International Journal of Current Social Science

| cusos.cumhuriyet.edu.tr | Founded: 2022 Available online Publisher: Sivas Cumhuriyet Üniversitesi

Corpus Linguistics in the Social Sciences: Method, Theory, and Epistemological Transformation

Mustafa Kayapınar[1,a,*]

[1] *Sivas Cumhuriyet Üniversitesi*
**Corresponding author*

Research Article

History

Received: 03/11/2025
Accepted: 02/12/2025

ABSTRACT

This article comprehensively examines the theoretical, methodological, and epistemological impact of corpus linguistics on the field of social sciences. Corpus linguistics does not merely study the structural properties of language but also investigates its social, cultural, and ideological dimensions. It conceptualizes language as an empirical record of human behavior, thought patterns, and social organization. Within this framework, the distinctions between corpus-based and corpus-driven approaches are clarified, and the application of methods such as KWIC, collocation, frequency, thematic, and semantic network analyses in social research is explained in detail. The contributions of leading scholars—Sinclair, McEnery, Hardie, Stubbs, Teubert, Biber, and Wodak—have played a decisive role in transforming corpus linguistics from an intuition-based interpretive discipline into a data-driven and replicable science. The paper also illustrates how corpus methods are applied in analyzing media discourse, gender representation, political communication, education, and religious texts. Moreover, critical discussions are presented on limitations such as representativeness, excessive quantification, semantic ambiguity, data ethics, and technological bias. The study emphasizes that corpus linguistics is not only a technical method but also an epistemological framework that requires ethical responsibility and cultural sensitivity. The findings suggest that corpus linguistics institutionalizes data-driven reasoning within the social sciences, enhances the measurability of meaning, and transforms language into an objective representation of social knowledge. Ultimately, corpus linguistics emerges as an integrative paradigm that bridges qualitative and quantitative methods, revitalizes cultural memory in the digital sphere, and redefines the epistemic foundation of contemporary research. As a result, the discipline provides not merely a methodological tool but a holistic and ethical approach for understanding the interaction between language, society, and knowledge in the digital humanities era.

Keywords: Corpus Linguistics, Social Sciences, Discourse Analysis, Digital Humanities, Epistemological Transformation

Sosyal Bilimlerde Derlem Dilbilimi: Yöntem, Kuram ve Epistemolojik Dönüşüm

Süreç

Geliş: 03/11/2025
Kabul: 02/12/2025

ÖZ

Bu makale, derlem dilbiliminin sosyal bilimlerdeki kuramsal, yöntemsel ve epistemolojik dönüşüm gücünü kapsamlı biçimde ele almaktadır. Derlem dilbilimi, dilin yalnızca yapısal yönünü değil, aynı zamanda toplumsal, kültürel ve ideolojik boyutlarını da inceleyerek, dili insan davranışının, düşünce biçimlerinin ve toplumsal düzenin ampirik bir kaydı hâline getiren bilimsel bir yaklaşım olarak tanımlar. Bu bağlamda, corpus-based ve corpus-driven yaklaşımların temel farkları açıklanmış; KWIC, eşdizimlilik, sıklık, tematik ve anlamsal ağ analizleri gibi yöntemlerin sosyal bilimler alanında nasıl uygulandığı ayrıntılı biçimde gösterilmiştir. Sinclair, McEnery, Hardie, Stubbs, Teubert, Biber ve Wodak gibi öncü araştırmacıların katkıları, derlem yöntemlerinin sezgisel ve yoruma dayalı analizlerden ampirik ve tekrarlanabilir veri analizlerine geçişte oynadığı belirleyici rolü ortaya koymuştur. Makalede ayrıca medya söylemi, toplumsal cinsiyet temsilleri, siyasal iletişim, eğitim ve dinî metin incelemelerinde derlem yöntemlerinin nasıl kullanıldığı örneklerle açıklanmıştır. Bunun yanında temsiliyet sınırlılığı, aşırı nicelcilik, anlamsal belirsizlik, veri etiği ve teknolojik önyargı gibi konular eleştirel bir bakış açısıyla tartışılmış; derlem dilbiliminin yalnızca teknik bir yöntem değil, etik sorumluluk ve kültürel farkındalık gerektiren bir epistemolojik çerçeve sunduğu vurgulanmıştır. Sonuç olarak çalışma, derlem dilbiliminin sosyal bilimlerde veri temelli düşünme biçimini kurumsallaştırdığını, anlamın ölçülebilirliğini artırdığını ve dili toplumsal bilginin nesnel bir temsiline dönüştürdüğünü ileri sürmektedir. Böylece derlem dilbilimi, dijital beşerî bilimler çağında nitel ve nicel yöntemleri birleştiren, kültürel belleği dijital düzleme taşıyan ve bilimsel araştırmada yeni bir paradigma inşa eden bütüncül bir yaklaşım olarak değerlendirilmektedir.

Anahtar Kelimeler: Derlem Dilbilimi, Sosyal Bilimler, Söylem Analizi, Dijital Beşerî Bilimler, Epistemolojik Dönüşüm

Copyright

This work is licensed under Creative Commons Attribution 4.0 International License

[a] *mustafakayapinar@yahoo.com* *https://orcid.org/0000-0001-9547-277X*

How to Cite: Kayapınar M (2025) Corpus linguistics in the social sciences: Method, theory, and epistemological transformation, International Journal of Current Social Science, 4(2): 51-60

Giriş

Konunun Sınırları

Derlem dilbilimi, yalnızca bir dil çözümleme tekniği değil, aynı zamanda modern sosyal bilimlerde veriye dayalı düşünme biçiminin somut bir yansımasıdır. Dilin sezgisel, normatif ya da spekülatif tanımlar yerine ampirik veriler üzerinden açıklanması gerektiği düşüncesi, 20. yüzyılın ortalarından itibaren bilim felsefesinde köklü bir dönüşümün parçası olmuştur (Kuhn, 1970, s. 10). Bu dönüşüm, dilin artık "yapısal bir sistem" değil, "toplumsal bir eylem biçimi" olarak ele alınmasını mümkün kılmıştır.

Derlem dilbilimi bu yönüyle, sosyal bilimlerdeki nicel ve nitel yöntem ayrımını aşan bir araştırma paradigması hâline gelmiştir. Sinclair'in öncülüğünde geliştirilen derlem yaklaşımı, dilin kendi içsel sezgisinden ziyade, "gerçek kullanım örnekleri" üzerinden anlam kazanması gerektiğini savunur (Sinclair, 1991, s. 12). Bu bakış, yalnızca dilbilimsel analizi değil, toplumsal ilişkilerin, kimliklerin ve ideolojilerin dilde nasıl inşa edildiğini de araştırma kapsamına dâhil eder.

Bu makale, derlem dilbiliminin tarihsel gelişimini, yöntemsel çeşitliliğini ve sosyal bilimlerle kesişim alanlarını incelemekte; böylece dilin toplumsal bağlamını veri temelli bir yaklaşımla açıklamayı amaçlamaktadır.

Çalışmanın Amacı ve Önemi

Bu çalışmanın temel amacı, derlem dilbiliminin yalnızca teknik bir araç değil, aynı zamanda epistemolojik bir paradigma olduğunu ortaya koymaktır. Modern sosyal bilimlerde dil, artık bireysel ifade aracı olmanın ötesinde, bilgi üretiminin ve güç ilişkilerinin merkezinde yer almaktadır (Fairclough, 1989, s. 23). Bu nedenle derlem dilbilimi, dilin toplumsal üretim süreçlerini ölçülebilir kılarak ideoloji, temsil ve kimlik gibi kavramların incelenmesinde yeni ufuklar açmıştır (McEnery & Hardie, 2012, s. 8).

Ayrıca derlem dilbilimi, sosyal bilimlerdeki "nesnellik" tartışmasına da yeni bir boyut kazandırmıştır. Çünkü bu yöntem, veriye dayalı olmasına rağmen bağlamdan bağımsız değildir; aksine, toplumsal pratikleri sayısal veriler aracılığıyla yeniden tanımlar (Teubert, 2005, s. 123). Bu yönüyle derlem dilbilimi hem nicel kesinliği hem de nitel yorumu birleştiren karma bir yöntemsel model sunar.

