



Türkçe Kısa Metinlerde Yazar Tanıma Yönteminin Sınanması ve Geliştirilmesi

Testing and Developing Author Identification Method for Short Texts in Turkish

Utku TANRIVERE*

Öz

Yazar tanıma, adli dilbilimin belki de en önemli uygulamalarından bir tanesidir. Türkçe alan yazınına bakıldığında ise bu konudaki çalışmaların nispeten kısa bir süredir yapıldığı ve bu çalışmalarda çok kısa metinlerin kullanılmadığı görülmektedir (Amasyalı ve Diri, 2006; Taş ve Görür, 2007; Doğan ve Diri, 2010). Oysaki adli dilbilim çerçevesinde sadece birkaç sözcükten ibaret olan metinler dahi bir inceleme konusu olabilmektedir. Geçmişte, Twitter isimli mikro-blog sitesinde yayınlanan metinler (tweet) üzerinde yapılan bir çalışma neticesinde Türkçe alan yazınında bir yöntem önerilmiştir (Tanrıvere, 2018). Bu çalışma kapsamında ise söz konusu yöntem, yine Twitter iletilerinden oluşan yeni bir bütüne ile sınanmış ve yöntemi oluşturan ölçütler sayıca artırılıp nitel bağlamda da detaylandırılarak geliştirilmeye çalışılmıştır. Yöntem, yazarının kim olduğu bilinmeyen iletilerin yazarı bilinen iletilerle karşılaştırılması neticesinde olumlu sonuçlar vermiştir. Uygulama sürecinde öngörülmemiş başka birtakım ayırt edici ölçütlerin de varlığı gözlenmiş ve bunlar detaylıca sunulularak, yöntemdeki ilgili ölçüt gruplarına dahil edilmeleri önerilmiştir. Elde edilen sonuçlar genel olarak, çok kısa metinlerde dahi yazar tespiti mümkün olduğunu ve yöntemin adli bağlamlarda da kullanılabileceğini göstermektedir.

Anahtar sözcükler: Adli dilbilim, yazar tanıma, yazar tespiti, birey dil, kısa metinler, sosyal medya.

Abstract

Author identification is probably one of the most important applications of forensic linguistics. Considering scientific literature in Turkish language, related studies have been carried out relatively recently and very short texts have not been used in these studies (Amasyalı & Diri, 2006; Taş & Görür, 2007; Doğan & Diri, 2010). However, texts involving only a few words may also be a subject of study in terms of forensic linguistics. In the past, as a result of a study on published texts (tweets) on a micro-blog site “Twitter”, a method in the scientific literature in the Turkish language had been proposed (Tanrıvere, 2018). In the present study, however, that very method has been tested with a new corpus consisting of tweets again, and the markers within the method were developed by increasing the number and making elaborations qualitatively. This method has brought about successful results by comparing texts with known and unknown authors. In the implementation process, there have been some distinctive but unpredicted markers. These were also given in detail and proposed to be added to the relevant marker group within the method. The results of the study have shown in general author identification is possible even in very short texts and the method may also be used in forensic contexts.

Keywords: Forensic linguistics, author identification, authorship attribution, idiolect, short texts, social media.

* Hacettepe Üniversitesi Adli Bilimler Anabilim Dalı doktora adayı, filolog. E-posta: utkutanrivere@hacettepe.edu.tr, ORCID: 0000-0003-2534-1701

Giriş

Soruşturma ve kovuşturma makamları tarafından adli dilbilim uzmanlarına yöneltilen talepler dikkate alındığında yazar tanıma, adli dilbilimin en önemli konularından biri olarak ortaya çıkmaktadır (Coulthard, 2005, s. 10). Birden fazla metin üzerinde yapılan incelemelerle yazarı aynı ve farklı olanların tespit edilmesi yoluyla gerçekleştirilen yazar tanıma faaliyetleri; başta fikrî mülkiyet hakkı ihlalleri ve dolayısıyla intihaller olmak üzere tehdit, hakaret, cinsel taciz ya da devlet sırlarının açıklanmasına benzer doğrudan yazılı dil aracılığıyla işlenebilecek suçların aydınlatılmasında rol oynayabileceği gibi başka birtakım suç ve hukukî uyuşmazlık konularıyla ilgili olarak da maddî gerçeğin ortaya çıkmasına katkı sağlayabilir. Bir kişinin hürriyetinden alıkonulduğu mekânda bulunan ve fail tarafından yazılmış bir not ya da bir yazılımın kaynak kodu içerisinde yer alan notların (source code comment) incelenmesi bu noktada örnek olarak gösterilebilir.

Yazar tanıma uygulamaları, mevcut yazılı dilsel veri bağlamında belirli birtakım koşullar altında gerçekleştirilmektedir. Bu nedenledir ki Grant (2008, ss. 216, 227), yazar tanıma çalışmaları kapsamında tek, belirli ve her koşulda, her türlü dilsel veriye uygulanabilecek bir yöntem ya da yaklaşım olmadığını ifade eder. Teknolojinin giderek gelişmesiyle birlikte internet erişiminin ve sosyal medya kullanımının kayda değer bir hızla artması, birtakım suç ve uyuşmazlık konularını bu platformlara da taşımış ve neticesinde bu ortamlardaki dilsel veri üzerinde de çalışma yapılabilmesini mümkün kılacak yeni yöntem ve yaklaşımlara gereksinim duyulmaya başlanmıştır. Twitter adlı mikro-blog sitesi ise kullanıcıların tek seferde yazabilecekleri iletileri 280 karakterle sınırlaması bakımından tüm sosyal medya platformları içerisinde adli dilbilim ve yazar tanıma bağlamında ayrı bir yere sahiptir. Zira, uzunluk yönünden herhangi bir kısıtlama bulunmadan yazılmış olan metinler ile 280 karakter gibi nispeten kısıtlı uzunluktaki metinler üzerinde uygulanabilecek yazar tanıma yöntemlerinin farklılık arz edeceği açıktır.

Uzunluk yönünden herhangi bir kısıtlamanın bulunmadığı ve görece fazla dilsel veri içeren metinlerde cümle uzunluğu, sözcük türleri, tekrar eden sözcük oranı gibi stilometrik ölçümler başarıyla uygulanabilse de gerek metin uzunluğuna getirilen kısıtlama gerekse adli dilbilim uygulamaları genelinde incelemeye konu edilen metinlerin oldukça kısa olması (Coulthard ve Johnson, 2007, ss. 162, 172), bahsi geçen stilometrik yöntemlerle istatistiksel açıdan anlamlı karşılaştırmalar yapmayı zorlaştırmaktadır (Tanrıvere, 2018, ss. 187, 189). Bu çalışmanın amacı ise, Türkçe kısa metinlerin (tweet) incelenmesi üzerine daha önce önerilmiş adli dilbilimsel bir yöntemin (Tanrıvere, 2018) geliştirilmesi, farklı hassasiyet düzeylerinde sınanması ve uygulamanın verimliliği ile birlikte kısıtlılıkları bağlamında sonuçların ortaya konulmasıdır.

Kuramsal Çerçeve

Verili birden fazla metnin arasında yazarları aynı ve farklı olanları tespit etmeye dair geliştirilen yöntemlerin temelinde “değişke” (*variety*) ve “lehçe” (*dialect*) kavramları yer almaktadır. Bu kavramlar, coğrafi bölge ve sosyal grup gibi farklı düzeylerde ortaya çıkarak belirtilen katmanlar çerçevesinde görülen farklı dil kullanımlarına işaret eder (Wardhaugh, 1998, s. 21). Bu noktada, coğrafi bölge ve sosyal gruptan çok daha özel bir düzlemde “kişisel lehçe” (*individual dialect; idiolect*) olarak da ifade edilen “bireydil” kavramı ortaya çıkar.

Yaygın olarak bilinmektedir ki, bireylerin buldukları bölge veya yer aldıkları sosyal gruplar gibi, kendi kişisel yaşam geçmişleri ve bu bağlamda oluşturdukları, sahip oldukları genel algıları da dilsel üsluplarını şekillendirmektedir. Bireydil, bir konuşucu veya yazarın kendi konuşması ya da metni ile başka kimselerin konuşma veya metinleri arasındaki farklılığa; ayırt edici ve özgün dil kullanımına gönderme yapmaktadır (Coulthard, 1998, 2004; Grant ve MacLeod, 2018; Olsson, 2008). Farklı kimselerin dil kullanımlarının birbiriyle tamamen örtüşmesinin mümkün olmadığı görüşü de bu kavramın temellerine dayandırılmaktadır (McMenamin, 2002, s. 62).

Olsson (2008, s. 33), yukarıda açıklanan bireydil kavramının bir sonucu olarak farklı kimselerin yansıtacakları farklı dil kullanımlarını “yazarlararası değişke” (*inter-author variation*) olarak nitelemektedir. Ne var ki iletişim sürecine katılan kimseler arasındaki sosyal statü veya yaş gibi farklılıklar (Wardhaugh, 1998, s. 48) ya da metin türü, metnin özel veya herkese açık oluşu, iki metin arasında geçen

süre (Olsson, 2008, ss. 33-34) gibi birtakım etkenler, aynı kişinin konuşma veya metinleri arasında farklılıklar görülmesine de sebep olabilecek niteliktedir. Bu bağlamda Olsson (2008, s. 33), bir kişiye ait üslup değişkenliğini ise “tek-yazar değişkesi” (*intra-author variation*) olarak ifade etmektedir.

Bütüncül olarak yazar tanıma çalışmaları da genel bir bakış açısıyla bu değişkelerin, daha özel bir yaklaşımla da yazarlararası değişkelerin tespit edilmesine odaklanmaktadır. İncelemeye alınan metinler arasında yazarlararası değişke görülmeysi, metinlerin aynı kişi tarafından yazıldığı; varlığı ise farklı kişiler tarafından yazıldığı görüşünün dilbilimsel bir gerekçeyle ortaya konulmasını sağlamaktadır.

