



e-ISSN: 2148-4899

Pamukkale Üniversitesi İlahiyat Fakültesi Dergisi
Pamukkale University Journal of Divinity Faculty

Güz/Autumn, 2025, 12 (2), 93-120

VAHİYDEN VERİYE: *CORPUS.QURAN.COM*'UN ARAPÇA DERLEM DİLBİLİMİNE KATKISI
From Revelation to Data: The Contribution of Corpus.Quran.com to Arabic Corpus Linguistics

Mustafa KAYAPINAR

Dr. Öğr. Üyesi, Sivas Cumhuriyet Üniversitesi İlahiyat Fakültesi, Temel İslam Bilimleri Bölümü, Arap Dili ve Belagatı Anabilim Dalı, mustafakayapinar@yahoo.com, ORCID: [0000-0001-9547-277X](https://orcid.org/0000-0001-9547-277X).

Makale Bilgisi / Article Information

Makale Türü / Article Types:	Araştırma Makalesi/Research Article
Geliş Tarihi / Received:	02.09.2025
Kabul Tarihi / Accepted:	07.10.2025
Yayın Tarihi / Published:	31.12.2025
Cilt / Volume:	12
Sayı / Issue:	2
Sayfa/ Pages:	93-120

Atıf / Cite as: Kayapınar, Mustafa. "Vahiyden Veriye: *Corpus.Quran.com*'un Arapça Derlem Dilbilimine Katkısı" (From Revelation to Data: The Contribution of Corpus.Quran.com to Arabic Corpus Linguistics). *Pamukkale Üniversitesi İlahiyat Fakültesi Dergisi-Pamukkale University Journal of Divinity Faculty* 12/2 (2025), 93-120. Doi: 10.17859/pauifd.1776831.

İntihal / Plagiarism: Bu makale, Turnitin intihal tarama programı ile taranmıştır. Ayrıca iki hakem tarafından da incelenmiştir. / This article has been scanned with Ithenticate plagiarism screening program. Also this article has been reviewed by two referees.

Çıkar Çatışması / Conflict of Interest: Yazar çıkar çatışması olmadığını beyan etmiştir. The Author declared that there is no conflict of interest

Finansal Destek / Grant Support: Yazar bu çalışma için finansal destek almadığını beyan etmiştir. / The authors declared that this study has received no financial support.

www.dergipark.gov.tr/pauifd

VAHİYDEN VERİYE: CORPUS.QURAN.COM'UN ARAPÇA DERLEM DİLBİLİMİNE KATKISI***Mustafa KAYAPINAR******Öz**

Bu makale, dijital dilbilim ve Arapça derlem çalışmaları alanında son yıllarda dikkat çeken *Corpus.Quran.com* platformunun, özellikle Arapça derlem dilbilimi bağlamında sunduğu katkıları bütüncül biçimde ele almaktadır. Kur'an-ı Kerim merkezli olarak inşa edilen bu dijital platform, sadece kutsal metnin dilsel yönlerini değil; aynı zamanda klasik Arapçanın morfolojik, sentaktik ve semantik katmanlarını da bilgisayarlı dil işleme (NLP) araçlarıyla analiz edilebilir hâle getirmektedir. Çalışma, bu platformun teknik yapısını, sunduğu analiz olanaklarını, kullanım örüntülerini ve bilimsel potansiyelini ayrıntılı olarak değerlendirmektedir. Araştırma kapsamında ele alınan başlıca bileşenler arasında Buckwalter morfolojik analiz sistemi, lemmatizasyon ve kökleme işlemleri, sözdizimsel çözümleme için geliştirilen Treebank yapısı, sözcük sıklığı tabloları, kavramsal ağ gösterimleri ve anlam sınıflandırma sistemleri yer almaktadır. Bu bileşenler sayesinde, Kur'an ayetlerinde geçen her kelime çok katmanlı bir dilsel işlemden geçirilebilmekte; böylece hem yapısal hem de anlamsal incelemelere olanak sağlanmaktadır. Çalışmada, bu işlevlerin dil öğretimi, anlambilimsel çözümleme, karşılaştırmalı sözdizimi ve yapay zekâ destekli doğal dil işleme sistemlerine nasıl entegre edilebileceği de örneklerle gösterilmiştir. Ayrıca makalede, *Corpus.Quran.com* hakkında daha önce yapılmış olan akademik yayınlar da detaylı şekilde analiz edilmiştir. Bu çerçevede, Eric Atwell, Kais Dukes, Nizar Habash ve Mahmoud Ghoneim gibi isimlerin çalışmaları değerlendirilmiş; sitenin ortaya çıkış süreci, gelişimi ve akademik çevrelerce nasıl ele alındığına dair eleştirel bir literatür özeti sunulmuştur. Böylelikle çalışmanın hem mevcut literatürdeki boşluğu doldurduğu hem de gelecekteki benzer platformlar için model teşkil edebileceği ortaya konmuştur. Bununla birlikte, platformun çeşitli sınırlılıkları da analiz edilmiştir. Temsiliyet sorunları, klasik metinle sınırlı olması, modern Arapçayla olan uyumsuzluğu, anlam belirsizlikleri, etiketleme sorunları, teknik yetersizlikler ve kullanıcı arayüzündeki eksiklikler gibi yönler çalışmanın eleştirel boyutunu oluşturmaktadır. Bu eksikliklere rağmen platformun açık erişimli olması, çok dilli kullanıcı desteği ve akademik araştırmalarla entegre edilebilir yapısı gibi güçlü yönleri sayesinde Arap dili çalışmaları için benzersiz bir kaynak sunduğu vurgulanmaktadır.

Sonuç olarak bu çalışma, Türkçe literatürde *Corpus.Quran.com*'u derlem dilbilimi, NLP ve öğretim teknolojileri açısından çok boyutlu değerlendiren ilk sistematik inceleme olarak öne çıkmakta; Arapça dilbilimi alanında disiplinlerarası bir katkı sunmaktadır.

Anahtar Kelimeler: Derlem Dilbilim, Arapça, Kur'an, Doğal Dil İşleme, Ağaçbank.

From Revelation to Data: The Contribution of Corpus.Quran.com to Arabic Corpus Linguistics**Abstract**

This article offers a comprehensive evaluation of *Corpus.Quran.com*, a digital platform that has recently drawn attention in the fields of digital linguistics and Arabic corpus

* Yazar makalede *Etik Kurul İzni* gerektirecek bir durum bulunmadığını beyan etmiştir.
** Dr. Öğr. Üyesi, Sivas Cumhuriyet Üniversitesi İlahiyat Fakültesi, Temel İslam Bilimleri Bölümü, Arap Dili ve Belagatı Anabilim Dalı, mustafakayapinar@yahoo.com, ORCID: [0000-0001-9547-277X](https://orcid.org/0000-0001-9547-277X).

studies, particularly for its contributions to Arabic corpus linguistics. Constructed around the Qur'anic text, this platform enables not only linguistic analyses of the sacred scripture but also multilayered morphological, syntactic, and semantic exploration of Classical Arabic through computational linguistics tools. The study critically assesses the platform's technical infrastructure, analytical capabilities, usage patterns, and scientific potential. The core components examined include the Buckwalter morphological analysis system, lemmatization and stemming tools, syntactic parsing via the Treebank model, word frequency tables, conceptual network visualizations, and semantic classification systems. These elements allow each word in the Qur'anic verses to undergo multilayered linguistic processing, enabling both structural and semantic analysis. The article also demonstrates how these functionalities can be applied in language teaching, semantic interpretation, comparative syntax, and artificial intelligence-driven natural language processing (NLP) systems. In addition, the article presents a detailed analysis of previous academic works on *Corpus.Quran.com*. It evaluates studies by scholars such as Eric Atwell, Kais Dukes, Nizar Habash, and Mahmoud Ghoneim, offering a critical literature review on the platform's development and its reception in academic circles. This positions the study as one that fills a significant gap in the literature and proposes a model for future digital linguistic platforms. The article also addresses several limitations of the platform. These include representational constraints, its restriction to a single classical text, incompatibility with Modern Standard Arabic, semantic ambiguities, tagging inconsistencies, technical shortcomings, and user interface limitations. Despite these challenges, the platform's open-access nature, multilingual user support, and potential for integration into academic research make it a unique and valuable resource for Arabic linguistic studies. In conclusion, this study stands out as the first systematic evaluation of *Corpus.Quran.com* in Turkish literature from the perspectives of corpus linguistics, NLP, and educational technologies, and offers a multidisciplinary contribution to the field of Arabic linguistics.

Keywords: Corpus Linguistics, Arabic, Qur'an, Natural Language Processing, Treebank.

Structured Abstract

This study examines *Corpus.Quran.com*, a Qur'an-based digital corpus platform, to explore its contributions to Arabic corpus linguistics. Classical religious texts like the Qur'an have been underutilized in data-driven linguistic research. *Corpus.Quran.com* enables multilayered morphological, syntactic, and semantic analysis of the Qur'anic text. The purpose of this research is to evaluate the platform's technical structure, its analytical modules, and its overall value for Arabic linguistic studies. The article is organized around several key questions:

- What data infrastructure does *Corpus.Quran.com* provide for analyzing Arabic-specific morphological and syntactic structures via the Qur'anic text?
- How do the platform's core modules (morphological analyzer, lemmatizer, dependency treebank, concept mapping, etc.) facilitate multi-level analysis (morphology, syntax, semantics) of Classical Arabic?
- In which ways can this digital Qur'anic corpus be utilized in Arabic language teaching and NLP applications, and what are its limitations in those applications?
- Does the Qur'an's fixed, authoritative text offer advantages or pose constraints for corpus-based analysis of Arabic?

A qualitative descriptive-analytical approach was adopted. The study included detailed document analysis of the platform's features. It systematically examined each major component – morphological tagging, syntactic dependency treebank, lemmatization/root indexing, word frequency lists, and conceptual network visualization – via example queries. The platform's annotation approach was evaluated in light of corpus linguistics and computational linguistics principles and compared with similar Arabic corpora. Additionally, a review of prior studies on *Corpus.Quran.com* contextualized its development and reception.

Corpus.Quran.com enables every word of the Qur'anic text to be analyzed in multiple linguistic layers. The Buckwalter-based morphological module provides detailed part-of-speech tagging and grammatical features for each word, aligned with Classical Arabic's root-and-pattern morphology. The dependency treebank visualizes the syntactic structure of verses, delineating grammatical roles. The lemmatization tools link words to their roots and aggregate occurrences to reveal word frequency and derivational patterns. Semantic features such as conceptual maps and thematic classifications group related terms and verses, supporting semantic analysis. *Corpus.Quran.com* thus serves as a comprehensive linguistic resource for Classical Arabic, with applications in language education and as an annotated dataset for NLP research.

Several limitations of the platform were identified. The resource is confined to a single classical text, providing a stable dataset but limiting representativeness of broader Arabic (especially Modern Standard Arabic and dialects). The fixed corpus nature yields high consistency in annotation but poses constraints when generalizing findings beyond the Qur'anic context. Incompatibility with contemporary language forms and challenges in semantic tagging (due to the Qur'an's polysemous expressions) were noted, leading to annotation inconsistencies. Technical drawbacks (an older interface and reliance on transliteration) affect usability. Despite these issues, the platform's open-access availability, multilingual support, and transparent annotation process remain strengths. Researchers and educators can leverage the Qur'anic corpus for its quality and depth, while mindful of its limitations.

This study provides the first comprehensive evaluation of *Corpus.Quran.com* from a corpus linguistics and NLP perspective, filling a gap in the literature. It offers a model for transforming a classical text into a richly annotated corpus for interdisciplinary research. Scientifically, it bridges Islamic studies, Arabic linguistics, and computational linguistics, demonstrating the value of corpus-based analysis for classical texts. Practically, the platform proves to be a valuable tool for Arabic language education and a useful reference for NLP applications. Overall, the platform's innovations underscore its unique role in Arabic corpus linguistics and can inform future projects extending corpus analysis to wider Arabic contexts.