Araştırma Soruları

Bu çalışma, aşağıdaki temel sorular etrafında yapılandırılmıştır:

- Derlem dilbilimi tarihsel olarak nasıl bir gelişim izlemiştir ve bu gelişim sosyal bilimlerdeki paradigma değişimleriyle nasıl ilişkilendirilebilir?
- Derlem yöntemleri (örneğin KWIC, eşdizimlilik, sıklık analizi) toplumsal anlam üretimini çözümlemede hangi düzeyde etkilidir?
- Derlem dilbiliminin ideoloji, temsil, kimlik ve güç ilişkilerini analiz etmedeki katkıları nelerdir?
- Bu yöntemin epistemolojik ve etik sınırlılıkları nelerdir?

Yöntem

Bu makale, nitel ve nicel yaklaşımı bütünleştiren bir derlem tabanlı söylem çözümlemesi metodolojisine dayanır. Literatürdeki klasik kaynaklar (Sinclair, Stubbs, McEnery, Hardie, Tognini-Bonelli, Fairclough, Baker, Wodak, Teubert) ayrıntılı olarak incelenmiş; farklı dönemlerde derlem yaklaşımının sosyal bilimlere nasıl entegre edildiği analiz edilmiştir (Tognini-Bonelli, 2001, s. 3).

Ayrıca çalışmada tümevarımsal ve tümdengelimsel analiz teknikleri birleştirilmiştir:

Tümdengelimsel kısımda, derlem dilbiliminin teorik çerçevesi tarihsel bağlamda ortaya konulmuştur.

Tümevarımsal kısımda ise, bu yöntemin toplumsal söylem analizinde sunduğu ampirik bulgular değerlendirilmiştir.

Yöntemsel olarak makale, üç temel eksende ilerler:

- Kuramsal zemin (derlem ve sosyal bilim ilişkisinin teorik çerçevesi),
- Yöntemsel araçlar (KWIC, eşdizimlilik, sıklık, tematik analiz vb.),
- Sosyal uygulama alanları (medya, siyaset, dinî söylem, toplumsal cinsiyet, eğitim).

Tarihsel Gelişim

Derlem dilbilimi, 20. yüzyılın ortalarından itibaren dilbilimde yaşanan ampirik dönüşümün en belirgin sonucu olarak ortaya çıkmıştır. Dönemin hâkim yaklaşımı olan Chomskyci dilbilim, dilsel yetiyi (competence) merkezine alarak "soyut ideal konuşucu" modeline yönelmişti (Chomsky, 1957, s. 13). Bu yaklaşım, dilin toplumsal yönünü ikinci plana iterken, "doğal veri"yi güvenilmez görmüştür. Oysa 1950'lerin sonunda bazı araştırmacılar, dilin yapay biçimde kurgulanmış cümlelerden değil, gerçek kullanım örneklerinden incelenmesi gerektiğini savunmaya başladılar. Bu araştırmacılar arasında özellikle John Firth'in bağlam merkezli yaklaşımı, M. A. K. Halliday'in işlevsel dil kuramı ve John Sinclair'in derlem tabanlı gözlem ilkeleri öne çıkmıştır (Firth, 1957; Halliday, 1991; Sinclair, 1991). Böylece derlem dilbilimi, doğal veriye dayalı incelemenin dil araştırmalarında yeni bir epistemolojik zemin oluşturduğu bu kırılmadan doğmuştur.

Erken Dönem: Brown Derlemi ve Bilgisayar Devrimi

Derlem temelli çalışmaların kurumsal başlangıcı, 1960'ların başında Brown University'de gerçekleştirilen Brown Corpus projesidir. Henry Kučera ve W. Nelson Francis tarafından hazırlanan bu derlem, 1 milyon kelimelik Amerikan İngilizcesi örneklerini dijital ortama aktararak dilin istatistiksel incelenmesine öncülük etmiştir (Francis & Kučera, 1964, s. 4). Bu proje, dilin artık "ölçülebilir bir sistem" olarak ele alınabileceğini göstermiştir.

Kučera ve Francis, dilin istatistiksel dağılımlarının, gramer kurallarından daha güvenilir bir dil betimlemesi sunduğunu belirtmişlerdir (Kučera & Francis, 1967, s. 9). Bu yaklaşım, Noam Chomsky'nin sezgisel veriye dayalı analizlerine doğrudan bir alternatif oluşturmuştur. Brown Corpus'un başarısı, kısa sürede Lancaster-Oslo/Bergen (LOB) Corpus, London-Lund Corpus ve Kolhapur Corpus gibi uluslararası derlem projelerinin önünü açmıştır (Leech, 1979, s. 7).

Bu dönemde bilgisayar teknolojisinin gelişmesi, metinlerin dijital ortama aktarılmasını kolaylaştırmış ve otomatik sözcük sayımı, morfolojik analiz gibi işlemleri mümkün kılmıştır. Böylece dilbilimsel araştırma ilk kez veri bilimiyle buluşmuştur (Biber, Conrad & Reppen, 1998, s. 13).

Sinclair Dönemi ve COBUILD Devrimi

1980'lere gelindiğinde, John Sinclair öncülüğünde Birmingham Üniversitesi'nde kurulan COBUILD (Collins Birmingham University International Language Database) projesi, derlem dilbilimini kurumsal bir araştırma alanı hâline getirmiştir (Sinclair, 1991, s. 1). COBUILD, sadece bir derlem değil, aynı zamanda gerçek dil verisine dayalı bir sözlükleme ve öğretim hareketiydi. Sinclair, "dil kullanımının gerçekliği"ni merkeze alarak, grameri sezgisel normlardan kurtarıp ampirik gözleme dayandırdı (Sinclair, 2004, s. 11).

Sinclair'in "*idiom principle*" kavramı, kelimelerin anlamının sadece tekil biçimlerinden değil, dilsel kalıplaşmalar ve eşdizimliliklerden doğduğunu savunur (Sinclair, 2004, s. 45). Bu anlayış, sözcüklerin bağımsız değil, bağlam içinde anlam kazandığı düşüncesini öne çıkarır. Bu hem pragmatik hem de sosyal bilimsel bir bakıştı. Çünkü dil, Sinclair'e göre, toplumsal pratiklerin tekrar yoluyla sabitlenen bir yansımasıydı (Sinclair, 2004, s. 45).

COBUILD projesi, aynı zamanda Collins COBUILD English Dictionary (1987) gibi modern sözlüklerin doğmasına zemin hazırladı ve "derlem tabanlı sözlükçülük" akımını başlattı (Stubbs, 1997, s. 30). Böylece dil, sezgisel yargılarla değil, toplumsal kullanımla tanımlanır hâle geldi.

Hallidayci Sosyal Dönüşüm

1980'lerin sonunda M.A.K. Halliday'in sistemik-fonksiyonel dilbilim yaklaşımı, derlem çalışmalarını toplumsal bağlama taşımıştır. Halliday, dilin yalnızca bir iletişim aracı değil, sosyal bir eylem biçimi olduğunu vurgular (Halliday, 1978, s. 108). Bu anlayış, derlem verilerinin "dilsel kalıplar"ın ötesinde "toplumsal anlam üretimi"ni yansıttığı düşüncesini güçlendirmiştir.

Halliday'in öğrencilerinden Hasan, metin türü, bağlam ve toplumsal anlam arasındaki ilişkiyi açıklayarak "kayıt (register)" kavramını derlem çalışmalarına entegre etmiştir (Hasan, 1985, s. 65). Bu gelişme, özellikle eğitim, medya ve kültürel analizlerde derlem kullanımını yaygınlaştırmıştır.

Bu dönemde Stubbs'un *Text and Corpus Analysis* (1997) adlı eseri, derlem yöntemlerinin kültürel çalışmalar ve eleştirel söylem çözümlemesiyle nasıl birleşebileceğini göstermiştir (Stubbs, 1997, s. 89). Stubbs, dilin ideolojik boyutunu istatistiksel kalıplar üzerinden okumanın, sosyal bilimlerde "nicel kanıt temelli eleştiri" dönemi başlattığını ileri sürer (Stubbs, 1997, s. 102).

Eleştirel Söylem ve Sosyal Bilimlerle Kesişim

1990'lardan itibaren derlem dilbilimi, eleştirel söylem çözümlemesi (CDA) ile birleşmiştir. Norman Fairclough, Ruth Wodak ve Teun van Dijk gibi araştırmacılar, dilin toplumsal güç ilişkilerini inşa etme biçimlerini incelemişlerdir (Fairclough, 1995, s. 23). Ancak bu çalışmalar, başlangıçta büyük ölçüde nitel gözleme dayalıydı. Paul Baker ve Alan Partington gibi isimler, derlem yöntemlerini CDA'ya entegre ederek bu eksikliği kapattılar (Baker, 2006, s. 34).

