



KORKUT ATA TÜRKİYAT ARAŞTIRMALARI DERGİSİ

Uluslararası Türk Dili ve Edebiyatı Araştırmaları Dergisi
The Journal of International Turkish Language & Literature Research

Sayı/Issue 13 (Aralık/December 2023), s. 919-956.
Geliş Tarihi-Received: 19.06.2023
Kabul Tarihi-Accepted: 29.12.2023
Araştırma Makalesi-Research Article
ISSN: 2687-5675
DOI: 10.51531/korkutataturkiyat.1317017

Türkçede Sözcük Demetleri: Kuram ve Uygulama

Lexical Bundles in Turkish: Theory and Practice

Nuh DOĞAN*
Ahmet DAĞLI**
Betül GÜVENDİ***

Öz

Bir dilin söz varlığında ayrı ve tek tek sözcüklerden ziyade birden fazla sözcükten oluşmuş dilde sıklıkla tekrarlanan, birbirine nispeten az ya da çok donuk çeşitli *kalıp dil birimleri* vardır. Kalıp dil birimleri dillerde yapı, anlam, işlev ve diğer dil bilimsel farklılıklarıyla çeşitlilik gösterir. *Sözcük demetleri* de kalıp dil birimlerinin bir türüdür. Sözcük demetlerinin belirli dil türüne, metin türüne ve uzmanlık alanına özgü metin oluşturmada etkili söylemsel işlevleri vardır. Türkçede sözcük demetlerinin çoğunlukla 3 sözcüklü olduğu ileri sürülmüş ve daha çok 3 sözcüklü sözcük demetleri belirlenmiştir. Bu çalışmada Türkçe sözcük demetlerinin 3 ve daha fazla sözcüklü oluşabileceği gibi 2 sözcüklü de olabileceği, özellikle 2 ve 3 sözcüklü sözcük demetlerinin önemli bir kısmının sıklığa ve donukluğa bağlı olarak belirli söylem işlevleriyle sözlükselleştiği ve deyimselleştiği ileri sürülmüştür. Sözcük demetlerinin bir kısmı dil bilimsel tamamlanmamış metin işlevi olan çok-sözcüklü birimleri teşkil ederken bir kısmı sözlükselleşmiş ve deyimselleşmiş kalıp dil birimlerinden oluşabilmektedir. Bu çalışmada 2 birimli sözcük demetlerinin belirlenmesinde Türkçe Ulusal Derlemi'nden (TUD) hareket edilmiş, derlem-tabanlı hibrit bir yaklaşım benimsenmiştir. Dil türü, metin türü ve uzmanlık alanından bağımsız olarak Türkçenin genel konuşma ve yazı dilinde pedagojik, leksikolojik ve psiko-dilbilimsel açıdan önemli iki-birimli sözcük demetleri belirlenmiştir.

Anahtar Sözcükler: Sözcük demetleri, eşdizimlilik, kalıpsallık, sözlükselleşme, derlem.

Abstract

In the vocabulary of a language, there are various formulaic sequences that are more or less fixed compared to each other, and are composed of more than one word, rather than isolated and individual words. Formulaic sequences vary across languages with their structure, meaning, function and other linguistic features. Lexical bundles are also a type of formulaic sequences. Lexical bundles have discursive functions that are effective in creating texts specific to a language or register, text type and domain. It has been suggested that lexical bundles in Turkish are mostly 3-words, and mostly 3-word lexical bundles have been identified. In this study, it has been suggested that Turkish lexical bundles can also be 2-word, contrary to what

* Doç. Dr., Ondokuz Mayıs Üniversitesi, İnsan ve Toplum Bilimleri Fakültesi, e-posta: nuhdogan55@hotmail.com, ORCID: 0000-0001-8935-8428.

** Doç. Dr., Ondokuz Mayıs Üniversitesi, İnsan ve Toplum Bilimleri Fakültesi, e-posta: ahmet.dagli@omu.edu.tr, ORCID: 0000-0003-3166-2544.

*** Doktora Öğrencisi, Ondokuz Mayıs Üniversitesi, İnsan ve Toplum Bilimleri Fakültesi, e-posta: betulguvendi2022@gmail.com, ORCID: 0000-0003-1343-3959.

is believed, and that a significant part of 2- and 3-word lexical bundles in particular are lexicalized and idiomatic with certain discourse functions depending on frequency and opacity. Some of the word bundles may consist of incomplete multi-word units which are textual functions, and some may consist of lexicalized and idiomatic formulaic sequences. In this study, Turkish National Corpus (TUD) was used to determine 2-unit lexical bundles and a corpus-based hybrid approach was adopted. Regardless of a language or register, text type and specific domain, pedagogically, lexically and psycholinguistically important 2-unit lexical bundles have been identified in the general spoken and written language of Turkish.

Keywords: lexical bundles, collocation, formulaicity, lexicalization, corpus.

Giriş

Bir dilin söz varlığında ayrı ve tek tek sözcüklerden ziyade birden fazla sözcükten oluşmuş dilde sıklıkla tekrarlanan, birbirine nispeten az ya da çok donuk çeşitli *kalıp dil birimleri* (formulaic sequences) vardır. Kalıp dil birimleri, Sinclair' in (1991) *açık seçim ilkesine* (open choice principle) koşut olarak tasarladığı ve *deyim ilkesi* (idiom principle) adını verdiği *kalıp dil* (formulaic language) eğiliminin ürünüdür¹. Sinclair' in deyim ilkesine göre konuşurlar, çözümlenebilir (analitical) görünseler de bir ve bütün olarak kullanılan, yarı-önceden üretilmiş, *sözcük-dilbilgisinin* (lexis-grammar) ortak seçimiyle oluşan birçok kalıp öbek ifadeye sahiptir (1991, s. 110). Ne dil bilgisinin ne de sözlüğün ürünü olan bu kalıp ifadeler, açık seçim ilkesi uyarınca söz dizimsel kurallara göre üretilen dil bilimsel öbeklere kıyasla dilde, özellikle metin üretiminde çok daha işlevseldir. Kalıp dil birimi alanyazında "dil grameri aracılığıyla analiz etmek ve türetmekten ziyade belleğe bütün olarak kodlanan ve kullanım sırasında bellekten yine bütün olarak getirilen; başka bir deyişle önceden üretilmiş veya öyle görünen sözcüklerin veya diğer unsurların bitişik veya ayrı dizisi" şeklinde tanımlanır (Wray 2002, s. 9). Kalıp dil birimlerinin söylemin akıcı, etkili, yetkin ve doğal olmasını sağlayan ve konuşurun dil bilimsel işlem yükünü azaltan önemli işlevleri vardır (Pawley ve Syder 1983, s. 191; Wood 2015, s. 12-16) ve bunlar konuşurların doğal dilde ve söylemde sıklıkla başvurdukları unsurlardır. Kalıp dil birimleri dillerde yapı, anlam, işlev ve diğer dil bilimsel farklılıklarıyla çeşitlilik gösterir. Kimi kalıp dil birimlerinin deyimisel olduğundan sezgisel olarak hemen fark edilebilecek bilişsel bir önemi vardır (ör. *eli sıkı, alın teri, at gözlüğüyle bakmak*). Kimi, saydam ya da geçirimsiz olmasına karşın bilişsel bir önem arz eder (ör. *ders almak, kaygı vermek, sabah ola hayrola* vs.). Metin ve söylem işlevleri olan kimi kalıp dil birimlerinin ise (ör. *bildiğim kadarıyla, bununla birlikte, belki de en, bundan böyle, ile ilgili, başta olmak üzere* vs.) bilişsel bir önemi ve deyimliliği olmadığı hâlde *kalıpsal* (formulaic) olduğu ileri sürülür. Kalıp dil birimleri sahip oldukları dil bilimsel nüanslarından dolayı alanyazında bazıları aynı türü ifade eden 42 farklı terimle karşılanmaktadır (Wray and Perkins, 2000). *Çok-sözcüklü birimler* (multi-word units), *sözcük birlikleri* (lexical phrases), *sözcük çerçeveleri* (lexical frames), *kalıp sözler* (formulas), *tekrarlanan sözcük kümeleri* (recurrent clusters), *klişeler* (routines), *donuk sözler* (fixed expressions), *önüretimli kalıplar* (prefabricated patterns), *ngramlar* (n-grams), *sözcük zinciri* (chains), *sözcük birlikleri* (clusters), *eşdizimler* (collocations), *birleşikler* (compounds), *deyimler* (idioms), *fosilleşmiş biçimler* (fossilized forms) vs. kalıp dil birimleri için kullanılan terimlerden yalnızca birkaçıdır. Kalıp dil birimlerinin belirlenmesinde araştırmacılar amaçlarının çeşitliliğine bağlı olarak farklı ölçüt ve yöntemler kullanır. Kalıp dil birimleri *sıklık* (frequency); *donukluk* (fixedness); *deyimleşme* (idiomaticity); *sözcük sayısı* (length of sequence); söz dizimsel, anlamsal ve pragmatik *tamamlanmışlık* (completeness); *sezgisel tanınurluk* (intuitive recognition) olmak üzere altı tipik özelliğe göre birbirinden farklılaşır (Conrad & Biber, 2005). Kalıp dil birimleri bu özelliklerden birine ya da bir bölümüne öncelik vererek birbirinden ayırt edilir. Söz konusu ölçütlerin tümünün ya da bir kısmının

¹ Bu iki ilke ya da eğilim dildeki *üretimlilik* (productivity) ve *kalıpsallık* (formulaicity) olgusuyla ilgilidir (Perkins, 1999).

sonucu olarak ortaya çıkan kalıp dil birimlerinin kapsayıcı ve başat özelliğinin *kalıpsalık* (formulaicity) ya da *önüretimlilik* (prefabricated) olduğu söylenebilir.

Sözcük demetleri (lexical bundles) de kalıp dil birimlerinin bir türüdür. Sözcük demeti, basitçe “bir dil türünde (register) çok sık tekrar eden *sözcük dizileri* (sequences)” şeklinde tarif edilir (Biber 2006, s. 134; Biber vd. 2004, s. 376). Sözcük demetlerinin metin oluşturmada önemli işlevleri vardır. Biber vd. (1999) sözcük demetlerinin bir dil türü olarak özellikle akademik konuşma ve yazmada söylemin önemli yapı blokları olarak sıklıkla ve yaygın şekilde kullanıldığını, dil öğrenenler için de önemli bilgi oluşturduğunu aktarır. Cortes (2004), diğer kalıp dil birimleri gibi sözcük demetlerinin, özellikle belirli dil türlerinde akıcı ve yetkin dil kullanımının önemli bir belirleyicisi olduğunu belirtir. Sözcük demetleri, dil bilimi alanyazında *sözcük birlikleri* (lexical phrases) adıyla sezgisel önem yaklaşımına göre daha önce Nattinger&DeCarrico (1992); derlem-çıkışlı yaklaşımla ise *tekrarlanan sözcük-birleşmeleri* (recurrent word-combinations) terimiyle ilk önce Altenberg (1998) tarafından çalışılmıştır. Biber vd. (1999) tekrarlanan sözcük-birleşmeleri için sözcük demetleri terimini kullanmış, sözcük demetlerini bilgisayar destekli ve derlem-çıkışlı yaklaşımla salt sıklığa göre belirlemiş, bazı yapısal özelliklerini tanımlamıştır. Devam niteliğindeki çalışmalarda Biber vd. (1999)’un yaklaşımı farklı dil türlerine uygulanmış ve sözcük demetlerinin söylem işlevleri belirlenmiştir (Biber, Conrad, Cortes 2004; Cortes 2004; Cortes 2008; Cortes 2013; Conrad & Biber 2005).

Türkçe alanyazında deyim, atasözü ve ikileme dışında sözcük demetleri ve kalıp dil birimlerinin diğer türleri pek önemsenmemiştir². Türkçede sözcük demetlerine ilk defa Aksan&Aksan (2013), Çalışkan (2014b) ve Aksan&Aksan (2015a) tarafından dikkat çekilmiştir. Bu yayınlarda sözcük demetleri için *sözcüksel set* ve *çok-sözcüklü birim* terimi kullanılmıştır. Türkçe sözcük demetlerinin belirlendiği ve derlemden çıkarıldığı bu ilk ve öncü çalışmalarda Biber vd. (1999)’un araştırma geleneğine uygun olarak “derlem-çıkışlı sıklık yaklaşımı” benimsenmiş ve sözcük demetlerinin belirli yapısal tipleri tanımlanmıştır. Bu çalışmalarda Biber vd. (1999)’da İngilizce için belirlenen sözcük demeti uzunluğu ölçütü (en az üç sözcüklü) esas alınmış ve üç-sözcüklü sözcük demetleri incelenmiştir. Türkçe iki-sözcüklü sözcük demetlerinin dikkate alındığı Aksan & Aksan (2015b)’de ise “sıklık kesme noktasının” çok yüksek tutulması olası iki-sözcüklü sözcük demetlerinin belirlenememesine neden olmuştur. İki-sözcüklü dil birimlerin diğer dillerde olduğu gibi Türkçe için de önemli bir veri olmadığı düşünülmüş (Aksan&Aksan 2013; Aksan&Aksan, 2015a; Aksan vd. (2016a); devam niteliğindeki çalışmalarda iki-sözcüklü sözcük demetleri genellikle ya göz ardı edilmiş ve konunun kapsam alanı dışında tutulmuş ya da iki-sözcüklü sözlüksel eşdizimler içinde betimlenmiştir (Aksan, 2016b). Sözcük demetlerinin İngilizce temelinde benimsenen tanımlayıcı ölçüt ve yaklaşımları eklemeli bir dil olan ve zengin biçim bilimsel araçlar kullanılan Türkçe gibi diller için belirleyici ve sınırlayıcı değildir³. Durrant’ın (2016) da ifade ettiği gibi İngilizcede çok-sözcüklü birimlerle sağlanan birçok anlatımın Türkçede tek bir sözcük biçimiyle gerçekleştirilebildiği dikkate alındığında özellikle iki-birimli Türkçe sözcük demetlerinin İngilizce sözcük demetlerinden yapısal açıdan farklılaştığı, Türkçede iki-birimli sözcük demetlerinin İngilizceye nispeten çok daha yaygın ve sıklıkla kullanıldığı ileri sürülebilir. İngilizce sözcük demetlerinin sözcük sayısının en az üç (3) olarak belirlenmesinin nedeni sözcük demetlerini eşdizimlerden ve deyimlerden ayırtmaktır (Altenberg 1998, s.103). Ancak

² Deyim, atasözü ve ikileme dışındaki diğer kalıp dil birimleriyle ilgili az sayıda araştırma yapılmıştır. Bunlardan Özkan (2007), Çalışkan (2014a) ve Doğan’ın (2016) eşdizimlerle, Gökdayı’nın (2011) ise kalıp sözlerle ilgili çalışmalarının dikkat çektiği söylenebilir.

³ Sözcük demetleriyle ilgili bilimsel araştırmaların alanyazında hemen hemen birkaç çalışma dışında İngilizceyle sınırlı olduğu görülmüştür. Bu nedenle sözcük demetlerinin tanımlayıcı özelliklerinin İngilizceye dayalı olması doğal bir sonuçtur.

Türkçede deyim ve eşdizimler üç-sözcükten oluşabildiği⁴ gibi sözcük demetleri de *sınırlı eşdizimlerden* (restricted collocations) farklı olarak iki-sözcüklü gerçekleşebilir⁵.