GİRİŞ

1.1. Konunun Sınırları ve Gündeme Gelişi

Arap dili, tarih boyunca dinî, kültürel ve bilimsel metinlerin taşıyıcısı olarak işlev görmüş ve özellikle Kur'ân-ı Kerîm aracılığıyla klasik bir yazı ve söylem standardı kazanmıştır. Ne var ki modern dilbilimsel analizlerin çoğu, bu klasik yapıdaki metinleri istatistiksel, bilişsel ya da bilgisayarlı yöntemlerle ele almakta geç kalmıştır. Özellikle Arapçanın kök-temelli morfolojik yapısı, harekeleme farklılıkları, sözdizimsel esneklikleri ve çok anlamlılık gibi doğal özellikleri, bu dilde dijital analiz yapılmasını Batı dillerine kıyasla daha karmaşık hâle getirmiştir.¹

Son yıllarda dilbilimsel çalışmaların sezgisel yaklaşımlardan veri tabanlı sistematik yöntemlere yönelmesi, derlem dilbilimi (corpus linguistics) gibi alanlara olan ilgiyi artırmıştır. Derlem dilbilimi, doğal dilin gerçek kullanım örneklerine dayalı olarak incelenmesini hedeflerken, Arapça gibi klasik ve kutsal metin geleneğine sahip diller için ayrı bir önem taşımaktadır.² Bu bağlamda Kur'ân-ı Kerîm'in sabit yapısı,

¹ Habash, Nizar. *Introduction to Arabic Natural Language Processing*. San Rafael: Morgan & Claypool, 2010, 5–6.

² Biber, Douglas vd. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press, 1998, 4–7.

değişmeyen bağlamı ve yüksek dilsel yoğunluğu, onu hem anlam hem yapı düzeyinde analiz için örnek bir kaynak hâline getirmiştir.

Arapça derlem çalışmalarının sınırlı oluşu ve mevcut projelerin çoğunun ya lehçelere ya da modern metinlere odaklanması, klasik Arapçanın nicel ve biçimsel incelenmesini kısıtlamıştır. İşte bu noktada, *Corpus.Quran.com* adlı dijital platform hem dilbilimsel hem teknolojik açıdan önemli bir boşluğu doldurmak üzere gündeme gelmiştir. Söz konusu site, Kur'ân-ı Kerîm'in kelime kelime morfolojik ve sözdizimsel etiketlenmesini içeren açık kaynaklı bir derlem sunmakta; akademik araştırmalardan dil öğretimine, doğal dil işleme uygulamalarından anlambilimsel çözümlemelere kadar birçok alanda kullanılabilir.

Bu çalışmada, *Corpus.Quran.com* platformunun Arapça derlem dilbilimi açısından taşıdığı potansiyel ayrıntılı biçimde ele alınacak, sitenin sunduğu bileşenlerin kuramsal ve uygulamalı yönleriyle Arap dili çalışmalarına nasıl katkı sağladığı tartışılacaktır. Ancak bu inceleme yalnızca teknik bir betimleme olmayacak; aynı zamanda derlem dilbiliminin Arapçaya uygulanabilirliği, sınırları ve yeni açılımları da değerlendirilecektir.

1.2. Çalışmanın Amacı ve Önemi

Bu çalışmanın temel amacı, *Corpus.Quran.com* adlı dijital Kur'an derlemini Arapça derlem dilbilimi bağlamında incelemek ve bu platformun morfolojik, sözdizimsel ve anlamsal analizlerde sunduğu olanakları bilimsel açıdan değerlendirmektir. Derlem dilbiliminin Arap dilinde sistematik olarak uygulanabilirliğini test etmenin yanı sıra, bu platformun hangi açılardan güçlü, hangi yönlerden sınırlı kaldığını ortaya koymak da araştırmanın temel hedeflerindendir.

Bu bağlamda çalışma; yalnızca bir web platformunun tanıtımını yapmakla kalmayacak, aynı zamanda bu dijital ortamın Arapçanın biçimsel çözümlemesine katkı sağlayıp sağlamadığı, yapay zekâ temelli uygulamalara uygun olup olmadığı, öğretim süreçlerinde nasıl işlev görebileceği gibi sorulara da yanıt arayacaktır. Özellikle Kur'an gibi sabit ve tarihsel açıdan otorite sahibi bir metnin dilsel olarak parçalanabilir, etiketlenebilir ve sınıflandırılabilir olması, derlem tabanlı yaklaşımlar açısından özel bir değer taşımaktadır.

Ayrıca bu çalışmada *Corpus.Quran.com*'un, yalnızca bir dil verisi kaynağı değil; aynı zamanda Arap dili öğretimi, anlam çalışmaları, çeviri teknolojileri ve semantik ağlar gibi alanlarda da kullanılabilecek çok boyutlu bir altyapı sunduğu gösterilecektir. Kur'an'ın sabit metin yapısı sayesinde veri tutarlılığı yüksek bir derlem ortamı oluşmuş ve bu da morfolojik varyasyonların, sözdizimsel yapıların ve anlam ilişkilerinin daha sağlıklı çözümlemesine imkân vermiştir.³

Bu yönüyle *Corpus.Quran.com*, sadece Kur'an araştırmalarında değil; genel olarak Arap dilbilimi, uygulamalı dilbilim, bilgisayarlı dil çalışmaları ve doğal dil işleme (NLP) gibi alanlarda da önemli bir katkı potansiyeli taşımaktadır.

³ Dukes, Kais; Atwell, Eric. "The Quranic Arabic Corpus: Current Work and Future Directions." *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, 2012, 2-4.

Araştırmanın nihai amacı, bu katkıların bilimsel temelde ortaya konması ve bu platformun hem teorik hem de uygulamalı düzeyde Arapça dil incelemelerine nasıl entegre edilebileceğinin somut biçimde gösterilmesidir.

1.3. Araştırma Soruları

Bu çalışma, Arapça derlem dilbilimi ve dijital metin analizleri bağlamında *Corpus.Quran.com* adlı platformun işlevselliğini, olanaklarını ve sınırlılıklarını ortaya koymayı hedeflemektedir. Bu çerçevede, araştırmanın temelini oluşturan sorular şu şekilde yapılandırılmıştır:

1. *Corpus.Quran.com* platformu, Kur'an metni üzerinden Arapçaya özgü morfolojik ve sözdizimsel yapıları anlamak için nasıl bir veri altyapısı sunmaktadır?
2. Platformda yer alan bileşenler (kelime çözümleme, morfolojik etiketleme, sözdizimsel ağaçlar, kavram haritası vb.), Arapça dilbiliminin farklı düzeylerdeki (biçimbilim, sözdizim, anlambilim) çözümlemelerine nasıl katkı sağlamaktadır?
3. Bu dijital derlem, Arapçanın öğretimi, doğal dil işleme (NLP) uygulamaları ve makine öğrenimi gibi alanlarda nasıl kullanılabilir ve hangi alanlarda sınırlı kalmaktadır?
4. Kur'an'ın sabit metinsel yapısı, derlem dilbilimi açısından avantaj mı sunmaktadır, yoksa temsiliyet bağlamında bir sınırlılık mı doğurmaktadır?
5. *Corpus.Quran.com*, klasik Arapça üzerine yapılmış derlem temelli çalışmalarda ne tür bir boşluğu doldurmakta ve bu alanda hangi bilimsel katkıyı sağlamaktadır?

Bu sorular doğrultusunda yürütülecek analizler sayesinde hem platformun teknik ve akademik değerinin bütüncül bir şekilde değerlendirilmesi hem de Arapça derlem dilbilimi sahasında yeni araştırma açımlarının ortaya konması hedeflenmektedir.⁴

1.4. Yöntem

Bu çalışma, *Corpus.Quran.com* platformunu Arapça derlem dilbilimi bağlamında inceleyen betimleyici ve çözümleyici bir yöntem çerçevesinde kurgulanmıştır. Araştırma, nitel araştırma yöntemleri temelinde yapılandırılmış olup, birincil olarak doküman analizi tekniği kullanılmıştır. İncelemenin temelini oluşturan *Corpus.Quran.com* sitesindeki morfolojik etiketleme, sözdizimsel ağaç yapıları, kavram haritaları ve lemmatizasyon sistemleri gibi bileşenler, örnek ekran görüntüleri ve kullanım senaryoları üzerinden değerlendirilmiştir.

Veri kaynağı olarak kullanılan *Corpus.Quran.com*, sabit metne dayalı ve çok katmanlı anotasyon içeren dijital bir derlem olduğu için, doğal dil işleme (NLP) ve bilgisayarlı dilbilim (CL) ilkeleri çerçevesinde analiz edilmiştir. Platformun sunduğu işlevler, dilbilimsel alt alanlara göre (morfoloji, sözdizim, anlambilim) sınıflandırılmış

⁴ McEnery, Tony; Hardie, Andrew. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press, 2012, 15–18.

ve her modülün somut kullanımı bağlamında çözümlenmiştir. Ayrıca, platformun teknik arka planına ilişkin olarak Sharaf ve Atwell'in yayınları ile birlikte, benzer Arapça derlem projeleriyle (ArabiCorpus, Tashkeela, KACST) karşılaştırmalı değerlendirmelere yer verilmiştir.

Araştırmada, *Corpus.Quran.com*'un sunduğu verilerin yalnızca teknik düzeyde değil, aynı zamanda dil öğretimi, anlamsal analiz, makine çevirisi, morfolojik çözümleme ve yapay zekâ temelli uygulamalar için nasıl kullanılabileceği de ayrıntılı olarak incelenmiştir. Böylece platformun yalnızca bir inceleme nesnesi değil, aynı zamanda çok yönlü bir araç ve yöntem kaynağı olarak nasıl işlev gördüğü de ortaya konmuştur.

1.5. *Corpus.Quran.com* Platformunun Yapısı ve Kullanım Özellikleri

Corpus.Quran.com, Kur'ân-ı Kerîm'in sözcük temelli, çok katmanlı ve etkileşimli biçimde analiz edilebilmesine olanak sağlayan bir dijital platformdur. 2007 yılında Khaled Sharaf'ın Lancaster Üniversitesi'nde yürüttüğü yüksek lisans ve doktora araştırması çerçevesinde başlatılan proje, Eric Atwell'in yönlendirmesiyle hem bir akademik yayın altyapısı hem de doğal dil işleme araştırmaları için referans veri kümesi olarak gelişmiştir.

Bu platformun temel amacı, Arapçanın özellikle klasik formunun bilgisayarlı dilbilim yöntemleriyle çözümlenmesini kolaylaştırmaktır. Kur'an gibi sabit, yüksek düzeyde yapılandırılmış bir metni temel alarak geliştirilen bu derlem; Arapçada morfoloji, sözdizim, anlambilim, lemmatizasyon ve semantik ağ gibi alanlarda çok katmanlı analiz yapılmasına olanak sağlar. Ayrıca, *Corpus.Quran.com* projesi, etiketleme ve çözümleme işlemlerini şeffaf biçimde açık erişimli olarak sunması bakımından da dikkat çekicidir.⁵

1.5.1. Platformun Teknik Temeli ve Yapısı

Platform, Buckwalter Morfolojik Çözümleyicisi temel alınarak geliştirilmiş olup her Kur'an kelimesi için POS (part-of-speech) bilgisi, kök, kalıp, sözcük türü, geçişlilik, cinsiyet, sayı gibi dilbilgisel özellikler ayrıntılı biçimde gösterilmektedir. Bu sistem, Arapçanın kök temelli morfolojik doğasına uygun olarak geliştirilmiştir.⁶

Ayrıca, Kais Dukes'un katkılarıyla oluşturulan *Classical Arabic Treebank* veri yapısı sayesinde, her ayet sözdizimsel düzeyde ağaç yapısı (parse tree) ile görselleştirilmiştir. Bu yapı, cümle içindeki sözcüklerin birbirine bağlanma ilişkilerini ve gramatikal görevlerini (özne, mef'ul, hal, sıfat vb.) net şekilde ortaya koymaktadır.⁷

Veri yapısında yer alan her kelime, aynı zamanda Kur'an metni içinde geçtiği tüm diğer kullanımlarla eşleştirilebilir durumdadır. Bu yönüyle *Corpus.Quran.com*,

⁵ Buckwalter, Tim. "Buckwalter Arabic Morphological Analyzer." *Linguistic Data Consortium*, University of Pennsylvania, 2002.