İlgili Alanyazın

Özellikle İngilizce alan yazınına bakıldığında hem genel bağlamda metinler hem de doğrudan kısa metinler çerçevesinde gerek dilbilimciler gerekse bilgisayar bilimciler tarafından çok çeşitli yazar tanıma çalışmalarının yapıldığı görülmektedir (Amasyalı ve Diri, 2006; Bhargava, Mehndiratta ve Asawa, 2013; Coulthard, 1998, 2004; Grant, 2008; MacLeod ve Grant, 2012; Taş ve Görür, 2007). Ne var ki, özellikle bilgisayar bilimcilerin kısa metinler özelinde gerçekleştirdikleri stilometrik çalışmalar, adli dilbilim bağlamında karşılaşıldığı belirtilen metinlerin metin uzunluğu yönünden niteliği ve dolayısıyla ihtiyaç duyulan yaklaşımdan uzak görünmektedir.

Örneğin Twitter üzerinde stilometrik yöntemlerle gerçekleştirilen bir yazar tanıma uygulaması kapsamında sözcüksel ve sözdizimsel unsurlar ile birlikte Twitter sitesine özgü özellikler ve duygu belirtmeye yarayan diğer işaretler (*emoji*) olmak üzere dört başlık içeren bir çalışma yapılmıştır (Bhargava ve diğerleri, 2013): Dört başlıkta yer alan ölçütler birlikte kullanıldığında yazar tanıma yöntemi 10 ileti içeren bir grupta %91,11 gibi önemli bir başarı sağlamışsa da çalışmanın her bir yazardan 300 farklı ileti (*tweet*) seçildikten sonra, 5-6 sözcük içeren kısa iletilerin hariç tutularak gerçekleştirildiği görülmektedir.¹ Dolayısıyla örneklem, yukarıda ifade edildiği gibi adli dilbilim yazını bağlamında, kısa metinler üzerine gerçekleştirilen yazar tanıma uygulamalarına ilişkin ihtiyaca karşılık vermemektedir.

Doğrudan adli dilbilim alanına katkıda bulunan yazarlar olarak MacLeod ve Grant (2012) ise farklı hassasiyet düzeylerinde, daha önce değinilen adli bağlama çok daha yakın bir uygulama gerçekleştirmişler. Bu çalışma kapsamında tekil iletiler için yapılan karşılaştırmada %90’a, gruplandırılmış birden fazla ileti için yapılan uygulamada ise %100’e varan sonuçlar elde edilmiş olmakla beraber çalışmanın hiçbir safhasında bir ileti yanlış bir yazara atfedilmemiş, diğer bir deyişle hiçbir durumda “hatalı pozitif” sonuç alınmamıştır (MacLeod ve Grant, 2012, s. 219-221). Uzmanlar, yazar tanıma için kullandıkları ölçütleri dilbilgisi, sözcük, noktalama ve üretim biçimine özgü diğer özellikler olarak dört ana başlık altında sunmaktadır (MacLeod ve Grant, 2012, s. 217-218). Mevcut çalışmada sınanmak istenen yöntem de oldukça yakın bir ölçüt gruplandırması önermektedir (Tanrıvere, 2018, s. 184).

Türkçe örneklem ile gerçekleştirilen çalışmalara bakıldığında bilgisayar bilimcilerin, yazar tanıma çalışmalarına dilbilimcilerden daha fazla ilgi duyduğu görülmekteyse de, İngilizce alan yazını için değinilen yöntem ve veri niteliği, bu çalışmalar genelinde de geçerliliğini korumaktadır (Tanrıvere, 2018, s. 179-180). Yazın kapsamında istatistik temelli (stilometrik) ölçüm yöntemleri, örneklem olarak gazetelerin köşe yazıları üzerinde uygulanmış ve bu yazıların köşe yazarlarıyla eşleştirilmeleri sağlanmıştır (Amasyalı ve Diri, 2006, s. 222, 224; Doğan ve Diri, 2010, s. 11-12; Taş ve Görür, 2007, s. 153-154). Nispeten daha yakın bir zamanda ise on köşe yazarının yazmış olduğu 50’şer köşe yazısı ile aynı yazarların kendi Twitter hesaplarında paylaştıkları 200’er ileti bir çalışmaya konu edilmiş, Twitter ortamında paylaşılmış her bir iletinin, tek tek, bir köşe yazısıyla eşleştirilmeye çalışılmasındansa, bu iletilerin bir araya getirilmesiyle oluşturulan bir grubun karşılaştırmaya konu edilmesinin, uygulama kapsamında daha verimli olacağı görüşü ifade edilmiştir (Mayda ve Amasyalı, 2016).

¹ Yazar tanıma için önerilen yöntemlerin başarılı olması adına çok kısa metinlerin çalışmadan hariç tutulması oldukça anlaşılardır. Bununla birlikte beş ya da altı sözcük içeren metinlerin de adli dilbilim uygulamaları kapsamında yazar tanıma faaliyetine konu olması muhtemeldir. Bu çalışma ile sınanacak yöntem için, en az üç sözcük içeren iletiler seçilecek, yalnızca bir veya iki sözcük içeren iletiler hariç tutulacaktır.

Değınilen alıřmalar Trke baėlamında nerilen yazar tanıma uygulamaları olarak olduka kayda deėer olsalar da, kısa metinlerin incelenmesinde aynı derecede geerli olduklarını sylemek mmkn olmamaktadır. Benzer řekilde, İngilizce alan yazınında da yazar tanıma konusundaki yntemlerin oėunun, kısa metinlerde yazar tanıma uygulaması iin uygun olmadıkları belirtilmiřtir (MacLeod ve Grant, 2012, s. 210).

Bu alıřma erevesinde geliřtirilmesi ve sınanması hedeflenen yntem ise doėrudan adli dilbilimsel yazar tanıma uygulamaları iin, Twitter adlı mikro-blog sitesinden seilen rneklemenin incelenmesi ile yakın bir zamanda nerilmiřtir (Tanrıvere, 2018, s. 177). Twitter zerinde iletilerini “herkese aık” (*public*) olarak paylařan rastgele 10 hesaptaki yine rastgele 5’er adet olmak zere toplam 50 ileti incelenmiř ve farklı yazarların oluřturdukları iletiler arasında grlen noktalama iřareti veya byk ve kk harf kullanımı gibi farklılıklar, karřılařtırmaya ynelik bir lte dnřtrlerek sunulmuřtur (Tanrıvere, 2018, ss. 180-181). Sz konusu ltler yntem bařlıėı altında detaylandırılmaktadır. Ayrıca alıřma kapsamında incelemeye konu edilen iletilerin tamamı, Twitter kullanıcıları tarafından Kasım 2017’den nce paylařılmıřtır (Tanrıvere, 2018, s. 181) ve belirtilen tarihten nce Twitter’de oluřturulabilecek iletilerin uzunluėu gnmzdeki gibi 280 deėil, 140 karakter ile sınırlıdır. Dolayısıyla burada sınanacak olan yntem, 140 karakter ile sınırlı iletilerin incelenmesiyle nerilmiřtir. Bu durumun, mevcut alıřma kapsamında herhangi bir olumsuzluk yaratmayacaėı deėerlendirilmekte, zira uzun metinler iin geliřtirilen yazar tanıma yntemleri kısa metinlerde kullanıřlı olmasa da bunun aksinin herhangi bir fark yaratmayacaėı varsayılmaktadır.

Veri

ncelikle Twitter zerinde, ařaėıda belirtilen varsayım, nitelik ve kısıtlamalar kapsamında, 50 farklı hesaptan paylařılan 20’řer olmak zere toplam 1000 adet ileti rastgele biimde toplanmıřtır. Elde edilen bu 1000 gnderinin ierisinden de, karřılařtırma uygulamasında kullanılmak zere 138 farklı ileti yine rastgele tespit edilmiřtir. Bu 138 ileti; 2’řer yazardan ve 5’er yazardan, her yazardan 1’er ve 5’er karřılařtırma metinli olmak zere seilmiřtir. İletiler, her bir grupta 3’er kez tekrarlanmak zere toplam 12 tekil uygulamada kullanılmak zere gruplandırılmıřtır.

İncelemeye alınan iletilerin paylařıldıkları tarihte retildikleri ve bir hesap aracılıėıyla paylařılmıř tm iletilerin aynı aygıt aracılıėıyla yazıldıėı varsayılmaktadır. İletilerdeki dil kullanımını etkileyebilecek olan aygıtlarla ilgili bu varsayım, aynı zamanda mevcut alıřma kapsamında olası “etki karıřımı” konusunu oluřturmaktadır. Zira aynı yazarın farklı aygıtlar zerinde oluřturduėu iletiler incelendiėinde, yazarların ayırt edilmesinde kullanılabilir bazı ltlerde hatalı sonular alınması sz konusu olabilecektir. (Tanrıvere, 2018, ss. 184, 190). Bununla birlikte iletileri karřılařtırılacak her bir Twitter hesabının yalnızca bir yazarının olduėu ve ayrıca her bir yazarın sadece bir hesaptan paylařım yaptıėı, incelenen iletilerin yayımlandıėı birden fazla hesapta aynı yazarın yer almadıėı varsayılmaktadır. Son olarak alıřma kapsamında yazarların, Trkenin anadil konuřucusu oldukları ve paylařılan iletilerin, kendi rnleri olduėu varsayımları da yer almaktadır.

rneklem iin Twitter hesaplarındaki paylařımlar toplanırken, aynı yazarların iletileri arasında en az iki gnlk bir sre bulunması gzetilmiř olup bylelikle birden ok karřılařtırma metniyle yapılacak deneyler iin yazarların anlık deėiřkenlerden olabildiėince az etkilenmiř yerleřik dilsel sluplarının tespit edilmesi hedeflenmiřtir. Yazarların kendilerine ait olmayan zl szler ve sair alıntılar, alıřmanın dıřında tutulmuřtur. Twitter’de bir ileti iin daha nce 140 olan karakter kısıtlaması, Kasım 2017’den sonra 280’e ykseltilmiř ve bu nedenle belirtilen tarihten nceki paylařımlar rnekleme dahil edilmemiřtir. Metin uzunluėu ynnden 2’den fazla szck ieren bu paylařımlar, yalnızca kendileriyle sınırlı olup herhangi bir medya (harici bir internet sitesine baėlantı, fotoėraf veya video, bařka bir kimsenin yaptıėı paylařım vb.) iermemektedir. alıřmanın niteliėi gereėi eřitli kurum ve kuruluřların ya da basın danıřmanı, szc, iletiřim kiřisi gibi birtakım kimselerin paylařımda bulunduėu tanınmıř kiřilerin hesaplarından yapılan paylařımlar rneklemden hari tutulmuřtur.

