

## ARAŞTIRMA-DAVETLİ YAZAR / RESEARCH ARTICLE-INVITED AUTHOR

### Dilbilimsel Otopsi: Yazar Tanımlama Ölçütleri ve Yöntemleri

#### Linguistic Autopsy: Author Profiling Criteria and Methods

Burcu İlkay Karaman

#### Öz

**Amaç:** Yazar tanımadan (=author attribution/identification) farklı olarak, yazar tanımlama (=authorship profiling/author characterisation) günümüzde oldukça ayrıntılı işlemler gerektirmektedir. Yazar tanımlama öğeleri, bir şüpheli metindeki yazarın cinsiyetini, yaş grubunu, meslek grubunu, anadili, lokasyonu ve demografik özelliklerini, kişilik özelliklerini, ruh sağlığını, eğitim düzeyini, etnik kökenini/ırkını, kültür düzeyini/yapısını, sosyal statüsünü, dini inanışını, politik eğilimini ve kültürel değerlerini belirlemeden ibarettir. Bu makalede yazar tanımlama öğelerinden şüpheli bir metindeki yazarın cinsiyetine, yaş grubuna, meslek grubuna, anadili, lokasyonu ve demografik özelliklerine, kişilik özelliklerine ve ruh sağlığına yönelik yöntemlerden bahsedilecektir.

**Sonuç:** Yöntem olarak adli dilbilimi dilbiliminin kuramsal ve uygulamalı altalanlarından faydalanmaktadır. Bu altalanlar göstergebilimi, sözcükbilimi, sesbilimi/sesbilgisi, biçimbilimi, anlambilimi, edimbilimi, terimbilimi, sözdizimi, metindilbilimi, psikodilbilimi, bilişsel dilbilimi, nörodilbilimi, klinik dilbilimi, toplumdilbilimi, derlem dilbilimi, bilgisayarlı dilbilimi şeklinde sıralanabilir. Söz konusu altalanlarda kullanılan yöntemlerden hareket ederek yazar tanımlama öğeleri için yazar tanımlama ölçütleri geliştirilmiştir. Bunlar şüpheli belgenin yapısal özelliklerini, yazarın ulam önceliğini, metnin okunabilirliğini, sözlükbirimsel özelliklerini, yazarın hatalı kullanımlarını ve metnin konu odaklılığını kapsamaktadır.

**Bulgular ve Tartışma:** Bu makalede yazar tanımlama (=authorship profiling/author characterisation) öğelerinden yazarın cinsiyetini, yaş grubunu, meslek grubunu, anadili, lokasyonu ve demografisini, kişilik özelliklerini ve ruh sağlığını belirlemede uygulanan yöntemlerden bahsedilirken, her bir öğe için değişken olarak, %60-80 doğruluk oranıyla teşhis konabildiğine de dikkat çekilmiştir. Bu oranın yükselmesi için yeni yöntemlerin geliştirilmesi dilbiliminin ilgili altalanlarındaki gelişmelere bağlı olarak adli dilbilimciler tarafından sağlanabilir. Ancak, Türkiye’de olgulara dair özgün kaynaklara erişim güçlüğü nedeniyle benzetim (=simulation) amaçlı kurgusal veri ve derlemlerle araştırmalar yapma zorunluluğu kaçınılmazdır.

**Anahtar Sözcükler:** Dilbilimi; Adli Dilbilimi; Yazar Tanımlama.

#### Abstract

**Objective:** Different from authorship attribution, i.e. authorship identification, authorship profiling, also referred to as author characterisation, requires somewhat more elaborated investigations to be performed. This article deals with the authorship profiling criteria involving methods for the prediction of the suspect’s gender, age group, profession, native language, location, and demography, personality traits/type, and mental health.

**Materials and Methods:** Authorship profiling elements involve the prediction of gender, age group, profession, native language, location, demography, personality traits/type, mental health, educational level, ethnic belonging/ethnicity, cultural level, social status, religious belief, political interests, and cultural values of a suspect through a thorough investigation of the text s/he produced. For these elements, which we refer to as authorship profiling criteria, several methods have been developed involving the analysis of the author’s part of speech usage, the suspected text’s readability features, lexical choice, errors, and topic specific features. The aforementioned criteria are of universal value, that is, can be applied to any language, however, may sometimes vary from language to language and from writing system to writing system, depending on language family and/or typology.

**Conclusion:** The article specifically deals with the investigation of methods, such as the prediction of gender, age group, profession, native language, location, and demography, personality traits, and mental health, concerning authorship profiling elements. It needs mentioning that for each element the accuracy rate for prediction varies between 60-80%. In order to increase this rate, forensic linguists need to develop new methods, which is dependent on the advances encountered in each subfield of linguistics. As it is considerably difficult to acquire original case material in Turkey, it is inevitable to carry out research with fictional data and corpus.

**Keywords:** Linguistics; Forensic Linguistics; Author Profiling.

DOI: 10.17986/blm.2019356618

Burcu İlkay Karaman: Doç. Dr.,  
Dokuz Eylül Üniversitesi Edebiyat  
Fakültesi Dilbilim Bölümü Genel  
Dilbilimi Anabilim Dalı, İzmir  
Email: burcu.karaman@deu.edu.tr  
ORCID iD: <https://orcid.org/0000-0003-0506-2035>

#### Bildirimler/ Acknowledgement

Yazarlar bu makale ile ilgili herhangi bir çıkar çatışması bildirmemişlerdir. The authors declare that they have no conflict of interests regarding content of this article.

#### Finansal Destek/Support Resources

Yazarlar bu makale ile ilgili herhangi bir finansal destek bildirmemişlerdir. The Authors report no financial support regarding content of this article.

\*Sorumlu Yazar/ Corresponding Author:

Geliş: 26.09.2019

Kabul: 27.09.2019

p-ISSN: 1300-865X

e-ISSN: 2149-4533

## 1. Giriş

Adli dilbilimi, uygulamalı dilbiliminin bir altalanı olarak adli olgu ve olaylara dilbilimsel yöntemlerin uygulandığı bir bilim dalıdır (1). Betimlemeli dilbilimi (=descriptive linguistics) ile adli dilbilimi ortak çalışmalara imza atar; bu bağlamda, derlem (=corpora) olsun, gerçek veri (=real data) olsun (veri hacmi ve veriyi edinme gücüyle ilintili olarak) araştırmaların odak noktasındadır (2). Adli dilbiliminin araştırma konuları arasında aşağıdakiler vardır (2):

1. Hukuki belgelerdeki dil kullanımı;
2. Kolluk kuvvetlerindeki dil kullanımı;
3. Çocuk ve korumaya muhtaç tanıklarla yapılan görüşmeler, bunların ifade alımları;
4. Mahkemede ifade alma/verme süreçleri;
5. Dilbilimsel kanıtlar ve bilirkişilerin mahkeme önünde ifade vermesi;
6. Yazar tanıma (=authorship attribution/identification) ve intihal tespiti (=plagiarism detection);

## 2. Gereç ve Yöntem

Erken adli dilbilimsel araştırmalar dilbilimi, hukuk, psikoloji, antropoloji ve sosyoloji gibi bir dizi bilim dalının ortak çalışmasıyla, elyazısı çözümlemesi, adli sesbilgisi ve dilbilimcinin mahkemede bilirkişi olarak rol alması gibi çeşitli konuların Avustralya, Avrupa ve Kuzey Amerika'daki bazı ülkelerde ele alınmasıyla gerçekleşmiştir (2). Nitekim adli dilbilimsel araştırmalar çeşitli dilbilimsel yöntemlerin hukuki sorunlara uygulanmasıyla başlamıştır da denilebilir. Dolayısıyla, adli dilbilimci dilbiliminin kuramsal olsun, uygulamalı olsun aşağıda belirtilmiş altalanlarından faydalanmaktadır:

1. göstergebilimi (grafolojik ve grafemik incelemeler, emoji kullanımı gibi),
2. sözcükbilimi (zihinsel sözlükçe (=mental lexicon), sözcük varlığı (=lexicon), sözcük düzlemi (=lexis) gibi),
3. sesbilimi/sesbilgisi (konuşucu tanıma (=speaker identification), konuşucu doğrulama (=speaker verification), konuşucu tanımlama (=speaker profiling) vs. gibi),
4. biçimbilimi,
5. anlambilimi (anlam belirsizliği, eşdeğerlik sorunları gibi),
6. edimbilimi (sözylem kuramı (=speech act theories), Grice Kuralları (=Gricean maxims) gibi),
7. terimbilimi (meslek grubunun belirlenmesi gibi),
8. sözdizimi (imla, dizimsel özellikler, tümce uzunluğu vs. gibi),
9. metindilbilimi (bağlaşıklık ve tutarlılık örüntüleri, bağlaç kullanımı, biçem gibi),

10. psikodilbilimi (anadillilik, dilsel yeterlilik, dilsel doğruluk ve gerçeklik, yani yalan tespiti gibi),
11. bilişsel dilbilimi (deneyimsel çözümleme (=experiential analysis), belirginlik çözümlemesi (=prominential analysis), dikkat çözümlemesi (=attentional analysis) gibi),
12. nörodilbilimi (örn. psikiyatrik olgular, alzheimer, demans gibi),
13. klinik dilbilimi (ses bozuklukları, ağız yapısındaki bozukluktan ileri gelen sesletim bozuklukları gibi),
14. toplumdilbilimi (söylem çözümlemesi (söylem imgelerinin kullanımı, onaylamalar, tekrarlar ve açılmalar, dolgular, belirteçler, argo birimler ve dilbilgisi kuralları dışında kalan birimler), lehçebilimi (dialektoloji), kesit dil araştırmaları (örn. argo ve çete jargonu), sözel şiddet unsurları (örn. küfür gibi),
15. derlemdilbilimi (veri madenciliği ve çözümlemesi, veri istatistiği gibi),
16. bilgisayarlı dilbilimi (veri madenciliği ve çözümlemesi, veri istatistiği gibi), vs.