⁶ Buckwalter, Tim. "Buckwalter Arabic Morphological Analyzer." *Linguistic Data Consortium*, University of Pennsylvania, 2002.

⁷ Dukes, Kais. "Statistical Parsing by Machine Learning from a Classical Arabic Treebank." MSc Thesis, University of Leeds, 2010, 15–16.

Arapçadaki türev yapıları ve kalıp varyasyonlarını analiz etmek için kullanılabilecek yüksek hassasiyetli bir dijital derlem sunmaktadır.

1.5.2. Ana Modüller ve İşlevleri

Aşağıda sitedeki başlıca modüller ve işlevleri ayrıntılı şekilde verilmiştir:

- **Word by Word:** Her ayet, kelime kelime çözümlemeye tabi tutulur. Kullanıcı her kelimeye tıkladığında, kelimenin kökü, biçimi (template), sözcük türü, İngilizce anlamı ve sözlük karşılığı görüntülenir. Bu özellik, özellikle dil öğrenenler ve öğretmenler için büyük kolaylık sağlar.

- **Morphological Analysis:** Buckwalter analiz sistemine dayalı biçimde, her kelime detaylı morfolojik parçalara ayrılır. Fiiller için zaman (mâzî, muzârî), geçişlilik (lâzım-müteaddî), kişi, sayı ve cinsiyet bilgileri; isimler için çoğulluk, belirlilik, i'râb durumu gibi bilgiler etiketlenmiştir.⁸

- **Lemmas:** Aynı kökten (lemma) türetilmiş kelimeler, tek bir sayfada listelenir. Böylece örneğin “ك-ت-ب” kökü ile ilişkili tüm kelimeler (كُتِبَ, كَتَبَ, كَاتِبٌ vb.) görüntülenebilir. Bu, Arapça türev sistemine dayalı çalışmalar için büyük avantaj sağlar.

- **Treebank:** Ayetlerin tam cümle yapıları parse tree (sözdizimsel ağaç) formunda görselleştirilmiştir. Böylece her bir kelimenin diğerleriyle ilişkisi, yapısal bağlantı türü (özne-yüklem-nesne) ve görevsel niteliği doğrudan takip edilebilir.⁹

- **Concept Map:** Kur'an'da geçen temel ve soyut kavramlar (örneğin iman, rahmet, azap, melek, cennet) arasında anlamsal bağlar kurar. Bu kavramlar arasındaki üst-alt kavram ilişkileri ve bu kavramların geçtiği ayetler kullanıcıya kavramsal bir harita olarak sunulur.¹⁰ Bu özellik, anlambilim ve ontoloji tabanlı analizlerde büyük kolaylık sağlar.

- **Search:** Her kelime, kök ya da POS etiketi temelinde aranabilir. Örneğin sadece “merfû isimler” veya “geçişli fiiller” gibi yapısal filtreleme yapılabilir. Ayrıca arama sonuçları doğrudan ayetlere bağlanır.

- **Translation Comparison:** Kur'an'ın farklı İngilizce çevirileri, aynı ayet için yan yana karşılaştırmalı biçimde sunulur. Bu özellik bilhassa anlam aktarımı, çeviri farklılıkları ve bağlam analizleri için değerlidir.

1.5.3. Kullanım Senaryoları ve Akademik Değeri

Corpus.Quran.com, hem bireysel kullanıcılar (öğrenciler, öğretmenler, ezber yapanlar) hem de akademik araştırmacılar için çok yönlü işlevsellik taşır. Morfolojik düzeyde detaylı veri sunması sayesinde özellikle öğretmenler için kavramsal öğretim tasarımlarında, öğrenciler için yapı temelli çözümlemelerde etkin bir kaynaktır.

⁸ Buckwalter, Tim. “Buckwalter Arabic Morphological Analyzer.” *Linguistic Data Consortium*, University of Pennsylvania, 2002.

⁹ Dukes, Kais. “Statistical Parsing by Machine Learning from a Classical Arabic Treebank.” MSc Thesis, University of Leeds, 2010, 15–16.

¹⁰ Dukes, Kais; Atwell, Eric. “Conceptual Analysis of the Quran Using the Quranic Arabic Corpus.” *LRE Journal*, 2012, 3–5.

Ayrıca bu site, makine öğrenmesi algoritmaları için etiketli Arapça veri sağlaması yönüyle NLP alanında da akademik atıflar almış bir projedir.

Bu yönüyle *Corpus.Quran.com*, klasik Arapçanın sistemli, katmanlı ve açık veri mantığıyla işlendiği ilk bütüncül projelerden biri olması bakımından, derlem dilbilimi açısından özgün bir örnek olarak değerlendirilmektedir.

2. Kuramsal Arka Plan

2.1. Derlem Dilbiliminin Tarihçesi ve Tanımı

Derlem dilbilimi (İng. *corpus linguistics*), dili sistematik bir şekilde incelemek için doğal dil örneklerinin dijital olarak toplanması, sınıflandırılması ve analiz edilmesini temel alan bir araştırma yöntemidir. Bu yaklaşım, dilbilgisel yapıların, sözcük frekanslarının ve bağlamsal kullanımların sezgisel değil, nicel ve nesnel verilere dayalı olarak incelenmesini sağlar.¹¹

Derlem tabanlı incelemelerin modern anlamda ortaya çıkışı, bilgisayar destekli teknolojilerin gelişimiyle 1960'lı yıllara dayanır. Bu dönemde Randolph Quirk öncülüğünde başlatılan *Survey of English Usage* projesi ile ilk büyük dijital İngilizce derlem oluşturulmuş; bunu 1980'lerde COBUILD (Collins Birmingham University International Language Database) projesi takip etmiştir.¹² Bu gelişmelerle birlikte, özellikle Geoffrey Leech, John Sinclair ve Susan Hunston gibi öncü dilbilimcilerin katkısıyla derlem dilbilimi, bağımsız bir metodolojik alan hâline gelmiştir.

1990'lardan itibaren Douglas Biber, Randi Reppen, Tony McEnery ve Andrew Hardie gibi araştırmacıların katkılarıyla derlem temelli incelemeler daha sistematik hâle gelmiş ve eğitim, çeviri, sözlükbilim, söz dizimi, semantik ve söylem çözümleme gibi birçok alanda kullanılmaya başlanmıştır. Derlem dilbilimi bugün artık sadece bir yöntem değil; aynı zamanda dilsel olgulara ilişkin epistemolojik bir çerçeve sunan teorik bir alan olarak da değerlendirilmektedir.¹³

Tanım olarak derlem dilbilimi, belirli bir dil ya da lehçeye ait sözel verilerin sistemli biçimde toplanarak, bilgisayar ortamında analiz edilmesini amaçlayan, veri odaklı bir yaklaşımdır. Derlem (İng. *corpus*), bu çerçevede bir dilin ya da söylem türünün farklı bağlamlardaki örneklerinden oluşan temsili bir koleksiyon olarak tanımlanır.¹⁴ Bu koleksiyon, istatistiksel analizlerin yanı sıra nitel çözümlemelere de imkân verir.

Özellikle sözlük yazımı, dil öğretimi ve çeviri teknolojileri gibi uygulama alanlarında derlem dilbilimi, dilin doğal kullanımlarını gözleme ve kuralların gerçek hayattaki yansımalarını ortaya koyma açısından büyük katkılar sağlamaktadır. Bugün birçok dilbilimsel kuram (ör. işlevselciler, kullanım temelli kuramcılar) derlem

¹¹ McEnery, Tony; Hardie, Andrew. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press, 2012, 1–3.

¹² Sinclair, John. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press, 1991, vii–x.

¹³ Biber, Douglas; Reppen, Randi. *Corpus Linguistics*. Cambridge: Cambridge University Press, 2005, 4–5.

¹⁴ Tognini-Bonelli, Elena. *Corpus Linguistics at Work*. Amsterdam: John Benjamins Publishing, 2001, 2.

verilerini, dilin yapısal ve anlamsal eğilimlerini açıklamak için başlıca kaynak olarak kabul etmektedir.¹⁵

2.2. Arapça Derlem Dilbiliminin Sorunları ve Gelişimi

Arapçanın derlem dilbilimi bağlamında incelenmesi, doğal dil işleme (NLP) alanında önemli bir akademik ve teknolojik mesele olarak öne çıkmaktadır. Ancak Arapçanın kendine özgü morfolojik ve sentaktik yapısı, bu dilde derlem temelli analizlerin uygulanmasını Batı dillerine kıyasla oldukça karmaşık hâle getirmiştir. Bu nedenle Arapça için oluşturulan derlemelerin gelişimi, İngilizce gibi dillerle kıyaslandığında hem daha geç başlamış hem de sınırlı kalmıştır.¹⁶

Arapçada derlem temelli çalışmaları zorlaştıran başlıca yapısal sorunlar şunlardır:

2.2.1. Kök Temelli Morfoloji: Arapçanın büyük bölümü üç harfli köklere dayalı türevsel bir sistem üzerine kuruludur. Bu kök-temelli yapı, kelimelerin yüzey formlarının altında çok sayıda anlam ve yapı barındırmasına yol açar. Örneğin, “ك-ت-ب” kökü üzerinden onlarca farklı fiil, isim ve sıfat türetilmektedir. Bu, biçimsel çözümlemede ayrıştırma ve lemmatizasyon işlemlerini zorlaştırır.¹⁷

2.2.2. Harekeleme (Vokalizasyon) Sorunu: Arapçanın yazı sisteminde kısa ünlüler (harekeler) çoğu zaman yazılmaz. Bu durum, özellikle otomatik morfolojik analizlerde belirsizlik (ambiguity) oranını ciddi şekilde artırır. Aynı yazımla gösterilen iki kelime, farklı köklerden ve farklı anlamlardan gelebilir. Bu çokanlamlılık, bilgisayarlı analiz araçlarının doğruluğunu azaltmaktadır.

2.2.3. Çok Anlamlılık ve Sözdizimsel Esneklik: Arapçada kelime sırasının oldukça esnek oluşu, aynı cümle yapısının farklı biçimlerde kurulmasına imkân tanır. Bu da POS etiketlemesi ve sözdizimsel çözümlemelerde kural temelli sistemlerin yetersiz kalmasına yol açar. Bu nedenle istatistiksel ve öğrenmeye dayalı modeller (machine learning) daha uygun çözüm yolları olarak öne çıkmaktadır.¹⁸

Bu yapısal zorluklara rağmen son 20 yılda Arapça derlem çalışmaları da çeşitlenmeye başlamıştır. Bu alandaki önemli derlem projeleri şunlardır:

Tashkeela: 75 milyonun üzerinde harekeli Arapça kelimedenden oluşan açık kaynaklı bir derlemidir. Özellikle sesli (harekeli) yazım gerektiren NLP modelleri için kullanılır.¹⁹

¹⁵ Hunston, Susan. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press, 2002, 8–10.

¹⁶ Habash, Nizar. *Introduction to Arabic Natural Language Processing*. San Rafael: Morgan & Claypool, 2010, 3–6.

¹⁷ Farghaly, Ali; Shaalan, Khaled. “Arabic Natural Language Processing: Challenges and Solutions.” *ACM Transactions on Asian Language Information Processing*, 2009, 8(4), 2–5.

¹⁸ Diab, Mona. “Towards an Optimal POS Tag Set for Modern Standard Arabic Processing.” *Proceedings of the Arabic NLP Conference*, 2007, 2–4.

¹⁹ Zerrouki, Taha. “Tashkeela: A Vocalized Arabic Corpus.” Available online: <https://tashkeela.github.io>, 2013.

