidir. Üstelik güncel bilgi teknolojilerinin araştırma yapma ve okuma alışkanlıkları üzerindeki

etkilerini bilimsel iletiřimi kolaylařtırma bağlamında da deęerlendirilmesi elzemdir. Akademik makalelerin yapısal özellikleri ve sunumlarından kaynaklanan sorunlar, üzerinde çalışılarak iyileřtirme yapılabilecek özelliktedir.

Bilimsel makaleler başlık, öz, anahtar kelimeler, makale tam metni, teřekkür, kaynakça ve eklerin belirli kurallara göre üretilip birleřtirilmesinden oluşur. Bu bölümler arasında özler, başlık ve anahtar kelimeler ile birlikte okuyucunun makaleye ulaşması ve tam metni okumaya karar vermesini içeren süreçte hayati önemdedir. Öz, tam metnin ne hakkında olduğunu belirli kurallar çerçevesinde, kısaca özetleyen bir metin parçasıdır. İyi yazılmış bir özü hızlı bir gözden geçirme ile inceleyen okura tam metnin içeriğini anlatması beklenir.

Makale tam metinleri çoęunlukla giriş, yöntem, bulgular, sonuç ve tartışma benzeri yapısal başlıklar takip edilerek yazılır. Öz, tam metinde kullanılan tüm bu bilindik yapısal bölümlerin önemli noktalarına değinmelidir. Ancak uygulamada çoęunlukla tam metnin içeriğini yansıtmada eksik kalabilmektedirler (Bitri ve Keseroęlu, 2015, s. 251; Hartley, 2008, s. 32). Makalenin yayımlandığı derginin politikasında herhangi bir yönlendirme bulunmuyorsa yazarlar öz yazarken metin organizasyonunda serbest davranma eğilimindedirler. Bu serbest bakış açısı ile yapılandırılmış özler, sunum, organizasyon ve okunurlukları açısından okuyucu ile metin arasındaki ilişkiyi negatif etkileyecek özelliklere sahiptir (Bitri ve Keseroęlu, 2015, s. 251; Day, 1996, s. 22; Gazni, 2011, s. 279; Hartley vd., 2003, s. 395; Hartley, 2008, s. 32; Ufnalska ve Hartley, 2009, s. 71, Meadows, 1985, s. 117). Bu doğrultuda, yapılandırılmamış makale özlerinin organizasyon ve dil açısından sorunlu bulunduğu dile getirilebilir. İyi yapılandırılmış bir öz hem ilgili makalenin arama motorlarınca doğru olarak dizinlenerek erişim sağlanabilmesini hem de konu ile ilgili arařtırmacıların tam metne vakit ayırmaya karar vermelerini sağlayacaktır. Bir bilimsel yayının üretilme amacı ilgili çalışmayı geniş kitlelere ulařtırmaktır. Özlerin iyi yapılanmıř olmasının bilimsel iletiřim bağlamındaki önemi açıktır.

Tez çalışmasında yukarıda kısaca özetlenen sorunlara çözüm aramak amacıyla, Türkçe kütüphanecilik ve bilginbilim alanında klasik özler ile yayımlanmış hakemli makalelerden oluşan derlem üzerinde otomatik yapısal öz elde eden bir sistem tasarlanmıştır. Genel olarak otomatik özetleme, çıkarıma ve yoruma dayalı olmak üzere iki farklı yaklaşımla yapılır. Yoruma dayalı yaklaşımların amacı metnin ana fikrini yansıtacak şekilde yeni cümleler üretmek üzere özet metni elde etmek iken, çıkarıma dayalı yaklaşımlar tam metnin önemli cümlelerini seçip bir araya getirerek özet metni ortaya çıkarır. Bu haliyle, yoruma dayalı yaklaşımların çıktıları insan tarafından üretilen özetlere daha benzerdir, ancak üst düzey dilbilimsel yöntemler kullandığı için uygulanabilirliği halen oldukça düşüktür. Çıkarıma dayalı otomatik özetleme yöntemleri tam metnin önemli noktalarını temsil eden cümleleri seçip bir araya getirerek metni kısaltma prensibine göre çalışırlar. Çıkarıma dayalı yöntemlerin, yazarın perspektifinden uzaklaşmadan, nesnel (Torres-Moreno, 2014, s. 4), dilden bağımsız, uygulama kolaylığına sahip (Gambhir ve Gupta, 2017, s. 3) öz hazırlama araçları üretmeye yatkın oldukları değerlendirilmektedir. Ancak çıkarıma dayalı yöntemlerde de seçilen önemli cümlelerin tam metindeki bağlamlarından kopartılarak sırayla bir araya getirilmesinden dolayı yaşanabilen bağlam ve tutarlılık sorunları yaşanmaktadır. Çalışma kapsamında geliştirilen çıkarıma dayalı otomatik özetleme sistemi tam metnin her bir yapısal bölümünden seçilen en önemli cümleleri üyesi oldukları yapısal bölümlerde bir araya getirecek şekilde tasarlanmıştır. Bu doğrultuda çıkarıma dayalı yöntemleri kullanırken tam metinde var olan yapısal düzeni koruyarak cümle seçmek ve çıktıda yine bu yapısal düzeni takip etmek yukarıda değinilen bağlam ve tutarlılık sorunlarına da çözüm olabilecek niteliktedir (Saggion ve Poibeau, 2013, s. 17). Temel prensibi kısaca özetlenen tez çalışmasında tek kaynaklı, tek dilli, çıkarıma dayalı ve genel özetlerin çıkarılacağı bir sistem üzerinde çalışılmıştır. Bu bağlamda çalışmanın temelini oluşturan hipotez: “Bilimsel makaleler için çıkarıma dayalı otomatik öz hazırlamada, tam metinlerin yapısal özelliklerini göz önüne almak, üretilecek özlerin kalitesini artıracaktır.” şeklinde belirlenmiştir.

Geliştirilen sistemin yazarlar ve editörler için öz üretme ve değerlendirme süreçlerinde bir yardımcı araç olarak kullanılabilir olması; elde edilecek yapısal özlerin daha kolay okunurluklu olmaları sebebiyle de potansiyel her düzeyden okurlara kolaylık sağlaması beklenmektedir.

Bu çalışma için kullanılan derlem, çeşitli ön işlemler ve veri işleme aşamalarından geçirilerek Hacettepe Üniversitesi Bilgi İşlem Dairesi tarafından tutulan veri tabanında saklanmıştır. Çalışmanın çıktılarında biri de Türkçe bibliyometrik analizler için de kullanılabilir bu işlenmiş veri kaynağıdır.

### **1.3. ARAŞTIRMANIN SORULARI**

Bu çalışma ile araştırılan sorular aşağıda sunulmaktadır:

1. Türkçe kütüphanecilik ve bilginilim literatüründe yer alan makalelerin tam metin organizasyonları bir bilimsel yayından beklenen temel yapısal özelliklere göre düzenlenmiş midir?
2. Türkçe kütüphanecilik ve bilginilim literatüründe yer alan makalelere ait özlerin okunurluk değerleri okunurluk ölçeğine göre nasıldır?
3. Türkçe kütüphanecilik ve bilginilim literatüründe yer alan makalelerden çıkarıma dayalı otomatik özetleme yöntemleri kullanılarak elde edilen yapısal ve klasik özlerin tam metni tarama başarımları nasıldır?
4. Türkçe kütüphanecilik ve bilginilim literatüründe yer alan makalelerden çıkarıma dayalı otomatik özetleme yöntemleri kullanılarak elde edilen yapısal ve klasik özlerin okunurluk değerleri okunurluk ölçeğine göre nasıldır?
5. Türkçe kütüphanecilik ve bilginilim literatüründe yer alan makaleler için çıkarıma dayalı otomatik özetleme yöntemleri ile öz elde etmede yapısal format kullanmanın sistem çıktılarının başarımlarına etkisi nasıldır?

### **1.4. ARAŞTIRMADA KULLANILAN YÖNTEM VE TEKNİKLER**

Bu çalışma kapsamında bir otomatik yapısal öz hazırlama sistemi önerilmektedir. Geliştirilen otomatik özetleme sistemi, tek dilli, tek kaynaklı ve çıkarıma dayalıdır. Sistemin çalışma prensibi makalelerin tam metinleri içindeki en önemli cümlelerin belirlenmesi ve belirlenen bu cümlelerin buldukları yapısal bölümlere göre sırasıyla seçilip otomatik yapısal öz (OYÖ) elde edecek şekilde özet dokümanlarına eklenmesidir. Yapısal özellikleri hesaplamalara katmanın sistem çıktılarına etkisinin değerlendirmesinde kullanılmak üzere otomatik klasik özlere (OKÖ) ihtiyaç duyulacağı



için sistem aynı cümle seçme fonksiyonu ile yapısal özellikleri devre dışı bırakarak da çalışacak şekilde tasarlanmıştır. Sistem çıktısı otomatik özler için kelime üst sınırı 250 olarak belirlenmiştir.

Sistem değerlendirmesi üç aşamalı planlanmıştır. İlk aşamada sistem tarafından OYÖ ve OKÖ'ler için seçilen cümlelerin tam metin içindeki yığılımları karşılaştırmalı olarak değerlendirilmektedir. İkinci aşamada derlem için tam metin, yazar tarafından üretilen öz (orijinal öz), OYÖ ve OKÖ okunurluk değerleri karşılaştırılmaktadır. Son aşamada, orijinal özler ile OYÖ ve OKÖ'ler arasında n-gram çakışmaları ile yapısal olan ve olmayan otomatik özlerin başarımları yine karşılaştırmalı olarak değerlendirilmektedir.

Bahsedilen otomatik yapısal özetleme sisteminin tasarlanması ve derlem üzerinde çalıştırılıp, çıktılarının değerlendirilmesi süreçlerinde çeşitli yöntem ve teknikler kullanılmıştır. Üçüncü bölümde detaylarıyla incelenen bu yöntem ve araçlar ana hatlarıyla Tablo 1'deki gibidir.

Metin ön işlemleri için öncelikle derleme ait makaleler temel bileşenlerine ayrılarak (başlık, öz, tam metin, kaynakça gibi) ilişkisel bir veri tabanında tutulabilecek hale gelmiştir. Daha sonra her bir makalenin tam metni işaretçiler tarafından yapısal özelliklerine göre işaretlenmiştir. İlişkisel veri tabanı paragrafların yapısal özelliklerinin hem işaretçi görüşlerine hem de yazarın kullandığı başlıklara göre cümlelerle ilişkisini koruyarak metnin cümlelerine bölünmesini sağlamıştır. Veri yapılandırma aşaması için metinler, okunurluk hesaplama, içeriklerin yapısal düzene uyumları gibi bazı analizler yapıldıktan sonra gövdeleme işleminden geçirilerek kelime sıklık vektörleri halinde ifade edilmiştir. Bu aşamadan sonra bu işlenmiş veri kullanılarak, derleme ait makaleler için çıkarıma dayalı otomatik özetleme yöntemleri ile otomatik özler üretilmiştir. Son olarak sistem çıktısı özlerin değerlendirme süreçleri Zemberek (Zemberek-NLP, 2018) kullanılarak elde edilen n-gram dizileri üzerinden, ROUGE 2.0 (Ganesan, 2018a) paketi kullanılarak otomatik olarak gerçekleştirilmiştir. Zemberek kullanılarak geliştirilen Java uygulamaları ile gerçekleştirilen durak kelimelerden arındırma vb. kelime ön işlemleri

ROUGE 2.0 paketini uygularken pasif hale getirilmiştir. ROUGE 2.0 sadece değerlendirme ölçümlerini uygulayan hesaplamaları gerçekleştirmek için kullanılmıştır.

Tablo 1. Tez çalışması kapsamında kullanılan yöntemler ve araçlar

Süreç	İşlem	Yöntem	Araç
Otomatik öz hazırlama sistemi tasarımı	Metin ön işlemleri	<ul style="list-style-type: none"> <li>• Web ara yüzünden yapılan işaretleme işlemleri</li> <li>• Veri tabanı işlemleri</li> <li>• Otomatik metin yapılandırma işlemleri</li> </ul>	PHP, MySQL, JavaScript, JQuery
Okunurlukların hesaplanması ve değerlendirilmesi	Okunurluk formülünün uygulanması ve sonuçların değerlendirilmesi	<ul style="list-style-type: none"> <li>• “Flesch Reading Ease” (FRE)’nin Türkçe’ye uyarlaması (Ateşman, 1997; Flesch,1948)</li> <li>• Korelasyon testleri, Saçılım grafikleri, Dağılım grafikleri</li> </ul>	MySQL, PHP, R Studio
Otomatik öz hazırlama sistemi tasarımı	Veri yapılandırma	Gövdeleme ve kelime sıklık vektörlerinin eldesi	Zemberek (Zemberek-NLP, 2018), Java, NetBeans, PHP, MySQL,
	Otomatik öz hazırlama	Çıkarıma dayalı yöntemler kullanarak öz için cümle seçimi yapılması	PHP, JavaScript, JQuery
Otomatik özlerin değerlendirilmesi	Otomatik özlerin n-gram birliktelik istatistiklerine göre değerlendirilmesi	ROUGE (Recall-Oriented Understudy for Gisting Evaluation) ölçümleri ile otomatik değerlendirme	R Studio, ROUGE 2.0, Zemberek, Java, Netbeans, MySQL

Yöntem bölümünde alt başlıklar halinde yapılan tüm işlemlerin detayları ele alınmaktadır. Tablo 1’de görülebileceği gibi, web tabanlı uygulamalar için PHP, JavaScript, JQuery; veri tabanı olarak MySQL; gövdeleme, öz nitelik vektörleri

oluřturma ve n-gram birliktelik hesaplamaları için Türkçe için açık kaynak kodlu bir proje olan Zemberek ile Java ve NetBeans IDE platformunun yanı sıra ROUGE 2.0 paketi kullanılmıřtır. İstatistiksel analizler ve verilerin görselleřtirilmesi için R Studio (R Core Team, 2020) kullanılmıřtır.

### **1.5. ARAŐTIRMANIN KAPSAMI**

Türkçe kütüphanecilik ve bilgibilim literatüründe yer alan yayınlar için hazırlanan derlem *Türk Kütüphanecilięi* dergisinde 1996-2015 yılları arasında yayınlanmış 249 ve *Bilgi Dünyası* dergisinde 2000-2015 yılları arasında yayınlanmış 172 hakemli Türkçe makaleyi içermektedir. Arařtırma kapsamı Türkçe ile sınırlı olduęundan bu dönemde yayınlanmış İngilizce makaleler arařtırma kapsamı dıřında bırakılmıřtır. Yine arařtırma kapsamında yazar tarafından üretilen özler deęerlendirme ařamasında kullanılacaęı için özü olmadan basılan makaleler kapsam dıřı bırakılmıřtır. Bu řekilde seçilen toplam 421 makalelik derlem, çeřitli metin ön iřlemlerinden geçirilerek iliřkisel veri tabanına kaydedilmiřtir.

Otomatik özetleme iři, çıkarıma dayalı yöntemlerle yapılmıřtır. Türkçe akademik makaleler üzerinde denenmiř çıkarıma dayalı üç farklı yöntemin birleřiminden oluřan doęrusal bir formül ile önemli cümle seçim iřlemi gerçekteřirilmifitir. Otomatik cümle seçim iřlemlerinde tam metinlerin yapısal özelliklerinin kullanılmasının çıktıya katkısı, otomatik yapısal ve otomatik klasik özlerin yazar tarafından üretilen özler ile karřılařtırılması ile deęerlendirilmiřtir.

### **1.6. ARAŐTIRMANIN DÜZENİ**

Çalıřmanın giriş bölümünde arařtırma konusunun önemi ele alındıktan sonra arařtırmanın amacı, hipotezi ve sorularına yer verilmiřtir. Arařtırmada kullanılan temel yöntem ve teknikler ile arařtırmanın düzeni üzerinde durulmuřtur.

İkinci bölümde bilimsel yayınların tarihçesi, makalelerin tam metinlerinde yapısal düzen kullanımını açasından ele alındıktan sonra bilimsel yayınların özleri üzerinde durulmuřtur.

Özler ile ilgili literatürün ardından genel hatlarıyla otomatik metin özetleme sistemleri incelenmiştir. Çıkarıma dayalı metin özetleme sistemleri, kullanılan yöntem ve teknikler ile ele alınmış ve değerlendirme yöntemlerine değinilmiş, Türkçe üzerinde gerçekleştirilen otomatik özetleme çalışmaları ile ilgili literatüre yer verilmiştir.

Araştırmada kullanılan yöntemlerin ele alındığı üçüncü bölümde öncelikle derlem ve özellikleri ile metin ön işlemleri açıklanmıştır. Metin ön işlemlerinden geçirilen makalelerin özlerinin ve tam metinlerinin okunurluk hesaplamalarının nasıl yapıldığı detaylandırıldıktan sonra otomatik cümle seçim işlemlerinin uygulanışına geçilmiştir. Bunun için öncelikle kelime sıklık vektörlerinin elde edilişi, daha sonra çıkarıma dayalı yöntemlerin uygulanışı ve cümle seçim işlemleri kullanılan yöntem ve tekniklerin detaylarıyla birlikte açıklanmıştır. En son sistem çıktısı olan yapısal özlerin değerlendirilme aşamasında kullanılan ROUGE paketleri ve uygulanmasına değinilmiştir.

Dördüncü bölümde uygulanan yöntem ve teknikler sonucunda elde edilen istatistiki bulgular ve yapısal özlerin başarımları analiz edilmektedir. Ayrıca derleme ait makalelerin tam metin yapısal özellikleri ve bu özelliklerin sistem çıktılarının başarımlarına etkisi üzerinde durulmaktadır. Bu bölümde hem derlemin hem de derlemden elde edilen özlerin okunurluklarına ilişkin bulgulara yer verilmektedir.

Beşinci ve son bölümde çalışmanın amacı doğrultusunda elde edilen bulgular yorumlanmaktadır. Ayrıca sonuçlar doğrultusunda getirilen önerilere de bu bölümde yer verilmektedir.

## 2. BÖLÜM: LİTERATÜR DEĞERLENDİRMESİ

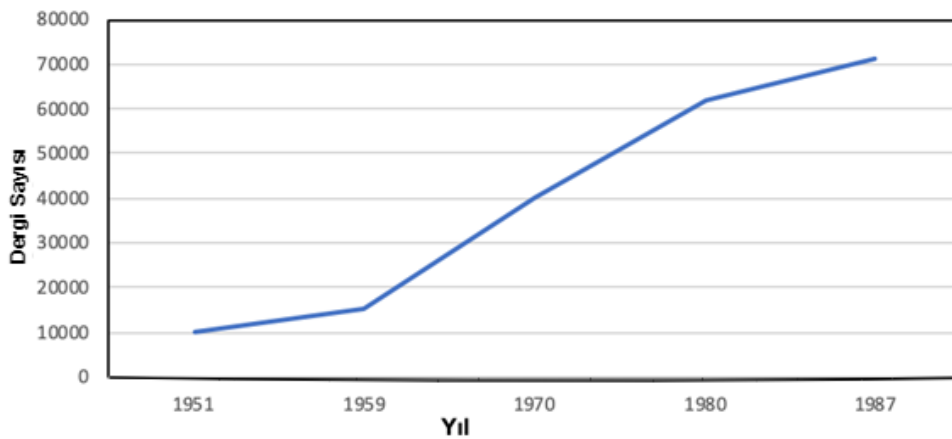
Bu bölümde çalışma kapsamında ele alınan iki temel konu olan akademik makalelerin yapısal düzenleri ve özleri ile otomatik metin özetleme ile ilgili literatür incelenmektedir. Bilimsel makalelerin başlangıcından günümüze geçirdiği değişimler yapısal düzenleri bağlamında ele alındıktan sonra özler ve çeşitleri üzerinde durulmaktadır. Otomatik metin özetleme konusu ise genel hatlarıyla ele alınarak çalışma kapsamında benimsenen çıkarıma dayalı otomatik metin özetleme yaklaşımı ve değerlendirme süreçlerine değinilmektedir. Hem özlerle hem de Türkçe otomatik metin özetleme çalışmaları ile ilgili yapılmış çalışmalar ele alınmaktadır.

### 2.1. BİLİMSEL YAYINLARIN EVRİMİ VE İMRAD

Bilimsel makalelerin ortaya çıkışı 17. yüzyıla kadar uzanmaktadır. İlk zamanlardan günümüze kadar güncel ihtiyaçlara ve bilimsel iletişimin koşullarına uyum sağlayarak değişen ve dönüşen makale yazım formatı ilk iki yüzyıl boyunca standardize edilmemiştir (Sollaci ve Pereira, 2004, s. 364). İlk tür bilimsel makaleler daha çok bilim insanlarının gözlemlerini aktardıkları, tarihsel olarak sıralanmış anlatılara dayalı formatlarda yayımlanmışlardır (Day, 1996, s. 4). Bu makaleler düz yazı formunu ve varsa deney raporlarını içeren, genellikle tek yazarlı, kibar bir dille yazılmış ve aynı anda birçok konuyu ele alan metinlerdir. Deneysel rapor ise genellikle olayların kronolojik sırada sunulmasından ibarettir (Sollaci ve Pereira, 2004, s. 364). 19. yüzyıl başlamak üzere iken, yöntem ve sonuçların makalelerde açıkça verilmesi ihtiyacı duyulmuş ve böylece bugün de kullandığımız yapısal formatlar ortaya çıkmıştır (Day, 1996, s. 3). Genel kabul görmüş Giriş, Yöntem, Bulgular, Tartışma ve Sonuç şeklinde ilerleyen temel yapısal başlıkların ilk harflerinden oluşan IMRAD (**I**ntroduction-Giriş, **M**ethods-Yöntem, **R**esults-Bulgular **and D**iscussion-Tartışma ve Sonuç) formatını 1876 yılında Pasteur'ün *Etudes sur la Biere* (fermantasyon çalışmaları) kitabında kullandığı kabul edilmektedir (Wu, 2011, s. 1346). Kitabında “giriş”, “yöntem” ve “tartışma” bölümlerini bu başlıkları açıkça kullanmadan da olsa belirgin biçimde ayırarak organize eden Pasteur'ün bu alışılmadık formatı bulmadaki motivasyonunun bilim camiasında çalışmalarının hesap verebilirliğini ve anlaşılabilirliğini artırmak olduğu ifade edilmektedir (Day, 1996, s. 4).

IMRAD hem dergilerde yer tasarrufu hem de editörler ve yazarlar için zaman ve hesap verebilirlik vadeden bir format olarak bilim camiasında kabul görmüş olsa da, ancak 1970’lerin sonunda araştırma makaleleri için baskın hale gelebilmiştir. Tıp makaleleri için standart biçim olarak kabul edilişi ise 1980’leri bulmaktadır (Wu, 2011, s. 1346).

Tıbbî bilimsel yazında IMRAD formatının geniş çapta benimsendiği dönemlere işaret etmek amacıyla yapılan bir çalışmada 1935 ile 1985 arasında geçen dönem 5 yıllık periyotlar halinde incelenmiştir. *British Medical Journal*, *American Medical Association*, *The Lancet*, ve *New England Journal of Medicine* dergilerinde yayımlanmış toplam 1297 makalenin yapısal içeriklerine, tam metinlerinde kullanılan başlıkların “yöntem”, “bulgular” ve “tartışma” kelimelerini açıkça içermesine göre karar verilmiştir. Yapılan değerlendirmeye göre ilk 20 yılda yayımlanmış makalelerde IMRAD kullanımı, hiç görülmemekten %20’lik bir dilimde gözlenmeye kadar ilerleyebilmiştir. 1955 ile 1975 arasında geçen 20 yıllık dönemde ise IMRAD formatının açıkça kullanıldığı makalelerin görülme sıklığı neredeyse 4 katına çıkarak derlemde %100’e ulaşmıştır. 1951 ile 1987 yılları arasında tüm dünyadan tahmini dergi sayılarını gösteren Şekil 1’i bu bağlamda incelediğimizde benzer bir artışın varlığından bahsedebilmekteyiz.



Şekil 1. 1951-1987 yılları arasında dünyada tahmini dergi sayıları (Meadows, 1997, s. 15)

Şekil 1’i 1951 sonrası, yani 2. Dünya Savaşı sonrasında yaşanan hızlı bilimsel gelişmelerin ardından gelen bilimsel yayın artışı olarak da değerlendirebiliriz. 1955 sonrası tıp menşeli yayınlarda yaşanan hızlı IMRAD benimsenme süreci, Şekil 1’de

görülebilen dünyada bilimsel dergi sayılarındaki artış eğrisi ile örtüşmektedir. Bu da makalelerin yapısal formatlarındaki gelişmelerin artan bilgi hacmiyle mücadele için atılan adımlar olarak da görülebileceği (Meadows, 1997, s. 13) görüşünü desteklemektedir.

IMRAD yapısının okurlara sağladığı yararlar içeriği evrensel başlıklar altında gruplayarak bilgiye kolay erişimi sağlamak ve bu modüler biçim ile okumayı kolaylaştırmaktır (Sollaci ve Pereira, 2004, s. 366). Okuyucu eğilimi metni doğrusal bir şekilde taramaktan çok, önceden belirlenmiş alanlarda bulunan özel bilgileri aramaktan yanadır (Meadows, 1985, s. 30; Sollaci ve Pereira, 2004, s. 366). Okuyucunun bilgi bulmada yaşadığı bu kolaylık şüphesiz bilgisayarlar için de geçerli olacaktır. İçeriği gruplanarak üretilmiş bir metnin düz metne göre daha iyi bir dizinlenme performansına sahip olacağı açıktır. Yazarlar için evrensel bir yapısal düzende üretim yapmanın başlıca faydaları ise Pasteur'ünküne benzer şekilde hesap verebilirliği ve anlaşılabilirliği artırmanın yanı sıra içeriği düzenlemede kolaylık ve öz kontrolüdür.

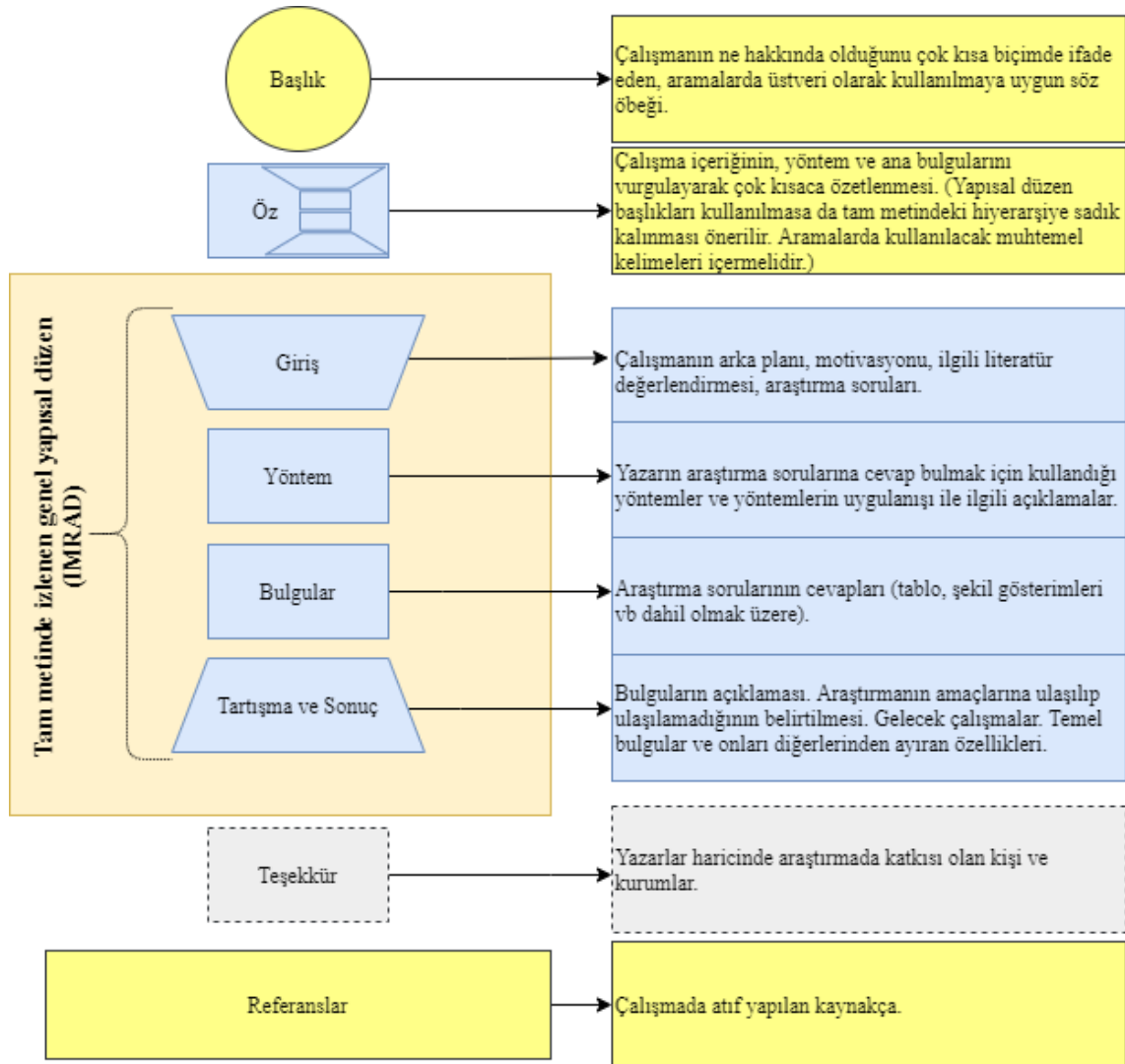
Makale tam metninin formatı bu şekilde evrilirken bir bütün olarak akademik makalelerin biçimi aşağıda başlıklarıyla verilen düzene kavuşmuştur (Meadows, 1997, s. 11):

- Başlık
- Yazar adı/adları ve adresi
- Öz
- Tam metin (Genellikle IMRAD benzeri bir formatı takip eder)
- Kaynaklar, tablolar, şekil ve altyazılar

Makalenin tam metni IMRAD'a göre yapılandırılırken, her bir yapısal bölüm, cevap bulması gereken sorularla ifade edilebilen, yazarlar, hakemler ve okurlar tarafından anlaşılması ve takip edilmesi kolay bir metin parçasını ifade eder:

- Giriş (Introduction): Hangi problem incelendi?
- Yöntem (Methods): Problem nasıl incelendi?
- Bulgular (Results): Neler bulundu?
- Tartışma (Discussion): Bulgular ne anlama gelmektedir? (Day, 1996, s. 5).

Bu çalışma kapsamında makalelerin çözümlenmesinde kullanılan yapısal bölümlerde IMRAD düzeni takip edilirken, Şekil 2'deki diyagram model temel alınmıştır.



Şekil 2. Makalelerin yapısal bölümleri diyagramı. (Wu'nun IMRAD diyagramından uyarlanmıştır (Wu, 2011, s. 1349))

## 2.2. BİLİMSEL YAYINLARIN ÖZLERİ

Öz, bilimsel bir makalenin içeriğinin kısa ve kapsamlı bir özetidir. Başlık ve varsa anahtar kelimeler ile beraber okurun ilk karşılaştığı ve incelediği kısımdır. En çok bilinen tanımı ile bir bilimsel yayının özü, iyi organize edilmiş, kendi kendine yeten, kısa ve bilgi yoğun, aynı zamanda okunabilir bir makale bölümüdür (APA, 2010, s. 26).



Özün makalede yer almasının amacı, okuyucuyu tam metni okumadan içerik hakkında bilgilendirerek araştırma sorusu ve içerik arasındaki ilişkilere yönelik karar vermesini sağlamaktır. Özler, aynı zamanda makalenin başlığı ve anahtar kelimeleriyle birlikte, belgeye ilgi duyabilecek kişilerin elektronik ortamlarda yapacağı taramalar için dizinleme materyali de sağlar. Makalenin özü, ana metnine göre daha küçük bir metin verisi olduğu ve tam metnin içeriğini yansıttığı için dizinlemede kullanımını son derece ekonomiktir.

Araştırma makalelerine ait özler ilk kez *Chemical Abstracts* gibi ikincil servislerde yayınlanmaya başlamıştır. Bu tür ikincil servislerden sonra araştırma makalelerinin özle başlamasının yaygınlaşması, 2. Dünya Savaşından hemen sonra, 1950’li ve 1960’lı yıllardan itibaren (Hartley ve Betts, 2009, s. 2010). Bir makalenin özü okuyucuların ilgili yayının tam metin içeriği hakkında hızlı bir şekilde bilgi edinmesine olanak tanır. Özlerin değişmez bir makale bölümü olarak sahneye çıkışı, bu dönemde yaşanan bilimsel yayın artışı karşısında bilim insanlarının güncel yayınları takip etme kaygılarına paralel bir gelişme olarak ele alınabilir.

APA tarafından yayımlanan ve özellikle sosyal bilimler alanında yaygın bir şekilde kabul gören yazım kurallarını belirleyen kılavuza göre (2010, s. 26), iyi bir öz aşağıdaki özellikleri taşımaktadır:

- Kesin olmalıdır: Tam metnin amacını ve içeriğini doğru şekilde yansıtmalı. İçerikte olmayan bir detaydan bahsetmemelidir.
- Değerlendirici olmamalıdır: Makale tam metnini değerlendirmemeli, rapor etmelidir.
- Tutarlı ve okunabilir olmalıdır: Süslemelerden uzak, net bir dille yazılmalıdır. Pasif yerine aktif cümle çatıları tercih edilmelidir.
- Özlü olmalıdır: Kısa ve mümkün olan en bilgilendirici cümlelere yer verilmelidir. Hedef kitlenin elektronik aramalarında kullanacağı muhtemel kelimeler kullanılarak yazılmalıdır.

Yukarıdaki özellikler doğrultusunda, iyi yazılmış bir öz ilgili yayının en kuvvetli paragrafı olmak durumundadır. İyi hazırlanmış bir özün araştırmanın kapsamı ve esas

amaçları, kullanılan yöntemler, bulgular ve ana sonuçlara dair bilgi verici olması beklenir (Day, 1996, s. 21; Indiana University Writing Tutorial Services, 2020). Özlerin kapsamlı olması bir zorunluluktur. Buna karşın, tam metne dair tüm ayrıntıların özet halinde verilmesi okuyucuların makalenin tamamını okumaktan vazgeçmelerine ya da yorulup dikkatlerinin azalmasına neden olabilir. Bir bilimsel yayının özünü formüle etmenin hassas bir dengesi vardır. Bir bilimsel yayının özünden beklenenler ve öz yazarken kaçınılması gerekenler Tablo 2’de sıralanmıştır.

Tablo 2. Öz yazımında yapılması ve kaçınılması gerekenler (APA, 2010, s. 26; Day, 1996, s. 21; Labaree, 2018)

Öz Yazımında Yer Verilmesi Gerekenler	Öz Yazımında Kaçınılması Gerekenler
<ul style="list-style-type: none"> <li>• Araştırmanın arka planı, amacı, motivasyonu</li> <li>• Araştırma problem(ler)i;</li> <li>• Çalışmanın temel tasarımı, yöntem(ler)i;</li> <li>• Ortaya çıkan önemli bulgular veya eğilimler;</li> <li>• Yorumların ve sonucun kısa bir özeti</li> </ul>	<ul style="list-style-type: none"> <li>• Çalışmanın arka plan bilgisinin detayı</li> <li>• Her türlü literatür atfı.</li> <li>• Bitmemiş cümleler (ör. “...” kullanımı)</li> <li>• Her çeşit şekil ya da tablo vb.</li> <li>• Sonuç kısmında “çok”, “küçük”, “kayda değer” gibi belirsiz ifadeler.</li> <li>• Okuyucunun kafasını karıştıracak kısaltma ya da jargon kullanımı</li> </ul>

Yazarlar öz hazırlarken format belirlemede makalelerini yayımlamak istedikleri dergilerin stillerine göre hareket ederler. Halen dergilerin baskın çoğunluğu klasik öz formatı kabul etse de, öz yazımı konusunda farklı yaklaşımlar mevcuttur. Örneğin, sosyal medyada yayımlarını duyurabilmek için kullanabilecekleri bir formatta ayrı bir öz talep eden ya da farklı okur kitlelerine de seslenebilmek için alan dışından her türlü okuyucunun anlayabileceği bir özet (lay summary) de isteyen dergilerin temel amacı her zaman geniş bir yelpazeden okur kitlesinin dikkatini çekebilmektir. Günümüzde dergilerin talep edebildikleri farklı öz formatları aşağıdaki gibi derlenebilir (Hartley ve Cabanac, 2017):

**Klasik Özler:** Klasik yaklaşımla, tek blok halinde yazılan özlerdir. Bu tür özler bilgi verici ya da tanımlayıcı olabilirler (Day, 1996, s. 22).

- Bilgi verici özler, araştırma makalesi, deneysel makale gibi iyi yapılandırılmış bilimsel yayınlarda yer alır. Bu tür özler genellikle dört bölümden oluşur:

- Amaç
- Yöntem
- Bulgular
- Yorumlar

Amaç bölümü araştırmanın birincil amacını dile getirirken, hipotezini de içerebilir. Yöntem bölümünde makalede geçen tüm yöntemler bütünüyle ele alınmaz sadece araştırmanın kapsamının anlaşılacağı detayda başlıca yöntemler dile getirilir. Bulgular da yöntemle benzer olarak, makalenin araştırma sorularının cevaplarını ele alınan tüm bulgularla değil, sadece en önemlileri hakkında bir bilgilendirme ile verilir. Yorumlar bölümü bulguları değerlendirmeye yönelik analizleri ya da deney sonuçlarını verir. Hipotezin doğrulanıp doğrulanmadığına da bu bölümde yer verilebilir (Day, 1996, s. 22; Indiana University Writing Tutorial Services, 2020).

- Tanımlayıcı özler, genellikle literatür tarama makaleleri, derleme, editöryal ya da kitap için yazılan daha az yapılandırılmış özlerdir. Genellikle,

- Kapsam,
- Kullanılan argümanlar
- Sonuç

olmak üzere üç bölümden oluşur. Kapsam orijinal dokümanda ele alınan materyalin sınırlarını ve dokümanın temel hedefini belirtir. Kullanılan argümanlar, dokümanda yer aldıkları sırayla verilmelidir. Eğer doküman argümanların sıralanışından oluşmuyorsa onun yerine ana hatları burada incelenir. Sonuç bölümünde ise dokümanın temel kapanış argümanına ve onun etkilerine yer verilir (Day, 1996, s. 22; Indiana University Writing Tutorial Services, 2020).

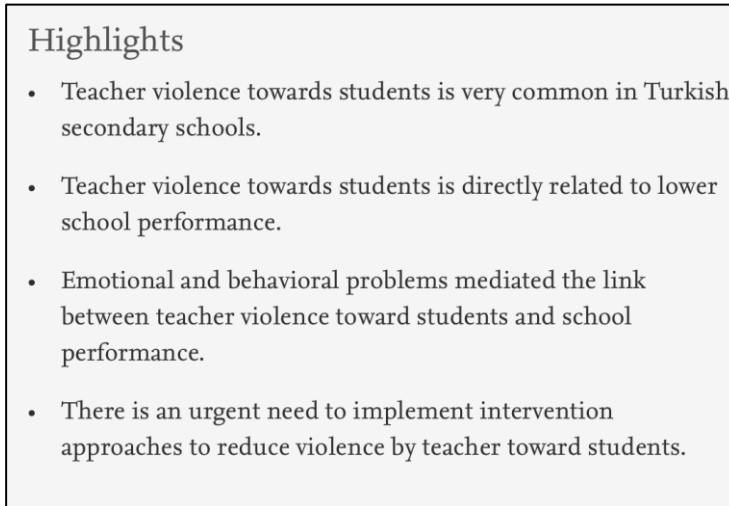
Günümüzde halen baskın olarak kullanılan öz formatı klasiktir. Buna karşın, klasik özler okunurluk, içeriği yansıtma, üretim, denetim, erişim kolaylığı gibi bilimsel iletişim sürecinde önemli yer tutan pek çok konuda yetersiz bulunmaktadır (Bitri ve Keseroğlu, 2015; Gazni, 2011; Hartley vd, 2003; Hartley, 2008; Hartley ve Betts, 2009; Mabe ve Amin, 2002).

