03. Tarihî Türkçe metinlerin doğal dil işleme yöntemleri ile incelenmesi

Yasemin KUBİLAY[1]

Meriç GÜVEN[2]

APA: Kubilay, Y. & Güven, M. (2023). Tarihî Türkçe metinlerin doğal dil işleme yöntemleri ile incelenmesi. *RumeliDE Dil ve Edebiyat Araştırmaları Dergisi*, (Ö12), 23-36. DOI: 10.29000/rumelide.1330375.

Öz

Son yıllarda bilişim teknolojileri alanındaki gelişmelere bağlı olarak dilleri, bilgisayarlı dilbilim ve doğal dil işleme (DDİ) yöntemleriyle incelemek, çözümlemek ve yapılandırılmış verilere dönüştürmek olanaklı hâle gelmiştir. Türkçenin tarihî dönemleri ile ilgili eş dizimlilik, birliktelik kullanımı, semantik prozodi ve semantik harita çalışmalarının azlığı bizi bu konuları hesaplamalı perspektifle inceleyen bir çalışma yapmaya sevk etmiştir. Araştırmamızda tarihî Türkçe ile yazılmış eserlerde de semantik ağları görüntülemek için münasip kelime hazinesinin mevcut olduğu görülmüştür. İki veya daha çok kelimenin alışkanlıktan kaynaklanan birlikte kullanımları şeklinde tanımlanan eş dizimlilik kavramına göre, bazı kelimeler yalnızca belirli sözcüklerle kullanılma temayülü gösterir. Birlikte sıkça kullanılan kelimeler zamanla aynı çağrışım özelliklerini kazanır, anlam ve biçim bakımından kalıplaşır. Bir biçime eş dizimlilikleri tarafından aşılanan kalıcı istikrarlı anlam aurası ise semantik prozodi olarak tanımlanır. Bu anlam aurası, incelememiz sırasında somut bir şekilde karşımıza çıkmıştır. Çalışmamızda seçilen her kavramın Python programlama dilinde, GraphViz Kütüphanesi ile oluşturulan semantik haritası üzerinde anlam aurası gösterilmiştir. Ayrıca "eş dizimli" ögeleri tespit etmek üzere makine öğrenmesinin bir alanı olan DDİ yöntemlerinden kelime yerleştirme (word embedding/vectorization) ile metin madenciliğinde, makine öğrenmede ve DDİ tekniklerinde yararlanılan GloVe kütüphanesi kullanılarak hazırladığımız yazılım ve Tensor Flow Kelime Yerleştirme Projektörü yazılımı kullanılmıştır. Metinler, bilgisayar ortamına düz metin dosyası (.txt) olarak aktarılmış, metinlerdeki eş dizimli kelime varlığı yazılımlara işlenmiş ve Log-likehood, MI değeri, T-skoru, Dice coefficient değeri gibi farklı istatistik analizleri ile de eş dizimlilikler incelenmiştir. Metinlerdeki eş dizimliliklere ait istatistikleri, kelime sıklıkları, vektörel temsillerinin DDİ'de kullanılan yazılımlar aracılığıyla genel olarak çıkarımları amaçlanmıştır. Sayısallaştırılan metin istatistikî değerleriyle işlenmiş ve ulaşılan veriler görsel olarak gösterilmiştir. Araştırmamız, özellikle Türkçenin tarihî dönem eserlerinin daha iyi anlaşılmasına katkı sağlayacak; eş dizimli yapıların tespiti sayesinde tarihî metinlerde satır arası sözcüklerin anlamlandırılması, anlam auralarının keşfi kolaylıkla yapılabilecektir. Gelişen bilgisayar teknolojileri sayesinde yeni yazılımlar ile anlam haritalarının, sözcüklerin vektörel şemalarının ve anlam auralarının somut gösterimi sağlanacaktır.

Anahtar kelimeler: Doğal dil işleme, birliktelik kullanımı, eş dizimlilik, semantik prozodi, semantik harita

1 Doktora Öğrencisi, Uşak Üniversitesi, Fen Edebiyat Fakültesi, Lisansüstü Eğitim Enstitüsü, Türk Dili ve Edebiyatı, Yeni Türk Dili (Uşak, Türkiye), ykubilay@yahoo.com, ORCID ID: 0000-0003-1478-7062 [Araştırma makalesi, Makale kayıt tarihi: 21.06.2023-kabul tarihi: 20.07.2023; DOI: 10.29000/rumelide.1330375]

2 Doç. Dr., Uşak Üniversitesi, Fen Edebiyat Fakültesi, Türk Dili ve Edebiyatı Bölümü (Uşak, Türkiye), meric.guven@usak.edu.tr, ORCID ID: 0000-0003-2533-5272

Analizing historical Turkish texts with NLP

Abstract

In recent years, depending on the developments in the field of information technologies, it has become possible to analyze and transform languages into structured data with computer linguistics and natural language processing (DDI) methods. The scarcity of collocation, use of association, semantic prosody and semantic map studies related to the historical periods of Turkish has prompted us to conduct a study that examines these issues from a computational perspective. In our research, it has been seen that there is a suitable vocabulary for displaying semantic networks in works written in historical Turkish. According to the concept of collocation, which is defined as the habitual use of two or more words, some words tend to be used only with certain words. Words that are frequently used together gain the same connotation features over time and become stereotyped in terms of meaning and form. The permanent stable aura of meaning instilled in a form by its collocations is defined as the semantic prosody. This aura of meaning came up concretely during our investigation. In our study, the meaning aura of each selected concept is shown on the semantic map created with the GraphViz Library in the Python programming language. In addition, the software we prepared using the word embedding/vectorization, a field of machine learning, word embedding/vectorization and the GloVe library used in text mining, machine learning and DDI techniques, and the Tensor Flow Word Placement Projector software were used to detect "collocation" elements. The texts were transferred to the computer as a plain text file (.txt), the collocation word presence in the texts was processed into the software, and the collocations were examined with different statistical analyzes such as Log-likehood, MI value, T-score, Dice coefficient value. It is aimed to infer the statistics of collocations in the texts, word frequencies and vectorial representations in general through the software used in DDI. The digitized text was processed with statistical values and the data reached were shown visually. Our research will contribute to a better understanding of the historical period works of Turkish; Thanks to the detection of collocation structures, the interpretation of interlinear words in historical texts and the discovery of meaning auras will be made easily. Thanks to the developing computer technologies, new software will provide concrete representation of semantic maps, vectorial schemes of words and meaning auras.

Keywords: NLP, comitatives, collocations, semantic prosody and semantic maps

Doğal dil işleme yöntemsel olarak esasen bilgisayarlı dil biliminin bir dalıdır. Hesaplamalı dilbilim olarak da literatürde yer alan istatistiksel verilerden yararlanan yapay zekâ yöntemleri dilimizin en eski eserlerinden başlayarak son dönem yazılı ve sözlü tüm eserlerin kullanıma uygundur.