Ortaya çıkan yeni alan, Corpus-Assisted Discourse Studies (CADS) olarak adlandırıldı (Baker, 2006, s. 47). CADS, ideolojik söylemlerin frekans, eşdizimlilik ve bağlamsal dağılım analizleriyle desteklenmesini öngörüyordu. Örneğin Baker'ın İngiliz basınında göçmen temsilleri üzerine çalışması, kelime sıklıkları ve eşdizimlilik modelleri üzerinden "öteki" kavramının nasıl inşa edildiğini göstermiştir (Baker, 2006, s. 47).

Wodak'ın Discourse-Historical Approach modeli ise tarihsel bağlamın, derlem verileriyle birlikte ele alınması gerektiğini savundu (Wodak & Meyer, 2016, s. 21). Böylece derlem dilbilimi, artık sadece dilin formunu değil, toplumun ideolojik haritasını da inceleyen bir araç hâline geldi.

Dijital Dönem ve Büyük Veri Çağı

2000'lerden itibaren dijitalleşme, derlem dilbilimini "büyük veri bilimi"yle buluşturdu. Web tabanlı derlemler (örneğin Corpus of Contemporary American English [COCA], British National Corpus [BNC] ve Sketch Engine platformu) milyarlarca kelimeye ulaşarak araştırma ölçeğini radikal biçimde genişletti (Davies, 2009a, s. 3).

McEnery ve Hardie'ye göre bu dönem, derlem dilbiliminin "mikro ölçekte istatistik, makro ölçekte kültürel analiz" çağıdır (McEnery & Hardie, 2012, s. 44). Derlem araçları, toplumsal değişimi uzun vadeli veri trendleri üzerinden izlemeyi mümkün kılar. Örneğin, Google Ngrams veya NOW Corpus verileri, kelime sıklıklarının on yıllar boyunca nasıl değiştiğini istatistiksel olarak gösterebilir (Michel et al., 2011, s. 176).

Bu teknolojik gelişmeler, derlem dilbilimini yalnızca dilbilim değil, sosyoloji, medya çalışmaları, siyaset bilimi, kültürel antropoloji ve hatta teoloji gibi alanlarla da buluşturmuştur (Teubert, 2005, s. 7). Özellikle Arapça, Çince, Türkçe gibi yapısal olarak karmaşık diller için geliştirilen yeni etiketleme sistemleri, derlem temelli kültürel karşılaştırma çalışmalarını güçlendirmiştir (Habash, 2010, s. 54).

Dijital Beşerî Bilimler Çağında Derlem

Günümüzde derlem dilbilimi, "dijital beşerî bilimler (*Digital Humanities*)" hareketinin temel bileşenlerinden biridir. Bu alanda amaç, büyük metin arşivlerini dijital araçlarla analiz ederek tarihsel, kültürel ve toplumsal eğilimleri ortaya koymaktır. Teubert'in ifadesiyle, "dilsel veri, artık yalnızca dilin değil, toplumsal hafızanın da kaydıdır" (Teubert, 2005, s. 15).

Baker ve Egbert'in çalışmaları, derlem yöntemlerinin sosyal medya, çevrim içi forumlar ve bloglardaki söylemleri analiz etmekte nasıl kullanılabileceğini göstermiştir (Baker & Egbert, 2019, s. 9). Böylece derlem, çağdaş toplumların dijital kimliklerinin de haritasını çıkaran bir araştırma aracına dönüşmüştür.

Bu dönüşüm, epistemolojik olarak "sezgiden veriye, bireyden topluma" yönelmiş yeni bir bilim anlayışını temsil eder. Artık dil, yalnızca bir ifade sistemi değil; veriyle ölçülebilen toplumsal bir olgudur.

Yöntemler ve Yaklaşımlar

Derlem dilbilimi, dilin işleyişini açıklamak için sezgilere değil, ampirik gözleme ve nicel veriye dayanır. Bu özellik, onu sosyal bilimlerdeki diğer araştırma yöntemlerinden ayırır; çünkü derlem analizi, dilin yalnızca biçimsel yönünü değil, aynı zamanda toplumsal üretim biçimlerini de ölçülebilir hâle getirir (Baker & McEnery, 2015, s. 89). Sinclair'in ifadesiyle, "derlem, dilin konuşulduğu biçimdeki doğal halinin bilimsel temsildir" (Sinclair, 2004, s. 22). Bu yaklaşım, dili sadece bir sistem değil, toplumun düşünme biçimlerini kodlayan bir yapı olarak görür.

Corpus-Based ve Corpus-Driven Ayrımı

Derlem çalışmalarında iki temel yönelim bulunur: Corpus-based (derlem destekli) ve corpus-driven (derlem güdümlü) yaklaşımlar. Corpus-based yaklaşım, önceden belirlenmiş kuramsal hipotezleri test etmek için derlem verisini kullanır; yani teori öncedir, veri onun doğrulayıcısıdır (Fairclough, 1995, s. 2). Örneğin, Fairclough'un medya söylemindeki güç ilişkilerine dair varsayımları, sözcük sıklıkları ve eşdizimlilik modelleri üzerinden test edilmiştir (Fairclough, 1995, s. 2).

Buna karşılık corpus-driven yaklaşım, hipotezleri veriden türetir. Sinclair'e göre, araştırmacının "dil verisine müdahale etmeden, onun kendi örüntülerini keşfetmesine izin vermesi" gerekir (Sinclair, 2004, s. 35). Bu yöntem, bilimsel önyargıyı minimize eder. Stubbs ise bu yaklaşımın "veri temelli sezgi" ürettiğini, böylece dilin kendisinin teoriye yön verdiğini belirtir (Stubbs, 1997, s. 101).

Her iki yönelim de sosyal bilimlerde tamamlayıcıdır. Örneğin, göçmen söylemini inceleyen bir araştırmada corpus-based yaklaşım, belirli kavramların (ör. "mülteci", "kaçak", "tehdit") sıklığını ölçerken; corpus-driven yaklaşım, bu kelimelerin çevresinde oluşan yeni anlam kümelerini ortaya çıkarabilir (Johns, 1991, s. 27).

KWIC (Key Word in Context)

KWIC (Bağlam İçinde Anahtar Kelime) yöntemi, derlem analizinin en temel araçlarından biridir. Bu teknik, belirli bir kelimenin geçtiği tüm bağlamları yan yana getirerek, kelimenin işlevsel ve anlamsal çeşitliliğini ortaya koyar (Teubert, 2005, s. 128). Sinclair, KWIC dizilimlerinin "dilsel davranışın mikroskobik düzeyde gözlemlenmesi" anlamına geldiğini vurgular (Sinclair, 2004, s. 45). Sosyal bilimler açısından bu yöntem, sözcüklerin ideolojik çağrışımlarını ve söylemsel pozisyonlarını görünür kılar.

Örneğin, medya derleminde "kadın" kelimesinin bağlamları incelendiğinde, "mağdur", "güzel", "duygusal" gibi sıfatlarla yüksek eşdizimlilik gösterdiği saptanmıştır (Leech, 1981, s. 21). Buna karşılık "erkek" kelimesi daha çok "lider", "başarılı" gibi güç çağrışımlı fiillerle birlikte kullanılmıştır. Bu tür KWIC analizleri, toplumsal cinsiyet rollerinin dilde nasıl kurulduğunu nesnel biçimde gösterir (Leech, 1981, s. 21).

Eşdizimlilik (Collocation) ve Sıklık Analizi

Eşdizimlilik *(collocation*), kelimelerin birlikte görülme eğilimini ölçen istatistiksel bir tekniktir. Firth'in ünlü ifadesiyle, "bir kelime, birlikte bulunduğu kelimelerle tanımlanır" (Firth, 1957, s. 11). Bu yöntem, toplumsal ideolojilerin dilde nasıl kalıplaştığını anlamada büyük önem taşır. Stubbs, "eşdizimlilik analizi, ideolojinin istatistiksel yüzüdür" der(Stubbs, 1997, s. 72).

