

KISA METİNLERDE YAZAR TANIMA: TWITTER İÇİN BİR YÖNTEM ÖNERİSİ

Utku TANRIVERE¹

Özet

Verili bir metnin yazarını tespit etmek ve birkaç metnin arasından yazarı aynı olanları bulmak, adli dilbilimin önemli uygulama alanlarından biridir. Günümüze dek yapılan çalışmaların çoğu, gazetelerin köşe yazıları gibi uzun ve nispeten fazla miktarda dilsel veri içeren bütüncü ile gerçekleştirildiğinden, kısa metinlerde yazar tanımaya ilişkin herhangi bir yöntem sunmamaktadır. Zira adli dilbilim kapsamında yapılan çalışmalarda, incelemelere konu edilen metinler oldukça kısadır. Bununla birlikte, mikro-blog siteleri de dâhil olmak üzere, sosyal medya üzerinde yapılan paylaşımlar da çoğu kez ceza veya hukuk davalarının konusu olabilmektedir. Dolayısıyla kısa metinler üzerinde yazar tanımaya ilişkin birtakım yöntem ve yaklaşımlara ihtiyaç duyulmaktadır. Bu çalışma, özellikle karakter kısıtı gözetilerek seçilen Twitter adlı mikro-blog sitesinden toplanan çeşitli metinlerin; dilbilgisi, noktalama, sözcük ve bağlam özellikleri kullanılarak yazarlarının tespitine ilişkin bir yöntem önermeyi amaçlamaktadır. Ayırt edici olduğu veya olabileceği görülen özellikler paylaşılma ve çoğunlukla bütüncüye göre biçimlendirilmesi gereken yöntemler, genel başlıklarıyla birlikte sunulmaktadır.

Anahtar Sözcükler: adli dilbilim, yazar tanıma, yazar tespiti, kısa metinler, sosyal medya

AUTHOR IDENTIFICATION IN SHORT TEXTS: A METHODOLOGICAL PROPOSAL FOR TWITTER

Abstract

Determining a given text's author and finding the texts whose authors are the same, through several texts is one of the most important application fields of forensic linguistics. Most of the studies, which are done up to the present, were conducted along with a long corpus, which also consists of relatively more linguistic data such as newspaper articles. Thus, they don't suggest a method in terms of author identification

¹ Hacettepe Üniversitesi Adli Bilimler Anabilim Dalı doktora adayı, filolog. utkutanrivere@hacettepe.edu.tr

of short texts. Yet, in forensic linguistic studies, the entreated texts are quite short. However, social media entries including the micro-blogs can be subject to criminal or law cases several times. Accordingly, there's a need for some methods and aspects related to author identification. This study aims to suggest a method to determine authors by using grammar, punctuation, lexis and context features of texts, which were collected from the micro-blog named Twitter, with a special regard to its character constraint. The features, which are seen to be or may be distinctive, are shared and the methods, which mostly need to be formatted related to the corpus, are presented along with their general titles.

Keywords: forensic linguistics, author identification, authorship attribution, short texts, social media

Giriş

Yazar tanıma, belirli bir metnin yazarının kim olduğunu, bazen de kim olmadığını belirlemeye yönelik gerçekleştirilen dilbilim temelli bir çalışmadır. Bu incelemeler, adli dilbilim sahası için yadsınamaz bir öneme sahiptir. Jan Svartvik (1968), ilk kez “adli dilbilim” adlandırmasını kullanarak bu alanın bilimsel yazına girmesini sağladığı “*The Evans Statemens: A Case For Forensic Linguistics*” isimli eseri ile adli dilbilime ilişkin kuramsal bir çerçeve oluşturmamış olsa da yazar tanıma çalışmalarına temel teşkil edebilecek bir örnek ortaya koymuştur. Svartvik’in (1968) eseri, esasen istatistiksel bir çalışma olarak nitelenmekte, sonrasında Malcolm Coulthard tarafından 1994 yılında gerçekleştirilen ve Derek William Bentley’in suçsuzluğunun ortaya çıkmasında önemli bir rol oynayan çalışma ise betimleyici bir dilbilimsel analiz olarak görülmektedir (Olsson, 2008, s. 20). Coulthard’a (2005) göre “bir metnin ne söylediği ve yazarının kim olduğu” (s. 10), adli dilbilim uzmanlarına en sık sorulan sorulardır. Bu uğraşların çok daha öncesinde Antik Yunan’da dahi dramaturglar, yazılan oyunların, kendi oyunlarından aşırımlar içerdiğini öne sürmüşlerse de bu iddiaların hiçbiri “adli” bir süreç içerisinde ele alınmamıştır (Olsson, 2008, s. 4).

Her insanın, dili kendine özgü bir şekilde kullanıyor oluşu ve bu farkların kolaylıkla ortaya konulabileceği inancı, süreç içerisinde “dilsel parmak izi” (İng. linguistic fingerprint) kavramının ortaya çıkmasına neden olsa da bu kavramı destekleyecek nitelikteki veriler henüz oldukça yetersizdir ve sistematik olarak sunulmamıştır (Olsson, 2008, s. 25-26; Coulthard, 2005, s. 14). Adli dilbilim ve yazar tanıma çalışmaları kapsamında yapılması gereken, her bir kişinin eşsiz olduğu varsayılan üslubunu ispatlamaya çalışmaktansa, incelemeye konu edilen metinler bağlamında yazarların aralarındaki üslup farkını ortaya koymaktır. Zira

McMenamin'in (2002) de belirttiği gibi "iki ayrı birey, dili tamamen aynı biçimde kullanmaz ve algılamaz" (s. 62), böylece bireylerin kendi üslup ve dilsel alışkanlıklarıyla oluşan bireysel dil (İng. idiolect) gelişir. Bunun sonucunda ise başka yazarlar tarafından kaleme alınmış iki ayrı metinde görülebilecek farklılıkların kaynağı olan "yazarlararası değişke" (İng. inter-author variation) oluşur (Olsson, 2008, s. 33), ki bu fark veya benzerlikler, fiziksel bir parmak izinin güvenilirliğiyle karşılaştırılmaz (Brückner, 2011, s. 15). "Tek yazar değişkesi" (İng. intra-author variation) kavramı ise bir yazar tarafından kaleme alınmış çeşitli metinlerin arasındaki farklılıkları niteler ve buna farklı nedenler yol açabilir (Olsson, 2008, s. 33).