Doğal Dil İşleme çalışmaları kapsamında aşağıdaki sıralanan konuları görmekteyiz:

- Yazım yardımcı araçlarının geliştirilmesi
- Yazım yanlışlarının düzeltilmesi
- Bul ve değiştir
- Basılı bir metni okuma (optik olarak metin okuma) ve okuma yanlışlarını düzeltme
- Bir metnin özetini çıkarma

- Metnin içerdiği bilgiyi çıkarma
- Bilgiye erişim
- Metni anlama
- Bilgisayarla sesli etkileşim
- Bilgisayarın konuşması (metni seslendirme)
- Konuşmayı anlama (konuşmayı metne dönüştürme)
- Soru yanıt dizgeleri
- Yabancı dil okuma yardımcı araçları
- Yabancı dilde yazma yardımcı araçları
- Doğal diller arası çeviri (Adalı 2020:19)

Bu anlamda çeşitli üniversitelerin söz gelimi ODTÜ, Sabancı, Hacettepe, Medeniyet Üniversitesi, Çukurova, Mersin, Ankara Üniversitesi, Hacı Bayramı Veli, Dokuz Eylül ve Uşak Üniversite'lerinin çalışmaları mevcuttur. Bu eksende bizim bilimsel araştırma ve TÜBİTAK projelerine başvurumuz olmuştur. İş birliğimizin bulunduğu Dokuz Eylül ve Uşak Üniversite'leri ile 2018 yılından bu yana çalışmalarımız sürmektedir. Yaptığımız ve yapmakta olduğumuz çalışmalardan sizleri bazı örneklerle bilgilendirmek istiyoruz:

1. Dede Korkut Hikayeleri'nin 13. Boyu: "Salur Kazan'ın Yedi Başlı Ejderhayı Öldürmesi"nde eş dizimli öge (küme ve dizi)lerin bilgisayarlı dil bilim yöntemleri ile incelenmesi[3]

"Salur Kazan'ın Yedi Başlı Ejderhayı Öldürmesi" adı verilen bu destanî anlatma Dede Korkut anlatmalarının sayısını on üçe çıkarmıştır. Öte yandan Azmun'a göre bu nüsha 27 soylama ile 2 ayrı (13. ve 14.) boylamadan müteşekkildir. Bunlar: 'Salur Kazan'ın Aras Suyu ile Kars Kalesi'ni Aldığı Boy' ile 'Salur Kazan'ın Yedi Başlı Ejderhayı Öldürdüğü Boy'dur. Bu çalışma dilimizde eş dizimlilik kavramını ve Dede Korkut Hikayeleri'nin 13. Boyu'nda geçen eş dizimli öge (küme ve dizi)lerin incelenmesini esas almıştır. İki temel konudan oluşan çalışma, alanyazınında eş dizimlilik araştırmalarını ve bilgisayarlı dil bilim yöntemleri ile "ikili ve üçlü eş dizimleri (2-gram ve 3- gram, bigram/trigram) saptamaya yönelik incelemeleri içermektedir.

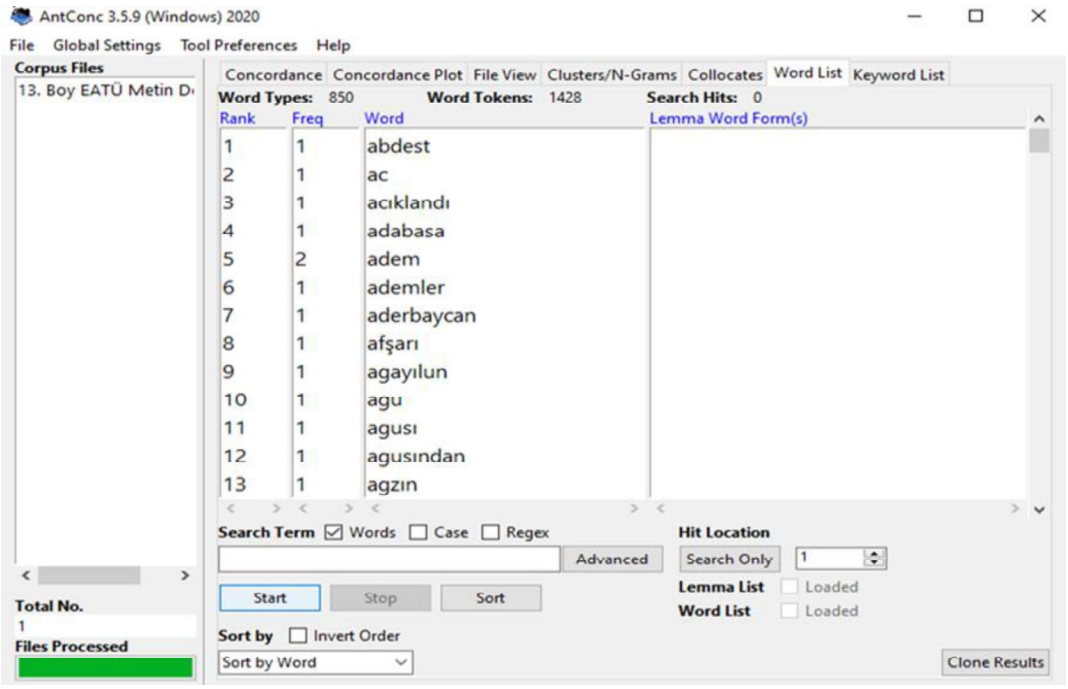
Buna göre, Python programlama dilinde NLTK (Natural Language Tool Kit-Doğal Dil Araç Kutusu) Kütüphanesi kullanılarak metin ön işlemeden geçirilmiş, metindeki noktalama işaretleri, sayılar ve Türkçe sık geçen kelimeler (stopwords) ayıklanmış, ardından eş dizimli kelimelerin saptanması amacıyla metinde ard arda geçen bütün 2-gram (bigram) ve 3-gram (trigram) eş dizimli ögeler çıkarılarak metinde kaç defa geçtikleriyle birlikte bir 'Virgül Ayrımlı Dosya'ya (Comma Separated File, CSV) yazılmıştır. Laurence Anthony tarafından hazırlanan AntConc 3.5.9 (Windows) 2020 adlı yazılım ile de hikâye yeniden bilgisayar ortamına geçirilmiş, metin dosyaları tek tek oluşturulmuş ve

[3] Adı geçen çalışma Turkish Studies - Language and Literature Dergisi'nde yayımlanmıştır.