Derlenmiř olan iletilerin tamamı, veri toplama alıřmasının gerekleřtirildiėi tarihlerde “herkese aık” olarak yayındadır. Twitter (2018), gizlilik politikası ile sayfasında, herkese aık olarak yapılan

paylaşımların üçüncü taraflar tarafından işlenip kullanılabileceği konusunda kullanıcılarını bilgilendirmekte, üye olan ve üyeliğini sürdüren kullanıcılar da bu şartları okumuş ve kabullenmiş sayılmaktadır.

Yöntem

Geliştirilmesi hedeflenen ve bununla beraber verimliliği, olumlu ve olumsuz yönleri değerlendirilecek olan ölçütlerin üç temel başlık altında gruplandırılması önerilmektedir (Tanrıvere, 2018, s. 184): “Yazım ve noktalama” başlığı altında iletilerin biçimsel boyutu, “sözcük ve bağlam” başlığı altında bu iletilerin içeriksel yönleri ve doğrudan bu şekilde adlandırılmamış da olsa, ayrıca gruplandırılması gerektiği belirtilmiş olan “diğer ölçütler” bağlamında ise kullanımda olan teknoloji ve sosyal medya platformunun niteliklerine özgü biçimde konu etiketleri (*hashtag*) ve duygu belirtmeye yarayan yüz ifadeleri (*emoticon*) ile benzeri diğer işaretler ele alınmaktadır.

Yazım ve Noktalama

Yazım ve noktalama bakımından karşılaştırmada kullanılabileceği ifade edilen ölçütler, noktalama ve yazım yanlışları olmak üzere iki temel alt başlık içermektedir (Tanrıvere, 2018, s. 186).

Noktalama işaretleri sırasıyla “konum”, “sayı”, “parantezler” ve “tırnak işaretleri” olmak üzere dört farklı ölçüt grubuna ayrılmakta ve noktalama işaretlerinin cümle içinde, cümle sonunda, tırnak içindeki cümle sonunda ve ileti sonunda kullanımı; tek işaret, birden çok işaret ve birbirinden farklı işaretler hâlinde kullanımları; parantezlerin şekil yönünden normal parantez, köşeli parantez, küme parantezi veya büyüktür-küçüktür simgeleriyle oluşturulması ve tırnak işaretlerinin tekli veya çift simge ile kullanımı ayrı birer ölçüt olarak belirtilmektedir (Tanrıvere, 2018, s. 186).

Rastgele seçilmiş ve yazarı bilinmeyen 1 ileti (buradan itibaren “karşılaştırma iletisi” olarak adlandırılacaktır) ile yazarı bilinen 12 farklı iletinin (buradan itibaren “kontrol iletisi” olarak adlandırılacaktır) dâhil olduğu bir uygulamada, yalnızca cümle sonundaki noktalama işaretleriyle birlikte cümlelerin başında büyük veya küçük harf kullanımının değerlendirilmesinin yeterli olduğu belirtilmiştir (Tanrıvere, 2018, s. 183). Buradan anlaşılmaktadır ki yöntem, büyük ve küçük harf kullanımına dair tespitleri, biçimsel ölçütlerin ikinci ana başlığı olan “yazım yanlışları” kapsamına almaktadır. “Konum” ve “sözcük” olarak iki alt başlık içeren yazım yanlışları bağlamında ayrıca, söz konusu yanlışın; münferiden yapılan bir hata mı, sürekli yapılan bir hata mı, yoksa yazarın kasten tercih ettiği bir üslup eğilimi mi olduğu önem taşımaktadır.

İlk durumda, yani yazarın yanlış bilgisinden veya dil kullanım eğiliminden kaynaklanmayan münferit bir hatayla karşılaşıldığında bu durumun, tam olarak tespit edilen yanlıştan önce ve sonra klavyede basılan tuşun ya da tuşların tespit edilmesiyle anlaşılacağı öngörülmektedir. Örneğin “Kısa metinler” yazmak isteyen bir kimsenin “Kİsa metinler”, “Kıssa metinler”, “Kısa metyinler” veya “Kısa merinler” yazmış olması, münferit birer hata olarak görülebilir. İlk örnekte klavyedeki *shift* tuşunu geç bırakan yazar, ikinci örnekte bir tuşa iki kere basmış, üçüncü örnekte “t” harfinden sonra Türkçe klavyede sağında bulunan “y” tuşuna da basmış, son örnekte ise “t” tuşu yerine Türkçe klavyede solunda bulunan “r” tuşuna basmıştır. Ayrıca noktalama işaretleri yönünden de “münferit hata” olarak değerlendirilmesi gereken bu yazımlarla, yazarın bir sonraki metninde muhtemelen karşılaşmayacağı varsayılmaktadır. Bu nedenle münferit hatalar, yazarlar bakımından ayırt edici olarak ele alınmamalıdır.

Diğer yandan süregelen hatalar ile tercih edilen üslubun içerdiği ve ölçünlü dilden ayrılan dil kullanımları, yazarın muhtemelen her metninde yer alabilecektir. Bilinçli bir eğilim olarak yer verilen kullanımların, yanlış dil bilgisi kaynaklı yapılan hatalardan farkı ise yazarın, herhangi bir nedenle farklı bir üslup tercih etmesi durumunda bu kullanımlarla karşılaşmayacağı varsayımıdır. Buna örnek olarak herhangi bir kimseye hitap etmeksizin, doğrudan kendisini takip eden diğer kullanıcılar geneline kendi vücut ağırlığıyla ilgili ileti yazan bir Twitter kullanıcısı sürekli “külo” yazımını tercih ederken, yazışma çerçevesinde kendisinin beslenme ve diyetetik uzmanı olduğu anlaşılan (ve hesap bilgilerinde bu yönde beyan bulunan) kişiye hitap ettiği (*mention*) iletilerde “kilo” yazmayı tercih ettiği görülmüştür. Dolayısıyla bilinçli eğilime dayalı biçimde ölçünlü dilden ayrı kullanımların varlığı, yazarın tercih ettiği üsluba, dolayısıyla iletişim sürecinde yer alan değişkenlere bağlı

görülmektedir. Yanlış bilgiye dayanan yazım hatalarında ise bu bilgi düzeltilmediği müddetçe yazar, aynı hatayı üslup seçiminden bağımsız olarak tekrarlamaya devam edecektir. Önerilmiş olan yöntemde, hata kavramından uzaklaşarak, yazarın bilinçli tercihi olduğu değerlendirilen yazımların, açıkça belirtilmese de örtülü biçimde sözcük ve bağlam yönünden yapılacak karşılaştırma içerisinde ele alındığı görülmektedir (Tanrıvere, 2018, s. 187).

Yazım yanlışları yönünden yapılan “konum” ve “sözcük” ayrımı; bir hatanın, iletinin belirli bir yerinde mi, yoksa belirli bir sözcük üzerinde mi tekrar ettiğini tespit etmeye yöneliktir. Örneğin “bas” yerine “bass” sözcüğünün sürekli kullanımı sözcük yönünden bir hatayken, cümlelere veya iletinin sadece ilk cümlesine küçük harfle başlamak ise konum yönünden bir hata olarak değerlendirilmektedir. Ayrıca, şimdide dek gerçekleştirilen önsel çalışmalarda sıkça karşılaşıldığı için belirtilmesi gerekli görülmektedir ki, sözcüklere getirilen eklerin yazımında görülebilecek hatalar konum, bağlaçların yazımında rastlanılan yanlışlar ise sözcük yönünden değerlendirilebilecek yazım hatalarıdır. Sözcüklerin arasında birden fazla boşluk bırakmak veya boşluk bırakmamak, bir cümleyi ardışık iki nokta ile sonlandırmak veya soru işareti ile sonlandırılması gereken cümlenin sonunda noktaya yer vermek de -münferit olmadığı sürece- konum yönünden yapılan hatalardır.

Sözcük ve Bağlam

Sözcük ve bağlam bakımından yapılacak karşılaştırma için belirli bir ölçütten bahsedilmemiş, verilen uygulama örneğinde ise “ölçünlü dilin dışındaki ifadeler”, “küfürler”, “yapıca bağımsız tümcelerin bir araya getirilişi” görgül ve betimleyici biçimde ele alınmıştır (Tanrıvere, 2018, s. 187). Bununla birlikte iletinin kime atfedildiğinin belirli ya da belirsiz oluşu, öznenin içerik yönünden açık ve anlaşılır oluşu, örnek metindeki sövme belirten yazımların içeriğinin yanı sıra cümle içerisinde buldukları konumlar ve oluşturulma biçimleri (morfolojisi), cümlelerde bağlaç kullanma durumu ve tekrar eden sözcük bulunup bulunmadığı da incelemeye konu edilmiştir (Tanrıvere, 2018, s. 188).