Devam niteliğindeki çalışmalarda Türkçe sözcük demetlerinin belirlenmesinde ve derlemden çıkarımında benimsenen yaklaşımla birlikte sözcük demetleri için kullanılan çok-sözcüklü birim teriminin içeriği ve kapsamı değişmiştir. İlk ve öncü çalışmalarda benimsenen derlem-çıkışlı sıklık yaklaşımı yerini sıklığın farklı *birliktelik ölçümleri* (associative measures) ve yapısal tiplerle birleştirdiği hibrit yaklaşımlara bırakmıştır. Sözcük demetlerin çıkarımında Aksan&Aksan (2015a)'da sıklığın yanı sıra ilk defa bir birliktelik ölçümü kullanılmıştır. Aksan, Mersinli ve Altunay (2015)'de ise birliktelik ölçümlerine başvurmadan önce dil bilgisel filtreleme ve sınıflandırmanın sayısal sıralı n-birim listesinden çok-sözcüklü birim olmayanları dışlayacağı düşünülmüştür. Mersinli (2015) Türkçe çok-sözcüklü birimlerin derlemden çıkarımında birliktelik ölçümlerinin dil bilgisel filtreleme safhasından önce ya da sonra kullanılabileceğini, kural-tabanlı ve istatistik tekniklerin uygulandığı böyle bir hibrit yaklaşımın zorunlu değilse bile bir gereklilik olduğunu ileri sürer. Aksan vd. (2016)'da bir çok-sözcüklü birim adayının sıklığına biçim-sözdizim özelliklerinin etki ettiği düşünülmüş, çok-sözcüklü birimlerin belirlenmesinde ve çıkarımında Aksan&Aksan (2015a-b)'de belirlenen belirli yapısal ve biçim-söz dizimsel tipler bir filtreleme aracı olarak kullanılmıştır. Biçim-sözdizimsel yapıların sıklığa eşlik etmesinin çok-sözcüklü birimlerin sözlükselleşmesine neden olduğu düşünülmüştür. Çok-sözcüklü birimlerin çıkarımı ve belirlenmesinde değişen metodolojik yaklaşıma paralel olarak bu süreçte çok-sözcüklü birim teriminden anlaşılan şey de değişmiştir. İlk ve öncü çalışmalarda çok-sözcüklü birim teriminin Biber vd. (1999)'un sözcük demeti kavramına eş bir içeriği varken sonraki çalışmalarda terimin kapsam alanı genişlemiştir. Çok-sözcüklü birim terimi artık sözcük demetleriyle birlikte sözlüksel eşdizimleri ve diğer kalıp dil birimlerini içerecek şekilde kullanılmıştır⁶ (Aksan, 2016). Bu tutum, Batı alanyazının aksine eşdizim ve sözcük demeti ayrımının ortadan kalkmasına, neredeyse deyim dışında tüm kalıp dil birimlerin tek bir kategori ya da dil bilimsel olgu etrafında incelenmesine ve içinden çıkılması zor bir terminoloji ve kavram karmaşasına yol açmıştır. Bu süreçte derlem verilerine ve istatistik analizlere dayanan bilgisayar destekli metodolojik uygulamalara öncelik verilmiş, sözcük demetlerini diğer kalıp dil birimlerden ayırt eden kuramsal ayrımlar ve dil bilimsel ölçütler çoğunlukla göz ardı edilmiştir. Bugün artık Türkçe alanyazında sözcük demetlerinin Batı alanyazında, Biber ve arkadaşlarının araştırma geleneğinde anlaşıldığı şekliyle anlaşılmadığı söylenebilir.

Türkçe sözcük demetleri araştırmalarında hibrit yaklaşımlar kullanılsa da Biber vd. (1999)'un etkisiyle sıklığın başat ölçüt olarak benimsenmesi kalıp dil ürünü olmayan, sıklığı yüksek olmasına karşın işlevsel ve psiko-dilbilimsel ya da pedagojik açıdan önemsiz, rastlantısal eşdizimlere ya da sözcük dizilerine değer atfedilmesine yol açmıştır. Derlem-çıkışlı sıklık yaklaşımı dilde sıklığı yüksek çok-sözcüklü birimleri ya da sözcük dizilerini derlemden çıkartmakta oldukça kullanışlı olmasına karşın her zaman kalıp dil ürünü çok-sözcüklü birimleri belirleme imkânı vermez. Wood (2019, s. 35), Biber vd.'nin (1999) derlem

⁴ Ayabakan (2019, s. 65) Türkçede özellikle 3 sözcükten oluşmuş birçok isim eşdizimi tespit eder. Üç sözcüklü isim eşdizimleri için özellikle bk. Ayabakan (2019).

⁵ Durrant (2016) ve Durrant (2013)'te Türkçede kalıpsallığın İngilizce eylemcil ya da cümlemsi sözcük demetlerinin yapısal özelliklerinin etkisinde kalarak sadece sözcüklerle değil sözcük içinde belirli ek ve ek dizileriyle de gerçekleştiği ileri sürülür. Bu olgu Doğan (2019) ve (2021)'de dil bilgisel eşdizimlilik açısından incelenmiş ancak bir söylem işlevi ve sezgisel önemi olmadığından kalıp dilin ürünü olarak değil sıklığa bağlı ancak dil bilgisel bir gelişme ve dil bilgisi kuralı olarak tanımlanmıştır.

⁶ Prof. Dr. S. Yeşim Aksan tarafından yürütülen "Türkçe'de çok sözcüklü birimler: derlem-temelli yöntem ile çıkarımları, sayısal dağılımlarının hesaplanması, yapı ve işlev özelliklerinin saptanması, sözlüğünün oluşturulması" başlıklı, 115K135 numaralı Tübitak projesi. Çevrimiçi: <https://search.trdizin.gov.tr/tr/yayin/detay/615654/>

ve sıklık çıkışlı yaklaşımıyla belirlenen sözcük demetlerinin metin oluşturmada önemli olmakla birlikte kalıpsallık ve önüretimlik niteliklerini taşımadığını, kalıp dil ürünü dil birimleri olmaktan ziyade çok-sözcüklü birlikler “multi-word units” olarak kabul edilebileceğini düşünür. Sözcük demetlerinin diğer kalıp dil birimlerinden ve sözcük dizilerinden ayırt edilmesinde sıklığın yanı sıra başka ölçütler de kullanılabilir. Cortes (2004) sözcük demetlerinin belirlenmesinde sıklığın yanı sıra deyimselleşme ve donukluğun da kullanılabilirliğini belirtir. Ellis vd. (2008) ve Simpson-Vlach & Ellis (2010) işlevsel ve pedagojik açıdan önemli kalıp dil birimlerini belirlemede sıklığın yanında sezgisel önem ya da tanınırlığı bir ölçüt olarak kullanmışlardır. Bu çalışmada dil bilimsel amaçlara yönelik olarak hem sözcük demetleri ve eşdizim⁷ arasında hem de sözcük demetlerinin kendi içinde betimleyici ve tanılayıcı bir ayırım yapılmıştır. Sözcük demetlerinin çıkarımında ve belirlenmesinde Türkçe Ulusal Derlemi’nden (TUD) hareket edilmiş, *derlem-tabanlı* (corpus-based) ancak özgün hibrit bir yaklaşım kullanılmıştır (bk. Bölüm 5). Türkçe sözcük demetlerinin yaygın kabulün aksine iki-sözcüklü de olabileceği, ister iki ister üç-sözcüklü olsun sözcük demetlerinin önemli bir kısmının her zaman belirleyici ve mutlak olmamakla birlikte nispi bir sıklığa ve donukluğa bağlı olarak belirli söylem işlevleriyle sözlükselleştiği (deyimselleştiği)⁸ ileri sürülmüştür. Sözcük demetleri sözlüksel ve işlevsel olmak üzere iki sınıfa ayrılmıştır. İşlevsel sözcük demetlerinin metin oluşturma ve söylem işlevi olan görece sık *sözcük birliklerini* (lexical chunks) teşkil ederken sözlüksel sözcük demetleri sözlükselleşmiş ve kalıpsal ifadelerden oluşabilmektedir. Türkçe Sözlük (2023) sözlükselleşmiş sözcük demetlerinin bir kısmını kayıt altına almasına ve sözlükbirimselleştirmesine karşın çok daha büyük bir bölümünün kayıt altına alınmadığı görülmektedir. Sözcük demetleri farklı ölçüt, ölçüm ve derlem dilbilimi uygulamalarıyla derlemden çıkarılmayı beklemektedir.

1. Sözcük Demetlerinin Ayrıncı Nitelikleri

Kalıp dil birimlerinden genellikle bir kısmı deyimsel (idiomatic); diğer kısmı ise deyimsel olmayan ancak sezgisel olarak önemli iki büyük sınıfı oluşturan çok-sözcüklü birimler anlaşılır. Kalıp dil birimleri üzerine yapılan çalışmalar da çoğunlukla bu kategoriler üzerine yoğunlaşmıştır (Pawley&Syder 1983, Moon 1998, Howart 1996, Nattinger & DeCarrico 1992, Wray&Perkins 2000, Wray 2002, Gökdayı 2011, Çalışkan 2014a, Doğan 2016, Doğan 2023). Sözcük demetleri bu iki sınıfın dışında ayrı bir kalıp dil birimi olarak düşünülmüştür. Sözcük demetleri basitçe “doğal bir söylemde en sık tekrar eden 3 ya da daha fazla sözcük biçimi dizileri (Biber vd. 1999, s. 990)” olarak tanımlanır. Alanyazında sözcük demetlerini diğer kalıp dil birimlerinden ayırt eden bazı temel tipik özelliklerinin olduğu kabul edilir. Bunlardan birincisi (i), sözcük demetleri aşırı derecede yaygın ve sıktır. İkincisi (ii) sözcük demetleri anlamca deyimsel değildir, bir başka deyişle “*bununla birlikte, her ne kadar, her şeyden önce*” gibi sözcük demetlerinde olduğu gibi anlamı bileşenlerinin anlamının toplamından anlaşılabilir ya da bütünü bileşenleri tümlenebilir durumdadır. Üçüncüsü (iii) sözcük demetleri deyimsel olmadığından çoğunlukla bilişsel ya da sezgisel önemi yoktur. Dördüncüsü (iv) ise sözcük demetleri genellikle iyi biçimlenmiş ya da tamamlanmış dil bilgisel yapılar değildir, çoğunlukla iki farklı dil

⁷ Eşdizim araştırmacıların amaçlarına ve metodolojik anlayışına göre oldukça farklı anlaşılmaktadır. Bir dil bilimci için terimin anlamı farklı, doğal dili işleme ve bilgisayar destekli dil mühendisi ya da bilişimci için farklı bir derlem dilbilimci için farklı anlamları vardır. Doğal dili işleme mühendisleri için bir yapıda gerçekleşen her tür sözcük birlikteliği eşdizim olarak görülebilmektedir (bk. Oflazar, Çetinoğlu ve Say 2004). Bir derlem bilimci için ise dil bilgisel yapıdan bağımsız ardışık her sözcük dizisi eşdizim olarak görülebilmektedir (bk. Özkan, 2020). Bir dil eğitimcisi için ise keyfi kullanım normlarının geliştiği ve bileşenlerinin değiştirim sınırlılığı olduğu iki sözcüğün kalıp kullanımını ifade eder (bk. Howard, 1996). Bu çok-sözcüklü ifadeler çoğunlukla alanyazında sınırlı eşdizimler olarak bilinir (bk. Doğan, 2023).

⁸ Sözlükselleşme ve deyimselleşme ilişkisi için bk. (Lipka 1999) ve (Doğan, 2023).

bilgisel yapı ya da birim arasında köprü vazifesi görür (Biber 2006, s. 34; Biber vd. 1999, s. 990; Biber&Barbieri 2007, s. 269). Türkçeden üçüncü (III) ve dördüncü (IV) tipik özellik için Aksan vd. (2016)'da belirlenen “olan bir başka; ya da olumsuz” gibi ardışık sözcük dizileri örnek gösterilebilir.

Sözcük demetlerini her ne kadar konuşma dilinin ve dilin genel söz varlığının bir niteliği olarak gören çalışmalar olsa da sözcük demetleri çoğunlukla farklı *dil türlerine* (registers), *metin türlerine* (genres) ve *uzmanlık alanlarına* (disciplines) bağımlı bir fenomen olarak görülmüştür. Sözcük demetleri farklı söylem ya da dil türlerinin, metin türlerinin ve uzmanlık alanlarının söylemsel özelliklerinin tanımlanmasında bir araç olarak kullanılır (Salazar, 2014, s. 20). Alanyazındaki çalışmaların da çoğunlukla akademik yayın ve konuşma gibi belirli dil türleri, tarih ve biyoloji gibi belirli uzmanlık alanları ya da üniversite ders ve çalışma kitapları gibi belirli metin türleri üzerine yapıldığı görülmektedir. Sözcük demetlerinin söz konusu alanların geleneksel ifade biçimlerini yansıttığı ve metin ve dil türlerinin kullanılan sözcük demetleri açısından farklılaştığı, belirli sözcük demetlerinin belirli dil türlerinde belirli işlevlerle daha sık ve alana özgü kullanıldığı ileri sürülür (bk. Biber, Conrad, and Cortes, 2004; Biber ve Barbieri, 2007). Sözcük demetlerinin sık kullanımının bir dil türünün yetkin kullandığına işaret ettiği, kullanılan sözcük demetlerinin dil türlerini, uzmanlık alanlarını ve metin türlerini ayırt edici ve betimleyici özelliklerinin olduğu düşünülür (Cortes, 2004, s. 398; Salazar, 2014). Dolayısıyla bir dil türünde, bir uzmanlık alanında ya da bir metin türünde yetkin olabilmek için bunlara özgü sözcük demetlerinin doğallığın, akıcılığın ve ifade gücünün edinilmesinde önemli olduğu ve bu tür yaygın sözcük dizilerinin özel olarak öğrenilmesi gerektiği ileri sürülür (Cortes, 2006). Sözcük demetlerinin yapısal ve deyimsel açıdan tamamlanmamış, bilişsel açıdan da önemli olmamasına karşın sistematik söylem işlevlerinin olduğu, söylemde önemli yapı blokları olarak kullanıldığı, özellikle metin oluşturmada önermenin ve yeni bilginin ifadesi için söylem çerçeveleri olarak işlev gördüğü, daha büyük söz öbekleri ve cümleler için bir edim bilimsel baş (pragmatic head) sağladığı, söylemin sürdürülebilmesini kolaylaştıracak yorumlayıcı ve açıklayıcı çerçeveler oluşturduğu ve metnin önermesine yönelik bir bakış açısı sunduğu ve metinsel anlamlar ifade ettiği kabul edilir (Biber, 2009, s. 284-5).

Sözcük demetlerinin belirlenmesinde kullanılan yegâne ölçüt gerçek anlamda sıklıktır. Sözcük demetleri araştırmalarında sıklığa diğer kalıp dil birimleri araştırmalarında olduğundan çok daha büyük önem ve öncelik verilir. Sözcük demetlerinin, kalıp dil birimlerinin belirlendiği ilk ve öncü çalışmaların⁹ aksine anadili sezgisine ve sezgisel öneme dayalı olarak değil dilin büyük ölçekli derlemelerine dayalı özel bilgisayar yazılımları kullanarak sıklık-çıkışlı bir yaklaşımla ancak belirlenebileceği ileri sürülür. Olası sözcük demetlerinin belirlenmesinde ilk ve en ayırt edici ölçüt sıklık olmasına karşın bir sözcük dizisinin sözcük demeti sayılması için tek başına yeterli değildir. Sıklık ölçütü ve ölçümleriyle elde edilen sözcük dizilerinin sözcük demeti kabul edilebilmesi için belirli bir sıklık kesme noktasına ulaşması ve metinsel dağılım göstermesi gerekir. Ancak tekrarlanma oranı ya da sıklık kesme noktası alanyazında araştırmacıdan araştırmacıya değişmekte ve keyfi olarak belirlenmektedir. Bir sözcük dizisinin sözcük demeti sayılabilmesi için gerçekleşme sıklığı Biber vd. (1999)'da her 1 milyonda 10 kez, Cortes (2004)'te her 1 milyonda 20 kez; Chen & Baker (2010)'da her 1 milyonda 25 kez; Biber & Barbieri (2007) ve Biber, Conrad ve Cortes (2004)'te her 1 milyonda 40 kez olarak belirlenmiştir. Türkçe alanyazında ise sıklık kesme sıklığı üç-sözcüklüler için Aksan&Aksan (2015a)'da her milyonda 15 kez; Aksan (2016)'da her milyonda 100 kez; iki-sözcüklüler için Aksan&Aksan (2015a)'da her milyonda 100 kez, Aksan (2016)'da her

⁹ Bk. Pawley&Syder 1983, Howart 1996, Wray&Perkins 2000, Wray 2002.

milyonda 200 kez olarak kararlaştırılmıştır. Sıklık kesme noktasının kullanışlı olması için sık tekrarlanan sözcük demetlerinin metinsel dağılımının dengeli olması beklenir. Bir sözcük dizisinin sözcük demeti sayılabilmesi sıklık kesme noktasının yanı sıra derlemde 5 farklı metne dağılmasını gerektirir. Bu kısıtlamayla bireysel konuşur veya yazarların kendine özgü kullanımlarının sözcük demeti sayılmasının önüne geçilir ve gerçek sözcük demetlerine koruma sağlanır (Biber, 2006, s. 134; Biber, Conrad ve Cortes, 2004, s. 376).