Eşdizimlilik analizi genellikle Mutual Information (MI), T-score veya Log-likelihood gibi istatistiksel ölçütlerle yapılır (Mauranen, 2009, s. 23). Örneğin, "terör" kelimesiyle yüksek MI değerine sahip olan "İslam", "radikal", "şiddet" gibi kelimeler, belirli bir medya ideolojisinin kalıplaşmış örüntüsünü ortaya çıkarabilir (Mauranen, 2009, s. 23). Bu yöntem, eleştirel söylem analizinde (CDA) kullanılan nitel çözümlemelere nicel doğrulama sağlar (Baker & Egbert, 2019, s. 44).

Sıklık (frequency) analizi ise, belirli sözcüklerin veya yapısal kalıpların derlem genelinde ne sıklıkta geçtiğini belirler. McEnery ve Hardie'ye göre "sıklık verisi, toplumsal öncelikleri yansıtır; çünkü toplumun neyi sık konuştuğu, neyi düşündüğünün göstergesidir" (McEnery & Wilson, 1996, s. 15).

Sıklık analizleri zaman serisiyle birleştirildiğinde, ideolojik dönüşümleri gösterebilir. Örneğin, "özgürlük" kelimesinin 1980 sonrası siyasi söylemlerde artışı, neoliberal dilin yükselişiyle paralellik gösterir (Biber, 1988, s. 4).

Tematik ve Anlamsal Analiz

Gelişmiş derlem analizlerinde artık yalnızca sözcük düzeyinde değil, tema ve anlam ağı düzeyinde de modelleme yapılmaktadır. Tematik analiz, büyük metin derlemlerindeki kelime kümelerini istatistiksel olarak gruplayarak, metinlerin hangi sosyal temalar etrafında örgütlendiğini gösterir (Michel et al., 2011, s. 176). Örneğin, sosyal medya derlemlerinde "aile", "ahlak", "kadın" gibi temalar birlikte kümelenerek, toplumun değer yönelimlerini nicel biçimde yansıtır.

Anlamsal ağ çözümlemesi (*semantic network analysis*) ise, kelimeler arasındaki ilişkileri bir ağ yapısı içinde görselleştirir. Bu yöntem, derlem verilerini grafiksel biçimde analiz ederek ideolojik kümelenmeleri ortaya çıkarır (Teubert, 2005, s. 131). Biber, bu tekniklerin "nitel anlamı nicel yapıya dönüştürme" kabiliyetiyle sosyal bilimlerde yeni bir paradigma sunduğunu ifade eder (Biber & Conrad, 2009, s. 91).

Bu yaklaşımlar, dijital beşerî bilimlerde özellikle kültürel analiz, edebî metin incelemesi ve tarihsel söylem çalışmalarında kullanılmaktadır (Teubert, 2005, s. 131).

Derlem Destekli Söylem Analizi

Derlem dilbiliminin sosyal bilimlerle kesiştiği en güçlü alan Corpus-Assisted Discourse Studies (CADS) olarak adlandırılır. Baker, CADS yöntemini "dilsel tekrarlardan toplumsal kalıplara giden bir köprü" olarak tanımlar (Baker, 2006, s. 33). Bu yaklaşımda nitel söylem çözümlemesi, nicel derlem istatistikleriyle desteklenir. Fairclough'un "güç söylemi" modelinde, belirli ideolojik kategorilerin dilde nasıl sistematik biçimde yerleştiği derlem araçlarıyla kanıtlanabilir (Fairclough, 1995, s. 20).

Partington, derlem destekli söylem çözümlemesini üç aşamada tanımlar: (1) veri çıkarımı, (2) istatistiksel örüntüleme, (3) nitel yorum (Partington, Duguid & Taylor, 2013, s. 6). Bu çok katmanlı yapı, ideolojik çözümlemeyi hem ampirik hem eleştirel kılar.

Örneğin Wodak'ın Discourse-Historical Approach modeli, tarihsel bağlamı derlem verisiyle birleştirerek, politik söylemlerde "biz"–"onlar" karşıtlığının tarihsel sürekliliğini gösterir (Wodak & Reisigl, 2001, s. 12). Böylece derlem yöntemleri, toplumsal kimlik ve ideoloji çalışmalarına ölçülebilirlik kazandırmıştır.

Etik ve Epistemolojik Boyut

Derlem yöntemlerinin gücü, aynı zamanda eleştiriye açık yanıdır. Çünkü veri seçimi, her zaman belirli bir ideolojik tercihi içerir. Teubert'in belirttiği gibi, "hiçbir derlem masum değildir; her derlem bir bakış açısının ürünüdür" (Teubert, 2005, s. 18). Bu nedenle çağdaş derlem çalışmalarında, veri etiği, temsiliyet ve kültürel bağlam konuları giderek önem kazanmıştır (Wodak & Meyer, 2016, s. 27).

Wodak ve Meyer, derlem verisinin sosyal bağlamdan koparılarak istatistiksel olarak yorumlanmasının, anlamın aşırı indirgenmesine yol açabileceğini vurgular (Wodak & Meyer, 2016, s. 27). Dolayısıyla etik yaklaşım, veriyi yalnızca nesne olarak değil, toplumsal eylemin kaydı olarak ele almayı gerektirir.

Yöntemsel Bütünlük

Sonuç olarak derlem dilbilimi, nicel ve nitel metodolojilerin bütünleştirildiği hibrit bir araştırma paradigmasıdır. McEnery ve Wilson'un ifadesiyle, "derlem hem bilimsel kesinlik hem de insanî yorumu aynı anda mümkün kılar" (McEnery & Wilson, 1996, s. 15). Bu nedenle derlem dilbilimi, sosyal bilimlerde yalnızca bir araç değil, düşünme biçimi hâline gelmiştir.

Bu metodoloji, dilin yüzeyinde görünen yapıları değil, toplumun bilinçaltı söylemsel kalıplarını açığa çıkarır. Sinclair'in de belirttiği gibi, "derlem, kelimelerin değil, toplumların nasıl düşündüğünün kaydıdır" (Sinclair, 2004, s. 42).

Katkılar ve Uygulama Alanları

Derlem dilbilimi, yalnızca dilbilimsel yapıların çözümlemesinde değil; sosyal bilimlerdeki çok sayıda alt disiplinin araştırma yöntemini dönüştürme gücüne sahip bir yaklaşımdır. Bu dönüşüm, dilin toplumsal üretim biçimlerini nesnel olarak ölçme imkânı sunmasıyla ilgilidir. McEnery ve Hardie'nin ifadesiyle "derlem, toplumun dilde bıraktığı ölçülebilir izlerin bütünü"dür (McEnery & Hardie, 2012, s. 53). Derlem tabanlı yöntemler, dilin ideolojik, kültürel ve politik katmanlarını istatistiksel modellerle çözümleme olanağı sağlayarak, sosyal bilimlerde veriye dayalı analiz paradigmasının öncüsü olmuştur. Bu yaklaşımı özellikle eleştirel söylem çözümlemesi (Fairclough, 1995; Wodak & Meyer, 2016), toplumsal değişim incelemeleri (Baker, 2006; Baker, Gabrielatos & McEnery, 2013) ve kültürel söylem araştırmaları (Stubbs, 1997) kapsamında güçlendiren çalışmalar dikkat çekici şekilde ortaya koymuştur.

Medya Söylemi ve İdeolojik Temsiller

Medya söylemi, derlem yöntemlerinin en etkin uygulama alanlarından biridir. Fairclough'a göre medya dili, toplumsal ideolojilerin hem üretildiği hem de meşrulaştırıldığı başlıca araçtır (Fairclough, 1995, s. 23). Derlem analizleri, bu ideolojilerin dilsel yüzünü ortaya koymak için büyük veri desteği sağlar. Örneğin Paul Baker'ın *Using Corpora in Discourse Analysis* adlı çalışmasında İngiliz basınında "göçmen" kavramının bağlam analizleri yapılmış; *immigrant* kelimesinin sıklıkla "illegal", "problem", "burden" gibi sözcüklerle eşdizimlilik gösterdiği belirlenmiştir (Baker, 2006, s. 45). Bu durum, göçmenliğin "tehdit" olarak sunulan bir ideolojik kalıba oturduğunu kanıtlamıştır.