Bir metnin oluşturulması sürecinde dört farklı yazarlık türünün bulunabileceğini tespit eden Grant (2008, s. 217-218), "uygulayıcı yazarlık" (İng. executive authorship) kavramının, tamamen özgün metnin oluşturulması süreciyle ilgili olduğunu belirtir.² Buna göre yazar tanıma çalışmalarında esas teşkil eden soru, uygulayıcı yazar ile ilgilidir ve bir metnin birden çok uygulayıcı yazarı olabileceği gibi, uygulayıcı yazar tarafından bir başkasına yazdırılmış bir metin de incelemenin konusu olabilir (Grant, 2008, s. 218).

Türkçe Metinler Odağında Yazar Tanıma Çalışmaları

Bilinmeyen bir kimse tarafından yazılan metinlerin yazarlarının tespit edilmesini amaçlayan yazar tanıma çalışmaları, temel olarak dilbilimci ve bilgisayar bilimcilerin ilgi konusu olagelmıştır. Türkçe bilimsel yazına (veya Türkçe metinlerde yazar tespitine ilişkin bilimsel yazına) bakıldığında ise bilgisayar bilimcilerin, konuya dilbilimcilerden daha çok ilgi duyduğu söylenilebilir. Bu noktada ortaya konulan çalışmaların ise çoğunlukla ikili (İng. bigram) ya da üçlü (İng. trigram) karakter gruplarının kullanım sıklıklarına dayalı bir yöntem olarak gelişen n-gram (Amasyalı & Diri, 2006, s. 224; Doğan & Diri, 2010, s. 11) ve cümle sayısı, sözcük sayısı, sözcük kategorileri ve bunun gibi başka özellik vektörlerinin (Taş & Görür, 2007, s. 154) birlikte veya ayrı ayrı kullanıldığı bilgisayar destekli analizlerden oluştuğu görülmektedir.

Sözü edilen bilgisayar destekli yazar tanıma çalışmalarının ortak noktası, derlem olarak gazetelerin köşe yazılarını incelemeleri ve netice olarak da köşe yazarlarıyla eşleştirmeleridir (Amasyalı & Diri, 2006, s. 222; Doğan & Diri, 2010, s. 12; Taş & Görür, 2007, s. 153). Bu çalışmaların, adli dilbilim bağlamında gerçekleştirilen yazar tanıma

² Diğer yazarlık türleri ise metinler-arası özelliklerinin; atıf, alıntı ve intihal dâhil olmak üzere, kronolojik olarak önce yazılan esere dayandırıldığı "ön-yazarlık" (İng. Precursory authorship), bir metni okuyarak veya bu metnin altına imza atarak metinde yer alan içeriği kabul ve beyan etmeye ilişkin "bildirimsel yazarlık" (İng. Declarative authorship) ve oluşturulmuş bir metnin dilbilgisi, yazım ve anlatımına ilişkin düzeltmelerini niteleyen "düzeltmenlik" (İng. Revisionary authorship) olarak belirtilir (Grant, 2008: 217, 218).

incelemeleri çerçevesinde kullanışlılıklarına dair veri bulunmamaktadır. Zira gazetelerdeki köşe yazılarının editör kontrolünden geçmeleri nedeniyle yazım ve noktalama hataları başta olmak üzere birtakım özelliklerin, yazarı ayırt etmekte kullanılamayacağını vurgulamak gerekir. Bununla birlikte çalışmalara konu edilen örneklem, çok sayıda ve nispeten uzun metinler içermektedir. MacLeod ve Grant, gerçekleştirdikleri bir çalışmada aynı konuya değinirler:

Yazar tanıma konusundaki geleneksel yaklaşımlar, çalışmanın genellikle başarıyla uygulanabileceği miktardaki veri ile gerçekleştirilmektedir ve bu nedenle kısa mesajlar (SMS), Twitter gibi mikro-bloglar ya da anlık yazışmalar gibi kısa metinlerin analizi için uygun değildir (2012, s. 210).³

Adli dilbilim bağlamında ise yazar tespiti gerektiren vakalar da dâhil olmak üzere eldeki örneklem (deney grubu) ve karşılaştırma verisi (kontrol grubu) genellikle oldukça kısıtlıdır: Metin kısadır, çok sayıda sözcük yoktur ve bu nedenle yazara ilişkin çok az miktarda bilgi sunarlar (Coulthard & Johnson, 2007, s. 162, 172). Örneğin, buna paralel olarak, Türkçe intihar mektupları üzerinde yapılan bir çalışma da; gerçek intihar mektuplarının, intihar etmeyecek kimseler tarafından “intihar mektubu” olarak yazılmış sahte mektuplar karşısında kayda değer miktarda daha az cümle ve sözcük içerdiğini ortaya koymuştur (Tanrıvere, 2016, s. 44). Ne var ki, yazar tanıma çalışmalarının karşılaştırma sürecinde yer alan “muhtemel yazarlar”, diğer bilgi ve ipuçlarının da doğrultusunda sayıca kısıtlanmış durumdadır ve adli dilbilim uzmanı, karşılaştırmasını, bu kısıtlı sayıdaki yazarlar arasında gerçekleştirir (Coulthard, 2005, s. 14).