2. Kalıpsallığın Ölçüt ve Ölçümleri: Sözcük Demetleri Kalıpsal (Formulaic) mıdır?

Deyim, atasözü ve kalıp söz gibi geleneksel kalıp dil birimlerinin tespitinde bilişsel belirginlik ve sezgisel önem ya da anadili sezgisi başat ölçütlerden biri olarak kullanılmıştır. Bilişsel belirginlik ve sezgisel önem, alanyazında ilk çalışmalarda bir sözcük dizisinin kalıpsal statüsüne ve potansiyeline işaret eden belirleyici bir nitelik olarak görülmüştür. Derlem-çıkışlı sıklık yaklaşımında ise bilişsel belirginlik ve sezgisel önem kalıpsallığın göstergesi olarak kabul edilmez. Sıklık, özellikle sözcük demetlerinin belirlenmesinde yegâne ölçüt olarak benimsenir. Biber, Conrad ve Cortes (2004)'e göre bir sözcük dizisinin ne ölçüde önüretimli ve kalıpsal olduğunu ve hafızada ne ölçüde bütün olarak saklandığını yüksek sıklık yansıtır. Yüksek sıklıktaki sözcük dizileri düşük sıklıktaki sözcük dizilerine kıyasla hafızada bütün olarak saklanır ve analiz edilemez bütünler (chunks) olma olasılığı çok daha yüksektir. Conrad ve Biber (2004)'e göre yüksek sıklıktaki donuk sözcük dizilerinin konuşur ve yazarlar için analiz edilemez bütünler olması metinlerde belirli söylem işlevleri kazanmasını sağlamaktadır. Sözcük demetlerinin söylemin temel yapı blokları olarak kullanıldığı güçlü işlevsel ilişkileri vardır ve bu, sıklığa göre belirlenebilir (2004, s. 58-59). Sıklık yaklaşımını benimseyen araştırmacılar, sıklık ölçütüne göre belirlenen sözcük demetlerinin belirli dil ve metin türlerine ve uzmanlık alanlarına özgü ayırt edici metinsel ve iletişimsel işlevlerinden dolayı rastlantısal olmadığını, aksine yüksek sıklığın onların önüretimli yahut kalıpsal statüsünün göstergesi olduğunu, analiz edilemez bütünler olarak hafızada saklandığını ve kullanıldığını, özellikle üniversite öğrencilerinin başarısını etkilediğinden pedagojik açıdan da önemli olduğunu, kendiliğinden edinmekten ziyade özellikle öğretilmesi gerektiğini iddia ederler (Biber ve Barbieri, 2007, s. 283-284).

Biber ve arkadaşlarının sözcük demetlerinin belirlenmesinde sıkı sıkıya benimsedikleri sıklık-çıkışlı yaklaşımına alanyazında itirazlar yükselmiş, sıklığın kalıpsal ya da önüretimli nitelikteki dil birimlerini her zaman belirleyemediği ileri sürülmüştür. Wray (2002, s. 31) kalıp dil birimlerinin belirlenmesinde sıklık ölçütüne önem vermemiş; sık tekrarlanan tüm sözcük dizilerinin kalıpsal olmadığını, bir sözcük dizisinin sıklığı olmasına karşın önüretimli yahut kalıpsal olmayabileceğini ileri sürer. Wray'ın (2002) sıklık yaklaşımına olan itirazı başka araştırmacılar tarafından da desteklenir. Hudson (1998, s. 26-27) sıklığın kalıp dil birimlerinin oluşmasında bir faktör olduğunu ancak zorunlu olmadığını belirtir. Wood (2015, 2019) salt sıklığa dayalı belirlenen sözcük demetlerinin bir kalıp dil birimi "formulaic sequences" olarak görülmesine karşı çıkar. Belirgin bir önemi ve bütüncül bir anlam ya da işlevi olan bazı kalıpsal sözcük dizileri derlemde gerçekten yüksek sıklık göstermeyebilir, aksine kalıpsal sözcük dizileri derlemde birkaç defa görünebilir (2019, s. 37-38). Dolayısıyla sözcük demetlerinin ya da herhangi bir çok-sözcüklü dil biriminin kullanım sıklığının olması onların önüretimli ya da kalıp dil ürünü oldukları anlamına gelmez (Wood 2019, s. 15). Bir kalıp dil birimi salt sıklıktan çok daha fazla bir şeydir; kalıpsal bir dil biriminin hafızada bir bütün olarak saklanması ve hafızadan bir bütün olarak getirilip bir bütün olarak kullanılması, özellikle de bir anlam ya da işlev birimi olması gerekir (Wood 2015, s. 20). Wood, salt sıklığa göre belirlenen sözcük

demetlerinin kalıp dil birimi olmaktan daha ziyade “sözcük dizileri”¹⁰ olarak düşünülebileceğini ileri sürer (2019: 35). Wood, sıklığın bütüncül işleme ya da kalıpsallığın hiçbir şekilde gerçek bir göstergesi olamayacağını, bu nedenle beraberinde psiko-dilbilimsel ölçüm, anadili sezgisi (sezgisel önem) gibi başka analiz ölçeklerinin kullanılabileceğini ifade eder (2019, s. 37). Schimitt vd. (2004) bunlardan en dikkat çekenidir. Onlar derleme dayalı sıklık yaklaşımıyla belirlenen sözcük demetlerinin psiko-dilbilimsel açıdan önemli olup olmadığını, hafızada kalıpsal ve bütün olarak saklanıp saklanmadığını test etmek için farklı metotlara¹¹ başvurmuştur. Schimitt vd. (2004, s. 138) deneysel araştırmalarında sıklık çıkışlı yaklaşımla belirlenen sözcük demetlerinin psiko-dilbilimsel açıdan geçerli ve önemli olmadığını, dolayısıyla usta konuşurların hafızasında kalıpsal ve bütün olarak saklanmadığını belirlemiştir. Benzer itirazlar başka araştırmacılardan da gelmiştir. Ellis vd. (2008, s. 380) sıklık çıkışlı yaklaşımla belirlenen n-birim ya da sözcük dizilerinin yüksek sıklığı olmasına karşın bu, onların belirli ve ayırıcı işlev ve anlamlarının olduğu anlamına gelmeyeceğini, bunların çoğunun dil bilgisel birimlerin ve işlev sözcüklerinin yüksek sıklığı dolayısıyla sıradan ve rastlantısal oluştuğunu ileri sürer. Ellis vd. (2008) ikinci dil edinimi ve akademik yazma ve konuşma gibi alanlarda pedagojik açıdan önemli ve psiko-dilbilimsel geçerliliği olan kalıp dil birimlerinin belirlenebilmesi için yüksek sıklığın yanı sıra bir derlem verisi olan MI (mutual information) ölçümlerinden faydalanılması gerektiğini savunur. MI ölçümü, çok-sözcüklü bir birimin sıklığını, onu oluşturan sözcüklerin her birinin genel sıklığıyla karşılaştırır, böylece iki sözcüğün bir nedenden dolayı ve sadece rastlantısal olarak şans eseri birlikte ortaya çıkma olasılığını yansıtır (Manning & Schütze, 1999). MI, sözcük dizilerindeki sözcüklerin rastlantısal olarak beklenenden daha sık birlikte gerçekleşme derecesini değerlendirmek için tasarlanmış, bilişim bilimi alanında yaygın olarak kullanılan istatistiksel bir ölçüm olup bir öbekteki sözcüklerin ne kadar uyumlu olduklarının veya ne kadar aynı dizimde gerçekleştiklerini ölçer. Daha yüksek bir MI skoru, kelimeler arasında daha güçlü bir ilişki ve birlikteliğin olduğu anlamına gelirken daha düşük MI skoru onların birlikteliğinin şansa bağlı gerçekleşme ihtimalinin daha yüksek olduğunu gösterir¹² (Ellis vd. 2008, s. 380). Ellis vd. (2008) derlemden MI skoru ile belirlenen n-birim ya da sözcük dizilerinin salt yüksek sıklığa göre belirlenenlerin aksine rastlantısal olmadığını ve birlikteliğin bağdaşık olduğunu ve bağdaşık birlikteliğin dil bilgisel açıdan iyi biçimlenmiş, ayırt edici ve belirli işlev ve anlamlara karşılık geldiğini ileri sürer. Biber vd.nin (1999 ve 2004) salt sıklığa dayanan sözcük demeti yaklaşımının basit ve doğru metodolojik bir avantajı olmasına karşın salt sık tekrar eden uzun sözcük dizileri listesiyle sonuçlandığı, sezginin önemli sayacağı ayrımları ortadan kaldırdığı, sezgisel olarak bilinebilecek psiko-dilbilimsel açıdan daha öncelikli ve önemli, dolayısıyla kalıpsal sözcük dizilerinin önemsiz diğer sözcük dizileriyle aynı listede sıraladığı görülür (Simpson-Vlach&Ellis 2010, s. 490). Dolayısıyla derlem-çıkışlı sıklık yaklaşımıyla sadece kalıpsal ya da psiko-dilbilimsel açıdan önemli çok-sözcüklü birimleri değil sıklığı yüksek ne kadar sözcük dizisi varsa hepsi aynı listede toplanır. Bu listelerdeki sözcük dizileri tümüyle aynı nitelikte olmadığından bazıları diğerlerine göre bilişsel ya da psiko-dilbilimsel açıdan çok daha önemli ve belirgin

¹⁰ Wood, her ne kadar çok-sözcüklü birim “multi-word unit” terimini kullansa da bu terim bizim bu çalışmada sözcük dizileri terimiyle karşıladığımız kategoriye kasteder.

¹¹ Schimitt vd. (2004) deneysel araştırmalarında *dikte testi*ni kullanmışlardır. Ancak bu çalışmanın amacı, kullanılan test ve ölçütlerin güvenilirliğini sınamak ya da birbirine üstünlüğünü göstermek olmadığından ölçütlerin ayrıntılarına ve nasıl uygulandığı meselelerine girilmemiştir.

¹² Sözcükler arası birliktelik ya da eşdizimlilik gücünü hesaplayan bir derleme dayalı ölçüm aracı bir dil için elverişli ve uygun olan bir eşdizimlilik ve birliktelik ölçümü başka bir dil için uygun olmayabilir. İngilizce için IM değeri uygun birliktelik gücünü hesaplayan bir eşdizimlilik ölçümü iken aynı ölçüm Türkçe için aynı sonuçları vermez. Mersinli (2015)’in Türkçe için birliktelik gücünü kullanışlı olarak hesaplayan ölçüm araçları içerisinde IM değeri yoktur. Türkçede iki-sözcüklü birimler için log-likelihood ölçeği daha iyi sonuçlar verir. Derlemdaki uygulamaları için 5. Bölüme bakınız.

olduğundan sıklık-çıkışlı yaklaşımla belirlenen sözcük dizileri arasında bir ayırım yapılması gerekir (Simpson-Vlach&Ellis, 2010, s. 490). Türkçe alanyazından Aksan vd. (2016)'da n-birimler "n-grams" ve çok-sözcüklü birimler "multi-word units" terimleriyle karşıladıkları iki farklı kategoriden söz edilir. Derlemde gözlenen sıklığı yüksek herhangi bir sözcük dizisinin n-birim; n-birimlerin biçim-sözdizimsel filtrelemeden geçirilmesiyle elde edilenlerin ise çok-sözcüklü birim olduğu ve bunların mental sözlükte (lexicon) tek bir birim olarak saklanan geçerli "sözlükselleşmiş sözcük dizileri" olduğu ileri sürülür¹³ (Aksan vd. 2016, s. 74). Aksan vd. (2016)'nın bu ayırımı de sıklık yaklaşımının farklı sözcük dizilerini aynı listede topladığını doğrulamaktadır. Biber vd. (2009)'un sıklık yaklaşımıyla derlemden çıkarılabilecek sözcük dizileri için *çok-sözcüklü örüntüler* (multi-word patterns), *sözlüksel-dilbilgisel birleşmeler* (lexico-grammatical combinations), *çok-sözcüklü diziler* (multi-word sequences), *çok-sözcüklü kalıp diziler* (multi-word formulaic sequences) gibi farklı terimleri kullanması da bununla ilgilidir. Bu çeşitlilik, sıklık çıkışlı yaklaşımın yanı sıra başka ölçüt ve ölçümlerin kullanılmasını zorunlu kılmaktadır. Simpson-Vlach&Ellis (2010) sezginin önemine vurgu yaparak tamamen sıklığa dayalı listelerin sezgisel olarak ayıklanmasının pedagojik açıdan ayırt edici, çok daha kısa kalıp dil birimleri listesi sağlayacağına inanırlar. Bu nedenle Simpson-Vlach&Ellis (2010, s. 490) psiko-dilbilimsel ve pedagojik açıdan önemli sözcük demetlerini tespit edebilmek için sıklık ve sezginin birlikte kullanılması gerektiğini savunur.

Salazar (2014) ise yüksek sıklık ile IM ölçümlerinin avantajlarından faydalanabilmek ve dezavantajlarından kaçınabilmek için her iki yaklaşımı birleştirerek eş zamanlı kullanmayı tercih etmiştir. Söz konusu niceliksel ve sayısal ölçümlerin daha ileri araştırmaların bir başlangıç noktası olabilecek sözcük dizilerinin ana listesini oluşturmak için güvenilir ve basit bir yöntem sağladığını ancak ölçümlerin hâlâ pedagojik açıdan önemsiz uzun sözcük demetleri listeleriyle sonuçlandığından tek başına yeterli olmadığını düşünmektedir. Salazar (2014, s. 46) bu nedenle, sıklık ve MI tabanlı yaklaşımla belirlenen ana listeyi daha da daraltmak ve bazı özelliklerinden dolayı listeye dâhil edilemeyecek sözcük dizilerini ortadan kaldırmak amacıyla "sezgisel seçime dayana belirli dışlama ölçütleri" oluşturulması gerektiğini iddia eder. Sezgisel seçimin akademik yazma eğitiminde pedagojik açıdan en yararlı sözcük demetleri listesinin oluşturulmasına büyük katkı sağladığı düşünülür. Salazar (2014) sezgisel seçim ölçütünü kullanarak İngilizcede benimsenen 9 (dokuz) farklı yapıdaki sözcük demetini dışarıda bırakmıştır. Bunlara Türkçede muadilleri olan Aksan vd. (2016)'da belirlenen şu 3 kategori örnek verilebilir.

i) Diğer sözcük demetlerinin ya da farklı dil bilgisel yapıların parçalarından oluşanlar: *belki de en, yer alan ve, diye bir şey, kamu kurum ve, için ne kadar vs.*

ii) Tamamen işlev sözcükleriyle oluşanlar: *ya da bir, ya da daha, ya da başka, ve belirli bir vs.* (Aksan vd. 2016)

iii) Belirtlilik işaretleyicisiyle oluşanlar: *çok önemli bir, çok büyük bir, kadar büyük bir vs.*

Salazar'ın (2014) sezgisel seçim işlemiyle Simpson-Vlach&Ellis (2010) gibi daha çok dil bilgisel açıdan tamamlanmamış yapıları çıkardığı görülür. Türkçede salt sıklık yaklaşımıyla Aksan vd. (2016)'da belirlenen "*bu da bu, sen de bu, o da bu, o zaman o, olan bir başka, ya da çok, ve yeni bir, ya da olumsuz*" gibi farklı tamamlanmamış sözcük diziler ya da

¹³ Aksan vd. (2016) sözcük demetlerinin sözlükselleşmesini n-birimlerin biçim-sözdizimsel filtrelemeyi aşmasına bağlar. Ancak biçim-sözdizimsel yapı tipleri n-birimlerin çıkarımında kullanılan yaklaşımla belirlendiği için biçim-sözdizimsel filtreleme işlemi de aslında tekrar n-birimleri ortaya çıkarmakta ya da toplamaktadır. Bundan dolayı biçim-sözdizimsel filtreleme yüksek sıklık kesme noktası kullanılmadığında çok geçerli sonuçlar vermediği görülmektedir. Araştırma sürecinde sıklık kesme noktasının 15'ten 100 ve 200'e çıkarılmasının sebebinin de bundan kaynaklandığı söylenebilir.

n-birimlerin kalıpsal olmadığı, pedolojik ve sözlüksel açıdan rahatlıkla göz ardı edilebileceği, buna bağlı olarak metinsel işlevinin de belirgin ve ayırt edici olmadığı söylenebilir. Bununla birlikte sözcük demetleri sezgisel önem gereği yapısal açıdan tamamlanmış olmamalarına karşın belirli sözlüksel, işlevsel ve pragmatik özellikleriyle sözlükselleşmiş dil birimlerinden oluşabilir. “hiç mi hiç, hem de çok, ama yine de, diğer yandan da” gibi sözcük demetlerinin bu nitelikte olduğu söylenebilir. Dolayısıyla tamamlanmışlık ya da belirli yapı temelli dışlama kriterlerinin de anadili sezgisi olmadan tek başına yeterli olmadığı söylenebilir.