**Yapısal Özler:** Bu tür öz tercih eden dergiler, yapısal özleri, klasik özlerle birlikte verebileceği gibi makalenin başında tek başına da konumlandırmayı tercih edebilir. Yapısal düzen içeriği daha anlaşılır ve okunabilir kılmak için kullanılan bir formattır. Yapısal öz kullanan dergiler çoğunlukla IMRAD'a benzer bir düzeni yazım kılavuzlarında detaylarıyla açıklayarak talep eder. Buna göre öz yazarken, makalenin, giriş, yöntem, araştırma, sonuç ve değerlendirme bölümlerinden önemli görülen noktalara yer verilmesi istenir. Böylece, tam metni modüler bir şekilde özetleyen, klasik özlerden daha kolay okunduğu halde daha fazla bilgi içerme kapasitesine sahip özler elde edilir. 1980'lerden itibaren zorunlu kullanıldığı tıp alanında klinik kararlara rehberlik etmede faydalı bir kaynak sunabilen yapısal özlere (Fontelo, Gavino ve Sarmiento, 2013, s. 209), başta psikoloji ve eğitim bilimleri olmak üzere diğer alanlarda da sıklıkla rastlanmaktadır. Yapısal öz kullanımının zorunlu tutulduğu dergiler öncelikle deneysel çalışma alanlarında yayımlananlardır. Oysa tüm disiplinler için yapılan araştırmaların bir temeli ve bir amacı olması gerektiği, amaca ulaşmak için kullanılan yöntemler (deneyden sistematik olmayan incelemeye kadar), elde edilen bulgular, sınırlamalar ve sonuçlar olacağı ve gerçek çalışma bağlamında tüm bunların dile getirilebileceği (Hartley ve Betts, 2009, s. 2016) kabul edilmektedir. Yapısal öz kullanımına bilimsel iletişime hız ve tutarlılık kazandırmak amacıyla başlanmıştır ve deneysel olan ya da olmayan tüm alanlarda da bu özelliklerinden faydalanmaya çalışmak akılcı bir yaklaşım olacaktır. Şekil 3'te verilen, *Library Review*'da yayımlanmış yapısal öz örneği, sosyal bilimler için alternatif bir yapı olarak incelenebilir.

<p><b>Abstract</b></p> <p><b>Purpose</b> – The purpose of this paper is to explore the benefits of a writing support programme in developing the skills and motivation of librarians to write for academic publication.</p> <p><b>Design/methodology/approach</b> – A brief review of the literature is presented. The model developed and implemented by this author is outlined. Findings from a survey of participants are discussed.</p> <p><b>Research limitations/implications</b> – The formal programme commenced in 2007. The publication process takes time, particularly in the case of peer-reviewed journals. This is exploratory work. It will take time to build up a body of information and a community of librarians writing for publication. Initial evidence indicates there is significant value to the programme.</p> <p><b>Practical implications</b> – The model is transferable and could help in building skills and confidence in academic writing. In addition academic writing could serve as a bridge between lecturing and library staff, addressing issues of common concern across the Academy.</p> <p><b>Originality/value</b> – This is the first formal writing support programme for librarians in Irish universities. Models exist in the USA. A similar model is used in the UK and Ireland to support lecturing staff writing for publication.</p> <p><b>Keywords</b> Librarians, Authorship, Serials, Skills training, Ireland, Northern Ireland</p> <p><b>Paper type</b> Case study</p>
---

Şekil 3. Yapısal öz örneği (Fallon, 2009)

**Önemli Cümle (Highlights) Özleri:** Makalenin önemli üç dört cümlesinin seçilerek listelenmesinden oluşur. Örneği Şekil 4’de incelenebilir.



Şekil 4. Önemli cümle öz örneği (Kızıltepe, Irmak, Eslek ve Hecker, 2020)

Önemli cümle özleri özellikle Elsevier tarafından yayınlanan dergilerde, bilinen özden önce ya da sonra yer alabilmektedir.

**Grafik Özler:** Klasik bir anlayışla öz yazarken her türlü grafik ya da şekil kullanımından kaçınılması gerekirken, grafik özler bunun tam tersidir. Grafik özler, bir makalenin ana bulgularını sadece grafik veya şekle dayanan görsel bir özet olarak sunar. Bu tür özler sadece tek bir grafikten oluşabileceği gibi görsel özete birkaç cümle yazılı olarak eşlik edebilir. Grafik özlerin savunucuları arasında, *International Journal of Pharmaceutics* ve *Journal of Ethnopharmacology* dergileri örnek olarak verilebilir.

**Video Özler:** Bu tür özleri kullanan dergilerde yayının klasik özüne ek olarak yazarın yayının sunumunu yaptığı kısa, genellikle beş dakikalık bir videonun bağlantısı da verilir. Örneğin: *Nature Chemistry*, *The New Journal of Physics* ve *The British Medical Journal*.

**Okunabilir Özler:** Konu hakkında şimdiye kadar bilinenler, makalenin konuya katkısı, makalenin uygulamaya ve/veya politikalara etkileri nelerdir gibi önceden belirlenmiş başlıklar altında bilgi sağlar. Yayın kurulu bu listelerin, alanda mesleki deneyimi olan saha çalışanlarına makale ile raporlanan çalışmaların kendi günlük aktiviteleri ile ilgisini kurmada, yazarlara ise başarılı bir makalenin planlamasında yardımcı olacağını düşünmektedir. Örneğin: *British Journal of Educational Technology*.

**Tweet Özler:** *Methods in Ecology and Evolution* ya da *An International Journal of Obstetrics and Gynaecology* örneğindeki gibi dergiler yayınların sosyal medyadaki yayılımını sağlamak için klasik özün yanı sıra tweet edilebilir karakter sınırında bir öz daha talep edebilmektedir.

**Bilgisayarca Üretilmiş Özler:** Günümüzde bilgisayarca öz hazırlama yaklaşımına denk düşen ilk uygulamalar *Chemical Abstracts* gibi öz dergilerinde basılan, profesyonel özetleyicilerin hazırladıkları özlerdir (Hartley ve Betts, 2009, s. 2010). 50 yılı aşkın zamandır devam eden otomatik metin özetleme çalışmalarına rağmen bilgisayarca öz üretme işlemi henüz yaygın kullanıma geçmemiştir. Ancak başka pek çok alanda yaygın kullanıma geçen otomasyon sistemleri ve hızlı teknolojik gelişmeler er ya da geç bilgisayarca hazırlanan makale özlerinin de bir norm olacağı zamanları işaret etmektedir. İyi bir öz, bilimsel iletişimin sürmesi için zorunlu olan araştırma sonuçlarının yayınlanması, etkileşimi ve yayılması yolundaki ilk adımdır. Araştırmacılar çoğu zaman alanlarında yayımlanan tüm makalelerin tam metinlerini okumaya yetişemedikleri için özler sayesinde bilimsel gelişmelerden güncel şekilde haberdar olmayı başarırlar. Bilimsel iletişim bağlamında, değişen kullanıcı alışkanlıklarını gözeterek yeni öz formatları üzerinde daha çok düşünülerek elektronik ortamın doğasına uygun sunum biçimleri araştırmak akademik yayıncılık için elzemdir.

### 2.3. ÖZLERLE İLGİLİ AKADEMİK ÇALIŞMALAR

Özler üzerine yapılmış bilimsel araştırmalar (i) özlerin içeriği ve farklı disiplinler için nelerin uygun olacağı, (ii) basımları ve düzenleri de dâhil olmak üzere okunurlukları ve

dil özellikleri (iii) bilgilerin sunumu ve sıralanışının resmileştirilmesi (yapılandırılmış özler) olmak üzere üç farklı şekilde gruplanmaktadır (Hartley ve Betts, 2009, s. 210-211).

Bilgi kullanıcılarının ihtiyaçlarını araştıran bir çalışmada (Nicholas, Huntington ve Jamali, 2007, s. 448) 1000'den fazla bilim insanının ScienceDirect'teki 17.000 oturumunun işlem kayıtları incelenmiştir. Buna göre, kullanıcıların en az %20'sinin sadece öz okuduğu ve ilgili makaleleri seçmeleri ve araştırmaları için gerekli ön bilgileri sağlamak adına özlere güvendikleri saptanmıştır.

Özlerin mantıksal yapıları ve yazımları ile ilgili öneriler üzerine yoğunlaşan çalışmalara göre özler de makalelerin tam metinleri gibi yapısal bir biçim takip etmelidir, ancak yazarlar nadiren bu konudaki önerilere uyum sağlamaktadırlar (Atanassova vd., 2016; Jamar vd., 2014; Šauperl vd., 2008). *PLOS* dergilerinde yayınlanmış 85.000'den fazla makalenin özleri ile tam metinlerini cümlelerinin benzerliklerine göre inceleyen bir çalışmaya göre, özlerin %84'ü tam metinle benzerliği tespit edilebilen en az bir cümle içermektedir (Atanassova vd., 2016, s. 643). Aynı çalışmada, gövde metinlerinin IMRAD düzeni de incelenmiş ve makalelerin gövde metinlerinde %98 oranında dört IMRAD bölümü kullanıldığı görülmüştür (Atanassova vd., 2016, s. 639). Gövde-öz cümle benzerliklerinin analizi sonucunda ulaşılan sonuç ise özlerin büyük oranda tam metinlerin ilk %4'lük dilimi ile son %5'lik kısımlarından oluşmasıdır (Atanassova vd., 2016, s. 646). Buna göre özlerin dilbilimsel materyallerinin büyük kısmının giriş bölümlerinin başlangıcından ve sonuç bölümlerinin son kısımlarından alıntılanarak şekillendirildiği sonucuna varılmıştır. Yöntem ve bulgular kısımlarında ise benzerlik oranlarında önemli bir düşüş görülmüştür (Atanassova vd., 2016, s. 645). Tam metin ve öz cümlelerinin benzerlikleri, kosinüs benzerliği, Levenshtein uzaklığı ve iki cümlenin birbirinin alt cümlesi olup olmamasına göre puanlanmıştır. Buna göre, öz cümlelerinin sadece %25'inin tam metin cümlelerine %60'ın üzerinde bir oranla benzerlik gösterdiği saptanmıştır (Atanassova vd., 2016, s. 642). Çalışma bu yönüyle bilimsel yayınların gövde metinlerinin retorik yapısı ile insan eliyle oluşturulan özlerin içeriği arasındaki ilişkiyi incelemesi ile önem kazanırken, tam metnin yapısal düzeninin özlerde takip edilmediğini de ortaya koymuştur. Slovenya'da yayımlanan Sosyoloji, Farmakoloji ve

Slovenya Dili ve Edebiyatı disiplinlerine ait çalışmaların, üç disiplinden toplam 100 makalelik bir derlemeyi inceleyen bir başka araştırma, yayınları hem ISO 214 yapısal düzeni, hem de Slovenyalı bir bilim insanının sosyoloji alanı için önerdiği yapısal düzen ile uyumlarına göre analiz etmiştir. Sonuç olarak yine yazarların çoğunlukla önerilen standartlara uyum sağlamaktan uzak oldukları saptanmıştır (Šauperl vd., 2008, s. 1430). Malzeme bilimi ve teknolojisi ile kütüphane ve bilginin alanlarındaki özetlerin mantıksal yapısının ISO 214 veya IMRAD formatlarına uyup uymadığını inceleyip, öz bileşenleri için kılavuz önermeyi amaçlayan bir diğer çalışmanın (Jamar vd., 2014, s. 21) sonuçlarına göre de yine yazarların özetlerde her zaman tüm IMRAD bölümlerini kullanmadıkları söylenebilir. Araştırmanın sonuçlarına göre öz oluştururken, yapısal bölümlere göre kullanılacak cümle sayıları tavsiye edilmiştir. Buna göre özetler, üç giriş, üç yöntem ve üç bulgular cümlesi içermelidir. İki tartışma ve sonuçlar cümlesi de bulguların çalışmanın bir parçası olmayan konular üzerindeki etkilerini, olası uygulama ve ileri çalışma önerileriyle birlikte araştırmanın bir değerlendirmesini sunmalıdır (Jamar vd., 2014, s. 29).

Özetlerin kaliteleri bir diğer araştırma konusudur. Genel anlamda herhangi bir özet için kalite değerlendirmesi görecelidir. Özetlerin kalitesini değerlendirme işi de bu bağlamda ele alınarak belli bazı ölçütlere göre gerçekleştirilir. Önemsenen belli başlı ölçütler okunurluk, yapı ve içeriğe dairdir. Bir dokümanın okunurluğu, okuyucu ile metin arasındaki ilişkiden hatta bu ilişkinin gerçekleştiği ortamdan etkilenebilen bir kavramdır (Çoban, 2014, s. 97). Bu yüzden metinde geçen kelimelerin hece uzunluğu ve cümlelerin kelime uzunluğunu temel alan sayısal hesaplamalar okunurluğa dair nesnel bir veri elde edilebildiği için değerlendirmelerde sıklıkla kullanılır. Bunlardan en bilinen ve yaygın olarak kullanılan Flesch okunurluk testini, kullanıcı görüşleri ile karşılaştırmalı olarak değerlendiren bir çalışma (Ufnalska ve Hartley, 2009) okunurluk hesaplamaları ve kontrol listesiyle kullanıcı görüşü toplamayı birlikte yürütmüştür. Araştırmayı üç farklı klasik özün okunurluğu bağlamında sınırlı tutan çalışmada okunurluk değerlendirmesinde kullanıcılara yönelik kontrol listesi için -Anlaşılabilirlik, -Dilbilgisi kurallarına uyum, -Yazım kurallarına uyum, -Yapı, -İçerdiği bilgi, -Kısalık, -Ana dili makalenin dilinden farklı olan okurlara uygunluk başlıklarına yer verilmiştir. Flesch okunurluk skoru en yüksek hesaplanan öz için, kontrol listeleri ile 5'li Likert ölçeğe göre yapılan



değerlendirmede en düşük puan toplanmıştır. Çalışmada okunurluk hesaplamalarının özlerin kalitesini değerlendirmede tek başına yeterli olamayacağı sonucuna varılmıştır. Çok atıf almış makalelerin özlerinin okunurluğunu değerlendiren bir diğer araştırma (Gazni, 2011) Harvard Üniversitesi, Johns Hopkins Üniversitesi, Stanford Üniversitesi, Washington Üniversitesi ve Max Planck Enstitüsü'nün 1975-2009 yılları arasındaki yayınlarını Flesch okunurluk formülünün sonuçlarına göre incelemiştir. Buna göre her bir eğitim kurumunun yayınları için, atıf sayıları ile özlerin okunurlukları arasında çeşitli derecelerde ters bir korelasyonun varlığı gösterilmiştir. Yani yayınların okunurluk değerleri düştükçe atıfları artmaktadır. Çalışmada belirtildiği üzere, araştırmacıların makale özlerini “çok zor” okunurlukta ürettiği olmaları akademik çevrelerde sorun olarak görülmemektedir. Bu durumun sebeplerinden biri okuyucu hedef kitlesinin o alanda yoğunlaşarak çalışmış, üst düzey eğitim görmüş kişiler olmasıdır. Bir diğer sebep olarak akademik camiada “zor” okunurluğun “tolere” edilir, “prestijli” bir gerçek olarak kabul edildiği düşüncesi öne sürülmüştür. Aralarında *Nature*, *Science*, *NEJM*, *The Lancet*, *PNAS* ve *JAMA*'nın da bulunduğu, sosyal bilimler, sağlık bilimleri, biyoloji vb. pek çok alandan 123 dergiden, 1881 ile 2015 arasından yayınlanmış 709.577 makalenin özlerinin okunurluğunu inceleyen bir çalışmanın sonuçlarına göre bilimsel okunabilirlik yıllara göre giderek azalmaktadır (Plavén-Sigray, Matheson, Schiffler, ve Thompson, 2017). Ayrıca özlerin okunurluk değerleri tam metinlere göre belirgin şekilde daha düşüktür. Bilimsel jargonun giderek artan bir kullanımının göstergesi olarak yorumlanan analiz sonuçlarına göre araştırma bulgularının hem tekrarlanabilirliği hem de erişilebilirliği bu eğilimden olumsuz etkilenmektedir. Alanında en yüksek atıf alan üç turizm dergisinin okunabilirliği üzerine yapılan bir çalışmaya göre (Dolnicar ve Chapple, 2015) 1993, 2003 ve 2013 yıllarında yayımlanmış sayılardan toplam 493 makale analiz edilmiştir. Çalışmanın sonuçlarına göre, makalelerin %75'inin FRE değerleri “çok zor” olarak bulunmuştur. Yıllara göre okunurluk değerleri ise belirgin bir şekilde düşmektedir. Öz taramalarında bilgiye erişim bağlamında daha çok okuyucuya hitap edebilmek için başlatılan ve kullanıcı/sistem arayüzlerini iyileştirme çalışmaları kapsamında yapılan bir okunurluk araştırması Flesch ölçümlerine göre özleri ve tam metinlerini analiz etmiştir. Bu çalışmanın sonuçlarına göre, (Dronberger ve Kowitz, 1975, s. 110) tam metinlerin okunurlukları özlerden daha düşük bulunmuştur. Okunurluk düzeyi olarak hem tam metinler, hem de özler zor okunur olarak rapor edilmiştir.

Yapılandırılmış özlerin etkinliğini başlık/alt başlık kullanımının etkisi ile birlikte araştıran bir çalışmada kullanıcı görüşleri değerlendirilmiştir (Hartley ve Betts, 2007). Kullanıcı grubu, bilgibilim ve psikoloji eğitimi alan 64 lisans öğrencisi, çeşitli alanlardan 225 akademisyen ve araştırmacı ile 250 bilgibilimciden oluşturulmuştur. Kullanıcılardan, altı makalenin alt başlıklara ayrılmış yapısal; alt başlıklara ayrılmış klasik; alt başlıkların kullanılmadığı yapısal; alt başlıkların kullanılmadığı klasik sunumlarını değerlendirmeleri istenmiştir. Sonuç olarak, başta bilgibilimciler ve öğrenciler olmak üzere yapılandırılmış sürümlere geleneksel olanlardan daha olumlu değerlendirmeler yapıldığı, başlıkların ise değerlendirmeye olumlu ya da olumsuz bir etkisi olmadığı gösterilmiştir (Hartley ve Betts, 2007, s. 238). Çalışmanın sonuçlarında yapılandırılmış özlerde sağlanan daha fazla bilginin, yapısal olarak hazırlanmış özlerin geleneksel olanlardan daha üstün bulunmasına yol açmış olabileceğine dikkat çekilmiştir (Hartley ve Betts, 2007, s. 239).

#### **2.4. GENEL HATLARIYLA OTOMATİK METİN ÖZETLEME**

Otomatik metin özetlemenin bir disiplin olarak geçmişi 1950'lerin sonunda Luhn'un ilk çalışmalarına dayanmaktadır ve temel motivasyonu günümüzdeki gibi giderek artan yayın kapasitesinin işlenmesidir. Luhn (1958) kimya alanındaki akademik yayınlara otomatik öz üreten bir sistem geliştirmiştir.

Otomatik bir doküman özetleme sistemi “çıkarıma” ya da “yoruma” dayalı olabilir. Yoruma dayalı yöntemler tam metindeki cümleleri yeniden yorumlayarak özet oluşturur. Bu şekilde özet oluşturulurken, orijinal metnin belirgin fikirlerini yakalayan cümle ya da cümle öbekleri içerdikleri önemli fikirleri iletecek şekilde kısaltılarak yeniden yazılmaya çalışılır. Yoruma dayalı özetleme yöntemleri, verilen belgelerde üstü kapalı geçilen konuları bile tamamen anlamaya çalışır. Bu yaklaşım, insan özetlemesine benzer bir yol izlese de pratikte gelinen noktada bilgisayarların bir insan özetinin performansına ulaşması zordur. Ayrıca yoruma dayalı otomatik özetleme yöntemlerinde metnin içeriğini aktarmada yeni cümleler üretmek için, dil bilgisi ve sözlük işlemeyi gerektiren doğal dil işleme süreçlerinin çok bileşenli ve maliyetli işlemlerine ihtiyaç duymaktadırlar. Çıkarıma dayalı otomatik özetleme sonraki bölümde detaylarıyla ele alınmıştır. Kısaca

metindeki en önemli cümleleri temsilci seçip bir araya getirerek özet meydana getirme prensibine göre çalıştığı söylenebilir.

Bu iki yaklaşımdan birinin diğerinden daha iyi olduğunu söylemek olanaklı olmamakla birlikte, her iki yöntemin de üstünlüklerinden ve zayıf yönlerinden bahsedilebilmektedir. Yorumlamaya dayalı yöntemler için daha gelişmiş özetler sağlamayı hedeflemesi ve yüksek sıkıştırma oranları uygulayabilmesi; çıkarıma dayalı yaklaşımların daha büyük miktarlarda veriye daha düşük maliyetle uygulanabilmesi üstün yönleridir. Buna karşın çıkarıma dayalı yöntemlerde ortaya çıkan özetlerin anlamsal bütünlüğünün bazen tutarsız olabilmesi; yoruma dayalı otomatik özetleme yöntemlerinin ise uygulamadaki karmaşıklığı yöntemlerin zayıf yönleridir. Uygulama kolaylığı sebebiyle çoğu otomatik metin özetleme sistemi çıkarıma dayalıdır ve bu kolaylık da çıkarıma dayalı yöntemlerin kendi standartlarını geliştirmesine yol açmıştır (Gambhir ve Gupta, 2017, s. 3).

Otomatik özetleme sisteminin bir diğer sınıflandırması üretilen özete “genel” ya da “sorgu temelli” olmasıdır (Torres-Moreno, 2014, s. 43). Genel özet, metnin tamamından çıkartılan ayrıntılı özet iken, sorgu temelli özet ise sadece kullanıcının yazdığı sorgu ile ilgili cümleleri seçerek çıkartılmış özettir (Lloret ve Palomar, 2012, s. 4). Özetlenecek doküman sayısına göre “tekli” ya da “çoklu” doküman özetleme yapılabilir. Tekli doküman özetleme her bir çıktı için tek bir doküman üzerinde işlem yaparken, çoklu doküman özetleme birbirleri ile ilgili olan/olduğu düşünülen birden fazla kaynağı tarar ve ortak bir özet çıkartır (Saggion ve Poibeau, 2013, s. 8). Ayrıca özetlenecek doküman yığını “tek bir dilde” yazılmış olabileceği gibi, özetleme işlemi “çoklu dille” de yapılabilir. Üretilen çıktı özelliklerine göre bir özet “belirtici” (indicative) ya da “bilgilendirici” (informative) olabilir. Belirtici özetler dokümanın ne hakkında olduğunun bilgisini verirken bilgilendirici özetler tüm bilgiyi daha ayrıntılı bir formda vermektedir (Lloret ve Palomar, 2012, s. 4).

## 2.5. ÇIKARIMA DAYALI OTOMATİK METİN ÖZETLEME YAKLAŞIMI

Çıkarıma dayalı özetleme, metindeki önemli, yani en çok bilgiyi taşıyan cümleleri tanımlamak, çıkarmak ve bunları istenen sıkıştırma oranına göre bir araya getirerek özet oluşturmaktır. Cümlelerin önem dereceleri çeşitli özellikler göz önünde bulundurularak ağırlıklandırılır ve bu ağırlığa bağlı olarak her bir cümleye puan verildikten sonra özeti üretmek için yüksek skorlu cümle seçimi yapılır. Yani oluşturulan özet, metnin orijinal cümlelerinin bir seçkisidir ve bu cümlelerin çıkarımı için kullanılan çeşitli yöntemler vardır. Yöntemleri uygulamaya geçmeden önce tam metin dokümanlar çeşitli ön işlemlerden geçirilmelidir.

Luhn (1958) öncü çalışmasında bir metinde sıklıkla geçen terimlerin metin içindeki önemini diğerlerinden çok olması gerektiği ve bu terimleri içeren cümlelerin de önemli cümleler olduğu düşüncesinden yola çıkarak bir otomatik metin özetleme sistemi geliştirmiştir. Geliştirdiği sistemde terimleri sıklıklarına göre puanlandırırken, alt ve üst kesme değerlerinin arasında kalan terimleri dikkate almıştır. Bu şekilde, hem günlük hayatta sık kullanılan ve genelde içerik belirtmeyen terimleri ayıklamayı, hem de sistemi gereksiz gürültüden arındırmayı hedeflemiştir. Daha sonra her bir cümle için puanı terim sıklıklarına göre hesaplanarak yüksek puanlı cümleler özete eklenmiştir. Alanda yol gösterici bir başka çalışmada (Edmundson, 1969) kelime sıklığı bilgisine ek olarak “ipucu sözcük öbekleri”, “başlık terimleri” ve “cümle konumu” gibi özellikler de hesaplanmıştır. Bu çalışma ile çoklu özellikler önemli cümle seçim işlemi için doğrusal bir fonksiyonun parametreleri olarak kullanmaya başlanmıştır. Aşağıdaki başlıklarda (Lloret ve Palomar, 2012, s. 6) çıkarıma dayalı otomatik özetleme sistemlerinde ele alınmakta olan belli başlı özellikler dâhil oldukları yaklaşımlara göre listelenmektedir:

### i. İstatistik Temelli Yaklaşımlar

Bu yaklaşımlar, kaynak metindeki önemli cümleleri puanlamak için istatistiksel hesaplamalar kullanırlar. Sık kullanılan istatistiksel özelliklerden bazıları, terim sıklığı (tf), ters terim sıklığı (idf), cümlelerin konumu, pozitif anahtar kelime, negatif anahtar kelime, cümlelerin merkeziliği (diğer cümlelerle benzerliği), cümlelerin başlığa benzerliği, cümlelerin göreceli uzunluğu, cümledeki sayısal verilerin varlığı, cümledeki isim soylu kelimelerin varlığı olarak sıralanabilir.

İstatistik temelli yaklaşımların kullandığı teknikler dilden bağımsız oldukları için herhangi bir ilave dil bilgisi veya karmaşık dil işleme süreci gerektirmezler. Bu sebeple daha az işlemci ve bellek kapasitesine ihtiyaç duyarlar (Gambhir ve Gupta, 2017, s. 5).

ii. Konu Temelli Yaklaşımlar

Konu temelli yaklaşımların ana fikri ana kavramları ve alt kavramları bulmaktır. SVD (Singular Value Decomposition) hesaplamalarından çıkarılan sonuç kavramları, dokümanın konuları olarak kabul edilir. Ancak, bu konular başka ayıklanmış konuların alt başlıkları olabilir. Bu yaklaşımda, bir grup alt konu olabilen başlıca konulara karar verdikten sonra, cümleler ana konuların özetinin bir parçası olarak toplanır (Özsoy, Çiçekli ve Alpaslan, 2010, s. 870).

iii. Grafik Temelli Yaklaşımlar

Grafik tabanlı yaklaşımlarda cümle veya sözcük, anlamsal olarak ilgili metin unsurlarını birbirine bağlayan düğümler ve kenarlar ile temsil edilir. Bu yaklaşımda iki cümle arasındaki benzerlik bulunur ve benzerlik belirli bir sınırın üstünde olursa, cümle arasındaki bağlantı da dikkate alınır. Bağlantı yapıldıktan sonra, grafik üzerinde rastgele yürüyüş yöntemine göre, önemli cümleler seçilir (Gambhir ve Gupta, 2017, s. 6).

iv. Söylem Temelli Yaklaşımlar (Discourse Based)

Bu yaklaşım, otomatik metin özetleme için dilbilimsel teknikler kullanarak metindeki söylem ilişkilerini keşfeder. Söylem ilişkileri ise bir metin içindeki cümleler ve bölümler arasındaki bağlantıları temsil eder (Gambhir ve Gupta, 2017, s. 6-7). Cümleler bağlantılarına göre ağırlıklandırılır. İlgili cümleler yüksek, ilgisizler düşük ağırlığa sahip olurlar.

v. Makine Öğrenmesi Temelli Yaklaşımlar

Makine öğrenme yaklaşımı denetimli, denetimsiz veya yarı denetimli olabilir. Cümlelerin önemini belirten özellikler, Naive-Bayes Sınıflandırıcı, Karar Ağaçları, Saklı Markov Modeli, Log-linear Modelleri, Sinir Ağları ve Bayes Sınıflandırıcı gibi tekniklerle birlikte kullanılır (Lloret ve Palomar, 2012, s. 9).

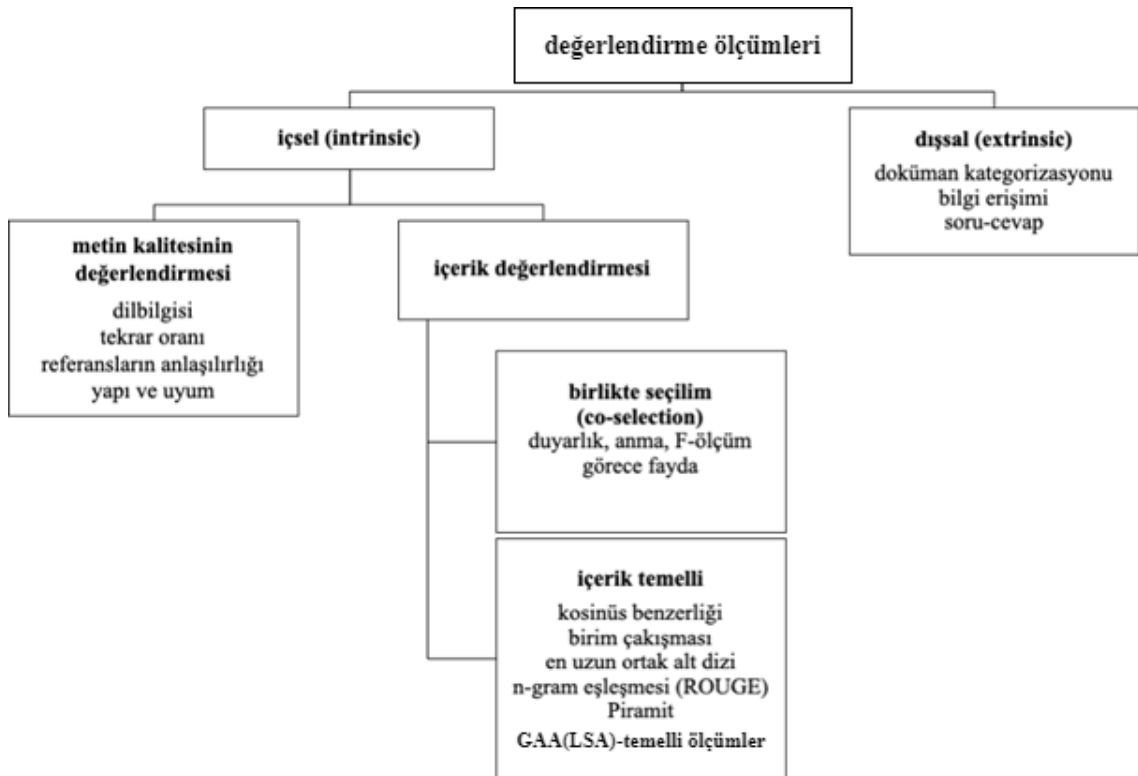
Bir otomatik özetleme sistemi işleyişine insan faktörünü ekleme derecesine göre denetimli, denetimsiz ya da yarı denetimli olarak sınıflandırılabilir. Denetimli sistemler,

cümlelerinin kullanışlı özelliklerini öğrenmek için insan tarafından üretilen özetleri de kullanılır. Bu tür sistemlerde insan tarafından üretilen özetlerden eğitim verileri olarak yararlanılır. Denetimsiz sistemler eğitim verileri içermez ve yalnızca hedef dokümanları işleyerek özet oluşturur. Yarı denetimli öğrenme sistemleri, dokümanlara uygun fonksiyon veya sınıflandırıcı üretmek için etiketli (insan tarafından önem derecesine göre işaretlenmiş cümle) ve etiketsiz verilerin hepsini kullanır (Gambhir ve Gupta, 2017, s. 7).

## **2.6. OTOMATİK ÖZETLEME SİSTEMLERİNİN DEĞERLENDİRİLMESİ**

Otomatik metin özetleme çok çalışılan ve teknolojik gelişmelere paralel olarak hızlı ilerlemeler kaydedilen bir alan olsa da, özellikle değerlendirme süreci hala birçok zorluğun olduğu açık bir araştırma konusudur (Saggion ve Poibeau, 2013, s. 15; Torres-Moreno, 2014, s. 50).

Şekil 5'te verilen yaygın kullanılan değerlendirme ölçümleri taksonomisinde de görüldüğü gibi, otomatik özetleme sistemlerinin karşılaştırılmasında kullanılan temel yaklaşımların genel olarak içsel değerlendirme ve dışsal değerlendirme olmak üzere, iki yaygın değerlendirme türünde sınıflandırıldığı görülmektedir (Torres-Moreno, 2014, s. 243). Genel çerçevesiyle içsel değerlendirme özetin kendisi hakkındadır ve özetin analizine dayanır (Steinberger ve Jezek, 2009, s. 1002). Üretilen otomatik özetin insan tarafından okunarak değerlendirilmesi (elle değerlendirme), tamamen otomatik ya da hem elle hem de otomatik yaklaşımları bir arada kullanan yarı otomatik yaklaşımlar kullanarak değerlendirme yapılması içsel değerlendirme çeşitleridir (Torres-Moreno, 2014, s. 243). İç değerlendirmeler içerik değerlendirmesi ve metin kalitesini değerlendirme üzerinde kurulabilir (Steinberger ve Jezek, 2009, s. 1002). İçerik değerlendirmeleri çıktılarının tam metinlerdeki temel konuları belirleme başarımlarını ölçerken, metin kalitesi değerlendirmeleri otomatik özlerin okunabilirliğini, dilbilgisi kurallarına uyumunu ve tutarlılığını ölçer. Dışsal değerlendirmede ise özet kalitesi, özetlerin belirli bir görev için ne kadar yararlı olduğuna göre değerlendirilir (Steinberger ve Jezek, 2009, s. 1002).



Şekil 5. Özet değerlendirme ölçümlerinin taksonomisi (Steinberger ve Jezek, 2009, s. 1006)

Şekil 5’te verilen her bir değerlendirme ölçümünün uygulanma sürecinin kendine has zorlukları ve üstün yönleri mevcuttur. Üretilen özetin değerlendirmesinin bire bir kullanıcısı tarafından yapılmasının otomatik süreçlere göre daha güvenilir sonuçlar vereceği açıktır. Bu tür elle değerlendirme süreçleri için ilk bakışta otomatik değerlendirmelere göre daha kolay uygulanabilir oldukları da düşünülebilir. Ancak, derlemin ve üretilen özetlerin okunması yorucu ve zaman alıcı bir değerlendirme sürecine sebep olmaktadır. Geçerli bir değerlendirme için uygun sayıda değerlendirici bulmak, elle değerlendirmenin bir başka zorluğudur (Torres-Moreno, 2014, s. 246). Diğer yandan bir doküman ya da doküman topluluğu için ideal özet nedir sorusunun cevabı son derece öznel ve geniş bir yelpazede değişkenlik gösterir (Torres-Moreno, 2014, s. 243). Bu sebeple insanların özet değerlendirmede bir fikir birliğine ulaşmaları da çoğunlukla beklenmez (Saggion ve Poibeau, 2013, s. 15). Ancak elle değerlendirme ile bir özet metninin kalitesini belirleyen, okunurluk, kullanılan dilin düzeyi ve üretilen özette ana metnin ana kavramlarının yer alıp almaması gibi pek çok önemli unsur bütüncül bir bakış açısıyla dikkatlice değerlendirilmiş olur (Torres-Moreno, 2014, s. 246).

Otomatik değerlendirme yaparken de insan değerlendirenden tümüyle bağımsız çalışmak her zaman mümkün olmayabilir. Yarı otomatik yöntemlerle değerlendirme yaparken, sistem özeti insan tarafından üretilmiş bir referans metin ile karşılaştırılması gerekmektedir. Akademik makalelerin özetlenmesi söz konusu olduğunda, tekli dokümandan elde edilen bir otomatik özeti yazar tarafından hazırlanmış öz ile karşılaştırılmasında otomatik içerik değerlendirme yöntemlerinden faydalanmak mümkündür. ROUGE (Lin, 2004), DUC<sup>1</sup> (The Document Understanding Conference) konferanslarından itibaren standart bir otomatik değerlendirme yöntemi olarak günümüzde de kullanılmaktadır. Kullanılan derlemde yazar tarafından üretilmiş özetler de mevcut ise; insan değerlendirmeleri ile yüksek korelasyon değeri veren ROUGE ailesini kullanmak tavsiye edilmektedir (Steinberger ve Jezek, 2009, s. 1022). Bu bağlamda sistem özetleri ile insan tarafından oluşturulan ideal özetler arasındaki karşılaştırma, n-gram, kelime dizileri ve kelime çiftleri gibi birimlerin üst üste gelme sayılarını hesaplar (Lin, 2004, s.74).

Bu çalışma kapsamında geliştirilen sistem çıktıları değerlendirilirken içsel değerlendirme ölçümleri kullanılarak metin kalitesi temelli ve içerik temelli değerlendirmeler yapılmıştır. Bu çalışmada çıktıların içerik değerlendirmeleri hız ve uygulama kolaylığı sebebiyle, ROUGE ölçümlerinin Java ile derlenmiş sürümü olan ROUGE 2.0 paketi kullanılarak gerçekleştirilmiştir. Yazar tarafından üretilmiş özetlerin standart öz olarak değerlendirme aşamasında kullanılması planlanmıştır. İlerleyen bölümlerde metin kalitesi ve içerik değerlendirmesi ölçümleri çalışmamız bağlamında ele alınmaktadır.

### **2.6.1. Metin Kalitesinin Değerlendirilmesi**

Bu çalışma kapsamında metin kalitesinin değerlendirmesinde çıktı özlerinin sayısal okunurluk değerleri hesaplanarak tam metin ve orijinal özetler ile karşılaştırılmıştır. Ayrıca ilgili tam metinlerin yapısal formatlarından faydalanılarak her bir makaleden elde edilen çıktıların tam metinleri tarama başarımları değerlendirilmiştir. Tez çalışması kapsamında

---

<sup>1</sup> DUC, 2001-2008 yılları arasında NIST (National Institute of Standards and Technology) tarafından yürütülmüştür. Amacı, araştırmacıların özetleme sistemlerinin hem geliştirilme hem de değerlendirilme aşamalarında geniş çaplı deneylerde yer almasını sağlamaktır. Düzenlenmiş bir dizi özetleme değerlendirmesi ile otomatik metin özetlemesinde daha fazla ilerleme kaydedilmesine çalışılmıştır.



makalelerin çözümlenmesinde kullanılan yapısal bölümler için Şekil 2'deki diyagram incelenebilir.

## 2.6.2. İçerik Değerlendirilmesi

Çalışmada kullanılan içerik değerlendirmesi ölçümleri bu bölümde kısaca ele alınmaktadır.