Stubbs, medya söylemindeki bu tür örüntülerin "gizli ideolojilerin istatistiksel görünümü" olduğunu söyler (Stubbs, 1997, s. 93). Dolayısıyla derlem analizi, yalnızca metinsel gözleme dayalı bir yorum değil; aynı zamanda ideolojik yeniden üretim mekanizmalarının matematiksel çözümlemesidir. Wodak'ın Discourse-Historical Approach yöntemi, bu bağlamda derlem verilerini tarihsel-toplumsal bağlamla birleştirerek medya dilinde süreklilik kazanan "biz–onlar" karşıtlıklarını ampirik olarak göstermiştir (Wodak & Reisigl, 2001, s. 12).

Toplumsal Cinsiyet, Kimlik ve Temsil

Toplumsal cinsiyet temsilleri, derlem dilbiliminin sosyal bilimlerdeki en verimli uygulama alanlarından biridir. Cameron'un belirttiği gibi, "dil, toplumsal cinsiyet rollerinin hem aynası hem aracıdır" (Cameron, 1992, s. 51). Bu bağlamda derlem temelli araştırmalar, kadın ve erkek temsillerinin dilde nasıl kalıplaştığını nicel verilerle ortaya koyar.

Baker ve McEnery'nin *Corpora and Discourse Studies* adlı eserinde, İngilizce basın derlemleri üzerinde yapılan analizlerde, "kadın" sözcüğünün daha çok özel alan (aile, duygusallık, güzellik) ile, "erkek" sözcüğünün ise kamusal alan (iş, güç, liderlik) ile ilişkilendirildiği saptanmıştır (Baker & McEnery, 2015, s. 89). Bu tür bulgular, toplumsal cinsiyet eşitsizliğinin dilin kendisinde nasıl doğal bir yapı gibi yeniden üretildiğini gösterir.

Stubbs, "derlem analizi, ideolojik doğallaştırmayı görünür kılarak sosyal eleştiriyi bilimsel zemine taşır" der (Stubbs, 1997, s. 72). Bu nedenle feminist dil çalışmaları, derlem yöntemleriyle desteklendiğinde yalnızca normatif bir söylem değil, istatistiksel doğrulanabilir bir olgu hâline gelir.

Eğitim ve Öğretim Alanı

Derlem dilbilimi, dil öğretimi ve eğitim sosyolojisi açısından da devrim niteliğinde katkılar sağlamıştır. Tim Johns'un geliştirdiği Data-Driven Learning (DDL) modeli, öğrencilerin dili öğretmenden değil, doğrudan derlem verisinden keşfetmesini öngörür (Johns, 1991, s. 27). Bu yaklaşım, dil öğretimini "keşif temelli öğrenme"ye dönüştürür.

McCarthy, derlem destekli öğretimin öğrencilerde "dil farkındalığını" (*language awareness*) güçlendirdiğini belirtir (McCarthy, 1991, s. 87). Çünkü öğrenici, sözcüklerin doğal bağlamlarda nasıl kullanıldığını bizzat gözlemler. Ayrıca derlem temelli ders materyalleri, dilin toplumsal ve kültürel yönlerini yansıtması bakımından geleneksel dilbilgisi kitaplarına kıyasla çok daha temsilîdir (McCarthy, 1991, s. 87).

Bu yöntem aynı zamanda eğitim sosyolojisinde temsil analizleri için de kullanılabilir. Ders kitaplarında kadın ve erkek karakterlerin konuşma oranlarının, meslek dağılımlarının veya eylem türlerinin derlem temelli olarak ölçülmesi, eğitimde örtük cinsiyet kalıplarını ortaya çıkarabilir (Nesi, 2000, s. 53).

Dinî Söylem ve Kültürel Derlemler

Derlem yöntemleri son yıllarda dinî metinlerin dijital incelemesinde de kullanılmaya başlanmıştır. Habash ve Dukes'un Quranic Arabic Corpus projesi, Kur'an'daki tüm kelimeleri morfolojik, sözdizimsel ve anlamsal düzeyde etiketleyerek derlem biçiminde erişime sunmuştur (Habash & Dukes, 2010, s. 123). Bu proje hem Arap dilbilimi hem de dijital beşeri bilimler açısından öncü bir modeldir.

Bu tür derlemler sayesinde, kutsal metinlerdeki tematik dağılımlar, kavramsal ağlar ve kök tabanlı anlam ilişkileri ölçülebilir hâle gelmiştir. Örneğin, "adalet" (*ʿadl*) kökünün hangi surelerde, hangi bağlamlarda tekrarlandığı ve hangi kavramlarla birlikte kullanıldığı nicel olarak incelenebilir (Habash & Dukes, 2010, s. 123). Böylece derlem dilbilimi, klasik filolojik incelemelere ampirik boyut kazandırır.

Benzer biçimde hadis, tefsir ve fıkıh literatürlerinin dijitalleştirilmesi, İslam araştırmalarında "dijital ilim" çağını başlatmıştır. Bu gelişme hem dilbilimsel hem de kültürel analizde metin verisinin yeniden keşfini mümkün kılar (Dukes, 2010a, s. 16).

Siyasal Söylem ve Kamu Alanı

Derlem dilbilimi, siyasal iletişim analizinde de güçlü bir yöntem hâline gelmiştir. Fairclough, siyasal söylemin "dilin ikna gücü üzerine kurulu bir toplumsal eylem biçimi" olduğunu belirtir (Fairclough, 1995, s. 20). Bu bağlamda siyasetçiler tarafından sık kullanılan kelimeler, temalar ve metaforlar derlem aracılığıyla çözümlenebilir.

Wodak ve Reisigl'in *Discourse and Discrimination* adlı eseri, Avrupa'daki sağ popülist söylemlerin kelime düzeyinde "ulusal değer", "güvenlik", "tehdit" gibi kavramlarla örüldüğünü göstermiştir (Wodak & Reisigl, 2001, s. 12). Bu kelimelerin sıklık ve eşdizimlilik analizleri, söylemin ideolojik çerçevesini nicel biçimde kanıtlar.

Ayrıca sosyal medya derlemleri üzerinden yürütülen çalışmalar, çevrim içi politik kutuplaşmanın dilsel izlerini ortaya koymuştur. Baker ve Egbert, #refugeeswelcome ve #stopimmigration etiketlerinin duygusal sözcük kümeleriyle birlikte analiz edilmesi sonucunda, empati ve korku söylemlerinin istatistiksel olarak kutuplaştığını göstermiştir (Baker & Egbert, 2019, s. 49).

Kültür, Mizah ve Duygusal Söylem

Derlem dilbiliminin bir diğer ilginç uygulama alanı kültürel söylem analizidir. Teubert'e göre "dilsel tekrar, kültürel hafızanın en güvenilir göstergesidir" (Teubert, 2005, s. 18). Bu nedenle ulusal marşlar, atasözleri, deyimler veya mizah metinleri gibi kültürel ürünler derlem yöntemiyle incelendiğinde, bir toplumun değer yapısı ve zihinsel çerçevesi çözümlenebilir.

Örneğin fıkralarda tekrar eden kelimeler, ulusal mizah anlayışının dilsel göstergeleri olarak değerlendirilmektedir (Raskin, 1985, s. 71).

Bu tür kültürel derlemler, dilin yalnızca iletişim değil, kimlik üretim aracı olduğunu kanıtlar. Böylece derlem dilbilimi, beşerî bilimlerle sosyal bilimleri birleştiren ortak bir paradigma oluşturur.

Dijital Beşerî Bilimler ve Yeni Ufuklar

21. Yüzyılda derlem dilbilimi, dijital beşerî bilimler (Digital Humanities) içinde yeniden konumlanmıştır. Bu alan, tarih, sosyoloji, dilbilim, felsefe ve bilişim teknolojilerini birleştiren disiplinler arası bir platformdur. Baker ve Egbert, "derlem yöntemleri dijital beşerî bilimlerin veri omurgasını oluşturur" der (Baker & Egbert, 2019, s. 9).

Günümüzde dijital derlemler, yalnızca metinleri değil; görsel, sesli, hatta çoklu ortam verilerini de kapsayacak şekilde genişlemektedir. Bu sayede kültürel etkileşim, dijital kimlik ve çevrim içi söylem gibi konular, derlem tabanlı yaklaşımla incelenebilir hâle gelmiştir (Teubert, 2005, s. 15).

Bu dönüşüm, sosyal bilimlerde "veri epistemolojisi" olarak adlandırılan yeni bir dönemin başlangıcını temsil eder. Artık dil, toplumsal deneyimin hem aracı hem verisidir.