Örneklem, Yöntem ve Uygulama

Yazar tanıma çalışmalarında her zaman; çok sayıda metinden hangilerinin yazarlarının aynı (veya farklı) olabileceği, iki metnin yazarlarının aynı olup olmadığı gibi tek ve belirli bir soru bulunmamakla beraber, bu soruların cevaplanması için de her zaman uygulanabilecek tek ve genel-geçer bir yöntem bulunmamaktadır (Grant, 2008, s. 216, 227). Bu konuda yaptığımız çalışmalar, araştırma sorusu aynı kalsa dahi, yazarın yönlendirici niteliği nedeniyle değişen bütüncüye göre farklı yöntem ve yaklaşımlar geliştirilmesi gerektiğini ortaya koymuştur. Bu konudaki çalışmalar kapsamında görgül bir yaklaşım esas alındığından, ilerleyen bölümlerde, ele alınan ileti örnekleri bağlamında deneysel olarak geliştirilen betimleyici inceleme ve karşılaştırma yöntemleri sunulmaktadır.

³ Yazar tarafından tercüme edilmiştir.

Veri Toplama ve Verilerin Niteliği

Bir mikro-blog sitesi olan Twitter (www.twitter.com) üzerinde; “herkese açık” (İng. public) biçimde paylaşımında bulunan rastgele 10 hesap tespit edilmiş, her bir hesaptan 5’er adet mesaj (İng. tweet) yine gelişigüzel biçimde seçilerek incelemeye konu edilmiştir. Yazarlar D1’den D10’a kadar numaralandırılmış, mesajlar ise yazar numarasından sonra “-n” ile etiketlenmiştir. Buna göre “D3-4”, üçüncü yazarın dördüncü mesajıdır.

Anılan internet sitesinde veri toplanırken, aynı yazar tarafından farklı zamanlarda oluşturulmuş mesajların seçilmesine özen gösterilmiştir. Mesajların büyük bir bölümünde, aynı yazar tarafından yazılan diğer bir mesaj ile en az bir haftalık bir süreç bulunmaktadır. Bu zaman aralığı bütün veri içerisinde en fazla üç güne kadar yaklaşmıştır. Mesaj tarihlerine ilişkin tabloya aşağıda yer verilmiştir.

Tablo 1

Yazarlara Göre Mesajların Üretilme Tarihleri

Yazar / Mesaj	1	2	3	4	5
D1	12/07/2016	04/07/2016	26/06/2016	18/06/2016	30/05/2016
D2	13/07/2016	13/07/2012	22/12/2014	16/07/2015	27/11/2012
D3	12/07/2016	04/07/2016	26/06/2016	19/06/2016	12/06/2016
D4	12/07/2016	09/07/2016	02/07/2016	24/06/2016	15/06/2016
D5	13/07/2016	05/07/2016	28/06/2016	21/06/2016	05/06/2016
D6	12/07/2016	27/05/2016	13/05/2016	05/05/2016	20/03/2016
D7	12/05/2016	14/04/2016	09/03/2016	01/03/2016	19/12/2015
D8	13/07/2016	06/07/2016	30/06/2016	23/06/2016	15/06/2016
D9	13/07/2016	05/07/2016	28/06/2016	21/06/2016	18/09/2015
D10	13/07/2016	06/07/2016	30/06/2016	23/06/2016	17/06/2016

Buna göre her bir yazarın ilk mesaj tarihi başlangıç kabul edildiğinde; 25 mesajda gün, 11 mesajda ay ve 4 mesajda yıl farkı bulunmaktadır. Mesajların üretilme tarihine ilişkin bu seçimin, yazarların yerleşik üslubunu ortaya koymakta fayda sağlayacağı varsayılmaktadır. D2 numaralı yazar, mesajlarını, tabloda belirtilen sıralamayla yeniden paylaşmıştır. (İng. Retweet)

Mevcut veriye ilişkin bir diğer varsayım ise mesajların, yayımlandıkları tarihte üretilmiş olduklarıdır. Ayrıca; mesajları incelemeye alınan her bir hesabın sadece bir yazarının bulunduğu ve hiçbir yazarın, analize konu edilen birden fazla hesabın yazarı

olmadığı da önsel olarak doğru ve geçerli kabul edilmektedir. Yazarların, Türkçenin “anadil konuşucusu” olduğu varsayılmaktadır.

Örnekleme için Twitter adlı mikro-blog sitesinin seçilme nedeni, daha önce de belirtildiği gibi, yazar tanıma analizleri de dâhil olmak üzere adli dilbilim incelemeleri kapsamındaki metinlerin kısa olmaları ve oldukça kısıtlı bir veri sunabilmeleridir. Anılan platformda, her bir mesaj, en fazla 140 karakter ile sınırlandırılmaktadır. Bazı kullanıcılar, bu sınırlama nedeniyle, art arda birden fazla mesaj yayınlamaktadır. Toplanan veride ise her bir mesaj yalnızca kendisi ile sınırlı olup mesajın öncesi veya sonrasında ilgili bir başka ileti bulunmamaktadır. Dolayısıyla, toplanılan örneklemin, metin uzunluğu yönünden, adli dilbilim çalışmalarında ele alınan metinlerin özelliklerini büyük ölçüde karşılayacağı öngörülmektedir.