Söz konusu uygulama örneğinde “milyar” sözcüğü yerine “milyor”, “euro” yerine “yuro” yazımı da ele alınmış (Tanrıvere, 2018, s. 187) ve böylelikle yöntem bakımından bir farklılık doğmuştur. Daha önce sunulmuş olan bilgiler bakımından bu yazımların, ilgili başlık altında (yazım ve noktalama > yazım yanlışları > sözcük) değerlendirilmesi gerektiği düşünülse de yazımların niteliği göz önünde bulundurulduğunda bunların birer “hata” değil, “eğilim” olarak değerlendirildiği ve incelenen ileti yazarının tercih ettiği üslubun bir parçası olarak sözcük ve bağlam kapsamında ele alındığı anlaşılmaktadır.

Yer verilen uygulama ile birlikte sunulan bilgiler değerlendirildiğinde, sözcük ve bağlam özelliklerine dair yapılacak incelemenin, büyük ölçüde, karşılaştırmayı yapacak kişinin değişiklik ve geliştirmesine açık olduğu görülmektedir. Başka bir deyişle, sunulmuş olan ayırt edici ölçütlerin haricinde, verili metinler arasındaki farkların veya benzerliklerin başka hangi yollarla ortaya çıkarılabileceğine veya pekiştirilebileceğine dair tespit, incelemeyi yapan kişiye kalmaktadır. Şimdide dek yapılan önsel çalışmalar ve sosyal medyadaki gözlemler ise aşağıda belirtilen konuların, sözcük ve bağlam yönünden ayırt edici bir ölçüt olarak değerlendirmelerde ele alınması gerektiğini göstermiştir. Burada yer verilmeyen sözcük ve bağlam ölçütleri, her bir karşılaştırma (ve örneklem) özelinde incelemeyi yapacak kişi tarafından geliştirilmeli ve çeşitlendirilmelidir.

Yazarın tercih ettiği eğilim sonucu ölçünlü dilden uzaklaşan sözcük ve diğer ifadeler ile bu ifadelerin oluşturulma biçimi, ayırt etmede kullanılabilir niteliktedir. Ölçünlü dilden ayrılan sözcüğün yazar tarafından tercih edilen bir kullanım mı, yoksa münferit ya da süregelen bir hata mı olduğu, daha önce de belirtildiği gibi incelemeyi yapan kişiye bırakılmaktadır. Ne var ki bu üç seçenekten ilkinin elenmesi durumunda söz konusu sözcük, “yazım ve noktalama” kapsamında bir yazım yanlışlığı olarak ele alınacak, sonuç olarak başka başlık altında da olsa değerlendirmeye dahil olacaktır.

Yazılı bir iletide vurgu yaratmak için tercih edilen kullanımlar da önem arz etmektedir. Bir sözcüğün ya da sözcük öbeğinin büyük harfle, tırnak içinde, ünlemle veya parantez içinde ünlemle ya da vurgulandığını belirtecek başka herhangi bir şekilde yazımı, bu noktada ayırt edici olabilecek niteliktedir. Bunun yanı sıra iletide yabancı bir dildeki sözcüğün kullanılması, bu sözcüğün kaynak dili, terminoloji dahil olmak üzere jargon ve ayrıca argo kullanımı, bağlaçlara yer verilmesi ve bu bağlaçların anlam değeri (olumlu-olumlu ya da olumlu-olumsuz gibi), özellikle muhtemel yazarlara ait birden fazla karşılaştırma

iletisi bulunan durumlarda ileti içeriğinin kontrol iletisi karşısındaki konu benzerliği ile anlaşılır ve tutarlı oluşu da yazarları ayırt etmede kullanılabilir. Zira gerçekleştirilen gözlemlerde, anılan sosyal medya platformundaki bazı kullanıcıların; insan ilişkileri, spor, dizi/film vb. belirli konu başlıkları özelinde ileti oluşturmaya eğilimli oldukları tespit edilmiştir.

Yansıma ses (sözcük) ve kısaltma kullanımı ile sayıların yazımı ise MacLeod ve Grant (2012, s. 218) tarafından önerilmiş olup Türkçe kısa metinler üzerinde de kullanışlı olacağı öngörülmektedir.

Diğer Ölçütler

“Diğer ölçütler” başlığı altında ise daha önce ifade edildiği gibi konu etiketleri ile duygu belirten yüz ifadelerinin ele alınması öngörülmektedir (Tanrıvere, 2018, s. 184).

Konu etiketleri bakımından bunların konumu, bir ya da daha fazla sözcükten ibaret veya cümle oluşları, birden fazla sözcük olması hâlinde her sözcüğün ilk harfinde büyük harf kullanma eğilimleri değerlendirilmektedir. Duygu belirtmeye yarayan yüz ifadelerinin ise öntanımlı grafikler veya noktalama işaretleriyle oluşturulması, bir veya çok sayıda aynı ya da farklı grafik veya noktalama işareti kullanılması gibi hususlar incelemeye konu edilmektedir. Bu noktada belirtmek gerekir ki, her ne kadar “yüz ifadeleri” olarak anılsa da öntanımlı grafikler çok çeşitli olabilmekte ve hayvanlar, bitkiler, taşıtlar, bayraklar gibi farklı görseller içerebilmektedir. Bu nedenle ilgili ölçüt, “şekil ve işaretler” olarak adlandırılmıştır.

Tüm bu bilgileri özetlemek gerekirse, uygulanacak yönteme dair aşağıdaki şekilde bir çizelge ortaya çıkmaktadır. Yazar tanıma çalışması, detayları bir sonraki başlık altında sunulmuş olan veri kullanılarak, bu çizelgede gösterilmiş olan ölçütler ile gerçekleştirilmektedir.

Tablo 1. Uygulamada kullanılacak ölçütler

		Konum	Cümle içi	Cümle sonu	
Yazım ve Noktalama	Noktalama		Tırnak içindeki cümle sonu	İleti sonu	
				Tekli	
		Sayı	Çoklu	Aynı	Farklı
		Parantezler		Normal parantez	Köşeli parantez
				Küme parantezi	Büyüktür/küçüktür simgeleri
		Tırnak işaretleri		Tek	Çift
		Yazım yanlışları		Konum	Sözcük
				Eğilim sonucu ölçünlü dilden uzaklaşan sözcük ve ifadeler ile bunların oluşturulma biçimi	
		Vurgu		Büyük harf	Tırnak işareti
			Ünlem		Parantez içinde
Sözcük ve Bağlam	Yabancı sözcük		Kaynak dil		
	Argo ve jargon		Terminoloji ise bilim/meslek alanı		
	Bağlaç		Anlam değeri		
			Konu, anlaşılabilirlik ve tutarlılık		
			Yansıma sözcükler		
	Kısaltmalar	Niteliği		Kalıplaşmış	Oluşturulma biçimi
				Kişisel	
				Büyük harf kullanımı	Nokta kullanımı
	Sayılar	Niteliği		Yazıyla	Rakamla
					Nokta kullanımı
Diğer Ölçütler	Konu etiketleri		Konum		
			Sözcük / sözcük öbeği / cümle		
		Büyük harf kullanımı	Etiket başında	Sözcük başında	
	Şekil ve işaretler	Biçim		Grafik	
				Noktalama işareti	
				Tekli	
	Sayı	Çoklu	Aynı	Farklı	

Bulgular

İki Yazar ve Birer İleti

Yazarı bilinmeyen birer ileti (M0) ile ikişer yazarın birer iletisini (rastgele numaralandırılmış M1 ve M2) içeren toplam üç farklı deney (rastgele numaralandırılmış D1, D2 ve D3) gerçekleştirilmiştir. M0 yazarını tespit edebilmek için bu ileti, M1 ve M2 iletileriyle ayrı ayrı karşılaştırılmış, sonucun güvenilirliğini arttırmak amacıyla M1 ve M2 mesajları da kendi aralarında mukayese edilmiştir. İki yazarın birer iletisiyle yapılan deneylerin ilkine detaylarıyla yer verilmiş, diğer iki deneyin sonuçları ise aşağıdaki tabloda özet olarak sunulmuştur.

Buna göre ilk deneydeki M0 ve M1 iletileri arasında iki ölçütte ayrılık tespit edilmiş, benzerlik gösteren ölçüte ise rastlanmamıştır. Noktalama yönünden ayrılık teşkil eden ölçüt, karşılaştırmada kendini üç kez tekrar etmiştir: M0 iletisinde bir cümle mevcutken cümle sonunda noktalama işareti kullanılmamış, M1 iletisinde yer alan üç cümlenin sonunda da uygun noktalama işaretlerine (sırasıyla nokta, nokta ve soru işareti) yer verilmiştir. M0 iletisinde Instagram adlı ve görsel veri paylaşımı üzerine kurulu sosyal medya platformundaki bir işlevden bahsedilirken “story’e” denilmiş, M1 iletisinde ise Twitter’deki ana sayfa (akış ya da “*timeline*”, kısaca “*time*”) kastedilerek “*tayma*” yazımı tercih edilmiştir. Bu iki ifade, gerek yabancı sözcüğün Türkçe metinde kullanımı gerekse sosyal medya platformlarının birer alanı olarak eşdeğerdir ve yazımlarında farklılık görülmesi, yazar tanıma çalışması bakımından kayda değer bulunmuştur.

M0 ve M2 iletileri arasında ise bir benzerlik bulunmuştur: Her iki iletide de birer cümle oluşturulmuş ancak sonlarında noktalama işaretine yer verilmemiştir. Bunun dışındaki ölçütler ise ya metnin biçim ve içeriği itibarıyla uygulanabilir değildir ya da her iki iletinin de büyük harfle başlaması gibi ölçünlü dil kuralları kapsamında benzer sonuç verdiği için “etkisiz” olarak değerlendirilmiştir.