Biber, sıklık ile kalıpsallık arasında doğru orantı olduğunu ileri sürmesine karşın salt sıklığın tüm kalıp dil birimlerini belirlemede her zaman kullanışlı olmadığı, çoğu kalıp dil biriminin düşük sıklık oranlarının olabileceği, ayrıca rastlantısal ya da önemsiz eşdizimlerin oldukça yüksek sıklık oranının olduğu yönündeki itirazları da bir taraftan kabul eder (2009, s. 286). Ancak Biber (2009), eşdizim ve sözcük demeti gibi farklı kalıp dil birimlerin belirlenmesinde farklı derlem-çıkışlı istatistik ölçümlerin kullanılması gerektiğini ileri sürer. Biber (2009), salt sıklığın sözcük çiftleri arasındaki eşdizimsel gücü ve ilişkiyi hesaplayamadığını, eşdizimleri belirlemede başarısız olduğunu, bununla birlikte sözcük demetlerinin derlemden çıkartılmasında kullanışlı olduğunu, bu nedenle eşdizimlerin ve sözcük demetlerinin belirlenmesinde farklı istatistik ölçümlerin ve derlem metodunun kullanılması gerektiğini savunur. Biber (2009), MI istatistik ölçümünün bir sözcük birleşmesinin ya da sözcük çiftlerinin eşdizimsel gücünü ve olasılığını hesapladığını, ancak sözcüğün sıra dizisi (order) dikkate alınmadığından sözcük demetlerini gözden kaçırdığını, dolayısıyla MI ölçümü ve istatistik hesaplamasının çok-sözcüklü birimlere uygulandığında sorun çıkaracağını ileri sürer. Diğer taraftan Biber (2009), MI ölçümünün daha çok iki içerik sözcüğü arasındaki eşdizimsel ilişkiyi ya da gücü hesapladığını; sözcük demetlerinin ise dil bilgisel ya da işlev sözcükleri ile içerik sözcüklerinden oluştuğunu, MI ölçümünün içerik ve işlev sözcüklerinden oluşan çok-sözcüklü birimlerin istatistiklerini hesaplayamadığı, dolayısıyla MI ölçümünün sözcük demetlerini belirlemede kullanışlı olmadığını ileri sürer. Ona göre farklı çok-sözcüklü kalıp ifadeler vardır, bunların belirlenebilmesi için farklı nicel metotlar kullanılmalıdır (Biber, 2009, s. 288). Biber (2009), İngilizce için içerik ve işlev sözcüklerinden oluşan kalıpsal çok-sözcüklü birimlerinin belirlenmesinde sıklık-çıkışlı yaklaşımın; iki içerik sözcüğünün eşdizimsel gücünün hesaplanmasında ise MI ölçeği ve yaklaşımının daha kullanışlı olduğunu ileri sürer.

Sıklık-çıkışlı yaklaşımda, sıklık önemli bir ölçüt olmakla birlikte tek başına kesin ya da açıklayıcı bir veri olarak kabul edilmez. Biber, Conrad ve Cortes (2004) sıklık verisinin açıklanması gereken sözcük örüntülerinin (pattern) belirlenmesine olanak sağladığı, araştırmacıların çoğunlukla fark edemediği kullanım şekillerini ortaya çıkardığını ifade eder. Sıklık çıkışlı yaklaşımın iki aşamalı olduğu söylenebilir. İlk aşamada derlemde sık tekrar eden sözcük dizileri belirlenir, ikinci aşamada sözcük dizileri belirli sıklık kesme noktasına göre ayıklanır. Cortes (2004) sıklık kesme noktasının 20 ila 40 kez olmasının sözcük demetlerini belirlemek için kullanılabilir ve garantileyecek kadar yüksek bir oran olduğunu belirtir. Dolayısıyla sıklık kesme noktalarının sözcük dizilerinin sözcük demeti kabul edilmesinde belirleyici olduğu, bir sözcük dizisinin kalıpsal olup olmadığını belirlemek için kullanıldığı söylenebilir. Ancak belirlenen sıklık kesme noktalarının keyfi olması kalıpsallığın ya da sözcük demeti olma ölçütünün de keyfi olduğu anlamına gelir. Sıklık kesme noktasının özellikle İngilizce 3 sözcüklü sözcük demetlerinin belirlenmesinde listelenemeyecek kadar fazla sözcük dizisi derlediğinden çok da kullanışlı olmadığı görülür (bk. Biber vd. 1999). Diğer taraftan Türkçe alanyazında iki-sözcüklü sözcük demetlerinin belirlenmesinde sıklık kesme noktasının her milyonda 100 ya da 200 defa olacak şekilde belirlenmesi keyfiliğin yanı sıra sözcük demetlerinin belirlenmesini de imkânsız kılmıştır.

Kalıpsallığın sıklıkla doğru orantılı olmadığı dikkate alındığında hiçbir sıklık kesme noktasının kalıpsal dil birimlerini tespit etmek için geçerli ve yeterli bir ölçüt olmadığı söylenebilir. Kaldı ki sözcük demetlerinin derlemden çıkartılmasında benimsenen sıklık kesme noktasının her milyonda 10'dan 200'e kadar uzanması sıklık kesme noktasının da en az sezgisel önem kadar keyfi olduğu anlamına gelir. Dolayısıyla benimsenen ölçüt ve yaklaşımın; bu ölçüt ve yaklaşımlarla elde edilen sonuçların birbirine üstün olmadığını gösterir.

3. Sözcük Demeti ve Eşdizim¹⁴ Ayrımı

Türkçede çok-sözcüklü birim terimi, artık sözcük demetleri ve eşdizimler de dâhil farklı kalıp dil birimlerini içerecek şekilde ve aralarında bir ayrım gözetilmeden kullanıldığı için ayırt edici ve kullanışlı değildir (bk. Aksan 2016). Kalıp dil birimleri arasında dil bilimsel amaçlara dönük bazı betimleyici ayrımların yapılması gerekir. Batı alanyazında da dil öğretimi ve sözlükçülük gibi belirli amaçlara yönelik olarak eşdizim ile sözcük demetleri arasında ayrım yapılması her zaman önemsenmiştir (Stubbs, 2002; Biber, 2009). Çok-sözcüklü birim araştırmalarında eşdizim ile sözcük demeti arasında dört önemli ayırt edici özellik belirlenir. İlki, yukarıda da ifade edildiği gibi sözcük demetlerinin uzunluk ya da sözcük sayısı açısından 3 ve daha fazla sözcükten oluştuğu düşünülür ve *genişletilmiş eşdizimler* (extended collocations) olarak adlandırılır (Biber vd. 1999). Ancak Türkçede sözcük demetlerinin sözcük sayısı ya da uzunluk açısından eşdizimlerden ayırtılabileceğini ileri sürmek güçtür. İkincisi sözcük demetlerinin söylemde çok sık tekrarlanmasıdır ancak benimsenen sıklık kesme noktalarının keyfi olması ve sıklığın her zaman belirleyici olmaması bunu da kullanışlı bir ölçüt olmaktan çıkarmaktadır. Üçüncüsü deyimleşmemiş olmak ancak eşdizimler de deyimleşmemiş¹⁵ olduğundan bunun da ayırtıcı olmadığı söylenebilir. Dördüncüsü dil bilimsel açıdan tamamlanmamış olmak ancak daha önce ifade edildiği gibi kimi sözcük demetleri kalıpsallığa bağlı olarak tamamlanmış ve sözlükselleşmiş de olabilmektedir. Dolayısıyla bu ölçütlerin sözcük demetlerini tam olarak eşdizimlerden ayırt ettiği söylenemez. Stubbs (2002) ve Biber (2009) eşdizim ve sözcük demetleri arasında önemli temel iki ayrım yapar. Onlara göre eşdizimler öncelikle iki *içerik sözcüğü* (content words) (isim, fiil, sıfat ve zarf) arasında gerçekleşir. Sözcük demetleri ise *dil bilimsel sözcükler* (grammatical words) adı verilen *işlev sözcükleri* (function words) ile içerik sözcüklerinin diziminden oluşur. Ancak burada özellikle de Türkçe açısından işlev ve içerik sözcüğü arasında kesin bir ayrım yapmanın güçlüğü baş göstermektedir. Buna karşın edatların, bağlaçların, bazı belirtme sıfatlarının ve zamirlerin işlev sözcüğü olduğu söylenebilir. *Saygı göstermek* bir eşdizim iken *bununla birlikte, her ne kadar, ama yine de* kalıp dil birimlerinin sözcük demeti olduğu söylenebilir. Bir eşdizim, sözcük birimleri arasındaki güçlü birliktelik ilişkilerinden oluşabilirken sözcük demetlerinin sözcük biçimleri arasında sıra düzen bağımlı gerçekleştiği, sözcük demetlerinin dizim ve sözcük biçimi açısından katı bir donukluğunun olduğu kabul edilir. Bir başka deyişle eşdizimlerde "sözcük birimsel donukluk" varken sözcük demetlerinde "sözcük biçimsel bir donukluk" vardır. Bu, genellikle *dizimsel* (syntagmatic) ve biçim-sözdizimsel bir donukluğu da içerir. Örneğin *tavır takınmak* fiil eşdizimi sözcede *takındığı*

¹⁴ Eşdizim dar ve geniş anlamda iki farklı anlamı vardır. Geniş anlamda eşdizim iki sözcüğün derlemde ardışık düzende bitişik ya da ayrı olarak belirli bir mesafe ya da pencere aralığında sıklık kullanımını ifade eder. Geniş anlamda eşdizim iki sözcüğün aynı dizimde yan yana ya da bitişik bulunduğu sözcük dizilerini ifade eder. Herhangi iki sözcüğün dil bilimsel yapıdan bağımsız aynı dizimde artarda kullanılması işlevsel ve anlamsal birim olmayı gerektirmez. Dar anlamda eşdizim ise belirli dil bilimsel yapılarda gerçekleşen, bileşenlerinin birbirini keyfi olarak kısıtladığı ve gerektirdiği, önüretimli ve bilişsel önemi olan anlamsal birim ya da bütün olarak kullanılan kalıp dil birimi türünü ifade eder ve yazın alanında çoğunlukla sınırlı eşdizimler kastedilir.

¹⁵ Biber vd. (1999) bu terimi daha çok sözlükselleşme anlamında kullanılır. Zira Lipka (1999) deyimleşmeyi sözlükselleşmeyle eşzamanlı gelişen bir süreç olarak değerlendirir.

tavır, tavır takınması, tavır takındı, takındığı kabul edilemez bu tavrın gibi sözcük birimleri arasında ve sözcük biçimlerinden bağımsız farklı dil bilgisel yapı ve dönüşümlerde kullanılabilir. Buna karşın *bundan böyle, ama yine de ve her ne kadar* sözcük demetlerinin **yine de ama, *böyle bundan, *ne kadar her* dizimleri kabul edilebilir değildir. Sözcük demetleri belirli sözcük biçimlerine bağımlıdır ve sözcük biçimlerinin belirli ve sabit bir dizimde donuk şekilde kullanılmasıyla gerçekleşir. Sıklık-çıkışlı yaklaşımda *diye geçirdi içinden*¹⁶ sözcük demetinin *diye geçirecek içinden, diye içinden geçirdi, diyerek geçirdi içinden* gibi farklı dizim ve biçimlerinin sıklık kesme noktası koşulunu sağlayamadığında sözcük demeti kabul edilmez. Sıra düzen bağımlı dizimsel ve biçim-sözdizimsel donukluk sözcük demetlerinin ayırıcı ikinci başat özelliğidir. Ancak *diye geçirdi içinden* ifadesinin donuk olmadığı, dizimsel ve biçim-sözdizimsel farklı kullanımlarının da sezgisel önemi olduğu dikkate alındığında bileşenlerinden biri işlev sözcüğü olsa da sözcük demeti olmadığı anlamına gelir. Sözcük demetlerini eşdizimlerden ayıran diğer bir özelliği ise onların metinsel ve söylemsel işlevleri olmasıdır. Buna bağlı olarak eşdizimlerin bir sözlüksel anlamı vardır ve içerik sözcükleri sınıfında yer alır; sözcük demetlerinin ise metinsel ve söylemsel işlevleri vardır ve işlev sözcükleri sınıfında yer alır. Dolayısıyla sözcük demetleri dil bilgisel kategorileri ve işlevleri ifade ederken eşdizimler sözlüksel anlam birimleri ifade eder.

4. Çalışmanın Kuramsal Temeli: Sözcük Demetleri Sözlükselleşmiş midir?

Sözcük demetlerinin çoğunlukla metinsel ve söylemsel işlevleri olmakla birlikte bir kısmı sözlükselleşmiş; bir kısmı sözlükselleşmemiş iki farklı kategoriden oluştuğu söylenebilir. Biber vd.nin (1999) araştırma geleneğinde sözcük demetleri dil bilgisel açıdan tamamlanmamış, deyimselleşmemiş, bilişsel ve sezgisel önemi olmayan, tümlenebilir ve şeffaf kabul edildiğinden söylem düzeyinde oluşan, pragmatik açıdan metin oluşturma ve söylem işlevleri olduğundan çoğunlukla sözlükselleşmemiş çok-sözcüklü birimler olarak düşünülmüştür. Ancak pedagojik, psiko-dilbilimsel ve leksikolojik açıdan önemli sözcük demetlerinin kayda değer bir kısmının belirli pragmatik ve söylemsel işlevleriyle sözlükselleştiği görülür. Lipka (1999) sözlükselleşmeyi¹⁷ deyimselleşmeyle eş zamanlı gelişen bir süreç olarak tasarlar. Cortes'in (2004, s. 400) sözcük demetlerinin belirlenmesinde deyimselleşmenin de bir ölçüt olarak kullanılabilceğini ileri sürmesi kimi sözcük demetlerinin sözlükselleştiği fikrini desteklemektedir. Bu görüş, Türkçe alanyazından Çalışkan (2014b, s. 51) ve Aksan vd. (2016, s. 74) tarafından da dile getirilir. Conrad ve Biber (2004, s. 58) sözcük demetlerinin betimlenmesinde sıklık, donukluk ve sözcük uzunluğu ölçütlerini kullanmasına karşın bu ölçütler diğer çok-sözcüklü birimlerin de özellikleri olabileceğinden özellikle sözlükselleşmiş sözcük demetlerini belirlemek ve ayırt etmek için kullanışlı ve yeterli değildir¹⁸. Kopaczky (2013) sözlükselleşmiş sözcük demetlerini kalıpsallığın diğer süreçleriyle ilişkilendirerek betimler. Kopaczky (2013) öncelikle çok-sözcüklü birim teriminin belirsizliğinden kaçınmak için *önüretimlilik* (prefabrication) ve *önüretimliler* (prefabs) terimlerini kullanır. Önüretimlilik dildeki kalıpsallık olgusunu ve aşağıda iki türde değerlendirdiğimiz donukluk¹⁹ sürecini ifade

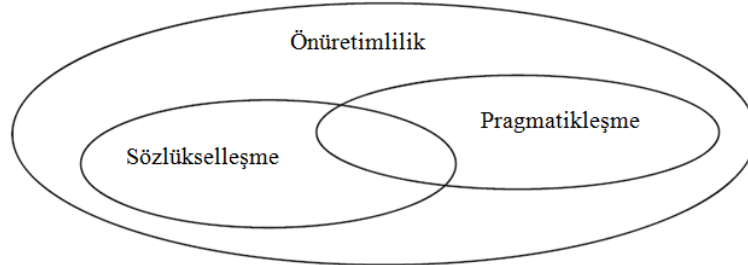
¹⁶ Aksan vd. (2006)'dan alınmıştır.