Dilbilimsel yöntemlerin kullanılarak suça dair bir **ses kaydı** ya da yazılı **belge incelemesi** yaparak, konuşucunun ya da yazarın tespitini yapmak adli dilbilimin temel görevidir. Adli soruşturma ve kovuşturma evrelerinde dilin, **sözlü** ve **yazılı** kanıt hazırlama ve sunma işlemlerini barındıran, ölçme, inceleme ve çalışma alanıdır. Kriminal ya da sivil sorunlarla ortaya çıkan ya da bir yasa ile ilişkili olan veya doğrudan yasanın kendisi olan herhangi bir **sözlü dil ögesi** veya **yazılı metin** üzerinde çalışma yapan bir alandır. Suçu çözme ve haksız yere suçlanmış bireylerin beraat ettirilmesini içeren suçla ilgili birçok alanla ilişkilendirilmiştir.

### 2.1. Adli Dilbiliminin İncelediği Metin Tipleri

Her türlü kriminal iddiaya ait olan sözlü ve/ya yazılı belgeler incelenebilir. Ancak, adli dilbilimi ile sözlü ve/ya yazılı metinlerin incelemeye alınmasıyla:

1. tehdit/gözdağı telefonları ve yazıları (tanıklara yönelik sözel/yazılı tehditler gibi)
2. nefret söylemleri ve yazıları,
3. şantaj telefonları ve yazıları,
4. sözel ve yazılı saldırı (argo, hakaret, küfür kullanımları),
5. intihar notları ve yazıları,
6. kısa mesajlar,
7. dilekçeler,
8. elektronik posta/mektup içerikleri,
9. sosyal medya gönderileri,

10. takma isimlerle açılan sosyal medya hesaplarındaki yazışmalar,
11. ankesörlü telefon ya da cep telefonu görüşmeleri (örn. bir başkasının adına yapılan sahte telefon görüşmeleri),
12. ele geçirilen telefon hatlarıyla faillerin yaptığı konuşmalar,
13. sahte vasiyetnameler,
14. fidyeye telefonları, notları, yazıları,
15. adli mercilerde alınmış ifade tutanakları (mahkemelerde/duruşmalarda ve güvenlik birimlerinde verilen ifadeler, örn. karakol tutanakları),
16. kurumsal tutanaklar (soruşturma tutanakları, toplantı tutanakları gibi),
17. cinayet notları, yazıları, itirafları,
18. hırsızlık ve soyguna dair notlar ve yazılar,
19. izinsiz konuşmalar,
20. deşifreyon (örn. terör üyeleri arasında yapılan şifreli konuşma ve yazışmalar, suç örgütü üyeleri arasında kullanılan çete jargonu gibi),
21. suikasta/suikast girişimine dair konuşma, görüşme ve yazışmalar,
22. şiddet söylemleri (örn. ekonomik şiddet, siyasi ya da kurumsal şiddet, cinsel şiddet, kültürel şiddet, dini şiddet, siber zorbalık, bilgi şiddeti, psikolojik şiddet, duygusal şiddet, zihinsel şiddet, fiziksel şiddet gibi),
23. taciz söylemleri (örn. psikolojik taciz, duygusal taciz, zihinsel taciz kapsamında reşit olmayan ile engelliye tehdit etmek, yersiz/gereksiz eleştirmek, korkutmak, aşağılamak, değersizleştirmek, suçlamak, kişiye bağırarak hakaret etmek gibi, fiziksel/bedensel taciz kapsamında cinsel tacize yönelik, özellikle reşit olmayanlarla, yani çocuklarla, yapılan cinsel taciz amaçlı ve çocuk pornosuna yönelik mesajlaşmalar, yazışmalar ve görüşmeler gibi),
24. istismar söylemleri (örn. fiziksel istismar, cinsel istismar, duygusal istismar, ihmal gibi),
25. iftira dilekçeleri,
26. itiraf dilekçeleri,
27. güvenlik birimlerine yapılan asılsız ya da şaşırtmacalı telefon ihbarları,
28. başkasının e-posta ya da sosyal medya hesabına yapılan yasadışı erişimlerle bu hesaplar üzerinden yapılan yazışmalar,
29. akademik makale, kitap, vs. (örn. intihal) gibi geleneksel yöntemler olsun elektronik ortamda ağ üzerinden olsun, iletişimin daha çağdaş değişimleri de artık çözümlenebilir duruma gelmiştir (Bazı değişikliklerle bkz: [http://www.thetext.co.uk/learn\\_about\\_forensic\\_linguistics.html](http://www.thetext.co.uk/learn_about_forensic_linguistics.html)). Bunlardan bazıları adli dilbilimin araştırmaları ve adli bilişimin ortak çalışmasıyla beraber

deşifre edilebilmektedir. Adli dilbilimi ile sözlü ve/yazılı metinlerin incelemeye alınmasıyla elektronik ortamda yapılan iletişimin daha çağdaş değişimleri de artık çözümlenebilir duruma gelmiştir.

## 2.2. Adli Dilbiliminin Araştırma ve İnceleme Konuları

Adli dilbilimi iki önemli konu etrafında şekillenmiştir (Bkz. <http://www.forensiclinguistics.eu/forensic-linguistics.html>):

1. Hukuk ve dilbiliminin kesiştiği tüm alanlarda araştırma ve uygulamanın yapılması.
2. Dilin hukuki ve adli olgularda rolü, özelliği ve kanıtsal değerlerinin araştırılması.

Bu bağlamda özellikle aşağıda belirtilen konular araştırma nesnesi olur:

1. Sözlü metin incelemeleri:
  - a. **Ses çözümlenmeleri (=voice analyses):**
    - i. sesbilimsel ve sesbilgisel çözümlenmeler (=phonetic and phonological analyses)
    - b. **Konuşmacı çözümlenmeleri (=speaker analyses)**
      - i. Konuşmacı tanıma (=speaker identification)
      - ii. Konuşmacı doğrulama (=speaker verification)
      - iii. Konuşmacı tanımlama (=speaker profiling)
    - c. **Dilsel doğruluk ve gerçeklik (=language veracity)**
  2. **Yazılı metin incelemeleri:**
    - a. **Grafolojik ve grafemik çözümlenmeler (=graphological and graphemic analyses)**
      - i. Elyazısı incelemeleri (fulaj izi incelemeleri, vs.)
      - ii. İmza incelemeleri
      - iii. Banknot ve değerli kağıtların incelenmesi (kaşe ve mühür izi incelemeleri, Mürekkep incelemeleri (yaş tayini, fiziksel incelemeler vs.), sürşarj incelemeleri, lekeleme, karalama, silme, kazma, eskitme, yakma, yırtma gibi tahriyat unsurlarının incelenmesi, vs.)
      - iv. İmla incelemeleri (büyük-küçük harf kullanımı, yazım hataları, noktalama işaretleri kullanımı, vs.)
      - v. Emoji/emotikon kullanımlarının incelenmesi
    - b. **Yazar çözümlenmeleri (=authorship analyses)**
      - i. Yazar tanıma/belirleme (=authorship attribution/identification)
      - ii. Yazar doğrulama (=author verification)
      - iii. Yazar tanımlama (=authorship profiling/author characterisation)
      - iv. Yazılararası Benzerlik Bulma/İntihal Tespiti (=similarity/plagiarism detection)

### c. Dilsel doğruluk ve gerçeklik (=language veracity)

Dilbilimsel yöntemlerle şüphelinin doğruyu ifade edip etmediğini saptamak.

### 2.3. Yazar Çözümlemeleri (=author analyses)

Adli dilbiliminin yazılı verilerin içeriğine ilişkin 5 (beş) temel soruşturması vardır (1, 3, 4, 5, 6, 7). Dilsel doğruluk ve gerçeklik incelemeleri hariç bunlardan dördü yazar sorgulamalarına ilişkindir:

1. Yazar Tanıma/Belirleme (=authorship attribution/identification):

**Yazar tanıma**, ya da **yazar belirleme**, şüpheli bir metnin yazarının kim olduğunun belli olmadığı, ancak bir grup yazar arasından birisi tarafından yazıldığı bilindiği durumlarda, doğru yazarı bulma işidir. Kısaca, bir iletiyi kimin aktardığı veya kimin aktarmadığını aydınlatmak için yazar tanıma/belirleme incelemesi yapılır. Birçok çalışmada **yazar tanıma** terimi ile **biçem ölçümü** anlamındaki **stilometri** (=stylometry) terimi aynı anlamda kullanılmıştır. Bu tip çalışmalarda genellikle yazarı biricik olarak niteleyebilecek ve yazı özelliklerinden çıkarılmaya çalışılan niceliksel veriler incelendiğinden, **niceliksel yazar tanıma** (=quantitative authorship attribution) terimi de sık sık karşımıza çıkmaktadır (8).