### 2.6.2.1. Birlikte Seçilim Ölçümleri

Birlikte seçilim değerlendirmesinde kullanılan temel ölçümler, duyarlık (precision), anma (recall) ve f-ölçümdür (f-score). Bunlardan duyarlık değeri ideal ve sistem tarafından üretilen özetlerde birlikte yer alan cümlelerin toplam sayısının sistem özetinin cümle sayısına bölümü ile elde edilir. Anma değeri ise ideal ve sistem özetlerinde birlikte yer alan cümle sayısının ideal özetin toplam cümle sayısına bölümü ile hesaplanmaktadır. f-ölçüm değeri duyarlık ve anma değerlerinin birleşimi ile elde edilir. En basit elde edilmiş şekli ise bu iki değerlerin harmonik ortalamasının hesaplanması ile olur (Mani vd., 1999, s.79).

### 2.6.2.2. İçerik Temelli Ölçümler

İçerik temelli benzerlik ölçümleri, argüman olarak karşılaştırılmak istenen iki metin gösterimini alır ve aralarındaki benzerliği 0 ile 1 arasında sayısal bir değerle ifade eder. Burada 0 değeri iki metnin tamamen farklı olduğu, 1 ise birbiri ile yakından ilişkili olduğu anlamına gelir (Saggion, Radev, Teufel, Lam ve Strassel, 2002. s. 752).

#### 2.6.2.2.1. Kosinüs Benzerliği

İçerik benzerliğini ölçmede kullanılan en temel ölçüm kosinüs benzerliğidir. Kosinüs benzerliği, iki metni karşılaştırarak aralarındaki benzerliğe karar vermede en çok tercih edilen yöntemlerden biridir.  $X = (x_1, x_2, \dots, x_p)$  ve  $Y = (y_1, y_2, \dots, y_p)$  karşılaştırılacak iki

cümlenin vektör gösterimleri olsun. X'in Öklid normu,  $\|X\| = \sqrt{x_1^2 + x_2^2 + \dots + x_p^2}$  ve X ile Y'nin vektörel çarpımı,  $XY = x_1y_1 + x_2y_2 + \dots + x_py_p$  ile tanımlanmak üzere iki vektör arasındaki  $\theta$  açısının kosinüs değeri bu iki vektörün temsil ettiği iki cümlenin benzerlik değerini verir ve Formül 1'deki gibi hesaplanır (Salton ve Buckley, 1988. s. 514).

$$\text{benzerlik}(X,Y) = \cos(\theta) = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \quad (1)$$

#### 2.6.2.2.2. Birim Çakışması

X ve Y iki farklı metin parçasının küme gösterimini temsil etmek üzere, aralarındaki çakışma Formül 2'deki gibi hesaplanır (Saggion vd. 2002. s. 752).  $\|X\|$  ilgili kümenin boyutunu ifade etmektedir.

$$\text{çakışma}(X,Y) = \frac{\|X \cap Y\|}{\|X\| + \|Y\| - \|X \cap Y\|} \quad (2)$$

#### 2.6.2.2.3. En Uzun Ortak Alt Dizi

X ve Y iki farklı metin parçasının dizi gösterimi olmak üzere,  $\text{lcs}(X,Y)$  değeri X ve Y arasındaki en uzun ortak alt dizinin (longest common subsequence) uzunluğudur ve Formül 3 ile hesaplanır (Saggion vd. 2002. s. 753).

$$\text{lcs}(X,Y) = \frac{\text{uzunluk}(X) + \text{uzunluk}(Y) - \text{edit}_{di}(X,Y)}{2} \quad (3)$$

Uzunluk değerleri ilgili metinlerin uzunluk değeri,  $\text{edit}_{di}(X,Y)$  X'in Y'ye dönüşmesi için gereken minimum silme ve ekleme işleminin sayısıdır (Crochemore ve Rytter, 1994, s.

188). İki metni karşılaştırırken, iki metnin cümleleri arasında normalleştirilmiş bir çift hesaplanmaktadır. Kosinüs benzerliği ve birim çakışmasının aksine, en uzun ortak dizi, metinde bilginin nasıl sıralandığına duyarlıdır (Saggion vd. 2002. s. 753).

#### 2.6.2.2.4. ROUGE-N: N-gram Birliktelik İstatistikleri

n-gramlar, n ile ifade edilen uzunlukta sıralı terimin birliktelik istatistiğini bulmada kullanılan terim dizgeleridir. n-gram kullanarak otomatik özetlerin değerlendirilmesi sürecinde aday özet ile referans özet(ler)i arasındaki anma değeri her bir ROUGE-N için Formül 4'te verildiği gibi hesaplanır.

$$\text{ROUGE-N} = \frac{\sum_{S \in \{\text{Referans Özetler}\}} \sum_{\text{gram}_n \in S} \text{Say}_{\text{eşleşme}}(\text{gram}_n)}{\sum_{S \in \{\text{Referans Özetler}\}} \sum_{\text{gram}_n \in S} \text{Say}(\text{gram}_n)} \quad (4)$$

Burada n, n-gramların uzunluğu olmak üzere;  $\text{gram}_n$ ,  $\text{Say}_{\text{eşleşme}}(\text{gram}_n)$  referans ve aday özetlerde görülen maksimum n-gram birlikteliği sayılarıdır (Lin, 2004, s.74).

Örneğin “Polis kaçan hırsız yakaladı” cümlesinin kök terimlerinin n-gramları aşağıdaki gibi elde edilir:

n=1 için 1-gram listesi : ‘polis’, ‘kaç’, ‘hırsız’, ‘yakala’

n=2 için 2-gram listesi: ‘polis kaç’, ‘kaç hırsız’, ‘hırsız yakala’

n=3 için 3-gram listesi: ‘polis kaç hırsız’, ‘kaç hırsız yakala’

Bu ilk cümlenin kelimelerini anlamı değiştirecek şekilde farklı bir sırada veren, “Kaçan hırsız polisi yakaladı.” cümlesi için n-gramlara bakacak olursak:

1-gramlar: ‘kaç’, ‘hırsız’, ‘polis’, ‘yakala’

2-gramlar: ‘kaç hırsız’, ‘hırsız polis’, ‘polis yakala’

3-gramlar: ‘kaç hırsız polis’, ‘hırsız polis yakala’

dizgeleri elde edilir. İlk ve ikinci cümlelerin n-gram birlikteliklerini toplam eşleşme sayıları bağlamında incelediğimizde, n=1 için ilk cümle ile ikinci cümle tamamen örtüşüyor olarak hesaplanmaktadır. n değerini büyütürsek cümlelerin birliktelik

toplamlarına bakmaya devam ettiğimizde,  $n=2$  için sadece ‘kaç hırsız’ birlikteliğine rastlanmaktadır. Yani, iki cümle için 2-gram birlikteliklerine göre örtüşme oranı  $1/3$ 'e düşmektedir.  $n=3$  için ise iki cümlede birlikte yer alan 3-gram mevcut değildir. Yani 3-gram örtüşmelerine göre iki cümle tamamen farklıdır.

Cümle bazında daha güvenilir sonuçlar için değerlendirme aşamasında hesaplayacağımız iki ROUGE değeri ROUGE-L ve ROUGE-SU4'tür. ROUGE-L, en uzun ortak alt dizi ölçümünü verir. ROUGE-L'in zayıf noktası da metindeki ana kelime dizisini yakalayıp daha kısa diziler oluşturan yan konuları atlayabilmesidir (Lin, 2004, s. 76). Cümle sıralamasında herhangi bir kelime çiftini keyfi boşluklara izin vererek değerlendirmeye alan ROUGE-SU4 ise en çok 4 tane 1-gram atlayarak oluşturulan 2-gram birlikteliği ölçmektedir. (Lin, 2004, s. 78).

Akademik makaleler için otomatik özetleme çalışmaları yapılırken yazar tarafından hazırlanmış özlemlerin değerlendirme sürecinde referans özet olarak kullanılması yaygın bir yaklaşımdır. İnsan tarafından özet hazırlama sürecinin oldukça göreceli bir konu olduğu bilinmektedir. Bu bakış açısıyla, bu çalışma kapsamında da yazarın kendi ürettiği özü referans/ideal özet olarak kullanmanın akla yatkın bir uygulama olduğu değerlendirilmiştir. Bu doğrultuda, geliştirilen sistemin çıktısı yapısal olan ve olmayan özlemler ile makalelerin yazarları tarafından üretilmiş özlemlerin  $n$ -gram çakışma oranları ROUGE ölçümlerine göre karşılaştırılmaktadır. Bu karşılaştırma, ROUGE-1, ROUGE-2, ROUGE-L ve ROUGE-SU4 ölçümlerine göre, ortalama anlama, ortalama duyarlılık ve ortalama f-ölçüm değerleri kullanılarak elde edilmektedir.

## 2.7. TÜRKÇE OTOMATİK METİN ÖZETLEME ÇALIŞMALARI

Çeşitli alanlarda ve dillerde otomatik metin özetleme, 1950'lerden bugüne çok sayıda yaklaşım ve teknik geliştirilen ve geliştirilmeye devam edilen sıcak bir çalışma alanıdır. Türkçe otomatik metin özetleme çalışmaları ise kısıtlı sayıdadır. Bu çalışmalar arasında dokümanların yapısal özelliklerini de ele alan çalışmalar oldukça sınırlıdır. Yapısal özellikler ile kullanıcıların bilgi isteklerini birleştiren özgün bir özetleme yaklaşımında,

cümle seçiminde cümle puanlama ile birlikte bölüm puanlama da yapılmıştır (Pembe, 2011). Araştırmada Internet aramaları için kullanıcı ihtiyaçlarına özel özet oluşturmak hedeflenmiştir. Derlem 500 doküman içermektedir. İlk aşamada Internet dokümanlarının bölüm ve alt bölüm hiyerarşilerinin ilgili başlık ve alt başlıklarla birlikte ortaya çıkarılması için yapısal olarak işaretleme yapılmıştır. İkinci aşamada ise, puanlamalar ve cümle seçimi gerçekleştirilmiştir. Başlık (başlıkta geçen terimlerin cümlede geçme sıklıkları), konum (herhangi bir alt bölümün ilk cümlesi ise 1 değilse 0), terim sıklığı (cümlede geçen terimlerin tüm dokümanda geçme sıklığından elde edilen değer) ve sorgu cümlesi (sorguda geçen terimlerin cümlede geçme sıklıklarının toplamı) metotlarını kullanarak cümleleri puanladıktan sonra, dokümanların her bir bölümüne cümlelerin önemine göre puan verilmiş ve cümle seçimine geçilmiştir. Elde edilen sonuçlara göre önerilen sistemin karar süresi ve doğruluk başarımlarının Google özet çıkarımından daha iyi olduğu kullanıcı değerlendirmeleri ile doğrulanmıştır.

Türkçe metinler üzerinde yapılan ilk otomatik metin özetleme araştırmasında ise, ekonomi alanından seçilmiş 50 Türkçe haber makalesinden oluşan bir derlem üzerinde çalışılmıştır (Altan, 2004). Tekli doküman özetleyen bu çalışmada, cümleler istatistiksel yöntemler kullanılarak (terim sıklığı bilgisi ve cümlelerin konum bilgisi gibi özelliklere göre) puanlanma gerçekleştirilmiştir. HTML formunda işlenen dokümanlar, HTML etiketleri yardımıyla başlık, cümle ve paragraflarına ayrıldıktan sonra, başlık terimleri incelenmiş, pozitif ve negatif cümle analizleri yapılmıştır. Çalışma, temelde istatistiğe dayanıyor olsa da cümle seçiminde kelimelerin negatif, pozitif kelime listelerini de kullanarak anlamsal bütünlüğü vurgulamanın yollarını araştırmıştır.

Türkçe derlem kullanılan bir başka çalışmada (Kutlu, Cıgır ve Çiçekli, 2010) kelime sıklığı, başlığa benzerlik, anahtar söz öbekleri (key phrases), cümle konumu, merkezilik özellikleri ele alınmıştır. Bu özelliklerin ağırlıkları makine öğrenmesi metotları ile puanlanmıştır. Derlem, 120 haber dokümanı ve 100 Türkçe bilimsel yayından oluşmuştur. Söz konusu çalışmada haber dokümanları için %40, akademik yayınlar için %5 sıkıştırma oranı ile özet oluşturulmuştur. Kullandıkları yöntemlerin değerlendirmesi sonucunda cümle konumu ve merkezilik özelliklerinin sonucu en çok etkileyen iki özellik olduğunu belirtmişlerdir (Kutlu vd., 2010, s. 1321). 50'şer metinden oluşan iki Türkçe

veri seti üzerinde gizli anlamsal analiz tekniđi denenmiřtir. Arařtırmada önerilen iki yeni yaklařımdan ‘‘Çapraz’’ (Cross) yaklařımının ‘‘Konu’’ (Topic) yaklařımına göre daha iyi sonuç verdiđi, girdi matrislerinin oluřturulma metotlarının sonuçlara etkisinin çok az olduđu saptanmıřtır. Bir bařka ifadeyle, dokümanları ayırıřtırırken cümlelerin kelimeleri ierme durumlarını, binary (0/1; var/yok), kelimenin görölme sıklıđı ya da tf/idf uygulanarak elde edilmiř sıklıklarla kullanmanın sonuçlara etkisinin ihmal edilebilecek düzeyde olduđu anlařılmıřtır (Özsoy vd., 2010, s. 826).

Bir dokümandaki en önemli cümleleri seçerek ilgili dokümanın özetini çıkartmayı hedefleyen bir bařka alıřmada (Güran, Arslan, Kılı ve Diri, 2014) 20 Türke haber dokümanı üzerinde ıkarıma dayalı 15 farklı cümle seçim yönteminin tekli ve birlikte kullanım performansları deđerlendirilmiřtir. Bu yöntemlerin ıktıları 30 deđerlendirici tarafından elle seçilen cümlelerden oluřan özetlerin bir deđerlendirme veri seti üzerinde analiz edilmiřtir. Sonuç olarak cümle konumu, ortak yakınlık sayısı ve isim soylu kelimeleri ierme durumu özelliklerinin üçlü kullanımını en iyi sonuçları vermiřtir (Güran vd, 2014, s. 195). Özellikler dođrusal bir fonksiyonda bir araya getirilirken ađırlıkları eřit tutulmuřtur. alıřma kapsamında deđerlendiricilerin cümle seçiminde en çok, sık rastlanan kelimeleri ieren cümleleri tercih ettikleri belirtilmiřtir (Güran vd, 2014, s. 194).

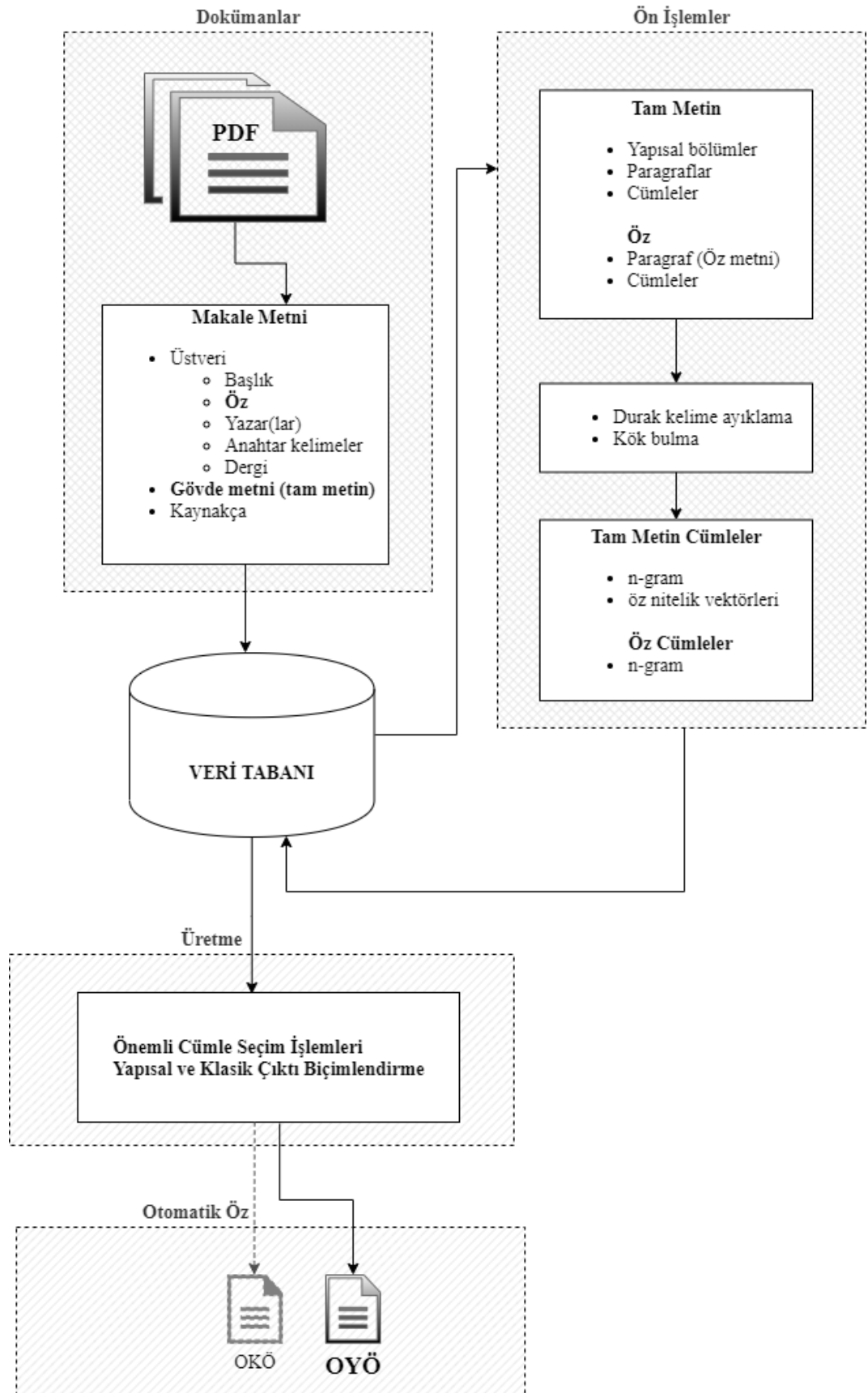
Bir önceki alıřmada (Güran vd, 2014) kullanılan derleme, 130 doküman ve elle hazırlanmıř ıkarıma dayalı özetlerinin de eklendiđi daha büyük bir derlem üzerinde ‘‘Fuzzy Analytical Hierarchy Process’’ tabanlı bir cümle skortlama fonksiyonu ile alıřma kapsamında geliřtirilen bir diđer genetik algoritma uygulanmıřtır (Güran, Uysal, Ekinci ve Güran, 2017). ıkarıma dayalı bir otomatik özetleme sistemi ile yapılan analizler sonucunda 15 farklı cümle seçim metodu kullanılmıřtır. Sonuç olarak, FAHP tabanlı sistemin genetik algoritma uygulayan sisteme göre daha üstün olduđu saptanmıřtır (Güran vd, 2017, s. 67). Ayrıca deđerlendiricilerin en çok ‘‘terim sıklıđı’’, ‘‘cümlelerin konumu’’, ‘‘cümle uzunluđu’’ özelliklerini kullanmayı tercih ettikleri de bu alıřmanın bulguları arasındadır (Güran vd, 2017, s. 64).

### 3. BÖLÜM: YÖNTEM

Bu bölümde ilk olarak başlangıçlarından 2015 yılına kadar *Türk Kütüphaneciliği* ve *Bilgi Dünyası* dergilerinin hakemli makalelerini içeren derlemin metin ön işlemlerinden geçirilerek kelime sıklık vektörlerine kadar indirgenmesi süreci ele alınmaktadır. Otomatik cümle seçim işlemleri ve okunurluk hesaplamaları hakkında bilgi verildikten sonra, bu çalışma için tasarlanan, otomatik yapısal özetleme yöntemi ve geliştirilen sisteminin değerlendirme süreci açıklanmaktadır. Çalışma kapsamında geliştirilen otomatik özetleme sistemi mimarisi Şekil 6'da gösterilmektedir. Geliştirilen sistemde kullanılan metin verisinin elde edilip işlenerek veri tabanına kaydedilmesi süreci 115K440 Proje numaralı TÜBİTAK SOBAG Projesi (Al ve Sezen, 2018) ile paralel yürütülmüştür. Tam metinlerin yapısal başlıklarına göre işaretlenmesi sürecinde projedeki uzman bursiyerler görev yapmıştır. Uzman bursiyerlerden üç tanesi, akademik makale yazım süreçlerine aşina, editörlük görevi yapan, Bilgi ve Belge Yöntemi Bölümü doktora öğrencisi araştırma görevlileridir. Derlem makalelerinin yapısal bölümlerine göre işaretlenmesi sürecinde bu üç uzman bursiyer ağırlıklı olarak görev almıştır. Derlemi oluşturan veri setine proje web sayfasından erişilebilir (<http://115k440.hacettepe.edu.tr>).

#### 3.1. DERLEM VE ÖZELLİKLERİ

Çalışma kapsamında geliştirilen otomatik yapısal özetleme sistemi, *Bilgi Dünyası* ve *Türk Kütüphaneciliği* dergilerinin tüm hakemli sayılarında yayımlanmış makaleler için yapısal özetler üretmek üzere tasarlanmıştır. Bu bağlamda 2000-2015 yılları arasında yayımlanmış *Bilgi Dünyası* ve 1996-2015 yılları arasında yayımlanmış *Türk Kütüphaneciliği* dergilerinin sayıları işlenerek, *Bilgi Dünyası* 'nda yayımlanmış 172, *Türk Kütüphaneciliği* 'nde yayımlanmış 249 olmak üzere toplam 421 makale yapılandırılmış veriye dönüştürülmüş ve veri tabanına kaydedilmiştir. *Bilgi Dünyası* 'nda yayımlanmış 23, *Türk Kütüphaneciliği* 'nde yayımlanmış 6 İngilizce makale araştırmanın Türkçe ile sınırlı olması sebebiyle derleme dâhil edilmemiştir. *Türk Kütüphaneciliği* 'nde yayımlanmış yazar özü bulunmayan iki makale de çıktı değerlendirmelerinde orijinal özetlerle karşılaştırma yapılacağı için derlem kapsamına alınmamıştır. Her iki dergi de açık erişimli olduğu için dergi sayılarına ve makalelere erişim sorunsuz gerçekleşmiştir.



Şekil 6. Otomatik yapısal özetleme sistemi mimarisi



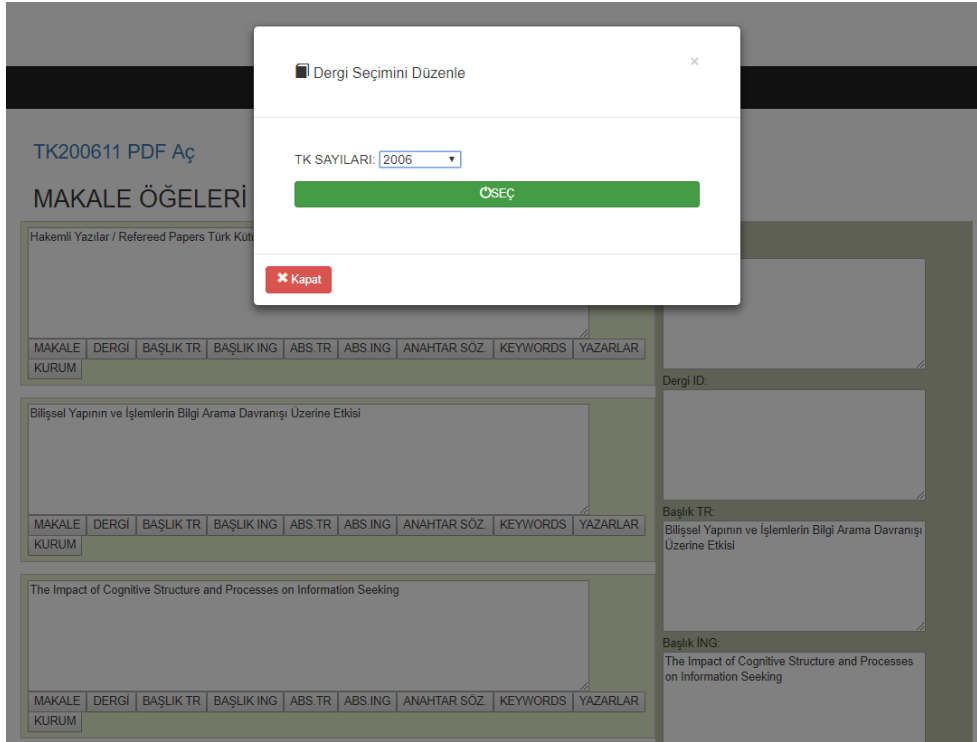
### 3.2. METİN ÖN İŞLEMLERİ

Metin ön işlemleri metin madenciliği sürecinin en önemli ve en çok zaman alan aşamasıdır. Standart bir formata dönüştürülmüş metin girdisi elde edildikten sonra sistem mimarisine eklenecek eklentiler özetleme sistemi tarafından kolaylıkla desteklenecektir (Hovy ve Lin, 1999, s. 84). Çalışma kapsamında otomatik yürütülen metin ön işlemlerine ek olarak bursiyer bir işaretçi ekip tarafından makalelerin yapısal özelliklerine göre işaretlemeler de yapılmıştır. Elde edilen işlenmiş metin verisi ilişkisel veri tabanında tutulmuştur ve bu çalışmanın dışındaki metin analizlerinde de girdi olarak kullanılmaya uygundur.

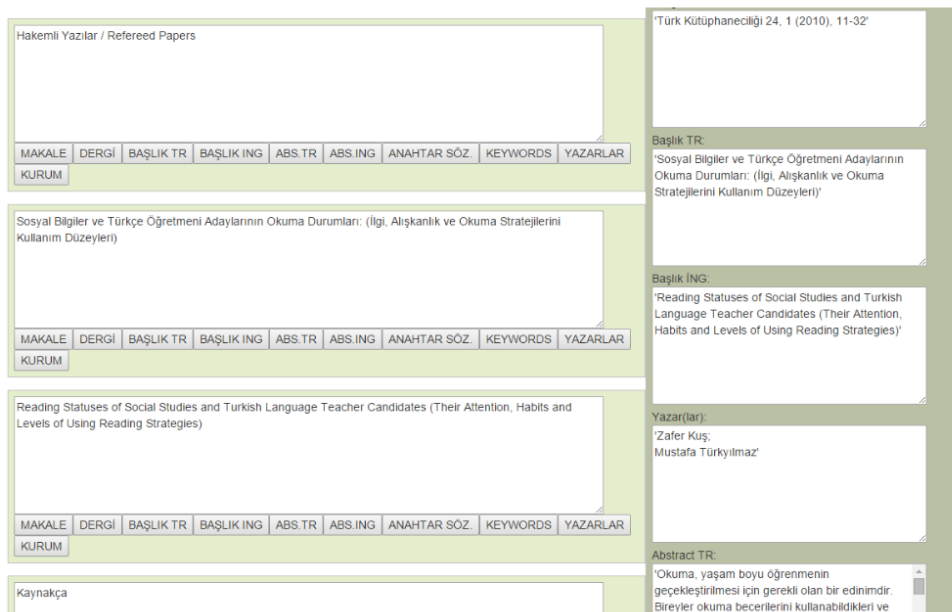
Araştırma kapsamında iki aşamalı planlanan metin işleme çalışmaları için web tabanlı arayüzler geliştirilmiştir (bkz. <http://115k440.hacettepe.edu.tr>). Elle yapılacak metin işleme çalışmaları için TÜBİTAK 115K440 numaralı “Türkçe atıflar için içerik tabanlı atıf analizi modeli tasarımı” başlıklı TÜBİTAK SOBAG Projesi bursiyerlerinin desteği sağlanmıştır. Bursiyerler akademik bir yayının yapısal düzeni hakkında ön bilgi sahibi, Bilgi ve Belge Yöntemi Bölümünden 2 lisans, 4 doktora öğrencisi 6 kişilik uzman bir ekiptir. 4 doktora öğrencisi bursiyerlerden biri yazılım ve veri tabanı sürecini üstlenirken diğerleri veri işaretlemelerini gerçekleştirmiştir. Doktora öğrencilerinden biri ayrıca, bu iki sürecin uyumlu ilerlemesi ve veri işaretlemesinin uyumunu gözetmiştir. Proje kapsamında, ekibe yapılacak işlemlere dair ayrıca eğitim desteği, her bir bursiyerin yeteneklerine göre kendi kaynaklarından sağlanmıştır.

Metin ön işlemlerine hazırlık aşamasında, PDF dosyaları halinde arşivlenmiş makaleler indirilmiş ve tek tek düz metin formatına dönüştürülmüştür. Bu aşamada sıklıkla karşılaşılan karakter seti problemlerine çözüm olabilen çeşitli doküman çeviriciler kullanılmıştır. Dokümanlar dönüştürüldükten sonra metin işleme ve yapılandırma işlemlerine başlanabilmektedir. Metin ön işlemlerinin ilk aşamasında, her bir dokümanın tam metni Şekil 7 ve Şekil 8’de örnekleri görülen web arayüzünü kullanan bursiyer ekibin gözetiminde Şekil 2’de verilmiş yapısal organizasyon çerçevesinde, başlık, öz, anahtar kelimeler, kaynakça ve gövdelerine yarı otomatik bir şekilde ayrılarak işaretlenmiştir. Bu

arayüz seçilen makalenin üstverilerini anlık olarak otomatik ayrıştırıp bursiyerin düzenlemesi ya da onaylaması için web ekranına getirmektedir.



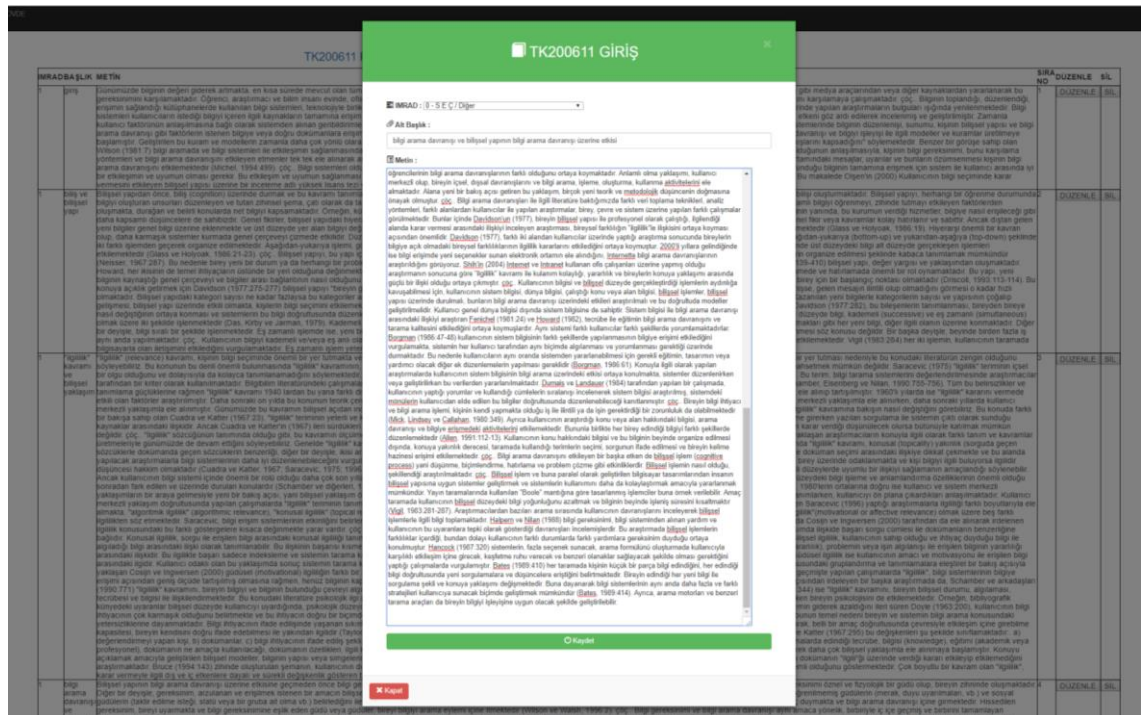
Şekil 7. Makale seçim arayüzü



Şekil 8. Makale üstveri ayrıştırma arayüzü

İkinci aşamada metin ön işleme için geliştirilen arayüzü yardımı ile gövde metinlerinin IMRAD'a göre işaretlenmesinin uzman bursiyer eliyle yapılması sağlanmıştır. Bu aşamada uzman ekipten bir kişi bursiyer ekibin makale tam metinlerine uygulamaları ön işlemlerde verdikleri kararların uzlaşma düzeylerini izleyebileceği bir panel kullanmıştır. Tam metin bölümlerinin içeriklerinin IMRAD başlıklarına uyumlarına göre yapılan işleme işlemlerinin sonucunda veri tabanına kaydedilen kararlar ekip içinde uzlaşma sağlananlardır.

Bu aşamanın sonunda gövde metni paragraf biriminde ele alınarak Tablo 3'teki yapısal bölüm başlıklarına ayrıştırılmıştır. Seçilen bir makalenin gövde metninden her bir paragraf Şekil 9'daki arayüz ile yazarın belirlediği bölüm başlığının yanı sıra IMRAD bölümlerine göre de adlandırılarak veri tabanına kaydedilmiştir.



Şekil 9. Makale gövde metni paragraf ayrıştırma/IMRAD işleme arayüzü

Otomatik özetleme işlemlerinde Tablo 3'teki ilk dört yapısal bölüm altındaki gövde metinleri tam metin olarak kullanılmıştır.

Tablo 3. Çalışmada kullanılan yapısal bölümler

Yapısal Bölüm No	Yapısal Bölümler
1	Giriş
2	Yöntem
3	Bulgular
4	Tartışma ve Sonuçlar
5	Teşekkür
6	Ekler
7	Dipnot
8	Dipnot Şeklinde Teşekkür

Bu iki aşamadan sonra her bir makalenin tam metni veri tabanında kaydedilmek üzere cümlelerine otomatik olarak ayrıştırılmıştır. Bu sırada kullanılan basit kurallar şu şekilde sıralanabilir:

- 1- Her cümle “.” ya da “!”, “?” gibi cümle sonuna konan ve cümlenin bittiğini işaret eden noktalama işaretleri ile son bulur. Ancak bazen cümle sonu olmadığı halde bu işaretlerden kullanılması gerekebilir. Örneğin: “Prof. Dr.”, “XII. yy” vb. Bu durumlar için bir saklı sözcükler kütüphanesi oluşturulmuştur. Burada “yy., Prof. Dr., bkz.” vb. kısaltmalarla birlikte “2.” ya da “XI.” rakam gösterimleri kural dışı tutulmuştur.
- 2- Her bir cümle büyük harfle başlar. Ancak her büyük harfle başlayan kelime cümle başı değildir. Örneğin: “Bugün Ankara’da hava sıcaklığı mevsim normalleri üzerindedir.” cümlesindeki özel isim olan Ankara’nın büyük harfle başlaması gibi. Bu durumlar için varsa bir önceki kelimedenden sonra gelen noktalama işaretinin fonksiyonuna göre cümlenin başlayıp başlamadığına karar verilmiştir.

Metin ön işlemlerinin sonunda elde edilen veri tabanında her bir makalenin üstverisi;

- Türkçe başlık
- İngilizce başlık
- Yazarlar
- Türkçe anahtar kelimeler
- İngilizce anahtar kelimeler
- Türkçe klasik öz
- İngilizce klasik öz

detayıyla elde edilmiştir.

Tam metnin cümleleri ise

- Paragraf no
- Cümle no
- Üst Başlık
- IMRAD bölümü
- Cümle metni

bileşenleriyle ilişkisel bir veri tabanında tutulmaya başlanmıştır. Bu ilişki her bir dokümanın tam metni için, öncelikle ilgili paragrafın ana metindeki yer almasına göre sıra numarası, yazarının uygun gördüğü bölüm adlandırması ve bulunduğu paragrafın işaretçi görüşlerine göre yer aldığı yapısal bölümlerine göre.

Bu aşamadan sonra cümlelere ayrılmış işlenmemiş metin parçaları: etiket ve özel formatlardan arındırıldıktan sonra kök bulma aşamasına geçilmiştir. Kök bulma aşamasına kadarki tüm işlemler geliştirilen PHP betikleri ve MySQL komutları aracılığıyla gerçekleştirilmiştir.

Kök bulma, aynı kelime kökünden türemiş olan terimleri tespit etmek için yapılır. Özellikle Türkçe gibi sondan eklemeli bir dilde eğer kelime sıklığı, benzerlik vb. analizler yapılacaksa, kelimelerin son ekleri (çekim ekleri, çoğul ekleri gibi) çıkartıldıktan sonra işlenmesinin sonuçlara etkisi büyüktür. Böylece metin içinde farklı şekillerde yer alan ancak kökleri itibarıyla aynı anlamda birleşen kelimeler tek bir şekilde temsil edilebilir. Örneğin, “kütüphane”, “kütüphaneler”, “kütüphanenin”, “kütüphanelerin” gibi dört farklı kullanım kök bulma aşamasından sonra “kütüphane” olarak kelime vektöründe yerini alır. Çalışmada kök bulma işlemi için Zemberek kullanılarak her bir cümle durak kelimelerinden<sup>2</sup> arındırıldıktan sonra köklerine ayrılmıştır.

---

<sup>2</sup> Durak kelimeleri bilgi taşıyıcılığı önemsiz bazı bağlaç, edat ve zamirlerden oluşan kelimelerdir (Türkçede ve, veya, ancak vb.).

### 3.3. DOKÜMAN VEKTÖRLERİ VE AĞIRLIKLANDIRMA

Çıkarıma dayalı otomatik metin özetleme işlemleri için kök bulma aşamasından sonra her bir cümle terimlerine göre ağırlıklanarak hesaplamalara katılır. Tez çalışması kapsamında cümleler, kelime sıklıklarına göre ağırlıklandırılmıştır. Bu bağlamda kullanılan sıklık ağırlıkları oluşturulurken her bir cümleye ait her bir terimin ilgili cümlede görünme sayısı hesaplanmıştır (Dumais, 1991, s. 232).

Örneğin, *Bilgi Dünyası Dergisi*, 2000 yılı 1. sayı 1. makalenin 2. ana başlığının altındaki, 5. paragrafının metninin 27. cümlesi: “Tarım toplumundan sanayi toplumuna geçiş eğitimi nasıl etkilediyse, sanayi toplumundan bilgi toplumuna geçiş de kurumların yapısında köklü değişiklikleri zorunlu kılmaktadır.”. İlgili cümle veri tabanında Şekil 10’daki gibi tutulmaktadır. Buna göre, cümlenin yer aldığı makalenin tanımlayıcı kimlik numarası (paper\_id), cümlenin bağlı bulunduğu paragrafın sıra numarası, cümlenin sıra numarası, cümlenin uzman işaretçilere göre içinde yer aldığı IMRAD bölümü ve yazar tarafından yer aldığı bölüme verilen isimden oluşan bir üstverisi Şekil 10’da görülmektedir.

paper_id	paragraf_no	cumle_no	imrad	baslik	text
BD200011	5	27	1	geleceğe yönelik tartışmalar	Tarım toplumundan sanayi toplumuna geçiş eğitimi nasıl etkilediyse, sanayi toplumundan bilgi toplumuna geçiş de kurumların yapısında köklü değişiklikleri zorunlu kılmaktadır.