¹⁷ Sözlükselleşme sadece birden fazla sözlüksel unsurdan oluşan anlam birimlerini ifade etmez. Çok-sözcüklü birimlerin sözlükselleşmesi sözlükselleşmenin çıktılarında yalnızca biridir. Sözlükselleşmenin bir alt kategori olarak kalıpsallıkla başladığı ve bazı kalıp dil birimlerinin sözlükselleşmenin ürünü olduğu söylenebilir.

¹⁸ Bu değerlendirme daha önceki açıklamalarla çelişiyor gibi gelebilir. Ancak çok-sözcüklü birim ya da kalıp dil birimlerinden söz ederken sadece sözcük demetleri ve eşdizimler kastedilmemektedir. Bu belirleme yapılırken birçok diğer kalıp dil birimleri kapsayacak şekilde ifade edilmiştir.

¹⁹ Biber vd. (1999) ilk çalışmasında sözcük demetlerinin donuk olmadığını, devam niteliğindeki çalışmalarında (Conrad ve Biber, 2004) ise donuk olduğunu belirtir. Biber ve arkadaşlarının donukluktan kastettiği şey sözlükselleşmeye bağlı gelişen değiştirilebilirlik kısıtlılıkları değildir. Daha çok sözcük demetinin bileşenlerinin

eder. Önüretimlilik²⁰ yazarın önüretimliler²¹ adını verdiği sözlükselleşmiş çok-sözcüklü birimler, donuk sözcük birlikleri (fixed chunks) ve sözcük demetleri olmak üzere çeşitli kalıp dil birimlerini üreten kapsayıcı bir süreçtir. Önüretimlilik süreci dilde gelişen sözlükselleşme ve pragmatikleşme olmak üzere iki alt süreci ve alanı içerir. Kopaczky (2013, s. 57) sözlükselleşme ile pragmatikleşme süreçlerinin dilde kesiştiği ve üçüncü ayrı bir alan tasarlar. Şekil 1, önüretimlilik, pragmatikleşme ve sözlükselleşme süreçlerinin birbiriyle söz konusu ilişkisini gösterir.



Şekil 1. Sözcük Donukluğu Süreçlerinin İlişkisi

Kopaczky (2013, s. 57) sözlükselleşmeyi Brinton and Traugott (2005)'ten hareketle "birden fazla sözlüksel unsurun birleşmesinden türeyen anlam birimleri" olarak tanımlar ve söz konusu çok-sözcüklü birimlerin sözlükselleşmenin bir ürünü ya da sonucu olduğunu düşünür. Sözlükselleşme farklı dereceleri olan bir süreç olup dil bilgisel açıdan tamamlanmış, geçirimsiz (opaque), başka deyişle anlamsal açıdan kapalı ya da tümlenemez birimlerle son bulur. Pragmatikleşme ise belirli söylem koşullarında, metin düzeyindeki gereksinimlerin yön verdiği sözcük biçimlerinin donuklaşmasına neden olan bir diğer süreçtir. Belirli bir metin türünün üretilmesinde ve belirli bir söylem işlevinin iletilmesinde kullanılan, söylem ve bağlamla yakından ilgili bu sözcük birliklerinin oluşumunun arkasındaki süreç pragmatikleşme²² adıyla anılır. Pragmatikleşme bir sözlüksel unsurun işlev sözcüğüne geçişinin bir türü olarak kabul edilir²³. Dolayısıyla bu iki farklı sürece bağlı olarak farklı kategoriler gelişir. Önüretimlilik bir tarafta sözlükselleşmeye bağlı olarak sözlükselleşmiş sözcük demetlerini; diğer tarafta pragmatikleşmeye bağlı olarak iletişimdeki rolü sadece sözlüksel anlamından kaynaklanmayan fakat belirli bir bağlamın ürünü olan dilin donuk sözcük birliklerini²⁴ içerir. Kopaczky (2013, s. 59), donuk sözcük birliklerini donukluk sürecinin ürünü olarak "pragmatik önüretimliler" olarak tanımlar. Kopaczky, Şekil 1'de sözlükselleşme ve pragmatikleşmenin kesiştiği özel alana özellikle dikkat çeker. Bu alanda sözcük demetlerinin sözlükselleşmiş, yarı-sözlüksel ve yarı-pragmatik sınıfının yer aldığını düşünür. Dil türü, metin türü ve uzmanlık alanlarında kullanılan bu kategorinin dilin standartlaşmasını sağladığı ve söylemin vazgeçilmez parçalarını oluşturduğu ileri sürülür (2013, s. 60).

belirli sözcük biçimleriyle yüksek sıklık kullanımının olması, diğer dil bilgisel eşdeğerlerin anlamsal koşulları sağlasa bile düşük sıklığı olması dolayısıyla göz ardı edilmesidir. Örneğin "dedi kendi kendine" bir sözcük demeti kabul edilebilirken "demiş kendi kendine" sözcük dizisinin sıklık kesme noktasını aşmadığından sözcük demeti olarak kabul edilmemesi bundan ileri gelir.

²⁰Bu terim, kalıpsallık ile eş anlamlı kullanılır.

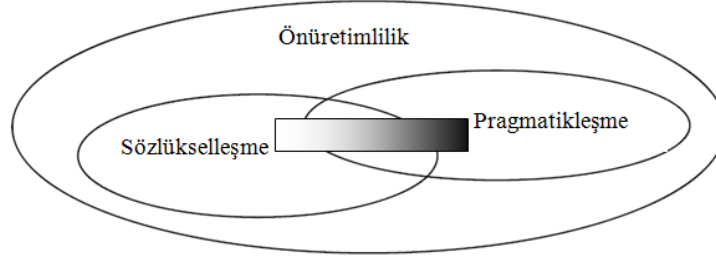
²¹ Kopaczky (2013), çok-sözcüklü birim ifadesinin belirsizliğinden ve kolayca sınıflandırılmayacak kadar çeşitlilik arz ettiğinden önüretimli (prefab) teriminin daha doyurucu ve kullanışlı olduğunu ileri sürer.

²² Pragmatikleşme sadece çok-sözcüklü birimlerden ibaret değildir. Pragmatik unsurlar bir sözcüklük ifadelerden oluşabileceği bağlamla türeyen çeşitli uzunluklarda sözcük birliklerini de içerir (Kopaczky 2013, s. 59).

²³ Diğer türü ya da sonucunun ise dilbilgiselleşme olduğu kabul edilir.

²⁴ Biber et al. (1999: 58) bu kategoriyi tek bir dil bilgisel birim olarak işlev göre sözcük biçimlerinin dizisi olarak tanımlar "a sequence of word forms which functions as a single grammatical unit".

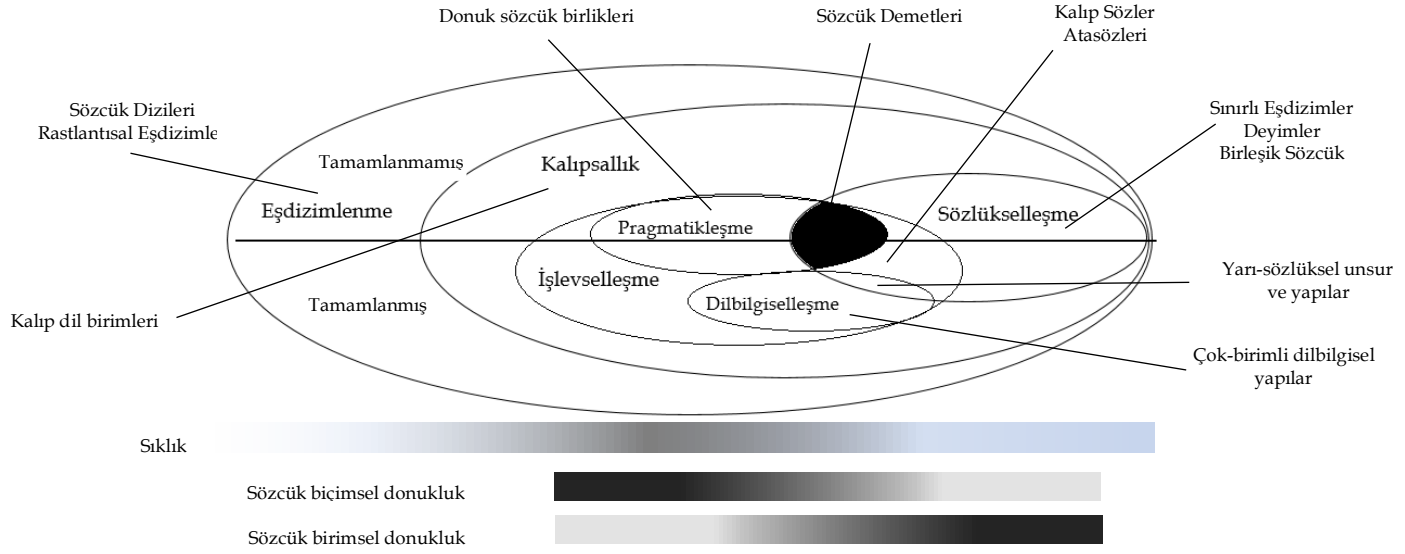
Kopaczky, sözcük demetlerinin Şekil 2’de aşamalı renk şeridiyle gösterilen yatay alanda bulunduğunu ancak sözlükselleşmeye ya da pragmatikleşmeye doğru yayılım ve yönelim gösterdiğini düşünür (2013, s. 62). Şekil 2’deki aşamalı yatay renk şeridi sözcük demetlerinin yapıya, işlevsel ve anlamsal özelliklerinin farklı olduğunu işaretler. Şekil 2’deki aşamalı yatay renk şeridi, bazı sözcük demetlerinin tamamlanmış yapıda ve sözlüksel olduğunu; bazılarının ise donuk sözcük birlikleri şeklinde daha çok söylem örüntüleri oluşturduğunu ve metin türü işaretlediğini gösterir (Kopaczky 2013, s. 62). Pragmatikleşme alanının donukluk açısından farklılaştığı düşünülür. Pragmatikleşmenin sözlükselleşmeye nazaran sözcük biçimsel donukluğu gerektirdiği, sözlükselleştikçe sözcük biçimsel donukluğun azaldığı söylenebilir²⁵.



Şekil 2. Sözcük Demetlerinin Sözlükselleşme ve Pragmatikleşmeyle İlişkisi

Kopaczky (2013, s. 59) derleme-çıkışlı sıklık yaklaşımıyla belirlenebilecek sözcük demetlerinin sözlükselleşme ve pragmatikleşme açısından farklılık gösterdiğini, özellikle bazı sözcük demetlerinin sözlükselleşmiş, bazı sözcük demetlerinin ise daha pragmatikleşmiş, bazılarının da yarı-sözlüksel ve yarı-pragmatik özellikler gösterdiğini ileri sürer. Biz pragmatikleşme ve sözlükselleşme süreçlerinin kesişim noktasında gelişen sözlükselleşmiş, yarı-sözlükselleşmiş ve yarı-pragmatik sözcük demetlerinin pedagojik açıdan önemli ve bilişsel açıdan belirgin sözcük demetleri olduğunu düşünüyoruz. Bu sözcük demetleri ilk aşamada derlem-çıkışlı sıklık yaklaşımıyla derlemden çekilebilir. Ancak derlem-çıkışlı sıklık yaklaşımı beraberinde salt sıklıkla belirlenebilecek farklı sözcük dizilerini de öne çıkaracağı için önemsiz sözcük dizilerinin sözcük demetleriyle karışmasına ve onların da sözcük demeti sayılmasına neden olur. Dilin pedagojik, psiko-dilbilimsel ve leksikolojik açıdan önemli sözcük demetlerinin sıklık bilgisi ve sözcükler arasındaki ilinti ya da eşdizimlilik gücünü hesaplayan kullanışlı bir birliktelik ölçümünün yanı sıra anadili sezgisi ve süzgeciyle ancak belirlenebileceğine inanıyoruz. Bu yaklaşım, sözcük demetlerinin sadece bu özel alanda oluşarlardan ibaret olduğu ve diğer daha pragmatik nitelikte sözcük demetlerinin olmadığı ya da diğer türü dışladığı anlamına gelmez. Bu çalışmada daha çok Türkçenin genel söz varlığında belirli söylem işlevleriyle sözlükselleşmiş, yarı-sözlükselleşmiş, yarı-pragmatik/ yarı-işlevsel sözcük demetleri dikkate alınmıştır.

²⁵ Donukluk, sözcük biçimsel donukluk ve sözcük birimsel donukluk olmak üzere ikiye ayrılabilir. Pragmatikleşmenin sözcük biçimsel donukluğu, sözlükselleşmenin ise sözcük birimsel donukluğu gerektirdiği söylenebilir.



Şekil 3. Kalıpsallığın Süreçleri, Etkenleri ve Kalıp Dil Birimlerinin Oluşumu

Yüksek sıklığın tek başına sözcük demetlerini belirlemeyeceği, sıklığın farklı sözcük dizilerinin özellikleri olduğu söylenebilir. Sıklığın görece etkili olmakla birlikte kalıp dil birimlerini birbirinden ayırt etmek için tek başına yeterli olmadığı söylenebilir. Kalıp dil birimlerinin oluşumunun görece sıklığa bağlı olarak farklı süreçlerden geçtiği söylenebilir. Bu süreçler Şekil 3'te resmedildiği gibi bir aşamalanma arz eder. Kalıp dil birimlerinin oluşmasında sözlükselleşme ve pragmatikleşmenin yanı sıra eşdizimlenme ve dilbilgiselleşme süreçlerinin de etkisi vardır. Her kalıp dil birimi bir önkoşul olarak geniş anlamda bir eşdizim²⁶ ilişkisini gerektirir. Eşdizimlenme sözcük dizilerinin oluşmasına neden olur. Eşdizimlenme yüksek sıklığa bağlı olarak kalıpsallık ya da önüretimlilikle sonuçlanır. Sözcük biçimsel donukluk ve sözcük birimsel donukluk kalıpsallığın sıklığa bağlı gelişen temel iki niteliğidir. Sözcük biçimsel donukluk bir sözcük dizisinin işlevselleşmesine; sözcük birimsel donukluk ise sözlükselleşmesine neden olur. Bir sözcük dizisinin işlevselleşmesi pragmatikleşme ve dilbilgiselleşmeyle sonuçlanır. Zira bir içerik sözcüğünün işlev sözcüğüne dönüşmesi ya da işlevselleşmesinin yollardan biri pragmatikleşme; diğeri ise dilbilgiselleşmedir (Claridge and Arnovick 2010'dan akt. Kopaczky, 2013). Sözcük dizilerinin işlevselleşmesi sıklığın bir sonucudur ancak sıklık farklı işlevsel ve sözlüksel kalıp dil birimlerinin de oluşmasına neden olduğundan tek başına ayırt edici değildir. Sıklık farklı dil bilgisel süreçlerin temel ve ortak bileşeni olarak ortaya çıkar (Şekil 3). Sıklık, daha çok dillerin betimlenmesi ve açıklanması gereken sözcük dizilerine odaklanmalarını sağlar ve önemli öncül bir veri sunar. Pedagojik, leksikolojik ve psiko-dilbilimsel açıdan önemli sözlükselleşmiş sözcük demetleri bu veri içinden uygun bir birliktelik ölçümünden sonra anadili sezgisiyle ancak ayıklanabilir. Bununla birlikte sözlükselleşmiş sözcük demetlerinin çıkartılmasında anadili sezgisinin dışında farklı ölçütler ve eşdizimlilik gücünü hesaplayan farklı birliktelik ölçümleri de kullanılmıştır (bk. Bölüm 5).

²⁶ Eşdizim alanyazında farklı anlaşılan bir terimdir. Geniş anlamda eşdizim iki sözcüğün dil bilgisel yapıdan bağımsız ayırık ya da bitişik ardışık biçimde aynı dizimde kullanımını ifade eder. Dar anlamda eşdizim ise belirli dil bilgisel yapılarda bitişik gerçekleşen, çoğunlukla da sınırlı eşdizimleri karşılar.