2. Yazar Doğrulama (=author verification):

**Yazar doğrulama**, şüpheli bir metnin, belirli bir yazar tarafından yazıldığı ya da yazılmadığının doğrulanmasıdır.

3. Yazar Tanımlama/Profilleme (=authorship profiling/author characterisation):

**Yazar tanımlama**, ya da **yazar profilleme**, bir iletiyi aktaran gerçek kaynağın belirlenemediği durumlarda mevcut kaynağın demografik özelliklerinin belirlenmesine dayanarak profil çıkarmadır. Şüpheli bir metnin yazarı hakkında yaşı, cinsiyeti, eğitim durumu, kültür düzeyi/yapısı, vs. gibi bilgilerin çıkartılmasıdır.

4. Yazılararası Benzerlik Bulma/İntihal Tespiti (=similarity/plagiarism detection):

**Yazılararası benzerlik bulma**, ya da diğer anlamıyla **intihal tespiti**, iki metin arasındaki benzerliklerin bulunması işidir. Eser sahipliğine ilişkin sorgulamalarda kullanılan yöntemdir. Genellikle intihal yapılıp yapılmadığına bakmak için uygulanır. Günümüzde bu tür sorgulara yanıt bulmak için bilgisayar yazılım programları mevcuttur (örn. *Turn it In* intihal yazılım programı).

### 2.4. Yazar Tanımlama Ölçütleri (=authorship profiling criteria)

Genel anlamda, yazar tanımlama ölçütleri,

1. şüpheli belgelerin yapısal özellikleri (=structural features),

2. ulam önceliği (=part of speech usage),

3. metnin okunabilirliği (=readability features),

4. sözlükbirimsel öncelikler (=lexical choice),

5. hatalı kullanımlar (=errors), ve

6. konu odaklılık (=topic specific features)

biçiminde sıralanabilir. Bu ölçütler hakkında anahatlarıyla aşağıdaki kısa bilgileri verebiliriz:

**1. Şüpheli Belgelerin Yapısal Özellikleri** (=structural features):

Yazar tanımlama/profilleme işlemlerinde veri madenciliği önemli rol oynamaktadır. İncelenmesi gereken şüpheli sözel/yazılı metinlerde **bildirişme** (=conversation), paragraf, tümce, tümce içerisindeki sözcük ile özel karakterlerin sayısının saptanması gibi işlemler gerektirir (9, 10). Şüpheli birden fazla bildirişme yapmış ise bildirişmenin uzunluğu, yani sayısı, hesaplanarak minimum, maksimum ve medium olarak derecelendirilir.

*Chatbot*'ların benzer biçimde bildirişme örüntülerine sahip oldukları dikkate alınır, şüpheli metnin **Jaccard Benzerlik Katsayısı** (=Jaccard similarity coefficient) da hesaplanmalıdır (9). Metinlerarasındaki ilişkinin ölçülmesi amacıyla geliştirilen ölçülerden (=metrics) birisi olan **Jaccard Endeksi** (=Jaccard index) ise iki metin üzerinden özellik çıkarımı (=feature extraction) yapıldıktan sonra ortak olan özelliklerin sayısının, iki metindeki toplam özellik sayısına bölünmesi ile elde edilir. Jaccard Endeksi'nin değerinin yüksek çıkması iki dizginin birbirine daha çok benzediği anlamına gelir. İki dizgi birbirine tamamen eşit ise Jaccard Endeksi 1 olarak bulunur. İki dizginin hiç ortak özelliği yoksa da değer 0 olarak bulunacaktır. Jaccard Endeksi 0 ile 1 arasında değişen bir değer alabilir. **Jaccard Endeksi** kullanılarak hesaplanabilen diğer bir değer de **Jaccard Mesafesi**'dir (=Jaccard distance). Literatürde **mesafe** (=distance) terimi, yunan alfabesindeki delta sembolü ile gösterildiği için indis olarak delta almış J harfi ile ifade edilir. Bu değer basitçe Jaccard Endeksi'nin 1'den çıkarılması ile elde edilir. Buradaki amaç, mesafe değeri olarak benzerlik değerinin tersi değer elde etmektir. Yani birbirine yakın olan dizgilerde, benzerliğin artması ancak mesafenin azalması gerekir. Bu anlamda mesafe, benzerliğin tam tersidir ve Jaccard Mesafesi de bu yaklaşıma göre, Jaccard Endeksi'nin tersini alır. Jaccard Endeksi'nde olduğu gibi 0 ile 1 arasında bir değer olup, Jaccard Endeksi'nin tam tersi şekilde, birbirinin aynı olan dizgiler için mesafe 0 ve birbiri ile hiç ortak özelliği bulunmayan dizgiler için mesafe değeri 1 çıkacaktır. **Jaccard Benzerlik Katsayısı** (=Jaccard similarity coefficient) **Jaccard Endeksi** ve **Jaccard Mesafesi**'nin **metin madenciliği** (=text mining) gibi ortamlarda kullanılması sırasında, çok sayıda metnin bulunacağı gerçeğinden yola çıkarak geliştirilmiştir.



Bu değerlerin hesaplanması için 4 (dört) farklı değer hesaplanması gerekir; bunlar iki farklı kümenin varlığındaki Küme 1'de bulunan M01 ile Küme 2'de bulunan M10, ve kümelerin kesişiminde bulunan M11 ile kümenin dışında kalan ve her iki kümede de bulunmayan değerlerin sayısı M00'dır (11).

## 2. Ulam Önceliği (=part of speech usage):

Cinsiyet belirleme işlemi için yapılması gereken işlemler arasında sözlükbirimsel özellikler, yani ulam (=part of speech), **önemli bir yer teşkil etmektedir. Yapılan çalışmalar göstermektedir ki, kadın ile erkek dili arasında belirgin farklılıklar vardır** (12,13). Ulam önceliği, isim, fiil, sıfat, zamir, edat, belirteç/zarf, bağlaç, niceleyici (*bazı, hepsi gibi*), ünlem, yansıma (=onomatopoeia) gibi kullanımların sıklığının belirlenmesi işlemidir (9, 14).

## 3. Metnin Okunabilirliği (=readability features):

Sözcük, tümce, hece sayısı ile zor sözcüklerin sayısının (*Dale-Chall Listesi*'nde olmayan sözcüklerin) belirlenmesidir (9, 15).

## 4. Sözlükbirimsel Öncelikler (=lexical choice):

Sözlükbirimsel önceliklerde, sözlükbirimleri ile sözlükbirimlerinin dışında kalan öğelerin kullanımına bakılır (9). Bu birimler aşağıda verilmiştir:

- kısaltmalar,
- emojiler/emotikonlar,
- kötü kullanımlar (çirkinleme, küfür gibi),
- duygulanım sözcükleri (İngilizce için bkz. *NodeBox* (Kaynak: <http://nodebox.net/code/index.php/Linguistics>),
- bağlaçlar,
- ikna ediciler (örn. *sen, para, yeni, sonuçlar, sağlık, kolay*, vs.)

## 5. Hatalı Kullanımlar (=errors):

Şüpheli metin içerisindeki hatalı kullanımları belirlemek amacıyla İngilizce için *LanguageTool* (Kaynak: <https://languagetool.org/>), *Grammarly* (Kaynak: <https://www.grammarly.com/>) gibi yazılım programları ile *ISO 27 Listesi*'ne (Kaynak: <http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#lqissue-typevalues>) bakılabilir (9). Türkçe için *Microsoft Word* programı üzerindeki *Yazım ve Dilbilgisi* ayarlarından da gidilebilir.

## 6. Konu Odaklılık (=topic specific features):

Konu odaklılığını belirlemek için *WordNet* (16) ile *LSA* (=Latent Semantic Analysis), yani *GAÇ* (=Gizli Anlam Çözümlemesi), yapılmalıdır (17, 10). *WordNet* ile eşanlamlıları, kapsayıcı sözcükleri ve ettirgen eylemleri (=causative verbs) bulmak mümkün iken, *LSA* ile elektronik ortamda anlam düzeyinde benzer sözcükleri yakalamak mümkün oluyor (9).

Yazar tanımlama ölçütleri anahatlarıyla yukarıda belirtilen işlemleri gerektirir, ancak bu tür işlemler yapılırken, şüpheli bir metnin hangi öğelerinin deşifre edilmesi gerektiğidir.