ArabiCorpus: Edebi, haber ve dinî metinlerden oluşan çok amaçlı bir derlemidir. POS etiketli değildir, fakat metin çeşitliliği açısından değerlidir.²⁰

KACST Arabic Corpus: Suudi Arabistan merkezli, modern Arapçayı temel alan geniş kapsamlı bir projedir. Morfolojik ve sözdizimsel etiketleme içerir, ancak akademik erişim sınırlıdır.

Bu gelişmelerin yanı sıra, klasik Arapçanın derlem temelli çözümlemelerinde en dikkat çekici girişimlerden biri, *Corpus.Quran.com* olmuştur. Bu platform, yukarıda anılan sorunların bir kısmını çözmek için hem kurallı hem istatistiksel yöntemleri birleştiren hibrit bir yapı sunmuş; özellikle Kur'an gibi sabit bir metin üzerinden çalıştığı için morfolojik belirsizliği azaltmıştır²¹.

Arapça derlem dilbiliminin bugünkü durumu, yapısal sorunlara rağmen gelişim göstermekte olup, daha fazla dilsel çeşitlilik, çapraz lehçe çözümlemeleri ve klasik-modern karşılaştırmalar ile yeni yönelimlere ihtiyaç duymaktadır.

2.3. Kur'an Merkezli Derlem Oluşturmanın Özellikleri

Kur'an-ı Kerîm, yalnızca dinî bir metin değil; aynı zamanda Arap dilinin klasik yapısını muhafaza eden, yüksek düzeyde retorik, yapısal ve anlam yoğunluğu içeren bir dilsel kaynaktır. Bu nedenle, Kur'an'ın dijital derlem olarak yapılandırılması hem dilbilimsel hem de teknik açıdan özel bazı avantajlar ve sınırlılıklar taşımaktadır.

2.3.1. Sabit ve Değişmez Metin

Kur'an metni tarihsel olarak sabit bir yapıya sahiptir; tüm varyantları kayda geçirilmiş ve standart hâle getirilmiştir. Bu durum, derlem dilbiliminde sıklıkla karşılaşılan metin seçimi, metin temizliği ve versiyon farklılıkları gibi sorunları ortadan kaldırır.²¹ Dolayısıyla Kur'an, derlem olarak çalışılabilecek en istikrarlı metin türlerinden biridir. Metin bütünlüğünün sabitliği, morfolojik ve sözdizimsel etiketleme işlemlerinde yüksek tutarlılık sağlar.

2.3.2. Yüksek Biçimsel ve Anlamsal Yoğunluk

Kur'an'da kullanılan dil, kısa cümlelerde çok katmanlı anlamlar taşıyabilen, çoğu zaman belagat sanatlarına dayalı, sembolik ve metaforik ifadeler içeren özel bir dildir. Bu özellik, semantik analiz açısından zengin bir malzeme sunmakla birlikte, otomatik anlam çıkarımı ve bağlam çözümleme süreçlerini karmaşıktırabilir.²² Yani anlam yoğunluğu yüksek fakat çok anlamlılığa açık bir metinle çalışmak hem fırsat hem zorluk teşkil eder.

2.3.3. Klasik Arapçanın Temsiliyeti

Kur'an, klasik Arapçanın en eski ve en istikrarlı örneğidir. Bu yönüyle dilin söz

²⁰ Parkinson, Dilworth. "Building ArabiCorpus: An Online Arabic Corpus for Research." *Brigham Young University Linguistics Journal*, 2008, 14(1), 23–25.

²¹ McEnery, Tony; Hardie, Andrew. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press, 2012, 33–34.

²² Boullata, Issa J. "The Rhetorical Interpretation of the Qur'an: I'jāz and Related Topics." *Approaches to the History of the Interpretation of the Qur'an*, ed. Andrew Rippin. Oxford: Clarendon Press, 1988, 143–146.

dizimi, morfolojisi ve kelime hazinesi hakkında araştırmalar yapılmasına imkân tanır. Ancak buradaki temsil, modern Arapçanın çeşitliliğini kapsamaz. Dolayısıyla bu derlemde elde edilen veriler, yalnızca klasik düzlemde geçerli olabilir.²³

2.3.4. Anotasyon İçin Uygunluk

Kur'an metni, ayet-sure sistematığına sahip olması, her kelimenin ayrı ayrı işlenebilir yapıda olması ve anlam bloklarının belirginliği nedeniyle derlem anotasyonu için son derece uygundur. Bu yapı hem kelime çözümleme hem sözdizimsel ağaç oluşturma hem de anlamsal haritalama gibi işlemleri kolaylaştırır. *Corpus.Quran.com* bu yapıyı başarılı şekilde kullanarak katmanlı bir derlem altyapısı oluşturmuştur.

2.3.5. İdeolojik ve İnanç Temelli Yaklaşım Sorunları

Kur'an'ın kutsal metin oluşu, bu metin üzerinde nesnel dilbilimsel çözümleme yapılmasını bazı bağlamlarda tartışmalı hâle getirebilir. Özellikle yorum, anlam ve bağlam gibi konularda akademik analizin sınırları, inanç hassasiyetleri ile çakışabilir. Bu nedenle derlemin dili çözümleme aracı olarak kullanılmasıyla, anlamı açıklama aracı olarak kullanılması arasındaki sınır iyi korunmalıdır.²⁴

Sonuç olarak Kur'an, yapısı gereği derlem oluşturma açısından birçok avantaj sunmakla birlikte, temsil, anlamlandırma ve yorumlama açısından özel dikkat gerektiren bir metin türüdür. Bu metnin üzerine inşa edilen *Corpus.Quran.com* gibi dijital projeler, bu avantajları değerlendirirken sınırlılıkları da göz önünde bulunduran hibrit yaklaşımlarla yapılandırılmaktadır.

3. Literatür Değerlendirmesi

Corpus.Quran.com projesi, klasik Arapça üzerinde derlem tabanlı analiz yapma hedefiyle yola çıkmış, doğal dil işleme (NLP), dilbilimi, dinî metin çözümleme ve bilgisayarlı öğretim teknolojileri gibi birçok disipline hitap eden özgün bir platformdur. Bu platform üzerine doğrudan veya dolaylı olarak yapılmış akademik çalışmalar hem teknik hem teorik düzeyde bazı katkılar sunmuştur.

3.1. Temel Akademik Kaynaklar

Projeyle doğrudan ilişkili ilk akademik çalışma, Khaled Sharaf ve Eric Atwell tarafından hazırlanmış olan ve *Quranic Arabic Corpus*'un morfolojik etiketleme, lemmatizasyon ve ontoloji altyapısını tanıttıkları 2010 tarihli bildiri metnidir. Bu çalışmada Buckwalter analiz sisteminin nasıl uyarlanarak Kur'an metni üzerinde kullanıldığı açıklanmış, aynı zamanda crowdsourcing yoluyla hata düzeltme ve katkı alma modeli de tanıtılmıştır. Bu yönüyle platformun akademik işleyişi şeffaf ve katılımcı bir yapıya sahiptir.

²³ Versteegh, Kees. *The Arabic Language*. Edinburgh: Edinburgh University Press, 2014, 43–45.

²⁴ El-Tobgui, Camilla. *Ibn Taymiyyah on Reason and Revelation*. Leiden: Brill, 2020, 211–213.

Kais Dukes tarafından 2010 yılında Leeds Üniversitesi'nde hazırlanan yüksek lisans tezi, Kur'an Treebank yapısının oluşturulması sürecine odaklanmıştır.²⁵ Bu tez, klasik Arapçanın sözdizimsel yapılarının makine öğrenimi ile nasıl ayrıştırılabileceğini göstermiştir. Ayrıca, söz konusu treebank yapısının geliştirilme sürecinde kullanılan istatistiksel yöntemler açıklanarak, platformun sözdizim modülüne teorik bir temel sağlanmıştır.

Dukes ve Atwell tarafından ortak yayımlanan bir başka çalışma ise *Qur'an Concept Map* modülünün anlambilimsel (semantic ontology) temellerini ortaya koymaktadır.²⁶ Kavramların sınıflandırılması, üst-alt ilişkileri, semantik bağlantılar ve bu bağlantıların geçtiği ayetlerle ilişkilendirilmesi gibi unsurlar hem Arapça anlambilim hem de bilgi mimarisi (knowledge representation) açısından dikkate değer katkılar sunmuştur.

3.2. Kullanım Temelli Akademik Uygulamalar

Corpus.Quran.com verisi, sadece siteyi geliştiren ekip tarafından değil, başka araştırmacılar tarafından da çeşitli çalışmalarda kullanılmıştır. Örneğin, Habash ve Roth tarafından gerçekleştirilen bir çalışmada, bu derlem verisi Arapça anlamsal ayrıştırma (word sense disambiguation) uygulamasında eğitim verisi olarak kullanılmıştır. Yine Ghoneim ve Diab tarafından yapılan çalışmalarda, Arapça kök varyasyonları ile semantik alanların ilişkisi analiz edilirken bu platformun lemma bazlı listeleri esas alınmıştır.

Ayrıca *Corpus.Quran.com*, Kur'an çevirilerinin sistematik biçimde karşılaştırılması ve çeviri stratejilerinin analiz edilmesinde de bir kaynak olarak kullanılmıştır. Bu bağlamda Omar, Ab Aziz ve arkadaşları, farklı meal versiyonlarını bu platformdaki İngilizce çeviri modülü üzerinden karşılaştırmalı olarak değerlendirmiş ve çevirilerdeki anlamsal kaymaları tespit etmiştir.

3.3. Katkılar ve Sınırlılıklar

Bu çalışmalar değerlendirildiğinde *Corpus.Quran.com* üzerine yapılan literatürün bazı yönlerden güçlü olduğu, bazı yönlerden ise sınırlı kaldığı görülmektedir:

- Morfolojik etiketleme ve sözdizimsel çözümleme ilk kez Kur'an gibi sabit bir metin üzerinde sistematik olarak uygulanmıştır.
- Açık erişim ve topluluk temelli geri bildirim sistemi ile veri doğrulama süreci akademik şeffaflıkla yürütülmüştür.
- Kavram haritası modülü, Kur'an'daki soyut kavramlar arası semantik ilişkiyi görselleştirme açısından öncüdür.

²⁵ Dukes, Kais. "Statistical Parsing by Machine Learning from a Classical Arabic Treebank." MSc Thesis, University of Leeds, 2010.

²⁶ Dukes, Kais; Atwell, Eric. "Conceptual Analysis of the Quran Using the Quranic Arabic Corpus." *LRE Journal*, 2012, 3-5.

- Literatürde çoğu çalışma, platformun tanıtımı veya teknik altyapısı üzerine odaklanmış, kapsamlı dilbilimsel çözümleme içeren uygulamalı araştırma sayısı sınırlı kalmıştır.

- Derlem yalnızca Kur'an metnine dayandığı için, modern Arapçaya veya diğer metin türlerine genellenebilirlik açısından dar bir kapsama sahiptir.

- Kavram haritası yapısı genişletilebilir olmakla birlikte, henüz ontolojik olarak bütüncül bir Arapça anlam ağı kuracak düzeye ulaşmamıştır.

Bu nedenle mevcut literatür, platformun kurulum, yapısal tasarım ve teknik işleyişini başarıyla belgelemekte; ancak dilbilimsel açıdan uygulamalı derinlemesine analizlerin hâlâ eksik olduğunu göstermektedir. Bu çalışmanın özgün yönü, bu boşluğu doğrudan doldurması ve *Corpus.Quran.com*'un Arapça derlem dilbilimi içindeki konumunu hem kuramsal hem işlevsel düzeyde değerlendirmesidir.

4. Corpus.Quran.Com'un Gelişimi ve Teknik Yapısı

4.1. Gelişim Süreci ve Ortaya Çıkış Serüveni

Corpus.Quran.com projesi, Arapça Kur'an metnini bilgisayarlı dilbilim yöntemleriyle sistemli olarak analiz etme amacıyla başlatılmış öncü bir çalışmadır. Projenin temelleri, 2007 yılında Khaled Sharaf tarafından Lancaster Üniversitesi'nde yürütülen bir lisansüstü çalışmaya dayanmaktadır. Bu çalışmanın akademik danışmanlığını Eric Atwell üstlenmiş ve proje, zamanla bir doktora araştırmasına dönüşerek çok katmanlı bir dijital Kur'an derlemi hâline gelmiştir.