Ayrıca yazarların yaş, cinsiyet vb. demografik bilgilerine ilişkin doğrulanmış hiçbir veri bulunmamaktadır. Kimi yazarların kullanıcı hesaplarında kullandıkları tanımlayıcı görseller (avatar, profil fotoğrafı) veya kullanıcı adları, yaş ve cinsiyete ilişkin bir fikir sunsa da kesin olarak bir bilgiye sahip olunmamakta, yapılan bu çalışma kapsamında ise belirtilen demografik değişkenlerin herhangi bir önem arz etmediği değerlendirilmektedir. Zira incelemenin temel amacı, pek çok farklı değişkenin bir araya gelerek oluşturduğu dinamikler çerçevesinde gelişen dilsel üslubun, farklı yazarlar arasında ayırt edilmesi ve buna ilişkin birtakım yöntem önermelerinin ortaya konulmasıdır.

Dilbilgisi ve Noktalama Bakımından Yöntem Geliştirme

Bireyin eğitim düzeyi, sosyal ve ekonomik durumu, metnin oluşturulduğu dolayım gibi çeşitli değişkenlerle bağlantılı olan dilbilgisi ile yazım ve noktalamaya ilişkin eğilim, tercih ve hatalar, yazarlar ve ürettikleri metinler arasında ayırt edici olabilmektedir.

Elde edilmiş olan 50 mikro-blog iletisinden rastgele seçilen 12 tanesi, yine örneklemin içerisinden gelişigüzel olarak saptanan ve incelemenin başlangıç noktası olması nedeniyle “M0” olarak anılan bir ileti karşısında incelemeye alınmıştır. Karşılaştırmada yer alan 12 mesaj da rastgele bir sıralamayla M1’den M12’ye kadar adlandırılmıştır. Bu esnada mevcut olan tek bilgi, M0 ile M1-M12 arasındaki mesajlarda bir yazarın tekrar ettiğiidir.

M0 ile karşılaştırma grubu mesajları incelendiğinde; büyük harf ve nokta kullanımları, aynı yazarın diğer mesajının tespit edilmesinde önsel olarak yeterli olmuştur.

Tablo 2

Mesajların İki Ölçütle Karşılaştırılması Örneği

Yazar	Mesaj	Cümle başlangıcı	Cümle sonu
D1	M0	Küçük harf	İki nokta
D9	M1	Küçük harf	—*
D2	M2	Büyük harf	Bir nokta
D8	M3	Büyük harf	Bir nokta
D4	M4	Büyük harf	Bir nokta
D3	M5	Büyük harf	Bir nokta
D9	M6	Küçük harf	—*
D10	M7	Büyük harf	Bir nokta
D6	M8	Küçük harf	—*
D2	M9	Büyük harf	İki nokta
D1	M10	Küçük harf	İki nokta
D5	M11	Büyük harf	Bir nokta
D7	M12	Küçük harf	Bir nokta

* İşaretsiz mesajların cümle sonlarında noktalama işareti kullanılmamaktadır.

Yukarıdaki tabloda verilen bilgilere göre M0 mesajında cümleler küçük harfle başlamakta, cümle sonunda ise iki nokta kullanılmaktadır. Cümleye küçük harfle başlanan diğer mesajlar M1, M6, M8, M10 ve M12'dir. Cümlenin sonunda noktalama işareti olarak ise iki nokta kullanılan sadece M9 ve M10 mesajları bulunmaktadır. İncelemede kullanılan bu iki temel karşılaştırma ölçütü ile elde edilen iki benzerlik grubunun kesişiminde yer alan ileti M10'dur. Yapılan kontrolde de M0 ve M10 iletilerinin, D1 yazarı tarafından yayımlandığı görülmüştür.

Karşılaştırmaya konu edilen 13 mesajda yapılan analiz, büyük harf kullanımı ve noktalama ölçütlerinin gerektiğinde (veri mevcut ise) detaylandırılmasının faydalı olacağını göstermiştir. Buna göre büyük harf kullanımını; "cümle başlangıcı, tırnak içindeki cümle başlangıcı, cümle başındaki özel adlar, cümle başında olmayan özel adlar" gibi alt kategoriler ile incelemek de yararlı olabilecektir. Aynı şekilde noktalama işaretleri; "cümle içindeki noktalama işaretleri, cümle sonundaki noktalama işaretleri, tırnak içindeki cümle sonunda kullanılan noktalama işaretleri, mesaj sonundaki noktalama işaretleri" benzeri gruplara ayrılabilir. Çalışmanın gerçekleştirildiği mesajlar genel olarak, örneklendirilen bu başlıklar

kullanıldığında, ayırt ediciliğin arttığına işaret etmektedir. Bunlarla birlikte MacLeod ve Grant'in de belirttiği gibi (2012, s. 218), noktalama işaretlerinin tekli (örneğin “!”), çoklu (“!!!!” gibi) veya karışık (“?!?!?!”) kullanımı, parantezlerin (parantez, köşeli parantez, küme parantezi, “büyüktür” ve “küçüktür” simgeleri) ya da tırnak işaretlerinin (“” veya ‘’) oluşturulma şekline ilişkin bir ayırım oluşturmanın da yazar tanıma çalışmaları için kullanışlı olacağı değerlendirilmektedir. Ayrıca, cümlenin belirli yerlerinde veya belirli sözcüklerin yazımında yapılan yanlışlar, yazarın tespitinde oldukça önemli bir rol oynamaktadır.