Şekil 10. Kelime sıklık vektörlerinin üstverileri

İlgili cümle kök bulma işlemi sonunda Şekil 11’deki terimler dizisi ve buna karşılık, her bir terimin cümle içinde görüldüğü sıklıkların dizisi veri tabanına kaydedilmiştir. Cümlelerin bu temsilci sıklık vektörleri otomatik cümle puanlama işlemleri için kullanılmaya hazırdır.

paper_id	cumle_no	imrad	kelimeler	kelime_vektor
BD200011	27	1	"toplum","sanayi","geç","eğitim","yapı","bilgi","tarım","kur","kıl","değiş","nasıl","etki","kök","zorunlu"	4,2,2,1,1,1,1,1,1,1,1,1,1,1,1

Şekil 11. Kelime sıklık vektörlerinin temsili

n-gram dizilerini elde etmede kullanılacak kelime kökleri de benzer anlayışla yeniden üretilerek veri tabanına kaydedilmiştir. Buna göre, “Tarım toplumundan sanayi toplumuna geçiş eğitimi nasıl etkilediyse, sanayi toplumundan bilgi toplumuna geçiş de kurumların yapısında köklü değişiklikleri zorunlu kılmaktadır.” cümlesinin n-gram dizisi “tarım, toplum, sanayi, toplum, geç, eğitim, nasıl, etki, sanayi, toplum, bilgi, toplum, geç, kur, yapı, kök, değiş, zorunlu, kıl” şeklindedir. Cümle içinde dört kez geçen toplum kelimesinin bu kez cümledeki sırasıyla temsil edildiğine dikkat edilmelidir.

### 3.4. OTOMATİK CÜMLE SEÇİM İŞLEMLERİ

Çıkarıma dayalı otomatik metin özetleme yöntemleri doküman içerisindeki cümleleri puanlama, sıralama ve seçme işlemidir. Otomatik metin özetleme yaklaşımları ve yöntemleri tam metinden belli başlı önemli temsilci cümleler belirlemek için işe koşular. Cümleler önceden belirlenen özelliklerine göre puanlanır ve her bir cümle için doküman içindeki önemi bu puanlarla belirlenmiş olur. Her bir özelliği ağırlıklandırarak bir araya getiren cümle seçme fonksiyonları çıkarıma dayalı otomatik özetleme sistemlerinin bir diğer aşamasıdır. Literatürde çeşitli diller için bu özelliklerin önemli cümle seçme işinde birbirlerine göre üstünlüklerini karşılaştıran çok sayıda çalışma mevcuttur.

Tez çalışması ile cümle seçme işleminde makalelerin yapısal özelliklerini kullanmanın çıktılara katkısının analiz edilmesi hedeflenmiştir. Bu doğrultuda, cümle seçiminde Türkçe veri setleri üzerinde daha önce denenmiş temel özelliklerin birleşiminden oluşan bir fonksiyon ile cümle puanlamaları gerçekleştirildikten sonra, cümle seçiminde yapısal özellikleri göz önüne alma ve almama yaklaşımlarının çıktılara katkısı test edilmiştir. Türkçe metinler üzerinde üstünlükleri daha önce denenip, gösterilmiş (Güran, vd. 2014) cümle konumu, ortak yakınlık sayısı ve isim soylu kelimeleri içerme özellikleri kullanılan fonksiyonun bileşenleridir. Bu özelliklerden, ortak yakınlık sayısı, cümle merkeziliği özelliği içinde ele alınmıştır. Cümle merkeziliği özelliği, cümlelerin tam metindeki diğer cümleler ile kosinüs benzerlikleri, ortak yakınlıkları ve n-gram çakışmalarının birleşimini cümle sayısına bölerek normalleştirmektedir (Binwahan, Salim ve Suanmali, 2010, s 572.). Bu özelliklerden, cümle konumunu puan ölçütü olarak kullanan başka birçok yaklaşım vardır (Fattah ve Ren, 2009; Ferreira vd. 2013; Kutlu vd., 2010; Pembe, 2011).

Ortak yakınlık sayısı hesabında temel olan kosinüs benzerliği ve n-gram çakışmaları ise özetleme sistemlerinin değerlendirmelerinde de standart olarak kullanılan ölçümlerdendir.

### 3.4.1. Cümle Puanlamada Kullanılan Özellikler

**Cümle Konumu:**  $C_i$  cümlesinde  $i$  cümlenin doküman içindeki sıra numarası,  $N$  dokümanın cümle sayısı olmak üzere,  $C_i$  cümlesinin cümle konumuna göre hesaplanan sıralama puanı:

$$CK(C_i) = \frac{N-i}{N-1} \quad (5)$$

Formül 5 ile hesaplanmıştır. Böylece her bir cümleye makaledeki görünme sırasına göre 1'den 0'a kadar sıralama puanları verilmiştir.

**Cümle Merkeziliği:** Merkezilik, çıkarıma dayalı otomatik metin özetlemede pek çok türde metin ve derlem için en çok uygulanan özelliktir. Cümleleri puanlamada, tam metinde verilen temel bilgiyi temsil etme derecelerini bulma prensibine dayanır. Tez kapsamında,  $n$  tane cümleye sahip bir doküman için her bir cümlenin merkeziliği, Formül 6'daki gibi elde edilmiştir (Binwahlan vd., 2010. s 572.).

$$CM(C_i) = \frac{\sum_{j=1}^{n-1} \cos(C_i, C_j) + \sum_{j=1}^{n-1} n\text{-yakınlık}(C_i, C_j) + \sum_{j=1}^{n-1} n\text{-gram}(C_i, C_j)}{n-1},$$

$$i \neq j \text{ ve } \cos(C_i, C_j) \geq 0,16 \quad (6)$$

Formül 6'ya göre merkezilik, ele alınan bir  $C_i$  cümlesi için, dokümandaki diğer  $C_j$ 'lerle aralarındaki kosinüs benzerliği, paylaşılan yakınlık sayısı ve paylaşılan ortak n-gram sayısının toplamından elde edilir. Bu toplam değeri  $n-1$  ile bölünerek normalize edilmiştir. Burada paylaşılan yakınlık sayısı, hem  $C_i$  hem de  $C_j$ 'ye benzer cümle kümeleri üzerinden



Formül 7'deki gibi hesaplanır. Paylaşılan n-gramlar için Formül 8 uygulanırken 2-gramlar kullanılmıştır. Paylaşılan yakınlık sayısı için eşik değeri  $\cos(C_i, C_j) \geq 0,16$ 'dır. Buna göre,  $|X|$ , X kümesinin eleman sayısını vermek üzere,

$$\text{n-yakınlık}(C_i, C_j) = \frac{|C_i(\text{yakınlık}) \cap C_j(\text{yakınlık})|}{|C_i(\text{yakınlık}) \cup C_j(\text{yakınlık})|}, i \neq j \quad (7)$$

$$\text{n-gram}(C_i, C_j) = \frac{|C_i(\text{n-gram}) \cap C_j(\text{n-gram})|}{|C_i(\text{n-gram}) \cup C_j(\text{n-gram})|}, i \neq j \quad (8)$$

olacak şekildedir (Binwahlan vd., 2010. s 573.).

**İsim Soylu Kelime İçerme:** Bu çalışma kapsamında ele alınan bir başka özellik de cümlelerin isim soylu kelime içerip içermemesidir. İsim soylu kelimeleri tespit etmek için Zemberek kullanılmıştır. Her bir cümlenin isim soylu kelime puanı (İSP) hesaplanırken, ilgili cümlenin içerdiği isim soylu kelime sayısı, cümlenin toplam kelime sayısının bir eksiğine bölünerek normalleştirilmiştir.

Cümle konumu, cümle merkeziliği ve isim soylu kelime içerme puanları hesaplandıktan sonra, her üç özelliğin ağırlığını eşit kabul eden doğrusal Formül 9 ile birleştirilerek her bir makalenin i. cümlesi  $C_i$  ile ifade edilmek üzere sıralama puanları  $SP(C_i)$ 'ler aşağıdaki gibi hesaplanmıştır.

$$SP(C_i) = CK(C_i) + CM(C_i) + İSP(C_i) \quad (9)$$

Hesaplama işlemi veri tabanında yer alan kelime sıklık vektörleri ve n-gram dizileri kullanılarak, PHP betikleri ile gerçekleştirilmiştir.

### 3.4.2. Cümle Sıralama ve Seçim İşlemi

Otomatik özetleme sistemlerinin çıktıları elde edilirken tam metinden hangi oranda sıkıştırılmış bir özet elde edileceğine sıkıştırma oranı denir. uzunluk(X) fonksiyonu X metin parçasının uzunluğun kelime sayısına göre vermek üzere, TM tam metin, Ö özet gösterimi ile bir otomatik özetleme sisteminin sıkıştırma oranı, Formül 10'daki gibi ifade edilebilir:

$$\text{sıkıştırma oranı} = \frac{\text{uzunluk (Ö)}}{\text{uzunluk (TM)}} \quad (10)$$

İyi bir özetleme sisteminden beklenen, çıktılarının sıkıştırma oranları 0'a, yakınsarken taşıdıkları bilgi miktarının artmasıdır (Hovy, 2005, s. 593). Sıkıştırma oranları kelime sayısı cinsinden sabit bir üst sınır olarak seçilebileceği gibi her bir girdi dokümanın uzunluğuna göre değişen kelime sayılarında, yani tam metnin belirli bir yüzdelik değerinde sıkıştırma olarak da belirlenmiş olabilir. Akademik makalelerin orijinal özleri tam metnin uzunluğundan etkilenmez. Dergilerin kılavuzlarında yer verilen özlerle ilgili kurallar arasında çoğunlukla belirli bir kelime sınırı da yer alır. Bu sınır genellikle 150-300 kelime arasında değişebilir (APA, 2010, s. 26). *TK* dergisinin yazım kılavuzunda özler için bir kelime sınırı belirlenmemiştir. *BD* dergisinin yazım kılavuzunda belirlenmiş 250 kelime sınırı genellikle kullanılan kelime sınırına göre makul bir değer olduğu için geliştirilen sistemin çıktıları için de üst sınır olarak kabul edilmiştir. Otomatik yapısal özler oluşturulurken, belirlenen bu 250 kelime sınırı makale tam metninde kullanılan her bir yapısal bölüme eşit olarak dağıtılarak kullanılmıştır.

Derlem tam metnlerinin yapısal bölüm işaretlemesi yaygın kabul gören ve bilinirliği olan IMRAD başlıkları ile uyumlu olacak şekilde gerçekleştiği için sistemimizin çıktısı olan yapısal özün düzeni de IMRAD ile uyumludur. Buna göre hedeflenen yapısal düzen ile elde edilen içeriğin özelliklerinin aşağıdaki gibi olması beklenmektedir:

**Giriş:** Çalışmanın arka planı ve motivasyonu nedir? Hangi problem incelendi?

**Yöntem:** Problem nasıl incelendi? Araştırma kapsamında neler yapıldı?

**Bulgular:** Neler bulundu?

**Tartışma ve Sonuç:** Çalışmanın bulguları ne anlama gelmektedir? Çalışmanın özgün sonuçları nelerdir?

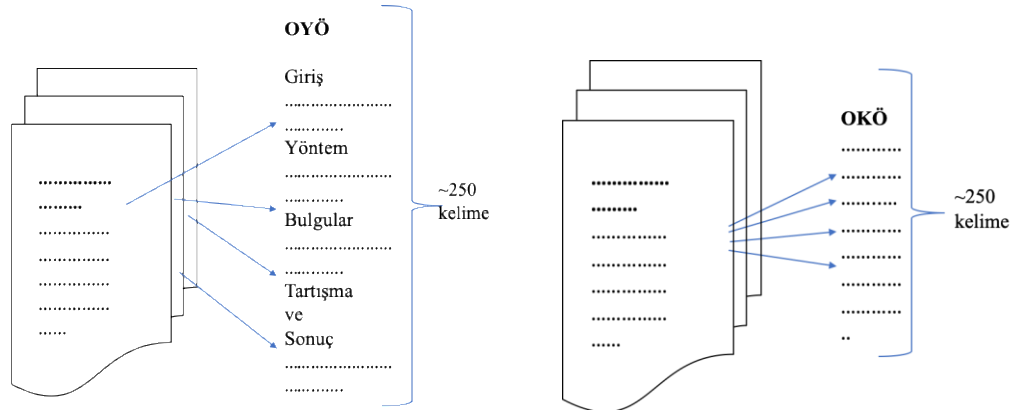
Sistemimizin hedeflenen çıktısı otomatik yöntemlerle elde edilen yapısal özlerdir (OYÖ). Bu çıktılarının yanı sıra, yapısal özellikleri kullanmadan elde edilen otomatik klasik özler (OKÖ) ile yapısal özellikleri göz önüne almanın çıkarıma dayalı otomatik metin özetleme sistemine etkisi değerlendirilmiştir.

OYO ve OKÖ elde etmek için öncelikle cümleler  $SP(C_i)$  fonksiyonu ile puanlanmışlardır. Puanlanmış cümleler aşağıdaki aşamalardan geçirilerek çıktılar elde edilmiştir:

- OYÖ elde ederken, cümleler öncelikle makalelerin insan işaretçiler tarafından işaretlenmiş her bir yapısal bölümüne, daha sonra  $SP(C_i)$  fonksiyonu ile hesapladığımız puanlarına göre sıralanmıştır. Sıralama aşamasından sonra her bir IMRAD bölümünden eşit karakter sınırına ulaşana dek önemli cümle seçme işlemi gerçekleştirilmiştir. Böylece her bir yapısal bölüm için o bölümün en yüksek puanlı cümleleri elde edilmiştir. Her bir yapısal bölümden en yüksek puanlı cümleleri seçme işlemi toplam kelime sınırının ilgili makalenin yapısal bölüm sayısına dağıtılarak kullanılmasıyla gerçekleşmektedir. Örneğin, iki yapısal bölümden oluşan bir makale için, her bir yapısal bölümde  $250/2=125$  kelime sınırına ulaşana kadar kelime seçim işlemine devam edilmiştir.
- Otomatik klasik özler (OKÖ) için ise, tam metnin yapısal özellikleri dikkate alınmamıştır.  $SP(C_i)$  fonksiyonu ile elde edilen puanları sıralandıktan sonra 250 kelime sınırına ulaşana değin cümle seçimi gerçekleştirilmiştir.

Özetleyecek olursak, otomatik cümle sıralama işlemleri derleme uygulandıktan sonra, cümle seçim işlemi Şekil 12'deki görselleştirilen iki farklı algoritma ile aşağıdaki gibi yapılmıştır:

- Dâhil olduğu yapısal bölümüne ve puanına göre sıralı cümleler dizisinden, makalelerin her bir yapısal bölümünü gezerek 250 kelime sınırına ulaşıncaya kadar yapılan cümle seçme işlemi ile OYÖ'ler,
- Puanına göre sıralı cümleler dizisinden en yüksek puanlı cümleleri 250 kelime sınırına ulaşıncaya kadar seçme işlemi ile OKÖ'ler elde edilmiştir.



Şekil 12. OYÖ ve OKÖ için tam metinden cümle seçimleri

Puanlarına göre sıralanmış cümlelerden seçim yaparak aynı kelime uzunluğunda metin çıktıları elde etmek doğal olarak mümkün değildir. Örneğin, cümle seçim işlemi devam ederken toplam kelime sayısı 245'e geldiğinde bir sonraki yüksek puana sahip cümlelerin kelime sayısı 15 ise o cümle öze katılmadan işlem sonlanacak ve 245 kelimelik bir öz elde edilmiş olacaktır.

Seçim işleminden önce her bir i cümlesi için sıralama puanı hesaplandıktan sonra skor tablosuna eklenmiştir. Şekil 15'te bir kesiti görülen hesaplama tablosunda skor\_2020 ile gösterilen alan bu toplam sıralama puanları diğer alanlar  $SP(C_i)$  fonksiyonunda ele alınan özelliklerinin tekil değerleridir. Bu haliyle sistem veri tabanı sıralama puanında farklı deneyler yapmak için uygun bir yapıdadır. Hesaplamalar sonucunda puanlanan cümleler yukarıda tarif edilen iki farklı şekilde sıralanarak sistem çıktıları elde edilmiştir.

paper_id	cumle_no	imrad	isim_soylu	cos_benzerligi	cos_yakinlari	ngram	siralama_puani	merkezlilik	skor_2020
BD200023	53	Giriş	3	35.5478	79.435178128251	0.061	0.747573	0.558466	2.80604
BD200023	82	Bulgular	5	43.3319	91.405170425248	0.0819	0.606796	0.654461	2.26126
BD200023	51	Giriş	2	27.3523	74.178638996143	0.0882	0.757282	0.493297	2.25058
BD200023	114	Bulgular	5	25.2735	74.428422857082	0.0552	0.451456	0.484258	2.18571
BD200023	47	Giriş	7	36.583	91.958562013043	0.0641	0.776699	0.624299	2.17878
BD200023	19	Giriş	9	38.3226	89.520941028837	0.0383	0.912621	0.620786	2.17626
BD200023	27	Giriş	4	21.0474	68.619566035282	0.0539	0.873786	0.435538	2.10932
BD200023	45	Giriş	12	43.5048	88.752748989614	0.0393	0.786408	0.642218	2.09529
BD200023	40	Giriş	15	31.5641	79.532140036973	0.0317	0.81068	0.539456	2.06442
BD200023	30	Giriş	6	36.4957	83.161037920859	0.0488	0.859223	0.581095	2.04032
BD200023	60	Giriş	7	36.8843	89.305111659372	0.0652	0.713592	0.612886	2.02648
BD200023	28	Giriş	1	5.975	24.654995723197	0	0.868932	0.148689	2.01762
BD200023	107	Bulgular	3	28.2197	76.526176495634	0.0615	0.485437	0.508774	1.99421
BD200023	20	Giriş	12	32.0797	78.847915276621	0.0305	0.907767	0.538632	1.99185
BD200023	1	Giriş	6	8.9839	37.97599116625	0	1	0.227961	1.97796
BD200023	42	Giriş	10	19.8853	71.983702027034	0.04	0.800971	0.44616	1.96142
BD200023	95	Bulgular	13	32.9724	80.207497418246	0.039	0.543689	0.549606	1.95996
BD200023	57	Giriş	3	10.0666	35.698222607234	0.0678	0.728155	0.222488	1.95064

Şekil 13. OKÖ cümlelerinin seçime hazırlanmış görüntüsü

Şekil 13 ve Şekil 14 çıktı cümlelerinin veri tabanında sıralanmış hallerini göstermektedir. Şekil 13'te ele alınan makalenin seçimi, tam metni dört IMRAD bölümünden oluşan ve OKÖ sürümüne tam metnin birden fazla IMRAD bölümünden cümle seçilebilmiş makaleler arasından rastgele gerçekleşmiştir. Bu ekran görüntüsü, skor tablosunun, paper\_id alanı 'BD200023' ve skor\_2020 alanı azalan sırada olacak şekilde 250 kelime sınırına kadar elde edilen cümlelerden oluşmuştur.

Şekil 14 ise OYÖ cümlelerinin veri tabanındaki kesitlerinin görünümüdür. OYÖ için cümleler, önce "imrad", sonra "skor\_2020" değerine göre sıralanmış, 250 kelime sınırı her bir IMRAD bölümü için eşit şekilde dağıtılarak cümle seçimi yapılmıştır. Şekil 14 her bir IMRAD bölümünün en yüksek puanlı cümlelerinin kelime sınırına ulaşana kadar sıralanmasıyla oluşmaktadır.

paper_id	cumle_no	imrad	isim_soylu	cos_benzerligi	cos_yakinlari	ngram	siralama_puani	merkezlilik	skor_2020
BD200023	53	Giriş	3	35.5478	79.435178128251	0.061	0.747573	0.558466	2.80604
BD200023	51	Giriş	2	27.3523	74.178638996143	0.0882	0.757282	0.493297	2.25058
BD200023	47	Giriş	7	36.583	91.958562013043	0.0641	0.776699	0.624299	2.17878
BD200023	19	Giriş	9	38.3226	89.520941028837	0.0383	0.912621	0.620786	2.17626
BD200023	27	Giriş	4	21.0474	68.619566035282	0.0539	0.873786	0.435538	2.10932
BD200023	66	Yöntem	13	56.9412	92.014050827385	0.0976	0.684466	0.723558	1.92802
BD200023	62	Yöntem	7	47.2129	91.262705310759	0.1127	0.703883	0.672759	1.9151
BD200023	82	Bulgular	5	43.3319	91.405170425248	0.0819	0.606796	0.654461	2.26126
BD200023	114	Bulgular	5	25.2735	74.428422857082	0.0552	0.451456	0.484258	2.18571
BD200023	107	Bulgular	3	28.2197	76.526176495634	0.0615	0.485437	0.508774	1.99421
BD200023	95	Bulgular	13	32.9724	80.207497418246	0.039	0.543689	0.549606	1.95996
BD200023	96	Bulgular	5	30.8162	85.99110128259	0.0553	0.538835	0.567294	1.93946
BD200023	165	Tartışma ve Sonuç	9	27.8256	78.485897434101	0.0427	0.203883	0.516282	1.72017
BD200023	159	Tartışma ve Sonuç	5	6.4946	31.045185369281	0.0455	0.23301	0.182453	1.66546
BD200023	166	Tartışma ve Sonuç	10	32.2954	88.41402802629	0.054	0.199029	0.58623	1.49955

Şekil 14. OYÖ cümlelerinin seçime hazırlanmış görüntüsü

BD200023 kimlik numaralı makale için orijinal öz, OYÖ ve OKÖ metinleri, Ek 1’de incelenebilir. Veri tabanı üzerinden çıktı örnekleri gösterilen tüm seçim işlemleri PHP betikleri ile otomatik olarak gerçekleştirilmiştir.

### 3.5 OKUNURLUK HESAPLAMALARI

Tez çalışması kapsamında derlem özellikleri incelenirken, her bir makalenin gövde metninin ve klasik özlerinin okunurluk değerleri de sayısal olarak hesaplanmıştır.

Okuma süreci okuyucunun anlam çıkarması, verilen mesajı kavraması ve nihayet yorumlamasını içeren karmaşık bir süreçtir (Özdemir, 2000, s. 21). Bir metnin okuyucu kitlesinin seviyesine uygunluğuna karar vermek önemlidir. Metnin hedef okuyucu kitlesine uygunluğu okunurluk hesaplamalarıyla belirlenir. Türkçe metinlerin okunurluklarını ölçmek için dile özgü bir formül üretilmemiş olsa da, 1997 yılından beri yaygın olarak “Flesch Reading Ease” (FRE) orijinal adıyla adlandırılan formülün bir uyarlaması (Ateşman, 1997) kullanılmaktadır. FRE, bir metnin okunurluğuna metinde geçen kelimelerin hece uzunluğu ve cümlelerin kelime uzunluğunu temel alarak hesaplayan bir formüldür. Ateşman’ın bu formülden yola çıkarak geliştirdiği, Türkçe için yaygın olarak kullanılan ve bu çalışma okunurluk hesaplamalarında yararlanılan eşitlikler Formül 11 ve 12, ölçek ise Tablo 4’teki gibidir:

$$\text{okunabilirlik değeri} = 98,825 - 40,175h_o - 2,610k_o \quad (11)$$

$$h_o = \frac{\text{Toplam hece sayısı}}{\text{Toplam kelime sayısı}}, \quad k_o = \frac{\text{Toplam kelime sayısı}}{\text{Toplam cümle sayısı}} \quad (12)$$

Tablo 4. Ateşman okunabilirlik ölçeği

Ateşman Değeri	Metnin Okunabilirlik Derecesi
90-100	Çok Kolay
70-89	Kolay
50-69	Orta Zorlukta
30-49	Zor
0-29	Çok Zor

Okunurluk hesaplaması bu haliyle kelime ve cümle uzunlukları üzerine kurulu olsa da, sondan eklemeli bir dil olan Türkçede kelimeleri uzatan yapım ekleri okunurluğu her zaman zorlaştırmayabilmektedir. Dahası metin özelliklerinin yanı sıra okuyucunun hazır bulunuşluğu ve bilgi düzeyi, okunan ortam, metnin sunumu (harflerin puntosu, italik ya da düz yazılması vb.), jargon kullanımı gibi okunurluğu etkilediği düşünülen başka faktörler de vardır (Ateşman, 1997; Çoban, 2014, s. 97). Kısa cümlelerin uzun cümlelere göre çoğu kez daha kolay okunacağı açıktır ancak kelimelerin okura tanıdık ya da yabancı olmasının da okunurluğa etkisi büyüktür. Örneğin, nicel özellikleri dikkate alındığında “Ben, annem ve babam pazartesi günü Kahramanmaraş’a uzun süredir görüşemediğimiz abimi ziyarete gideceğiz.” cümlesinin okunurluğunun “Galat-ı meşhur fasih-i mehcurdan evladır.” cümlesinin okunurluğundan zor olması gerekirken, çoğunluk için ikinci cümlelerin anlaşılabilir olduğu dile getirilmektedir (Çoban, 2014, s. 108).

Türkçe için kolay ve zor sözcükleri araştırıp sözlük oluşturmak ve bu sözlüğü hesaba katarak metinlerin okunma zorluklarına karar verecek bir hesaplama geliştirmek büyük bir kazanç olacaktır. Ancak henüz bu konuda yapılmış kapsamlı bir araştırma mevcut değildir. Akademik metinlerde jargon kullanımının yaygın olduğu, kullanılan dilin genellikle zor olmasının alışkanlık haline geldiği bilgileri ışığında metnin cümle ve kelimelerin uzunluğu üzerine bir ölçekte değerlendirilmesinin ayırt edici olacağı, bu sebeple Türkçe akademik metinlerin okunurluğunu karşılaştırmada Ateşman Okunurluk Değeri (AOD)’nin kullanılabilir olacağı düşünülmüştür.

Çalışma kapsamında makalelerin klasik özlerinin, gövde metinlerinin ve geliştirilen otomatik yapısal özetleme sistemi ile üretilen özlerin okunabilirliği PHP betikleri ile hesaplanmıştır. Bunun için veri tabanında metin parçası olarak tutulan orijinal yazar özlerinin okunurluğu hesaplanırken her bir öz metni cümlelerine ayrıştırılmış, daha sonra her bir cümledeki boşluk karakterleri sayılarak toplam kelime sayıları ve her bir kelimedeki ünlü harfler sayılarak toplam hece sayıları bulunmuştur. Hece sayıları hesaplanırken Türkçe kelimelerde her bir ünlü harfin bir heceyi belirlediği bilgisi kullanılmıştır. Tam metin ve sistem çıktılarının cümleleri veri tabanında cümlelerine ayrıştırılmış halde mevcuttur. Kalan işlemler sistem çıktılarına uygulanan ile aynıdır. Bu şekilde cümle başına kelimeler, kelime başına heceler elde edilmiş ve sonuç olarak

Formül 11 ve 12'den elde edilen okunurluk değerlerinden Tablo 4'e göre okunurluk dereceleri (AOD) elde edilmiştir.

### **3.6 OTOMATİK ÖZLERİN DEĞERLENDİRİLMESİ**

Tez çalışması kapsamında otomatik yöntemler ile öz elde ederken tam metnin yapısal özelliklerini gözeterek cümle seçmenin sistem çıktılarına etkisinin ölçülmesi de otomatik yöntemler ile yapılmıştır. Çıktı başarımları, okunurluk, tam metni tarama düzeyi ve ROUGE ölçümleri ile değerlendirilmiştir edilmiştir. ROUGE değerlendirmesi için geliştirilen çıkarıma dayalı otomatik özetleme sisteminin yapısal çıktılarının (OYÖ) yanı sıra klasik çıktılar (OKÖ) da kullanılmıştır. OYÖ ve OKÖ'lerin yazar tarafından üretilmiş orijinal özlerle n-gram çakışmaları ROUGE 2.0 paketi ile ROUGE-1, ROUGE-2, ROUGE-L ve ROUGE-SU4 ölçümleri kullanılarak hesaplanmıştır. Okunurluk değerlendirmeleri AOD hesaplanarak yapılmıştır. Sistem çıktılarını tam metni tarama yetenekleri ise çıktıların tam metnin yapısal bölümlerine yığılım grafikleri ile görselleştirilerek analiz edilmiştir.



## 4. BÖLÜM: BULGULAR

Bu bölümde öncelikle *BD* ve *TK* dergilerinin yazım kılavuzları incelenerek dergi politikalarında yazar alışkanlıklarına etki edecek farklar olup olmadığı anlaşılmasına çalışılmıştır. Ardından derlem kapsamındaki makale tam metnlerinin yapısal bölümleri analiz edilerek, IMRAD düzenine uygun yapısal bölümlerin bu tez çalışması çerçevesinde yeterliliği sorgulanmıştır. Daha sonra, dergiler, kullanılan dil açısından özlerin ve tam metinlerin okunurluğunu etkileyecek unsurlara göre karşılaştırmalı olarak incelenmiştir. Derlemin okunurluğuna dair incelemelerde makalelerin okunurluk katsayılarına dağılımları hem *TK* ve *BD* dergilerine göre hem de tüm derlem bağlamında incelenerek, derlemin okunurluk özelliklerini tanımlayıcı bulgular ortaya konmuştur.

Bir sonraki aşamada geliştirilen otomatik yapısal özetleme sisteminin çıktılarına ilişkin yapılan incelemeler ele alınmıştır. Bunun için ilk olarak, OYÖ ve OKÖ'lerin meydana geldiği cümlelerin tam metindeki dağılımları grafikler eşliğinde gösterilmiştir. Böylece akademik makaleler için çıkarıma dayalı otomatik özetleme yöntemleri ile öz hazırlarken metinlerin yapısal özelliklerini sisteme eklemenin tam metni taramadaki katkısı gösterilmiştir. OYÖ ve OKÖ çıktılarının okunurluk değerleri hesaplandıktan sonra, bir önceki aşamada elde edilen tam metin ve orijinal özlerin okunurluk değerleri ile karşılaştırmalı olarak incelenmiştir. Son olarak, geliştirilen sisteme yapısal özellikler eklemenin çıktı başarımlarına katkısını görmek için OYÖ ve OKÖ çıktıları ile yazarların ürettiği orijinal özlerin n-gram çakışmaları her bir ROUGE ölçümüne göre hesaplanmıştır. Yapısal özelliklerin sistem başarımına katkısı incelenirken tam metinde kullanılan yapısal bölüm sayısının çıktıların başarımı üzerindeki etkisi de ortaya konmuştur.

### 4.1. BD VE TK YAZIM KURALLARI

Yazarlar hakemli dergilere yayın gönderirken hedefledikleri derginin yazım kurallarına uygun içerik üretmek durumundadırlar. Dergiler ise kendi yayın kurallarını oluştururken, makaleler için belirli yapısal biçimlere uygun olacak şekilde şablonlar üretebilirler.

Dergiler, söz konusu kuralları herkesin erişebileceği şekilde web sayfalarından duyururlar. *BD* ve *TK* dergileri de kurallarını web sayfalarında ana sayfadan tek tıklamayla ulaşılabilecek şekilde yayımlamaktadır. *BD* ana sayfasında kurallar bağlantısı dikkat çekici bir buton halinde “YAZAR REHBERİ” başlığı ile verilmiştir (Bilgi Dünyası, 2012). *TK* web sayfasında kurallara erişim birkaç farklı bağlantıdan düzenlenmiştir. *TK* ana sayfasından hem ana menüden “GÖNDERİLER” bağlantısı ile hem de ana sayfanın bitiminde yer alan “Yazar Rehberi” bağlantısı ile kurallara ulaşılabilmektedir. Bu ikisinin yanı sıra, yine ana sayfadan ulaşılabilen, “Yazarlar İçin” bağlantısı altından “Yazar Rehberi” bağlantısı, iki tıklama ile kurallara ulaşabilmeyi mümkün kılmaktadır (Türk Kütüphaneciliği, 2018).

*TK* yazar rehberlerinde dergiye gönderilecek makalelerin ana bölümleri (Türkçe ve İngilizce makale adı, yazar adları, Türkçe ve İngilizce öz) ve yazının biçimsel özellikleri (fontlar, puntolar, vurgular vb.) ile ilgili kurallar listelenmişken öze ya da gövde metniyle ilgili yapısal bir biçim tavsiyesi yer almamaktadır. *BD* yazar rehberi *TK*'ye göre daha detaylıdır. “Yayın Politikaları” başlığı altında sayfa düzeni ve font ailesinin yanı sıra makalede yer alması istenen her bir yapısal bölüm IMRAD formatına uygun bir şekilde sıralanarak bu başlıklar altında nelerin yer almasının beklendiği detaylı olarak açıklanmıştır. İstenen başlıklar, “Giriş”, “Yöntem”, “Bulgular”, “Tartışma, Sonuç ve Öneriler” şeklinde IMRAD ile uyumlu yapısal biçimdir. *BD*, öz yazımı ile ilgili yapısal bir format önermese de içerik olarak “Çalışmanın amacını, kapsamını, yöntemini, sonuçlarını, öne çıkan yanlarını ve özgün değerini açık ve kısa bir şekilde yansıtmalıdır.” ifadesi ile içeriğin yapısal özellikleri gözeterek üretilmesini istemektedir.

Her iki derginin web sayfalarında yukarıda belirtilen yönlendirmelere göre, *BD* ve *TK* yayınları için belirlenmiş bir yapısal düzen kuralı olmadığı anlaşılmaktadır. Bunlardan *BD* dergisinin belirli bir taslak önermeden de olsa içeriğin düzenlenmesiyle ilgili IMRAD ile uyumlu bir çerçeve çizdiği görülmektedir. Her iki dergi de özleri için IMRAD ya da benzeri belirgin bir şablon belirlememiştir ve klasik öz kabul etmektedir.

## 4.2. TÜRKÇE KÜTÜPHANECİLİK VE BİLGİBİLİM ALANINDA YAYINLANMIŞ MAKALELERİN YAPISAL BİÇİMLERİ

Bu bölümde uzman işaretçiler tarafından yapılan yapısal bölüm etiketlemesine göre tam metinlerinin IMRAD'a uyumları ele alınmıştır. Çalışma kapsamında makalelerin yapısal bölümlerine karar vermede uzmanlar tarafından yapılan bu etiketleme kullanılmıştır. Tablo 5 içerdikleri IMRAD bölümü sayısına ve bu bölümlerin birlikte yer alış desenlerine göre makale sayılarını göstermektedir.

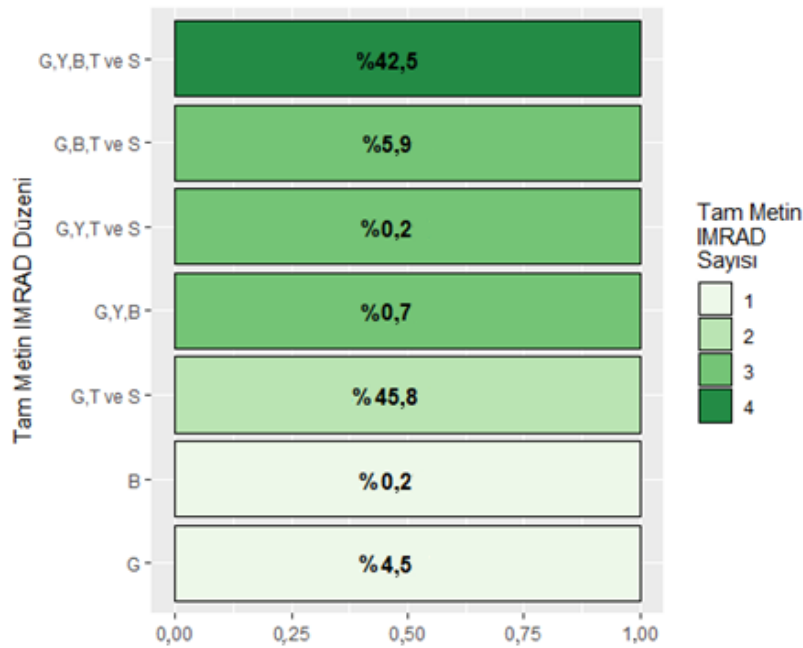
Tablo 5. IMRAD'a uyumlu bölüm sayıları ve yapısal düzenlerine göre makale sayıları

IMRAD Bölümü	Makale Sayısı	IMRAD Düzeni	Makale Sayısı
1	20	Giriş Bulgular	19 1
2	193	Giriş, Tartışma ve Sonuçlar	193
3	29	Giriş, Yöntem, Bulgular Giriş, Yöntem, Tartışma ve Sonuçlar Giriş, Bulgular, Tartışma ve Sonuçlar	3 1 25
4	179	Giriş, Yöntem, Bulgular, Tartışma ve Sonuçlar	179

Tablo 5'te görüldüğü gibi, derlemde en yaygın görülen yapısal düzen giriş ile tartışma ve sonuçlar olmak üzere iki IMRAD bölümünden oluşmaktadır. Bu tür kullanımın derlemdeki yaygınlığı derlem bağlamındaki yazarların iki IMRAD düzeni ile yazma alışkanlığına sahip olduğunu düşündürmektedir. Sosyal bilimler alanında yazılmış makalelerinin iki IMRAD bölümünden oluşması kimi zaman kabul görüyor olsa da her araştırmanın bir problemi ve bu problemi çözmek için bir yöntemi olmasının gerekliliği de bir diğer yaygın görüştür.

Derlem genelinde iki IMRAD bölümünden oluşan makalelerden sonra en yaygın kullanım, dört IMRAD bölümünün tamamına yer verilmiş olan gruptur. Buna göre, Tablo 5'te görüldüğü gibi, derlemi oluşturan 421 makale içinde 179 tanesi tüm IMRAD bölümlerine sahiptir.

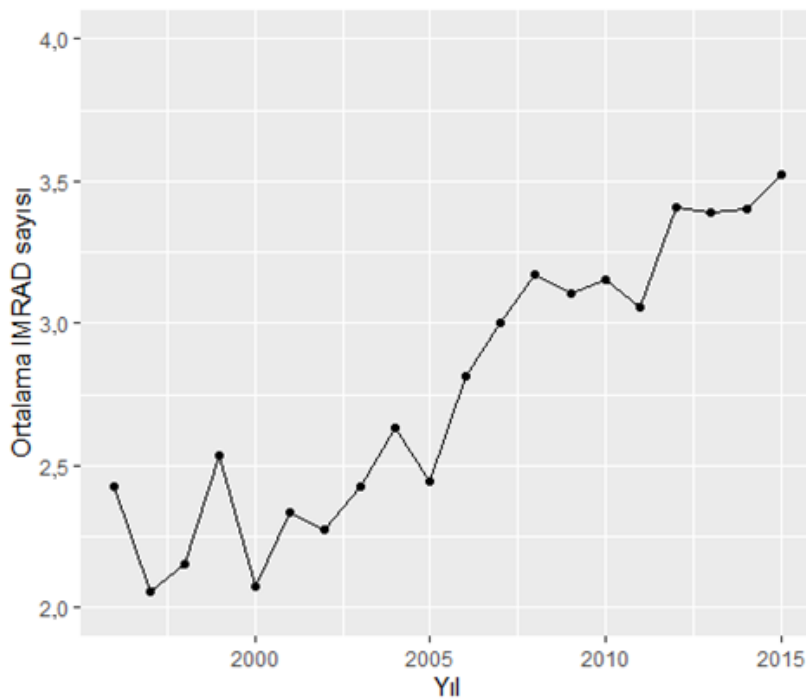
Tek IMRAD bölümünden oluşan makaleler arasında sadece bulgulardan oluşan bir makalenin ve üç IMRAD'lı makaleler arasında yöntemi olmayan 25 makalenin bilimsel araştırmanın doğasına uygun bir yapısal düzene sahip olmadığı ilk bakışta söylenebilir. Sadece bulgular bölümünden oluşan bir akademik makalenin yapısal olarak sorunlu olduğu oldukça açıktır. Derlemdaki üç IMRAD bölümü içeren, yöntemsiz 25 makale özelinde, bulgular ile tartışma ve sonuçları olan bir akademik makalenin araştırma aşamasında belirli bir yöntemin de takip edilmiş olması gerekliliğinden yola çıkarak yine önemli bir yapısal sorunun varlığından bahsetmek mümkündür. Önceki bölümlerde değerlendirildiği gibi bilimsel makalelerde yöntemin açıkça yazılmaması literatürde de sık dile getirilen bir sorundur.