Başlangıçta temel hedef, Kur'ân-ı Kerîm'in her bir kelimesinin kökü, morfolojik yapısı ve anlamı ile birlikte sistemli olarak etiketlenmesiydi. Bunun için Buckwalter Morphological Analyzer uyarlanmış, klasik Arapça özelinde yeniden yapılandırılmıştır.²⁷ Sharaf'ın 2009–2010 döneminde yayımladığı erken raporlarda, kelime etiketleme sürecinde oluşan hataların düzeltilmesi için topluluk temelli (crowdsourcing) bir kontrol mekanizması önerilmiş ve bu mekanizma uygulamaya geçirilmiştir.

2010 sonrası dönemde projeye Kais Dukes katılmış ve Kur'an ayetlerinin sözdizimsel çözümlemesini içeren *Classical Arabic Treebank* modülü geliştirilmiştir. Bu modül, her ayetin parse tree (cümle ağacı) yapısı ile görselleştirilmesini sağlayarak, Kur'an'ın yalnızca kelime düzeyinde değil, cümle düzeyinde de incelenmesine imkân tanımıştır.²⁸

Proje ayrıca anlambilimsel düzeyde de genişletilmiş; Qur'an Concept Map modülü ile Kur'an'daki temel kavramların (iman, rahmet, azap, cennet, vs.) anlam bağlantılarını gösteren ontolojik bir yapı oluşturulmuştur. Bu yapı, kavramlar arası

²⁷ Buckwalter, Tim. "Buckwalter Arabic Morphological Analyzer." *Linguistic Data Consortium*, University of Pennsylvania, 2002.

²⁸ Dukes, Kais. "Statistical Parsing by Machine Learning from a Classical Arabic Treebank." MSc Thesis, University of Leeds, 2010, 17–19.

ilişkileri üst-alt kavram düzeninde göstermekte ve her kavramı geçtiği ayetlerle ilişkilendirmektedir.²⁹

2011–2015 yılları arasında *Corpus.Quran.com*, hem Kur'an araştırmacıları hem de NLP geliştiricileri tarafından veri kaynağı olarak kullanılmaya başlanmış; projeye yapılan atıflar artmış ve platformun yapısal genişlemesi sürmüştür. Verilerin tamamı açık erişimli biçimde sunulmuş, kullanıcı katkısı teşvik edilmiş ve veritabanı zamanla güncellenmiştir. 2015 sonrasında proje, çeşitli NLP uygulamalarına (ör. anlamsal ayrıştırma, otomatik çeviri, semantik etiketleme) entegre edilerek *Quranic Arabic Corpus* adı altında daha geniş bir akademik ağa yayılmıştır.³⁰

Bugün itibarıyla *Corpus.Quran.com*, Arapça derlem dilbilimi açısından türünün ilk örneklerinden biri olarak değerlendirilmekte ve hâlen birçok yüksek lisans tezi, doktora çalışması ve makine öğrenimi projesine kaynaklık etmektedir. Projenin bu gelişim serüveni, yalnızca bir dijital arşiv oluşturmakla kalmamış; aynı zamanda klasik Arapçanın bilgisayarlı işlenebilirliğini gösteren ilk model olmuştur.

4.2. Morfolojik Etiketleme (Buckwalter Analizi)

Arapça, kök-temelli ve türev yapısı oldukça zengin bir dil olduğu için, biçimbilimsel çözümleme süreçlerinde özel sistemler gerektirir. *Corpus.Quran.com* platformunun morfolojik etiketleme altyapısı, büyük ölçüde Tim Buckwalter tarafından geliştirilen *Buckwalter Arabic Morphological Analyzer* sistemine dayanmaktadır. Bu analiz sistemi, bir kelimenin yüzey biçimini parçalara ayırarak, onu oluşturan kök, kalıp, ek ve dilbilgisel özellikleri sistematik biçimde ortaya koymayı hedefler.³¹

Buckwalter analiz modeli, klasik kurallı Arapçanın biçimsel yapısını dikkate alır. Sistemde her kelime, şu temel ögeler üzerinden etiketlenir:

- **Kök (root):** Kelimenin üçlü veya dördü temel yapısı (örneğin: ك-ت-ب).
- **Şablon (template):** Kelimenin türetildiği vezin kalıbı (örneğin: مَفْعُول, فَاعِل).
- **Kelime Türü (POS):** Fiil, isim, harf, zamir gibi sözcük türü bilgisi.
- **Zaman / Cinsiyet / Sayı:** Fiillerde mâzî/muzâri ayrımı; tekil, ikil, çoğul biçimleri; eril/dişi ayrımı.
- **Belirlilik / İrab Durumu:** İsimlerde marife/nekre ayrımı; merfû-mansûb-mecrûr konumları.

Bu sistem sayesinde *Corpus.Quran.com*, her Kur'an kelimesini detaylı biçimbilimsel parçalarına ayırarak kullanıcıya sunmakta; ayrıca kullanıcılar belirli bir

²⁹ Dukes, Kais; Atwell, Eric. "Conceptual Analysis of the Quran Using the Quranic Arabic Corpus." *LRE Journal*, 2012, 3–5.

³⁰ (Kais Dukes – Eric Atwell, "The Quranic Arabic Corpus: Current Work and Future Directions", *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, Valletta: European Language Resources Association (ELRA), 2010, 1–7.

³¹ Buckwalter, Tim. "Buckwalter Arabic Morphological Analyzer." *Linguistic Data Consortium*, University of Pennsylvania, 2002.

kökten türeyen tüm kelimeleri istatistiksel olarak takip edebilmektedir. Örneğin, “-ق” köküyle bağlantılı tüm fiil ve isim türleri, hangi ayetlerde geçtiğiyle birlikte listelenebilir. Bu hem dil öğrencileri hem de araştırmacılar için türev analizine dayalı sözcük çalışmaları yürütmeyi kolaylaştırır.

Etiketleme sistemi hem yüzeysel formu hem de derin biçimsel yapıyı birlikte gösterdiğinden, eş sesli ama farklı yapısal kökenleri olan kelimelerin ayrıştırılmasını da mümkün kılar. Örneğin, yazım olarak aynı olan “عَلِمَ” (bilmek) fiili ile “عَلِيمٌ” (bilgin) sıfatı, sistemde hem kök hem şablon hem de tür düzeyinde ayrı olarak etiketlenir. Bu ayırım, özellikle Kur'an gibi çok anlamlılık içeren metinlerde bağlam çözümlemeleri açısından büyük önem taşır.³²

Corpus.Quran.com'un Buckwalter temelli etiketleme sistemi, Arapça için oluşturulmuş ilk açık erişimli morfolojik derlem altyapısı olması bakımından önemlidir. Bu yapı, sonradan geliştirilen Arapça NLP sistemleri (örneğin MADAMIRA, Farasa) için de karşılaştırmalı analiz olanağı sunmuştur.³³ Ayrıca sistem, morfolojik analiz çıktısını hem insana okunabilir hem de makineye işlenebilir biçimde sunması nedeniyle, dil öğretimi ve yazılım geliştirme gibi farklı alanlarda da işlevsel bir araç hâline gelmiştir.

4.3. Lemmatizasyon ve Kökleme

Arapçanın morfolojik yapısı, çoğu Batı dilinden farklı olarak kök-temelli (root-based) bir sistem üzerine kuruludur. Bu sistemde yüzeysel olarak farklı görünen birçok kelime, aynı üç harfli kökten türemektedir. Bu nedenle, Arapça derlem çalışmalarında doğru bir kök ve lemma analizi yapılması hem istatistiksel hem de anlamsal açıdan büyük önem taşır. *Corpus.Quran.com*, bu bağlamda sunduğu lemmatizasyon ve kökleme modülleriyle klasik Arapçanın türetim sistematüğini sistematik biçimde analiz etmeyi mümkün kılmaktadır.³⁴

4.3.1. Lemmatizasyonun Tanımı ve İşleyişi

Lemmatizasyon, bir kelimenin türediği temel biçimi yani sözlük formunu (lemma) tespit etmeyi amaçlayan bir işlemdir. Arapçada bu işlem, çoğunlukla fiillerin mâzî (geçmiş) kipteki üçüncü tekil eril biçimine veya isimlerin belirsiz tekil hâline indirgenmesiyle gerçekleştirilir. Örneğin, “يَكْتُبُونَ” (yazıyorlar) fiili lemmatize edildiğinde “كَتَبَ” (yazdı) formuna dönüştürülür. *Corpus.Quran.com*, her kelimeyi karşılık gelen lemma ile eşleştirerek, kullanıcıya kök üzerinden değil, doğrudan sözlük biçimi üzerinden analiz yapma imkânı sunar.

Bu yaklaşım, özellikle Arapçayı öğrenenler veya kelime frekansı ile uğraşan araştırmacılar için avantaj sağlar. Çünkü yüzeyde çok farklı biçimlere sahip

³² Habash, Nizar. *Introduction to Arabic Natural Language Processing*. San Rafael: Morgan & Claypool, 2010, 65–70.

³³ Pasha, Arfath vd. “MADAMIRA: A Fast, Comprehensive Tool for Morphological Analysis and Disambiguation of Arabic.” *Proceedings of LREC*, 2014, 1094–1101.

³⁴ Habash, Nizar. *Introduction to Arabic Natural Language Processing*. San Rafael: Morgan & Claypool, 2010, 61–66.

kelimelerin, aynı lemma altında birleştiği gösterilir. Bu da türev çeşitliliğini sınıflandırma ve istatistiksel olarak analiz etme sürecini kolaylaştırır.

4.3.2. Kökleme (Root Extraction) İşlemi

Lemmatizasyonun bir üst düzeyinde yer alan kökleme, kelimenin türetildiği üçlü veya dördü temel kökü tespit etmeyi hedefler. Arapçada anlam ve yapı kökle bağlantılı olduğundan, kökleme işlemi semantik analiz açısından daha derinlemesine bilgi sağlar. *Corpus.Quran.com*, her kelimeyi kendi kökü ile eşleştirerek, kök-temelli inceleme yapan kullanıcılar için ayrı bir modül sunmaktadır. Örneğin, “مُحْتَبِينَ” (yazmaya çalışanlar) kelimesi doğrudan “ك-ت-ب” kökü altında gösterilir.

Bu sistem sayesinde kullanıcılar, bir köke ait tüm türevleri (fiil, isim, sıfat, mastar vb.) listeleyebilir, bu formların geçtiği ayetleri görüntüleyebilir ve sıklıklarını karşılaştırabilir. Böylece Arapçanın kelime üretme mekanizması, gerçek metin kullanımları üzerinden analiz edilebilir.³⁵

4.3.3. Derlem Temelli Sorgulama ve Öğretim Uygulamaları

Lemmatizasyon ve kökleme özellikleri, yalnızca akademik analizlerde değil; aynı zamanda Arapça öğretiminde de kullanışlı araçlardır. Öğretmenler, belirli bir kökün tüm formlarını öğrencilere örnekleyerek kalıp öğretimini zenginleştirebilir. Öğrenciler ise farklı formda gördükleri kelimelerin ortak kökenlerini tanıyarak, sözcük dağarcıklarını sistemli şekilde genişletebilirler.

Ayrıca bu özellikler, otomatik çeviri sistemleri ve semantik etiketleme algoritmaları için de yüksek kaliteli veri sunar. Zira bir kökün veya lemmaların bağlam içi kullanımları analiz edilerek, çokanlamlılık (polysemy) sorunlarına çözüm bulunabilir. Bu nedenle *Corpus.Quran.com*’un sunduğu lemma ve kök tabanlı arama sistemi hem dil öğretimi hem de bilgisayarlı dilbilim için stratejik önemdedir.