Yukarıda değinilen ölçütlerin yanı sıra, yer verilmiş ise duygu belirtmeye yarayan noktalama işaretleri (:P, :, ;), :’(((, :* gibi) ve ön tanımlı yüz ifadeleri (İng. Emoji, emoticon) ile konu etiketi (İng. hashtag) kullanımının da ayırt edici olmakla birlikte, ayrı bir başlık altında değerlendirilmesi gerektiği görülmektedir. Zira bu çalışmada dilbilgisi ve noktalama bağlamında gerçekleştirilen incelemenin odağını biçim, sözcük ve bağlam yönünden yapılan incelemenin odağını ise içerik oluşturmaktadır. Noktalama işaretleri kullanılarak oluşturulan veya ön tanımlı yüz ifadeleri ile konu etiketleri ise hem biçim hem de içerik düzleminde ele alınabilecek niteliktedir. Örneğin, bir yüz ifadesinin noktalama işaretleriyle mi yoksa ön tanımlı biçimiyle mi oluşturulduğu kadar, ifadenin anlamı da önem taşımaktadır. Aynı şekilde bir konu etiketinin kullanılıp kullanılmadığı kadar, bu etiketteki sözcüklerin hem dilbilgisi hem de sözcük ve bağlam yönünden incelenmesi söz konusu olacaktır. Dolayısıyla ayırt edici olduğu görülen bu özelliklerin her iki bağlamda da ele alınması gerekeceğinden, farklı bir başlık altında incelenmelerinin yöntem bakımından yerinde olacağı değerlendirilmektedir.

Bununla birlikte aynı yazar, dilbilgisi ve noktalama bağlamında, metni oluşturduğu aygıt ile birlikte değişen bir dilsel alışkanlık da gösterebilir. Örnek vermek gerekirse, yazarın kullandığı aygıtlardan birisi, kullanılan noktalama işaretinden sonraki harfin büyük veya küçük olacağını otomatik olarak tayin ediyor veya yazarın yazım yanlışlarını otomatik olarak düzeltiyor olabilir. Bu durumda, yazarın sözcük kullanımı ve üslubu ayırt edicilik bakımından daha büyük bir önem taşıyacaktır.

Yazarların genel dilsel alışkanlıklarını karşılaştırmak adına, her yazarın 5'er mesajından elde edilen veriler, yukarıda belirtilen 2 ana ve toplam 8 alt kategori kapsamında değerlendirildiğinde; diğer yazarlar karşısında en fazla benzerlik unsuru tespit edilen D1 kodlu yazarın benzerlik/farklılık tablosu Şekil 1'de gösterilmiştir.

Anılan şeklin sunumundan önce belirtmek gerekir ki, iki metnin yazarının aynı olup olmadığı incelenirken; ölçünlü dil benzerlikleri, diğer bir deyişle dil kurallarına uygun olan benzerlikler, karşılaştırmaya esas alınmamalıdır. Değerlendirmeyi temel olarak etkileyen asıl

unsur, kural dışı dil kullanımı ve dolayısıyla yapılan hatalar, görülen farklılıklardır. Yazar tanıma çalışmalarının temel amacı yerleşik üslubu tespit etmek, ölçmek ve karşılaştırmak olduğundan, ölçünlü dil kuralı kapsamındaki benzerlikler, inceleme açısından yanıltıcı olabilecek niteliktedir. Zira ölçünlü dil kuralları, bütün eğitim-öğretim kurumlarında “doğru” ve “kullanılması gereken” olarak sunulmaktadır.

Örnek vermek gerekirse; iki ayrı metnin birer nokta ile sonlandırılmış olması, yazarların aynı olduğuna dair bir izlenim uyandırmamaktadır. Buna karşın her iki metnin de iki nokta ile sonlandırılması, yazarlar açısından bir benzerlik unsuru olarak işaretlenebilir. Aynı şekilde, iki metinde cümlelere büyük harfle başlanmış olması, yazar açısından gösterilebilecek bir benzerlik unsuru olmamakla birlikte; iki metinde de cümlelerin küçük harfle başlaması, diğer metinler karşısında ayırt edici bir göstergedir.

D1 yazarının karşılaştırmaya konu edildiği Şekil 1’de; diğer yazarların dil kullanımına benzerlik teşkil eden ölçütler yeşil, kararsızlığa yol açan veya iki yazarın da ölçünlü dil kuralına uyması gibi ayırt edici olmayan ölçütler mavi, farklılık gösteren ölçütler ise kırmızı renkle işaretlenmiştir. Noktalama ölçütü “değişken” olarak belirtilen D6 ve D10 yazarları hiçbir zaman iki nokta kullanmadıkları için “farklı unsur” olarak işaretlenmiştir.

YAZAR	BÜYÜK HARF KULLANIMI				NOKTALAMA			
	CÜMLE BAŞI	TIRNAK İÇİNDEKİ CÜMLELER	ÖZEL ADLAR		CÜMLE İÇİ	CÜMLE SONU	TIRNAK İÇİNDEKİ CÜMLE SONU	MESAJ SONU
			CÜMLE BAŞI	CÜMLE ORTASI				
D1	Küçük	Küçük	Küçük	Büyük	Virgül, tırnak	İki nokta	Kullanmıyor	İki nokta
D2	Büyük	—	—	Büyük	Virgül, tırnak	İki nokta	—	İki nokta
D3	Büyük	Büyük	Büyük	—	Virgül, tırnak	Bir nokta	Nokta yok, bir ?	Kullanmıyor
D4	Büyük	Küçük	Büyük	Büyük	Virgül, tırnak	Bir nokta	Nokta yok, bir ?	Bir nokta
D5	Büyük	Küçük	Büyük	Büyük	Virgül, tırnak	Bir nokta	Kullanmıyor	Bir nokta
D6	Küçük	—	Küçük	Küçük	Virgül	Değişken	—	Değişken
D7	Küçük	—	Küçük	Küçük	Virgül	Bir nokta	—	Bir nokta
D8	Büyük	Küçük	Büyük	Büyük	Virgül, tırnak	Bir nokta	Nokta yok, ? Yok	Bir nokta
D9	Küçük	—	Küçük	Küçük	Kullanmıyor	Kullanmıyor	—	Kullanmıyor
D10	Büyük	—	—	Büyük	Kullanmıyor	Bir nokta	—	Değişken