Sınırlılıklar ve Eleştiriler

Derlem dilbilimi, sosyal bilimlerde ampirik yöntemin önemli bir dayanağı olmakla birlikte, bazı epistemolojik ve teknik sınırlılıklardan tamamen arınmış değildir. Bu sınırlılıklar hem yöntemsel hem de ideolojik düzeyde ortaya çıkar. Çünkü her derlem, seçilen verinin sınırları kadar anlamlıdır; seçilmeyen, dışarıda bırakılan her metin, sessiz bir önyargı üretir (McEnery & Wilson, 1996, s. 15). Bu nedenle McEnery ve Wilson'un ifadesiyle "derlem dilbilimi kesinlik değil, olasılık üretir" (McEnery & Wilson, 1996, s. 15).

Temsiliyet Sorunu

Her derlem, belirli bir veri seçim sürecinin sonucudur. Bu seçim süreci, çoğu zaman araştırmacının kültürel, politik ya da dilsel tercihlerine bağlı olarak şekillenir. Teubert, "hiçbir derlem masum değildir; derlemin oluşturulduğu bağlam, sonuçlarını belirler" diyerek bu sınırlılığı açıkça dile getirir (Teubert, 2005, s. 16). Örneğin İngilizce merkezli derlemler (BNC, COCA gibi) küresel İngilizce çeşitliliğini tam olarak temsil etmez. Afrika, Asya ya da Orta Doğu varyantlarının yeterince yer almaması, "küresel" olarak sunulan bir derlemin aslında Batı-merkezci bir seçimi yansıttığını gösterir (Biber, 1988, s. 4).

Bu sorun, Arapça derlemler için daha da belirgindir. Çünkü Modern Standart Arapça (MSA) dışındaki lehçelerin (Mısrî, Şâmî, Mağribî vb.) çoğu hâlen yeterli düzeyde dijitalleştirilmemiştir (Habash, 2010, s. 98). Dolayısıyla Arapça derlem araştırmaları, çoğu kez resmi, yazılı ve medyatik dil verilerine dayanmakta; halk söylemi, dini retorik ve mahallî sözlü gelenekler dışarıda kalmaktadır (Dukes, 2010b, s. 17).

Temsiliyet sorunu, yalnızca dilsel değil, aynı zamanda sosyolojik bir meseledir. Çünkü derlemler, "toplumun hangi kesimlerinin sesi duyulacak?" sorusuna fiilen cevap verir (McEnery, Baker & Hardie, 2012, s. 12).

Nicelcilik Eleştirisi

Derlem yöntemlerine yöneltilen en köklü eleştirilerden biri aşırı nicelciliktir (quantitative bias). Stubbs'un uyarısıyla, "sayılabilen her şey anlamlı değildir, anlamlı olan her şey de sayılabilir değildir" (Stubbs, 1997, s. 93). Derlem araştırmalarında sıklık ve eşdizimlilik gibi nicel göstergelere aşırı güven, bağlamın derin yapısını gölgede bırakabilir. Fairclough, derlem destekli eleştirel söylem analizinin "sayılara tutsak olmadan" yürütülmesi gerektiğini vurgular (Fairclough, 1995, s. 20). Çünkü söylem, yalnızca kelime sıklıklarından ibaret değildir; sessizlikler, imalar ve pragmatik stratejiler de ideolojik anlam üretir.

Bu eleştiri, özellikle medya ve siyaset derlemlerinde geçerlidir. Örneğin "özgürlük" kelimesinin artan sıklığı, gerçekten özgürlük bilincinin arttığını değil, bu kavramın söylemde araçsallaştırıldığını gösterebilir (Fairclough, 1995, s. 28). Dolayısıyla derlem verilerinin yorumu, daima nitel analizle dengelenmelidir.

Anlamsal Belirsizlik ve Çokanlamlılık

Her dil, çokanlamlılık (polysemy) ve bağlamsal çeşitlilik içerir. Bu özellik, otomatik etiketleme ve lemmatizasyon süreçlerinde hatalara yol açar. Sinclair'in belirttiği gibi, "kelimeler istatistiksel olarak değil, bağlamsal olarak anlam kazanır" (Sinclair, 2004, s. 42). Arapça gibi kök-temelli dillerde bu sorun daha karmaşık hâle gelir. Çünkü aynı kökten türeyen kelimeler, bağlama göre farklı anlam alanlarına taşınabilir. Örneğin "ʿilm" kökü hem "bilgi" hem "ilim" hem "bilim" anlamlarında kullanılır; bu tarihsel anlam genişlemesi klasik sözlüklerde ayrıntılı biçimde açıklanmıştır (İbn Manzur, 1999, c. 12, s. 417; Firuzabadi, 2011, s. 1025). Derlem bu farkı istatistiksel olarak ayırt edemez (Habash, 2010, s. 105). Bu bağlamsal bulanıklık, özellikle semantik ağ ve anlamsal alan analizlerinde yanlış genellemelere neden olabilir (Dukes, 2010a, s. 16). Bu nedenle anlam çözümlemelerinde, insan denetimi hâlâ vazgeçilmezdir.

Teknik Kısıtlar ve Dilsel Önyargılar

Derlem analiz araçları (Sketch Engine, AntConc, WordSmith vb.) genellikle Latin alfabesi ve Avrupa dillerinin morfolojisine göre tasarlanmıştır (Habash & Rambow, 2005, s. 41). Bu durum, Arapça gibi sağdan sola yazılan, eklemeli olmayan ama türevsel zenginliğe sahip dillerde teknik kısıtlar doğurur.

Örneğin, Buckwalter morfolojik çözümleyicisi Arapça kökleri ayırmada büyük kolaylık sağlasa da, bazen "lam–elif" birleşimleri veya diakritik farklılıklar nedeniyle kelimeleri yanlış sınıflandırır (Habash & Rambow, 2005, s. 41).

Ayrıca otomatik etiketleme sistemleri, ideolojik önyargıları da yansıtabilir. Belirli kelimelerin "olumlu" ya da "olumsuz" olarak işaretlenmesi, yazılımın eğitim aldığı veri kümesinin kültürel kodlarına bağlıdır (Mauranen, 2009, s. 23). Dolayısıyla teknoloji, görünmez bir ideolojik filtre işlevi görebilir.

Epistemolojik Eleştiriler

Teubert, derlem dilbiliminin yalnızca "gözlenen metinlerle sınırlı" olduğunu, gözlenemeyen anlam potansiyellerini ihmal ettiğini söyler (Teubert, 2005, s. 16). Bu, epistemolojik olarak pozitivist bir sınırdır. Çünkü dil, yalnızca görünen istatistiksel kalıplardan değil, aynı zamanda gizli bilişsel kategorilerden oluşur.

Tognini-Bonelli, bu eleştiriyi kabul etmekle birlikte, "derlem, gözlenenin bilimi olarak değerli, fakat gözlemlenemeyenin bilimi değildir" der (Tognini-Bonelli, 2001, s. 12). Yani derlem analizi, toplumsal dilsel davranışların yüzey yapısını çözümleyebilir ama derin bilinç düzeyindeki anlam üretim süreçlerine doğrudan erişemez.

Bu nedenle bazı araştırmacılar, derlem dilbilimini "veri pozitivizmi" (*data positivism*) olarak eleştirir (Teubert, 2005, s. 18). Onlara göre, derlem yalnızca mevcut söylemi ölçer; alternatif, bastırılmış veya görünmez söylemleri tespit edemez.

Dinî ve Kültürel Metinlerde Aşırı Yorum Riski

Dinî metinlerin derlem tabanlı incelenmesinde karşılaşılan bir diğer sınırlılık, metin kutsallığı ile veri nesnelliği arasındaki gerilimdir. Habash, Kur'an dilinin "gramatik bir sistem olmanın ötesinde, teolojik bir bağlama sahip" olduğunu hatırlatır (Habash, 2010, s. 98). Bu nedenle Kur'an ayetlerini yalnızca morfolojik veya sözdizimsel yapı olarak ele almak, metnin hermenötik derinliğini yitirmeye yol açabilir.

Ayrıca kültürel ve tarihsel bağlamdan kopuk derlem analizleri, dini kavramların anlam alanlarını yanlış temsil edebilir. Örneğin "rahmet" kökünün sadece sıklık değeri üzerinden incelenmesi, onun teolojik ve duygusal boyutunu ihmal eder (Dukes, 2010b, s. 17).