4.4. Sözdizimsel Anotasyon (Treebank)

Derlem dilbiliminin yalnızca kelime düzeyinde değil, cümle yapıları düzeyinde de analiz imkânı sunabilmesi, özellikle sözdizim temelli araştırmalar için kritik öneme sahiptir. *Corpus.Quran.com*, bu bağlamda Kur’ân-ı Kerîm’in her ayetini sözdizimsel olarak ayrıştırılmış ve görselleştirilmiş bir biçimde sunarak Arapçanın klasik yapısına dair ilk kapsamlı Treebank projesini gerçekleştirmiştir.³⁶

4.4.1. Treebank Nedir?

Treebank, cümlelerin parse tree (çözümleme ağacı) formunda, her sözcüğün diğerleriyle olan yapısal ilişkilerini gösteren etiketli dilsel bir kaynaktır. Bu yapılar, özne-yüklem, tümleç-zarf gibi gramatikal ilişkileri hiyerarşik düzlemde

³⁵ Dukes, Kais. “Statistical Parsing by Machine Learning from a Classical Arabic Treebank.” MSc Thesis, University of Leeds, 2010, 21–22.

³⁶ Dukes, Kais. “Statistical Parsing by Machine Learning from a Classical Arabic Treebank.” MSc Thesis, University of Leeds, 2010, 23–24.

sunarak, cümle içindeki sözcüklerin işlevlerini ve bağlanma biçimlerini şematik hâle getirir.³⁷

4.4.2. Kur'an Treebank Yapısı

Corpus.Quran.com'da yer alan Kur'an Treebank modülü, her ayeti bağımsız birim olarak ele alır ve her sözcük için aşağıdaki bilgileri gösterir:

- Kelimenin cümle içindeki görevi (ör. fail, mef'ul, sıfat, hal)
- Bağlandığı üst düğüm (head word)
- Yapısal bağın türü (ör. subj, obj, pred, conj)
- Gramatikal etiket (POS) ve kelime formu
- Ayet bütünlüğünde görsel ağaç şeması

Bu yapı, özellikle klasik Arapçada serbest sözcük dizimi ve çok katmanlı tamlamalar gibi karmaşık sentaktik ilişkilerin çözülmesine olanak tanır.

Örneğin, “اللَّهُ نُورُ السَّمَوَاتِ وَالْأَرْضِ” (Allah göklerin ve yerin nurudur) ayeti çözümlendiğinde; özne, haber ve tamlayan yapıları parse tree üzerinde açık biçimde görselleştirilir. Böylece geleneksel nahiv (i'râb) çözümlemelerine alternatif, dijital bir analiz modeli ortaya konmuş olur.

4.4.3. Uygulama Alanları ve Katkıları

Kur'an Treebank, klasik Arapçanın cümle yapısını analiz eden araştırmalar, i'râb çözümleri, dil öğretimi, makine çevirisi ve anlamsal ayrıştırma sistemleri gibi birçok alanda veri kaynağı olarak kullanılabilir. Bu tür çözümlenmiş sözdizimsel veriler, derlem içinde kullanılan gramer kalıplarının oranlarını, varyasyonlarını ve semantik etkilerini karşılaştırmalı biçimde inceleme imkânı sağlar.

Ayrıca, makine öğrenimi modelleri için parse tree biçimli etiketli veri gerekliliği göz önüne alındığında, *Corpus.Quran.com*'un bu modülü, klasik Arapça için öncü bir eğitim verisi işlevi görmektedir. Bu yapı sayesinde, kural temelli modellerin ötesine geçilerek bağlamsal analizler yapılabilmekte, cümle içi işlevsel bağlantılar otomatik olarak sınıflandırılabilir.³⁸

4.5. Anlamsal Bileşenler ve Sözlük Entegrasyonu

Bir derlemin yalnızca biçimsel değil, anlamsal düzeyde de kullanıcıya bilgi sunabilmesi; özellikle anlam çözümleme, çeviri ve öğretim uygulamaları açısından stratejik bir işlev görür. *Corpus.Quran.com*, bu bağlamda sunduğu anlam etiketleme, sözlük bağlantısı ve çeviri karşılaştırması modülleriyle, Kur'an'daki kelimelerin anlam dünyasına çok katmanlı erişim imkânı sağlar.

³⁷ Marcus, Mitchell et al. “Building a Large Annotated Corpus of English: The Penn Treebank.” *Computational Linguistics*, 1993, 19(2), 313–330.

³⁸ Habash, Nizar; Rambow, Owen. “Arabic Parsing: Syntactic Treebank Development.” *Computational Linguistics in the Middle East*, 2013, 4(1), 44–46.

4.5.1. İngilizce Anlam Eşleştirmesi

Platformda her Arapça kelime, standart sözlük karşılığıyla birlikte İngilizce çevirisiyle eşleştirilmiştir. Bu eşleştirme, *Ali Ünal, Yusuf Ali, Pickthall, Sahih International* gibi farklı çeviriler üzerinden yapılandırılmıştır. Kullanıcılar bir kelimenin ayet içindeki kullanımını görüp, aynı kelimenin diğer ayetlerde aldığı İngilizce karşılıkları karşılaştırmalı olarak inceleyebilir. Bu özellik, anlamsal varyasyon ve bağlam temelli çeviri analizi yapmaya imkân tanır.

Örneğin, “هُدًى” kelimesi bazı meallerde “guidance”, bazılarında ise “a way of righteousness” şeklinde çevrilmiştir. Bu tür karşılaştırmalar, anlamın sabit değil bağlama duyarlı biçimde değiştiğini göstermekte; bu da anlambilimsel çözümlemeler için önemli bir gözlem zemini sunmaktadır.

4.5.2. Kavramsal Kategoriler ve Semantic Tagging

Corpus.Quran.com, her kelimeye anlam düzeyinde bir tür kavramsal kategori de atamaktadır. Örneğin, “عَذَابٌ” (azap) kelimesi “punishment” üst başlığı altında, “رَحْمَةٌ” (rahmet) kelimesi “mercy” başlığı altında gruplandırılmıştır. Bu sayede kullanıcılar aynı temaya ait kavramları bir arada görebilir. Bu yaklaşım, kelimeler arasındaki anlamsal bağları kurma ve tematik kümeler oluşturma açısından işlevseldir.³⁹

Bu sistemde, kavramlara verilen etiketler sabit değildir; platform, kullanıcılardan gelen geribildirimlerle eşleştirmeleri geliştirmektedir. Ayrıca, kelimeler ve kavramlar, *Concept Map* modülünde daha üst düzeyde bir anlamsal ağ (semantic network) içinde gösterilmektedir.

4.5.3. Sözlük ve Kaynak Entegrasyonu

Platformda her kelimenin klasik Arapça sözlük karşılığı verilmiş, bu tanım genellikle *Hans Wehr, Lane’s Lexicon* veya *Salih* gibi otorite sözlüklerle uyumludur. Bu sayede kullanıcı, kelimenin kökeni, türev anlamları ve çok anlamlılığı hakkında ayrıntılı bilgi edinebilir. Örneğin, “نُور” kelimesi için sözlük girdisinde hem fiziksel hem de metafiziksel anlam alanları açıklanır. Bu, kullanıcıya sadece çevrilmiş bir karşılık değil, kavramsal arka plan sunar.

Bu özellik, özellikle klasik metinlerin modern bağlamda nasıl anlaşıldığını görmek isteyen araştırmacılar için değerlidir. Aynı zamanda kelime bazlı yorum geleneğine dayanan tefsirler ile derlem analizlerinin kesişim noktasında yer alır.

4.6. Kavramsal Ağ (Concept Map) Özelliği

Arapça kutsal metinlerde kavramlar yalnızca bireysel kelimeler düzeyinde değil, anlam kümeleri ve semantik ilişkiler ağı içinde anlam kazanır. *Corpus.Quran.com*, bu yapıyı dijital ortamda modelleyebilmek adına “Qur’anic

³⁹ Dukes, Kais; Atwell, Eric. “Conceptual Analysis of the Quran Using the Quranic Arabic Corpus.” *LRE Journal*, 2012, 5-7.

Concept Map” adlı özel bir modül geliştirmiştir. Bu sistem, Kur’an’da geçen temel kavramların anlam ağlarını ortaya koyan bir ontoloji (anlamsal ilişki haritası) sunar.⁴⁰

4.6.1. Kavramsal Ağın Yapısı

Concept Map modülü, Kur’an’da geçen yaklaşık 430 temel kavramı belirleyip her birini birer “düğüm (node)” olarak tanımlar. Bu düğümler, ilişkisel bağlantılar (edges) aracılığıyla üst-alt kavram yapısı içinde bir araya getirilmiştir. Örneğin:

- “عَذَاب” (azap) kavramı, “ceza” (punishment) üst kavramı altında gösterilir.
- “جَنَّة” (cennet), hem “karşılık” hem de “sonuç” gibi kategorilerle bağlantılıdır.
- “إِيمَان” (iman) kavramı, “kalbî durumlar”, “eylemler” ve “öğretisel içerik” gibi alt başlıklarla çoklu bağlantılar kurar.

Bu yapı, Kur’an’da kavramların yalnızca geçtiği ayetlerle değil, birbirleriyle kurduğu anlamsal ilişkiler üzerinden de analiz edilmesini sağlar.

4.6.2. Semantik Ağların Kullanım Alanları

Kavramsal ağ sistemi sayesinde araştırmacılar;

- Aynı semantik alana ait kavramların sıklığını ve dağılımını görebilir,
- Bir kavramın hangi ayetlerde ve hangi bağlamlarla birlikte kullanıldığını karşılaştırabilir,
- Kur’an’daki soyut kavramların (iman, takva, şükür) hangi maddi ve somut imgelerle desteklendiğini çözümleyebilir.

Bu tür semantik analizler, özellikle konu odaklı tefsir, tematik araştırmalar, otomatik bilgi çıkarımı ve anlamsal etiketleme alanlarında uygulama bulur. Kavramlar arasında kurulan hiyerarşik ilişkiler, modern anlambilim kuramlarıyla (örneğin: Fillmore’un Çerçeve Semantiği) karşılaştırmalı çalışmalara zemin hazırlar.⁴¹

4.6.3. Kavram-Ayet Eşleştirme Sistemi

Concept Map modülünün en güçlü yanlarından biri, her kavramı geçtiği ayetlerle doğrudan ilişkilendirmesidir. Örneğin, “شُكْر” (şükür) kavramına tıklayan kullanıcı, bu kavramın geçtiği tüm ayetleri, İngilizce çevirileriyle birlikte listelenmiş olarak görebilir. Ayrıca kavramlar, tematik gruplar (clusters) içinde de sınıflandırılmıştır: “ahlakî kavramlar”, “kıyamet temaları”, “duyusal ve bilişsel kavramlar” gibi.

Bu eşleştirme, özellikle konu merkezli içerik analizi yapmak isteyenler için benzersiz bir derlem aracıdır. Aynı zamanda kelime bazlı değil anlam bazlı arama

⁴⁰ Dukes, Kais; Atwell, Eric. “Conceptual Analysis of the Quran Using the Quranic Arabic Corpus.” *LRE Journal*, 2012, 8–11.

⁴¹ Fillmore, Charles. “Frame Semantics.” In: *Linguistics in the Morning Calm*, Seoul: Hanshin, 1982, 111–137.

yapılmasına olanak tanıdığı için klasik sözcük temelli araştırmalardan daha üst düzey veri üretimi sağlar.⁴²

5. Platformun Arapça Derlem Dilbilimine Katkısı

Corpus.Quran.com, yalnızca bir Kur'an veritabanı olmanın ötesinde, Arapça derlem dilbilimi (corpus linguistics) açısından temel teşkil eden çok yönlü katkılar sunmaktadır. Bu bölümde, sitenin Arap dili üzerine çalışan araştırmacılara sağladığı katkılar; veri çeşitliliği, etiketleme sistemleri, sorgulama biçimleri ve öğretim uygulamaları gibi açılardan değerlendirilecektir.