Şekil 15. Tam metinlerde rastlanan yapısal düzenlerin makale sayılarına göre ağırlıkları

Şekil 15'te y ekseninde yer alan kısaltmalarda G, "Giriş"; Y, "Yöntem"; B, "Bulgular; T ve S, "Tartışma ve Sonuçlar"a karşılık gelmektedir. Buna göre, Tablo 5'de giriş, yöntem, bulgular ve tartışma ve sonuçlar olmak üzere tüm yapısal bölümlere rastlanan 179 makale Şekil 15'e göre tüm derlemin %42,5'ine karşılık gelmektedir. Tablo 5'e göre 193 makalede görülen giriş ile tartışma ve sonuçlardan oluşan ikili IMRAD kullanımını ise, Şekil 15'e göre derlemin %45,8'ini oluşturmaktadır.

Şekil 15'e göre, tam metin IMRAD düzenleri arasında (G,B,T ve S) ile gösterilen Tablo 5'teki yontemsiz üçlü IMRAD kullanılan 25 makale derlemin yaklaşık %6'sını oluşturmaktadır. Bir ve üç IMRAD bölümünden oluşan makalelerin tamamı ise derlemin %11,5'ini oluşturmaktadır. Bu oranlardan yola çıkarak derlem genelinde rastlanan belirgin yapısal sorunların nispeten az rastlanır oldukları söylenebilir. Derlem bağlamındaki makale içeriklerinin uzman görüşlerine göre IMRAD'a uyumunun yıllara göre değişimini veren Şekil 16'ya göre, 2000'lerin ilk yıllarından itibaren makalelerde kullanılan yapısal bölüm sayısı kullanımının artan bir eğilime girdiği görülmektedir.



Şekil 16. Makalelerdeki yapısal bölüm düzeninin yıllara göre değişimi

Türkçe kütüphanecilik ve bilgibilim alanında yapısal şekilde yazma alışkanlığı üzerine elde ettiğimiz bulguları yazarların yazım alışkanlıkları, dergilere göre makalelerin yapısal düzenleri ve alana özgü yapısal düzen önerileri üzerinde devam ettirmek mümkündür. Bu yönde ileri çalışmaların faydalı olacağı görüşü ile bu çalışma çerçevesinde elde edilen bulguları değerlendirecek olursak, genel hatları ile alandaki yapısal düzen kullanımının geliştirilen sistem için yeterli veriyi sağlayacak düzeyde olduğunu söyleyebiliriz. Derleminde literatürdekine benzer şekilde özellikle yöntem bölümünün eksikliği sorun olarak göze çarpmaktadır. Ancak makalelerin yarıya yakını, dört IMRAD bölümünden

oluşmaktadır. Tüm IMRAD bölümlerine sahip olan bu makaleler otomatik öz üretiminde yapısal düzene uygun seçim yapmanın etkisini test edebilecek düzeydedir (%42,5). Bir, iki ve üç IMRAD bölümü içeren makalelerin ise yapısal düzenin çıktılara etkisini değerlendirme aşamasında kullanışlı olacağı düşünülmektedir.

Yıllara göre yazarların yapısal yazma alışkanlıklarında artış gözlenmesi, alana özgü yapısal içerikler üzerinde çalışmanın faydalı olduğu düşüncemizi desteklemektedir. Yapısal yazma alışkanlıklarındaki bu artış ilerleyen yıllarda devreye girecek bilgi teknolojileri ile daha iyi işlenecek metin verisinin işareti olarak da kabul edilebilir.

Buraya kadar olan incelemeler ışığında Türkçe kütüphanecilik ve bilgilim alanında yayımlanmış makalelerden oluşan derlemin, tam metinlerin yapısal içeriğini IMRAD'a göre işleyen bir otomatik özetleme sistemi geliştirmeye uygun olduğu değerlendirilmektedir.

### **4.3. TÜRKÇE KÜTÜPHANECİLİK VE BİLGİBİLİM ALANINDA YAYINLANMIŞ MAKALELERİN ÖZ-TAM METİN OKUNURLUKLARI**

Akademik makale özlerinin okunurluklarının bilimsel iletişime engel oluşturacak düzeyde zor yazıldığı önceki bölümlerde değerlendirilmişti. Bu bölümde öncelikle derlem genelindeki tam metin, öz okunurluk düzeyleri analiz edilmiştir. İlerleyen bölümlerde bu bulguların sistem çıktılarının okunurluk değerleri ile karşılaştırılması yapılmaktadır. Bu bölümde ilgili değerlendirmenin ön hazırlığı olarak tam metin ve özlerin okunurluğuna etki edecek faktörler ve okunurluk değerleri karşılaştırmalı olarak ele alınmaktadır. Karşılaştırma dergilere ve okunurluk düzeylerine göre olmak üzere tam metin ile orijinal özler arasında gerçekleştirilmiştir.

Derlemi oluşturan makalelerin tam metinlerinin ve özlerinin okunurluklarını analiz etmek için İngilizce için dünyada yaygın olarak kullanılan FRE okunurluk ölçeğinden Türkçe için uyarlanmış ölçek kullanılmıştır. Hesaplamanın detayları yöntem bölümünde ele

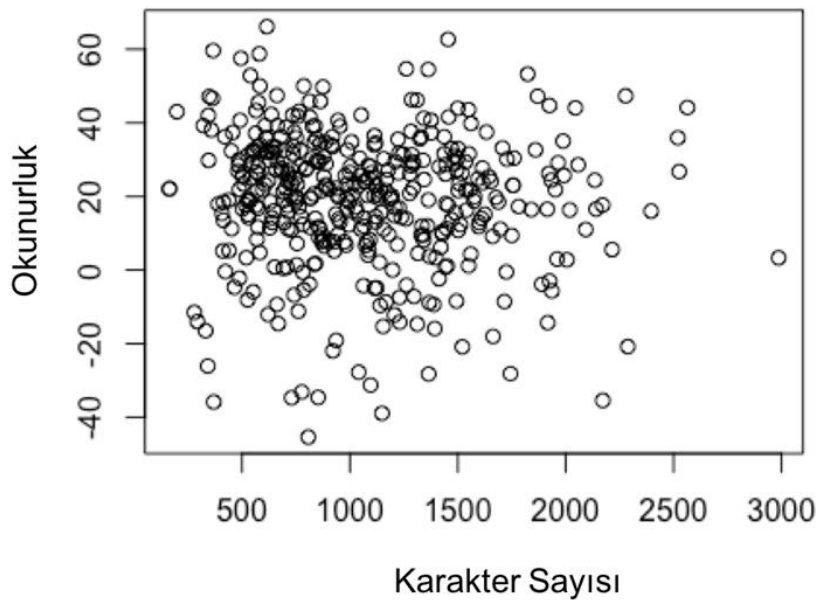
alınmıştır. Bu hesaplamada kullanılan kelime sayıları, cümle sayıları ve Ateşman Okunurluk Değerleri (AOD) ile karakter sayılarına ait özet Tablo 6’da verilmiştir.

AOD, dokümanları okunurluk ölçeğine göre sınıflayarak metnin okunurluğu hakkında yorum yapmayı mümkün kılmaktadır. Bu ölçeğe göre AOD, 0 ile 29 arasında iken metnin okunurluk değeri “çok zor”, 30 ile 49 arasında iken metnin okunurluğu “zor”, 50 ile 69 arasında iken okunurluk değeri “orta zorlukta”, 70 ile 89 arasında “kolay” ve 90 ile 100 arasında “çok kolay”dır. Bu doğrultuda, Tablo 6’ya göre derlemin tam metninin okunurluk değerleri “çok zor” ile “kolay” arasında değişmektedir. Makalenin yazarı tarafından yazılmış klasik özlerin okunurluk değerleri ise 0’dan küçük değerlerden başlayarak “çok zor” ve “zor” olarak sınıflanan değerlere kadar yükselmektedir.

Tablo 6. Makale tam metin ve orijinal özlerinin okunurluk özet tablosu

			<b>Cümle S.</b>	<b>Kelime S.</b>	<b>Karakter S.</b>	<b>Okunurluk</b>
Öz	Bilgi Dünyası	En Küçük	2	34	321	-38,94
		Ortalama	6	120	1140	18,09
		Ortanca	8	116	1150	20,86
		En Büyük	18	279	2564	57,55
	Türk Kütüphaneciliği	En Küçük	1	17	163	-45,39
		Ortalama	6	107	1019	19,93
		Ortanca	5	95	878	21,71
		En Büyük	20	301	2987	66,16
	Derlem	En Küçük	1	17	163	-45,39
		Ortalama	6	113	1069	19,18
		Ortanca	6	105	995	21,58
		En Büyük	20	301	2987	66,16
Tam Metin	Bilgi Dünyası	En Küçük	97	1471	13	6,6
		Ortalama	296	4597	42	37,52
		Ortanca	280	4404	40	36,97
		En Büyük	719	11	90	61,11
	Türk Kütüphaneciliği	En Küçük	49	695	6164	6,72
		Ortalama	300	4541	41	38,93
		Ortanca	267	4084	37	39,23
		En Büyük	979	12	2987	85,24
	Derlem	En Küçük	49	695	6164	6,6
		Ortalama	298	4564	42	38,56
		Ortanca	272	4228	38	38,56
		En Büyük	979	11.661	100	85,24

Negatif okunurluk değerleri Ateşman Okunurluk Ölçeği'nde yer almamaktadır. Yani, hesaplanan katsayılar formülün tanımlanan aralıklarının dışında kalmaktadır. İlk bakışta bu durumun öz metinlerinin kısalığı sebebiyle gerçekleştiği düşünülebilir. Derlemi öz uzunlukları ve okunurluk değerlerine göre incelediğimizde, en uzun öze sahip ilk iki *BD* makalesinden 2564 karakterli olan özün okunurluğu 44,18 (zor) hesaplanırken 2288 karakterli olan ikinci en uzun özün okunurluk değerinin ise -20,80 olduğu görülmektedir. Şekil 17'de verilen, derlemin tamamındaki özlerin okunurluklarının karakter uzunluklarına göre saçılım grafiği incelendiğinde ise negatif okunurluk değerlerinin tüm karakter uzunluklarındaki özlerde hesaplanabildiği gözlenebilir.



Şekil 17. Öz okunurluklarının karakter sayısına göre saçılımı

Çalışma kapsamında dilbilimi alanına bir katkı yapılması hedeflenmediği akılda tutularak bu durumu irdelemeye devam edeceğiz olursak; aynı alanda yazılmış bilimsel yayınlara ait özlerin geneli kısa, yoğun jargon kullanımlı, eşit derecede yabancı kelime içeren metinler olarak kabul edilebilir. O halde okunurluk formülleri haricinde metin özelliklerinden kaynaklanan dikkate alınabilecek diğer unsurlar eşit olarak kabul edilebilir özellikler sergilemektedir. Ayrıca derlemimizi oluşturan iki derginin özlerinin yapısal biçimleri ve sunum özellikleri de temelde birbirinden farklı değildir. Bu durumda okunurluk hesaplamaları -zayıf yönleriyle beraber- yine de okunurluğa dair dikkate alınabilecek



anamlı bir gösterge olarak göze çarpmaktadır. Aşağıda *TK* ve *BD* dergilerinden her bir AOD'ye sahip örnekler sunulmuştur. Negatif AOD'ye sahip özlere göre daha uzun ve peş peşe sıralanan jargonlardan oluşan cümleleriyle bu görüşü destekleyen izlenimi kuvvetlendirecek özelliktedir. Özellikle -45,39 AOD ile derlemin en düşük okunurluğuna sahip hesaplanan *TK* makalesine ait özde 57 kelimedenden oluşan bir cümleye rastlanmaktadır. -38,94 okunurluk değeri ile en düşük ikinci okunurluk değerine sahip olan *BD* makalesinin özünde de benzer şekilde 69 kelimedenden meydana gelen bir cümle mevcuttur. Her iki cümlede de Türkçe dilbilgisi kurallarına uygunluk ve kolay okunurluk değil; çok bağlaçlı, uzun ve devrik yazma alışkanlığı göze çarpmaktadır. Diğer özlere göre her bir özün okunurluk değeri arttıkça kısaltılmakta ve dilbilgisi kuralları yönünden iyileşmektedir.

TK örneği, AOD=-45,39

“Batı ülkelerinde ve Türkiye'de halk kütüphanelerinin mevcut durumu ve uygulanan politikalar medyada yer alan haberler, resmi politika belgeleri ve literatüre dayalı olarak gözden geçirilmiş ve Türkiye'nin halk kütüphanesi politikası irdelenmiştir. Ana hatları ile Avrupa ile aynı paralelde hareket ettiği izlenimi vermekle birlikte, Türkiye'de geçmişten gelen boşluğun göz ardı edilmesi ve halk kütüphanelerinin toplumsal içe alma, varsılla yoksul arasındaki uçurumu azaltma, halkı yaşam boyu öğrenme olanaklarından yararlandırma gibi konularda oynayabileceği sosyo-ekonomik rolden çok, ağırlıklı olarak, kültürel boyuta yoğunlaşmaları nedeniyle bilgi toplumuna dönüşüme potansiyel katkılarının yeterince değerlendirilmediği sonucuna varılmıştır.” (57)

BD örneği, AOD=-38,94

“Bu çalışmanın ana amacı, temel enformasyon okuryazarlığı standartlarına dayanarak enformasyon profesyonellerinin Webde enformasyon okuryazarı olabilmelerinin genel koşullarını ortaya koymaktır. Günümüzde hızla değişen ve gelişen enformasyon alanında ve özellikle Internet ortamında, türleri ve sayıları artan kütüphane ve enformasyon örgütlerinin (elektronik, sayısal ve sanal kütüphanelerin) yer aldığı ve yer alacağı düşünüldüğünde, Web

üstünden kullanıcının enformasyon gereksinimini karşılamak durumunda bulunan enformasyon profesyonellerinin, Web sitesi içerik ve tasarımı planlama ve yönetme dahil tüm eylemlerini kaliteli, doğru ve güvenilir olarak gerçekleştirebilmesi, Webde enformasyon okuryazarı olma koşullarını bilmesi ve onları yerine getirmesi ile mümkündür. Çalışmada, Association of College and Research Librariesin yükseköğretim için geliştirdiği temel enformasyon okuryazarlığı standartlarına dayanarak geliştirilen bu koşullar, enformasyon profesyonellerinin Webde enformasyon okuryazarı olabilmelerinin genel koşulları olarak kabul edilebilir.” (69)

BD örneği AOD=25,70, Çok Zor

“Kitap ilk sayfalarından hatta satırlarından başlayarak papirüs, parşömen tahta veya metal tabakalar üzerine yazılmış, rulo halinde ya da ciltlenmiş olsun iki yönüyle ele alınması gereken bir araç -bilgi taşıyıcısı olmuştur. Kitap, yazılmış ya da basılmış yaprakların bir araya getirilmesinden oluşan ve bir konuyu belirli bir düzen içinde sunan yapıt olarak tanımlanmaktadır. Kitabı oluşturan bu iki unsur, onun içeriği ve fiziksel yapısıdır. İçerik, anlatılmak istenen konuyu/temayı yani kitabın amacını ifade eder, bilgilendirici, öğretici, yol gösterici, eleştirici olanı aktarır. Bilinen en eski el yazması kitapların günümüzden yaklaşık 3000-3700 yıl öncesine tarihlenen ve Hinduizm'in ilk kutsal metinlerinin yer aldığı Rig Veda'dan günümüzün elektronik kitaplarına kadar geçen süreç içerisinde her ne biçimde ve içerikte olsun “kitap” kavramı zihinlerimizde yer etmiştir. Dünyanın neresinde olursa olsun hangi dil ile söylenirse söylensin “kitap” her zaman aynı fiili ve formu ifade etmektedir. Günlük yaşantımızın son 30 yılında hemen her alanda bireylerin yaşantısını değiştiren/kolaylaştıran teknoloji, uzun bir geçmişe sahip olan kitap için de pek çok değişimi ortaya koymuştur. Ancak bu değişim kavramlardan ve kitabın bileşenlerinden çok ona erişimi, onu üretmeyi/yazmayı ve ondan faydalanmayı daha keyifli kılmayı amaçlamaktadır. E-kitap ile ilgili son 10 yıl içerisinde başta eğitim olmak üzere, elektronik kaynaklar, kütüphaneler, bilgi hizmetleri, okuma alışkanlığı, iletişim gibi alanlarda çok sayıda akademik ve sosyal çalışmalar

yapılmıştır. Bu çalışmada e-kitap teknolojisi, temel bileşenleri olan yazar, yayıncı ve okuyucu ile beraber gündelik yaşamda bireylerin kitap anlayışındaki farklılaşma, görsel ve yazılı basın örnekleri 2 bağlamında farklı bir bakış açısı ile değerlendirilecektir.”

TK örneği, AOD=33,98, Zor

“Okuma alışkanlığı bireysel ve toplumsal gelişme ölçütleri arasında yer almaktadır. Üniversite öğrencileri güçlü okuma alışkanlığına sahip olması beklenen bir gruptur. Bu çalışma, Hacettepe Üniversitesi Tıp Fakültesi ve Bilkent Üniversitesi Güzel Sanatlar, Tasarım ve Mimarlık Fakültesi'nde öğrenim gören dördüncü sınıf lisans öğrencilerinin kitap okuma alışkanlığı düzeylerini ve bu alışkanlıklarına ilişkin çeşitli özellikleri belirlemek amacıyla yapılmıştır. Bu çerçevede, bu üniversitelerin ilgili fakültelerindeki toplam 104 öğrenciye, 2008 Aralık ayında anket uygulanmıştır. Araştırma sonucunda her iki üniversite öğrencilerinin de zayıf bir okuma alışkanlığına sahip oldukları anlaşılmıştır. Ayrıca, öğrencilerin düzenli okumadıkları, okuyacakları kitapları daha çok satın aldıkları görülmüştür. Öğrenciler, genelde zaman yetersizliği nedeniyle okuyamamaktadırlar. Kütüphane kullanma öğrencilerin okuma alışkanlığı ile bağlantılı olarak gerçekleştirdikleri bir etkinlik değildir.”

BD örneği AOD=57,55, Orta Zorlukta

“Bu çalışmada Tebriz Üniversitesinde 2005-2009 yılları arasında tamamlanan 627 tezin atıf analizi yapılmıştır. Seçilen 627 tezde yer alan toplam 44.488 atıf, yayınların türü, dili ve yılı gibi özellikler açısından incelenmiştir. Bu özelliklerin zaman içinde ne gibi farklılıklar gösterdiği araştırılmıştır. Çalışmada ayrıca, tezlerde yer alan atıfların dönemler itibarıyla yaşlanma hızlarında bir farklılık olup olmadığı araştırılmıştır.”

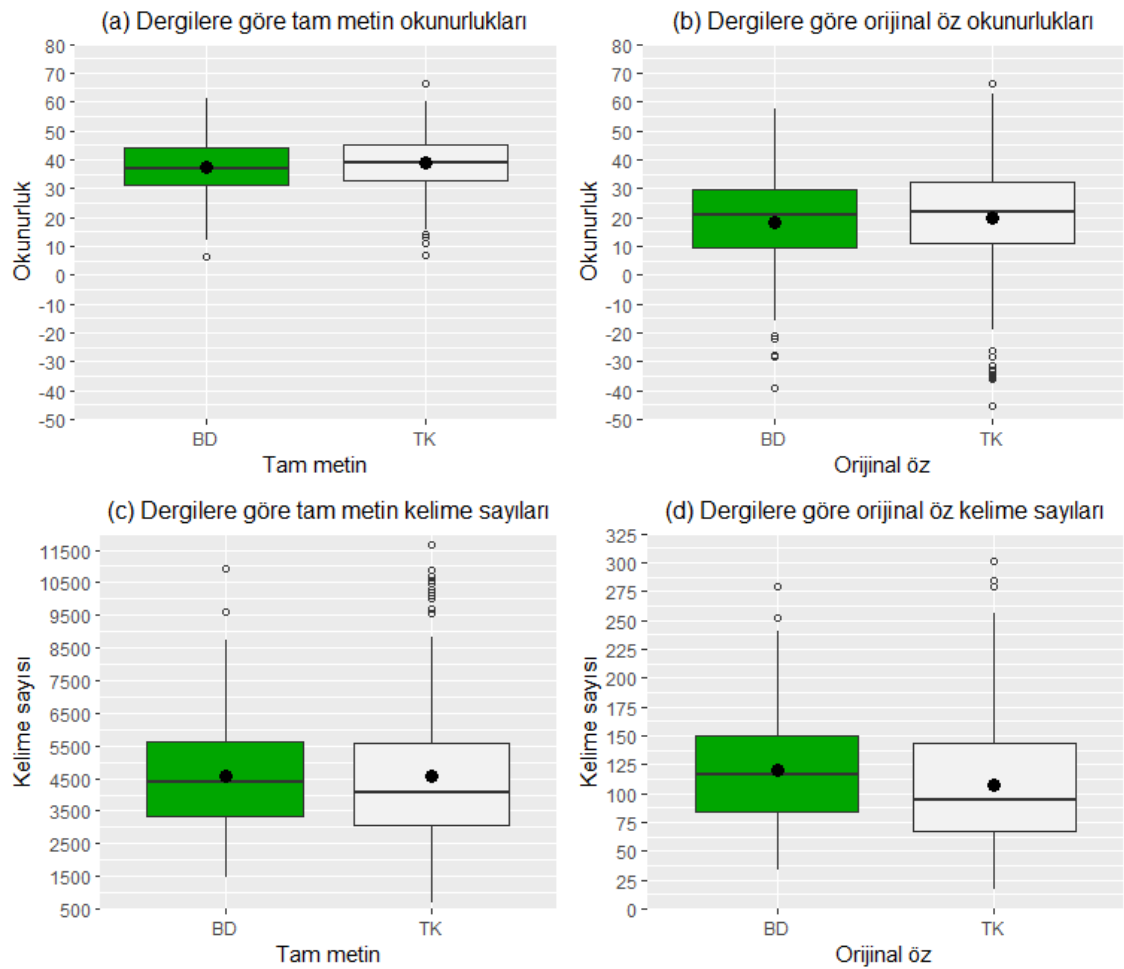
Bu koşullar altında negatif AOD'nin okunurluk için iyi bir işaret olmayacağı kabulüyle metnin okunurluğu açısından “çok zor”dan bir sonraki zorluğa işaret ettiği kabul

edilmiştir. Hesaplamalarda negatif okunurluk değerleri “negatif” başlığı altında gruplanmıştır.

Bu kabulün ardından *BD* ve *TK* dergilerinden oluşan derlemin makale okunurluklarının ve kelime sayısına göre uzunluklarının dergilere göre farklılık gösterip göstermediği incelenmiştir. Bu değerlendirme için her iki dergiye ait makalelerin kelime sayısına göre uzunlukları ve okunurluk değerleri (AOD); dergilere göre, yazar tarafından üretilmiş orijinal özleri ve tam metinleri için ayrı ayrı karşılaştırılmıştır.

Şekil 18 (c)’de her iki derginin tam metinlerindeki toplam kelime sayılarının birbirine çok yakın olduğu ancak Şekil 18 (d)’de *TK* dergisinin orijinal özlerinin *BD* dergisinden az daha düşük toplam kelime sayılarında da yığılım sergiledikleri görülmektedir. Aradaki bu fark kutu grafiklerinde siyah nokta ile belirlenen ortalama değerlerinde biraz daha azalmıştır. Ayrıca *TK* dergisinin kelime sayılarını gösteren kutu grafiğine göre dağılımın yönü *BD* makalelerine ait özlerin yığıldığı bölge ile örtüşecek şekilde pozitif çarpıktır. *TK* özlerinin uzunluk dağılımının *BD* özlerine göre daha heterojen olduğu göze çarpmasına rağmen çalışmanın ilerleyen bölümlerinde planlanan hesaplamaları etkileyecek bir farklılık olduğu düşünülmemiştir.

İki derginin orijinal öz ve tam metin okunurluk değerlerinin (AOD) yığılımlarını incelediğimizde (Şekil 18 (a) ve (b)) kelime sayılarına göre tespit ettiğimiz küçük farklılığın daha da azaldığı görülmektedir. Şekil 18 (b)’de verilen, her iki derginin orijinal öz okunurlukları ve Şekil 18 (a)’daki tam metin okunurlukları benzer şekillerde dağılım göstermiştir. Tablo 6’daki ham değerlere göre incelendiğinde de sonuç aynıdır. Dergilere göre orijinal öz ve tam metin okunurluklarının hem ortalama hem de ortanca değerleri birbirine çok yakındır. Bu doğrultuda, her iki derginin okunurluk değerleri arasında kayda değer bir farklılık olmadığı gösterilmiştir.



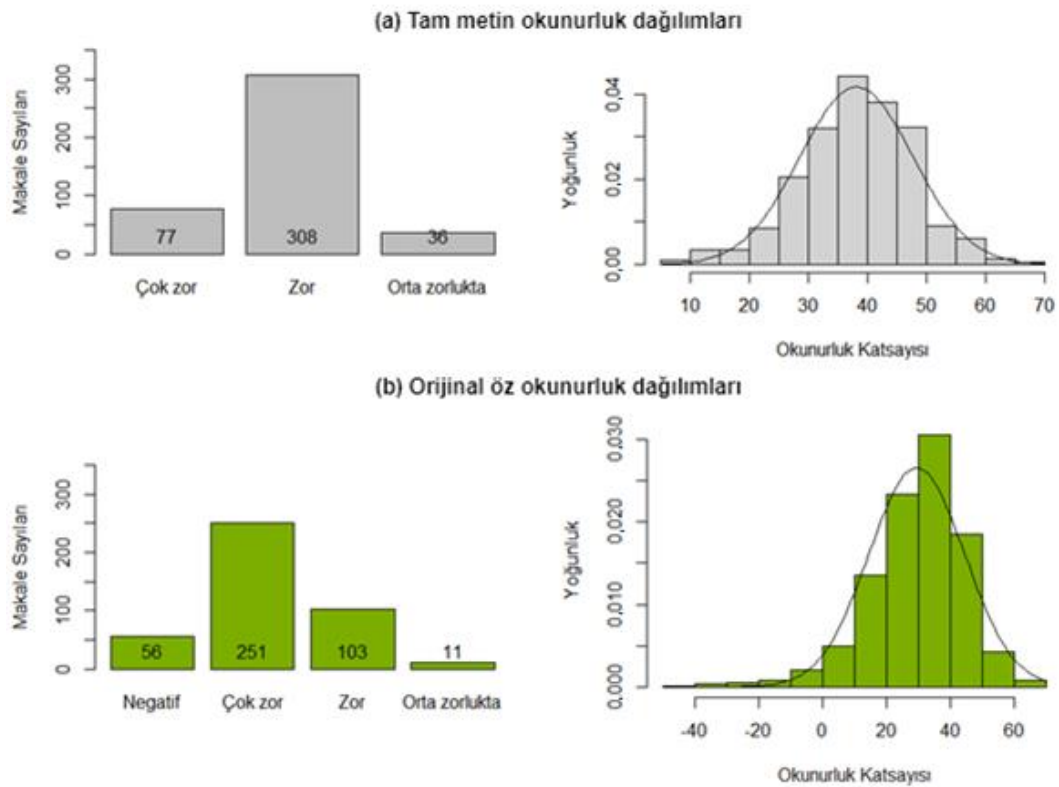
Şekil 18. Tam metin okunurluk (a), orijinal öz okunurluk (b), tam metin kelime sayısı (c), orijinal öz kelime sayısı (d) yığılımları

*BD* ve *TK* dergilerinin orijinal özleri ve tam metinleri ile hesaplanan okunabilirlik düzeylerinin derlemdeki dağılımları ise Şekil 19'daki gibidir. Makalelerin hem orijinal özleri, hem de tam metinleri okunabilirlik düzeylerine göre yaklaşık simetriktir.

Şekil 19 (b)'de verilen orijinal özlerin okunurluk dağılımlarında AOD'ye göre "çok zor" olarak sınıflanan makale sayısı 251'dir ve bu da derlemin %60'una karşılık gelmektedir. Orijinal özünün okunurluğu en kolay/en yüksek çıkan makalelerin sayısı ise 11'dir ve bu makaleler AOD'ye göre "orta zorlukta" sınıflanmaktadır. Orijinal özlerde negatif okunurluk değerleri hesaplanmış makaleleri de "çok zor" okunurluk değerine sahip makalelere eklediğimiz takdirde, derlemdeki 421 makalenin orijinal klasik özlerinden %73'e karşılık gelen 307 tanesi AOD'ye göre "çok zor" okunurlukludur denilebilir.

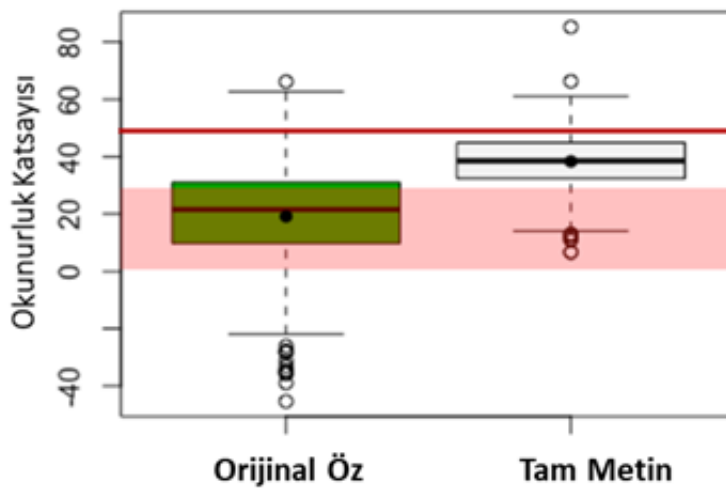
Türkçe kütüphanecilik ve bilgibilim literatüründe yer alan makalelerin orijinal özleri arasında AOD'ye göre “kolay” ya da “çok kolay” okunurluklu bir makaleye rastlanmamıştır. “Orta zorlukta” sınıflanan makale özlerinin yüzdesi ise sadece %2,6’dır.

Şekil 19 (a)’da görüldüğü gibi, makale tam metinlerinde de “kolay” ve üzeri okunurluk değerine sahip makaleye rastlanmamıştır. 77 makalenin tam metni “çok zor” okunurluk değerine sahiptir. 421 makalenin 308 tanesinin tam metni “zor” okunurluklu olarak hesaplanmıştır. Buna göre “zor” okunurluklu makale yüzdesi tüm derlemin %73’lük baskın çoğunluğuna karşılık gelirken, “çok zor” okunurluklu makaleler derlemin %18’i, “orta zorlukta” makaleler derlemin %8,5’ine karşılık gelmektedir. Derleme ait makalelerin yazar tarafından üretilen orijinal klasik özlerinin %73’ü “çok zor” ve “çok zor”dan daha da zor okunurluklu kabul ettiğimiz “negatif” okunurluk değerine sahiptir. Tam metinlerin ise yine %73’ü AOD’ye göre “zor” okunurluklu olarak sınıflanmıştır. Yazar özlerinin okunurlukları “çok zor”da yığılırken tam metinlerinin okunurlukları bir alt kolay okunurluk düzeyi olan “zor”da yığılmıştır.



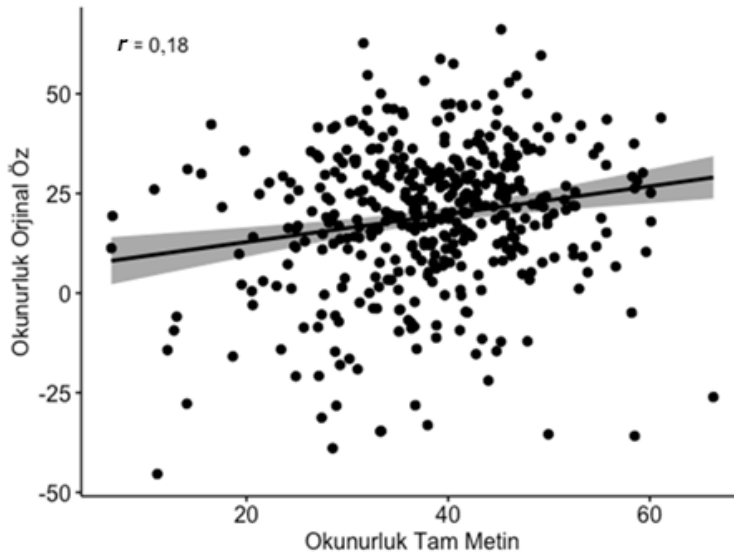
Şekil 19. Tam metin (a), orijinal öz (b) okunurluk dağılımları

Makalelerin orijinal özleri ile tam metnlerinin okunurlukları arasındaki farkı kutu grafikleri ile gösteren Şekil 20, derlem genelinde yazar tarafından üretilen orijinal özlerin okunurluklarının tam metnlerinden daha zor olduğunu açıkça ortaya koymaktadır. Bu bulgu literatüre benzer şekilde Türkçe kütüphanecilik ve bilginin alanında da yazarlar tarafından üretilmiş orijinal klasik özlerin okunurluk değerlerinin tam metnlerinden daha zor olduğunu göstermektedir. Şekil 20’de pembe taralı alan AOD’ye göre “çok zor” olarak sınıflanan değerlere karşılık gelmektedir. Taralı alan ile kırmızı çizgi arasında kalan renksiz alan ise AOD’ye göre “zor” olarak sınıflanan okunurluk düzeyinin yer aldığı alandır. Buna göre klasik özlerin okunurluklarının baskın çoğunluğu “çok zor” olarak sınıflanırken gövde metnlerinin neredeyse tamamı okunurluk düzeylerine göre bir kolay okunurluğa karşılık gelen “zor” sınıfında yer almıştır.



Şekil 20. Tam metin, orijinal öz karşılaştırmalı okunurluk değerleri

*BD* ve *TK* dergilerinde yayınlanan makalelerin özlerini de tam metinleri gibi yazarlar kendileri üretmektedir. Ancak, Şekil 21’de görüldüğü üzere, makalelerin tam metin ve öz okunurlukları için hesaplanan AOD değerleri arasında istatistiksel açıdan güçlü bir ilişki görülmemektedir. Başka bir ifadeyle makalelerin tam metnlerinin ve özlerinin okunurluk katsayıları her zaman birlikte azalıp artmamaktadır. Buna göre, yazarların makalelerin tam metin ve özlerinin okunurluklarını etkileyecek unsurlar bağlamında benzer bir yaklaşım sergilemediği anlaşılmaktadır.



Şekil 21. Tam metin, orijinal öz AOD ilişkisi

Bu doğrultuda, derlem bağlamında yazarların, makale özlerinde okunurluk ölçeğine göre “çok zor”, tam metinlerinde ise daha kolay bir okunurluk düzeyi olan “zor” yazma eğiliminde oldukları anlaşılmaktadır. Yazarların tam metinde ve özlerde kullandıkları dil özelliklerinin hesaplanan okunurluk değeri üzerindeki etkisi birlikte azalıp artmamaktadır. Yani yazarlar tam metinlerde kullandıkları dil özelliklerinden bağımsız olarak öz yazımında zor okunurluklu bir dil tercih edebilmektedirler.

#### 4.4. OTOMATİK YAPISAL ÖZLERİN DEĞERLENDİRİLMESİ

Bu bölümde çalışmanın araştırma sorularına paralel olarak, geliştirilen otomatik özetleme sisteminin çıktıları üç alt bölümde değerlendirilmiştir. Bunlardan ilkinde, sistem çıktılarını oluşturan cümlelerin tam metinler içindeki konumları incelenmiştir. İkinci aşamada otomatik özlerin okunurlukları tam metin ve orijinal özler ile karşılaştırmalı olarak analiz edilmiştir. Son olarak, çıktısı başarımları, yazar tarafından üretilen orijinal özler ile ROUGE ölçümlerine göre n-gram çakışmaları hesaplanarak karşılaştırılmıştır.

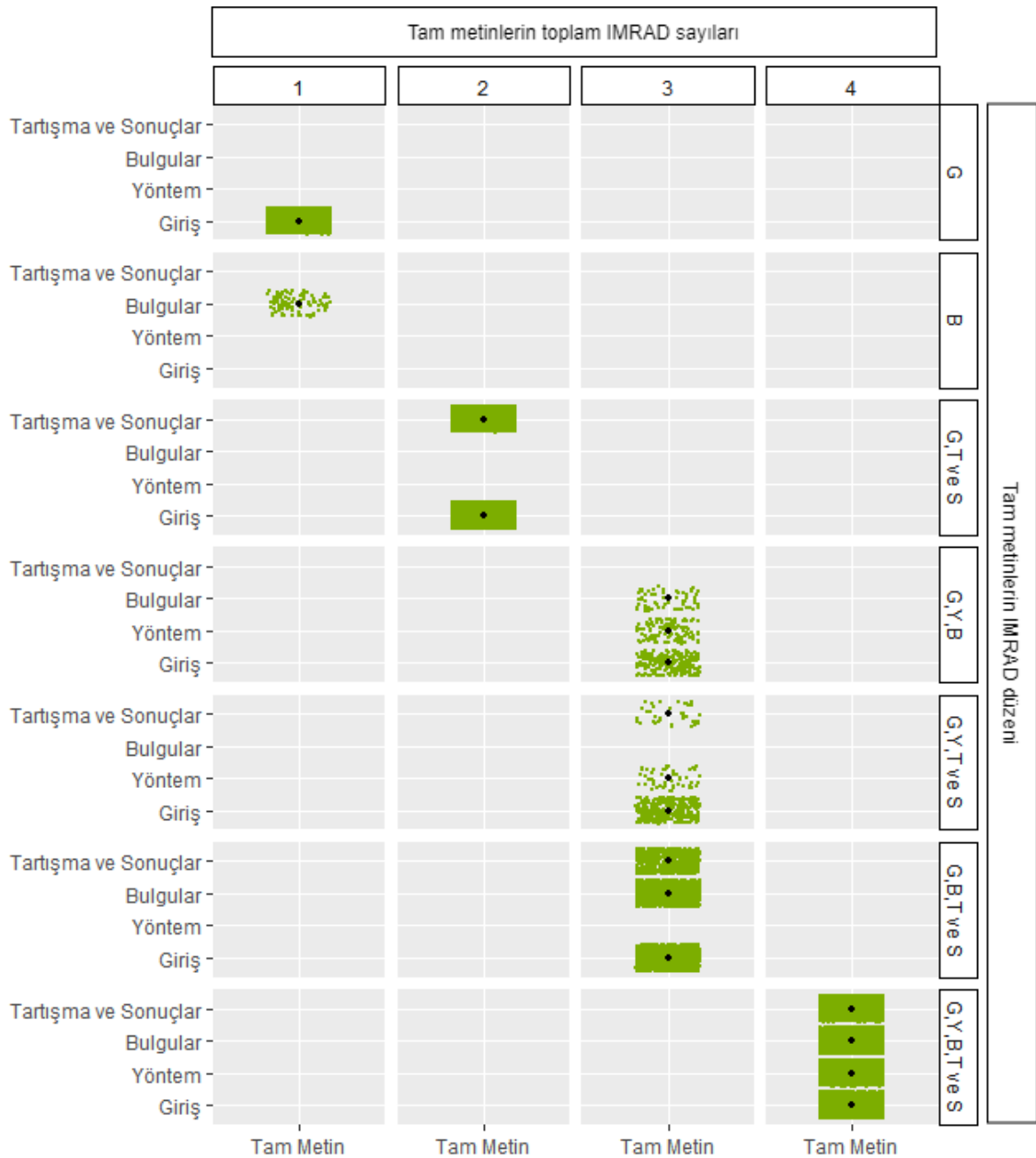


#### 4.4.1. OYÖ ve OKÖ Cümlelerinin Tam Metin İçerisindeki Dağılımları

Daha önceki bölümlerde ele alındığı gibi çıkarıma dayalı otomatik özetleme yöntemleri özet cümlelerini tam metinlerden olduğu gibi alıp kullanmaktadırlar. Bu doğrultuda, OYÖ ve OKÖ cümlelerinin tam metin içindeki dağılımlarını incelerken, cümlelerin tam metindeki IMRAD bölümleri göz önüne alınarak çıktılarının tam metni tarama düzeyine karar verilebilmiştir. Şekil 22, incelediğimiz derlem tam metinlerinin cümlelerini her bir makalede görülen IMRAD bölümü sayılarına ve bu IMRAD bölümlerinin birliktelik desenlerine göre gruplamaktadır.