Şekil 1. D1 yazarının diğer yazarlar karşısında benzerlik/farklılık tablosu

Yazar, D2 ile aynı şekilde, cümle ve mesajlarını ikişer nokta ile sonlandırır da D2’nin aksine cümleye küçük harfle başlamaktadır. D4 yazarı ile tek benzerlik ise tırnak içindeki cümlelerin küçük harfle başlamasıdır, başka benzerlik bulunmamaktadır. D5 ile aynı şekilde tırnak içindeki cümleler küçük harfle başlamakta ve tırnak içindeki cümlelerin sonunda noktalama işareti kullanılmamaktadır. Buna karşın D5; cümleleri ve cümle başındaki özel adları büyük harfle başlatmakta, cümle ve mesajlarını birer nokta ile sonlandırmaktadır. D8, tırnak içindeki cümleleri küçük harfle başlatmakta ve tırnak içindeki cümle sonunda

noktalama işareti kullanmamaktadır. Ayırt edici olarak ise cümleyi ve cümle başındaki özel adları büyük harfle başlatmakta, cümle ve mesajlarını birer nokta ile sonlandırmaktadır. D6, D7 ve D9 yazarları da cümleleri ve cümle başındaki özel adları küçük harfle başlatmaktadır ancak cümle ortasındaki özel adları da küçük harfle yazmakta ve noktalama eğilimleriyle D1'den ayrılmaktadır.

Bu karşılaştırma, nicel olarak ifade edilmeye çalışıldığında benzerlik unsurlarına (+1), etkisiz unsurlara (0) ve farklılık unsurlarına (-1) değeri verilerek her bir karşılaştırma için toplama işlemi yapılmıştır. Değerlendirme cetveli olarak ise (+8), (+7), (+6), (+5) sonuçları için “kuvvetle muhtemel aynı”; (+4), (+3), (+2) sonuçları için “muhtemelen aynı”; (+1), (0), (-1) sonuçları için “kararsız”; (-2), (-3), (-4) sonuçları için “muhtemelen farklı” ve (-5), (-6), (-7), (-8) sonuçları için “kuvvetle muhtemel farklı” tanımlamaları tayin edilmiştir. Ancak bu hesap ve tanımlama cetvelinin uygulanmasıyla, nitel yöntemle ayırt edilen yazarlar için “hatalı pozitif” sonuçlar alındığı görülmüş, bu nedenle elde edilen sonuçlar için herhangi bir şekilde nicel ifade kullanılmamıştır.

Dilbilgisi ve noktalama incelemeleri kapsamında kullanılarak yazar tanıma için verimli olduğu görülen veya mevcut örneklem için uygun olmaması nedeniyle kullanılmaması da ayırt edici olduğu değerlendirilen ölçütler, aşağıdaki tabloda özet olarak sunulmaktadır.

Tablo 3

Biçim Yönünden Ayırt Edici Olduğu Değerlendirilen Ölçütler

İnceleme konusu	Ölçütler
Konum	Cümle içi
	Cümle sonu
	Tırnak içinde cümle sonu
	Mesaj sonu
Sayı	Tekli
	Çoklu
	Karışık
Noktalama	Normal parantez
	Köşeli parantez
	Küme parantezi
	Büyüktür/küçüktür sembolleri
Tırnak işaretleri	Tek
	Çift
Yazım yanlışları	Konum
	Sözcük

Örneklemede yer alan cümle türlerine (basit, sıralı, birleşik), cümlenin eylem çatılarına (etken, edilgen) ve eylemin zamanına (geçmiş, şimdiki, gelecek, geniş) yönelik gerçekleştirilen karşılaştırma, yazarın eğilimine ilişkin ayırt edici bir sonuç vermemiş olup bu durumun metin uzunluğuna ilişkin kısıtlamadan kaynaklandığı değerlendirilmektedir.

Sözcük ve Bağlam Bakımından Yöntem Geliştirme

Hem dilbilgisi ve noktalama ile ulaşılan ayrımların her zaman yukarıdaki örnekteki kadar açık olmayacak oluşu dolayısıyla ortaya çıkan farklı ölçüt gereksinimini karşılamak hem de gerçekleştirilen karşılaştırmanın güvenilirliğini arttırmak üzere yazarların sözcük seçimleri, yazımları ve üslupları; bu çalışma kapsamında gerçekleştirilen karşılaştırmanın ikinci başlığını oluşturmaktadır.

Aşağıdaki tabloda, sözcük ve bağlam yönünden görgül olarak yapılacak karşılaştırma için rastgele seçilen mesajlara yer verilmiştir.

Tablo 4

Sözcük ve Bağlam Yönünden Karşılaştırma İçin Seçilen Üç Farklı Mesaj

Mesaj	İçerik
M0	“farkında olmadan her şeyi yalnız yapmaya alışmışım aq beni niye bu kadar yalnız bıraktınız orospu çocukları”
M1	“kime mikrofon uzatsalar kırgınım diyo aq biz de kırgınız da banka hesabımızda 10 milyar yuromuz olmadan kırgınız”
M2	“Dev Polonya bayrağı açtılar. Dikmesi baya zor olmuştur amk dhsjsjs”

Yukarıda sunulan mesajlara bakıldığında, öncelikle sözcük düzeyinde ölçünlü dilin dışındaki ifadeler ile küfürlere yer verilmesi gerektiği görülmektedir. Yapıca bağımsız tümcelerin bir araya getirilişi de yazar karakteristiğine ilişkin kayda değer bir veri sunmaktadır.