Bu nedenle dini metinlerde derlem yöntemi, filolojik ve tefsirsel okuma ile birlikte kullanılmalıdır. Aksi hâlde sonuç, nesnel değil, indirgemeci olur.

Etik Sorunlar

Veri toplama ve derlem oluşturma süreçlerinde etik meseleler, giderek daha fazla gündeme gelmektedir. Özellikle sosyal medya derlemleri, kullanıcıların rızası alınmadan toplanan verilerle oluşturulduğunda etik tartışmalara yol açar (Wodak, 2015, s. 43). Wodak, "veri her zaman bir insanın ifadesidir; bu nedenle her veri bir özneye aittir" der (Wodak, 2015, s. 43). Dolayısıyla anonimleştirme, izin ve bağlam ilkeleri, sosyal bilimlerde derlem kullanımının vazgeçilmez koşullarıdır.

Bazı araştırmalarda, politik veya dini görüşleri açık eden tweetlerin derleme dahil edilmesi, bireylerin mahremiyetini ihlal riski taşır (McEnery, Baker & Hardie, 2012, s. 12). Bu yüzden derlem araştırmalarında araştırma etiği (research ethics) protokollerinin yazılı olarak oluşturulması önerilmektedir (Dukes, 2010a, s. 16).

Çok Disiplinli Yaklaşımın Zorlukları

Derlem dilbilimi sosyal bilimlerle birleştiğinde, farklı disiplinlerin kavramsal çatışmaları da gündeme gelir. Sosyoloji "anlam"ı toplumsal yapı üzerinden tanımlarken, dilbilim "anlam"ı dilsel biçimlerden türetir. Bu iki yaklaşım arasında metodolojik köprü kurmak her zaman kolay değildir (Habash & Rambow, 2005, s. 41). Biber ve Conrad, derlem tabanlı araştırmalarda "disiplinler arası yorum farklılıklarının sonuçları değiştirebileceğini" belirtir (Biber & Conrad, 2009, s. 91).

Bu nedenle, derlem çalışmaları tek bir disipline ait olmaktan çıkıp, metodolojik diyalog gerektiren ortak zeminlere yönelmelidir. Aksi hâlde, teknik mükemmeliyet bilimsel bütünlüğün önüne geçer.

Yöntemsel Yeniden Değerlendirme

Bu eleştirilerin tümü, derlem dilbilimini değersizleştirmez; aksine, yöntemsel öz-eleştiri bilincini derinleştirir. Sinclair, "her veri, bir sessizliği de içinde taşır" diyerek, derlem çalışmalarının kendi kör noktalarını fark etmesi gerektiğini vurgular (Sinclair, 2004, s. 42). Dolayısıyla çağdaş derlem dilbilimi, artık salt veri yığını değil; etik, kültürel ve bilişsel boyutları içeren refleksif bir yöntem hâline gelmiştir.

Sonuç

Derlem dilbilimi, modern sosyal bilimlerin epistemolojik yöneliminde köklü bir dönüşüm yaratmıştır. Çünkü bu yaklaşım, dilin yalnızca iletişim aracı değil, aynı zamanda toplumsal bilgi üretim biçimi olduğunu ortaya koyar. McEnery ve Hardie'nin belirttiği gibi, "her derlem, bir toplumun düşünme biçiminin ampirik izdüşümüdür" (McEnery & Hardie, 2012, s. 121). Bu nedenle derlem dilbilimi, dilin yapısal analizinden çok daha fazlasını; insan davranışlarının, ideolojilerin ve kültürel temsillerin sistematik ölçümünü mümkün kılar.

Derlem temelli araştırmalar, sosyal bilimlerde sezgisel yaklaşımların yerine veri temelli düşünme biçimini yerleştirmiştir. Geleneksel sosyolojik yöntemlerde "gözlemci yorumu" ağırlıklı iken, derlem yaklaşımı ampirik tekrarı merkeze alır. Sinclair'in ifadesiyle, "dil, insanların söylediği şeydir; derlem ise bu söylenmiş olanın hafızasıdır" (Sinclair, 2004, s. 42). Bu yaklaşım, sosyal bilimlerde metin merkezli araştırmaları ölçülebilir, doğrulanabilir ve yeniden üretilebilir hâle getirmiştir. Özellikle medya, siyaset, toplumsal cinsiyet ve dinî söylem alanlarında derlem tabanlı analizler, nicel verinin nitel yorumu desteklediği melez bir epistemoloji ortaya çıkarmıştır (Wodak & Meyer, 2016, s. 29).

Derlem dilbiliminin gücü, nitel gözlemi ortadan kaldırmasında değil, onu nicel doğrulamayla pekiştirmesinde yatar. Biber ve Conrad'ın vurguladığı gibi, "derlem verisi sayı değildir; sayılar üzerinden anlam kurmaktır" (Biber & Conrad, 2009, s. 84). Bu bağlamda corpus-based ve corpus-driven yaklaşımlar, nitel hipotezlerle nicel doğrulama arasında metodolojik köprü işlevi görür.

Örneğin medya söylemi analizinde, belirli kelimelerin sıklık artışı ideolojik bir eğilime işaret ederken; bağlamsal KWIC incelemesi, bu eğilimin anlam yönünü somutlaştırır. Böylece sosyal bilimlerde ölçülebilir anlam kavramı ortaya çıkar (Biber, Conrad & Reppen, 1998, s. 94).

Derlem dilbilimi, yalnızca dilbilimsel değil, aynı zamanda eleştirel-sosyolojik bir araçtır. Fairclough'un "eleştirel söylem çözümlemesi" modelinde olduğu gibi, ideolojik kalıplar artık sezgisel biçimde değil, istatistiksel kanıtlarla teşhis edilebilmektedir (Fairclough, 1995, s. 28). Bu, özellikle medya ve politika söylemlerinde "hakikat-sonrası çağ"ın dilsel stratejilerini çözümlemek açısından belirleyici bir yöntemdir.

Ayrıca derlem tabanlı veri üretimi, bilgi politikaları açısından da yeni sorumluluklar doğurmuştur. Teubert'in uyarısıyla, "her veri bir seçimdir, dolayısıyla her derlem bir ideolojidir" (Teubert, 2005, s. 136). Bu, sosyal bilimcinin artık yalnızca gözlemci değil, aynı

zamanda veri küratörü olmasını gerektirir. Veri toplamak, aynı zamanda anlam inşa etmektir.

Dijital beşerî bilimler çağında derlem yöntemleri, kültürel belleğin dijital temsiline olanak sağlamaktadır. Bu durum hem dilsel hem kültürel kimliklerin korunması açısından stratejik önem taşır. Baker ve Egbert'in belirttiği üzere, "dijital derlemler, bir toplumun kültürel mirasını yeniden kodlama biçimidir" (Baker & Egbert, 2019, s. 58). Arapça, Türkçe ve Farsça gibi tarihsel derinliği olan dillerde derlem çalışmaları, artık yalnızca dilbilimsel değil, medeniyet temsili boyutunda değerlendirilmektedir (Wodak & Meyer, 2016, s. 38).

Bu çerçevede, Corpus.Quran.com, Shamela Derlemi veya ArabiCorpus gibi projeler, klasik metinlerin dijitalleştirilmesiyle hem tarihsel dil çözümlemelerine hem de dinî söylemin yeniden okunmasına zemin hazırlamıştır (Dukes, 2010b, s. 17). Bu süreç, Arap dilbiliminde veriye dayalı çağdaş bir hermenötiğin doğuşuna işaret etmektedir.

Veri temelli her bilimsel model gibi, derlem dilbilimi de etik farkındalık gerektirir. Wodak'ın belirttiği gibi, "veri, insanın sosyal eyleminden doğar; dolayısıyla etik sorumluluktan azade değildir" (Wodak, 2015, s. 18). Sosyal medya derlemleri, mahremiyet, temsil ve rıza konularını beraberinde getirir. Bu nedenle çağdaş derlem araştırmalarında artık "etik tasarım" kavramı öne çıkmıştır: veri toplanmadan önce kültürel bağlamın ve bireysel mahremiyetin korunması esas alınmalıdır (Wodak & Meyer, 2016, s. 27).

Etik boyut, yalnızca insan verisiyle ilgili değildir; aynı zamanda kutsal ve kültürel metinlerin ele alınış biçiminde de geçerlidir. Kur'an veya hadis gibi metinlerin derlem analizi, sadece istatistiksel değil, manevî sorumluluk bilinciyle yürütülmelidir (Wodak & Meyer, 2016, s. 27).