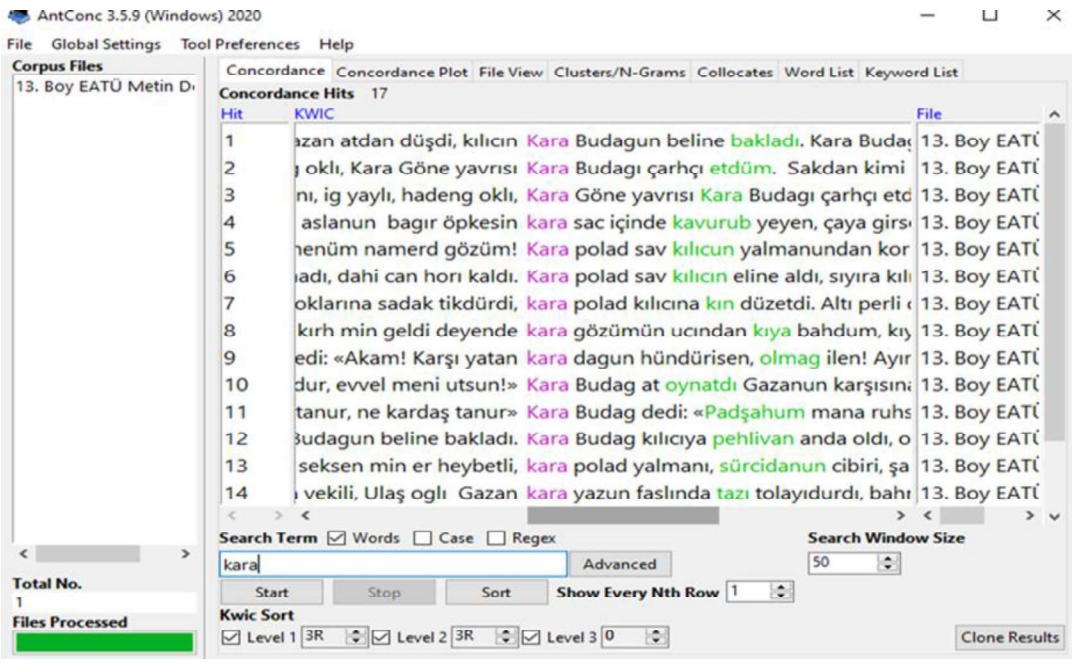
incelenmiştir. Bu uygulamada ikili ve üçlü eş dizimli sözler bilgisayar tarafından tespit edilmiş; frekansları ve cümle içindeki dizilimleri, günümüz Türkçesindeki ve Eski Anadolu Türkçesindeki kullanımları üzerinden karşılaştırılmış; eş dizimlilikler Log-likehood, MI değeri, T-skoru, Dice coefficient değeri gibi farklı istatistik analizleri ile incelenmiştir. AntConc 3.5.9 yazılımı, eş dizimliliklerin nicel olarak çıkarımında MI değeri ve T-skoru olmak üzere iki tür hesaplama yöntemi sunmaktadır. İstatistiksel önem kriteri esas alarak kullanılan yöntemde, temelde bir kelimenin toplam kullanımından ne kadarının belirlenen kelimelerle gerçekleştiğini esas alan algoritmalar uygulanmıştır. 3-gram yapıdaki eş dizimlerde eş dizim kümeleri ve dizileri tek tek saptanmıştır. Böylelikle herhangi bir metinde eş dizimli kelimeler bilgisayara otomatik olarak buldurularak eş dizimli yapıların daha pratik bir yolla tespiti sağlanmıştır. Bu noktada istatistiksel yöntemler çok sayıda veriyi doğru değerlendirebilme ve zamandan tasarruf etme yönünden ayrı bir önem kazanmıştır.

Çalışmada araştırmacı tarafından dil bilimsel tanımla saptanan eş dizimli söz varlığı bilgisayarca bulunan eş dizimli söz varlığı ile karşılaştırılmış ve istatistiksel değerler çıkarılmıştır. Buradan doğan rakamsal sonuçları sergilemek için de grafikler oluşturulmuştur. İncelemeye konu olan Dede Korkut Hikâyeleri'nden 13. Boy'un Eski Anadolu Türkçesi ile transkripsiyonlu metni, düz metin dosyası haline getirilmiş ve aşağıdaki şekilde görüldüğü üzere yazılıma yüklenim sırasında ilgili programdaki Türkçe kodlama (Turkish iso-8859-9) tarafından doğru anlaşılabilmesi için transkripsiyon alfabesindeki özel işaretli harflerin (ġ, ḫ, ḫ, ṣ, é, ż, ḳ) noktasız yazımı sağlanmıştır. Aksi takdirde program özel işaretli harfleri (ġ, ḫ, ḫ, ṣ, é, ż, ḳ) "?" olarak okumakta ve hata yapmaktadır.

Bu yazılımdaki word list (kelime listesi) sekmesinden bütün kelime biçimlerini sıklıklarına göre ya da alfabetik olarak listelemek mümkündür. Concordance (bağlamlı dizin/uyumluluk) sekmesinden de kelimeleri bağlam içinde 2 ya da 3 sağa yönelimli, 2 ya da 3 sola yönelimli aralıklar belirleyerek incelemek mümkün olmuştur. Çalışmamızda öncelikle kelimeler alfabetik sıraya göre dizilmiş, daha sonra tek tek bağlam içinde incelenmiştir. Kelimeler bağlam içinde bir kelime birlikteliği oluşturuyorlarsa, bu birliktelikler sıklık temelli yaklaşımda kullanılan istatistiksel ölçütlere göre incelenmiştir.



Şekil:1 AntConc 3.5.9 (Windows) 2020 Yazılımında Alfabetik Sıralı Kelime Listesi Ekran Görüntüsü

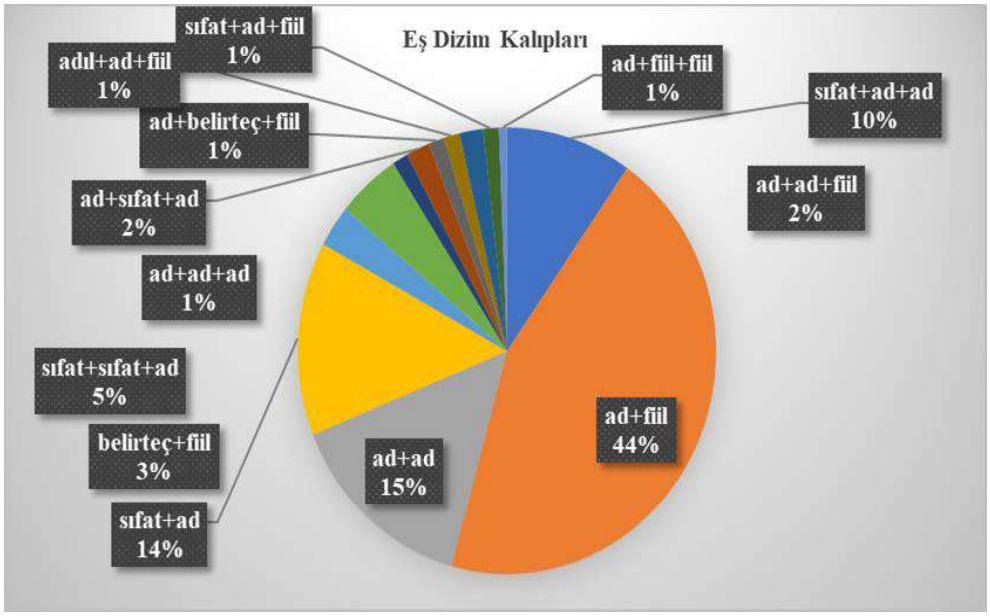


Şekil 2: AntConc 3.5.9 (Windows) 2020 Yazılımında Bağlamlı Dizin/Uyumluluk Analizi Ekran Görüntüsü