Sonuç olarak iki yazarla yapılan ilk deneyde, M0 ile M2 iletilerinin aynı yazar tarafından yayımlandığı değerlendirilmesi yapılmış ve bu karşılaştırma doğrulanmıştır.

Tablo 2. İki yazar ve birer iletiyle gerçekleştirilen ikinci ve üçüncü deneylerin özeti

Deney	Karşılaştırma	Ayırt edici ölçüt	Benzerlik sayısı	Ayrılık sayısı	Sonuç
D2	M0-M1	5	0	5	M0=M2
	M0-M2	2	2	0	
D3	M0-M1	3	3	0	M0=M1
	M0-M2	3	0	3	

Yukarıdaki tabloda ikinci deneyle ilgili sunulan M0-M2 benzerliğiyle ilgili olarak, her iki iletinin de sonunda nokta kullanılmasıyla birlikte her iki örnekte de noktanın ikişer kez ardışık olarak kullanıldığı ve bunun, benzerliği pekiştiren bir unsur olarak değil, iki ayrı benzerlik olarak değerlendirildiğini belirtmek gerekir. Zira yazarın her iki iletide noktalama işareti kullanımı ile birlikte bunların sayısına dair eğilimi de tutarlıdır. M0 ve M2 iletilerinin aynı yazara ait olduğu değerlendirilmesi yapılarak gerçekleştirilen kontrolde bu tespit doğrulanmıştır.

Üçüncü deneyde ise M0 ve M2 iletileri arasında 3 ayırt edici ölçüt kullanılmış ve bunların tamamı farklı sonuç vermişken, M0 ve M1 iletileri arasında yapılan karşılaştırmada kullanılan 3 ölçütün tamamından benzeşen sonuç alınmıştır. Karşılaştırmanın ardından M0 ve M1 iletilerinin aynı hesaptan yayımlandığı görülmüştür.

İki Yazar ve Beşer İleti

İkinci aşamada yazarı bilinmeyen bir M0 iletisi karşısında, iki yazarın beşer farklı iletisi karşılaştırmaya konu edilmektedir. Bu noktada önceki deneylerin aksine M0 iletisi ile diğer iletiler değil, yazarlar (rastgele numaralandırılmış Y1 ve Y2) karşılaştırılacaktır. Diğer bir deyişle beş ayrı iletisi toplanan her iki yazarın da, bu beş paylaşımındaki genel dil kullanımları tespit edilerek M0 iletisi karşısında ele alınmaktadır.

Bir kontrol iletisi karşısında iki yazarın beşer farklı iletisinin mukayese edildiği ilk deneyde, öncelikle M0 iletisinin özellikleri tespit edilmiştir. Buna göre ileti, bir sıralı cümleden ibaret olup tümceler noktalama işareti (virgül) ve bağlaç ile bir araya getirilmiştir. Bu noktalama işareti dışında cümle içinde veya sonunda noktalama işareti kullanılmamıştır. Cümleye küçük harfle başlanan iletinin konusu romantik ilişkilerdir. İletide konu etiketi veya şekil ve işaret bulunmamaktadır.

Y1 yazarının üslubu incelendiğinde, bazı sıralı cümleler noktalama işaretiyle bir araya getirilmişse de bu eğilimin tutarlı (sürekli) olmadığı görülmüştür. Cümle ve ileti sonlarında noktalama işareti kullanmayan yazar, tırnak içindeki cümlelerin içinde veya sonunda da noktalama işareti kullanmamıştır. Tırnak işaretlerini çift (standart: “ ”) kullanan yazar, iletilerine küçük harfle başlamaktadır. Hezeyan içeren ifadelerde sözcüklerin tamamı büyük harfle yazılmış, bunun dışında dikkat çekici bir yabancı sözcük ya da argo ve jargon kullanımına rastlanmamıştır. Sıralı cümle olarak kurgulanan iletilerin konusu romantik ilişkilerdir. Sayılar, yalnızca zaman ölçüsü belirtilirken (“5 dakika”) rakamla ifade edilmiştir. Konu etiketi veya şekil ve işaret kullanılmamıştır.

Y2 yazarı, hiçbir iletisinde noktalama işareti kullanmamıştır. Yazarın “giriolar”, “sövcen”, “davranıcan”, “benimde”, “getirde”, “yazmıcaksan” (sırasıyla “giriyorlar”, “söveceksin”, “davranacaksın”, “benim de”, “getir de”, “yazmayacaksın”) ifadeleri, ölçünlü dilden çeşitli bakımlardan uzaktır. Yazar, oluşturduğu iletilerin sonunda “aq” ifadesine yer vermektedir. Konuları çeşitlilik gösteren iletelerde bağlaç kullanılmamış olup bir iletide “2 şeker” denilerek sayı, rakamla yazılmıştır. Konu etiketi veya şekil ve işaret görülmemiştir.

M0 iletisi ve Y1 yazarı arasında dört benzerlik bulunmuştur: Cümle ve ileti sonundaki noktalama işaretleri, iletilere küçük harfle başlanması, iletelerde argo kullanılmayışı ve konu içeriklerinin romantik ilişki çağrışımlı oluşu benzerdir.

M0 iletisi ile Y2 yazarı karşılaştırıldığında ise dört ayırt edici ölçütten ikisi benzer, ikisi ayrı sonuç vermiştir. Buna göre noktalama işaretleri ve iletilere küçük harfle başlanması yönünden benzerlik görülmüş ancak ölçünlü dilden uzaklaşan ifadeler ile argo kullanımı bakımından ise ayrılık tespit edilmiştir.

Sonuç olarak iki yazarla yapılan bu deneyde, M0 iletisinin Y1 yazarı tarafından yayımlandığı değerlendirilmesi yapılmış ve bu karşılaştırma doğrulanmıştır.

Tablo 3. İki yazar ve beşer iletiyle gerçekleştirilen ikinci ve üçüncü deneylerin özeti

Deney	Karşılaştırma	Ayırt edici ölçüt	Benzerlik sayısı	Ayrılık sayısı	Sonuç
D2	M0-Y1	3	3	0	M0=Y1
	M0-Y2	3	0	3	
D3	M0-Y1	5	1	4	M0=Y2
	M0-Y2	5	5	0	

Tabloda sunulmuş olan ikinci deneyin karşılaştırma iletisinde (M0) noktalama işareti kullanılmamıştır. Y1 kodlu yazar ise yalnızca bir iletisinde birer kez cümle içinde ve sonunda noktalama işareti kullanmış olup diğer dört iletisinde hiçbir noktalama işaretine yer vermemiştir. Y2 kodlu yazarın, tüm iletisinde noktalama işareti kullanması dolayısıyla Y1 yazarının münferit olarak yalnızca bir iletide noktalama işareti kullanmış oluşu göz ardı edilmiştir. Y2 yazarı ise bir iletide cümleye küçük harfle başlamış ve bir iletiyi ardışık iki nokta ile sonlandırmıştır. Ayrı ayrı birden çok iletide tekrar etmediğinden, bu özellikler de yazara ait ileti ve cümlelerin büyük çoğunluğuna göre belirlenmiştir. M0 iletisinin, Y1 kodlu yazar tarafından yayımlandığı tespiti doğrulanmıştır.

İki yazarın beşer iletisiyle gerçekleştirilen üçüncü deneyde, tanımlı ölçütlerle birlikte Türkçe karakter (ç, ı, ğ, ö, ş, ü) kullanımı ve “bir” veya “bi” yazımında görülen farklılıklar da öne çıkmıştır. Kontrol iletisi olan M0 ile Y1 yazarı arasında görülen tek benzerlik iletilerin küçük harfle başlamasıyken, dört farklı ölçütte ayrılık görülmüştür. Y2 yazarının iletisinde ise ayırt edici olarak kullanılan beş ölçütte de benzer sonuç alınmıştır. Karşılaştırmanın ardından M0 iletisinin, Y2 kodlu yazar tarafından yayımlandığı görülmüştür.

Beş Yazar ve Birer İleti

İki yazarın birer ve beşer iletiyle yapılan ilk deneylerde incelemenin detaylarına yer verilmiştir. Bu nedenle beş yazarın birer iletiyle yapılan karşılaştırmaların sonuçları, aşağıdaki tabloda genel olarak sunulmaktadır. Yazar tanıma uygulaması bağlamında önem arz eden konular ile üçüncü deneyin detaylı açıklamasına ise tablonun altında ayrıca değinilmektedir.

Tablo 4. Beş yazar ve birer iletiyle gerçekleştirilen deneylerin özeti

Deney	Karşılaştırma	Ayırt edici ölçüt	Benzerlik sayısı	Ayrılık sayısı	Sonuç
D1	M0-M1	5	3	2	M0=M4
	M0-M2	4	1	3	
	M0-M3	4	0	4	
	M0-M4	5	4	1	
	M0-M5	3	2	1	
D2	M0-M1	2	0	2	M0=M3
	M0-M2	4	0	4	
	M0-M3	3	3	0	
	M0-M4	2	0	2	
	M0-M5	2	0	2	
D3	M0-M1	3	2	1	M0=M2
	M0-M2	2	2	0	
	M0-M3	2	1	1	
	M0-M4	2	1	1	
	M0-M5	2	1	1	

Gerçekleştirilen ilk deneydeki iletiler, kontrol ileti karşısında değişen ayırt edici ölçütlerle yer yer benzerlik ve farklılıklar göstermiştir. Bu ölçütlerin niteliği ile birlikte karşılaşılan benzerliklerin fazlalığı ve ayrılıkların azlığı, bunların iletilerde kaç kez ve ne sıklıkla tekrar ettiği göz önünde bulundurulmuştur. Bu noktada M0 ve M4 iletileri, yalnızca en fazla benzerlik ve en az ayrılık gösteren ölçüt sayısı ile değil, bu ölçütlerin iletilerdeki tekrarı bakımından da öne çıkmıştır ve bu iki iletinin aynı hesaptan yayımlandığı tespit edilmiştir.