Derlem dilbilimi, yapay zekâ ile birleştiğinde yeni bir epistemik evreye geçmiştir. Büyük dil modelleri (LLM), artık derlem verilerini yalnızca analiz etmekle kalmayıp, yeni dilsel örüntüler üretmektedir. Ancak bu durum, anlamın insandan makineye kaydığı yeni bir bilgi krizini de beraberinde getirir (McEnery, 2021, s. 411).

McEnery, "geleceğin derlemcisi, yalnızca veriyi değil, algoritmayı da anlamalıdır" diyerek bu dönüşümün yönünü belirtir (McEnery, 2021, s. 411). Yapay zekâ tabanlı dil işleme sistemlerinde, derlem temelli yaklaşım hem eğitim verisinin doğruluğunu hem de etik sınırlarını belirleyen temel ölçüt olacaktır.

Bu bağlamda sosyal bilimler, artık "metin okuma"dan "veri okuryazarlığı"na geçmek zorundadır. Çünkü çağımızın toplumsal gerçekliği, kelimelerde değil, veri örüntülerinde saklıdır.

Derlem dilbilimi, dilin nicel haritasını çıkarırken, aynı zamanda insanlığın ortak hafızasını da kayıt altına alır. Sinclair'in şu cümlesi, bu yaklaşımın özünü özetler: "Derlem, insanın söylediğini değil, hatırladığını gösterir" (Sinclair, 2004, s. 35). Dolayısıyla derlem tabanlı yöntem, yalnızca akademik değil, etik ve toplumsal bir projedir.

Bu çalışma, sosyal bilimler bağlamında derlem dilbiliminin;

- Veri temelli düşünmeyi güçlendirdiğini,
- İdeolojik söylemleri görünür kıldığını,
- Dijital beşerî bilimler ile klasik filoloji arasında köprü kurduğunu,
- Etik farkındalık ve kültürel temsil bilinci kazandırdığını,
- Ve nihayetinde bilgiyi nesnel ölçekte demokratikleştirdiğini ortaya koymuştur.

Sonuç olarak, derlem dilbilimi sosyal bilimler için artık bir yöntem değil, bir epistemolojik devrimdir. Bu devrim, dilin doğasını anlama çabasını, insanın kendi toplumsal varlığını anlama sürecine dönüştürmüştür.

Kaynakça

Baker, P. (2006). *Using corpora in discourse analysis*. London: Continuum.

Baker, P. (2010). *Using corpora in discourse studies*. London: Routledge.

Baker, P., & Egbert, J. (2019). *Triangulating methodological approaches in corpus-assisted discourse studies*. London: Routledge.

Baker, P., & McEnery, T. (2015). *Corpora and discourse studies: Integrating discourse and corpora*. London: Palgrave Macmillan.

Baker, P., Gabrielatos, C., & McEnery, T. (2013). *Discourse analysis and media attitudes: The representation of Islam in the British press*. Cambridge University Press.

Bender, E. M. (2013). *Linguistic fundamentals for natural language processing*. San Rafael, CA: Morgan & Claypool.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge: Cambridge University Press.

Cameron, D. (1992). *Feminism and linguistic theory*. London: Palgrave Macmillan.

Cameron, D. (1995). *Verbal hygiene*. London: Routledge.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Davies, M. (2009a). *The corpus of contemporary American English (COCA): Design, architecture, and linguistic insights*. Provo, UT: Brigham Young University.

Davies, M. (2009b). *The 400 million word corpus of contemporary American English (COCA)*. Provo, UT: Brigham Young University.

Dukes, K. (2010a). Integrated knowledge resources for the Arabic language: The Quranic Arabic Corpus. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC)* (p. 16). Valletta, Malta.

Dukes, K. (2010b). The Quranic Arabic Corpus: Linguistic and semantic annotation. In *Proceedings of LREC 2010* (p. 17). Valletta, Malta.

Fairclough, N. (1989). *Language and power*. London: Longman.

Fairclough, N. (1995a). *Critical discourse analysis: The critical study of language*. London: Longman.

Fairclough, N. (1995b). *Media discourse*. London: Edward Arnold.

Firth, J. R. (1957). A synopsis of linguistic theory, 1930–1955. In *Studies in Linguistic Analysis* (pp. 1–32). Oxford: Philological Society.

Firth, J. R. (1957). *Papers in linguistics 1934–1951*. Oxford University Press.

Firuzabadi, M. b. Y. (2011). *el-Kâmûsu'l-Muhît*. Beyrut: Daru'l-Kütübi'l-İlmiyye.

Francis, W. N., & Kučera, H. (1964). *A standard corpus of present-day edited American English*. Providence, RI: Brown University.

Gries, S. T. (2009). *Quantitative corpus linguistics with R*. London: Routledge.

Habash, N. (2010). *Introduction to Arabic natural language processing*. San Rafael, CA: Morgan & Claypool.

Habash, N., & Dukes, K. (2010). The Quranic Arabic Corpus: Using linguistic and semantic annotation to analyze the Quran. *Language Resources and Evaluation, 44*(1), 123–138.

Habash, N., & Rambow, O. (2005). Arabic tokenization, part-of-speech tagging and morphological disambiguation in one fell swoop. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL)* (pp. 41–47). Ann Arbor, MI: Association for Computational Linguistics.

Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. London: Edward Arnold.

Halliday, M. A. K. (1991). Corpus studies and probabilistic grammar. In K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics* (pp. 30–43). Longman.

Hasan, R. (1985). The structure of a text. In M. A. K. Halliday & R. Hasan (Eds.), *Language, context and text: Aspects of language in a social-semiotic perspective* (pp. 52–69). Oxford: Oxford University Press.

Hanks, P. (1990). Corpus evidence and lexicography. *International Journal of Lexicography, 3*(3), 137–146.

İbn Manzur, M. b. M. (1999). *Lisânu'l-Arab* (15 cilt). Beyrut: Dâru Sadir.

Johns, T. (1991). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *ELR Journal, 4*(New Series), 27–46.

Kilito, A. (2008). *Arabe et le Coran: Études linguistiques*. Paris: Maisonneuve.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Leech, G. (1979). Corpora and theories of linguistic performance. *Linguistics, 17*, 7–26.

Leech, G. (1981). *Semantics: The study of meaning* (2nd ed.). Harmondsworth: Penguin.

Mauranen, A. (2009). Features of data-driven research. *Corpora, 4*(2), 23–44.

McCarthy, M. (1991). *Discourse analysis for language teachers*. Cambridge: Cambridge University Press.

McEnery, T., Baker, P., & Hardie, A. (2012). Corpus linguistics and social change. *Journal of English Linguistics, 40*(1), 5–28.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.

McEnery, T., & Wilson, A. (1996). *Corpus linguistics: An introduction*. Edinburgh: Edinburgh University Press.

Michel, J.-B., Shen, Y. K., Presser Aiden, A., Veres, A., Gray, M. K., The Google Books Team, ... & Lieberman Aiden, E. (2011). Quantitative analysis of culture using millions of digitized books. *Science, 331*(6014), 176–182.

Nesi, H. (2000). *The use and abuse of EFL dictionaries*. Tübingen: Niemeyer.

Partington, A., Duguid, A., & Taylor, C. (2013). *Patterns and meanings in discourse: Theory and practice in corpus-assisted discourse studies (CADS)*. Amsterdam: John Benjamins.

Raskin, V. (1985). *Semantic mechanisms of humor*. Dordrecht: Reidel.

Scott, M. (2018). *WordSmith Tools* (Version 8) [Computer software]. Oxford: Lexical Analysis Software.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. London: Routledge.

Stubbs, M. (1997). *Text and corpus analysis: Computer-assisted studies of language and culture*. Oxford: Blackwell.

Teubert, W. (2005). Meaning, discourse and corpora. *Applied Linguistics, 26*(2), 113–136.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: John Benjamins.

Wodak, R. (2015). *The politics of fear: What right-wing populist discourses mean*. London: Sage.

Wodak, R., & Meyer, M. (2016). *Methods of critical discourse studies* (3rd ed.). London: Sage.

Wodak, R., & Reisigl, M. (2001). *Discourse and discrimination: Rhetorics of racism and antisemitism*. London: Routledge.