## 3. Bulgular ve Tartışma

**Yazar tanımlama** (=authorship profiling/author characterisation), yazılı bir metnin yazarının meçhul olduğu, yani şüphelinin olmadığı, durumlarda yazar hakkında bilgi edinmek amacıyla yapılan bir işlemidir. Bu nedenle sıklıkla yazar tanımlama, yazar profillemeye adı altında da geçer. Burada amaç, yazara ait bir profil oluşturmaktır. Bu işlem ile nokta atışı yapılamadığından, yazar hakkında ancak genel bilgiler edinilebilir. Bu işlemle örneğin, yazarın yaş grubu (10'lu yaşlar, 30-40 yaş aralığı, 65+ gibi) belirlenebilir, ancak kesin yaşı (28, 53, gibi) belirlenemez. Ya da meslek grubu belirlenebilir (sağlık sektörü gibi), ancak hangi mesleği icra ettiği (doktor, hemşire, laborant, teknisyen gibi) net söylenemez. Dolayısıyla, adli dilbilimciler şüpheli bir kaynağın yazarının:

- cinsiyetini,
- yaş grubunu,
- meslek grubunu,
- anadili, lokasyonu ve demografisini,
- kişilik özelliklerini,
- ruh sağlığını
- eğitim düzeyini,
- etnik kökenini/ırkını,
- kültür düzeyini/yapısını,
- sosyal statüsünü,
- dini inanışını,
- politik eğilimini,
- kültürel değerlerini,

dil kullanımına bakarak belirleyebilmektedirler. Bu makalede yalnızca ilk 6 (altı) öğeye yer vererek, şüpheli bir kaynağın cinsiyeti, yaş grubu, meslek grubu, anadili (dilsel kökeni) ile ağız ya da lehçe özellikleri, kişilik özellikleri, ve ruh sağlığının tespitinde başvurulması gereken yöntemlere dair bilgi verilecektir.

### 1. Cinsiyet Belirleme (=gender prediction):

**Göreceli sıklık** (=relative frequencies) özellikle şüpheli metnin yazarının cinsiyetinin belirlenmesinde kullanılan yöntemlerden biridir. Metnin içerisinde yer alan ulamların sayımı ile ortalaması hesaplanarak ulam sıklığına göre cinsiyet belirlenmesi yapılır. Bunun için, **doğal dil araçtakımı** (=natural language toolkit, nltk toolkit)'in *NLTK 3.4.5 Documentation* (<https://www.nltk.org/>) içerisinde olan **pos tagger** (=part-of-speech tagger), yani **ulam etiketleme** yazılım programı, kullanılır. Cinsiyet

belirleme ölçütlerinde gözetilen sözlükbirimsel kullanım sıklığı **Çizelge 01**'de verilmiştir. Kullanım sıklığı **yoğun** ise ↑, **düşük** ise ↓ işaretleriyle belirtilmiştir (18, 19, 20, 21). Koppel et. al. (12) yapmış oldukları çalışmalardan cinsiyet belirlemede %80 doğruluk oranı yakalamıştır (Daha fazla bilgi için ayrıca bkz. 13, 14, 22).

**Çizelge 1.** Cinsiyet belirleme ölçütlerinde gözetilen sözlükbirimsel kullanım sıklığı.

	Eril Dil Kullanımı:	Dişil Dil Kullanımı:
Ad/Özel Ad	↑	↓
Fiil	↑	↓
Sıfat	↑	↓
Belirteç/Zarf	↓	↑
Zamir	↓	↑
Bağlaç (için, ve/ile)	↓	↑
Bağlaç (diğer)	↑	↓
İlgeç/Edat	↓	↑
Soru	↓	↑
Ünlem	↑	↓
Emir/Yönerge	↑	↓
Olumsuzlama	↓	↑
Kesme	↑	↓
Şimdiki Zaman Kullanımı	↓	↑

## 2. Yaş Grubu (=age group):

Yaş grubunun belirlenmesinde birçok etmen aynı anda önemli rol oynamaktadır. Rangel et. al.'e göre (21) 10'lu yaşlarda bir birey daha çok arkadaşları ve duygularındaki iniş-çıkışlarından bahsederken, 20'li yaşlardaki bir birey üniversite yıllarından, 30'lu yaşlardaki bir birey ise evlilik hayatından (eşi ve çocuklarından), iş ortamından, mesleğinin gerekliliklerinden, ve politikadan bahsetmektedir. Bu bilgiler dikkate alınarak, yaş grubu belirleme işlemi için yapılması gerekenler aşağıda sıralanmıştır (9):

- Biçem (noktalama, büyük/küçük harf kullanımı, vs. dahil olmak üzere),
- Ulam,
- Akıcılık ve okunabilirlik/anlaşılabilirlik (İngilizce için: *Gunning Fog Index*, *SMOG Index*, *Flesch Reading Ease*, *Coleman Liau Index*, *Flesch Kincaid Grade Level*, *LIX*, *ARI (Automated Readability Index)*, *RIX (Readability Index)*),
- Sözlükbirimler, doğaçlamalar, ödünclemeler vs.,
- Bağlaşıklık örüntüleri (=collocational patterns),
- Argo, çirkinleme, hakaret ve küfür kullanımları,
- Arkaik kullanımlar (Türkçe için: Arapça, Farsça, Fransızca),
- Emoji/Emotikon kullanımı,
- Duygu sözcüklerinin kullanımı (İngilizce için: *MRC Psycholinguistic Database* [http://websites.psychology.uwa.edu.au/school/MRCDataBase/uwa\\_mrc.htm](http://websites.psychology.uwa.edu.au/school/MRCDataBase/uwa_mrc.htm)),

## 3. Meslek Grubu (=profession):

Terimlerin kullanımı bir bireyin uzmanlık alanının ya da hangi meslek grubundan olduğunun ipucunu verebilir. Bu işlemi gerçekleştirebilmek için, verinin büyüklüğüne göre manuel ya da elektronik tarama yapılabilir. Eğer verinin hacmi manuel taramaya elverişli değilse, elektronik tarama için *System Quirk* (<http://www.computing.surrey.ac.uk/SystemQ/>) benzeri terim ayıklama (=term extraction) yazılım programları kullanılabilir. Burada sözcük sıklık kullanımı (=frequency count) işlemi yapılarak, terimler sözcüklerden ayıklanır ve hangi bilimdalına ait olduğu belirlenir.

## 4. Anadillik, Lokasyon ve Demografi (=native language, location, and demography):

**Şüpheli metnin anadili** ürünü olup olmadığının sorgulanması (anadilinden sapmaların varlığı), yazarının ağız ya da lehçe kullanımı üzerinden yaşamakta ya da daha önce yaşamış olduğu coğrafyanın belirlenmesi, dilliliğinin (tekdilli, çiftilli, üçdilli, **çokdilli**, anadili ile ikinci, üçüncü, dördüncü vs. anadilinin varlığı, anadili ile yabancı dil(ler)in varlığı vs.) tespiti, amacıyla aşağıda belirtilen öğeler sorgulanmalıdır:

- Yazı sistemi
- Yazı biçimi
- Yazı boyutu
- Yazı akışkanlığı
- Satır eğimi
- Noktalama işaretleri
- Satıra sığdırma
- Yazım hataları
- Yazım farklılıkları
- Ağız ve/ya lehçe farklılıkları
- Anlam farklılıkları

**Şüpheli kaynağın anadili** (dilsel kökenini) ile ağız ya da lehçe özelliklerinin tespitinde Türkiye Türkçesi için yazarın el yazısına bakılarak Latin **yazı sistemine** alışkanlığı sorgulanır. Burada yeri gelmişken bunun grafolojik bir inceleme olmadığını da belirtmek gerekir. **Yazı biçimi**, el yazısı yerine daktilografik olması önemli bir gösterge sayılabilir. **Yazı boyutu**, sorgulanması gereken diğer önemli bir ölçüttür. Harfler değişik boyutlarda, ancak biçim itibarıyla tutarlı olmalıdır. Harf boyutlarının değişkenliği el yazısı bilen ancak yazı özelliğini gizlemek ya da yazarı okur-yazarlığı zayıf, erişkin olmayan/çocuk izlenimi yaratmak amacıyla büyük harf kullanan birisinin el ürünü, ya da anadili Türkiye Türkçesi veya Türkiye Türkçesinde eğitim görmüş biri izlenimi vermek için yapılmış olabilir. **Yazının akışkanlığı**, yazarın Latin yazı sistemine hâkim olup olmadığının göstergesidir; özellikle

sözcük içi duraksama ve kopukluklar, harfler arası düzensiz boşluklar, satır düzeninin gözetimi gibi unsurlar kişinin yazı sistemine aşinalığını gösterir. **Satır eğimi**, var ise yazarın oryantasyon sorunu yaşamakta, soldan sağa yazmakta zorlanmış görünüyor olmasına işaret edebilir; özellikle bazı harflerin aşağıya inen çizgilerinin daha uzun olması satırın aşağıya doğru eğim kazanmasına sebep olur, yani bazı harflerin sol çizgisi sağ çizgisinden daha kısa, ya da sağdaki çizgi soldakine göre aşağı doğru eğimli başlamıştır, dolayısıyla satır aşağı doğru eğimlenmiştir. Cetvel ile çizildiğinde satıra eğim veren etmenler harflerin sağ çizgilerinin sol çizgiye göre daha kısa olmasından kaynaklanmasındandır. Her bir satırın tek yönde eğimli olup olmadığı, bu eğimin sürekliliği ya da değişkenliği sorgulanmalıdır. **Noktalama işaretlerinde**, nokta, virgül, ünlem, soru işareti, üç nokta (./;/!/?/...) gibi öğelerden birinin diğeri yerine kullanımı, fazladan (./././...) ya da aşırı kullanımı (... , ... , ... , ...;...), veya kullanımının yokluğu gibi belirtiler dikkate alınmalıdır. Yazının **satıra sığması** dikkate alınması gereken diğer önemli göstergelerden biridir; zira Latin alfabesine alışkın olmayan bireylerde uzamsal kestirimlerde hata payı oldukça yüksektir. **Yazım hatalarında**, herhangi bir sözcüğün yanlış yazımı dil karmaşasından ileri gelebilir; özellikle uluslararasılaşmış sözcükler, yani ödünçlemelerde, bu durum belirgin bir hal alabilir (örn. *sempatik-sempatetik*); bunlar karma dil ürünü görünümündedirler. **Yazım farklılıklarının** ortaya konması için ise, şüpheli metnin içerisinde ve şüpheli metin ile karşılaştırılan/sorgulanan diğer metinlerin içerisinde yapılan tutarsızlıklar tespit edilmelidir. Bunlar sözcüklerin başharflerinin büyük/küçük yazılmış olması, noktalama işaretlerinin alışlagelmiş örüntüden sapması, boşluklardaki düzensizlikler, paragraf başlarında girintilerin varlığı/yokluğu gibi, unsurlardır (10). **Ağız ve/ya lehçe farklılıklarının** (örn. *İzmir'de katmer* (tur) – *Eskişehir'de çibörek* (tur); *kemik* (tur) – *sümük* (azj) gibi) tespiti ile **anlam farklılıklarının** saptanması için ise anlam çözümlemesi (=semantic analysis) ya da bileşen çözümlemesi (=componential analysis) yapılması gerekir.