İkinci deney grubundaki beş iletiden yalnızca bir tanesinde benzerlik taşıyan ve hiçbir farklılık göstermeyen özellikler görülmüştür. Daha önce ancak pekiştirici unsur olarak değerlendirilebileceği, ancak karşılaştırmaya esas teşkil etmemesi gerektiği belirtilen (Tanrıvere, 2018, ss. 184, 190) ölçünlü dil benzerlikleri de dikkate alındığı takdirde M0-M3 iletilerindeki benzerliğin arttığı tespit edilmiştir. Yapılan kontrolde de bu iki iletinin aynı kişi tarafından yazıldığı anlaşılmıştır.

Üçüncü deney grubunda her ileti, kontrol ileti karşısında en az bir ölçütte benzer ve en az bir ölçütte farklı sonuç vermiştir. İletilerin çok kısa olması nedeniyle sonuçların birbirine oldukça yakın olduğu görülmektedir: Kontrol ileti M0 toplam 6 sözcük içermekte, deneydeki en uzun ileti 17 ve en kısa ileti 4 sözcükten oluşmaktadır. Kontrol ileti bir cümleden ibarettir ve cümle içinde gerekmesine rağmen bir virgül kullanılmamıştır. Ne var ki sosyal medyada yapılan gözlemlerde, “... değil, ...” biçiminde kurgulanan ifadelerin çoğunda virgül kullanılmadığı tespit edilmiştir. Bu nedenle kontrol ileti içindeki “cümle içi noktalama eksikliği”, ayırt edici bir ölçüt olarak karşılaştırmada kullanılmamıştır. Bu deneyde kullanılan her bir iletinin karşılaştırma sonuçlarına ayrıca yer verilmesi gerekli görülmektedir.

M5 ileti cümle sonunda noktalama işareti kullanılması yönünden ayrılık ve ileti sonunda noktalama işareti kullanılmaması yönünden de benzerlik taşımaktadır. Bu iletide ikişer noktalama işareti kullanılması, noktalama işaretinden önce bir boşluk bırakılmasının ve diğer işaretler bakımından öntanımlı yüz ifadelerine yer verilmesinin karşılaştırmaya bir etkisi yoktur. Zira M0 ileti içinde noktalama işareti kullanılmadığından, bu işaretten önce boşluk bulunması veya bulunmaması ayırt edici değildir. Şekil ve diğer işaretlerin kullanımı ise aynı yazarın iletileri arasında da değişkenlik göstermekle birlikte, ancak kullanılması hâlinde biçim ve sayı yönünden karşılaştırma konusu edilmektedir.

M4 iletisinde ise cümle (ve ileti) sonunda noktalama işaretine yer verilmemiş ancak cümleye büyük harfle başlanmış, bu nedenle bir ölçütte benzerlik ve bir ölçütte farklılık tespit edilmiştir.

M3 iletisinde cümle (ve ileti) sonunda noktalama işareti bulunması bir farklılık, cümleye küçük harfle başlanması ise bir benzerlik unsurudur.

M1 ve M2 iletisinde ise cümle (ve ileti) sonunda noktalama işareti kullanılmaması ile cümleye küçük harfle başlanması, bu iletiler için tespit edilen iki benzerlik ölçütüne karşılık gelmektedir. Ne var ki, kontrol iletisi olan M0’da “güzellik” konusundan bahsedilerek “ruh güzelliği”ne değinilirken, M1 iletisinde bir sporcudan bahsedilerek “*at ağzını s*kerim*”², “*i*ne*”, “*yallah arabistana*” ifadelerine yer verilmesi; sözcük ve bağlam özellikleri bakımından bu iki iletinin birbirinden kuvvetle uzaklaşmasına neden olmuştur. Böylelikle en fazla benzerlik ve en az farklılık ölçütü gösteren ileti, M2 olarak tespit edilmiştir.

Beş Yazar ve Beşer İleti

Beş farklı yazarın beşer iletisinden elde edilen üslup özellikleriyle kontrol iletisi M0’ın karşılaştırıldığı üç ayrı deneyin özetlenmiş sonuçları, aşağıdaki tabloda sunulmuştur. Bu deneylerle ilgili vurgulanması gerekli görülen konulara tablonun altında yer verilmektedir.

Tablo 5. Beş yazar ve beşer iletiyle gerçekleştirilen deneylerin özeti

Deney	Karşılaştırma	Ayırt edici ölçüt	Benzerlik sayısı	Ayrılık sayısı	Sonuç
D1	M0-Y1	4	0	4	M0=Y4
	M0-Y2	3	3	0	
	M0-Y3	4	2	2	
	M0-Y4	6	6	0	
	M0-Y5	6	1	5	
D2	M0-Y1	1	0	1	M0=Y3
	M0-Y2	1	0	1	
	M0-Y3	1	1	0	
	M0-Y4	1	0	1	
	M0-Y5	3	0	3	
D3	M0-Y1	1	0	1	M0=Y4
	M0-Y2	2	1	1	
	M0-Y3	2	0	2	
	M0-Y4	2	2	0	
	M0-Y5	2	0	2	

İlk deneyde sadece benzer sonuç verip farklılık göstermeyen yazarlar, 3 ölçütü Y2 ve 6 ölçütü Y4 olarak tespit edilmiştir. Y2 yazarı; yalnızca soru cümlelerinin sonunda ikişer kez soru işareti kullanıp diğer cümlelerin sonunda noktalama işaretine yer vermemesi, cümleye küçük harfle başlaması ve argo kullanımı³ yönünden benzerlik göstermiştir. Y4 yazarı ise tıpkı M0 iletisinde olduğu gibi bir defa kesme işareti (‘) kullanıp cümle içinde başkaca noktalama işaretine yer vermemiş, cümle ve ileti sonunda noktalama işareti kullanmamıştır. Yazar aynı zamanda kontrol iletisinde olduğu gibi cümlelere küçük harfle başlamış, özel isimlerin ilk harflerini küçük yazmış ve Türkçe karakter (ç, ı, ğ, ö, ş, ü) kullanmamıştır. Bununla beraber M0 iletisinde yabancı bir dizi karakterinden bahsedilirken, Y4 yazarının bir iletisinde bu karakterin yer aldığı diziden bahsedilmiş, bir başka iletide ise “dizi hayranlarına yönelik internet sitesi” (“*fansite*”) konu edilmiştir. Dolayısıyla konu, özel isimlerin küçük harfle yazılması ve Türkçe karakter kullanılmayışı; Y4 yazarını M0 iletisi ile benzerliği bağlamında öne çıkarmıştır.

Yapılan ikinci deneyde toplam sekiz sözcük içeren kontrol iletisi, yalnızca bir ölçütte ayırt edici sonuç vermiştir ve bu, yazarın cümle (ve ileti) sonunda noktalama işareti kullanmayışıdır. Y1, Y2, Y3 ve Y4

² Bu ileti, yazarın kendisi tarafından sansürlenmiştir.

³ Argo olarak M0 iletisinde “gerizekalı”, Y2 yazarının bir iletisinde ise “mal” ifadelerine yer verilmiştir. Bunlar, daha “ağır” kabul edilebilecek sövme ve küfürlere nispeten, yazar profili bakımından benzeşen ifadelerdir.

yazarlarında ise muhtelif ayırt edici ölçütlerin varlığı tespit edilse de M0 iletiyle karşılaştırılabilecek yalnız bir ölçüt bulunduğu görülmüştür: Cümle ve ileti sonu noktalama işaretlerinin varlığı. Diğer ölçütlerin hiçbiri, M0 iletinin biçim ve içeriği itibariyle karşılaştırılabilir durumda değildir. Y5 yazarının ise noktalama işareti kullanmakla birlikte her iletişinde cümle içindeki bazı sözcüklerin büyük harfle başladığı ve tüm iletilerin, tıpkı şiirlerde olduğu gibi mısra düzeniyle ve her satır arasında birer boş satır bırakarak oluşturulduğu görülmüştür. Cins isimlerin başında büyük harf kullanılması ve iletinin biçimsel düzeni, ayrı ayrı ayırt edici birer ölçüt olarak not edilmiştir. Y3 yazarının cümle sonlarında noktalama işareti kullanmaması ve diğer tüm yazarların tesadüfi biçimde cümle ve ileti sonunda ilgili noktalama işaretine sürekli olarak yer vermesi, bu incelemede belirleyici olmuştur. Ne var ki bu tespit in hem istatistiksel hem de bilimsel olarak güvenilirliği tartışmaya son derece açık görünmektedir. Dolayısıyla yazar tanıma çalışmaları için mümkün olduğunca çok veri içeren, uzun metinlerin elde edilmesinin önemi bir kez daha ortaya çıkmaktadır.