Derlem bağlamındaki makalelerde görülen IMRAD bölümü düzenleri Tablo 5'te "IMRAD düzeni" sütununda incelenebilir. Şekil 22'de sağ dış kenardaki etiketlerde görünen kısaltmalar bu düzenin açılımını ifade etmektedir. Üst kısımdaki rakam etiketleri ise makalelerin sahip oldukları 1, 2, 3 ve 4 IMRAD bölümü sayılarını belirtmektedir. Buna göre grafikteki her bir nokta bir cümlenin ait olduğu makaledeki IMRAD bölümünü y eksenini ile belirlerken, üst kısımdaki rakam etiketleri cümlenin ait olduğu makalenin IMRAD bölümü sayısını gösterir. Noktanın sağ dış kenardaki etiketler ile izdüşümü ise ilgili cümlenin ait olduğu makalenin IMRAD düzeninin açılımını göstermektedir.

4.2. bölümde elde edilen bulgulardaki gibi Şekil 22'de de iki IMRAD bölümünden oluşan makalelerin tamamının giriş ile tartışma ve sonuçlar bölümlerinde yığıldığı göze çarpmaktadır. Üç IMRAD bölümünden oluşan makalelerde en az renklenen alanın yöntem bölümlerinde oluşu yine açıkça gözlenmektedir. Tek bir IMRAD bölümünden oluşan makalelerden sadece bir tanesi bulgular bölümünden oluştuğu için ilgili alan az renklenmiştir. Giriş bölümünden oluşan 19 makale derlemde sayısal olarak üstünlük göstermiyor olsa da, ilgili makalelerin tüm cümleleri giriş bölümünden geldiği için bu alanı koyu renklendirecek cümle yoğunluğu gözlenmektedir.



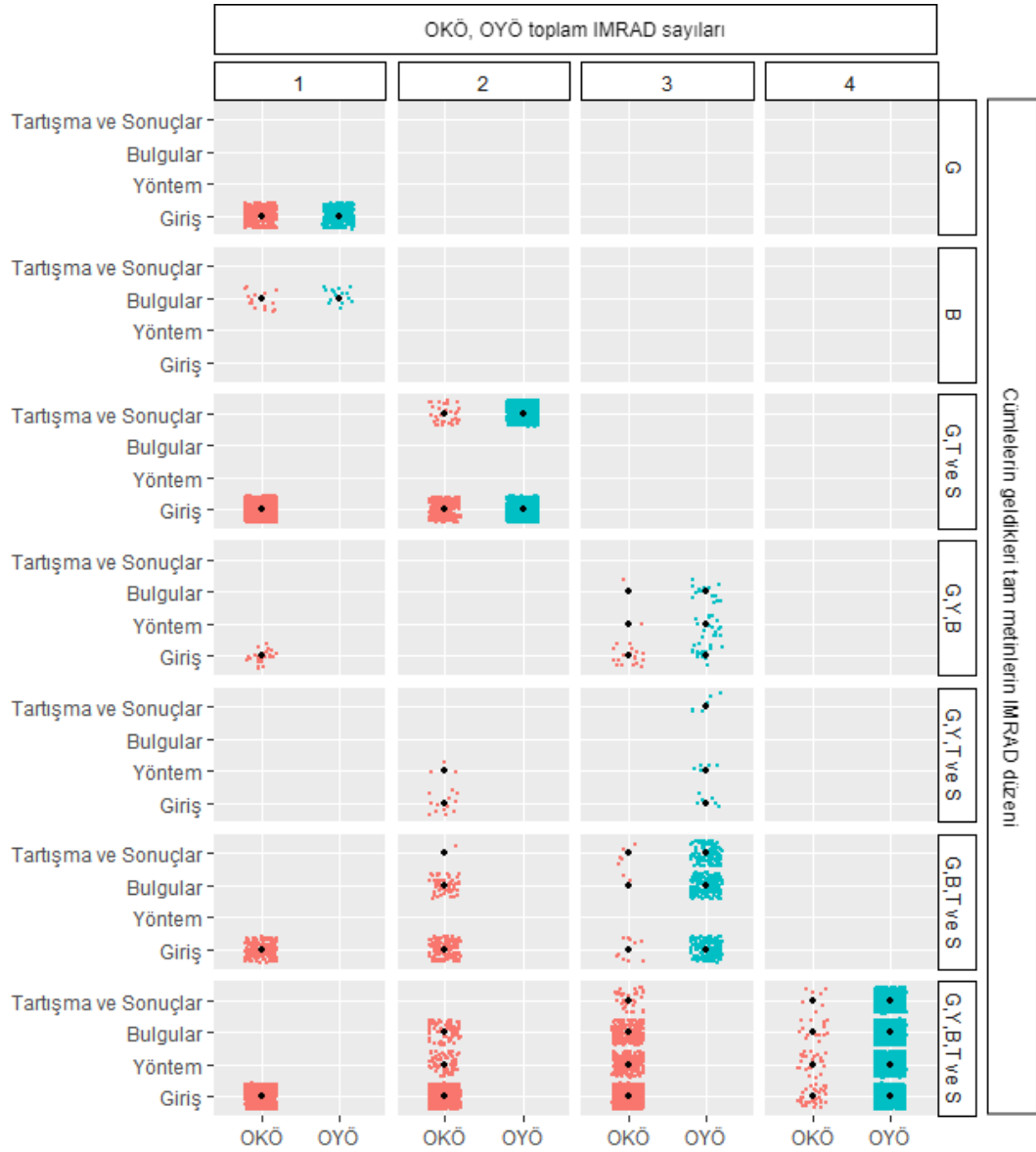
Şekil 22. Tam metin içeriklerinin IMRAD'a göre sınıflandırılması

Şekil 22 tam metin cümlelerini IMRAD'a göre gruplarken, Şekil 23'te her bir noktaya karşılık gelen cümleler otomatik özlere aittir. Şekil 23'te de, benzer bakış açısıyla, her bir cümlenin tam metnin "giriş", "yöntem", "bulgular" ile "tartışma ve sonuçlar" bölümlerinden hangisine ait olduğunu y eksenini belirlemektedir. Grafiğin sağ dış kenarındaki etiketler ilgili cümlenin ait olduğu tam metnin IMRAD düzenini sınıflandırmaktadır. Bu desen yine Tablo 5'te verilen, tam metinde görülen IMRAD düzenleridir. Sütunların üstlerindeki 1, 2, 3 ve 4'ten oluşan etiketler, ilgili cümlenin yer

aldığı özün kaç IMRAD bölümünden cümle seçimi ile oluştuğunu göstermektedir. x eksenini ise OYÖ ve OKÖ'leri gruplandırmaktadır. Kırmızı kutular OKÖ, mavi kutular OYÖ'leri temsil etmektedir. Bu doğrultuda, Şekil 23'te sistem çıktısı özlerin IMRAD düzenlerinin yanı sıra, ait oldukları tam metinlerin IMRAD düzenlerine göre cümlelerinin doküman içindeki dağılımı gösterilmektedir.

Şekil 23'te OYÖ'lerin grafikte belirlediği her alan Şekil 22 ile benzerdir. Bunun anlamı OYÖ cümlelerinin ilgili makalelerin yapısal düzenindeki dağılımlarının tam metnin yapısal düzenlerinin dağılımları ile uyumlu olmasıdır. Örneğin, tam metni iki IMRAD bölümünden oluşan makalelerin OYÖ'leri de aynı bölümlerden benzer yoğunlukta cümle içeriğine sahip olabilmişlerdir. Tam metinlerden cümle seçerken yapısal bölümleri gözetmenin böyle bir sonuç doğurması olağandır. Bu sonuç otomatik öz elde ederken tam metinlerin yapısal özelliklerinin göz önüne alınmadığı durumdaki dağılım ile karşılaştırmalı olarak incelendiğinde anlam kazanmaktadır. Grafikteki kırmızı kutulardan görülebileceği gibi yapısal bölümleri hesaplamalara katmayan OKÖ'ler için cümlelerinin tam metinlerden getirdikleri IMRAD bölümü sayıları, ilgili tam metninin IMRAD düzenindeki yapısal bölüm sayılarından tamamen bağımsız şekillenmiştir. Örneğin OKÖ'ler için 1 numaralı sütun incelendiğinde; tek bir IMRAD bölümünden oluşan OKÖ'lere ait cümlelerin aslında 4 (“G, Y, B, T ve S”), 3 (“G, Y, B”, “G, B, T ve S”), 2 (“G, T ve S”) ve 1 (“B” ve “G”) IMRAD bölümünden oluşan tam metinlerden gelebildiği görülmektedir. Yani düzenleri parantez içinde verilen, 1, 2, 3 ve 4 IMRAD'a sahip tam metinler için elde edilen otomatik klasik özlerde cümlelerin tek bir IMRAD bölümünden geldiği çıktı grupları mevcuttur. Ele alınan örnekte bu tek IMRAD bölümünün de hemen hemen her zaman giriş bölümü olarak gerçekleştiği görülmektedir. Grafikte her bir satırı sağa doğru ilerleyerek incelediğimizde ilgili satırın en sağındaki IMRAD düzenindeki tam metinler için elde edilen çıktıların içerikleri gözlemlenebilmektedir. Örneğin, dört IMRAD'dan oluşan tam metin düzenine ait olan cümleleri gruplayan en alt satırda bu kez “G, Y, B, T ve S” düzenine sahip tam metinlerden elde edilen OKÖ'lerin 1, 2, 3 ve 4 IMRAD bölümünden oluşan çıktılar olabildiği anlaşılmaktadır. Burada üç IMRAD toplamına sahip çıktıların “G, Y, B, T ve S” yani dört IMRAD'a sahip tam metinler ile kesiştiği alanda dört kırmızı kutunun belirmiş olması kafa karıştırabilir. Buradaki dört kırmızı kutu bu bölgedeki OKÖ'lerin tüm IMRAD bölümlerinin üçlü birliktelikleri ile

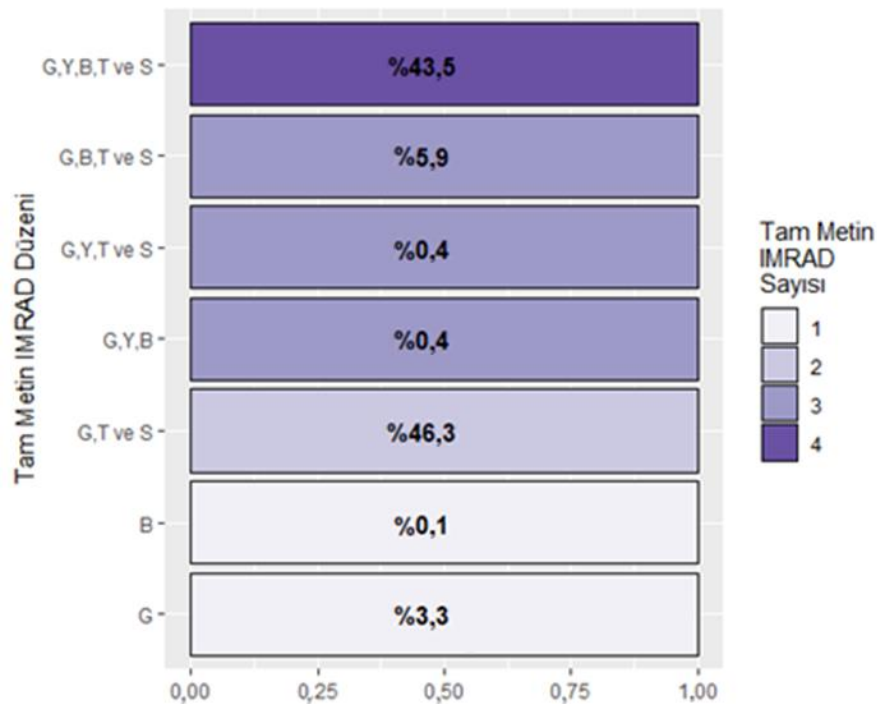
oluşabildiğini göstermektedir. Bu alanda en soluk renkli alanın y ekseninde “tartışma ve sonuçlar”a karşılık gelen kutu olmasından, bu birliktelikler içinde en az rastlanan IMRAD bölümünün “tartışma ve sonuçlar” olduğu anlaşılmaktadır.



Şekil 23. OKÖ (kırmızı), OYÖ (mavi) IMRAD düzenlerinin, tam metin IMRAD düzenlerine göre sınıflandırılması

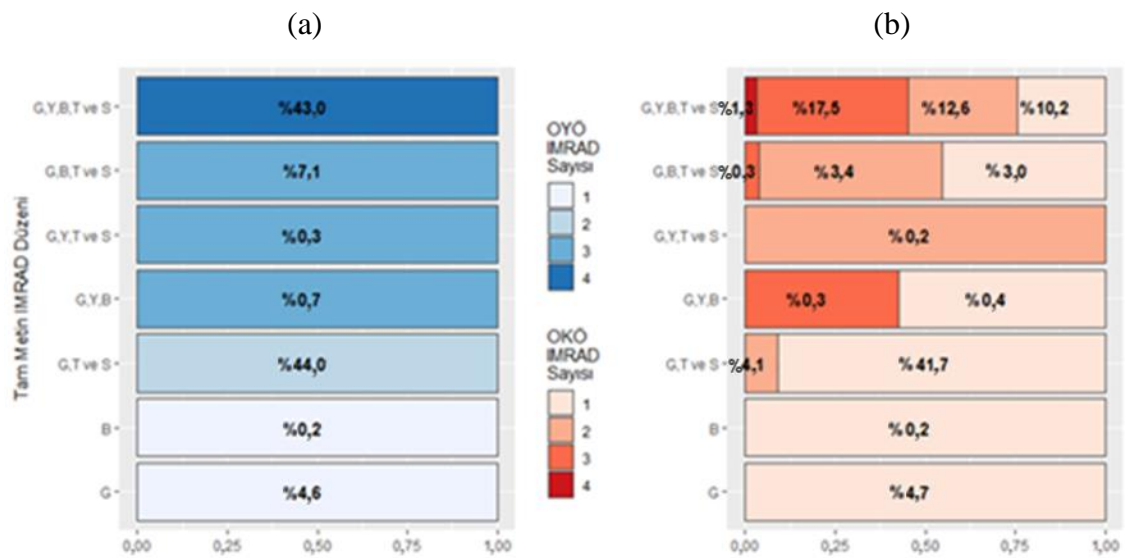
Satırlar ve sütunlar benzer şekilde incelenmeye devam edildiğinde OKÖ'ler için tek IMRAD bölümünden oluşan makaleler dışında tam metnin yapısal içeriği ile tam olarak üst üste gelen bir dağılımın gerçekleşmediği görülmektedir.

Bu bakış açısıyla tam metin, OYÖ ve OKÖ'lerin toplam IMRAD bölüm sayısının, tam metin IMRAD düzenine dağılımını veren Şekil 24 ile Şekil 25-a ve Şekil 25-b karşılaştırıldığında yığılım grafiklerindeki durumun sayısal açılımları da ortaya çıkmaktadır. Şekil 24'te tam metinlerde rastlanan IMRAD düzenlerine ait cümlelerin derlemde temsil ettikleri yüzdeler üzerinden ağırlıkları görülmektedir. OYÖ cümlelerinin sahip oldukları IMRAD bölümü sayısının, ait oldukları tam metinlerin IMRAD düzenlerine göre dağılımlarını Şekil 25-a'da; aynı gösterimin OKÖ karşılığını ise Şekil 25-b'de incelemek mümkündür. Buna göre, örneğin 25-a'da "G, Y, B, T ve S" düzenine sahip makaleler için elde edilen yapısal çıktılarının tamamı 4 IMRAD bölümünden cümlelere sahip özlerdir ve bu cümlelerin tüm derleme ait OYÖ cümleleri içindeki ağırlıkları %43'tür. Bu oran tam metinleri "G, Y, B, T ve S" yani 4 IMRAD'a sahip makale cümlelerinin tüm derlem cümlelerine oranı ile (%43,5) hemen hemen aynıdır. OYÖ cümleleri için tam metinlerin tüm IMRAD desenlerine renklerine göre dağılımları ilgili IMRAD düzenindeki IMRAD sayısı ile aynıdır. Bu dağılımlar Şekil 24 ve Şekil 25-a'da oransal olarak da birbirlerine çok benzerdir.



Şekil 24. Derlem cümlelerinin IMRAD sayısı ve düzenine göre ağırlıkları

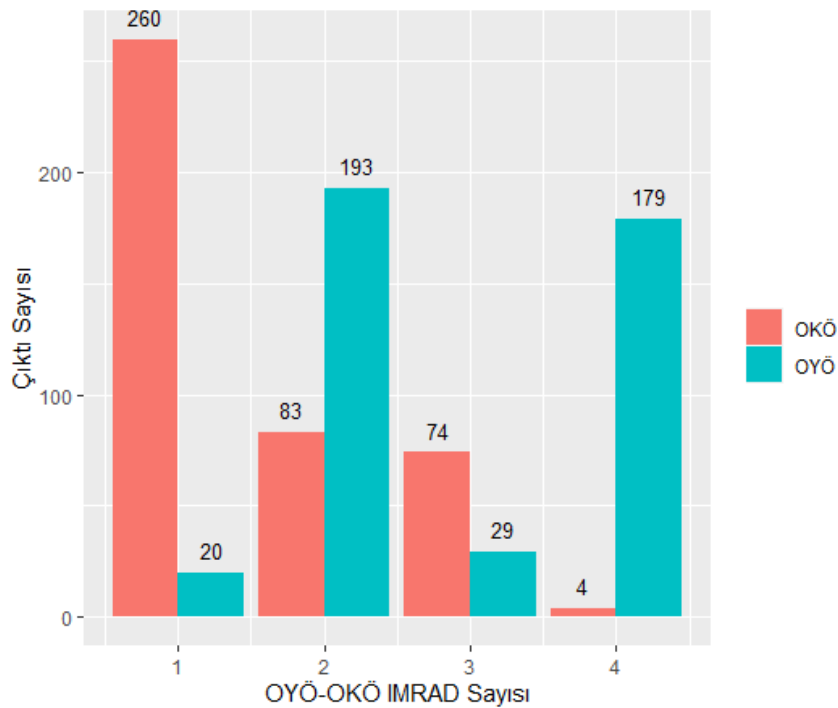
OKÖ cümlelerinin yapısal dağılımlarını veren Şekil 25-b görünüm olarak, ilk bakışta Şekil 24 ve 25-a'dan ayrılmaktadır. OKÖ'ler için yine 4 IMRAD bölümünden oluşan (G, Y, B, T ve S) düzenindeki makalelerden elde edilen çıktı cümlelerinin yer aldığı en üst satırı inceleyecek olursak, bu alanda OKÖ cümlelerinin toplam (%1,3+%17,5+%12,6+%10,2) %41,6'sı yer almıştır. Bu dilimde temsil edilen OKÖ'lerin, cümleleri 4 IMRAD bölümünden seçilebilen örnekleri tüm OKÖ cümlelerinin ancak %1,3'ü kadardır. 3 IMRAD bölümünden cümle seçebilen OKÖ'ler tüm OKÖ cümlelerinin %17,5'ine, 2 IMRAD'lılar, %12,6 ve tek IMRAD'lılar, %10,2'sine karşılık gelmektedir. OKÖ'ler için tam metinlerin IMRAD desenlerinde yer alan IMRAD bölüm sayıları ile grafiği renklendiren OKÖ İMRAD bölüm sayıları birlikte incelendiğinde, sadece "G" ve "B" IMRAD desenlerine sahip tam metinler ile aynı IMRAD bölümü sayılarına sahip olabildikleri görülmektedir. Bu durumda OKÖ'ler sadece giriş ve sadece bulgulardan oluşan makaleler dışında tam metinlerin yapısal düzenleri ile tam bir uyuşma sergilemekten uzaktır. Bu iki makale grubu için de başka yapısal bölümden cümle seçim ihtimali zaten mevcut değildir.



Şekil 25. OYÖ (a), OKÖ (b) cümlelerinin derlem IMRAD desenine göre IMRAD bölümü ağırlıkları

Derlemin tamamında OYÖ ve OKÖ'lerin IMRAD bölümü sayılarına göre nasıl dağıldığı ise Şekil 26'da incelenebilmektedir. Buna göre, OYÖ'lerin beklendiği gibi, içerdikleri

IMRAD bölümü sayılarına göre dağılımlarının Tablo 5'teki tam metinlerin IMRAD bölüm sayılarına göre gerçekleştiği görülmektedir. OKÖ'ler ise yine tam metinlerden farklı dağılmaktadır. Bu farklılaşma, 1 ve 3 IMRAD bölümüne sahip çıktılarda OKÖ'lerin OYÖ'lerden belirgin bir şekilde üstünlüğü ile gerçekleşirken, 2 ve 4 IMRAD sayıları için OYÖ'lerin üstünlüğü ile göze çarpmaktadır. Önceki bulgularda, 1 ve 3 IMRAD bölümü içeren tam metinler, derlemdeki yapısal düzene ilişkin en problemliler olarak değerlendirilmektedir. Bu durumda, sistemin klasik çıktılarının ilgili tam metinlerinde söz konusu olan yapısal sorunları artırarak sahiplendiği söylenebilir. Özellikle 4 IMRAD bölümüne sahip çıktılarda ise yapısal olanların sayısal üstünlüğü açıkça tespit edilmektedir.



Şekil 26. OYÖ, OKÖ IMRAD bölümü sayıları

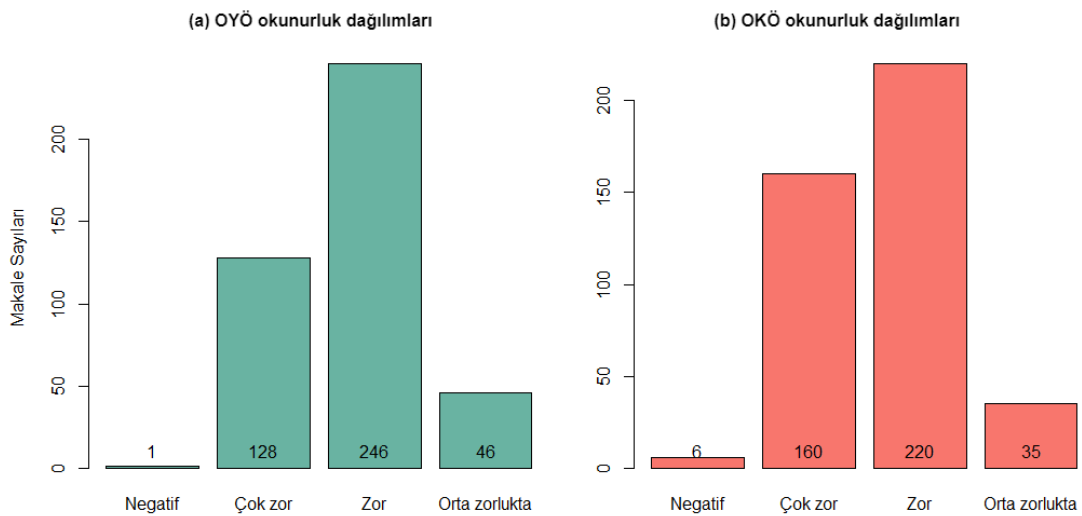
Yöntem bölümünde detaylı şekilde anlatıldığı gibi sistemde kullanılan özelliklerin ağırlıkları eşit olarak belirlenmektedir. Bu bölümde elde edilen bulgulara göre, OYÖ'lerin başka bir işlem ya da insan görüşüne başvurulmadan, tam metin cümlelerini yapısal olmayan çıktılara göre daha dengeli bir ağırlıkla tararken makalenin tüm yapısal bölümlerinden geniş bir yelpazeden cümle seçimi yapabildiği gösterilmektedir. Bilimsel iletişim bağlamında yazar özlerinde yaşanan en büyük sorunlardan biri özlerde tam metin

yapısal içeriğin eksik kalabilmesidir. Bu bölümdeki bulgulardan yola çıkarak, tam metnin yapısal özelliklerini göz önüne almayan otomatik sistemlerde de eksik yapısal içerik sorununun gözlenmeye devam edebileceği söylenebilir.

Tam metin ve otomatik özlerden elde edilen bulgulara göre çıkarıma dayalı otomatik özetleme yöntemleri kullanılarak elde edilen yapısal özlerin tam metni tarama başarımlarının otomatik klasik özlerin tam metni tarama başarımlarına göre belirgin bir şekilde üstün olduğu değerlendirilmektedir.

#### 4.4.2. Otomatik Yapısal Özlerin Okunurluğuna Dair Bulgular

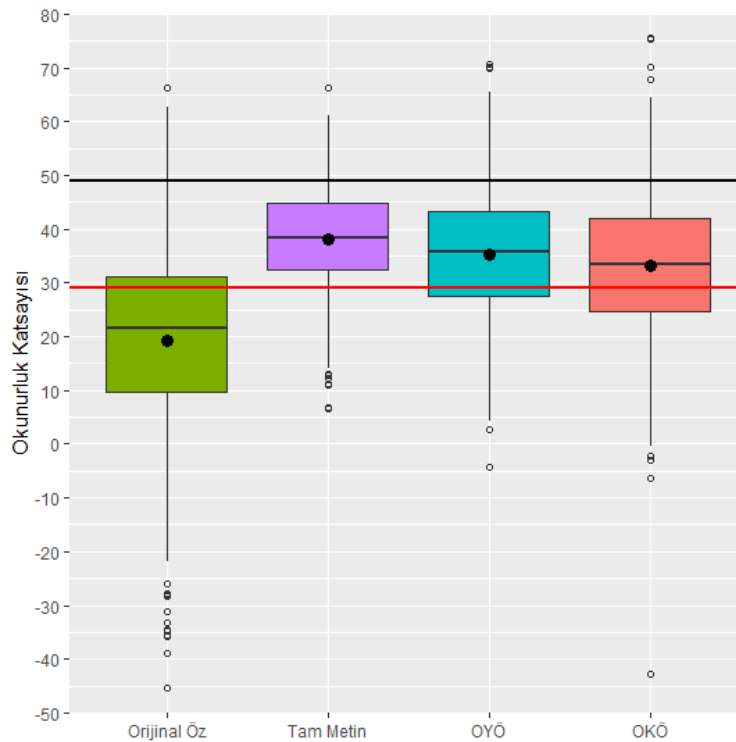
OYÖ ve OKÖ'lerin AOD'lerinin okunurluk ölçeğine göre gruplandırılmasından oluşturulan grafikler Şekil 27 (a) ve (b)'deki gibidir. Buna göre OYÖ ve OKÖ'lerin okunurluklarının birbirine benzer şekilde “Çok Zor” ve “Zor” değerinde yığıldığı görülmektedir. OKÖ'lerin hem negatif okunurluklu hem de orta zorlukta okunurluklu temsilci sayısı OYÖ'lerden biraz daha fazladır. Sistemin her iki çıktısının da çok zor ve negatif okunurluk değerlerine sahip temsilci sayılarının Şekil 19'daki klasik özlerin benzer değerlere sahip temsilci sayılarına göre belirgin bir şekilde azaldığı tespit edilmiştir.



Şekil 27. Sistem çıktısı OYÖ (a) ve OKÖ (b)'lerin okunurluk dağılımları



Derlemedeki tam metin, orijinal öz, OYÖ ve OKÖ okunurluk değerlerinin okunurluk katsayısına ve okunurluk değerlerine göre yığılımlarını karşılaştırmalı olarak gösteren kutu grafiği Şekil 28’de sunulmaktadır. Şekil 28, daha önce, Şekil 20’de ele alınan tam metin ve orijinal öz okunurluklarının grafiğine sistem çıktılarının da eklenmesiyle edilmiştir. Şekil 28’e göre, sistem çıktılarının ürettiği okunurluk değerlerinin, yazar tarafından üretilen orijinal özler ile tam metin okunurluklarının arasında yer aldıkları açıkça görülmektedir.

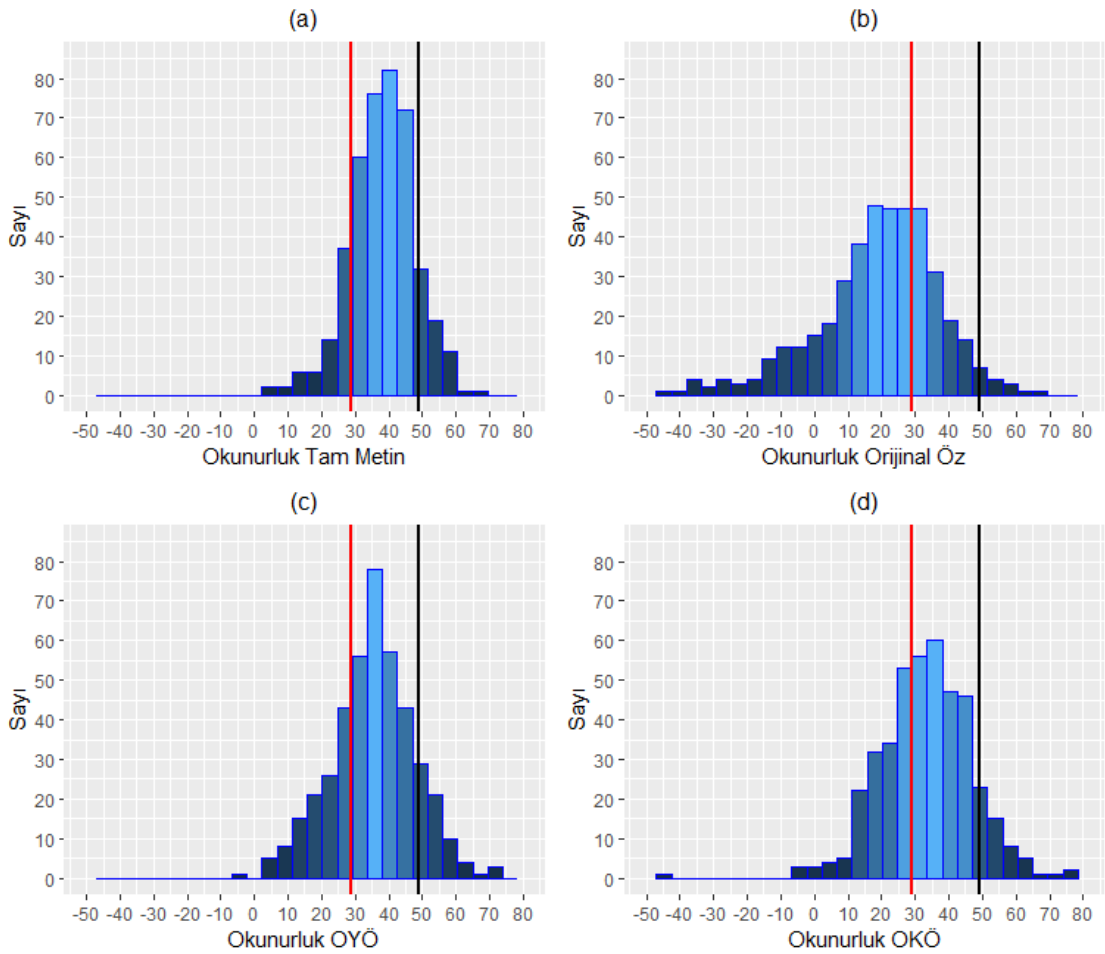


Şekil 28. Derlem okunurluk değerleri

Kırmızı yatay çizgi ( $y=29$ ) ile siyah yatay çizgi ( $y=49$ ) arasında belirlenen alan grafikte okunurluk ölçeğine göre “zor” olan alanı sınırlamaktadır. Kırmızı çizginin altı “çok zor”, siyah çizginin üstü “orta zorlukta” okunurluk düzeylerine aittir. Yazar tarafından üretilmiş orijinal özlerin yığılımı grafiğe göre en altta (en zor okunurluk düzeyinde), neredeyse tamamen “çok zor” olarak sınıflanan bölümde yer almıştır. Tam metinler ise tamamen ağırlıklı olarak “zor” okunurluklu bölgede yığılmıştır. Sistem çıktılarının okunurluklarının tamamına yakını özellikle OYÖ’ler için “zor” okunurluklu alanda olmak üzere “zor” ve “çok zor” okunurluklu alanlarda yığılım gösterirken; ortalama okunurluk değerleri de “zor” okunurluklu alanda yer almıştır. OYÖ’lerin OKÖ’lere göre

daha fazla temsilcisi “zor” okunurluklu bölgededir ve OYÖ okunurluk ortalamaları OKÖ’lerden daha yüksek AOD değerindedir.

Tam metin, orijinal öz, OYÖ ve OKÖ okunurluklarının kendi içlerinde nasıl dağıldığı Şekil 29 (a), (b), (c) ve (d)’de gösterilmektedir. Şekil 28’deki ile benzer şekilde, kırmızı dikey çizgi ( $x=29$ ) ile siyah dikey çizgi ( $x=49$ ) arasında belirlenen alan grafikte okunurluk ölçeğine göre “zor” olan alana işaret etmektedir. Buna göre, geliştirilen sistemin çıktılarının Şekil 29 (c) ve 29 (d)’de verilen okunurluk değerlerinin okunurluk ölçeğine göre dağılımının Şekil 29 (a)’da görülen tam metinlerin dağılımına benzer desende olduğu anlaşılmıştır. Şekil 29 (b)’de verilen yazar tarafından üretilen orijinal özlerin ise bu üç metin grubuna göre belirgin şekilde zor okunurluklu bir görünüm sergilediği ve farklı bir desende dağıldığı açıkça görülmektedir.



Şekil 29. Tam metin (a), orijinal öz (b), OYÖ (c) ve OKÖ (d) okunurluk dağılımları

Bu bölümde elde edilen bulgulara göre, çıkarıma dayalı otomatik özetleme yaklaşımları ile tam metinden temsilci cümle seçildiğinde yazar tarafından uygun görülen öz cümlelerine göre daha kolay okunurluklu cümle seçimi yapılabildiği ortaya konulmuştur. Geliştirilen sistemde cümle seçim yaklaşımı belirlerken seçilen özellikler için tek ölçüt olarak daha önce Türkçe derlem üzerinde denenmiş ve başarılı bulunmuş özellik seti olması göz önüne alınmıştır. Ancak, kullanılan cümle seçme fonksiyonunda yer alan, cümledeki isim soylu kelime sayısına bağlı gerçekleşen hesaplama, doğal olarak cümledeki kelime sayısını artırmaktadır. Çıktıların daha iyi okunurluklu olabilmesi için uygun özellik grupları ve bu özellikler için uygun katsayılar seçmek mümkündür. Geliştirilen sistemin sonuçlarına göre (uzun cümle elde etmeye neden olabilecek bir bileşene sahip) standart bir özellik seti ağırlıksız uygulandığında dahi yazar tarafından üretilen orijinal özlere göre daha iyi okunurluk değerlerine sahip çıktılar elde edilebildiği anlaşılmıştır. Diğer tüm değişkenler aynı olduğu halde yapısal özellikleri dikkate almayan otomatik özleri (OKÖ) de orijinal özlere göre daha iyi okunurluğa sahip olacak şekilde elde etmek mümkün olmuştur.

#### **4.4.3. Sistem Çıktılarının Yazar Tarafından Üretilen Orijinal Öz Üzerinden ROUGE Değerlendirmesi**

Bu bölümde, çalışma kapsamında geliştirilen otomatik özetleme sisteminin yapısal ve klasik çıktıların başarımları değerlendirilmektedir. Bu değerlendirme kapsamında, OYÖ ve OKÖ'lerin orijinal özlere göre başarımlarının karşılaştırılmasının yanı sıra, otomatik yöntemlerle öz elde etmede tam metnin yapısal özelliklerinin etkisi de derlem üzerinde test edilmiştir.