Çalışmamızda Metin Ekici'nin "Dede Korkut Kitabı Türkistan/Türkmen Sahra Nüshası Soylamalar ve 13. Boy Salur Kazan'ın Yedi Başlı Ejderhayı Öldürmesi" adlı kitabında yer alan "Salur Kazan'ın Yedi Başlı Ejderhayı Öldürmesi" adlı 13. Boyun orijinal metninden yola çıkılarak hikâyedeki eş dizimli yapılar

bilgisayarlı dil bilim yöntemleri ile tespit edilip incelenmiştir. Metnin günümüz Türkçesine aktarımı üzerinde inceleme yapmaktansa, Eski Anadolu Türkçesinin transkripsiyonlu aktarımı üzerinde eş dizimli yapıların tespit ve tasnif edilmesi uygun görülmüştür. Araştırmanın materyali Dede Korkut Hikâyeleri'nin sonuncusu olan "Salur Kazan'ın Yedi Başlı Ejderhayı Öldürmesi"dir. Demir (2019) tarafından yazılan Dede Korkut Destanı, Ekici (2019) tarafından hazırlanmış Dede Korkut Kitabı Türkistan/Türkmen Sahra Nüshası Soylamalar ve 13. Boy Salur Kazan'ın Yedi Başlı Ejderhayı Öldürmesi, Azmun (2019) tarafından kaleme alınmış Dede Korkut'un Üçüncü El Yazması: Türkmen Sahra ve Shahgoli ve diğerleri (2019) tarafından yazılmış Dede Korkut Kitabı'nın Günbed Yazması: İnceleme, Metin, Dizin ve Tıpkıbasım adlı eserler 13. Boy ile ilgili çalışılmış son örneklerdir. Çalışmamızda Eski Anadolu Türkçesi ile yazılmış 13. hikâye bilgisayar ortamına metin dosyası olarak aktarılmıştır. Ardından bilgisayar destekli dil bilimin istatistiksel yöntemleri ile eş dizimli söz varlığı açısından incelenmiştir. Hikâyedeki eş dizimler önce araştırmacı tarafından tespit edilmiş daha sonra da otomatik olarak belirlenmesi için bilgisayarlı dil bilim alanından yararlanılmıştır. Böylelikle hem dil bilimsel metotlarla araştırmacı tarafından saptanan bulgular hem de doğal dil işleme yöntemleri ile elde edilen bulgular karşılaştırılarak değerlendirilmiştir.

Dede Korkut'un Günbed nüshası, yalnız soylama ve yeni bir anlatma (boy) barındırması açısından değil dil açısından da bir o kadar önemlidir. Eser üzerinde gerek dil hususları açısından karşılaştırmalı çalışmalar gerekse söz varlığının yeni metotlarla incelenmesine dair birçok araştırma yapılabilir. Bizim bu çalışmadaki amacımız dil bilim alanında son yıllarda giderek popülerlik kazanan eş dizimlilik konusunu yeni keşfedilen 13. boylama üzerinde tespite ve tetkike çalışmaktır.



Tablo 1: Metinde Geçen Eş Dizim Kalıplarının Dağılımı

Çalışmamızda eş dizimlerin tespiti için önce araştırmacının sonra da bilgisayarlı dil bilim yöntemlerinin tespiti sağlanmış ardından metinde tespit edilen her iki eş dizim listesi karşılaştırılmıştır. Bilgisayarın eş dizimleri dil bilgisel olarak doğru tespit edip etmediği araştırmamızda cevabını aradığımız sorulardan birisi idi. Söz gelimi "kara" sözünün eş dizimlerini biz frekans sayılarını hariç tutarak 7 adet bulmuşken AntConc yazılımı 10 adet eş dizim öbeği bulmuştur. Yüzdelik hesaplama yapıldığında ise "kara" kelimesi

için bilgisayarın başarısı %60 iken metnin bütünündeki 163 adet eş dizimli ögenin tespitinde yaptığımız tutarlılığa dair karşılaştırmada ise bilgisayarın başarısı %75.61'dir. Bu bulgular ışığında eş dizimliliğin üye sayısını, yapısını, frekansını ve çeşidini saptamak için araştırmalarda bilgisayarlı dil bilimin istatistiksel yöntemlerinin kullanılmasının araştırmacı için büyük bir fayda ve kolaylık sağlayacağı söylenebilir.

2. Dede Korkut Hikâyeleri'ndeki "Öl-" fiilinin eş dizimlilik ve semantik prozodi yönünden bilgisayarlı dil bilim yöntemleri ile incelenmesi[4]

Bu çalışmada da yine bilgisayarlı dilbilim ve semantik prozodi ilişkisi ele alınarak istatistiksel değerlendirme yapılmıştır. Çalışmada materyal olarak Sadettin Özçelik'in hazırladığı 'Dede Korkut' ve 'Dede Korkut -Günbed Yazması- Kazan Bey Oğuznâmesi' adlı eserlerinden yararlanılmıştır. Bu doğrultuda söz konusu fiillerin hikâyelerdeki eş dizimlilikleri tespit edilmiş, seçilen sözcüklerin semantik haritaları ve anlam auraları incelenmiştir. İncelenen sözcüklerin kendi bağlamları içinde eş dizimsel özellikleri gösterilerek tablolar halinde sunulmuş; farklı tekniklerle hazırlanan tablolarda sözcüklerin semantik auraları verilmiştir.

Bilişim teknolojileri alanındaki gelişmelere bağlı olarak dillerin işlenmesi ve bilgisayar yazılımlarıyla incelenmesi kolaylaşmış; ses özelliklerinden yapı özelliklerine, söz varlıklarından söz dizimlerine kadar bütün dil birliklerinin istatistiksel veya kural tabanlı bilgisayarlı dil bilim yöntemleriyle çözümlenmesi olanaklı hale gelmiştir. Günümüzde web ortamından da veri seti oluşturmayı mümkün kılan uygulamalar ve yazılım teknolojileri sayesinde dil bilimciler muazzam bir kaynak ile karşı karşıyadır. Çalışma bu uygulama ve yazılımlardan yararlanılarak yapılmıştır.