**Şüpheli metin üzerinde yapılan manuel taramalar** (=manual scanning) yukarıda belirtilen öğelerden başka, dar anlamda **şu özelliklerin de gözden geçirimini gerektirir:**

- Tanımlık (=article)
- Özne (=subject)
- Nesne (=object)
- Belirteçimsi (=adverbial)
- Şahıs zamirleri (=personal pronouns)
- Çekim (=tense markers)
- Sonekler (=suffixes)

- Çoğul ekleri (=plural endings/suffixes) (-ler/-lar)
- Derilme (*not* → *no* (eng); *-mek/-mak* → *-me/-ma* (tur))
- Yardımcı fiiller (*yapmak, etmek, olmak* vs.)
- Koşaç (=copula)
- Eşdizimlilik/Söz öbekleri (=collocations/collocational patterns)
- Kalıplar (=phrases)
- Deyimler (=idioms)
- Atasözleri (=proverbs)

Şüpheli metni manuel tararken özellikle **uyuşmazlıkların** saptanması gerekir. Bunlar, tekil/çoğul halindeki özne ile ilgili eklerin eşgüdümlü olarak doğru/hatalı kullanımı, fiil (=verb) ile tümleyici (=complementiser) arasındaki uyuşmazlıklar, sayılamayan/kitle isimler (=mass nouns) ile sayılabilen/birey isimlerde (=count nouns) kullanım hataları (örn. *tuzları versene* (tur), *bilgiler birikimi* (tur) gibi), eksiltme (=deletions) ile eklemelerin (=additions) varlığı/yokluğu ve bunlardaki düzensizlikler, tümce içi öğelerin yer değiştirmesi, karmaşık dilbilgisel yapılarda (örn. eylemlilikler (=infinitives), ulaçlar (=gerunds), sıfat tamlamaları (=adjective clauses), girişirmeler/girişik tümceler (=sentence embeddings), dilek kipleri (=subjunctives) gibi) hatalardan oluşmaktadır. Yazar tanımlama açısından bu öğelerde belirgin örtüşmelerin ya da örtüşmezliklerin tespit edilmesi söz konusudur.

#### 5. Kişilik Özellikleri (=personality traits/type):

Kişilik örüntülerinin tespitinde İngilizce için *Personality Recognition at ICWSM 2013* ya da daha güncel olan *PAN 2015* (<http://pan.webis.de/clef15/pan15-web/author-profiling.html>) listesinden faydalanılabilir (15, 20):

- Açıklık (=openness)
- Disiplin (=conscientiousness)
- Dışavurumluluk (=extraversion)
- Uyuşma (=agreeableness)
- Duygudurum (=stability)

#### 6. Ruh Sağlığı (=mental health):

Dil ile düşünce arasında sıkı bağın varlığından, şüphelinin ruh sağlığına ilişkin ipuçları yakalamak mümkündür. Örneğin, duygudurum bozuklukları kapsamına giren depresyonda en yaygın belirtiler, sözcük ve tümce düzeyindeki olumsuzlamalardır (=negations). Psikoz grubunun altında yer alan şizofrenide ise durum daha karmaşık bir hal almakta ve düşünce sürecine yönelik belirtiler kendini konuşma sürecine yönelik belirtilerde son derece belirgin bir biçimde göstermektedir.

Şizofrenide görülen düşünce ve dil bozukluklarını ta-

nımlamak için, düşünce bozukluğu (=thought disorder), **formal düşünce bozukluğu** (=formal thought disorder), **düşünce kayması** (=thought splinage), çağrışım çözüklülüğü (=associative looseness), **dil bozukluğu** (=language disorder), **konuşma bozukluğu** (=speech disorder), **de-zorganize konuşma** (=disorganized speech), **bildirişim bozukluğu** (=communication disorder), gibi birçok farklı tanım kullanılmıştır (23).

Yukarıda sıralanmış **dil, konuşma ve bildirişme süreçlerine** yönelik bozukluklar aslında dört gruba ayrılır (24):

1. Ritim, ton, ve sesletim (boğumlama) kusurları
2. Kavram ve anlam bozuklukları
3. Dizimsel bozukluklar
4. Karşılıklı iletişimde (ikili konuşmalar gibi) bozukluklar

Aşağıda her grubun altında ilgili bozukluk hakkında bilgi verilecektir:

(a) Ritim, ton, ve sesletim (boğumlama) kusurları

**Ritim kusuru** (=rhythm abnormality): Şizofrenide ritim bozulmuştur. Kişi hızlı bir şekilde konuşurken birdenbire yavaş yavaş konuşmaya başlayabilir. Arada sırada konuşmada nedensiz duraklamalar da olur.

**Ton kusuru** (=intonation defect): Yavaş, fısıltı halinde konuşma, dişlerinin arasından konuşma, genizden konuşma gibi sorunlar gözlemlenebilir.

**Sesletim/boğumlama kusuru** (=pronunciation defect): Kekeleme, peltek konuşma, harfleri yanlış sesletme şeklinde ortaya çıkar. Sesletilen sözcükler uzamış ve incelmış hecelere, yani ezgisel (=prosody) olarak uygun değildir.

**Parafazik Hatalar** (=paraphasic error): Konuşmada sesbirimsel (=phonemic) ve anlambilimsel (=semantic) hataların yapılması. Sesbirimsel hatalar sözcüklerin hatalı sesletimi (=articulation) ile dil sürçmelerini (=slips of the tongue) kapsarken, anlambilimsel hatalar uygunsuz sözcük (=inappropriate word) kullanımına işaret etmektedir.

(b) Kavram ve anlam bozuklukları (=conceptual and semantic disorganisation)

**Ses çağrışımı** (=clang associations/clangings): Sözcüklerin, anlatımların anlamlarından ve dil kurallarından çok, ses uyumuna ve uyağa göre ortaya konmasıdır. Normal kişilerin düşünce akımını türlü fikirler arasındaki çağrışım bağlantıları yönlendirir. Şizofrenlerde ise bu çağrışım sözcüklerin anlamlarına göre değil, ses örüntülerine göre yapılır. Bu nedenle düşünce ve onun aksi olan konuşma parçalanmakta, bütünlüğünü kaybetmektedir. Kısacası, konuşma mantıklı değildir. Konuşmayı, söz-

cüklerin seçimini ve tümceler dizilişini sesel çağrışım lar yürütür.

**Çanta sözcükler** (=mot valise): Farklı anlamlar taşıyan sözcüklerin oluşturulmasıdır. Bazen de birkaç sözcük birleştirilerek bir sözcük haline getirilir. Bu tür sözcüklerin alıcının anlamlandırabilme olanağı yoktur.

**Yenileşim/Uydurma sözcükler/Sözcüklerin tikel kullanımını** (=neologisms/peculiar use of words): Kişinin yeni bir kavramı anlatmak istediğinde yeni anlatımlar, sözcükler ortaya koymasındır (25). Bu tür sözcükler kişi tarafından uydurulmuş, yani belirli bir kavramı karşılamak için üretilmiş yeni oluşumlardır. Bu tür kullanımlar kişiye özgü uydurma sözcüklerden oluşmaktadır. Uydurma sözcükler 4 (dört) grupta toplanır:

1. Sembolik varlıklar ve kişiler için kullanılan sözcükler,
2. yalancı terimler (=pseudoterms) kişinin hezeyanlar sonucu yapmış olduğu keşif ve icatları adlandırması için kullanılır,
3. ruhsal ve fiziksel durumlar, arzular ve cinsel istekler için kullanılan sözcükler,
4. anlamsız ve herhangi bir sisteme uymayan, özellikle kişinin kendisi için dahi anlam taşımayan sözcükler. (26, 27, 28)

**Sözcük Yakıştırmaları**(=word approximations): Morfolojik düzeyde gerçekleşen yenileşim olgusunun tam tersine sözcük yakıştırması semantik düzeydedir ve dilde var olan sözcüklerin kişi tarafından kastettiğinden başka bir kavramı karşılamak için kullanılması anlamına gelmektedir. Aynı zamanda sıradışı ve kişiye özgü (=idiosyncratic) kullanımlardır. Anlamca yakın gelen iki sözcükten birinin diğerinin yerine kullanılmasıdır (örn. ısı-sıcaklık, ısı-ateş, salam-sosis, salam-sucuk gibi).