5. Sözcük Demeti Çıkarma Yöntemleri ve Çalışmanın Yöntemi

Alanyazında kalıp dil birimlerinin önemi üzerine genel bir anlaşma olmasına karşın kalıp dil birimlerinin ayırıcı nitelikleri, kalıp dil birimlerini belirleme yöntemi, onların nasıl adlandırılacağı ve hangi terimlerle karşılanacağı ve hangi tür çok-sözcüklü birimlerin araştırılmaya değer olduğu hususlarında bir fikir birliği yoktur (Biber, Conrad ve Cortes, 2004, s. 372). Kalıp dil birimlerinin herhangi bir sınıfının belirlenmesinde araştırmacının amaçlarına dönük olarak alanyazında farklı yaklaşımlar kullanılmaktadır, dolayısıyla yöntem açısından şu yaklaşımın kesinlikle daha iyi olduğunu ileri sürmek mümkün değildir. Türkçe alanyazında konuyla ilgili ilk çalışmalarda Biber vd. (1999)'un araştırma geleneği takip edilmiş, çoğunlukla derlem-çıkışlı sıklık yaklaşımı kullanılmıştır. Derlem çıkışlı sıklık yaklaşımında önce dil bilgisel yapıda olması aranmaksızın eşdizimli, bir başka deyişle aynı dizimde gerçekleşen bütün sözcük dizileri özel bilgisayar yazılımları aracılığıyla derlemden çekilir ve sözcük dizilerinin istatistik olarak ham sıklık listesi oluşturulur, sonra ham sıklık listesine belirli sıklık kesme noktası uygulanarak ikinci bir liste oluşturulur. Yazara özgü kullanımlardan kaçınmak için sıklık gösteren üç-birimli sözcük dizileri en az 5 farklı metinde geçmesi koşuluyla kısıtlanır. Sıklık kesme noktasını aşan ve en az 5 farklı metne dağılım gösteren sözcük dizileri sözcük demeti kabul edilir (bk. Biber vd. 1999 ve Aksan&Aksan, 2015a). Bu yaklaşım özellikle üç-birimli sözcük demetlerinin belirlenmesinde güçlü ve yetkin metodolojik uygulama olmasına karşın iki-birimli sözcük demetlerinin belirlenmesinde kullanışlı olmadığı deneyimlenmiştir. İki-birimli sözcük demetleri derlem-çıkışlı sıklık yaklaşımıyla belirlenmek istendiğinde işlevsel ve sözlüksel açıdan önemsiz sayısız iki-birimli sözcük dizisi elde edilir. Bu yöntemle sözcük demeti olmaya aday dikkate alınamayacak kadar fazla sözcük dizisinin elde edilmesi iki-birimli sözcük demetlerinin çoğunlukla ilgi alanı dışında kalmasına, üç-birimli sözcük dizilerinin Türkçe sözcük demetleri için daha ilgili ve dikkate değer adaylar olarak görülmesine neden olmuştur (Aksan&Aksan, 2013; Mersinli, 2015). Bu zorluğun ve engelin önüne geçmek ve derlemden çıkartılan iki-birimli sözcük dizilerinden muhtemel sözcük demetlerini ayıklamak için yüksek sıklık kesme noktası belirlenmiştir (bk. Aksan&Aksan, 2015a-b; Aksan vd., 2016; Aksan, 2016). Buna rağmen aday sözcük demeti listeleri söylem, metin ve dil türlerine özgü işlevsel niteliklerden uzak rastlantısal birliktelikleri ya da sözcük dizilerini içermeye devam etmiştir.

Derlem çıkışlı sıklık yaklaşımı dildeki kalıp dil birimlerin çıkartılmasında yegâne yöntem ve yaklaşım değildir. Derlemden çekilen eşdizim ya da sözcük dizilerini sıklığa göre sıralamak ve bir sıklık kesme noktası uygulamak sözcük demetlerini her zaman ortaya çıkarmaya yetmemiştir. Alanyazında kalıp dil birimlerinin belirlenmesinde salt sayısal sıklık yaklaşımının yanında başta kural-tabanlı ve istatistiksel yöntem olmak üzere birden fazla ölçüt ve yaklaşımın birlikte kullanıldığı hibrit yaklaşımlar denenmiştir. Evert (2004, s. 20) "her iki sözcük de sık kullanılıyor olsa bile her iki sözcüğün birlikteliğinin rastlantısal olabileceğini", bu nedenle birliktelik ölçümleri kullanılmadan ham sıklık değerlerinin yetersiz kaldığını belirtmektedir. Mersinli (2014)²⁷, kalıp dil birimlerin çıkarılmasında eşdizimlerin birliktelik güçlerinin çeşitli istatistik yöntemlerle hesaplanması ve eşdizimlerin buna göre sıralanması; eşdizimler içinde kalıp dil birimleri yukarıya çıkaracak biçimde sıralamak ve sözcüklerin birlikteliğinin rastlantısal olmadığını ölçmek için birliktelik ya da eşdizim ölçümlerinin²⁸ kullanılması gerektiğini aktarır. Dolayısıyla ham sıklığa dayalı gözlem ve belirlemelerin birliktelik ölçümleri gibi ek ölçütler olmadan

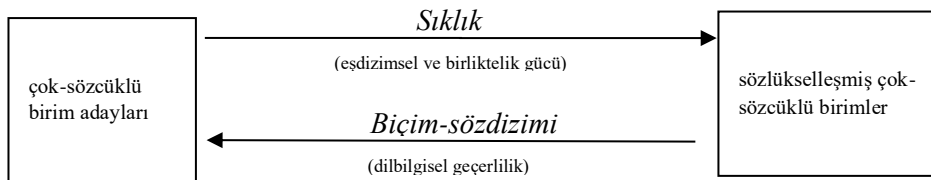
²⁷ 28. Ulusal Dil Bilimi Kurultayı, çevrimiçi:

https://www.researchgate.net/publication/278404832_Turkce%27de_Cok_Sozcuklu_Birim_Cikarimi

²⁸ Birliktelik ya da eşdizim ölçümleri (associative measures) iki ya da daha fazla sözcüğün, bir derlemden tek başlarına ve birlikte kullanımlarına göre aralarındaki ilintinin gücünün matematiksel formüllerle belirlenmesi" şeklinde tanımlanır (Pecina, 2009:138'ten akt. Mersinli, 2015).

yetersiz kaldığı düşünülmüştür. Son dönem Türkçe alanyazında kalıp dil birimlerin belirlenmesinde her ne kadar derlem-çıkışlı olduğu iddia edilse de hibrit yöntemlerin daha çok kullanıldığı söylenebilir²⁹. Derlem çıkışlı sıklık yaklaşımının dezavantajlarından kaçınmak ve olası sözcük demetlerini daha iyi belirlemek için Türkçe alanyazında farklı hibrit yaklaşımlar sınanmıştır. Aksan&Aksan (2015b)'de çok-sözcüklü birimler sayısal sıralı sıklığın yanı sıra birliktelik ölçümleri açısından da değerlendirilmiştir. Bununla birlikte alanyazında hem birliktelik ölçümleri hem de dil bilgisel yapıları dikkate alan kural-temelli yaklaşımı içeren hibrit yöntemlerin salt sayısal sıklık yöntemlerine göre daha başarılı olduğu ileri sürülür. Bir eşdizimi oluşturan sözcükler arasındaki birliktelik gücü, çok-sözcüklü birimlerin çıkarımında gerekli koşul olmasına rağmen yeterli koşul olmadığından Türkçede birliktelik ölçümlerinin de kullanıldığı salt sıralama yöntemiyle ulaşılan sonuçlar da tartışmalı bulunmaktadır (Mersinli, 2014). Sözcük dizileri arasındaki birliktelik gücü ve ölçümleri çok-sözcüklü birimlerin oluşumunu tek başına açıklamaya yeterli bulunmaz (Aksan vd. 2016, s. 80.) Birliktelik ölçümleriyle elde edilen sayısal sıralı listelerin ikinci aşamada kural-temelli olarak dil bilgisel filtrelemelerle ayıklanması gerektiği düşünülür (Mersinli, 2014; Mersinli&Aksan, 2016). Çok-sözcüklü birimlerin diğerlerinden ayıklanmasında Aksan &Aksan (2015a)'da tespit edilen biçim-sözdizimsel bilgi adı da verilen belirli *dil bilgisel örüntü* (colligational patterns) ya da yapısal tiplerin bir ölçüt olarak kullanılabilirliği düşünülür (Mersinli&Aksan, 2016). Aksan vd. (2016)'da biçim-sözdizimsel filtrelemenin çok-sözcüklü birimlerin oluşmasında sıklığa eşlik ettiği düşünüldüğünden belirli yapı tipleri çok-sözcüklü birimlerin belirlenmesinde ve diğerlerinden ayıklanmasında bir ölçüt olarak kullanılmıştır.

Aksan vd. (2016) bir sözcük dizisinin ya da çok-sözcüklü birim adayının sözlükselleşmesinde sıklığa biçim-sözdizimsel yapı tiplerinin eşlik ettiği ve etkili olduğu çift yönlü bir tabiatının olduğunu ileri sürer. Şekil 4'teki betimlemeye göre çok-sözcüklü birim adayları ya da sözcük dizileri sıklığa bağlı olarak önce sözlükselleşir. Sözlükselleşmiş çok-sözcüklü birimlerin geçerliliği ise belirli biçim-sözdizimsel ya da yapısal tiplere uygunluğuna göre belirlenir. Ancak bu yaklaşım iki açıdan kusurludur. Birincisi, çok-sözcüklü birimlerin yapısal tipleri (10 tane) daha önce de ifade edildiği gibi Aksan&Aksan (2015a-b)'de salt sıklığa göre belirlenmiştir. Sözcük demetlerini sözcük dizilerinden ayıklanmasında gözlenen sıklığa göre belirlenen biçim-sözdizimsel yapı tiplerinin filtreleme işleviyle kullanılması yine sıklığın öncelenmesi anlamına gelir. Yüksek sıklığa göre belirlenen yapı tipleri yine yüksek sıklığı olan çok-sözcüklü birim adaylarını ortaya çıkaracaktır. İkinci ise Şekil 4'te görülebileceği gibi yüksek sıklığın doğrudan sözlükselleşmiş çok-sözcüklü birimleri belirlediği düşünülmüştür. Ancak daha önce ifade edildiği gibi yüksek sıklık her zaman kalıpsal ve sözlükselleşmiş çok-sözcüklü birimleri doğurmaz (bk. Bölüm 2). Kimi pragmatik işlev ve kullanımlar sıklığa bağlı sözlükselleşir ancak sıklığa dayalı yapısal tipler sıklığı yüksek bir sözcük dizisinin sözlükselleşip sözlükselleşmediğini ortaya çıkarmaz.



Şekil 4. Çok-sözcüklü Birim Oluşumunun Çift Yönlü Tabiatı (Aksan vd. 2016)

²⁹ Tübitak Projesi: Türkçede Çok-sözcüklü Birimler Projesi, Proje yürütücüsü Prof. Dr. Sera Yeşim Aksan, Proje kodu: 115K135

Her biçim-sözdizimsel yapı tipinin kalıp dil birimlerini diğerlerinden ayırma potansiyeli yoktur. Mersinli ve Aksan (2016)'da 3 kategoride önerdikleri biçim-sözdizimsel ya da yapısal tiplerden özellikle bir kategorinin ayıklama potansiyeli olduğu aktarılır: 1. Kategori: SIFAT+yalın BELİRLEYİCİ+yalın İSİM+yalın durum: *kısa bir süre*; SIFAT+yalın BELİRLEYİCİ+yalın İSİM+yalın durum: *etkin bir şekilde*. Bu kategorinin tamamlanmış dil bilgisel yapılardan olduğu dikkate alındığında sözlükselleşmiş sözcük demetlerinde dil bilgisel tamamlanmışlık koşulu rahatlıkla aranabilir. Ancak bu sözlükselleşmiş sözcük demetlerinin muhakkak tamamlanmış dil bilgisel yapılar olacağı anlamına gelmez. Tamamlanmış dil bilgisel yapısı olmamasına karşın bazı sözcük dizilerinin belirli bir söylemde işlevsel ve pragmatik kullanımlarının donuklaştığı ve pragmatik işlevleriyle sözlükselleştiği söylenebilir. Bu, sözcük demetinin yapısal tipinin sık kullanılmasını da gerektirmez. Örneğin Türkçe Sözlük (2023)'te de kayıtlanan *daha bir* ve *daha da* sözcük demetlerinin tamamlanmış dil bilgisel yapısı olmamasına karşın söylemde görece sıklıkla kullanıldığı işleviyle sözlükselleştiği görülür. Ancak bu sözcük demetlerinin kullanıldığı biçim-sözdizimsel yapı tipi Aksan (2016)'da Tablo 1'de sunulan en sık rastlanan 20 yapısal tip içinde bulunmamaktadır.

Tablo 1: İki sözcüklü birimlerin dil bilgisel işlev dizileri (Aksan, 2016)

Sıra	İşlev dizisi	Gözl. Sıklık	Örnek	Sıra	İşlev dizisi	Gözl. Sıklık	Örnek
1	BEL__AD	456090	bir şey	11	SF__AD+buln	50875	aynı zamanda
2	BLC__SIF	195793	daha fazla	12	AD__BLC	47967	ne de
3	BLC__BLC	165868	daha çok	13	AD__AD+iylk+buln	45520	yıl içinde
4	AD__AD+iylk	144997	söz konusu	14	SIF__SIF	44402	yavaş yavaş
5	BEL__AD+buln	136277	bir şekilde	15	SIF__EYL+arak	43068	genel olarak
6	SIF__AD	108350	son derece	16	BEL__AD+yonl	41055	bir araya
7	BAG__BLC	91250	ya da	17	BEL__AD+cıkm	40501	bu yüzden
8	AD__AD	87030	ne zaman	18	AD__ILG	40148	ne kadar
9	EYL+mak__ILG	52162	olmak üzere	19	BEL__AD+ile	37561	bu nedenle
10	AD__EYL+an	51980	yer alan	20	AD__EYL+gecm	35396	devam etti

Türkçe Sözlük (2023)'te *daha bir* sözcük demeti "farklı", *daha da* sözcük demeti ise "karşılaştırma derecesini vurgular" şeklinde sözlüksel ve işlevsel anlamıyla tanımlanmıştır. Dolayısıyla tamamlanmamış yapıdaki kimi sözcük dizileri söylemde belirli işlevleriyle sözlükselleşebilir. Ancak bu yapıların sözcük demetlerinin oluşumunda ve sözlükselleşmesinde doğrudan ve belirleyici bir etkisi yoktur, sadece pragmatik işlevleriyle sözlükselleşmiş sözcük demetlerinin yapısal özelliklerinden biridir. Zira sözlükselleşme dil bilgisel yapıların özelliklerini sözlükselleşme derecesine bağlı olarak yitirmesine neden olabilmektedir. Şekil 3'te betimlendiği gibi sözcük dizileri iki sözcüğün aynı dizimde kullanılmasını ifade eden eşdizimlenme aşamasında yapısal açıdan tamamlanmış ya da tamamlanmamış olabilir ancak tamamlanmamış yapıların sözlükselleşmiş sözcük demeti olabilmesi öncelikle sıklığa ve donukluğa bağlı olarak işlevselleşmeyi ve belirli işlevde pragmatik bir kullanımı gerektirir.