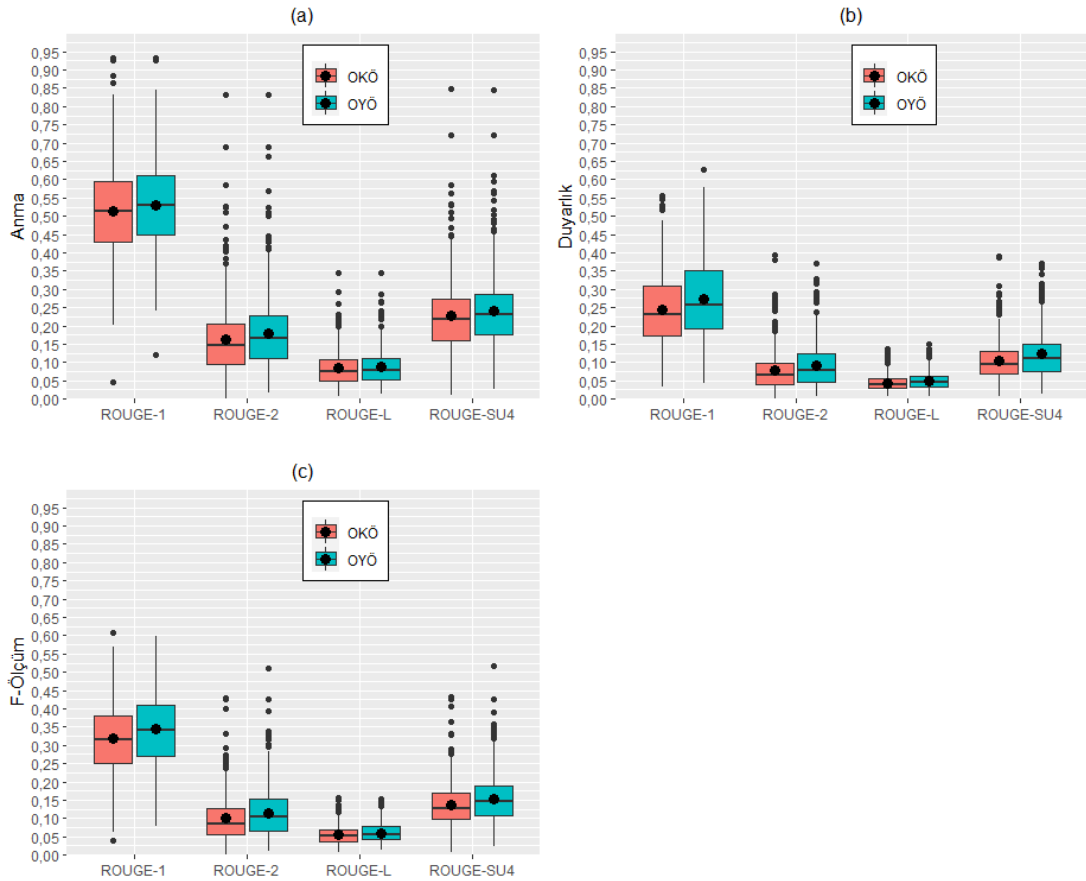
Değerlendirme için öncelikle, sistemin ürettiği yapısal ve klasik özlerin yazar tarafından üretilen orijinal özler ile n-gram çakışmaları yöntem bölümünde detaylarıyla yer verilen ROUGE ölçümlerine göre hesaplanmıştır. Şekil 30, derlemdeki tüm makaleler için sistem çıktıları ve yazar özlerinin çakışma oranları ROUGE-1, ROUGE-2, ROUGE-L ve ROUGE-SU4 ölçümlerine göre anma, duyarlık ve f-ölçüm değerlerinin yığılım grafikleri ile verilmiştir. Sistem çıktısı bir öz ile yazar özü arasındaki karşılaştırmadan elde edilen bir ROUGE anma değerinin örneğin 0,35 çıkması sistem çıktısı cümlelerinin yüzde 35

oranında yazar özü ile örtüştüğünü gösterir. Şüphesiz bu oldukça düşük bir benzerlik oranı olarak yorumlanabilir. Ancak akademik makaleler için otomatik öz ile yazar özü karşılaştırması yapıldığı bir senaryoda yazar özünün tam metindeki cümlelerden farklı şekillerde yazılabileceği açıktır. Bu fark anlam ve içerikten kaynaklanabileceği gibi eş anlamlı sözcük ve kavramlar sebebiyle de gerçekleşebilir. Eş anlamlılık kavramını da sistemde değerlendirebilmek için ilgili dilde hazırlanmış geniş sözlüklere ihtiyaç vardır. Bu sebeple, özellikle eş anlamlılık kavramını göz önüne almadan, sadece akademik metinler için değil, herhangi bir özetleme sistemi geliştirirken çıktılarının sadece ROUGE puanlarına bakarak ne kadar başarılı olduğunu söylemek mümkün değildir (Ganesan. 2018a, s. 1). Ayrıca, yapılmış çalışmalara göre, yazar özlerinin tam metni yansıtma oranları düşük çıkmakta, tam metnin tüm yapısal bölümlerinden bilgi içermediği de görülmektedir (Atanassova vd., 2016; Jamar vd., 2014, s.21; Šaupperl vd., 2008, s. 1430). Ancak, ROUGE puanları geliştirilen otomatik özetleme sistemlerinde denenen farklı yaklaşımların sistem çıktılarına katkısı olup olmadığını değerlendirmek için uygundur ve yaygın olarak bu amaçla kullanılmaktadır (Ganesan. 2018a, s. 1). Tez çalışması kapsamında da ROUGE puanları yapısal olan ve olmayan çıktıların başarımları ve tam metnin yapısal bölüm sayısının çıktı başarımına etkisi hakkında fikir sahibi olmak için kullanılmıştır.

Şekil 30 (a), (b) ve (c) geliştirilen sistem çıktıları olan OYÖ ve OKÖ'lerin ROUGE-1, ROUGE-2, ROUGE-L, ROUGE-SU4 ölçümlerine göre anma, duyarlık ve f-puanlarının tüm derlemdeki yığılımlarını göstermektedir. Her üç puan türünde de OYÖ, OKÖ çıktı gruplarının dağılımları benzer özellik göstermektedir. Mavi kutularla gösterilen OYÖ'lerin daha yüksek skorlar elde edebildiği grafiklerden anlaşılmaktadır. En yüksek puanlar ROUGE-1, ROUGE-2 ve ROUGE-SU4'e göre elde edilirken, en düşük puanlar belirgin bir şekilde ROUGE-L'e aittir. Yapısal özlerin klasiklere göre başarımlarının en düşük fark değerleri de ROUGE-L'e göre elde edilmiştir. OYÖ ve OKÖ'lerin yazar özleri ile n-gram çakışmalarının en yüksek değerleri her seferinde ROUGE-1'e göre elde edilmiştir. Çok kısa özetler için ROUGE-1'in tek başına yeterli olabileceği tavsiye edilmektedir (Ganesan, 2018b).

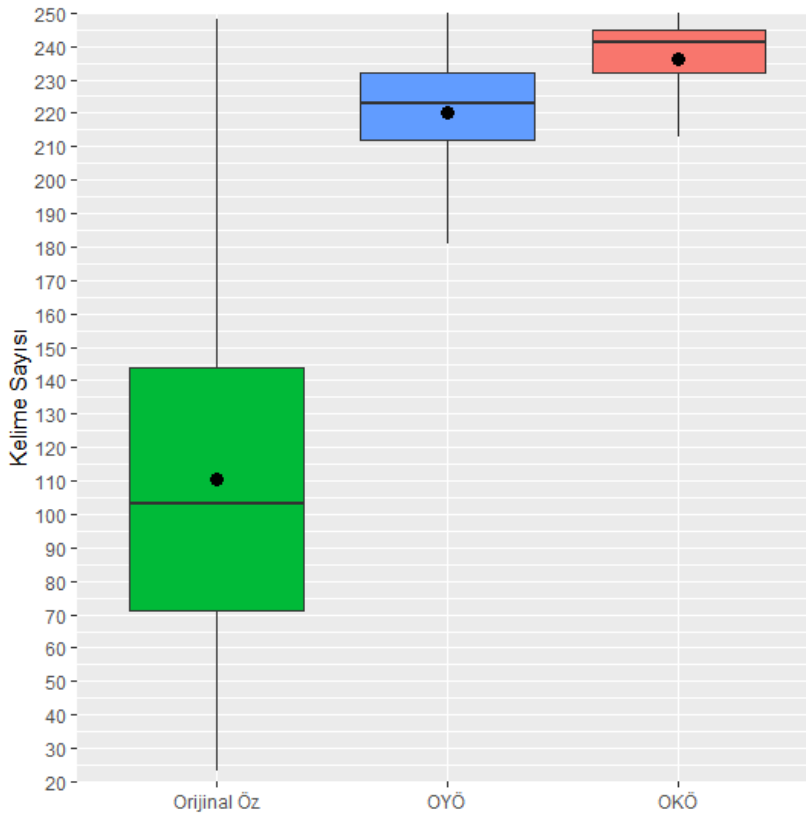
Şekil 30 (a)'da verilen her bir ROUGE ölçümünün anma puanları, Şekil 30 (b)'deki duyarlık puanlarından belirgin şekilde yüksek hesaplanmıştır. Duyarlık puanlarının düşük olması Şekil 30 (c)'de görülen f-ölçüm değerlerine de yansımıştır.

Duyarlık puanı sistem tarafından üretilen özetin gürültüsünden etkilenmektedir. Akademik makaleleri ele alan sistemlerin değerlendirme aşamasında tam metinden seçilen cümlelerle oluşmuş bir ideal özet değil, yazar tarafından üretilmiş orijinal makale özleri kullanılması yaygın bir yaklaşımdır. Bu yaklaşımın sebeplerine 3.6. Otomatik özlerin değerlendirilmesi bölümünde değinilmiştir. Ancak, literatür incelemesinde ele alındığı gibi, yapılan çalışmalara göre, yazar tarafından kaleme alınan öz cümlelerinin sadece %25'inin tam metin cümlelerine %60 üzerinde bir oranla benzerlik gösterdiği bilinmektedir (Atanassova vd., 2016, s. 642). Bu durumun, geliştirilen sistem çıktılarına gürültü ekleyeceği açıktır.



Şekil 30. IMRAD'a göre ROUGE-1, ROUGE-2, ROUGE-L ve ROUGE-SU4 ölçümlerinin anma (a), duyarlık (b) ve f-ölçüm (c) değerleri

Çıktıların duyarlık puanlarını düşürebilecek bir diğer önemli durum da yazar özlerinin 250 kelime sınırına çoğunlukla uzak olmasıdır. Şekil 31’de açıkça görüldüğü gibi orijinal özlerin kelime sayıları oldukça heterojen bir dağılım sergileyerek, geniş bir aralıkta değişim gösterebilmektedir (çoğunlukla 70 ile 145 arasında). Sistem özlerinin uzunlukları ise çok daha homojen bir şekilde, 210 ile 245 arasında yığılmıştır. Bu uzunluk farkının ROUGE değerlendirmesinde anma puanlarının duyarlık puanlarından yüksek hesaplanmasına sebep olabileceği akılda tutulmalıdır.



Şekil 31. Orijinal öz, OYÖ, OKÖ kelime sayıları

Yazar özlerinin görece kısa olması sebebiyle içerik olarak da kısıtlı olabilecekleri ve bu yüzden sistem çıktıları ile yüksek çakışma puanları elde edilemeyeceği de göz önüne alınmalıdır. Bu tür bir kısıtlılığın, duyarlılığın yanı sıra anma puanlarına da etki edecek bir diğer önemli faktör olduğu düşünülmektedir. ROUGE-L ölçümlerindeki belirgin düşüklük de bağlamda açıklanabilir. ROUGE-L, cümle seviyesi yapı benzerliğini doğal olarak hesaba katan bir ölçümdür ve karşılaştırdığı metinler arasında, meydana gelen en

uzun n-gram birlikteliği dizisini tanımlamaktadır. Bu doğrultuda, yukarıda ele alınan orijinal özler ile sistem çıktıları arasındaki benzerliği etkileyecek tüm unsurların cümleler arasındaki n-gram birliktelik dizilerinin boyutlarını da düşüreceği düşünülebilir.

Şekil 31'e göre, OYÖ ve OKÖ uzunlukları arasında da fark gözlenmektedir. Aradaki fark nispeten küçük de olsa yapısal çıktıların uzunluklarının klasik çıktılarından düşük olduğu anlaşılmıştır. Ancak OYÖ ve OKÖ sistem çıktılarının anma puanlarını karşılaştırdığımızda her iki grubun duyarlık puanları arasındaki fark anma puanları arasındaki farktan nispeten büyük de olsa otomatik yapısal özlerin üstünlüğü değişmemiştir.

Bu tez çalışmasının amacı otomatik bir öz hazırlama sisteminde tam metinlerin yapısal içeriğinin sistem çıktılarına etkisini araştırmaktır. Bu çerçevede, ROUGE puanları ile değerlendirilecek unsur, kullanılan yöntemlerin performans derecesi değil, yaygın bilinen yöntemlerle birlikte tam metinlerin yapısal özelliklerini ele almanın çıktıların kalitesine katkısıdır. Elde edilen puanların, geliştirilen sistemin yapısal ve klasik çıktıların birbirlerine üstünlüklerini ayırt etmede kullanılmaya uygun olduğu değerlendirilmiştir. ROUGE-1, ROUGE-2, ROUGE-L ve ROUGE-SU4 ölçümlerinin anma, duyarlık ve f-ölçüm değerlerinin hepsinde OYÖ'lerin OKÖ'lerden daha yüksek sonuçlar elde ettiği Şekil 30 (a), (b) ve (c)'de gösterilmiştir.

Çıktı başarımlarına tam metinlerin yapısal düzene uygun bölüm sayılarının etkisi Tablo 7'de daha iyi anlaşılmaktadır. Tablo 7 derlemde, IMRAD bölüm sayılarına göre gruplandırılmış makalelerin her bir ROUGE ölçümüne göre ortalama f-ölçüm değerlerini vermektedir. Tabloda IMRAD sayısı olarak "Tümü" ile ifade edilen satır, derlemin tamamının IMRAD'a göre gruplandırma yapılmadan hesaplanan ölçüm değerleridir. Buna göre tüm f-ölçüm puanları içinde her zaman en yüksek değere sahip olan çıktı grubunun dört IMRAD bölümüne sahip makale grubundan elde edilen OYÖ'ler olduğu görülmektedir.

Tablo 7. OYÖ ve OKÖ'lerin IMRAD'a göre ROUGE ölçümlerinin ortalama f-ölçüm değerleri

Tam Metin IMRAD Sayısı	ROUGE Ölçümü	OYÖ ort. f-ölçüm	OKÖ ort. f-ölçüm
4	ROUGE-1	<b>0,38540</b>	0,34480
3	ROUGE-1	0,32330	0,31640
2	ROUGE-1	0,31421	0,29663
1	ROUGE-1	0,22790	0,22790
Tümü	ROUGE-1	<b>0,34250</b>	0,31659
4	ROUGE-2	<b>0,13619</b>	0,11080
3	ROUGE-2	0,10776	0,10244
2	ROUGE-2	0,10083	0,09198
1	ROUGE-2	0,06265	0,06265
Tümü	ROUGE-2	<b>0,11493</b>	0,09967
4	ROUGE-L	<b>0,06687</b>	0,06021
3	ROUGE-L	0,05719	0,06300
2	ROUGE-L	0,05471	0,05016
1	ROUGE-L	0,05639	0,05639
Tümü	ROUGE-L	<b>0,06014</b>	0,05561
4	ROUGE-SU4	<b>0,17809</b>	0,15076
3	ROUGE-SU4	0,14524	0,13947
2	ROUGE-SU4	0,13795	0,12808
1	ROUGE-SU4	0,10350	0,10350
Tümü	ROUGE-SU4	<b>0,15392</b>	0,13733

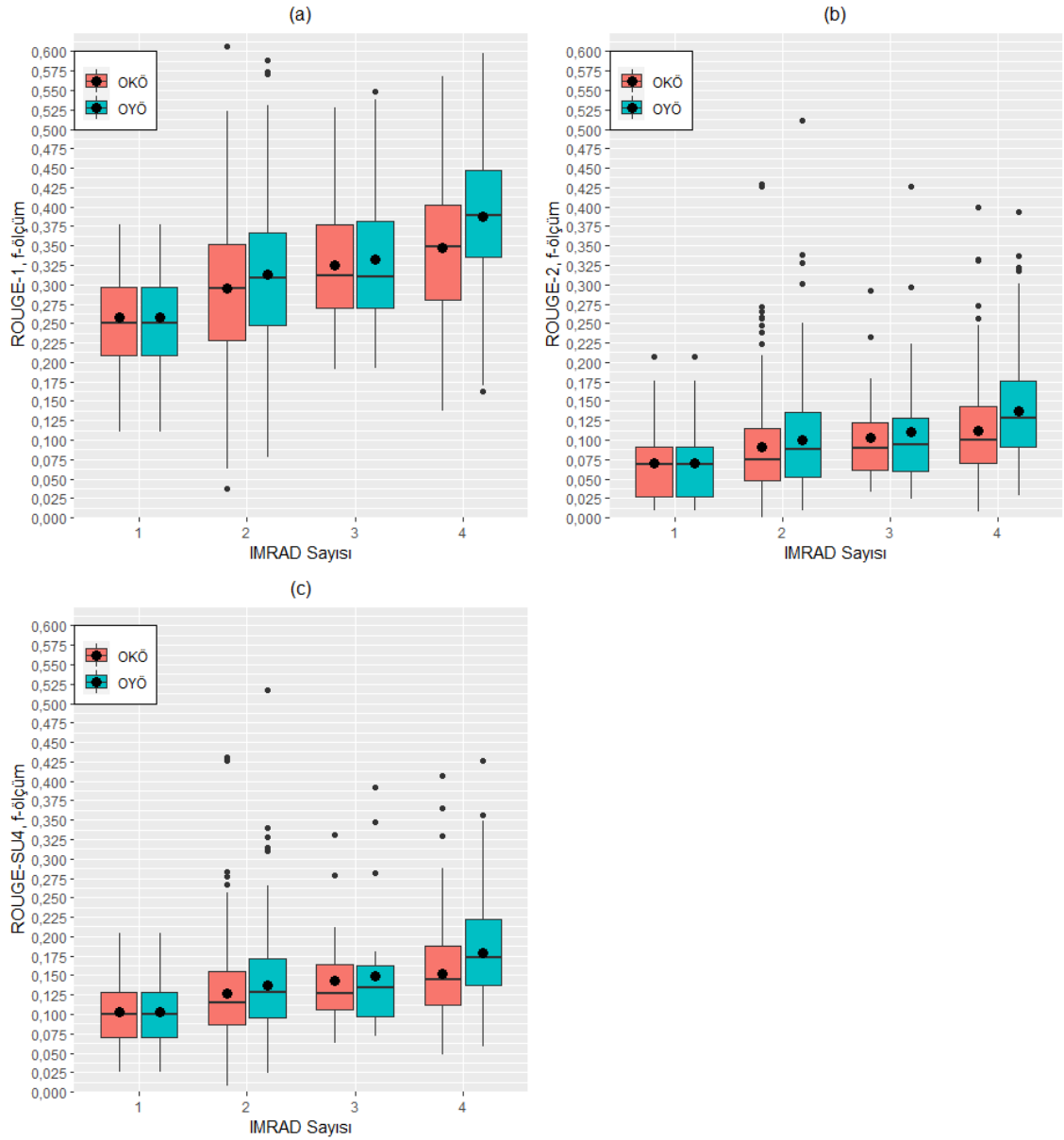
Şekil 32 (a), (b) ve (c), Tablo 7'de ortalamaları verilen IMRAD bölüm sayılarına göre gruplanmış f-ölçüm puanlarının kutu grafiklerinden oluşmaktadır. Buna göre mavi ve kırmızı kutular, x eksenine göre, 1, 2, 3, 4 IMRAD bölümünden oluşan makaleler için OYÖ ve OKÖ çıktı başarımlarını temsil etmektedir. Şekil 32 (a), (b)ve (c)'ye göre, tam metinlerdeki IMRAD bölümü sayısındaki artışa paralel olarak çıktılarının ROUGE-1, ROUGE-2 ve ROUGE-SU4 ölçümlerinin f-ölçüm değerlerinin de arttığı görülmektedir. ROUGE-L ölçümüne göre de en yüksek başarımlar dört IMRAD bölümünden oluşan sistem çıktılarında da %10'un altında çakışma oranları elde edildiği için bu ölçümün IMRAD sayılarına göre değişimi grafik eşliğinde incelenmeye gerek duyulmamıştır.

Tablo 7 ve Şekil 32'de göze çarpan ilginç bir bulgu, sistemin tüm çıktılarında; yani hem yapısal hem de yapısal olmayan otomatik özler için tam metinlerdeki IMRAD bölüm



sayısındaki artışa paralel olarak başarımlarda da artış gözlenmesidir. Tam metinlerdeki yapısal bölüm sayısı OKÖ elde etmede bir fark yaratmadığı için tam metinlerin yapısal düzeni iyileştikçe insan tarafından oluşturulan orijinal öz ile çıkarıma dayalı otomatik bir sistem tarafından oluşturulan özün benzerliğinin arttığı söylenebilir. Ayrıca tüm grafiklerde ve Tablo 7’de, iki ve dört IMRAD’a sahip tam metinlerden elde edilen çıktı gruplarında OYÖ’lerin OKÖ’lere üstünlüğünün benzer oranda ve daha belirgin şekilde arttığı görülmektedir. Bu durum, iki IMRAD’a sahip makalelerin, dört IMRAD bölümünden oluşan makaleler ile yapısal içerik bakımından benzer zenginlikte olması ile yorumlanabilir. Ancak, bir ve üç IMRAD bölümünden oluşan makaleler sayıca azınlıktadır; elde edilen bulgular iki IMRAD bölümüne sahip makalelerin bir ve üç IMRAD bölümünden oluşan makalelerden daha kabul edilebilir bir yapısal düzende olduğunu düşündürse de kesin sonuca varabilmek için makalelerin yapısal düzenleri özelinde ileri araştırmalara ihtiyaç vardır.

Şekil 32 (a), (b) ve (c)’de yer alan grafikler çıktı gruplarına göre karşılaştırmalı olarak incelendiğinde, OYÖ’lerde IMRAD sayısına göre artan başarımın ÖKO’lerden daha fazla olduğu açıkça görülmektedir. Dört IMRAD bölümünden oluşan makaleler için çıktı başarımları, diğer IMRAD bölümü sayılarına sahip makalelere göre yine belirgin bir şekilde yüksektir.



Şekil 32. IMRAD'a göre ROUGE-1 (a), ROUGE-2 (b), ROUGE-SU4 (c) ölçümlerinin f-ölçüm değerleri

Bu bölümde elde edilen bulgulara göre çıkarıma dayalı otomatik özetleme sistemleri ile akademik makaleler için öz elde ederken tam metinlerin yapısal özelliklerini de işleme dâhil etmenin sistem çıktılarının başarımlarını artırdığı açıkça gösterilmiştir. Klasik (OKÖ) ve yapısal (OYÖ) sistem çıktılarının başarımları incelenirken tespit edilen dikkat çekici bulgu ise tam metinlerde kullanılan yapısal bölüm sayısındaki artışın (yapısal çıktılar için daha fazla olmak üzere) her iki çıktı grubu için de başarımları arttırması olmuştur.

## 5. BÖLÜM: SONUÇ VE ÖNERİLER

İnternet ve bilgi teknolojilerinin yaygın kullanımı ile birlikte basılı bilgi miktarının artışı geçmiş dönemlere göre abartılı bir hal almıştır. Bu durum bireylerin okumaya ayırdıkları süre içinde daha fazla bilgi edinme gereksinimini de beraberinde getirmektedir. Bilimsel yayın hacmindeki artış karşısında kullanıcıların ilgi alanlarına giren tüm makalelerin tam metinlerini okumaya yetişememeleri sorunu da geçmişe göre farklı bir boyuttadır. Bilimsel makalelerin elektronik sürümleri bilgiye erişimde büyük kolaylıklar getirirken, çevrimiçi ortamın ve dijital ekranların doğası gereği makalelerin okunurlukları zorlaşabilmektedir. Özler bu zorlukları aşmada araştırmacılara yardımcı olduğu düşünülen makale bölümleridir. İyi yapılandırılmış bir öz, araştırmacının tam metnin tamamını incelemesi gerekmeden içeriğe dair bilgi sahibi olmasını sağlayan; kısa, özlü ve açık bir dille yazılmış makale bölümüdür. Ancak uygulamada, makalelerin yapısal bölümlerinden bilgi aktarmada eksiklikleri ve zor okunurluklu örnekleri sıklıkla göze çarpmaktadır. Bu sorunları aşmada yazarlara ve okurlara yardımcı olması için yapısal öz formatı ortaya çıkmıştır. Kullanıma geçtiği alanlarda hızla kabul gören bu formatlarda yazar belirli bir yapısal düzen şablonunu doldurduğu için içeriği tam yansıtırken, okur alt başlıklara bölünmüş bir metin okuduğundan dolayı daha kolay ve hızlı tarama yapabilmektedir. Ancak özellikle sosyal bilimler alanında en çok kullanılan format olan klasik özler tek bir blok paragrafta makalenin tüm önemli kısımlarını, mümkün olan en kısa haliyle yansıtmalıdır. Bunun da hem yazar hem okur hem de hakemleri yoran bir format olduğu değerlendirilmektedir.

Bu çalışma ile teknolojinin insanın bilgi ile etkileşimini devamlı dönüştürdüğü günümüz koşullarında, bilimsel iletişime özlerde yaşanan bu sorunlar bağlamında katkı yapmanın yolları irdelenmektedir. Kullanılan derlem kütüphanecilik ve bilgibilim alanında yer alan *Bilgi Dünyası* ve *Türk Kütüphaneciliği* dergilerinde yayımlanmış 421 makaleden oluşmaktadır. Çalışma kapsamında geliştirilen çıkarıma dayalı otomatik yapısal öz hazırlama sistemi ile makalelerin tam metinlerinin tüm yapısal bölümlerinden temsilci cümleler seçerek otomatik yapısal öz çıktılarını elde edilmiştir. Cümle seçmede, Türkçe metinler üzerinde üstünlükleri daha önce denenip, gösterilmiş cümle konumu, cümle merkeziliği ve isim soylu kelimeleri içerme özellikleri eşit ağırlıklı olarak kullanılmıştır.

Yapısal özelliklerin otomatik özetleme sürecine etkisini değerlendirmek için tam metnin yapısal bölümlerini göz önüne almadan cümle seçimi yapılarak klasik çıktılar da üretilmiştir. Otomatik değerlendirme sürecinin ideal özleri olarak ise yazar özleri kullanılmıştır. Araştırma kapsamında ele alınan kütüphanecilik ve bilgilim derleminin yapısal özellikleri incelenirken giriş, yöntem, bulgular, tartışma ve sonuçlar şeklinde genel kabul görmüş IMRAD başlıklarına göre değerlendirme yapılmıştır. Çalışmanın amacı makalelerin yapısal içeriğine göre cümle seçimi yapan bir sistem geliştirmek ve çıktıları değerlendirmek olarak belirlenmiştir. Bu doğrultuda, bir makalenin yapısal düzenine karar vermede yazar başlıkları değil tam metin içeriğinin niteliği dikkate alınmıştır. Derlemin IMRAD'a uyumunu anlamak için makale tam metin paragrafları uzman işaretçi görüşlerine göre içeriklerinin yapısal nitelikleriyle işaretlenmiştir. İşaretleme sonrası yapılan analize göre, derlem genelinde özellikle yöntem bölümünün eksikliği tespit edilmiştir. Hacmen baskın olarak tespit edilen yapısal bölüm ise giriş olarak öne çıkmaktadır. Giriş bölümü olmadığına karar verilen, sadece bulgulardan oluşan tek bir makale derlemin %2'sini temsil etmektedir. Yöntem, bulgular ve tartışma ve sonuç bölümlerinden birinin her zaman eksik kaldığı, üç IMRAD bölümünden oluşan makalelerin derleme oranı %6,8 olarak saptanmıştır. Sadece giriş bölümünden oluşan makaleler ise derlemin %4,5'ine karşılık gelmektedir. Derlemin %13,3'üne karşılık gelen bu makale grubunda yapısal düzen kullanımının belirgin bir şekilde sorunlu olduğu değerlendirilmiştir.

Derlemin baskın çoğunluğu iki ve dört IMRAD bölümüne sahip makalelerden oluşmaktadır. Dört IMRAD bölümünün tamamının yer aldığı, yapısal düzene tam uygun yazılmış makaleler derlemin %42,5'ini oluşturmaktadır. Giriş ile tartışma ve sonuçlar bölümlerinden oluşan iki IMRAD bölümüne sahip makaleler ise %46,3 oranında derlemde temsil edilmektedir.

Yöntem bölümünün eksik kullanımı derlem bağlamında en sık rastlanan problem olarak göze çarpmaktadır. Eksikliklere rağmen, sonuç olarak, derlemde yer alan makalelerin mevcut yapısal düzen kullanım oranlarının geliştirilen sistem için yeterli düzeyde olduğu değerlendirilmiştir.

Türkçe kütüphanecilik ve bilgibilim alanında, yazılan makaleler için, tespit edilen sorunlara rağmen, önemli ölçüde yapısal düzene uygun içeriğe sahip oldukları söylenebilir. Makalelerin baskın çoğunluğu giriş, yöntem, bulgular ile tartışma ve sonuçlar bölümlerinin tamamını ya da giriş ile tartışma ve sonuçlar bölümlerini içermektedirler. Sadece giriş ile tartışma ve sonuçlar bölümlerini içeren bilimsel makaleler için, sosyal bilimler alanında her zaman yöntem ve bulgular bölümlerinin yer almak zorunda olmadığına dair görüşler mevcuttur. Ancak her bilimsel makalenin mutlaka bir yönteminin ve yönteme bağlı bölümlerinin olması gerektiği görüşü de literatürde yer bulmaktadır. Giriş ile tartışma ve sonuç bölümlerinden oluşan makalelerin derlemde baskın kullanılan yapısal düzenlerden biri olması sebebiyle bu tür bir yapısal formatın hatalı ya da bir amaç doğrultusunda kullanımına ileri araştırmalar ile karar verilebilir.

Genel anlamda okuyucu ile metin ilişkisine etki eden pek çok unsur vardır. Bu unsurlar bazen okuru metne çekerken bazen de okumayı zorlaştırarak okuru, okuma eyleminden uzaklaştırmaktadır. Okurun hazır bulunuşluğu, konuya ilgisi, metnin yazıldığı dile hâkimiyeti, yaşı, okuma eylemi gerçekleştirilirken bulunulan ortamın ışık düzeni vb. gibi pek çok değişken bu bağlamda sayılabilir. Okunurluğa etki eden unsurları araştırırken tüm bu değişkenleri göz önüne almak her zaman mümkün değildir. Bu sebeple yaygın kullanımda olan sayısal okunurluk değerlerine göre belirlenen okunurluk düzeyleri yazılı iletişimde bir metnin okura uygunluğu ile ilgili dikkate alınan önemli faktörlerden biridir. Bu okunurluk değerlerinin en yaygın kullanılanları metinlerde cümlelerin kelime, kelimelerin hece uzunluğuna bağlı değişen bir kıyaslama yapar. Türkçe kütüphanecilik ve bilgibilim alanında yayımlanan makalelerin tam metinlerinin okunurluk düzeyleri hesaplandığında, derlemin büyük ölçüde, okunurluk ölçeğine göre “zor” okunurluklu olarak sınıflandığı görülmüştür. Yazar tarafından üretilmiş orijinal makale özlerinin okunurluk değeri ise bundan bir düşük okunurluk derecesi olan “çok zor” düzeyinde yığılmıştır.

Bilimsel makalelerin özleri, okura, ilgili tam metinlerin tamamını incelemeyen çalışmanın içeriği hakkında kısa, özlü ve kolay anlaşılır bir ön bilgi sağlamak amacıyla üretilen metinlerdir. Okurların, bu ön bilgi ile tam metnin araştırma konularıyla bağını

kurarak ilgili makaleyi okuyup okumamaya karar vermeleri kurgulanmıştır. Özlerde kullanılan dilin tam metinlerden daha sade ve anlaşılır olmasının gerekliliği açıkça biliniyor olmasına rağmen, derlem bağlamında, yazarların öz yazımında, tam metinlerde kullandıkları yazım dillerinden bağımsız olarak zor okunurluklu yazmayı tercih ettikleri görülmüştür. Bu bulgu literatür ile de uyumludur. Türkçe bilgibilim ve kütüphanecilik alanında yazan araştırmacıların öz yazımı özelinde okunurluk ölçeğine göre “çok zor” okunurluklu bir dil kullanma alışkanlığına sahip oldukları anlaşılmaktadır.

Geliştirilen sistem çıktılarının okunurluk değerleri incelendiğinde, hem otomatik yapısal hem de otomatik klasik özlerin okunurluk değerlerinin ortalamalarının birbirine yakın ve yazar tarafından üretilmiş orijinal özlerden belirgin bir şekilde daha kolay okunurluklu olduğu saptanmıştır. Orijinal özlerin okunurluk düzeylerinin ortalaması “çok zor”, sistem çıktısı özlerin ortalaması ise tam metinlere benzer şekilde “zor” olarak ölçülmüştür. Buna göre, orijinal yazar özlerinden daha kolay okunurluklu üretilen tam metinlerden otomatik yöntemlerle cümle seçimi yapıldığında orijinal özlerden belirgin bir şekilde daha kolay okunurluklu özler elde etmenin mümkün olduğuna dikkat çekilmektedir.

Yapısal özellikleri kullanarak üretilen otomatik özlerin, yapısal özellikleri göz ardı eden sürümlerine göre tam metni belirgin bir şekilde daha geniş bir aralıkta taradığı görülmüştür. Otomatik yapısal öz cümlelerinin derlem tam metinlerindeki dağılımlarının IMRAD’a göre yığılımları, makale tam metninin derlem cümlelerine dağılımlarının IMRAD’a göre yığılımları ile benzerlik göstermektedir. Geliştirilen sistemde yapısal özellikleri göz önüne almadan elde edilen özlerin, büyük çoğunlukla tam metnin yapısal bölümlerini eksik yansıttığı gösterilmiştir. Otomatik klasik özlerde yapısal düzene ilişkin sorunlu durumların tam metinlerdekinden daha yoğun gözlemlendiği tespit edilmiştir. Bu doğrultuda, geliştirilen sistemde çıkarıma dayalı otomatik öz elde ederken yapısal bölümleri kullanmanın tam metni tarama başarımını belirgin bir şekilde artırdığı anlaşılmıştır.

Yapısal çıktıların metin kalitelerinin, okunurlukları ve tam metni tarama başarımlarına göre yapısal olmayan çıktılara göre üstünlükleri gösterildikten sonra, ROUGE 2.0

değerlendirme paketi ile yapılan otomatik değerlendirmeleri ile içerik başarımları yorumlanmıştır. Değerlendirmede sistem çıktılarının, yazar tarafından üretilen orijinal özler ile n-gram çakışmaları ROUGE ölçümlerine göre elde edilen anma, duyarlık ve f-ölçüm puanlarıyla karşılaştırılmıştır. Buna göre, sistem çıktılarını elde ederken yapısal özellikleri hesaba katmanın başarımı artırdığı saptanmıştır. Yapısal bölümlerin özlere etkisini daha detaylı görmek için, tam metinde kullanılan IMRAD bölümü sayısına göre çıktı başarımları da ayrıca incelenmiştir. Sonuç olarak, otomatik yapısal özlere daha belirgin olmak üzere, tam metindeki IMRAD sayısı arttıkça, çıktılarının her iki grubunun ROUGE ölçümlerine göre f-ölçüm değerlerinin belirgin bir şekilde artış gösterdiği anlaşılmaktadır. Buna göre, çıkarıma dayalı bir otomatik özetleme sisteminde makalelerin yapısal özelliklerini hesaba katmanın başarımı artırdığı gösterilmiştir. Ek 1, Ek 2 ve Ek 3'te orijinal özler ile birlikte ilgili makalelerden elde edilen sistem çıktıları örnek olarak incelenebilir.

Özetleyecek olursak, araştırma kapsamında geliştirilen çıkarıma dayalı otomatik özetleme sisteminin çıktıları makalelerin tam metinlerinin orijinal özlerine göre belirgin bir şekilde daha iyi okunurluğudur. Buna ek olarak otomatik özetleme sisteminde makalelerin yapısal özelliklerini hesaba katmanın hem çıktılarının tam metni tarama başarımını, hem de ROUGE ölçümlerine göre orijinal özler ile n-gram çakışmalarını artırdığı bilgisi elde edilmiştir. Tam metnin yapılandırılma oranı arttıkça yapısal olanlarda daha belirgin olmak üzere orijinal özlere benzerlik oranlarının artması araştırmanın ilgi çekici bulgularından birisidir.

Çalışma kapsamında makale cümleleri tam metinlerin IMRAD bölümlerine göre işaretlenmiştir. Bu sayede, geliştirilen özetleme sistemi, tam metnin yapısal özelliklerinin çıktı başarımı üzerindeki etkisine göre değerlendirilebilmiştir. Bu değerlendirmenin, okunurluk, tam metni tarama aralığı ve orijinal özler ile n-gram çakışması boyutlarıyla gerçekleştirilmesi bu çalışmanın özgün yönüdür. Tam metin yapısal içeriklerinin çıktı başarımlarına katkısına dair elde edilen tüm bulgular birlikte değerlendirildiğinde; bilimsel makaleler için çıkarıma dayalı otomatik öz hazırlamada, tam metinlerin yapısal özelliklerini göz önüne almanın, üretilecek özlerin kalitesini artırdığı şeklindeki araştırma hipotezi doğrulanmaktadır.

Çalışmadan elde edilen sonuçlara göre, çıkarıma dayalı otomatik yöntemler ile tam metin içeriğini yapısal bölümlerine göre iyi tarayan, orijinal yazar özlerinden daha iyi okunurluklu otomatik yapısal özler elde etmenin mümkün olduğu gösterilmiştir. Elde edilen bu sonuç bilimsel iletişim bağlamında yaşanan pek çok soruna çözüm olabilecek niteliktedir. Çıkarıma dayalı yöntemlerle seçilen, bu iyi okunurluklu yapısal öz cümleleri, ilgili yapısal bölüm başlıkları altında, dijital ortamın doğasına daha uygun sunum özelliklerine de sahip olacaktır. Yazarlar için de tam metnin çok kısa bir özetini, tam metni çok iyi tarayarak ve olası tüm okurların ilgisini çekerek hazırlamak kolay bir iş değildir. Otomatik yapısal öz sistemlerinin çıktıları orijinal öze alternatif olarak değerlendirileceği gibi, yazarlar için yardımcı taslaklar olarak da kullanılabilir. Bu tür sistemlerin yaygın kullanımının bilimsel iletişime farklı bir bakış açısı getirerek, araştırmacıların özlerde kullandıkları zor okunurluklu dilin özelliklerini de değiştireceği düşünülmektedir. Bilim insanları daha çok bilgiyi daha kolay ve daha kısa sürede işlemek istemektedirler. Buna karşın artan bilgi hacmi karşısında yaşadıkları zorluklar her gün artmaktadır. İnsanın bilişsel kapasitesinde büyük gelişmeler yaşanması bugün için beklenmediğine göre bu tez ile geliştirilen otomatik yapısal öz sistemi ile bilimsel iletişim sürecinde hissedilen bir soruna çözüm getirmek mümkündür.

Bu çalışmadan elde edilen bulgular ve oluşturulan veri seti kullanılarak daha üstün başarıma sahip öz elde etme performansı için farklı deneyler gerçekleştirilebileceği değerlendirilmektedir. Bu bağlamda tez çalışmasının sonuçlarının ve Türkçe kütüphanecilik ve bilginbilim alanında yayımlanmış 421 makaleden oluşan, derlemde elde edilen yapılandırılmış verinin başka araştırmalarda kullanılabileceği düşünülmektedir. Bu çalışmanın devamında aşağıda sıralanan konularda ek çalışmalar yapılmasının faydalı olacağı değerlendirilmektedir:

- Derlemde dört IMRAD bölümü kullanılan makaleler ile sadece giriş ile tartışma ve sonuçlar bölümleri içeren makalelerin yoğunluklarının birbirine yakın olması ilgi çekici bulgulardan biridir. Sadece giriş ile tartışma ve sonuçlar bölümlerinden oluşan makalelerin ne kadarında yapısal düzenleme hatası yapıldığı, ne kadarının ilgili makalenin türüne özgü bir format olarak şekillendiğinin irdelenmesi gereklidir. Bu bulgudan yola çıkarak sosyal bilimler alanına özgü yapısal



formatlar üzerine yoğunlaşılacak ileri çalışmaların anlamlı olacağı değerlendirilmektedir.