Bu çalışma kapsamında gerçekleştirilen son deneyde kontrol iletiş, oldukça az miktarda dilsel veri içermekte olup toplam 5 sözcükten ibarettir. İletinin (ve cümle nin) sonunda noktalama işaretine yer verilmeyişi, yöntem bağlamında yer verilen ölçütler içerisindeki tek ayırt edici özellik olarak görülmektedir. Ayrıca içerik ele alındığında, yazarın erkek olduğu anlaşılmaktadır. Buna göre Y1 yazarı noktalama yönünden, Y3 ve Y5 yazarları ise hem noktalama hem de cinsiyet yönünden ayrılık göstermektedir. Zira Y3 ve Y5 yazarlarının oluşturduğu iletilere bakıldığında, yazarların kadın olduğu açıkça tespit edilebilmektedir. Y1, Y2 ve Y4 yazarlarının karşılaştırma iletileri ise bu konuda net bir tespit yapılmasına elverişli değildir. Netice olarak, Y2 ve Y4 yazarları, noktalama işareti kullanmamaları bakımından M0 iletiş ile benzerlik taşımaktadır. Bu esnada, her ne kadar yöntem bağlamında bir ölçüt olarak sunulmamışsa da iletilerin uzunluğunda açık bir farklılık olduğu görülmüştür. Kontrol iletiş 5 sözcükten ibarettir; Y2 yazarının iletileri en az 8, en fazla 27 ve ortalama 15,8 sözcükten oluşmaktadır. Y4 yazarının iletileri ise en az 5, en fazla 12 ve ortalama 6,8 sözcük içermektedir. Y2 yazarı, iletilerinde genellikle sebep-sonuç ilişkileri kurmuş, sıralı ve birleşik cümleler tercih etmiştir. Y4 yazarı ise daha yalın ve basit biçimde birtakım tespitlere yer verirken bunları basit cümlelerle aktarmıştır. Bu yönüyle, sözcük sayısı ve cümle kurgularının ayrı bir ölçüt olarak sayılmasıyla Y4 yazarının iletileri ile M0'ın benzerliği artmış ve yapılan kontrolde, M0 iletişinin Y4 tarafından yazıldığı doğrulanmıştır.

Özgül olarak bu deney, yöntem dahilindeki ölçütlerin ortalama sözcük sayısı ve ileti kurgusunu (cümle yapıları ile bir konunun ele alınış biçimlerini) da içerecek şekilde geliştirilmesinin, özellikle kısa metinlerde yazar tespiti açısından önemini ortaya koymaktadır. Bunlarla birlikte ileti içeriklerinden anlaşılabilmesi hâlinde, yazar cinsiyetinin veya yazara ilişkin başka birtakım bilgilerin belirlenmesi de uygulamaya ve dolayısıyla yazarın tespit edilmesine katkı sağlayacaktır.

Sonuç ve Öneriler

Çalışma kapsamında iki yazarın birer ve beşer iletiş ile beş yazarın birer ve beşer iletiş üzerinde, üçer farklı uygulama olmak üzere toplam 12 karşılaştırma gerçekleştirilmiştir. Yapılan bu incelemelerde, yöntem başlığı altında sunulan ölçütlerin bazıları hiç kullanılmamış, bazıları ise sıkça belirleyici bir özellik olarak öne çıkmıştır. Ölçütler arasındaki bu farklılığın hem örneklemin rastgele oluşu nedeniyle bazı ölçütlerin kullanılabilmesi ya da ayırt edici olabileceği iletilerle karşılaşılmamasından hem de bazı ölçütlerin ayırt edicilik gücünün daha yüksek oluşundan kaynaklandığı değerlendirilmektedir.

Özellikle incelemeye konu olan ileti sayısının azaldığı ve yazar sayısının arttığı deneylerde, önceden tahmin edilebileceği gibi, yazar tespitinin zorlaştığı ve karşılaştırmanın, daha çok sayıda iletiş ile yapılan deneylere göre, daha az güvenilir sonuç verdiği ortaya çıkmıştır. Diğer yandan bu çalışmanın örneklemini için veri toplayan, tasnif eden ve incelemeyi gerçekleştiren bir kişi olması dolayısıyla; karşılaştırmayı yapan, her deneyde kontrol iletişinin yazarı ile bir karşılaştırma iletişinin yazarının mutlaka eşleşeceğinin bilincindedir. Diğer bir deyişle iki yazarlı deneylerde %50, beş yazarlı deneylerde ise %20 ihtimalle yazar, hiçbir ölçüt olmaksızın doğru tespit edilecektir. Ne var ki 6'sı iki, diğer 6'sı beş yazarlı deneylerin hiçbirinde hatalı bir tespit yapılmamış ve hiçbir deneyde "sonuç" kısmı boş bırakılmamış, karşılaştırma sonucunda kararsız kalınmamıştır. Kontrol iletiş yazarının, karşılaştırma iletilerini oluşturan yazarlar arasında bulunmayabileceği durumlarda ise kimi deneylerde sonuç verilmeyebileceği, kararsız kalınabileceği değerlendirilmektedir. Bu durumda dahi, hatalı bir tespit yapılmamış olacaktır.

Yazar sayısının fazla olması ya da karşılaştırma iletilerinin az olmasının yanı sıra, kontrol iletilerinin çok kısa olup az miktarda dilsel veri içerdiği durumlarda da mukayesenin zorlaştığı görülmektedir. Karşılaştırma iletileri kısa olsa dahi, sayıca fazla olduklarında ölçütler yönünden belirgin birtakım farklılıklar ortaya koymaktadır. Ne var ki yazarı tespit edilmeye çalışılan kontrol iletileri, varoluşları itibariyle, ikame edilemez ve genişletilemez durumdadır. Kontrol iletileri için ayırt edici olmayan bir ölçüt ise ancak örneklemin elverişli olduğu, karşılaştırmada yer alan diğer başka yazarların birbirlerine olan benzerliklerini tespit etmede kullanılmakta ve böylelikle ikincil bir işleve sahip olabilmektedir.

Yapılan deneyler ve detaylarına yer verilen bulgular doğrultusunda, aşağıda maddeler hâlinde sıralanan konuların da birer ölçüt olarak yönteme eklenmesi gerektiği değerlendirilmektedir:

Büyük harf kullanımı: Yazarların iletiye, cümleye, tırnak içindeki cümleye başlarken veya özel adları (cümlelerin başında veya içinde) yazarken büyük harf kullanma eğilimleri, kaynak gösterilen yöntemde ayrı bir başlık olarak ele alınmışsa da (Tanrıvere, 2018, s. 185) bu çalışma çerçevesinde konum yönünden bir yazım yanlışı olarak değerlendirilmiştir. Ne var ki bu ölçütün sıkça kullanılması, ayırt ediciliğinin yüksek olması ve bir “yazım yanlışı” olmanın ötesinde yazarlar tarafından tercih edilen üslubun bir parçası olması dolayısıyla, kaynak yöntemdeki gibi ayrı bir başlık olarak kullanılması gerektiği değerlendirilmektedir. Bu ölçüt, yazım ve noktalama başlığı altına eklenmiştir.

Sözcük sayısı: Özellikle birden fazla karşılaştırma iletilerinin bulunduğu deneylerde, iletilerin uzunluğuna dair yazar eğilimlerinin ayırt edici olabildiği görülmüştür. Bu noktada, yazara ait iletilerdeki en az, en fazla ve ortalama sözcük sayıları, kontrol iletileri karşısında anlam kazanabilecek bir özelliktir. Bu ölçüt, sözcük ve bağlam başlığının konusu olarak eklenmiştir.

Kurgu: İletilerde sıklıkla karşılaşılan cümle türleri (basit, sıralı, birleşik) ile bir konunun nasıl ele alındığı, sözcük sayısı ile de yakından ilişkili olmakla birlikte, yazarların ayırt edilmesinde kullanılabilecek niteliktedir. Zira deneyler kapsamında iletileri incelenen yazarlardan bazılarının, bu yönde belirgin eğilimlerinin bulunduğu gözlemlenmiştir. Bu konu, içeriksel özelliği gözetilerek sözcük ve bağlam başlığında değerlendirilmektedir.

Vurgu > Diğer: Yöntemde vurgularla ilgili bir alt başlık bulunmakta ve burada büyük harf, tırnak işareti ve ünlem işaretiyle oluşturulan vurgular öngörülmektedir. Yapılan çalışmada, bazı harflerin tekrar edilmesi (“accayıp” gibi) ya da harfler arasında boşluk bırakılması (“n e d e n ?” gibi) gibi yollarla da vurgu oluşturulduğu görülmüştür. Bu nedenle, ilgili başlığın altına, öngörülmemiş olan ifadeler için “diğer” seçeneğinin eklenmesi kullanışlı olacaktır.

Profil bilgileri: Gerçekleştirilen son deneyde, ileti içeriklerinden hareketle yazarların cinsiyetlerine dair bir çıkarımda bulunulmuştur. Yazar tanıma uygulamasının temel amacı, metni oluşturan kişilerin ayırt edici dil kullanımlarını tespit etmek olsa da yazarların kimliklerine yönelik ayırım yapılabilesine olanak sunan bilgiler de göz ardı edilmemelidir. Bu bağlamda yazarın cinsiyeti, yaşı, yaşadığı yer veya memleketi, mesleği gibi birtakım bilgileri açığa çıkaran ileti içerikleri de yazar tanıma çalışmaları yönünden ayrıca kayda değerdir. Sözcük ve bağlam düzeyinde, kullanılması hâlinde argo ile birlikte jargon ve terminoloji örneklerinin hangi konuya, bilim ya da meslek alanına dahil olduğu sorusu da yazara ilişkin profil bilgilerini elde etmeye yönelik olduğundan; “profil bilgileri” başlığının “argo ve jargon” kullanımını da kapsayacak biçimde, sözcük ve bağlam başlığı altında genişletilmesinin yerinde olacağı değerlendirilmektedir.

Şekil ve işaretler > Konum: Deneylerin gerçekleştirilmesinden önce yer verilen yöntemde şekil ve işaretlerle ilgili bir başlık oluşturulmuşsa da bu işaretlerin kullanıldığı konumlar itibariyle de yazarların ayırt edilebileceği görülmüştür. Yazarların iletinin başında, ortasında, sonunda ve hatta iletideki bir sözcük yerine çeşitli şekil ve işaretler kullanması dolayısıyla bunların konumları, ayrı bir alt başlık altında ele alınmalıdır.