Bu çalışmada öncelikle Türkçe alanyazında göz ardı edilen iki-sözcüklü sözcük demetleri bir dil türü, metin türü ya da uzmanlık alanından bağımsız genel Türkçe söz varlığı açısından araştırılmıştır. İki sözcüklü sözcük demetleri alanyazında Aksan&Aksan (2015a)'da derlem-çıkışlı sıklık yaklaşımıyla incelenmesine karşın sıklık-çıkışlı yaklaşım Türkçe iki-birimli sözcük demetlerinin derlemden çıkartılmasında kullanışlı olmadığından ve sıklık kesme noktasının iki-birimli sözcük demetleri için her milyonda 200 olarak kullanılmasından gözden kaçmasına neden olmuştur. Her ne kadar birliktelik ölçümleri kullanılsa da yüksek sıklık kesme noktasının belirlendiği ve öncelendiği sıklık-çıkışlı yaklaşımın iki-birimli sözcük demetlerinin derlemden çıkartılmasında kullanışlı bir yöntem olmadığı anlaşılmaktadır. Sıklık-çıkışlı yaklaşım üç birimli sözcük demetlerinin çıkartılmasında daha kullanışlı bir yöntem olmasına karşın dökümü yapılan eşdizimler

arasından iki-birimli sözcük demetlerinin seçilmesinde eşdizimin bileşenleri arasındaki birliktelik gücünü ölçen uygun eşdizim ölçümlerinin tamamlanmış yapılarla ayıklanması ve sezgisel önem ölçütünü kullanmanın daha elverişli ve tutarlı sonuçlar verdiği söylenebilir. Mersinli (2015)'te ifade edildiği gibi iki-birimli çok-sözcüklü birimlerin çıkartılmasında sıklık yanında farklı birliktelik ölçümleri belirli amaçlara yönelik olarak rahatlıkla kullanılabilen ve verimli sonuçlar alınabilmektedir. Bu çalışmada iki-sözcüklü sözcük demetlerinin belirlenmesinde anadili sezgisiyle ayıklama aşamasından önce derlemin sunduğu birliktelik ölçümlerinden *log-likelihood* "log-olabilirliği" değeri kullanılmıştır.

Kalıp dil birimlerinin derlemden çıkarımında *karşılıklı* (symmetric) ve *yönelimli* (directional) olarak bilenen iki temel yaklaşım geçerlidir (Evert, 2004, s. 27). Kalıp dil birimlerinin derlemden çıkarımında her iki yaklaşımda etkin biçimde kullanılabilir (Mersinli&Demirhan, 2012). Karşılıklı yaklaşımda derlemdaki farklı çok sözcüklü birimler sıklık-çıkışlı yaklaşımla iki, üç ya da dört-birimli olarak sıklık frekansına göre sıralanır. Yönelimli yaklaşımda ise önceden belirlenen bir "çıkış sözcüğünden" hareket edilir. Çıkış sözcüğünün derlemda hangi diğer sözcüklerle çok daha sık birlikte kullanıldığını belirlemek için sağ ve sol bağlam aralığı 1'den 5'e kadar kısıtlanır ve çıkış sözcüğünün birliktelik dökümü elde edilir. Sağ ve sol bağlam kısıtlılığıyla elde edilen birliktelik dökümü birliktelik ölçümleriyle ilinti ya da eşdizimlilik gücünü vermek üzere sıralanır. Bu çalışmada iki-sözcüklü sözcük demetlerinin belirlenmesinde yönelimli, dolayısıyla derlemtabanlı bir yaklaşım³⁰ benimsenmiştir. Bu kapsamda 50 milyon sözcük kapasiteli Türkçe Ulusal Derlemi³¹ (TUD) kullanılmıştır. İki sözcüklü sözcük demetleri birkaç aşamada derlemden çıkartılmıştır. Birinci aşamada sözcük biçimlerinin salt sıklığı dikkate alınmış; çıkış sözcüğü olarak kullanılmak üzere *Türkçe Ulusal Derlemi Sözcük Sıklığı Listelerindeki Sayısal Sıralı Sözcük Listesinden* sıklığı en yüksek 200 sözcük biçimi seçilmiştir. Bir sözcük biçiminin derlemdaki sıklığı Stubbs'ın (2002) da ifade ettiği gibi sadece kendinden değil, aksine gerçekte bileşeni olduğu çok-sözcüklü birimlerin sıklığından kaynaklanır. Dil bilgisel ya da işlevsel sözcükleri doğal olarak dilde içerik sözcüklerine nispeten çok daha sık kullanılır. Bunun işlev sözcükleriyle teşkil edilen ve dilde sıklıkla tekrarlanan kalıp dil birimlerinden ileri geldiği söylenebilir. Dolayısıyla sözcük sıklığı listesi kalıp dil birimlerinin belirlenmesinde bir "çıkış sözcüğü listesi" olarak kullanılmıştır. Tablo 2'de çıkış sözcüğü listesinden *bunun* sözcük biçimi derlemden sorgulanır. Sorgulamanın sonunda Tablo 3'te *bunun* sözcük biçiminin sağ ve sol 5 sözcüklük mesafede diğer sözcüklerle birliktelik kullanımları alınır. Sonra birliktelik ölçümlerini ve ilinti güçlerini hesaplamak için Tablo 3'teki aynı ekrandan eşdizimlilik bilgisi istenir ve Tablo 4 elde edilir. Sonra eşdizimlilik sonuçları listesi sözcük pencere aralığı sağ ve sol +/-1 olacak şekilde Tablo 4'teki aynı ekrandan tekrar sorgulanır. Sonra eşdizim sonuçları ekranından sözcük biçiminin ilintilik gücünü hesaplamak için birliktelik ölçümlerinden "log-likelihood" değeri Tablo 5'te yüksek sıklıktan düşük sıklığa olacak şekilde sıralanır. Bu veri, iki sözcüğün üç-sözcüklük bir dizimde ne kadar sık kullanıldığını sunmakta ve güçlü dizimsel birlikteliği olup olmadığını değerlendirme imkânı sağlamaktadır. Zira Mersinli (2015) log-likelihood ölçümünün iki-birimli çok-sözcüklü birimlerin çıkartılmasında en kullanışlı eşdizimlilik ve birliktelik ölçümü olduğunu belirtir.

³⁰ Biber (2009)'da ifade edildiği gibi kalıp dil birimlerinin belirlenmesinde derlem tabanlı ve derlem çıkışlı yaklaşımların birbirine üstünlüğü yoktur. Araştırmacılar amaçları doğrultusunda belirli bir yaklaşımı seçebilir. Derlem tabanlı ve derlem çıkışlı yaklaşımların ayırıcı metodolojik özelliklikleri için bk. Biber, 2009.

³¹ TUD Sürüm 3.0 50 milyon sözcükten oluşan, 24 yıllık bir dönemi (1990- 2013) kapsayan, günümüz Türkçesinin çok sayıda farklı alan ve türlerden yazılı ve sözlü örneklerini içeren, geniş kapsamlı, dengeli ve temsil yeterliliğine sahip, genel amaçlı bir referans derlemdir. <https://www.tnc.org.tr/tr/> (18.04.2023)

Sözcük biçimlerinin aynı dizimde iki sözcüklük bağlamdaki çok sık tekrar eden ardışık sözcük dizileri sezgisel önem ölçütüyle değerlendirilmiş ve bir sözlükselleşmiş sözcük demeti olup olmadığı belirlenmiştir. İki-sözcüklü sözcük dizileri eğer üç-sözcüklü sözcük demetlerinin bileşeni ise bunlar üç-sözcüklü sözcük demeti olarak tanımlanmıştır. Sözcük demeti olup olmadığı değerlendirilen sözcük dizilerinin biri içerik diğeri işlev sözcüğünden oluşması koşulu gözetilmiştir. Türkçede sözcükleri işlev ve içerik sözcüğü olarak tasnif etmek güç olmakla birlikte edatların, bağlaçların, ünlemlerin, belirtilik işaretleyen belirtme sıfatlarının, zamirlerin işlev sözcüklerini teşkil ettiği söylenebilir. Ancak işlev sözcükleri Türkçede İngilizceye kıyasla farklı özellikler gösterir. Türkçede birçok içerik sözcüğü belirli sözcük biçimleriyle *yarı-sözlüksel unsurlar* (semi-lexical items) olarak işlev sözcüğü görevinde kullanılır. Örneğin dilde sıklıkla kullanılan *sonuç olarak, açık bir şekilde, bu durumda, büyük olasılıkla, başka bir deyişle* sözcük demetlerinde *olarak, şekilde, durumda, olasılıkla, deyişle* sözcük biçimlerinin ne tümüyle içerik ne de tümüyle işlev sözcüğü olduğu söylenebilir. Bu sözcük biçimleri pragmatikleşme sürecinin etkisiyle bir bileşen olarak sıklıkla kullanıldığı kalıp dil birimlerinde sözlüksel potansiyelini yitirmeye ve dilbilgiselleşmeye; daha sonra yarı-sözlüksel bir unsur olarak işlev sözcüğü gibi kullanılmaya başlar. Özellikle Mardela'nın (2012) da ifade ettiği gibi edatların da bir yarı-sözlüksel kategori olduğu söylenebilir. Türkçede *de, dahi, ise* vs. sona gelen bazı kadim edatlar dışında edat olarak tanımlanan, özellikle *her şeyden önce* ve *bir zaman sonra* sözcük demetlerindeki *önce* ve *sonra* son çekim edatlarının çoğunlukla yarı-sözlüksel kategori ve unsurların tipik örneği olduğu söylenebilir. Dolayısıyla Türkçede sözcük demetlerinin İngilizceden farklı olarak çoğunlukla bir içerik sözcüğü ile yarı-sözlüksel kategori ya da unsurla oluştuğu söylenebilir. Türkçe sözcük demetleri işlev+işlev sözcükleriyle de oluşmasına karşın işlev+içerik ve içerik+yarı-sözlüksel unsurlar ile oluşan sözcük demetleri genellikle bilişsel ve psiko-dilbilimsel açıdan daha belirgin ve yaygındırlar ve dil bilgisel açıdan da çoğunlukla tamamlanmışlık sergilerler.

Tablo 2. *bunun* sözcük biçiminin arandığı basit sorgu ekranı

Tablo 3. *bunun* sözcük biçiminin bağlamlı dizini

Standart Arama: **bunun** Arandı

Metin Sayısı: **3998** 5445 Metin içerisinde arandı

Gözlenen sıklık: **39694** 50678199 Sözcük içerisinde arandı

Normalleştirilmiş Sıklık: **783.26** -5.5 Sözcük aralığında arandı

Bağlı Dizini Eşdizimlilik

Dizilim Sonuçları

Bağlam İçerisinde Anahar Sözcük Görünümü Tümce Görünümü Sıralı Görünümü

Show 50 entries Copy CSV TSV Excel Search:

Tür	Künye	Sol	Sağ
W-UDRCIA-2639-244	"geniatçı" İBDAC terör örgütü ve	bunun	karşısında mağdur olarak laik halk
W-PE3RCBA-9249-625	şuydu "Bati ittifakına girmek istiyorsanız,	bunun	tek yolu var Kore'ye asker
S-SEABXW-0353-398	şunlar şunlar öldü mü öldü.	Bunun	hepsi sana demiş tek tek
S-ADABCo-0230-82	şube müdürlüğünde çalışan Kenan Temur'du.	Bunun	dışında bize ulaşan bilgilere göre
S-SEABXO-0342-415	şu, müzisyen, kemancı hep	bunun	yeğenleri bilmem neleri, amca çocukları
W-MG03AAA-1036-1224	şu anki gerçeği bilmek değil,	bunun	nereye doğru gittiğini bilmektir. Toplumcu

Tablo 4. *bunun* sözcük biçiminin eşdizimlilik sonuçları

Eşdizimlilik Sonuçları

Sayfada 100 Kayıt Göster Copy CSV Excel

#	Sözcük	Metin içerisinde arandı	Beklenen Eşdizimlilik Sıklığı	Gözlenen Eşdizimlilik Değeri	Birlikte oldukları doküman sayısı	Log-likelihood Değeri	M1 Değeri	M3 Değeri	T-Score Değeri	Dice Coefficient Değeri	Log Dice Coefficient Değeri
812952	İçin	284706	223.0033	7952	2466	43219.0516	5.1562	31.0704	88.673	0.049	9.6497
353653	bir	1332821	1043.9663	11789	2854	38900.5096	3.4936	30.5366	105.587	0.0171	8.1331
1059304	ya	1138904	892.1384	9264	2665	29151.8319	3.3918	29.7776	94.5929	0.0159	8.0242
172897	İsabet	33815	26.4865	2800	1184	20986.0348	6.724	29.6264	52.8797	0.0762	10.2856
772868	da	407420	319.1222	5115	2137	19456.5121	4.0026	28.6436	70.9443	0.0229	8.5502
1129640	da	398862	312.4189	4308	2002	16183.3433	3.8509	28.1275	66.6132	0.0206	8.3959
494389	nodet	8444	6.614	1537	901	14043.5251	7.8604	29.0322	39.1981	0.0639	10.031
1019771	olduğunu	63231	49.5273	2022	1161	11217.6499	5.3514	27.3145	44.9105	0.0393	9.3303
663587	yanında	16601	13.0032	1228	689	9808.5066	6.6742	27.4243	36.4298	0.0472	9.5943
1341998	olarak	219349	171.8107	2546	1345	9150.5566	3.8893	26.5174	50.2395	0.0197	8.3312
887530	dışında	19091	14.9535	1214	748	8390.7961	6.3431	26.8342	34.8294	0.0413	9.4024
160453	sonucu	13677	10.7129	1828	626	7452.8569	6.5844	26.5956	32.0538	0.0385	9.3018
1288030	sonucunda	7254	5.6819	895	540	7411.907	7.2994	26.9109	29.9123	0.0381	9.2869
535676	çık	227939	178.5391	2089	1200	6414.859	3.5276	25.5431	45.1721	0.0154	7.9778
672794	an	139945	109.6155	1488	1085	6156.019	3.9448	25.387	40.9718	0.0188	8.2664
708903	ama	149398	117.0198	1658	961	5785.6371	3.8246	25.2151	40.5985	0.0175	8.1665
1337715	var	92760	72.6567	1333	803	5293.6563	4.1974	24.9584	36.4434	0.0201	8.3653
326192	yani	10495	8.2205	742	456	5280.726	6.4961	25.5666	27.234	0.0296	8.9202

Tablo 5. bunun sözcük biçiminin log-likelihood birliktelik hesaplaması

#	Sözcük	Metin içerisinde arandı	Beklenen Eşizimlilik Sıklığı	Gözlenen Eşizimlilik Değeri	Birlikte oldukları doküman sayısı	Log-likelihood Değeri	MI Değeri	MI3 Değeri	T-Score Değeri	Dice Coefficient Değeri	Log Dice Coefficient Değeri
812952	için	284706	223.0033	6096	2176	29631.0935	4.7727	29.92	77.6383	0.0376	9.2662
172897	üzermine	33815	26.4865	2607	1104	19141.7558	6.621	29.3173	51.0247	0.0709	10.1825
663587	yanında	16601	13.0032	1265	647	9214.3447	6.6041	27.214	35.5552	0.0449	9.5242
494389	nedeni	8444	6.614	933	619	7509.9366	7.1402	26.8717	30.54	0.0388	9.3108
887530	üzerinde	19091	14.9535	1104	684	7413.7502	6.2061	26.4232	33.214	0.0376	9.2653
1288030	sonucunda	7254	5.6819	869	532	7141.7112	7.2568	26.7833	29.4746	0.037	9.2444
160453	sonucu	13677	10.7129	757	471	5009.5641	6.1429	25.2712	27.5062	0.0284	8.8604
326192	yanı	10495	8.2205	661	395	4546.5302	6.3293	25.0663	25.7046	0.0263	8.7534
1059304	ve	1138994	892.1384	3334	1662	4069.1046	1.9019	25.308	56.4431	0.0057	6.5343
772868	da	407420	319.1222	1809	1172	3359.9144	2.503	24.145	42.1904	0.0081	7.0507
302017	antamı	4746	3.7174	384	249	2648.051	6.6135	23.6291	19.077	0.0164	8.0682
1018629	yanına	2081	1.63	381	177	2590.4141	7.5288	23.996	17.3486	0.0144	7.8832
834389	sebebi	3032	2.3749	224	253	2580.2185	7.092	23.7717	17.9989	0.0152	7.957
54117	yerine	28140	22.0414	473	384	2023.4882	4.4296	22.2132	21.7824	0.014	7.842
672794	an	139945	109.6155	634	599	1953.1342	2.9276	22.3354	28.7993	0.0093	7.2492
1043297	için	1380	1.0809	221	132	1950.0815	7.6756	23.2514	14.8657	0.0108	7.4619
1157440	ancak	69486	54.4267	569	437	1632.3057	3.386	21.6906	23.821	0.0104	7.4159
708903	ama	149398	117.0198	747	557	1522.2516	2.6744	21.7643	27.2507	0.0079	7.0162
1232890	nedenini	1521	1.1914	160	150	1268.2484	7.0693	21.7132	12.6487	0.0078	6.991
348293	tarafı	1533	1.2008	154	135	1225.6782	7.0215	21.5923	12.4896	0.0076	6.9541

6. Türkçede İki ve Üç-Sözcüklü Sözcük demetleri

Türkçe iki-birimli sözcük demetleri yukarıda betimlenen yöntemle derlemden çıkarılmış; pedagojik, psiko-dil bilimsel ve sözlüksel açıdan önemli iki-sözcüklü sözcük demetleri listesi oluşturulmuştur. Listede sözcük demetlerinin derlemde gözlenen sıklığı, sözcük demetlerinin alanyazındaki durumu ve Türkçe Sözlük (2023)'te yer alıp almadığı kodlanmıştır. Sözcük demetleri söz konusu ölçü ve ölçümler kullanılarak işlevsel ve sözlüksel olmak üzere iki temel sınıfa ayrılmıştır. Sözcük demetlerinin sözcük dizilerinden ayıklanmasında belirgin bilişsel bir önemi olmadığı durumlarda Salazar'ın (2014) kullandığı ayıklama/ dışlama kriterleri ve dil bilgisel tamamlanmışlık ölçütü kullanılmıştır. Belirlenen sözcük demetlerinin Türkçe Sözlük (2023)'te ve alanyazında tespit edilip edilmediği karşılaştırılarak kuram ve uygulanan yöntemin kullanılabilirliği sınanmıştır.