- **M0:** “aq” (a**na koyayım), “orospu çocukları”
- **M1:** “diyo”, “aq” (a**na koyayım), “milyor”, “yuro”
- **M2:** “baya”, “amk” (a**na koyayım), “dhsjsjs”

M0 mesajında, birbiriyle anlam ilişkisi bulunan iki tümce kullanılmış fakat bağlaç kullanılmamıştır. Bağlaç kullanılması gereken yerde küfür (“aq”) olduğu görülmektedir. Mesajda “yalnız” sözcüğüne iki kez yer verilmiştir ve mesajın alıcısı belirsizdir. Yazarın kime seslendiği bilinmemektedir.

M1 mesajında, anlam ilişkisi içerisinde üç tümce kullanılmış, tümceler arasında bağlaca yer verilmemiştir. Bağlaç ihtiyacı bulunan iki tümce arasında küfür (“aq”) kullanılmıştır. Bu mesajda, “kırgınız” sözcüğü iki defa görülmektedir. Mesajın belirli bir alıcısı, diğer bir deyişle yazarın seslendiği bir kimse bulunmamaktadır. Bununla birlikte mesaj içeriğinde işaret edilen kişilerin kim oldukları belirsizdir.

M2 mesajında birbiriyle anlam ilişkisi bulunan iki ayrı tümce kullanılmış ve tümceler nokta ile birbirinden ayrılmıştır. Mesajdaki hiçbir sözcük birden fazla kez kullanılmamıştır. Mesajın yöneltildiği bir alıcı bulunmamakla birlikte, ilk tümcenin öznesi belirsizdir. Mesajın sonunda küfür (“amk”) kullanılmış ve internet ortamında “random gülüş” olarak tabir edilen biçimde gülünmüştür.

Tablo 5

M0 Karşısında M1 ve M2 Mesajlarının Benzerlikleri İle M2'nin Farkları

M1	M2
Benzerlikler (7 adet)	Benzerlikler (3 adet)
<ul style="list-style-type: none">• Mesaj alıcısı belirsizdir.• Özne içeriği belirsizdir.• Küfür aynıdır.• Küfür, bağlaç gereken yerdedir.• İki tümce arasında bağlaca yer verilmemiştir.• Küfür, aynı şekilde yazılmıştır.• Bir sözcük, iki defa kullanılmıştır.	<ul style="list-style-type: none">• Mesaj alıcısı belirsizdir.• Özne içeriği belirsizdir.• Küfür aynıdır.
	Farklar (4 adet)
	<ul style="list-style-type: none">• Küfür, cümlelerin sonundadır.• İki tümce arasında bağlaca ihtiyaç yoktur, nokta kullanılmıştır.• Küfür, farklı şekilde yazılmıştır.• Tekrarlanan sözcük bulunmamaktadır.

M0 ve M1 mesajlarında kızgınlık veya öfke gibi olumsuz tepkiler söz konusudur ve bu olumsuz tepkinin belirsiz ve genel bir kitleye yöneltildiği görülmektedir. M2 mesajında ise bir olay ile dalga geçilmekte, olumsuz bir tepkiye yer verilmemektedir. M1 ve M2 mesajlarında bazı sözcükler, konuşma dilindeki sesletimiyle yazılmıştır ancak sorgulanan M0 metninde böyle bir örnek bulunmamaktadır. M0 ve M1 mesajlarının ikisinde de görülen küfür içerikli sözcük (“aq”) türetiminin, iki sözcükten oluşan küfrün baş harflerinin bir araya getirilmesi ve ikinci harfin sonradan, benzer başka bir sesin simgesine (harfine) dönüştürülmesiyle gerçekleştirildiği görülmektedir. Buna karşın, M2 mesajında da aynı küfür kullanılmış

olmasına karşın; iki sözcükten birincisinin ilk iki harfi, ikinci sözcüğün ise yalnızca ilk harfi seçilerek bir araya getirilmiştir (“amk”).

Sözcük ve bağlam özellikleri, yukarıda yer verilen fark ve benzerlikler ışığında dikkate alındığında; M0 mesajının, M1 mesajının yazarı tarafından yayınlanmış olması daha muhtemel görünmektedir. Yapılan kontrolde ise M0 ve M1 mesajlarının aynı yazara ait olduğu görülmüştür.

Muhtemelen karakter kısıtlamasından dolayı bütün örnekleme yer alan sözcük türleri (isim, eylem, sıfat, bağlaç vb.) ile öbekleri, varlıkları ve birbirlerine olan uzaklıkları (aralarındaki sözcük sayısı) bakımından anlamlı ve ayırt edici bir fark görülmemiştir. İletilerin örnek ve örnekçe oranları (farklı sözcük sayısının toplam sözcük sayısına oranı) da yazarların tespitine ilişkin kullanışlı bir veri sunmamıştır.

Değerlendirme ve Sonuç

İnternetin hızla yaygınlaşmasının ardından sosyal medya ortamlarının ve kullanımının da artması ile geleneksel metinlerdekinden farklı dilsel alışkanlıklar içeren yazılarla karşılaşılmaktadır. Sosyal medya ortamlarındaki metinler için yazar tanıma çalışmalarına başta tehdit (Türk Ceza Kanunu, m. 106) ve hakaret (TCK, m. 125) olmak üzere; devletin güvenliğine ve siyasal yararlarına ilişkin bilgileri açıklama (TCK, m. 329), gizli kalması gereken bilgileri açıklama (TCK, m. 330), yasaklanan bilgileri açıklama (TCK, m. 336) gibi pek çok çeşitli suç ya da başka hukukî uyuşmazlıklar kapsamında ihtiyaç duyulabilmektedir. Zira bireylerin yüz yüze söyleyemeyecekleri şeyleri internet ortamında rahatlıkla ifade edebildikleri bilinmektedir (Yıldırım, Özdemir & Alparslan, 2018, s. 46) ve bilişim uzmanlarının yapacağı çalışmalar, metinlerin yazıldığı cihaz ile elektronik ve fiziksel adresleri tam olarak tespit etmekle sonuçlanacaksa da metnin yazarına ilişkin net bir bilgi sunmayacaktır. Diğer yandan adli bilimler, ceza muhakemesi hukukuna paralel olarak (Toroslu & Feyzioğlu, 2016, s. 7), maddi gerçeği araştırarak ortaya çıkarmaya çalışmaktadır.