İleti düzeni: Yapılan çalışmada, iletilerin biçimsel düzeninde yer yer boş satırların bulunduğu veya yazılanlara bir şiir gibi mısra düzeninde yer verildiği görülmüştür. Boş satırlar kimi zaman birbirinden bağımsız cümleleri, kimi zaman ise ileti içeriği ile konu etiketlerini ayırmak için ya da son sözcüğün veya sözcüklerin vurgusunu arttırmak için kullanılabilir. Bu bağlamda iletinin şekil yönünden düzenine dair görülen eğilim, yazarları ayırt etmede kullanılabilir. Bu son ölçüt ise ne doğrudan yazım ve noktalama ne de sözcük ve bağlam özellikleriyle ilintili görüldüğünden, konu etiketleri ile şekil ve işaretler konularının da yer aldığı “diğer ölçütler” başlığına eklenmiştir.

Yukarıda önerilen ölçütlerin de yönteme dahil edilmesiyle birlikte genişletilmiş ölçüt tablosu, aşağıda sunulmuş olan Tablo 6’daki gibi geliştirilebilir.

Daha önce yöntem oluşturmaya yönelik gerçekleştirilen çalışmada 10 farklı yazardan 5’er adet olmak üzere toplam 50 ileti incelenmişken (Tanrıvere, 2018, s. 181), bu çalışmada 50 farklı yazarın 20’şer iletisi olmak üzere toplam 1000 gönderi göz önünde bulundurulmuş ve hem gerçekleştirilen deneylere bulgularıyla birlikte hem de bu deneyler sonucunda yöntem bağlamında faydalı olacağı görülen iyileştirmelere yukarıda yer verilmiştir.

Özellikle yazar sayısının fazla ya da ileti sayısının az veya kontrol iletisinin oldukça kısa olduğu durumlarda karşılaşılan güçlükler yukarıda değinilmiştir. Her bir muhtemel yazarın kontrol iletisi karşısındaki benzerlik ve farklılıklarının tespit edilmesiyle en yakın muhtemel sonucun ortaya çıkarıldığı bu çalışmalarda, her iki yöndeki (benzerlik/ayrılık) ölçütlerin tekrar etme sıklıklarının da son derece önemli olduğu görülmüştür. Tıpkı dilbilgisel hataların münferit oluşu ya da tekrar edişi gibi, ölçütlerden birinde tespit edilen ayırt edici bir unsurun da münferit oluşu, her iletide yer alışı ya da her bir iletide birden fazla kez gözlenmesi, bu ölçütün ilgili inceleme özelindeki ispat gücünü etkilemektedir. Bu çalışmalar çerçevesinde sadece benzer veya sadece ayrı sonuç veren ölçütlere odaklanılmamalı, her iki yönden gerçekleştirilen incelemede en fazla benzerlik ve en az farklılık sonucunu veren yazarlar aranmalıdır.

Diğer yandan bu çalışma, çeşitli suç veya hukuki uyuşmazlık konuları bağlamında ihtiyaç duyulabilecek adli dilbilim uygulamalarını geliştirmek üzere gerçekleştirilmiştir. Soruşturma ve kovuşturma makamlarının da hiçbir zaman bir tek delili esas alarak hüküm vermemesi gerektiği açıktır. Dolayısıyla yukarıda örneklendirilen yazar tanıma çalışmalarından alınan sonuçlar, başka birtakım delillerle de desteklenerek bir örüntü oluşturduğunda anlam kazanacaktır. Bu nedenle, adli dilbilim çerçevesinde yazar tanıma çalışmasını yürüten kişi, yetersiz veriyle karşı karşıya kaldığında ya da yaptığı inceleme sonucunda kararsız kaldığında, bu durumu, görevlendirme makamına açıkça belirtmelidir.

Tablo 6. Uygulamalar ile birlikte genişletilmiş ölçütler

		Cümle içi		
		Cümle sonu		
		Tırnak içindeki cümle sonu		
		İleti sonu		
		Tekli		
		Çoklu	Aynı Farklı	
Yazım ve Noktalama	Noktalama	Parantezler	Normal parantez	
			Köşeli parantez	
			Küme parantezi	
			Büyüktür/küçüktür simgeleri	
	Tırnak işaretleri		Tek	
			Çift	
	Yazım yanlışları		Konum	
			Sözcük	
	Büyük harf kullanımı	Cümle başlangıcı	Tırnak içindeki cümle	
		Özel adlar	Cümle başında Cümle içinde	
Eğilim sonucu ölçünlü dilden uzaklaşan sözcük ve ifadeler ile bunların oluşturulma biçimi				
Sözcük sayısı (En az, en çok ve ortalama)				
Kurgu (Cümle türleri ve konunun ele alınışı)				
Vurgu		Büyük harf		
		Tırnak işareti		
	Ünlem	Parantez içinde		
		Diğer		
Sözcük ve Bağlam	Yabancı sözcük	Kaynak dil		
	Profil bilgileri	Argo ve jargon	Bilim/meslek/ilgi alanı	
	Bağlaç	Anlam değeri		
	Konu, anlaşılabilirlik ve tutarlılık			
	Yansıma sözcükler			
	Kısaltmalar	Niteliği	Kalıplaşmış	
Kişisel			Oluşturulma biçimi	
		Büyük harf kullanımı		
Sayılar	Niteliği	Yazıyla		
		Rakamla		
		Nokta kullanımı		
Diğer Ölçütler	Konu etiketleri	Konum		
		Sözcük / sözcük öbeği / cümle		
	Büyük harf kullanımı	Etiket başında		
		Sözcük başında		
	Biçim	Grafik		
		Noktalama işareti		
Şekil ve işaretler	Sayı	Tekli		
		Çoklu	Aynı Farklı	
		Konum		
İleti düzeni				

Türkçe alan yazınına bakıldığında hem adli dilbilim hem de yazar tanıma ve özgül olarak kısa metinlerde yazar tanıma çalışmalarının nispeten oldukça yeni olduğu görülecektir. Bu nedenle, anılan bu konular üzerinde daha fazla çalışılması gerektiği açıktır. Zira gerçekleştirilen bu araştırmalar yalnızca bilimsel bir boşluğu doldurmakla kalmayacak, aynı zamanda adaletin etkin ve doğru, diğer bir deyişle kısa sürede ve maddi gerçeğe uygun biçimde yerini bulmasını sağlamaya hizmet edecektir. Yazar tanıma uygulamaları özelinde ise her örneklem için uygun olduğu iddia edilebilecek tek ve genel-geçer bir yöntem olmadığı daha önce belirtilmiştir (Tanrıvere, 2018, s. 189). Dolayısıyla yukarıda sınanmış ve birtakım ölçütlere yönelik ekleme önerileri sunulmuş bu yöntem, her zaman için geliştirilmeye açıktır ve incelemeye konu edilecek örneklem özelinde uyarlanabilir niteliktedir. Gelecekte bu yöntemi geliştirip güçlendirmeye ve yapılan tespitlerin ispat gücünü ifade etmeye yarayan ölçekler oluşturmaya yönelik çalışmalar yapılması, hem bilimsel hem de mesleki alanda kayda değer bir bilgi birikimi oluşturmaya yarayacaktır.

Kaynakça

- Amasyalı, M. F. ve Diri, B. (2006). Automatic Turkish text categorization in terms of author, genre and gender. C. Kop, G. Fliedl, H. C. Mayr ve E. Métais (Eds.), *11th International Conference on Applications of Natural Language to Information Systems* içinde (ss. 221-226). Berlin, Heidelberg: Springer.
- Bhargava, M., Mehndiratta, P. ve Asawa, K. (2013). Stylometric analysis for authorship attribution on Twitter. *BDA 2013 Proceedings of the Second International Conference on Big Data Analytics*, 8302, 37-47.
- Coulthard, M. (1998). Identifying the author. *Cahiers de Linguistique Française*, 20, 139-161.
- Coulthard, M. (2004). Author identification, idiolect and linguistic uniqueness. *Applied Linguistics*, 25 (4), 431-447.
- 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. J. Gibbons ve M. T. Turell (Ed), *Dimensions of Forensic Linguistics* içinde (ss. 215-229). Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Grant, T. ve MacLeod, N. (2018). Resources and constraints in linguistic identity performance: A theory of authorship. *Language and Law / Linguagem e Direito*, 5 (1), 80-96.
- MacLeod, N. ve Grant, T. (2012). Whose tweet? Authorship analysis of micro-blogs and other short-form messages. In S. Tomblin, N. MacLeod, R. Sousa-Silva ve M. Coulthard (Eds.), *Proceedings of The International Association of Forensic Linguists' Tenth Biennial Conference* (ss. 210-224). Birmingham: Aston University.
- Mayda, İ. ve Amasyalı, M. F. (2016). Yazar tanımadaki köşe yazısı ve tweet'lerin çapraz kullanımı. 2. *Elektrik-Elektronik, Bilgisayar, Biyomedikal Mühendislikleri Bilimsel Toplantısı'nda* sunulan bildiri. İstanbul.
- McMenamin, G. R. (2002). Linguistic variation. In G. R. McMenamin (Ed.), *Forensic linguistics: Advances in forensic stylistics* (ss. 44-64). Boca Raton, London, New York, Washington D.C.: CRC Press.
- Olsson, J. (2008). *Forensic linguistics* (2. ed.). London, New York: Continuum.
- Tanrıvere, U. (2018). Kısa metinlerde yazar tanıma: Twitter için bir yöntem önerisi. *Humanitas*, 6 (12), 177-192. doi:10.20304/humanitas.439458
- 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.
- Twitter. (2018). Privacy Policy. <https://twitter.com/en/privacy> adresinden erişildi.
- Wardhaugh, R. (1998). *An introduction to sociolinguistics* (3. basım). Massachusetts, Oxford: Blackwell.