Araştırmamızın sonucunda metinde 2-gram, 3-gram, 4-gram, 5-gram, 6-gram ve 7-gram yapılar tespit edilmiştir. Dizgide karakterlerin bitişik olmaları gerekmediği, ancak peş peşe sırada olmaları gerektiği için (Sankur, 2005, s. 72) çalışmamızda belirli bir metin örneğindeki N adet karakterden oluşan dizgi veya N adet ögenin bitişik dizisi karşılığındaki N-gramlar yan yana gelen sözcük öbekleri olarak değerlendirilmiştir. Çalışmada ikili, üçlü, dörtlü, beşli, altılı ve yedili eş dizimli gruplar kronolojik olarak sıralandıktan sonra üçten fazla sayıdaki eş dizim öbeklerinde eş dizimli dizi ve eş dizimli küme tespiti yapılmıştır. Buna göre eş dizimsel dizi ve eş dizimsel küme arasındaki temel fark sözcüklerin birleşim özellikleri ile ilgilidir. Eş dizimsel dizilerde birleşimi oluşturan sözcüklerin birbirinden bağımsız olarak birleştiği gözlemlenirken; eş dizimsel kümeyi oluşturan sözcüklerin birbirinden bağımsız olarak birleşemediği gözlemlenmiştir.

Bilgisayarın eş dizimli yapıları dil bilgisel olarak doğru tespit edip etmediği araştırmamızda cevabını aradığımız sorulardan birisi idi. Bu itibarla çalışmamızda eş dizimlerin tespiti için önce araştırmacının sonra da bilgisayarlı dil bilim yöntemlerinin belirlediği eş dizimler kaydedilmiş; ardından metinde tespit edilen her iki eş dizim listeleri karşılaştırılmıştır. Buna göre biz, "öl-" fiilinin eş dizimlerini ve semantik prozodilerini frekans sayılarını hariç tutarak 173 adet bulmuşken AntConc yazılımı da 173 adet eş dizim öbeği bulmuştur. "Öl-" fiiline ilave edilen eklerle genişletilen sözcük, yazılıma her seferinde yeniden arattırılarak istenilen sonuca ulaşılmıştır. Bu bulgular ışığında eş dizimliliğin üye sayısını, yapısını, frekansını ve çeşidini saptamada bilgisayarlı dil bilimin istatistiksel yöntemlerinin kullanılmasının araştırmacı için büyük bir fayda ve kolaylık sağladığı ve sağlayacağı bir tespit ve kanaat olarak söylenebilir.

4 Uluslararası Türkçe Edebiyat Kültür Eğitim Dergisi (TEKE)'nde yayımlanmıştır.

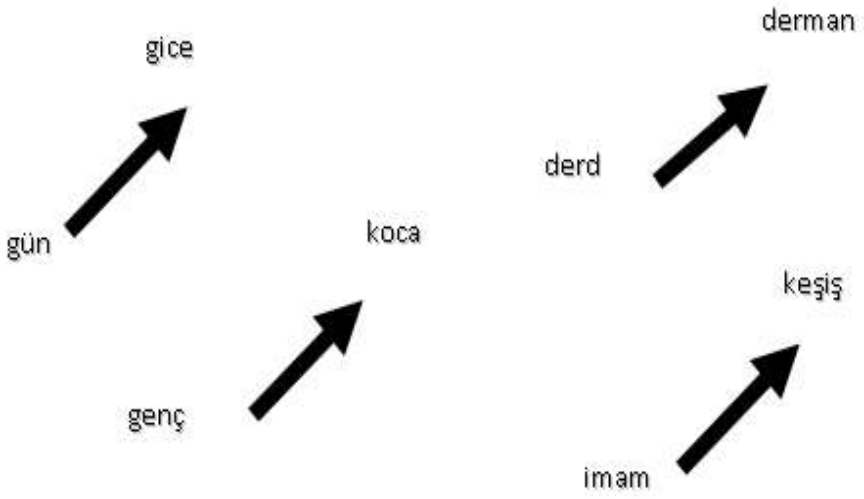
Python programlama dilinde, GraphViz Kütüphanesi ile oluşturulan semantik haritası üzerinde anlam aurası gösterilmiştir.

Şekil 3: "Öl-" Fiilinin Anlam Ağacı

3. Bugün artık, bilgisayarlar aracılığı ile büyük derlemler oluşturulup veriler elde edilebilmekte ve işlenebilmektedir (Aksan ve Yaldır 2011:377). Günümüzde web ortamındaki çeşitli uygulamalar ve yazılım teknolojileri aracılığı ile dilbilimciler zengin bir araştırma sahasına sahiptir. Çalışmamız bu uygulama ve yazılımlardan yararlanılarak yapılmıştır.