Yukarıda iki ayrı temel başlık altında, Twitter isimli mikro-blog sitesindeki mesajların yazarlarını tespit etmeye yönelik bir çalışma gerçekleştirilmiştir. Yapılan iki incelemede de, sorgulanan metnin, diğer metinler karşısındaki fark ve benzerlikleri ortaya konularak, karşılaştırma metinleri üzerinden yazarı tespit edilmektedir. Bu çalışma kapsamında, daha önce de değinildiği gibi yazar tanıma üzerine tek ve genel-geçer uygulanabilir bir yöntem olmadığından, sorgulanan metinler ile karşılaştırma metinleri arasında görülen ayrılık ve benzerliklerden yola çıkılarak, deneysel biçimde kontrol ölçütleri ortaya konulmuştur. Ayrıca,

gerçekleştirilen çalışmadaki sorgulanan metinler bağlamında kullanılsa dahi, ayırt edici olacağı öngörülen başka birtakım ölçütlerden de bahsedilmiştir. Her ne kadar söz konusu ölçütlerin kesin olarak her bir yazar tanıma incelemesinde ayırt edici olmayacağı öngörüsünde bulunulabilecek olursa da, bu ölçütlerin üst başlıklarının her karşılaştırmada faydalı olacağı değerlendirilmektedir.

Ölçünlü dil kuralları dâhilindeki benzerliklerin karşılaştırmaya esas alınmaması, yazar tanıma çalışmalarının en temel kurallarından biridir. Elektronik ortamdaki yazar tanıma çalışmaları kapsamında ise dilsel davranışın, farklı aygıtlar (bilgisayar, tablet, akıllı telefon gibi) üzerinde yazarın kendisi veya aygıt ayarları dolayısıyla birtakım değişiklikler gösterebileceği unutulmamalıdır. Bu durumda yazar tanıma çalışmasının sağlıklı biçimde yapılabilmesi için, aynı aygıtlar üzerinden yazılmış aynı türde metinler incelenmelidir.

Kaynakça

- Amasyalı, M. F. ve Diri, B. (2006). Automatic Turkish text categorization in terms of author, genre and gender. *11th International Conference on Applications of Natural Language to Information Systems* içinde (s. 221-226). Berlin, Heidelberg: Springer.
- Brückner, T. (2011). Gibt es einen "sprachlichen Fingerabdruck"?: Kritische Anmerkungen zum forensischen Textvergleich. *Sprachwissenschaftliche Kriminalistik und Sprachprofilung: Anfänge, Kontroversen, Meilensteine, Fallbeispiele* içinde (s. 49-54). Graz: Grazer Linguistische Monographien.
- Coulthard, M. (2005). Some forensic applications of descriptive linguistics. *Veredas - Revista de Estudos Linguísticos*, 9, 9-28.
- Coulthard, M. ve Johnson, A. (2007). *An Introduction to Forensic Linguistics: Language in Evidence*. London, New York: Routledge.
- Doğan, S. ve Diri, B. (2010). Türkçe dokümanlar için N-gram tabanlı yeni bir sınıflandırma (Ng-ind): yazar, tür ve cinsiyet. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 3(1), 11-19.
- Grant, T. (2008). Approaching questions in forensic authorship analysis. *Dimensions of Forensic Linguistics* içinde (s. 215-229). Amsterdam, Philadelphia: John Benjamins Publishing Company.
- MacLeod, N. ve Grant, T. (2012). Whose Tweet? Authorship analysis of micro-blogs and other short-form messages. *Proceedings of The International Association of Forensic Linguists' Tenth Biennial Conference* içinde (s. 210-224). Birmingham: Aston University.
- McMenamin, G. R. (2002). Linguistic variation. *Forensic Linguistics: Advances in Forensic Stylistics* içinde (s. 44-65). Boca Raton, London, New York, Washington D.C.: CRC Press.
- Olsson, J. (2008). *Forensic linguistics* (2. baskı). London, New York: Continuum.
- Svartvik, J. (1968). *The Evans statements: a case for forensic linguistics*. Göteborg: University of Göteborg.

Tanrıvere, U. (2018). Kısa Metinlerde Yazar Tanıma: Twitter İçin Bir Yöntem Önerisi. *Humanitas*, 6(12), 177-192

Tanrıvere, U. (2016). *İntihar Mektuplarının Psikodilbilimsel İçerik ve Tür Özellikleri Bağlamında İncelenmesi: Adli Dilbilimsel Bir Yaklaşım*. (Yayımlanmamış Yüksek Lisans Tezi). Hacettepe Üniversitesi, Ankara.

Taş, T. ve Görür, A. K. (2007). Author identification for Turkish texts. *Çankaya Üniversitesi Fen-Edebiyat Fakültesi, Journal of Arts and Sciences*, 7, 151-161.

Toroslu, N. ve Feyzioğlu, M. (2016). *Ceza Muhakemesi Hukuku* (16. Baskı). Ankara: Savaş Yayınevi.

Türk Ceza Kanunu (TCK). (2004). <http://www.mevzuat.gov.tr/MevzuatMetin/1.5.5237.pdf>

Yıldırım Ş., Özdemir M. ve Alparslan E. (2018). Kullanımlar ve doyumlar kuramı çerçevesinde bir sosyal paylaşım ağı incelemesi: Facebook örneği. *Intermedia International e-Journal*, 5 (8), 42-65.