**İkili duygu - Ambivalans** (=ambivalence): Bir duyarlılık bozukluğudur. İstemek ve istememek, sevmek ve nefret etmek, bilmek ve bilmemek, harekete geçmek ve geçmemek gibi zıt duyum ve haller andaş ve birlikte bulunur. Bu durumdaki hastalarda düşünce de iki kutup arasında dalgalanır. Bu durumlarda ikili duyguyu ifade eden **zıtanamlı sözcükler** birleştirilerek kullanılır.

(c) Dizimsel bozukluklar (=peculiar sentences)

**Dikişsiz konuşma** (=asyntaxie): Tümcenin yokluğu halinde dilbilgisi kurallarının tamamen ortandan kalkmasıdır. Kimi zaman eşanamlı sözcüklerin arka arkaya sıralandığı dikkati çeker (=stereotypy). Tümceler kopuk kopuk gelir. Böylece konuşmada bilgiyi aktarma gücü ortadan kalkar. Bu bozukluk gittikçe birbiri ile ilgisi olmayan sözcüklerin ardı ardına sıralanması durumunda sözcük salatasına dönüşür.

**Tutarsızlık/Enkoherans/Sözcük salatası/Şizofazi**



(=*incoherence/word salad/schizaphasia*): Kişinin kullandığı tümcelerdeki sözcüklerin çoğunun diziliş ve anlam yönünden birbiriyle ilişkili olmaması, tutarsızlık göstermesidir; yani hastanın sorulara yanıt verirken anlatımının karışık olması ve tutarlı bir sonuca ulaşamamasıdır. Anlaşılması dinleyici için çok güç olan, birçok çağrışımın veya birçok konunun bir arada bulunduğu ve art arda dizildiği, sözcükler veya tümceler arasındaki bağlantılarda mantıklı geçişlerin yapılmadığı bir konuşmadır. Konuşma adeta bir “*sözcük salatası*” niteliği taşır. Dilbiliminde metinsellik (=cohesion) ölçütlerinden sözcüksel (=lexical) ve dilbilgisel (=grammatical) anlamda kuraldışılıklar vardır. Sözcüksel düzeyde seyreden bağdaşıklık (=lexical cohesion), dilbilgisel düzeyde seyreden tutarlılık (=grammatical cohesion) diye adlandırılmıştır. Bu kuraldışılıklar kendini sözcüksel düzeyde yineleme (=repetition) ve eşdizimlilik (=collocations) hataları olarak gösterirken, dilbilgisel düzeyde ilgi (=reference), belirteç (=identification), eksiltme (=ellipsis), ve bağlaç (=conjunction) hataları olarak gösterir. Bir başka sınıflandırmaya göre **bağdaşıklık** (=coherence) bir metindeki anlamsal ve mantıksal uyumu araştırır. Prag Okulu’ndan Kıta Avrupası Yapısalcıları’nın önerdikleri tema (=theme) ve rema (=rheme) öğelerinin dikkate alınarak tematik ilerlemenin (=thematic progression) aşağıda belirtilen beş ölçüte göre dikkate alınması gerektiğidir (29, 30):

1. Tekdüze ilerleme (=simple linear progression)
2. Değişmeyen tema (=constant theme)
3. Türevli tema (=derived theme)
4. Ayrık rema (=split rheme)
5. Tema sürçmesi/sekmesi (=thematic lapse)

**Bağlaşıklık** (=cohesion) ise, sözcüksel ve dilbilgisel öğelerin birbiri ile bağlantılanmasını ve uyum içerisinde kullanımını araştırır. *Sözcüksel* ve dilbilgisel bağlaşıklık diye ikiye ayrılır. **Sözcüksel bağlaşıklıkta** bağdaştırmalar (=collocations), yineleme (=reiteration/repetition), tümel eşanlamlılık ile tikel eşanlamlılık (=absolute synonym and near-synonymy), zıtanlamlılık (=antonymy), üstterim (=superordinate term), genel ad (=general noun) ve eşdizimlilik (=collocation) gibi öğeler dikkate alınırken, **dilbilgisel bağlaşıklıkta** gönderim (=reference) (özellikle içgönderimde (=endophora) artgönderim (=anaphora) ile öngönderim (=cataphora) ve dışgönderimde (=exophora) durumsal ve metinsel gönderimler, değiştirme (=substitution), eksiltme (=ellipsis) ve bağlaç (=conjunctions) gibi öğeler dikkate alınır (29, 31, 32). Tutarlılık (=consistency) ise kesit dilin, zamanın, şahıs zamirlerinin, çoğul takılarının kullanımında tutarlılık gibi sözlü/yazılı metin içerisindeki tutarlılıkların gözetilmesidir (29, 30).

(d) Bildirişim bozuklukları/Karşılıklı iletişimde (ikili konuşmalar gibi) bozukluklar (=communication disorders/dialogical disorders)

**Konuşmama** (=mutism): Bilinçli ya da bilinçdışı nedenlerle konuşmayı reddetmedir.

**Konuşma Yoksunluğu** (=poverty of speech): Konuşma gereğinden kısa, sorulara verilen yanıtlar gelişigüzelidir. Kurulan tümceler kısa ve kullanılan sözcüklerin miktarı azdır.

**Konuşma İçeriği Yoksunluğu** (=poverty of content of speech): Sözcük sayısı fazla, tümceler uzun olsa da verilen yanıtların sorulan sorular için yetersiz kalması, konuşmanın yeteri kadar bilgilendirici olmaması durumudur.

**Basıncılı konuşma** (=pressured speech/pressure of speech/logorrhea/tachylalia/tachyphasia/lalorrhea): Konuşma miktarında kendiliğinden artıştır. Kişi hızlı ve çok konuşur, konuşmayı kesme yetersizliği vardır. Yeni konulara geçmek için sabırsızlık gösterir ve tümceleri tam olarak bitiremeyebilir. Basıncılı konuşma kısacası, spontan konuşma miktarında olağan veya toplumsal normlara göre artma olmasıdır. Kişi hızlı konuşur ve konuşmayı kesmek mümkün değildir.

**Çevresel konuşma/Hedefe ulaşamama** (=circumstantiality/weakening of goal): Sorulan sorunun bir hayli ayrıntılı düşünülüp gereksiz birçok bilginin sunulduğu ancak bir türlü hedefe ulaşamadığı konuşmadır. Aşırı dolaylı/imalı anlatımlarla, gerektiğinden fazla genellemenin yer aldığı bu tür konuşma, gereksiz ayrıntılarla da donatılmıştır.

**Ayrıntıcılık** (=pedantism): Kişinin konuşmasının gereksiz ayrıntılarla donatılmış olması ve bu nedenle amaca ulaşmada gecikmesidir.

**Teğet konuşma/geçme ya da Amacın kaybolması** (=tangentiality): Sorulara doğrudan değil, dolaylı olarak ya da tamamen ilgisiz yanıtlar vermedir. Konuşma başlangıçta anlaşılır gelirse de tümceler arasındaki bağlantılar mantıksaldır ve uygun dilbilgisel yapılar kullanılmıştır, ancak birey amaca varmada başarısızdır. Kısacası, belirli bir düşünce izlenmesine rağmen hasta asıl amaca ulaşmakta sıkıntı yaşar.

**Raydan çıkma/Çağrışım gevşekliği** (=derailment/loosening of associations): Özellikle spontane konuşmalarında, tümceler arasında mantıklı geçişin olmamasıdır. Kişi farkında olmadan konuştuğu konudan sapar ve başka bir konuya geçiş yapar.

**Mantıkdışılık** (=illogicality/non-logical reasoning/peculiar reasoning): Dinleyen için konuşma örüntüsünde bir mantıksallık yoktur, ancak verilen yanıt sorulan soru için mantıksız olsa da kendi içerisinde tutarlı bir yanı vardır. Örnekseme ya da çıkarsama açısından mantıksal sorunlar gözlemlenir.

**Benci Konuşma (=egocentric speech/self reference):** Kişi her konuyla kendisi arasında bir bağlantı kurar. Konuşmanın ana konusunu kendisi ile ilişkilendirir.

**Distraktibl Konuşma/Odaklanamama (=distractibility):** Normal seyirinde devam eden bir konuşma çevresel uyaran(lar) sebebiyle birden kesilir ve kişi etraftaki bu uyaran(lar) ile ilgili konuşmaya başlar. Kişinin düşüncelerine odaklanma sorunlarından ileri gelen bir konuşma bozukluğudur.

**Blok (=blocking):** Konuşma akışının görünür bir neden olmadan, bilinçdışı nedenlerle aniden kesilmesidir. Kişi birden durur, konuşmasını kaldığı yerden sürdürmez, ancak toparlama çabası içerisindedir. Blok, kısacası bir düşünce akışının birden sonlanması, düşünce akışının yarıda kesilmesi durumudur.

**Perseverasyon (=perseveration):** Bir sözcüğün, bir tümcenin, bir konunun ya da bir düşüncenin kalıplaşmış biçimde ve ısrarla yinelenmesi, konuşma süresince o sözcüklere, tümcelere ya da düşüncelere geri dönülmesi durumudur.