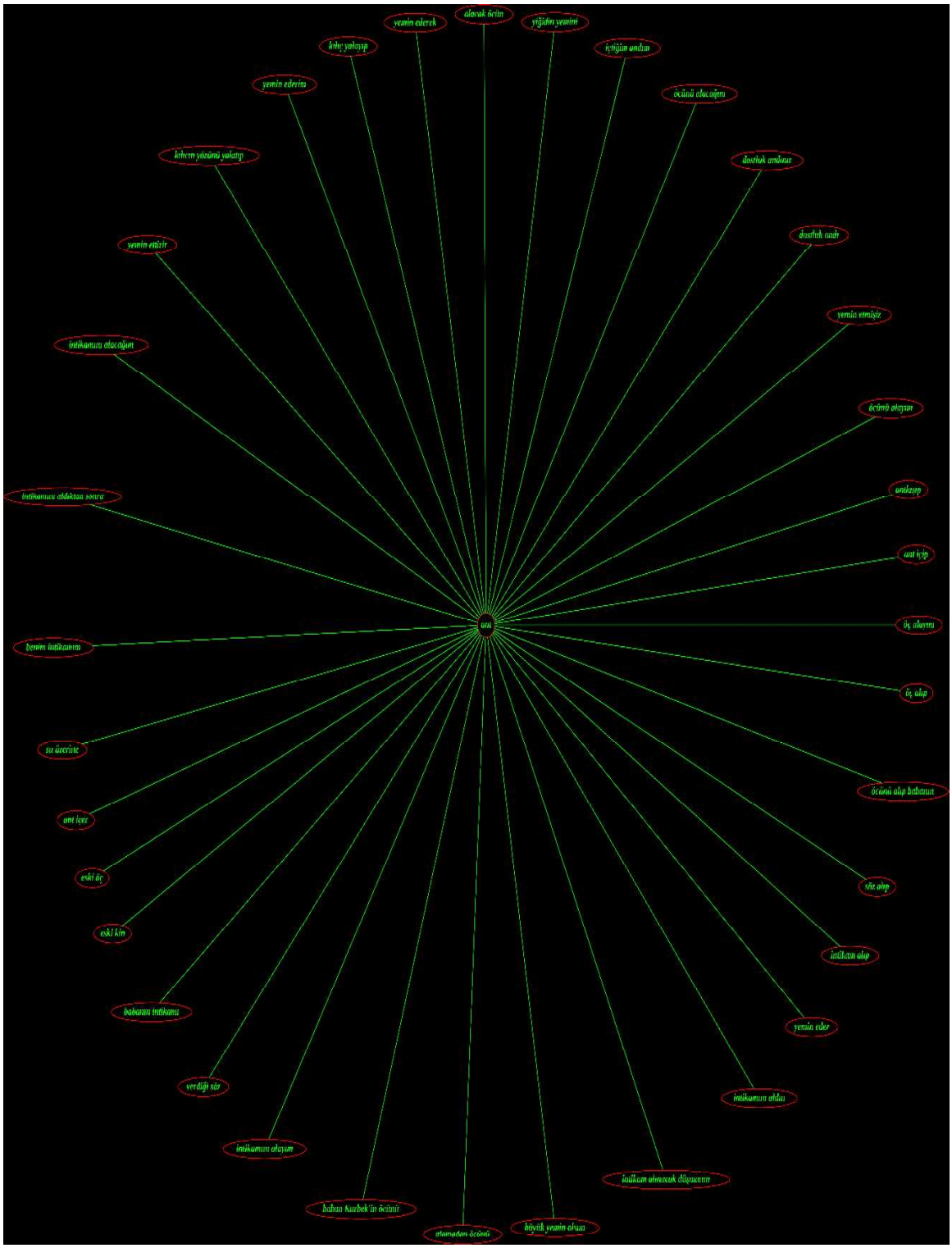
Çalışmada kullandığımız doğal dil işleme yöntemlerinden biri de vektörizasyondur. Bir dildeki kelimeler ana parçalar olarak alındığında bu kelimelerin bilgisayar tarafından matematiksel işlemeye daha uygun, görece daha basit ve tercihen daha az yer kaplayan bir şekilde temsil edilmesi DDİ'de sık kullanılan bir yöntemdir. Buna göre, önce bir vektör boyutu seçilir. Bu boyut temsil ve ayırt etme yeteneği olacak kadar

büyük, ancak gereksiz/fazladan yer kaplamayacak kadar da küçük olmalıdır. Bu boyut seçildikten sonra her kelimeye o boyutta bir vektör atanır. Vektörün elemanları başlangıçta rastgele doldurulur. Ancak metin üzerinde işlendikçe değerleri değiştirilen vektörler atandıkları kelimeleri daha iyi temsil eder hâle gelirler. Mesela bazı kelimeler anlamsal olarak yakın oldukları kelimelerle yakın vektörlere sahip olurlar. Bu anlamsal yakınlık eş dizimleri oluşturan kelimelerin tespit edilmesinde önem taşır. Bu vektörel temsil, yönlendirme yapan çizgilerle temsil edilir ve aşağıdaki şekilde görüldüğü gibi anlamsal bağlantıları olan kelimelerin (kosinüs uzaklığı ölçütüne göre) lineer yapı içinde olan vektörlerle gösterilmesini sağlar.



Şekil 4: Metinde anlamsal olarak karşıtlık bildiren kelimeler

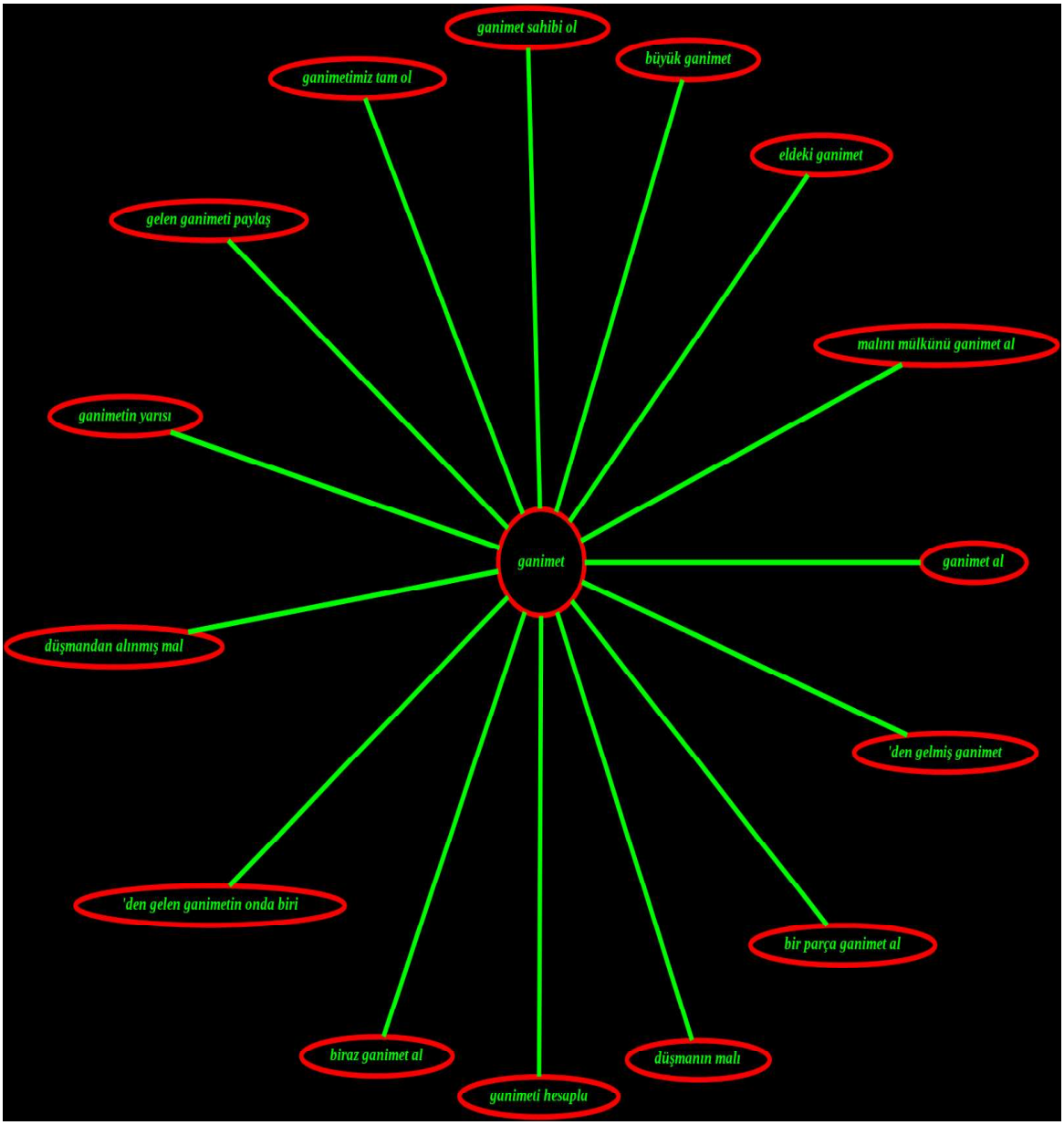
4. Seyitbek Destanı'ndan Ant ve Ganimet Motifleri