- Makalelerin tam metinlerinde yapısal düzene uygunluğun artışıyla, otomatik yöntemlerle tespit edilen temsilci cümlelerin tam metni yansıtma başarımlarının da artmasının sebepleri yapısal olan ve olmayan sistem çıktıları bağlamında değerlendirildiğinde:
  - Otomatik yapısal özler tam metnin tüm yapısal bölümlerinden cümle seçtiğine göre, yazar özleri ile aralarındaki benzerliğin artması ilgili orijinal özlerin yapısal içeriğinin de artışının bir işareti olarak kabul edilebilir. Buna göre, sadece tam metinlerde yapısal yazma alışkanlığının daha fazla teşvik edilmesi ile yazar tarafından üretilen orijinal özlerin yapısal içeriğinin iyileşmesine de katkı sağlanabileceği düşünülmektedir.
  - Bu bulgu her iki çıktı grubu için bilgiye erişim bağlamında ele alındığında yapısal düzene uygun üretilen tam metinlerde yazarın aktarmayı amaçladığı bilgiyi duyarlık ve anma kavramlarına göre daha isabetli içeren cümlelerin yer aldığı anlamına da gelir. Sistem çıktılarının bu bağlamdaki başarımlarına karar vermede yazarların ürettikleri orijinal özler ile benzerlikleri dikkate alınmaktadır. Orijinal özleri de tam metinlerdeki yapısal içeriklere göre irdeleyecek ek çalışmalar ile bu sebeplere ışık tutulabilir.
  - Bu bakış açısıyla, yapısal düzene uygun yazan ve yazmayan yazar profilleri üzerinde durulacak yeni araştırma konuları ile yapısal formatların tam metin ve orijinal özlerin kalitesine etkilerini irdelemenin mümkün olacağı değerlendirilmektedir.
- Derlem tam metinlerinin yıllara göre IMRAD düzenine uyumlarının arttığı da göz önüne alınarak, IMRAD'a uygun format kullanımının deneysel araştırma alanlarıyla sınırlı olduğu görüşünün yeniden gözden geçirilmesi önerilmektedir. Yapısal içeriklere dair elde edilen sonuçların sosyal bilimlere ait başka başlıklarda da makale tam metinleri üzerinde yapılacak yapısal içerik incelemeleri ile birleştirilerek alana özgü yapısal formatlar üretilmesinin mümkün olacağı düşünülmektedir.
- Bir otomatik özetleme sisteminde ağırlıklı olarak hangi özelliklerin kullanılması gerektiği uzman görüşleri ile belirlenebilir. Yapılacak kullanıcı araştırmaları ile

cümle seçme yaklaşımı daha iyi okunurluklu temsilci cümleler seçilecek şekilde inşa edilebilir. Bu şekilde, mevcut özellik setinde yer alan niteliklerin ağırlıklarını değiştirerek ya da örneğin cümlelerin kelime sayıları ya da kelimelerin hece sayılarının temsilci cümle seçme algoritmasına etki etmesini sağlayacak yeni özellikleri sistemde deneyerek okunurluk başarımları artırılabilir. Çalışma kapsamında geliştirilen sistemde, seçilecek temsilci cümlelerin kelime sayılarını artırmaya sebep olabilecek isim soylu kelime sayısı özelliği, diğer özelliklerle eş ağırlıklı olarak kullanılmıştır. Buna rağmen, çıktılar, orijinal özlerden daha kolay okunurluk düzeyindedir. Bu bakış açısıyla daha sonraki çalışmalar için cümle ve kelime uzunluklarını da hesaba katarak daha da iyi okunurluklu sistem çıktıları elde edebilmenin mümkün olacağı düşünülmektedir.

- Otomatik özetleme sistemleri tarafından üretilen özet metinlerin henüz insan özetleyicilerin dil ve anlatım düzeyine ulaşmadığı bilinmektedir. Bunun en önemli sebeplerinden biri seçilen cümlelerin, aralarında bir bağlam kurulmadan sıralanmasıdır. Çalışma kapsamında geliştirilen çıkarıma dayalı yapısal otomatik özetleme sisteminde yapısal başlıklar kullanarak öz üretmenin bu bağlam sorununu çözmeye katkı sunabileceği düşünülmektedir. Yapısal özlerin elle değerlendirmeleri için kullanıcı araştırmaları planlanabilir. Bu değerlendirme orijinal öz, OKÖ ve OYÖ'ler ile karşılaştırmalı olarak yapıldığında yapısal özelliklerin bağlam sorununa katkısının da ortaya konulabileceği değerlendirilmektedir.
- Bu çalışma kapsamında elde edilen, yapısal özelliklerine göre işaretlenmiş veri setini kullanacak öğrenen sistemler ile Türkçe akademik makalelerden oluşan derlemelerin yapısal bölümlerini otomatik yöntemlerle tespit edebilmenin mümkün olabileceği düşünülmektedir. Böylece, daha kolay okunurluklu ve tam metnin tüm yapısal bölümlerini içeren özler, tamamen otomatik yöntemler ile üretilebilir. Çalışmanın bu yönüyle başta bilginlik ve kütüphanecilik alanı olmak üzere diğer alanlarda da Türkçe bilimsel iletişime katkı sağlayacağı düşünülmektedir.
- Geliştirilen sistemin çıktıları ile planlanacak kullanıcı araştırmalarının, bilimsel iletişim sürecinde okur beklentilerine yeni bir bakış açısı getirmesinin mümkün olacağı görülmektedir. Çalışma kapsamında elde edilen bulgular, kullanıcı araştırmaları ile desteklenerek, bilimsel makalelerin hangi yapısal bölümlerinden

hangi ağırlıkta cümle seçildiğinde okuyucu için daha bilgilendirici özler elde edilebildiği araştırılabilir. Özler üzerinde, tam metnlerinin yapısal özelliklerini de ele alan ileri çalışmalar ile bilgi çağında bilimsel iletişime sağlanacak farklı bakış açısının yanı sıra otomatik özetleme literatürüne de katkı sağlamak mümkündür.

Çalışma kapsamında elde edilen bulgular doğrultusunda, bilimsel makalelerin sadece üstverisinin değil tam metnlerinin de araştırmacılar için ümit verici bir uygulama alanı olabileceği değerlendirilmektedir. Akademik makalelerin tam metnlerinden elde edilen bulguların otomatik yapısal öz elde eden sistemde kullanımının otomatik metin özetleme sürecine katkısı göz önüne alındığında, akademik makaleler üzerinde yapılan otomatik özetleme araştırmalarında derlemlerin tam metin yapısal özelliklerinin dikkate alınması önerilmektedir. Bu öneriye ek olarak, bu tür uygulamalarının sadece bilgisayar bilimleri ya da dilbilimi alanında değil kütüphanecilik ve ilgili diğer alanlarda da devam ettirilmesinin bilgi çağında bilimsel iletişimin gelişimi için yararlı olacağı düşünülmektedir.

## KAYNAKÇA

- Aktan, C.C. ve Vural, İ.Y. (2005). Bilgi toplumu, yeni temel teknolojiler ve yeni ekonomi. *Bilgi Çağı, Bilgi Yönetimi ve Bilgi Sistemleri*, 88 (1), 31-81.
- Al, U. ve Sezen, U. (2018), *Türkçe atıflar için içerik tabanlı atıf analizi modeli tasarımı*. (TÜBİTAK Sosyal Bilimler Araştırma Grubu - Proje No: SOBAG 115K440). Hacettepe Üniversitesi Bilgi ve Belge Yönetimi Bölümü.
- Altan, Z. (2004). A Turkish automatic text summarization system. *IASTED International Conference on Artificial Intellegence and Applications*. 311-316.
- APA. (2010). *The publication manual of the American Psychological Association* (6. Baskı b.). Washington DC: American Psychological Association.
- Atanassova, I., Bertin, M. ve Mayr, P. (2015). Mining scientific papers for bibliometrics: a (very) brief survey of methods and tools. *15th International Conference of the International Society for Scientometrics and Informetrics (ISSI) on Scientometrics and Informetrics (ISSI)*. 1091-1092. Turkey, Istanbul: Bogazici Univ.
- Atanassova, I., Bertin, M. ve Larivière, V. (2016). On the composition of scientific abstracts. *Journal of Documentation*, 72(4), 636-647.
- Ateşman, E. (1997). Türkçede okunabilirliğin ölçülmesi. *Dil Dergisi*, 58, 71-74.
- Athreya, B.H. ve Mouza, C. (2017). Introduction to thinking skills for the digital generation. *Thinking Skills for the Digital Generation* içinde, 1-10, Springer, Cham.
- Bilgi Dünyası. (2012). Yazar Rehberi. Bilgi Dünyası. Erişim adresi: <http://bd.org.tr/index.php/BD/pages/view/guideForAuthors#writingRules>
- Binwahlen, M., Salim, N. ve Suanmali, L. (2010). Fuzzy swarm diversity hybrid model for text summarization. *Information Processing & Management*, 46(5), 571-588.
- Bitri, E. ve Keseroğlu, H. (2015). Türk Kütüphaneciliği ve Bilgi Dünyası dergilerinin özlerine eleştirel bir bakış. *Türk Kütüphaneciliği*, 29(2), 241-257.
- Bloom, B.S. (1956). Taxonomy of educational objectives. *Bloom Taxonomy of Educational Objectives*, New York: David McKay Company. Inc.
- Bornmann, L. ve Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, 66(11), 2215-2222.
- Carr, N. (2008). Is Google making us stupid? *The Atlantic Monthly*. Erişim adresi: <https://www.theatlantic.com/magazine/archive/2008/07/is-google-making-us-stupid/306868/>
- Carr, N. (2010). The web shatters focus, rewires brains. *Wired*. Erişim adresi: [http://www.wired.com/2010/05/ff\\_nicholas\\_carr](http://www.wired.com/2010/05/ff_nicholas_carr)

- Crochemore M. ve Rytter W. (1994). *Text algorithms*. Oxford Univ. Press, Oxford, UK.
- Crowther, J.G. (2013). *British scientists of the nineteenth century* NY: Routledge.
- Çoban, A. (2014). Okunabilirlik kavramına yönelik bir derleme çalışması. *Dil ve Edebiyat Eğitimi Dergisi*, 9, 96-111.
- Day, R.A. (1996). *Bilimsel makale nasıl yazılır ve yayımlanır?* (A. G. A., Çev.) Ankara: TÜBİTAK Matbaası.
- Dewan, P. (2012). Are books becoming extinct in academic libraries? *New Library World*, 113(1/2), 27-37.
- Dolnicar, S. ve Chapple, A. (2015). The readability of articles in tourism journals. *Annals of Tourism Research*, 52, 161–166.
- Dronberger, G.B. ve Kowitz, G.T. (1975). Abstract readability as a factor in information systems. *Journal of the American Society for Information Science*, 26(2), 108-111.
- Drucker, P. F. (2012). *Managing in a time of great change*. Yayın yeri: Routledge.
- DuBay W.H. (2004) *The principles of readability. Impact Information*. Costa Mesa, CA: Yayınevi.
- Dumais, S.T. (1991) Improving the retrieval of information from external sources. *Behavior Research Methods, Instruments & Computers*, 23(2), 229–236.
- Edmundson, H. (1969). New methods in automatic extracting. *Journal of the ACM*, 16(2), 264-285.
- Fallon, H. (2009). A writing support programme for Irish academic librarians. *Library Review*, 66(8-9), 414-422.
- Fattah, M.A. ve Ren, F. (2009). GA, MR, FFNN, PNN and GMM based models for automatic text summarization. *Computer Speech & Language*, 23(1), 126-144.
- Ferreira, R., de Souza Cabral, L., Lins, R. D., e Silva, G.P., Freitas, F., Cavalcanti, G.D., Lima, R., Simske, S.J. ve Favaro, L. (2013). Assessing sentence scoring techniques for extractive text summarization. *Expert systems with applications*, 40(14), 5755-5764.
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32(3), 221–233.
- Fontelo, P., Gavino, A. ve Sarmiento, R. (2013). Comparing data accuracy between structured abstracts and full-text journal articles: implications in their use for informing clinical decisions. *Evidence-based medicine*, 18(6), 207-211.
- Gambhir, M. ve Gupta, V. (2017). Recent automatic text summarization techniques: a survey. *Artificial Intelligence Review*, 47, (1), 1-66.
- Ganesan K. (2018a). ROUGE 2.0: Updated and improved measures for evaluation of summarization tasks, Erişim adresi: <https://arxiv.org/abs/1803.01937>

- Ganesan K. (2018b). What is ROUGE and how it works for evaluation of summaries? Erişim adresi: <https://kavita-ganesan.com/what-is-rouge-and-how-it-works-for-evaluation-of-summaries/#.YHROkOgza00>
- Gazni, A. (2011). Are the abstracts of high impact articles more readable? Investigating the evidence from top research institutions in the world. *Journal of Information Science*, 37(3), 273-281.
- Güran, A., Arslan, S., Kılıç, E. ve Diri, B. (2014). Metin özetleme için cümle seçim metotları. *22nd Signal Processing and Communications Applications Conference (SIU)*, 192-195. Piscataway/NJ.
- Güran, A., Uysal, M., Ekinci, Y. ve Güran, C. (2017). An Additive FAHP based sentence score function for text summarization. *Information Technology And Control*, 46(1), 53-69.
- Hartley, J. (2008). *Academic writing and publishing: A practical handbook*. London: Routledge.
- Hartley, J. ve Betts, L. (2007). The effects of spacing and titles on judgments of the effectiveness of structured abstracts. *Journal of the American Society for Information Science and Technology*, 58(14), 2335-2340.
- Hartley, J. ve Betts, L. (2009). Common weaknesses in traditional abstracts in the social sciences. *Journal of the American Society for Information Science and Technology*, 60(10), 2010-2018.
- Hartley, J. ve Cabanac, G. (2017). Thirteen ways to write an abstract. *Publications*, 5(2), 1-7.
- Hartley, J., Pennebaker, J. ve Fox, C. (2003). Abstracts, introductions and discussions: How far do they differ in style? *Scientometrics*, 57(3), 389-398.
- Hovy, E. (2005) Automated Text Summarization. *The Oxford Handbook of Computational Linguistics*. Oxford University Press, Oxford, United Kingdom, 583-598.
- Hovy, E. ve Lin, C. (1999). Automated text summarization in SUMMARIST. *Advances in Automatic Text Summarization*, 14, 81-94.
- Indiana University Writing Tutorial Services. (2020). Writing Tutorial Services. Indiana University. Erişim adresi: <https://wts.indiana.edu/writing-guides/>
- Issa, T. ve Isaias, P. (2016). Internet factors influencing generations Y and Z in Australia and Portugal: A practical study. *Information Processing & Management*, 52(4), 592-617.
- Jamar, N., Šauperl, A. ve Bawden, D. (2014). The components of abstracts: The logical structure of abstracts in the areas of materials science and technology and of library and information science. *New Library World*, 115(1/2), 15-33.

- Kızıltepe, R., Irmak, T. Y., Eslek, D. ve Hecker, T. (2020). Prevalence of violence by teachers and its association to students' emotional and behavioral problems and school performance: Findings from secondary school students and teachers in Turkey. *Child Abuse & Neglect*, 107, 104559.
- Kurbanoğlu S. (2013). An analysis of concept of information literacy. *Media and Information Literacy for Knowledge Societies*, 78-87.
- Kutlu, M., Cıgır, C. ve Çiçekli, İ. (2010). Generic text summarization for Turkish. *The Computer Journal*, 1315-1323.
- Labaree, R. V. (2018). Organizing Your Social Sciences Research Paper. Erişim adresi: USC Libraries: <https://libguides.usc.edu/writingguide/>
- Lin, C.Y. (2004). Rouge: A package for automatic evaluation of summaries. Text Summarization Branches Out: Proceedings of the ACL-04 Workshop, 74-81.
- Lloret, E. ve Palomar, M. (2012). Text summarisation in progress: a literature review. *Artificial Intelligence Review*, 37(1), 1-41.
- Luhn, H. (1958). The Automatic Creation of Literature Abstracts. *IBM Journal of Research and Development*, 2(2), 159-165.
- Mabe, M. A. ve Amin, M. (2002). Dr. Jekyll and Dr. Hyde: Author-reader asymmetries in scholarly publishing. *Aslib Proceedings*, 54(3), 149-157.
- Mangen, A., Walgermo, B. ve Brønck, K. (2013). Reading linear texts on paper versus computer screen: Effects on reading comprehension. *International Journal of Educational Research*, 58, 61-68.
- Mani, I., House, D., Klein, G., Hirschman, L., Firmin T. ve Sundheim, B. (1999). The TIPSTER SUMMAC text summarization evaluation. In Proceedings of the Ninth Meeting of the European Chapter of the Association for Computational Linguistics (EACL-99), 77-85.
- Meadows, A.J. (1985). The scientific paper as an archaeological artefact. *Journal of Information Science*, 11(1), 27-30.
- Meadows, A.J. (1997). *Communicating research*. New York: Academic Press.
- Merzenich, M. (2008). Going Googly. on the brain. Erişim adresi: <http://onthebrain.com/2008/08/going-googly/>
- Mulrow, C., Thacker, S. ve Pugh, J. (1988). A proposal for more informative abstracts of review articles. *Annals of Internal Medicine*, 108, 613-615.
- Nicholas, D., Huntington, P. ve Jamali, H.R. (2007). The use, users, and role of abstracts in the digital scholarly environment. *The Journal of Academic Librarianship*, 33(4), 446-453.
- Özdemir, E. (2000). *Eleştirel okuma*. Ankara: Bilgi Yayınevi.

- Özsoy, M., Çiçekli, İ. ve Alpaslan, F. (2010). Text summarization of turkish texts using latent semantic analysis. *23rd International Conference on Computational Linguistics (Coling 2010)*, 869-876.
- Pembe, C. (2011). *Automated query-biased and structure-preserving document summarization for web search tasks*. (Yayımlanmamış doktora tezi). İstanbul: Boğaziçi Üniversitesi.
- Plavén-Sigray, P., Matheson, G. J., Schiffler, B. C. ve Thompson, W. H. (2017). The readability of scientific texts is decreasing over time. *elife*, 6, e27725.
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>.
- Rodrigues, V. (2013). How to write an effective title and abstract and choose appropriate keywords. *Editage Insights*. Erişim adresi: <https://www.editage.com/insights/how-to-write-an-effective-title-and-abstract-and-choose-appropriate-keywords>
- Saggion, H., Radev, D. R., Teufel, S., Lam, W. ve Strassel, S. M. (2002). *Developing infrastructure for the evaluation of single and multi-document summarization systems in a cross-lingual environment*. Language Resources And Evaluation 2002’de sunulan bildiri, 747-754.
- Saggion, H., ve Poibeau, T. (2013). Automatic text summarization: Past, present and future. Multi-source, multilingual information extraction and summarization, Springer Berlin Heidelberg, 3-21.
- Salton, G., ve Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5), 513-523.
- Šauperl, A., Klasinc, J. ve Lužar, S. (2008). Components of abstracts: Logical structure of scholarly abstracts in pharmacology, sociology, and linguistics and literature. *Journal of the American Society for Information Science and Technology*, 59(9), s. 1420-1432.
- Sharma, S., ve Harrison, J. (2006). Structured abstracts: do they improve the quality of information in abstracts? *American Journal of Orthodontics and Dentofacial Orthopedics*, 130, 523-530.
- Singer, L. ve Alexander, P. (2017). Reading on paper and digitally: What the past decades of empirical research reveal. *Review of Educational Research*, 87(6), 1007-1041.
- Sollaci, L. B., ve Pereira, M. G. (2004). The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey. *Journal of the Medical Library Association*, 92(3), 364-371.
- Steinberger, J. ve Jezek, K. (2009). Evaluation measures for text summarization. *Computing and Informatics*, 28, 251-275.



- Taşkın, Z. ve Al, U. (2018). A content-based citation analysis study based on text categorization. *Scientometrics*, 114(1), 335-357.
- Tenopir, C. ve King, D. (2014). The growth of journals publishing. *The Future of the Academic Journal*, Oxford: Elsevier Limited, (2. bs.) 159-178.
- Toffler, A. (1980). *The third wave*. New York: Morrow.
- Torres-Moreno, J.M. (2014). *Automatic text summarization*. Wiley. Erişim Adresi: <https://onlinelibrary.wiley.com/doi/book/10.1002/9781119004752>
- Türk Kütüphaneciliği. (2018). Yazar Rehberi. Türk Kütüphaneciliği. Erişim adresi: <http://www.tk.org.tr/index.php/TK/about/submissions#authorGuidelines>
- Ufnalska, S. ve Hartley, J. (2009). How can we evaluate the quality of abstracts? *European Science Editing*, 35(3), 69-71.
- Wästlund, E. (2007). *Experimental studies of human-computer interaction: Working memory and mental workload in complex cognition*. Goteborg, Sweden: Gothenburg University, Department of Psychology.
- Wu, J. (2011). Improving the writing of research papers: IMRAD and beyond. *Landscape Ecology*, 26(10), 1345-1349.
- Zemberek-NLP. (2018). [Yazılım]. Tedarik edilebileceği adres: <https://github.com/ahmetaa/zemberek-nlp>

## EKLER

### **EK 1. BD200023, Yazar tarafından üretilmiş orijinal öz, OYÖ ve OKÖ**

#### Orijinal Öz

Bu makalede, Türkiye'deki 37 üniversite kütüphanesinin Web sitelerinin içerikleri, seçilen 19 ölçüte dayalı olarak değerlendirilmiştir. Üniversite kütüphane Web sitelerinin çoğu, kütüphane ve sunulan hizmetlere ilişkin bilgi içermektedir. Kütüphanelerin yaklaşık %60'ının çevrimiçi katalogları Web üzerinde bulunmaktadır. Üniversite kütüphane Web sitelerinin çoğunda yararlı olabilecek bazı başka sitelere bağlantı olmamasına karşın kendi sitesi içinde arama fonksiyonu olanların sayısı çok azdır. Üniversite kütüphanelerinin Web sitelerinin daha fazla bilgi, yararlı dokümanlar ve hizmetler ile zenginleştirilmesi önerilmektedir.

#### OYÖ

##### **Giriş**

Bu makalede, Türkiye'de üniversite kütüphanelerine ait Web sayfalarının içeriği, özellikle kütüphanelere yönelik ölçütlere uygunluğu açısından incelenmektedir. Web sitelerinin değerlendirilmesinde iki nokta ön plana çıkmaktadır: Stover (1997, s. 55) kütüphane Web sitelerinin amaç ve işlevlerinin kütüphane türüne bağlı olduğunu ifade etmektedir. Temel kütüphane bilgileri: Kütüphane ve bilgi hizmetleri:

##### **Yöntemler**

Bu amaçla 37 üniversite kütüphane Web sitesi (bkz. Ek 1. Üniversite Kütüphanelerinin Web Sitelerinin Adresleri) 19-23 Ekim 2000 tarihleri arasında ziyaret edilmiştir. İlk grupta yer alan dört değerlendirme ölçütü Web sitesiyle ilgilidir (üniversitenin ana sayfasından kütüphanenin Web sitesine, kütüphanenin Web sitesinden üniversitenin ana sayfasına bağlantılar, kütüphane ana sayfasının son güncelleme tarihi ve yazarı).

##### **Bulgular**

Dolayısı ile bu iki üniversitenin ve kütüphanelerinin Web siteleri değerlendirilmemiştir. Web tasarımı açısından dikkat edilen ölçütlerden birisi de sitede/sayfalarda Web tasarımcısına ait isim, e-posta adresi gibi erişim bilgilerinin bulunmasıdır. Böylece kütüphaneye ait siteyi/sayfaları kimin hazırladığı belirlenebilmektedir. Sayfalarında çalışma saatlerini veren 23 kütüphane (%62) vardır. Ziyaret edilen kütüphane sayfaları bu açıdan da değerlendirilmiştir.

##### **Tartışma ve Yorumlar**

Alt sayfalarda telif hakkı tarihi verilmektedir. Kütüphanenin ana sayfası kısa ve basit olarak tasarlanmış, kütüphane bilgileri ve hizmetlerine ait bağlantılar

kolaylıkla görülecek biçimde listelenmiş ve bu bilgiler sayfa altında da tekrarlanmıştır. Söz konusu ölçütler genellikle kütüphane hakkında sayfalar ve bilgi hizmetlerine yöneliktir. Eksik birkaç sayfa/hizmet dışında bu kütüphanelerin sitelerinde çoğunlukla yardım veya SSS sayfaları ve site içinde arama fonksiyonuna rastlanmamaktadır.

OKÖ

Bilgi toplumu olmayı hedefleyen ülkeler için Internet ve bilgi teknolojisi giderek önem kazanmaktadır. Bu makalede, Türkiye'de üniversite kütüphanelerine ait Web sayfalarının içeriği, özellikle kütüphanelere yönelik ölçütlere uygunluğu açısından incelenmektedir. Aynı sunucu üzerine konulan -genellikle birbirleriyle ilgili- sayfaların koleksiyonu "Web sitesi", HTML formatında erişilebilecek bir dosya "Web sayfası", bir sayfa veya bağlantılı sayfalar "Web dokümanı" olarak anılmaktadır (Poulter ve diğerleri, 1999, s. 13). Web sitelerinin değerlendirilmesinde iki nokta ön plana çıkmaktadır: Tasarım ve içerik. Web sitelerinin içerik olarak değerlendirilmesinde ise, sitenin amacı, sunulan bilgi ve hizmetler ile bağlantılar önem taşımaktadır. Web sayfaları ile ilgili temel bilgiler, kurallar, öneriler, Web tasarımında kullanılan resim ve benzeri görüntülere ilişkin dosyalar ve bağlantılara sahip ücretli/ücretsiz Web siteleri giderek çoğalmaktadır. İnternetin yaygınlaşması kütüphane hizmetlerinde yeni olanaklar yaratmakta, Web araçlarının gelişmesi de bilgiye yerel ve uluslararası boyutta erişimi kolaylaştırmaktadır. Web sitelerinin kütüphane ve bilgi hizmetleri alanında da yaygınlaşması ile içerik ve tasarım açısından yapılan çalışma, makale, Web sayfası ve elektronik tartışma listesi sayısı artmaktadır. Stover (1997, s. 55) kütüphane Web sitelerinin amaç ve işlevlerinin kütüphane türüne bağlı olduğunu ifade etmektedir. Temel kütüphane bilgileri: Kütüphane ve bilgi hizmetleri: Uzaktaki bilgi kaynaklarına bağlantı: Çalışmamızda yukarıdaki gruplamaya dayanarak oluşturulan 19 ölçüte göre üniversite kütüphanelerinin Web siteleri değerlendirilmiştir. Dolayısı ile bu iki üniversitenin ve kütüphanelerinin Web siteleri değerlendirilmemiştir. Web tasarımı açısından dikkat edilen ölçütlerden birisi de sitede/sayfalarda Web tasarımcısına ait isim, e-posta adresi gibi erişim bilgilerinin bulunmasıdır. Sayfalarında çalışma saatlerini veren 23 kütüphane (%62) vardır. Ziyaret edilen kütüphane sayfaları bu açıdan da değerlendirilmiştir.

## **EK 2. BD201213, Yazar tarafından üretilmiş orijinal öz, OYÖ ve OKÖ**

### **Orijinal Öz**

Bu çalışmada, Türkiye'nin 2000 yılı sonrası bilim-teknoloji politikalarında kütüphane kurumuna ilişkin yaklaşımın değerlendirilmesi amaçlanmaktadır. Öncelikle konuyla ilgili kavramsal bilgiler verilmiş, daha sonra Teknoloji Geliştirme Bölgeleri Kanunu (2001), Araştırma ve Geliştirme Faaliyetlerinin Desteklenmesi Hakkında Kanun (2008), Bilim, Sanayi ve Teknoloji Bakanlığının Teşkilat ve Görevleri Hakkında Kanun Hükmünde Kararname (2011), Teknoloji Geliştirme Bölgeleri Uygulama Yönetmeliği (2002), Araştırma ve Geliştirme Faaliyetlerinin Desteklenmesine İlişkin Uygulama ve Denetim Yönetmeliği (2008), 5746 Sayılı Araştırma ve Geliştirme Faaliyetlerinin Desteklenmesi Hakkında Kanun Genel Tebliği (2008), Sanayi ve Ticaret Bakanlığı 2010-2014 Stratejik Planı, Uzun Vadeli Strateji ve Sekizinci Beş Yıllık Kalkınma Planı 20012005, Dokuzuncu Kalkınma Planı (2007-2013), Bilgi Toplumu Stratejisi (2006-2010) ve Eylem Planı, Orta Vadeli Program (2012-2014), DPT Stratejik Planı (2009-2013), Türk Bilim ve Teknoloji Politikası (1993-2003), Ulusal Bilim ve Teknoloji Politikaları: 2003-2023 Strateji Belgesi (Vizyon 2023), Bilim ve Teknoloji Politikaları Uygulama Planı (2005-2010), Uluslararası Bilim, Teknoloji ve Yenilik (BTY) Stratejisi Uygulama Planı (2007-2010), Ulusal Yenilik Stratejisi (2008-2010), Bilim ve Teknoloji Yüksek Kurulu 22.Toplantısı (2010), Ulusal Bilim, Teknoloji ve Yenilik Stratejisi (2011-2016), Bilim ve Teknoloji İnsan Kaynağı Stratejisi ve Eylem Planı (2011-2016) ve TÜBİTAK 2008-2012 Stratejik Planı, Türkiye Ulusal Programı (2008), Türkiye'nin AB Müktesebatına Uyum Programı (2007-2013), 59, 60 ve 61. Hükümet Programı, Adalet ve Kalkınma Partisi (AKP), Cumhuriyet Halk Partisi (CHP), Milliyetçi Hareket Partisi (MHP) ve Barış ve Demokrasi Partisi (BDP) Programları incelenmiştir. Çalışma sonucunda, Türkiye'nin bilim-teknoloji politikalarında kütüphane kurumuna ilişkin yaklaşımın son derece zayıf olduğu ve kütüphane kurumunun bu politikaların zorunlu bir parçası olarak görülmediği anlaşılmıştır. Kütüphanecilerin çağdaş bir bilim-teknoloji politikası oluşturulması için politika süreçlerine müdahale etmeleri önerilmiştir.

## OYÖ

### **Giriş**

Doğrudan ilgisi, organik parçalardan birisi olması bağlamında kütüphane kurumunun bilim-teknoloji politikaları içinde ele alınması, değerlendirilmesi ya da bu kuruma ilişkin güçlü bir yaklaşımın oluşturulması zorunlu görünmektedir. Literatürde, bilim-teknoloji politikaları bağlamında kütüphane kurumunu doğrudan ele alan çalışma sınırlılığı bulunmakla birlikte eğitim, kültür ve kütüphane ilişkisi üzerine bazı çalışmalar vardır.

### **Yöntemler**

Betimleme yöntemiyle gerçekleştirilen ve 2000 yılı sonrasını kapsayan bu araştırmada, Türkiye'nin bilim-teknolojiye yönelik politika/strateji belgelerinde kütüphane kurumu algısı ve yaklaşımının değerlendirilmesi amaçlanmıştır. Bu çerçevede bilim-teknoloji politikaları olarak değerlendirilebilecek yasalar, yönetmelikler, yönergeler, tebliğler, kalkınma planları, ulusal programlar, politika ve strateji belgeleri, toplantılar, stratejik planlar ve bazı parti programları incelenmiştir.

### **Bulgular**

Bilim ve teknoloji üretiminde kütüphaneyi merkeze alan bu yaklaşımlar kuşkusuz kütüphanenin bilimsel ve teknolojik gelişmeler için anahtar rolüne işaret eder. Bilim-teknoloji politikası içinde kütüphanenin yükleneyeceği rol ve işlevler yasal dayanaklara sahip olmalıdır. Bilim-teknoloji politikası kütüphane örgütlenmesinin nasıl olacağını açıklamalıdır. Ancak, kütüphane kurumu ile ilgili konular bağlamında ele alıp almadığı değerlendirilebilir. Ancak, bu kısımlarda kütüphaneden söz edilmemektedir.

### **Tartışma ve Yorumlar**

3. Kütüphane kurumu, Hükümet'in ve mecliste bulunan siyasi partilerin bilim-teknoloji politikaları kapsamında değerlendirilmemiştir.4. İncelenen Kalkınma Planlarında kütüphane kurumuna ilişkin olarak bilim-teknoloji politikaları bağlamında herhangi bir değerlendirilme görülmemiştir.10. Türkiye'nin bilim-teknoloji politikalarında kütüphane, genelde araştırmalar için destek hizmetleri sınıfında ve yardımcı bir unsur olarak düşünülmektedir.Bu bir anlamda, bilim-teknoloji politikalarında kütüphaneye ilişkin işlevlerin de sınırlanması anlamına gelmektedir.

## OKÖ

Söz konusu politikaların bu konuda hedefler, stratejiler, ilkeler, planlar, programlar ve projeler olarak sıralanabilecek politika unsurlarını içerdiği açıktır. Doğrudan ilgisi, organik parçalardan birisi olması bağlamında kütüphane kurumunun bilim-teknoloji politikaları içinde ele alınması, değerlendirilmesi ya da bu kuruma ilişkin güçlü bir yaklaşımın oluşturulması zorunlu görünmektedir. Literatürde, bilim-teknoloji politikaları bağlamında kütüphane kurumunu doğrudan ele alan çalışma sınırlılığı bulunmakla birlikte eğitim, kültür ve kütüphane ilişkisi üzerine bazı çalışmalar vardır. Betimleme yöntemiyle gerçekleştirilen ve 2000 yılı sonrasını kapsayan bu araştırmada, Türkiye'nin bilim-teknolojiye yönelik politika/strateji belgelerinde kütüphane kurumu algısı ve yaklaşımının değerlendirilmesi amaçlanmıştır. Bu çalışmada politika kavramı “yöntem” anlamıyla ele alınmaktadır. Bilim ve teknoloji üretiminde kütüphaneyi merkeze alan bu yaklaşımlar kuşkusuz kütüphanenin bilimsel ve teknolojik gelişmeler için anahtar rolüne işaret eder. Kütüphane kurumu ise bilimsel bilgi üretme sürecinin ana bileşenlerinden birisi olup bu işleviyle teknolojinin de altyapı unsurudur. Bilim-teknoloji politikası içinde kütüphanenin yükleneceği rol ve işlevler yasal dayanaklara sahip olmalıdır. Bilim-teknoloji politikası, hedefler çerçevesinde uygulanmak üzere kütüphanelere ilişkin kısa, orta ve uzun vadeli plan, proje ve program ve stratejiler geliştirmeye çerçeve oluşturmalıdır. Bilim-teknoloji politikası, kütüphaneleri ulusal kütüphane sistemi bütünlüğü içinde ve aynı zamanda türsel ayrımlarına bağlı işlevsel farklılıkları gözetilerek değerlendirmelidir. Bilim-teknoloji politikası kütüphane örgütlenmesinin nasıl olacağını açıklamalıdır. Bilim-teknoloji politikası kütüphanelerin elektronik kaynak ve hizmetlerine ilişkin yaklaşım ve standartları da belirlemelidir. Bir ülkede var olan bilim-teknoloji politikalarının kütüphane kurumuna ilişkin yaklaşımını değerlendirirken bu ilkelerin varlığı-yokluğu ölçüt olarak kullanılabilir. Ancak, kütüphane kurumu ile ilgili konular bağlamında ele alıp almadığı değerlendirilebilir. Ancak, bu kısımlarda kütüphaneden söz edilmemektedir.

### **EK 3. TK200221, Yazar tarafından üretilmiş orijinal öz, OYÖ ve OKÖ**

#### **Orijinal Öz**

Bilgi toplumunun bireylerine bilgi okuryazarlığı becerilerinin öğretilmesi büyük önem taşımaktadır. Bu yazıda, ilköğretim ve ortaöğretim öğretmenlerinin bilgi okuryazarlığı becerileriyle donatılmalarının gerekliliği tartışılmaktadır. Öğretmenlere bilgi okuryazarlığı becerilerini kazandırmaya yönelik bir çalışma ile bu çalışmadan elde edilen bulgular da aktarılacaktır.

#### **OYÖ**

##### **Giriş**

Okulların yeniden yapılandırılması çalışmaları bilgi okuryazarlığı ile doğrudan ilintilidir. Burada hem öğrencilerin hem de onları yönlendirme rolünü üstlenen öğretmenlerin söz konusu becerilere sahip olması gereği ortaya çıkmaktadır. Bilgi okuryazarlığı konusunda verilecek bir eğitimin öğretmenlerin bu alandaki bilgi düzeylerinde anlamlı bir farklılık yaratacağı varsayımından hareket edilmiştir.

##### **Yöntem**

Veri toplama aşamasında gözlem ve test tekniklerinden yararlanılmıştır. Eğitim süreci araştırmacılar tarafından yürütülmüş ve denekler programa karşı ilgileri sordukları sorular uygulama sırasında karşılaştıkları güçlükler açısından gözlemlenmiştir. Öntest ve sontest ortalamaları arasındaki fark başarı ölçütü olarak alınmıştır. Denek grubu söz konusu iki ilköğretim ve ortaöğretim kurumunda görev yapan toplam 154 öğretmen ve uzmandan oluşmaktadır. Eğitim araştırmacılar tarafından verilmiştir.

##### **Bulgular**

Eğitim sırasında konuya en çok ilgi gösteren ve en çok soru soran grup da sınıf öğretmenleri olmuştur. Testteki birinci soruda bir kütüphane kataloğunda bir kitaba ait ne tür bilgiler bulunabileceği sorulmaktadır. % 6 oranında bir artış söz konusudur. Deneklerin bu konulara eğitim sırasında ilgi gösterdikleri gözlenmiştir. Bu alanda bir eğitim ihtiyacı vardır.

##### **Tartışma ve Yorumlar**

Öğretmen yetiştiren kurumların programlarına konulacak olan bilgi okuryazarlığı dersleri için eğitim fakülteleriyle kütüphanecilik bölümleri arasında işbirliğine gidilmelidir. Okul kütüphanelerinde mutlaka kütüphanecilik bölümü mezunları çalıştırılmalıdır. Bilgi okuryazarlığı eğitimi verilecek grupların bilgisayar okuryazarlık düzeyleri de göz önüne alınarak oluşturulmaları daha iyi sonuçlar verecektir.

## OKÖ

21. yüzyıl toplumlarının gereksinim duyduğu insan profili değişmiştir. Eğitim kurumlarının en önemli işlevlerinden biri çağın gereksinim duyduğu insan gücünü yetiştirmektir. Okulların yeniden yapılandırılması çalışmaları bilgi okuryazarlığı ile doğrudan ilintilidir. Burada hem öğrencilerin hem de onları yönlendirme rolünü üstlenen öğretmenlerin söz konusu becerilere sahip olması gereği ortaya çıkmaktadır. Bu tür öğretim ortamlarının hazırlanması ve bilgi okuryazarlığının öğretim süreci içine alınması öğrencilerin eğitimi kadar öğretmenlerin ve kütüphanecilerin eğitimini, program geliştirme çalışmalarını ve öğretmen - kütüphaneci - eğitimci işbirliği çalışmalarını da kapsamına alan bir konudur. Artık yaşam boyu değişmeden devamlılığını sürdürebilen bir meslek grubu kalmamıştır. Bilgi okuryazarlığı becerilerine sahip olmak ise sadece yaşam boyu öğrenmenin ön koşulu olduğu için değil bilgiyi bulma, kullanma, değerlendirme ve iletme aşamalarında öğrencileri yönlendirebilmek açısından da öğretmenler için farklı bir önem taşımaktadır. Sözü edilen ve birbirini destekleyen iki öneriden biri öğretmen eğitiminin bilgi okuryazarlığını içermesinin sağlanması, diğeri bilgi okuryazarlığının lisans ve lisansüstü düzeyde öğretmen eğitimi programlarına dahil edilmesi konusunda akredite kurumlarının desteğinin alınması olarak belirtilmiştir. Bilgi okuryazarlığı konusunda verilecek bir eğitimin öğretmenlerin bu alandaki bilgi düzeylerinde anlamlı bir farklılık yaratacağı varsayımından hareket edilmiştir. Veri toplama aşamasında gözlem ve test tekniklerinden yararlanılmıştır. Eğitim süreci araştırmacılar tarafından yürütülmüş ve denekler programa karşı ilgileri sordukları sorular uygulama sırasında karşılaştıkları güçlükler açısından gözlemlenmiştir. Denek grubu söz konusu iki ilköğretim ve ortaöğretim kurumunda görev yapan toplam 154 öğretmen ve uzmandan oluşmaktadır. Eğitim araştırmacılar tarafından verilmiştir. Bu alanda bir eğitim ihtiyacı vardır.