**Konuşma yankılanması/Ekolali (=echolalia):** Kişinin konuşmalarında verici konumundaki kişilerin kullandığı sözcük, söz öbeği ya da tümcelerin çoğunu, kimi zaman aynıysa yineleyerek kullanmasıdır (33). Yani, kişinin kendi sözcüklerini ve konuşmalarını değil de, görüşmecinin sözcüklerini ve konuşmalarını tekrarladığı görülür.

**Verbijerasyon (=verbigeration):** Bazı sözcük ve anlamların amaçsızca, uzun süre yinelenmesidir. Bu durum uzun süredir hastanede yatmakta olan kronik ve regressive bireylerde görülür; toplum içindeki bireylerde görülmez (33).

**Küfür ve argo (=coprolalia):** Şizofrenlerde müstehcen ve küfürlü konuşmaya sıkça rastlanır. Bunun nedeninin aşırı duygu yüklü olmakla ilgisi olduğu düşünülmektedir.

**Tumturaklı Konuşma (=stilted speech/magniloquence):** Tuhaf olan bu dil kullanımı arkaik, görkemli/gösterişli, ağdalı/süslü, gerektiğinden nazik veya resmi olabilir.

**Özellikle konuşma yoksunluğu ile hedefe ulaşamama unsurları düşünce ve konuşmanın fakirleşmesi ölçümü için gereklidir. Gevşeklik, sözcüklerin tikel kullanımı, dizimsel bozukluklar ve mantıkdışılık ise psikotik bozukluklarda sıkça görülen dezorganize düşünce ve konuşmanın ölçümü için önemlidir; ayrıca psikotik bozukluğu teşhisini koymada dikkate alınan dört önemli kriterdir. Perseverasyon ve odaklanamama unsurları ise psikotik bozukluklara özgü olmasa da, psikotik hastalıklarda sıkça rastlanan olgulardır ve bu yüzden değerlendirmede dikkate alınır.**

Şizofrenik bireylerin üzerinde yapılan bazı araştırmalar sonucu dil ve düşünce performanslarına yönelik sonuçlar alabilmek ve çıkarımlarda bulunabilmek adına

birçok farklı araştırmacı tarafından dil ve düşünce bozuklukları ölçekleri oluşturulmuştur (34):

1. Scale for the Assessment of Thought, Language and Communication (TLC) – (35, 36)
2. Thought Disorder Index (TDI) – (37)
3. Bizarre-Idiosyncratic Thinking Scale (BITS) – (38)
4. Latent Semantic Analysis (LSA) – (17)
5. Communication Disturbances Index (CDI) – (39, 40)
6. Clinical Language Disorder Rating Scale (CLANG)– (41)
7. The Schizophrenia Communication Disorder Scale (SCD) – (42)
8. Thought and Language Index (TLI) – (43)
- i. Düşünce ve Dil Ölçeği (DDÖ) – (23, 44)
9. Formal Thought Disorder Scales (FTD) – (45)
10. Thought and Language Disorders Scale (TALD) – (46)
- Düşünce ve Dil Bozukluğu Ölçeği (DDBÖ) – (47)

Bu ölçekler klinik kullanımlarda yer alıp tanıda yardımcı unsur olarak dikkate alınmaktadır. *Thought and Language Index (TLI)* Liddle et. al. tarafından 2002 yılında yayımlanmış (43) akabinde Ulaş (23) ile Ulaş et. al. (44) tarafından geçerlilik ve güvenilirlik çalışması yapılarak Türkiye’de kullanıma sunulmuştur. *Thought and Language Disorders Scale (TALD)* ise Mutlu et. al. (47) tarafından geçerlilik ve güvenilirlik çalışması yapılarak Türkiye’de kullanıma sunulmuştur. Her iki ölçeğin de, İngilizceden çevirisi ile geçerlilik ve güvenilirlik çalışması yapılmış ise de Türkiye Türkçesine özgü bir ölçek niteliği barındırmadığından Türkçedeki dilsel ipuçlarını ıskalamaktadır. Bu nedenle, Türkiye’de yapılacak gerek adli dilbilimsel, gerek klinik dilbilimsel çalışmalar için Türkiye Türkçesinde hazırlanmış özgün bir düşünce, dil ve bildirişim bozuklukları ölçeğinin zihinsel sözlükçe uzmanlık konusu olan dilbilimciler tarafından geliştirilmesine gereksinim duyulmaktadır.

#### 4. Sonuç

Yazar tanımlama öğelerinden cinsiyet, yaş grubu, meslek grubu, anadillilik, lokasyon ve demografi ile kişilik özellikleri ve ruh sağlığını belirlemeye dair betimlemiş olduğumuz tanımlama ölçütleri ile yöntemler kimi kurgulanmış, kimi özgün kaynaklardan elde edilmiş verilerin araştırılması sonucu birikim ve deneyimlere dayalı geliştirilmiş ve günümüzde kullanılagelmekte olan ölçüt ve yöntemlerdir. Daha önce belirttiğimiz üzere, yazar tanımadan farklı olarak yazar tanımlamada nokta atışı yapılamamaktadır. Ancak şüpheli kişinin vermiş olduğu ipuçlarından hareket ederek, her bir öge için değişken olarak, %60-80 doğruluk oranıyla teşhis konabilmektedir. Önemli olan bundan sonra yapılacak çalışmalarla sonuçların güvenilirliğini arttırmak amacıyla bu oranın

yükselmesini sağlayabilecek yeni yöntemlerin geliştirilmesidir. Türkiye’de adli kurumlarda adli dilbilimi hakkında henüz bir farkındalık oluşmadığından, özgün kaynaklara erişim gücü nedeniyle benzetim (=simulation) amaçlı kurgusal veri ve derlemlerle araştırmalar yapma zorunluluğu kaçınılmazdır.

## Kaynaklar

1. Olsson J. Forensic Linguistics. London & New York: Continuum, 2008. ISBN-10: 1441170766. ISBN-13: 978-1441170767.
2. Coulthard M. & Johnson A. An Introduction to Forensic Linguistics: Language in Evidence. London & New York: Routledge, Taylor & Francis Group, 2007. ISBN-10: 1138641715. ISBN-13: 978-1138641716.
3. Aslantürk O. TAMGACI: Artırımsal ve Geri Beslemeli Türkçe Yazar Çözümleme. Turkish Authorship Analysis With An Incremental And Adaptive Model. Doktora Tezi. Tez Danışmanı. Prof. Dr. Hayri SEVER. Ankara: Hacettepe Üniversitesi, 2014.
4. Coulthard M. Author Identification, Idiolect, and Linguistic Uniqueness. Applied Linguistics. In: Applied Linguistics, Volume 25, Issue 4, December 2004, pp. 431–447. Published: 01 December 2004. DOI: <https://doi.org/10.1093/app-lin/25.4.431>, 2004.
5. Coulthard M. The Routledge Handbook of Forensic Linguistics. New York: Taylor & Francis, 2010. ISBN-10: 0415837235. ISBN-13: 978-0415837231.
6. Grant T. Authorship Attribution in a Forensic Context. Yayımlanmamış Doktora Tezi. Unpublished Doctoral Thesis. Birmingham: Birmingham University, 2005.
7. McMenamin G. R. Forensic Linguistics – Advances in Forensic Stylistics. Boca Raton, London, New York, Washington D. C.: CRC Press, 2002. ISBN: 0849309662.
8. Juola P. Authorship Attribution. Boston & Delft: now – The Essence of Knowledge, 2008. DOI: <https://doi.org/10.1561/1500000005>.
9. Meina M., Brodzińska K., Celmer B., Czoków M., Patera M., Pezacki J., Wilk M. Ensemble-based Classification for Author Profiling Using Various Features Notebook for PAN at CLEF 2013. <https://www.uni-weimar.de/medien/webis/events/pan-13/pan13-papers-final/pan13-author-profiling/meina13-notebook.pdf>, 2013.
10. Stamatatos E. A Survey of Modern Authorship Attribution Methods. In: Journal of the American Society for Information Science and Technology, Vol. 60, Issue 3, March 2009, pp. 538-556, DOI: <https://doi.org/10.1002/asi.21001>, 2009.
11. Şeker S. E. Veri Madenciliği – Bilgisayar Kavramları <http://bilgisayarkavramlari.sadievrenseker.com/2013/07/01/jaccard-indeksi-jaccard-index/>, 2013.
12. Koppel M., Argamon S., Shimoni A. R. Automatically Categorizing Written Texts by Author Gender. In: Literary and Linguistic Computing, Vol. 17, No. 4, pp. 401–412, DOI: <https://doi.org/10.1093/lc/17.4.401>, 2002.
13. Love H. Attributing Authorship. Cambridge: Cambridge University Press, 2002.
14. Argamon S., Koppel M., Fine J., Shimoni A. R. Gender, Genre, and Writing Style in Formal Written Texts. In: Text & Talk – An Interdisciplinary Journal of Language, Discourse and Communication Studies, Vol. 23, Issue 3, pp. 321–346. Ed. Sarangi Srikant. Ağ Erişimi: <https://www.degruyter.com/view/j/text.1.2003.23.issue-3/text.2003.014/text.2003.014.xml>. DOI: <https://doi.org/10.1515/text.2003.014>, 2006.
15. Reddy R., Vardhan V., Reddy V. A Survey on Authorship Profiling Techniques. In: International Journal of Applied Engineering Research, Volume 11, Number 5 (2016), pp. 3092-3102. ISSN 0973-4562. Research India Publications. <http://www.ripublication.com>, 2016.
16. Fellbaum C. WordNet: An Electronic Lexical Database. Cambridge: MIT Press, 1998.
17. Deerwester S., Dumais S., Furnas G.W., Landauer T. K., Harshman R. Indexing by Latent Semantic Analysis. In: Journal of the American Society for Information Science, 41(6), pp. 391-407, 1990. DOI: [https://doi.org/10.1002/\(SICI\)1097-4571\(199009\)41:6<391::AID-AS11>3.0.CO;2-9](https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-AS11>3.0.CO;2-9).
18. Küçükyılmaz T., Cambazoğlu B. B., Aykanat C., Can F. Chat Mining for Gender Prediction. In: Lecture Notes in Computer Science, Vol. 4243, p. 274283, 2006.
19. Pennebaker J. W. The Secret Life of Pronouns: What Our Words Say About Us. Bloomsbury Press, 2011.
20. Rangel F. & Rosso P. Use of Language and Author Profiling: Identification of Gender and Age. In: NLPCS 2013 CIRM – 10th International Workshop on Natural Language Processing and Cognitive Sciences, Marseille, France, 13-17 October 2013, pp. 177-186, 2013.
21. Rangel F., Rosso, P., Koppel M., Stamatatos, E., Inches G. Overview of the Author Profiling Task at PAN 2013. In: CEUR Workshop Proceedings Vol. 1179, Forner P., Navigli R., Tufis D. (eds.), Notebook Papers for PAN at CLEF 2013 LABs and Workshops, CLEF-2013, Valencia, Spain, September 23-26, <https://CEUR-WS.org>, 2013.
22. Nini A. Authorship Profiling in a Forensic Context. PhD Thesis. Birmingham: Aston University, 2014.
23. Ulaş H. Düşünce ve Dil Ölçeği Türkçe Formunun Şizofreni Hastalarında Türkçede Geçerlilik ve Güvenilirlik Çalışması. Uzmanlık Tezi. Dokuz Eylül Üniversitesi, Psikiyatri Anabilim Dalı, İzmir, 2005.
24. Songar A. Dil ve Düşünce. Cerrahpaşa Tıp Fakültesi Psikiyatri Kliniği Vakfı Yayınları. No: 7. İstanbul: Gür-Ay Matbaası, 1986.
25. Başkak B., Özel T. E., Atbaşoğlu E. C., Başkak S. Peculiar Word Use as a Possible Trait Marker in Schizophrenia. In: Schizophrenia Research, 2008; 103:311-317 DOI: <https://doi.org/10.1016/j.schres.2008.04.025>.
26. Aksan D. Anlambilimi ve Türk Anlambilimi: Ana Çizgileriyle. Ankara: Ankara Üniversitesi Dil Tarih ve Coğrafya Fakültesi, 1971.