Şekil 5: Ant Motifinin Semantik Gösterimi

Adres
RumeliDE Dil ve Edebiyat Araştırmaları Dergisi
Osmanağa Mahallesi, Mürver Çiçeği Sokak, No:14/8
Kadıköy - İSTANBUL / TÜRKİYE 34714
e-posta: editor@rumelide.com
tel: +90 505 7958124, +90 216 773 0 616

Address
RumeliDE Journal of Language and Literature Studies
Osmanağa Mahallesi, Mürver Çiçeği Sokak, No:14/8
Kadıköy - ISTANBUL / TURKEY 34714
e-mail: editor@rumelide.com,
phone: +90 505 7958124, +90 216 773 0 616



Şekil 6: Ganimet Kavramının Anlamsal Grafiği

5. "A Virtual Vocabulary Teacher for Language Distinction Training of Immigration Officers"[5]

Göçmenlerin ve seyahat edenlerin havaalanlarında yaşadığı çeviri sorunlarına çözüm olabilecek diller arası özellikle de Türkçe ve diğer Dünya dilleri arasında anlık çeviri yapabilecek bir arayüz ve yapay zekâ yöntemi ile geliştirdiğimiz bir çözüm önerisi idi. Bu konuda da yine örnek bir model üzerinde çalışmaktayız.

5 Research Leap Dergisi'nde yayımlanmıştır.

Sonuç ve Değerlendirme

Bilgisayarın otomatik çıkardığı çoklu yapıların (n-gramlar) %96'sının anlamlı eş dizimler olduğu görülmekte, bu da metinden eş dizimlilerin elle çıkarılmasına nazaran büyük bir hız ve kolaylık sağlamaktadır. Metnin bütünündeki 163 adet eş dizimli ögenin tespitinde yaptığımız tutarlılığa dair karşılaştırmada ise bilgisayarın başarısı %75.61'dir. Bu bulgular ışığında eş dizimliliğin üye sayısını, yapısını, frekansını ve çeşidini saptamak için bilgisayarlı dil bilim yöntemlerinin araştırmalarda kullanılması büyük fayda ve kolaylık sağlayacaktır. El ile etiketlenmiş olan bu eş dizimli sözler, ilerideki çalışmalarda kullanılacak yapay zeka yöntemleriyle:

a) Çoklu yapının/n-gramın eş dizimli olup olmadığının,
b) Eş dizimli ise hangi kategoride değerlendirilmesi gerektiğinin bilgisayar tarafından otomatik olarak bulunması için kıymetli bir veri kümesi özelliği taşımaktadır.

Metindeki çoklu yapılar/n-gramlar bu şekilde çıkartılarak insan değerlendiriciye sunulacak, çoğu doğru çıkacak olan bilgisayar önerilerinde az sayıda düzeltme yapılarak metin için hızlı bir eş dizim sözlüğü çıkarılabilecek olup bu bilgisayarlı sistem özellikle çok sayıda metnin incelenerek eş dizim sözlüklerinin çıkartılması yolunda büyük fayda sağlayacaktır.

İncelememizde biçim birimler dizi ve küme olarak sıklık temelinde otomatik yollarla belirlenmiş ve dil bilgisel örüntüler AntConc yazılımındaki uyumluluk satırlarını okuyarak çıkartılmıştır. Türkçe verileri işleyebilecek skip-gram ve congram gibi yazılımların bulunmaması ya da Türkçenin tarihsel derleminde Eski Anadolu Türkçesi döneminin metinlerine halihazırda yer verilmemiş olması veri ayıklama hususunda el emeğiyle çalışmayı gerektirmiştir. Aynı eş dizimli kelime köküne gelen farklı eklerin oluşturduğu birliktelikleri her seferinde bilgisayarın yeni bir eş dizimlilik olarak işaretlemesi de başlı başına bir sınırlılıktır. Bu durum kökleme (stemming) denilen bilgisayarlı dil bilim yönteminin yeni yazılımlarının Türkçeye özel hazırlanması ile çözümlenebilecektir. Eş dizimin kelime birimleri arasında mı yoksa kelime biçimleri arasında mı olduğu alanyazınında tartışmalıdır. Ejdehanuñ derisin soydurdı ile ejdehanuñ derisin soygil biçimleri aynı eş dizimlilik özelliği mi taşımaktadır yoksa her bir fiil biçimi ayrı eş dizimlilik özelliği mi taşımaktadır? Araştırmamızda fiillerin farklı biçimleri genellikle aynı eş dizimsel özellikleri taşıdığından yeni bir eş dizim olarak gösterilmemiştir. Konur atlı ve Konur at örneklerinde de ada gelen eklerle oluşan yeni biçimler, eş dizimsel anlamın belirli kavram alanında yer almasından ötürü aynı eş dizim olarak belirtilmiştir. Dönemin dil özelliklerine ait başlıca seslerin (ġ, ḫ, ḫ, ṣ, é, ż, ḳ) yazılımda metin dosyasına aktarılamayıp özel işaretleri olmadan işaretlenmek durumunda kalınması da bir başka sınırlılıktır. Tüm bunlar araştırmacının iş yükünü artırmakla birlikte Türkçenin bilgisayar programlarıyla işlenmesi sırasında nelere ihtiyaç duyulabileceği ya da nelerin sınırlılık oluşturabileceğini ortaya çıkarmıştır.

İncelememizde biçim birimler dizi ve küme olarak sıklık temelinde otomatik yollarla belirlenmiş ve dil bilgisel örüntüler AntConc yazılımındaki uyumluluk satırlarını okuyarak çıkartılmıştır. Türkçe verileri işleyebilecek skip-gram ve congram gibi yazılımların bulunmaması ya da Türkçenin tarihsel derleminde Eski Anadolu Türkçesi metinlerine hâlihazırda yer verilmemiş olması veri ayıklama hususunda el emeğiyle çalışmayı gerektirmiştir. Aynı eş dizimli sözcük köküne gelen farklı eklerin oluşturduğu birliktelikleri her seferinde bilgisayarın yeni bir eş dizimlilik olarak işaretlemesi de başlı başına bir sınırlılıktır. Bu durum kökleme (stemming) denilen bilgisayarlı dil bilim yönteminin yeni yazılımlarının Eski Anadolu Türkçesine özel hazırlanması ile çözümlenebilecektir.