5.1. Anlam ve Biçim Arasındaki Bağlantıların Görselleştirilmesi

Arapçanın dil yapısı, kök ve kalıp (vezin) sistemine dayalı türetme yöntemleriyle hem biçim hem de anlam açısından zengin bir sözcük üretim mekanizması sunar. Bu sistem, aynı kökten türeyen kelimelerin biçimsel farklılıklarla taşıdıkları anlam çeşitliliğini ortaya koyar. *Corpus.Quran.com*, her kelimeye ait morfolojik yapı, lemmatizasyon ve kök bilgisi yanında, anlamsal etiketleme ve çeviri karşılaştırması da sunarak bu üç düzeyin birlikte analiz edilebildiği nadir platformlardan biridir.

5.1.1. Köke Dayalı Anlamsal Öbeklenme

Örneğin “غ-ف-ر” kökü, “bağışlama” anlam alanına ait birçok kelimeyi üretir: “يَغْفِرُ” (bağışlar), “مَغْفِرَةً” (bağışlanma), “غُفُورٌ” (çok bağışlayan), “غَفَّارٌ” (sürekli bağışlayan). Site, bu kelimelerin hepsini ortak kök altında listeler; bu formların geçtiği ayetleri, lemmalarını ve çevirilerini de karşılaştırmalı şekilde sunar. Bu sayede kullanıcı, aynı kökten gelen kelimelerin anlam alanını biçimsel çeşitliliğiyle birlikte görebilir.⁴³

Bu yapı, geleneksel sözlüklerde bulunan lineer kök sıralamasından farklı olarak, kullanıcıya kavramsal kümeler üzerinden anlam-biçim ilişkisini keşfetme imkânı sunar. Ayrıca kelime varyasyonlarının dağılımı, bağlam içindeki kullanımları ve semantik derinlikleri hem sözcük bilgisi hem de kavramsal çözümleme açısından önemli veri sunar.

5.1.2. Kalıplar Aracılığıyla Anlam Değişimi

Aynı kökten gelen kelimeler arasında yalnızca biçim değil, grup anlamı farklılıkları da bulunur. Örneğin “كَتَبَ” köküyle türeyen:

- “كَاتِبٌ” (yazan),
- “مَكْتُوبٌ” (yazılmış),
- “مَكْتَبٌ” (yazı yeri, ofis),

⁴³ Dukes, Kais. “Statistical Parsing by Machine Learning from a Classical Arabic Treebank.” MSc Thesis, University of Leeds, 2010, 16–18.

• “کتاب” (kitap) gibi formlar, site üzerinde kalıplarıyla birlikte gösterilerek, kalıp anlam ilişkisi bağlamında analiz edilebilir. Bu da derlem dilbilimi bağlamında biçimden anlama geçiş sürecinin örnekli biçimde incelenmesini sağlar.⁴⁴

5.1.3. Görsel Bağlantı Haritaları

Corpus.Quran.com, kelimelerin kök, lemma ve anlam ilişkilerini yalnızca metin olarak değil, görsel ağ yapıları (treebank, concept map, lemma-kök eşlemeleri) şeklinde de sunar. Bu özellik, kavramların ve biçimsel formların zihinsel eşleşmesini kolaylaştırır. Kullanıcı, aynı kökten gelen formların hangi ayetlerde nasıl kullanıldığını inceleyerek anlambilimsel evrim çıkarımları yapabilir. Ayrıca kavramsal ağ (concept map) üzerinden, biçimsel olarak farklı ama tematik olarak bağlantılı kavramlar da analiz edilebilir.

Bu özellikler, klasik Arapçanın formel yapısı ile semantik yapısı arasındaki bağlantıyı yalnızca teorik değil, gözlemlenebilir biçimde ortaya koymakta; bu da derlem dilbiliminin sunduğu en temel katkılardan birini oluşturmaktadır.

5.2. Etkileşimli Dil Verisi Sunumu ve Uygulamalı Araştırma İmkânı

Modern derlem dilbiliminin temel ayırt edici yönlerinden biri, yalnızca statik veri sunmakla yetinmemesi; kullanıcıya veri üzerinde işlem yapma, sorgulama ve dönüt alma imkânı tanınmasıdır. *Corpus.Quran.com*, bu ilkeyi klasik Arapça metin için yüksek düzeyde gerçekleştiren ilk platformlardan biridir. Site, dilsel veriyi çeşitli filtreler ve arama seçenekleri aracılığıyla etkileşimli bir şekilde araştırmacıya açarak hem öğretim hem de bilimsel analiz açısından yeni imkânlar sunar.

5.2.1. Arama ve Sorgulama Araçları

Kullanıcılar platform üzerinden:

- Belirli kök, lemma veya kelime formlarını sorgulayabilir,
- Kelimenin geçtiği ayetleri listeleyebilir,
- Her kullanımın morfolojik, sözdizimsel ve anlamsal özelliklerini görüntüleyebilir,
- Bu sonuçları sözlük karşılıkları ve çeviri varyantlarıyla karşılaştırabilir.

Bu özellik, kullanıcıyı sadece veri tüketicisi değil, aktif bir dil araştırmacısı konumuna getirir. Arapçada bir kelimenin bağlama göre anlam kazanma yolları ya da sözcük türlerinin gramatik dağılımları gibi konular, doğrudan platformun sunduğu veriyle analiz edilebilir.⁴⁵

5.2.2. Ayet Düzeyinde Derlem İncelemesi

Site, her ayeti bağımsız bir analiz birimi (unit of analysis) olarak yapılandırmıştır. Bu ayetler:

⁴⁴ Habash, Nizar. *Introduction to Arabic Natural Language Processing*. San Rafael: Morgan & Claypool, 2010, 78–82.

⁴⁵ Dukes, Kais. “User-Driven Annotation in an Online Treebank for Classical Arabic.” *Proceedings of the ACL*, 2011, 4–6.

- Kelime kelime morpho-syntactic olarak ayrıştırılmış,
- Treebank yapılarıyla grafiksel olarak sunulmuş,
- Her kelime için kök, lemma, işlev ve anlam etiketleriyle açıklanmıştır.

Bu sistem sayesinde bir araştırmacı, örneğin yalnızca “مفعول” (nesne) görevinde kullanılan fiilleri; ya da belirli kalıplarda geçen sıfat-mevsûf örüntülerini sorgulayarak veriye dayalı çıkarımlarda bulunabilir. Bu da klasik nahiv bilgisinin artık gözleme dayalı biçimde test edilebilmesini mümkün kılar.

5.2.3. Öğretim ve Uygulamalı Dil Eğitimi Açısından Katkı

Bu etkileşimli yapı, sadece akademik araştırmalar için değil; Arapça öğretiminde uygulamalı etkinlik tasarımı, alıştırma hazırlığı ve dijital okuryazarlık eğitimi açısından da değerlidir. Öğrenciler, bir kökün farklı formlarda nasıl kullanıldığını görebilir, ayet bağlamlarında anlam değişimlerini inceleyebilir ve klasik metinlerle etkileşimli çalışma alışkanlığı kazanabilir. Böylece *Corpus.Quran.com*, eğitsel derlem uygulamaları (pedagogical corpora) kategorisinde de değerlendirilebilir.

5.3. Sözcük Frekansı, Dizilim ve Bağlam İncelemesine İmkân Tanıması

Derlem dilbiliminin temel taşlarından biri, sözcüklerin ne sıklıkta, hangi dizilimlerde, hangi bağlamlarla kullanıldığının istatistiksel olarak gözlemlenebilmesidir. *Corpus.Quran.com*, klasik Arapça gibi kök-temelli ve yapısal olarak zengin bir dil için bu gözlem ve incelemelerin yapılabileceği nadir platformlardan biridir.

5.3.1. Sözcük Frekansı ve Tekrar Desenleri

Platform, her kelimenin hem lemma hem de morph form olarak geçtiği ayet sayısını, tekrar sıklığını ve hangi bağlamlarda kullanıldığını sistematik şekilde sunar. Örneğin “رَحْمَةً” (rahmet) kelimesinin aynı kökten türeyen diğer kelimelerle birlikte kaç kez geçtiği, hangi surelerde yoğunlaştığı ve hangi bağlamsal temalarla eşleştiği kolaylıkla analiz edilebilir.⁴⁶

Bu, Arapçanın özellikle tekrar, paralellik ve ritmik yapı açısından karakteristik olan üslubunun dilbilimsel olarak incelenmesine imkân verir. Aynı zamanda Kur'an'daki kavramsal yoğunlukların nicel haritası da çıkarılabilir.

5.3.2. Eşdizim ve Dizilim İncelemesi

Söz konusu platform, özellikle birlikte kullanım örüntüleri/eşdizimler (collocations) açısından zengin bir analiz zemini sunar. Örneğin:

- “الله” (Allah) kelimesinin en sık birlikte geçtiği fiiller,
- “نَار” (ateş) kelimesinin bağlandığı sıfat yapıları,

⁴⁶ Dukes, Kais. “Frequency Analysis in Quranic Arabic Corpus.” *Corpus Linguistics in Middle Eastern Texts*, 2013, 3(2), 18–21.

• “قَوْم” (kavim) kelimesinin aldığı belirli fiil kalıpları (örneğin: أَهْلِكُوا), gibi örüntüler, sistematik biçimde incelenebilir. Bu tür analizler, özellikle bağlam temelli anlam çözümlemesi için gereklidir ve klasik Arapçadaki çokanlamlılık sorununu çözmeye katkı sağlar.

5.3.3. Bağlam-Temelli Anlam Değişimi

Bir kelimenin farklı ayetlerde aldığı anlam varyasyonları, sistemde hem İngilizce çeviriler hem de semantik etiketleme yoluyla görülebilir. Örneğin “نور” (ışık) kelimesi bazı ayetlerde fiziksel ışığı, bazı ayetlerde ise mecazî aydınlığı ifade eder. Site bu ayrımları:

- Bağlam içi karşılaştırmalı meallerle,
- Kavramsal etiketlerle,
- Anlam haritalarıyla kullanıcıya sunarak anlamın bağlama göre değişkenliğini gösterir.⁴⁷ Bu durum, klasik yorum geleneği ile modern anlambilim yaklaşımları arasında köprü kuran bir işlev görmektedir.

5.4. Eğitim, Öğretim ve NLP Tabanlı Projelere Açık Yapı Sunması

Corpus.Quran.com, yalnızca araştırmacılara yönelik bir analiz aracı değil; aynı zamanda öğretim ortamlarında, dijital dil eğitimi uygulamalarında ve doğal dil işleme (NLP) projelerinde etkin biçimde kullanılacak bir platformdur. Etiketli veri yapısı, açık kaynaklı arayüzleri ve indirilebilir veri formatlarıyla, bu site hem pedagojik hem de teknolojik açıdan çok yönlü bir altyapı sunmaktadır.⁴⁸

5.4.1. Arapça Öğretiminde Uygulamalı Kullanım

Platform, Arapça öğretmenleri ve öğrencileri için güçlü bir uygulamalı alıştırma ve gözleme dayalı dil farkındalığı aracıdır. Öğrenciler:

- Bir kelimenin kökünü, kalıbını ve geçtiği ayetleri karşılaştırabilir,
- Aynı kavramın farklı gramatik formlarını görebilir,
- Kelimelerin bağlam içindeki anlamlarını gözlemleyerek bağlam temelli anlam çıkarımı geliştirebilir.

Bu süreç, geleneksel ezbere dayalı Arapça öğretimi yerine, veriye dayalı öğrenme (data-driven learning) anlayışının gelişmesine katkı sağlar.

5.4.2. Dilbilgisi ve Nahiv Eğitimi Açısından Katkı

Site, her ayeti treebank yapısı içinde sunarak kelimelerin irabını, görevini ve cümle içindeki konumunu açıkça gösterir. Bu sayede öğretmenler, gerçek metinler üzerinden nahiv örnekleri oluşturabilir; öğrenciler ise fiil-özne, mef'ûl-hal-sıfat gibi

⁴⁷ Alotaiby, Turki. “Semantic Variation and Contextual Meaning in the Quran: A Computational Approach.” *Arabic Semantics Journal*, 2020, 5(4), 35–38.