27. Aksan D. Anlambilim: Anlambilim Konuları ve Türkçenin Anlambilimi. Ankara: Engin Yayınevi, 1998.
28. Cruse A. D. Meaning in Language: An Introduction to Semantics and Pragmatics. Oxford: Oxford University Press, 2000.
29. Brinker K. Linguistische Textanalyse: Eine Einführung in Grundbegriffe und Methoden. 6th ed. Berlin: Erich Schmidt Verlag, 2005.
30. Daneš F. Zur linguistischen Analyse der Textstruktur. In: Folia Linguistica 4: 72-78, 1970.
31. Brown G. & Yule G. Discourse Analysis. Cambridge: Cambridge University Press, 1993.
32. Hellmann C. The Notion of Coherence in Discourse. In: Focus and coherence in discourse Processing, pp. 190-202. Eds. Rickheit, Gert & Christopher Habel. Berlin & New York: de Gruyter, 1995.
33. Manschreck T. C., Maher B., Hoover T., Ames D. Repetition in Schizophrenic Speech. In: Language and Speech, Vol: 28, No: 3, Pp. 255-268. Kingston Press, 1985.
34. Kleiger J. H & Khadivi A. Assessing Psychosis: A Clinician's Guide. New York & London: Routledge, Taylor & Francis Group, 2015.
35. Andreasen N. C. Thought, Language, and Communication Disorders. I. Clinical Assessment, Definition of Terms, and Evaluation of Their Reliability. In: Archives of General Psychiatry, 1979 Nov; 36(12): 1315-1321, 1979a.
36. Andreasen N. C. Thought, Language, and Communication Disorders. II. Diagnostic significance. In: Archives of General Psychiatry, 1979 November; 36(12):1325-1330, 1979b.
37. Johnston M. H. & Holzman P. S. Assessing Schizophrenic Thinking. San Francisco: Jossey-Bass, 1979.
38. Marengo J. T., Harrow M., Lanin-Kettering I., Wilson A. Evaluating Bizarre-idiosyncratic Thinking: A Comprehensive Index of Positive Thought Disorder. In: Schizophrenia Bulletin 12(3):497-511. February 1986. DOI: <https://doi.org/10.1093/schbul/12.3.497>, 1986.
39. Docherty N. M., DeRosa, Andreasen N. C. Communication disturbances in schizophrenia and mania. In: Archives of General Psychiatry, No: 53, pp. 358-364, 1996.
40. Docherty N. M. Communication Disturbances Index: Rating Manual, 1996.
41. Chen E. Y. H., Lam, L. C. W., Kan C. S., Chan C. K. Y., Kwok C. L., Nguyen D. G. H., Chen R. Y. L. Language Disorganisation in Schizophrenia: Validation and Assessment with a New Clinical Rating Instrument. In: Hong Kong J. Psychiatry, 6 (1), pp. 4-13, 1996.
42. Olivier V., Hardy-Baylé M. C., Lancrenon S., Fermanian J., Sarfati Y., Passerieux C., Chevalier J. F. Rating Scale for the Assessment of Communication Disorders in Schizophrenics. In: European Psychiatry, Vol. 12, Issue 7, Ss. 352-361, 1997.
43. Liddle P. F., Elton T. C., Stephanie L. Thought and Language Index: An Instrument for Assessing Thought and Language in Schizophrenia. In: Br J Psychiatry, 181: 326-330, 2002.
44. Ulaş H., Alptekin K., Özbay D., Akdede B. B., Çakır E., Türmüklü M., Şimşek S., Arkar H., Akvardar Y. Düşünce ve Dil Ölçeğinin Türkçe Formunun Geçerlilik ve Güvenilirlik Çalışması. Klinik Psikiyatri. Sayı: 10, Ss. 77-85, 2007.
45. Barrera A., McKenna, P., Berrios, G. E. Two New Scales of Formal Thought Disorder in Schizophrenia. In: Psychiatry Research 157(1-3):225-34, February 2008, DOI: <https://doi.org/10.1016/j.psychres.2006.09.017>, 2008.
46. Kircher T., Krug A., Stratmann M., Ghazi S., Schales Ch., Frauenheim M., Turner L., Fahrman P., Hornig T., Katzev M., Grosvald M., Müller-Isberner R., Nagels A. A Rating Scale for the Assessment of Objective and Subjective Formal Thought and Language Disorder (TALD). In: Schizophrenia Research, Volume 160, Issues 1-3, December 2014, pp. 216-221. DOI: <https://doi.org/10.1016/J.SCHRES.2014.10.024>, 2014.
47. Mutlu E., Yazıcı M. K., Barışkın E., Ertuğrul A., Gürel Ş. C., Gürkan, Ş., Göka E. Yağcıoğlu A. E. A. Examination of Formal Thought Disorder and its Clinical Correlates with the Turkish Version of The Thought and Language Disorder Scale (TALD-TR) in Schizophrenia. In: Comprehensive Psychiatry, Volume 93, 2019, pp. 7-13, ISSN 0010-440X, DOI: <https://doi.org/10.1016/j.comppsy.2019.06.003>, 2019.

#### Ağ Sayfaları:

**Dale-Chall List of Simple Words** <https://www.usingenglish.com/resources/wordcheck/list-dale-chall+list+of+simple+words.html>

**Forensic Linguistics** <http://www.forensiclinguistics.eu/forensic-linguistics.html>

**Grammarly** <https://www.grammarly.com/>

**ISO 27 Listesi** <http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#lqissue-typevalues>

**Language Tool** <https://languagetool.org/>

**MRC Psycholinguistic Database** [http://websites.psychology.uwa.edu.au/school/MRCDataBase/uwa\\_mrc.htm](http://websites.psychology.uwa.edu.au/school/MRCDataBase/uwa_mrc.htm).

**NLTK 3.4.5 Documentation** <https://www.nltk.org/>

**NodeBox** <http://nodebox.net/code/index.php/Linguistics>

**NTV Haber** <http://arsiv.ntv.com.tr/news/111935.asp>

**PAN 2015** <http://pan.webis.de/clef15/pan15-web/author-profiling.html>

**System Quirk** <http://www.computing.surrey.ac.uk/SystemQ/>

**The Text** [http://www.thetext.co.uk/learn\\_about\\_forensic\\_linguistics.html](http://www.thetext.co.uk/learn_about_forensic_linguistics.html)