Eş dizimlilik kavramı, ana dili ve yabancı dil sözlüklerinin hazırlanmasında, deyimbilim araştırmalarında, yabancı dil öğretiminde, bilgisayarlı dil bilim ile derlem dil bilim çalışmalarında ve sözcük sıklığı araştırmalarında büyük bir yekûn tutmaktadır. Kişinin söz varlığında %40-%70 oranında yer kaplayan eş dizimlilik, alanyazınında da mühim bir yere sahiptir. Türkçede deyim ve birleşik sözcüklerin içinde değerlendirilerek göz ardı edilen eş dizimlerin sayısı gerçekte deyim ve birleşik sözcüklerden daha fazladır. Günümüzde hâlâ deyimlerle, birleşik sözcüklerle ve birliktelik kullanımlarıyla karıştırılan eş dizimli sözcüklerin tespiti, tasnifi ve incelenmesi bu birlikteliklerdeki kavram karışıklığını gidermek için faydalı olacaktır. Bir yandan da Türkçenin tarihî dönemlerine ait eş dizim sözlüklerinin hazırlanması, özel derlem çalışmalarının ve dönemler arası karşılaştırmalı incelemelerin yapılması metinlerin daha iyi anlamlandırılması açısından fayda sağlayacaktır.

Eş dizimlilik çalışmaları, özellikle Türkçenin tarihî dönem eserlerinin daha iyi anlaşılmasına katkı sağlayacak; eş dizimli yapıların tespiti sayesinde tarihî metinlerde satır arası sözcüklerin anlamlandırılması, anlam auralarının keşfi kolaylıkla yapılabilecektir. Gelişen bilgisayar teknolojileri sayesinde yeni yazılımlar ile anlam haritalarının, sözcüklerin vektörel şemalarının ve anlam auralarının somut gösterimi sağlanacaktır. Metinlerin tamamında sözcüklerin kullanış sırası, anlamsal örüntü ve kurgusal bütünlük ustaca nakşedilmiştir. Eş dizimlilik ve semantik prozodi üzerine yaptığımız bu araştırmada metinlerdeki örüntüler arasında anlamlı bir bağ olduğu gözlemlenmiş, sözcük birliklerinin rastlantıyla yan yana gelmediği saptanmıştır.

Eş dizimliliği yalnızca Batılı dil bilimcilerin araştırma konusu edinmeyip Türk dil bilimcilerin de son 20-25 yılda bu konuya ağırlık vermesi alanyazınında yeni yapılacak çalışmalara ışık tutacaktır. Türk dilinin tarihî metinlerindeki zengin sözcük hazinesinin ışığında eş dizimlilik ve semantik prozodi alanında gerçekleşecek yeni araştırmalar literatüre mühim bir katkı sağlayacaktır.

Kaynakça

Adalı, E. Türkçe Doğal Dil İşleme. Akçağ Yayınları, 2021.

Adalar, D. Anadili Olarak Arapça ve Türkçenin Öğretiminde Kullanılan Metinlerin Karşılaştırılması: Bir Eş Dizimsel Çözümleme Örneği, Ankara Üniversitesi Sosyal Bilimler Enstitüsü, Doğu Dilleri ve Edebiyatları Anabilim Dalı, Yüksek Lisans Tezi, 2004.

Aksan, Y. ve Y. Yaldır. Türkçe Sözvarlığının Nicel Betimlemesi, ed. Ç. Sağın Şimşek ve Ç. Hatipoğlu. 24. Ulusal Dilbilim Kurultayı Bildiri Kitabı. ODTÜ Basım İşliği, 2011.

Alsarray, M. Türkçe Ulusal Derlemi'nde Yüksek Sıklıkta Kullanılan Adların Eş Dizimliliği, Yıldırım Beyazıt Üniversitesi Sosyal Bilimler Enstitüsü Dilbilim Anabilim Dalı, Yüksek Lisans Tezi, 2015.

Ayabakan, M. Türkçe Sözlükte Eş Dizimli Ögelerin Sunumu ve Görünümleri, Ankara Üniversitesi Sosyal Bilimler Enstitüsü Dilbilim Anabilim Dalı, Yüksek Lisans Tezi, 2015.

Ayverdi, İ. Misalli Büyük Türkçe Sözlük. Kubbealtı Neşriyatı, 2020.

Dönmez, İ. ve E. Adalı. "Türkçe Tümce Çözümlemede Vektör Yaklaşımı." Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi, no.15, 2015, ss. 1-11.

Girgin, M. Süheyl ü Nev-Bahâr'daki Fiillerin Eş Dizim Sözlüğü, Kütahya Dumlupınar Üniversitesi Lisansüstü Eğitim Enstitüsü Türk Dili ve Edebiyatı Anabilim Dalı, Yüksek Lisans Tezi, 2022.

Gökdayı, H. Türkçede Öbekler. Kriter Yayınları, 2020.

Korkmaz, Z. Dede Korkut Hikayelerinde Dil-Üslup Bağlantısı. TDAY Belleten, no.46, 1998, ss.101–112.

Oğuzlar, A. Temel Metin Madenciliği. Dora Yayınları, 2011.

Önder, Ş.G. Arap Dilinde Eş Dizim, Marmara Üniversitesi Sosyal Bilimler Enstitüsü Temel İslam Bilimleri Anabilim Dalı, Yüksek Lisans Tezi, 2014.

Özçelik, S. Dede Korkut Üzerine Yeni Notlar, Gazi Kitabevi, 2006.

Özçelik, S. Dede Korkut -Dresden Nüshası- Giriş-Notlar Türk Dil Kurumu Yayınları, 2016.

Özçelik, S. Dede Korkut -Dresden Nüshası- Metin-Dizin Türk Dil Kurumu Yayınları, 2016.

Özçelik, S. Dede Korkut -Günbed Yazması- Kazan Bey Oğuznamesi, Ankara: Türk Dil Kurumu Yayınları, 2021.

Özkan, B. Türkiye Türkçesinde Belirteçlerle Fiillerin Birlikte Kullanımı ve Eş Dizimliliği-Derlem Tabanlı Bir Uygulama. Türk Dil Kurumu Yayınları, 2011.

Pennington, J. ve diğerleri. "GloVe: Global Vectors for Word Representation." Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014, ss. 1532-1543.

Sankur, B. Otomatik Dil Sınıflandırma ve Türkçe Bilişim Dili. Bilgisayar Destekli Dil Bilimi Çalıştayı Bildirileri. Türk Dil Kurumu Yayınları, 2006.

Şeker, S. E. "Doğal Dil İşleme (Natural Language Processing)." YBS Ansiklopedi, no.4, 2015, ss. 14-22.