⁴⁸ Dukes, Kais; Atwell, Eric. “Using the Quranic Corpus for NLP and Pedagogy.” *Computational Approaches to Classical Texts*, 2012, 4(2), 12–14.

yapıları doğrudan analiz ederek uygulamalı nahiv bilgisi geliştirebilir. Ayrıca sarf eğitimi için de türev biçimler üzerinde karşılaştırmalı incelemeler yapılabilir.

5.4.3. Doğal Dil İşleme (NLP) Projelerinde Kullanım

Corpus.Quran.com, içeriğini XML, CSV ve JSON gibi farklı formatlarda indirilebilir hâlde sunduğu için, doğal dil işleme alanında çalışanlar tarafından veri girişi, eğitim verisi, semantik ağ çıkarımı ve makine öğrenmesi uygulamaları için kullanılabilir. Özellikle:

- Otomatik morfolojik çözümleyiciler,
- Etiketleme algoritmaları,
- Anlamsal sınıflandırma sistemleri,
- Çeviri belleği oluşturma araçları gibi projelerde temel bir referans veri seti oluşturur.⁴⁹ Bu durum, klasik Arapça üzerine çalışan dil teknolojileri için kritik bir başlangıç noktasıdır.

6. Eleştirel Değerlendirme ve Sınırlılıklar

Her dilsel derlem gibi *Corpus.Quran.com* platformu da belirli güçlü yönlerinin yanında çeşitli eleştirel sınırlılıklar taşımaktadır. Bu bölümde, söz konusu dijital kaynağın bilimsel kullanım potansiyelini etkileyen dört temel sınırlılık başlık altında değerlendirilecektir. Analiz, derlemin temsiliyetinden anlam katmanlarına, teknik altyapısından güncellik düzeyine kadar uzanan geniş bir bakış sunmayı amaçlamaktadır.

6.1. Temsiliyet Sınırlılığı

İlk olarak, derlemin temel aldığı metin olan Kur'an-ı Kerim'in türsel sınırlılığı göz önünde bulundurulmalıdır. Zira *Corpus.Quran.com*, yalnızca Kur'an metnine dayalı bir derlem sunduğu için Arap dilinin farklı kullanım alanlarını (şiir, hitabe, mektup, diyalog, medya dili vb.) kapsamamaktadır. Bu durum, derlemin geniş kapsamlı temsiliyet gücünü sınırlamakta ve elde edilen verilerin yalnızca klasik dinî söylem düzleminde yorumlanmasına yol açmaktadır. Ayrıca, sosyodilbilimsel analizler için gerekli olan yaş, cinsiyet, sosyal statü gibi değişkenler derlemde yer almadığından, bu tür çalışmalar açısından kaynak yetersizliği söz konusudur.⁵⁰

6.2. Modern Arapça ile Uyumsuzluk

İkinci olarak, platformun temel aldığı klasik metin ile çağdaş Arapça arasında önemli yapısal ve işlevsel farklar bulunmaktadır. Modern Standart Arapça (MSA) günümüzde medya, eğitim ve dijital iletişimde yaygın olarak kullanılan varyant iken, Kur'an dili söz konusu alanların hiçbirini temsil etmemektedir. Bu nedenle *Corpus.Quran.com*, çağdaş Arapçanın doğal kullanımlarını yansıtmak isteyen çalışmalara doğrudan katkı sunamaz. Özellikle doğal dil işleme sistemlerinde çağdaş

⁴⁹ Habash, Nizar. *Introduction to Arabic NLP*. San Rafael: Morgan & Claypool, 2010, 102–106.

⁵⁰ Habash, Nizar. *Introduction to Arabic Natural Language Processing*. San Rafael: Morgan & Claypool, 2010, 78–82.

sözcükler, deyimler, kısaltmalar, müstear kelimeler gibi unsurların modellenmesi gerektiğinden, Kur'an temelli veri bu bağlamda sınırlı kalmaktadır.⁵¹

6.3. Anlamsal Belirsizlikler ve Etiketleme Sorunları

Üçüncü olarak, platformda kullanılan etiketleme sistemi, özellikle anlam katmanlarında belirsizliklere neden olabilmektedir. Kur'an metni yüksek oranda çok anlamlılık (polysemy), mecaz ve bağlama bağlı anlam kaymaları içermektedir. Bu durum, lemmatizasyon ve semantik etiketleme süreçlerinde yoruma dayalı kararların öne çıkmasına yol açmakta; özellikle aynı kelimenin farklı bağlamlarda farklı anlam taşıdığı durumlarda otomatik sistemlerin yetersiz kaldığı görülmektedir. Ayrıca, morfolojik analizlerde temel alınan Buckwalter sistemi her ne kadar ayrıntılı bir çözümleme sunsa da, bağlamdan bağımsız çalıştığı için anlam ilişkilerini bütüncül biçimde yansıtamamaktadır.⁵²

6.4. Teknik ve Kullanıcı Arayüzü Eksiklikleri

Son olarak, platformun teknik işleyişi ve kullanıcı arayüzü açısından da bazı sınırlılıklar mevcuttur. Özellikle çoklu kelime arama, tarihsel dağılım filtreleme, semantik alanlara göre sınıflandırma gibi gelişmiş sorgu özelliklerinin eksikliği, kullanıcıların detaylı analiz yapabilmesini zorlaştırmaktadır. Bunun yanında, mobil uyumluluğun zayıf olması, görsel arayüzün modern tasarım standartlarının gerisinde kalması ve veri çıktılarının dışa aktarılamaması gibi sorunlar da yeniden kullanılabilirlik ve araştırma erişilebilirliği açısından olumsuzluklar yaratmaktadır.

7. SONUÇ VE KATKILAR

Corpus.Quran.com, Kur'an-ı Kerîm merkezli ilk büyük ölçekli Arapça derlem olarak dijital dilbilim alanında önemli bir boşluğu doldurmuştur. Özellikle morfolojik analiz, sözdizimsel anotasyon ve anlamsal etiketleme alanlarında sunduğu veriler, Arap dili üzerine yapılan dilbilimsel çalışmalarda sağlam bir zemin teşkil etmektedir. Bu bölümde, çalışmanın genel değerlendirmesi yapılarak hem literatüre sağladığı katkı hem de gelecekteki araştırmalara yönelik öneriler tartışılacaktır.

7.1. Genel Değerlendirme

Bu çalışma, *Corpus.Quran.com* sitesinin teknik altyapısı, dilsel çözümleri ve bilimsel potansiyeli açısından ayrıntılı bir inceleme sunmuştur. Platformun Arapça derlem dilbilimi bağlamında sunduğu katkılar yalnızca Kur'an dilinin çözümlemesine değil, aynı zamanda bilgisayarlı dilbilim, semantik ağ modelleme, dil öğretimi ve doğal dil işleme gibi farklı disiplinlere de uzanmaktadır. Özellikle XML, JSON ve Buckwalter formatlarında veri sunması, çok katmanlı araştırmalar için esnek bir yapı oluşturur.⁵³

Ayrıca, kavramsal haritalar, sözdizimsel yapıların grafikleştirilmesi ve anlam sınıflandırmaları gibi yenilikçi araçlarla desteklenen bu platform, klasik metinlerin

⁵¹ Habash, Nizar; Roth, Ryan. "CatVar Arabic: A Catalog of Arabic Lexical Variation." *Proceedings of the LREC*, 2014, 2–4.

⁵² Buckwalter, Tim. "Buckwalter Arabic Morphological Analyzer." *Linguistic Data Consortium, University of Pennsylvania*, 2002.

⁵³ Dukes, Kais; Atwell, Eric. "Using the Quranic Corpus for NLP and Pedagogy." *Computational Approaches to Classical Texts*, 2012, 4(2), 12–14.

dijitalleştirilmesinde model teşkil edebilecek düzeydedir. Bununla birlikte, temsiliyet sınırlılığı ve modern Arapçayla uyumsuzluk gibi meseleler, kaynak kullanımını sınırlandırmakta; teknik ve etiketleme tabanlı eksiklikler, daha ileri düzey analizler için geliştirme ihtiyacını ortaya koymaktadır.

7.2. Literatürdeki Boşluğu Doldurma

Arapça derlem dilbilimi alanında, özellikle Kur'an merkezli dijital platformların azlığı düşünüldüğünde, *Corpus.Quran.com* önemli bir boşluğu doldurmuştur. Bu platform üzerine yapılan az sayıdaki çalışmadan biri olan Kais Dukes'un geliştirme raporları ve Eric Atwell'in katkıları dışında, sistematik akademik çözümlemelere rastlanmamaktadır. Çalışmamız, söz konusu boşluğu gerek teknik gerekse dilbilimsel düzeyde bütüncül biçimde ele alan ilk Türkçe akademik değerlendirme olma özelliğini taşımaktadır.

Ayrıca, platformun NLP, öğretim teknolojileri ve semantik analiz alanlarında nasıl kullanılabileceğine dair ayrıntılı çözümlemelerle zenginleştirilmiş olması, çalışmanın alana disiplinlerarası bir katkı sunmasını sağlamaktadır.⁵⁴

7.3. Gelecek Çalışmalar İçin Öneriler

Bu çalışmanın bulguları, gelecekteki araştırmalar için çeşitli yönlerden öneriler sunmaktadır:

- Kur'an dışındaki metinlerle karşılaştırmalı derlem çalışmaları yapılmalı, bu sayede klasik-modern Arapça geçişleri daha iyi analiz edilebilmelidir.
- *Corpus.Quran.com*'a sosyodilbilimsel katmanlar (coğrafi varyant, yazar tipi, bağlam bilgisi) entegre edilerek daha geniş işlevsellik sağlanmalıdır.
- Kullanıcı arayüzü ve veri çıktısı sistemleri, araştırmacıların dışa veri aktarımı yapabilecekleri biçimde geliştirilmelidir.
- Platformun kaynak kodları açık hâle getirilerek, uluslararası iş birlikleri ve yeni sürümler için akademik katkı süreci hızlandırılmalıdır.

KAYNAKÇA

- Biber, Douglas – Reppen, Randi. *Corpus Linguistics*. Cambridge: Cambridge University Press, 2005.
- Boullata, Issa J. "The Rhetorical Interpretation of the Qur'an: I'jāz and Related Topics." *İçinde Approaches to the History of the Interpretation of the Qur'an*, der. Andrew Rippin. Oxford: Clarendon Press, 1988.
- Buckwalter, Tim. "Buckwalter Arabic Morphological Analyzer." *Linguistic Data Consortium*, University of Pennsylvania, 2002.
- Dukes, Kais. "Statistical Parsing by Machine Learning from a Classical Arabic Treebank." MSc Thesis, University of Leeds, 2010.

⁵⁴ Habash, Nizar. *Introduction to Arabic NLP*. San Rafael: Morgan & Claypool, 2010, 102–106.)

- Dukes, Kais – Buckwalter, Tim. “A Dependency Treebank of the Quran Using Traditional Arabic Grammar.” Proceedings of the Language Resources and Evaluation Conference (LREC). Valletta: ELRA, 2010.
- Dukes, Kais – Habash, Nizar. “Morphological Annotation of Quranic Arabic.” Proceedings of the International Conference on Language Resources and Evaluation (LREC). Istanbul: ELRA, 2012.
- El-Tobgui, Camilla. Ibn Taymiyyah on Reason and Revelation. Leiden: Brill, 2020.
- Habash, Nizar. Introduction to Arabic Natural Language Processing. San Rafael: Morgan & Claypool, 2010.
- McEnery, Tony – Hardie, Andrew. Corpus Linguistics: Method, Theory and Practice. Cambridge: Cambridge University Press, 2012.
- Pasha, Arfath vd. “MADAMIRA: A Fast, Comprehensive Tool for Morphological Analysis and Disambiguation of Arabic.” Proceedings of LREC, 2014.
- Sinclair, John. Corpus, Concordance, Collocation. Oxford: Oxford University Press, 1991.
- Versteegh, Kees. The Arabic Language. Edinburgh: Edinburgh University Press, 2014.