Sözcük Demeti	Sıklık Değeri	Aksan & Aksan (2015a; 2015b), Aksan vd. (2016), Aksan vd. (2017), Mersinli (2015), Çalışkan (2014)	Türkçe Sözlükte Bulunma Durumu	İşlevsel Sözcük Demetleri	Sözlüksel Sözcük Demetleri
Adam gibi	638	-	+		+
Akşama kadar	705	-	-	+	
Alışılmışın dışında	88	-	-		+
Anlaşılacağı gibi	462	-	-	+	
Anlaşılacağı üzere	416	-	-	+	
Anlaşıldığı üzere	95	-	-	+	
Aracılığı ile	1126	-	-	+	
Asıl neden	63	-	-	+	

Aslında belki	61	-	-	+	
Aslında evet	80	-	-	+	
Aşağıdaki gibi	451	-	-	+	
Aynen öyle	274	-	-		+
Aynı anda	2726	-	-		+
Aynı biçimde	503	-	-		+
Aynı şekilde	2950	-	-		+
Aynı zamanda	11522	+	+		+
Az bir şey	45	-	-		+
Az biraz	301	-	-		+
Az buçuk	116	-	+		+
Az çok	5083	-	+		+
Az önce	1824	-	-		+
Az sayıda	861	-	-		+
Az sonra	1973	-	+		+
Bahsi geçen	189	-	-	+	
Bal gibi	250	-	-		+
Başka bir deyimle	120	-	-	+	
Başka bir deyişle	1437	+	-	+	
Başka bir ifadeyle	268	-	-	+	
Başka bir şey	3275	+	-	+	
Başka türlü	1239	-	-		+
Belirtildiği gibi	710	-	-	+	
Belirtildiği üzere	197	-	-	+	
Belirttiğim gibi	225	-	-	+	
Belki de	12653	+	+	+	
Belli belirsiz	820	-	+		+
Benzer şekilde	820	-	-	+	
Bıktım artık	81	-	-	+	
Bildiği gibi	241	-	-	+	
Bildiğin gibi	141	-	-	+	
Bildiğiniz gibi	378	-	-	+	
Bilindiği gibi	1500	-	-	+	
Bilindiği üzere	315	-	-	+	
Bin bir	2243	-	+		+
Bin kere	141	-	+		+
Bin kez	146	-	+		+
Binlerce kez	119	-	-		+
Binlerce yıl	378	-	-		+
Bir an	8592	-	+		+
Bir an önce	2444	+	+		+
Bir anda	5814	+	+		+
Bir ara	3823	-	+		+
Bir bakıma	1874	-	+		+
Bir başka	35873	+	-		+
Bir daha	48449	-	+		+
Bir dahaki	273	-	-		+
Bir de	65604	-	+		+
Bir diğeri	818	-	+		+
Bir gün	18681	-	-		+

Bir güzel	9293	-	+		+
Bir husus daha	13	-	-	+	
Bir kere daha	733	+	-		+
Bir kez daha	4044	+	-		+
Bir şey	31327	-	-	+	
Bir süre	15721	+	+		+
Bir süre daha	517	+	-		+
Bir süre önce	65331	-	-		+
Bir süre sonra	4509	+	-		+
Bir sürü	3382	-	+		+
Bir sürü şey	136	-	-		+
Bir şekilde	15677	-	-		+
Bir şey daha	411	-	-		+
Bir tek	13805	-	+		+
Bir türlü	6984	-	+		+
Bir yandan	8262	-	+		+
Biraz daha	6918	-	-		+
Biraz evvel	122	-	-		+
Biraz fazla	698	-	-		+
Biraz olsun	499	-	-		+
Biraz önce	1363	-	-		+
Biraz sonra	2751	-	-		+
Birazcık olsun	27	-	-		+
Birçok defa	55	-	-		+
Birçok kez	276	-	-		+
Birçok kimse	46	-	-	+	
Birçok şey	249	-	-	+	
Birden fazla	1478	-	-	+	
Birinci olarak	547	-	-	+	
Birkaç defa	257	-	-		+
Birkaç gün	2092	-	-		+
Birkaç kere	200	-	-		+
Birkaç kez	1263	-	-		+
Birkaç yıl	1105	-	-		+
Bitti artık	77	-	-		+
Bomba gibi	205	-	+		+
Böyle böyle	1974	-	+		+
Bu açıdan	2048	-	+	+	
Bu akşam	1119	-	-		+
Bu anlamda	2128	-	-		+
Bu arada	10356	-	+	+	
Bu bağlamda	2314	-	-		+
Bu bakımdan	1998	-	+	+	
Bu defa	1415	-	-		+
Bu denli	1339	-	+		+
Bu durum	6878	-	-	+	
Bu durumda	6086	-	-		+
Bu doğrultuda	685	-	-		+
Bu gece	2286	-	-		+
Bu hususta	614	-	-		+
Bu kadar	13685	-	+		+
Bu kez	8380	-	-		+
Bu nedenle	14143	-	-		+
Bu noktada	2673	-	-		+

Bu sayede	1395	-	-		+
Bu sebeple	1624	-	+		+
Bu sefer	2822	-	+		+
Bu şekilde	4630	-	-		+
Bu tür	10238	-	-		+
Bu uğurda	310	-	-		+
Bu yıl	5509	-	-		+
Bu yüzden	9930	-	+		+
Bugün yarın	91	-	+		+
Bugüne kadar	3446	-	-		+
Buna bağlı olarak	937	+	-	+	
Buna benzer	921	-	-		+
Buna dayanarak	53	-	-	+	
Buna ilaveten	75	-	-	+	
Buna ilişkin	141	-	-	+	
Buna göre	3955	-	-	+	
Buna karşı	598	-	-	+	
Buna karşılık	3160	-	+	+	
Buna karşın	1222	-	-	+	
Buna rağmen	1511	-	-	+	
Buna yönelik	69	-	-	+	
Bundan başka	507	-	-		+
Bundan böyle	1244	-	+	+	
Bundan dolayı	989	-	+	+	
Bundan evvel	14	-	-	+	
Bundan evvelki	15	-	-	+	
Bundan önce	183	-	-	+	
Bundan önceki	175	-	-	+	
Bundan ötürü	153	-	-		+
Bundan sonra	3795	-	-		+
Bundan sonraki	918	-	-	+	
Bundan sonrası	217	-	-	+	
Bunların dışında	373	-	-	+	
Bunun aksine	77	-	-	+	
Bunun anlamı	384	-	-	+	
Bunun dışında	1104	-	-	+	
Bunun için	6096	-	-		+
Bunun karşılığında	146	-	-	+	
Bunun karşısında	89	-	-	+	
Bunun nedeni	933	-	-		+
Bunun neticesinde	76	-	-	+	
Bunun ötesinde	157	-	-	+	
Bunun sebebi	324	-	-		+
Bunun sonucunda	869	-	-	+	
Bunun tersine	61	-	-	+	
Bunun üstüne	81	-	-		+
Bunun üzerine	2607	-	-		+

Bunun yanında	1265	-	-		+
Bunun yanı sıra	301	-	-		+
Bunun yerine	475	-	-	+	
Bununla birlikte	2518	-	+	+	
Buz gibi	722	-	+		+
Bütün bütün	1728	-	+		+
Bütün çıplaklığıyla	57	-	+		+
Bütün gece	616	-	-		+
Büyük bir çoğunluğu	205	-	-		+
Büyük bir ihtimalle	188	+	-		+
Büyük bir olasılıkla	349	+	-		+
Büyük çapta	201	-	-		+
Büyük çoğunluğu	664	-	-		+
Büyük ihtimalle	565	-	-		+
Büyük küçük	1289	-	-		+
Büyük olasılıkla	814	-	-		+
Büyük oranda	723	-	-		+
Büyük ölçüde	3464	-	-		+
Cin gibi	117	-	+		+
Çeşitli açılardan	85	-	-	+	
Çeşitli biçimlerde	148	-	-	+	
Çeşitli nedenlerle	320	-	-	+	
Çeşitli sebeplerle	67	-	-	+	
Çeşitli şekillerde	219	-	-	+	
Çeşitli yollarla	117	-	-	+	
Çığ gibi	133	-	-		+
Çocuk gibi	1138	-	+		+
Çoğu kez	1697	-	+		+
Çoğu zaman	2442	-	-		+
Çok çeşitli	764	-	-		+
Çok eski	721	-	-	+	
Çok farklı	2689	-	-	+	
Çok fazla	4225	-	-	+	
Çok geçmeden	852	-	-	+	
Çok güzel	5577	-	-	+	
Çok iyi	9838	-	-	+	
Çok yakın	1185	-	-	+	
Daha az	5878	-	-	+	
Daha çok	26871	+	-		+
Daha da	18548	-	+		+

Daha da önemlisi	534	-	-	+	
Daha doğrusu	3190	-	-		+
Daha fazla	12929	+	-	+	
Daha ilk günden	113	-	-	+	
Daha iyi	9202	-	-	+	
Daha önce	11647	+	-		+
Daha önceki	1691	-	-	+	
Daha önemlisi	965	-	-	+	
Daha sonra	15594	+	-		+
Daha sonra da	1224	+	-	+	
Daha sonraki	1598	-	-	+	
Daha sonraları	723	-	-		+
Daha sonrasında	104	-	-		+
Daha şimdiden	320	-	-		+
Daha yakın	852	-	-	+	
Daha ziyade	775	-	-		+
Dediğim gibi	697	-	-	+	
Dediğin gibi	207	-	-	+	
Değindiğimiz gibi	150	-	-	+	
Değindiği gibi	192	-	-	+	
Deli gibi	525	-	+		+
Deliler gibi	230	-	-		+
Dev gibi	186	-	+		+
Diğer bir deyimle	107	-	-	+	
Diğer bir deyişle	938	+	-	+	
Diğer bir ifadeyle	327	-	-	+	
Diğer taraftan	2595	-	-		+
Diğer yandan	2893	-	-		+
Diye bağırdı	1095	-	-	+	
Diye düşündü	1176	-	-	+	
Diye fısıldadı	249	-	-	+	
Diye konuştu	2975	-	-	+	
Diye mırıldandı	447	-	-	+	
Diye seslendi	317	-	-	+	
Diye sordu	3294	-	-	+	
Diye söylendi	391	-	-	+	
Doğal olarak	3129	-	-		+
Doğru bir şekilde	159	-	-	+	
Doğru şekilde	456	-	-		+
Dolayısı ile	622	-	-	+	
Dün gece	952	-	-		+
Düne kadar	169	-	-		+
Dünya üzerinde	351	-	-	+	
Eğer böyleyse	28	-	-	+	
Eğer ki	374	-	-		+

Eğer mümkünse	31	-	-	+	
Eğer öyleyse	91	-	-		+
Ek olarak	1474	-	-		+
En az	7604	-	-		+
En azından	5437	-	+		+
En büyük	12289	-	-	+	
En çok	11250	-	-		+
En eski	932	-	-		+
En fazla	3728	-	-		+
En güzel	4132	-	-		+
En iyi	7902	-	-		+
En iyisi	1396	-	-	+	
En önemli	11272	+	-		+
En önemlisi	1620	-	-	+	
En son	4113	-	-		+
En sonunda	1058	-	-		+
En yakın	1892	-	-		+
Eninde sonunda	457	-	+		+
Ertesi gün	3020	-	-		+
Esas olarak	1293	-	-		+
Eski çağlarda	56	-	-		+
Eski dost	71	-	-		+
Eski günler	50	-	-		+
Eski moda	46	-	-		+
Eski usul	47	-	-		+
Eski zamanlarda	79	-	-		+
Eskisi gibi	935	-	-		+
Eskisi kadar	315	-	-		+
Faltaşı gibi	83	-	-		+
Fena değil	191	-	-	+	
Fırtına gibi	105	-	+		+
Gece boyu	69	-	-		+
Gece boyunca	306	-	-		+
Geçen gece	107	-	-		+
Geçen gün	2066	-	-		+
Geçen hafta	1324	-	-		+
Geçen sefer	78	-	-		+
Geçen sene	559	-	-		+
Geçen sezon	197	-	-		+
Genç yaşında	121	-	-	+	
Genç yaşta	354	-	-		+
Genel anlamda	442	-	-		+
Genel çerçeve	40	-	-		+
Genel geçer	136	-	-		+
Genel görünüm	34	-	-		+
Genel hatlarıyla	117	-	-	+	
Genel olarak	5467	-	-		+
Gereğinden fazla	384	-	-	+	
Gereksiz yere	270	-	+		+
Gerektiği gibi	420	-	-		+

Geri dönülmez	65	-	-	+	
Gibi gibi	62	-	-	+	
Gibi görünse de	277	-	-	+	
Gördüğünüz gibi	290	-	-	+	
Görüldüğü gibi	2705	-	-	+	
Görüldüğü üzere	435	-	-	+	
Görüleceği gibi	389	-	-	+	
Görüleceği üzere	204	-	-	+	
Güzel güzel	1852	-	+		+
Haddinden fazla	98	-	+		+
Haksız yere	246	-	+		+
Harita üzerinde	81	-	-	+	
Hatta bazen	322	-	-	+	
Hatta belki de	106	-	-	+	
Hatta mümkünse	36	-	-	+	
Hatta neredeyse	34	-	-	+	
Hemen ardından	880	-	-	+	
Hemen hemen	7140	-	+		+
Hemen hepsi	590	-	-		+
Hemen herkes	361	-	-		+
Hemen sonra	1626	-	-		+
Hep aynı	1673	-	-		+
Hep beraber	688	-	+		+
Hep bir ağızdan	307	-	+		+
Hep birden	141	-	+		+
Hep birlikte	1769	-	-		+
Her açıdan	233	-	-		+
Her an	3047	-	-		+
Her bir	21896	-	+		+
Her biri	2584	-	+		+
Her biri için	88	-	-		+
Her birinin	868	-	-		+
Her daim	312	-	+		+
Her defasında	647	-	-		+
Her durumda	845	-	-		+
Her fırsatta	739	-	-		+
Her geçen gün	1158	+	-		+
Her gün	7356	-	+		+
Her ikisi de	815	-	-	+	
Her ikisinin de	277	-	-	+	
Her koşulda	232	-	-		+
Her nasılsa	210	-	+		+
Her ne hikmetse	70	-	+		+
Her ne kadar	3035	+	-		+
Her nedense	300	-	+		+

Her neyse	627	-	+		+
Her seferinde	1126	-	-		+
Her şey	9408	-	+		+
Her şeyden çok	105	-	-		+
Her şeyden evvel	102	-	-		+
Her şeyden önce	1782	+	-		+
Her şeyden önemlisi	75	-	-	+	
Her şeyden öte	14	-	-		+
Her şeye karşın	178	-	-		+
Her şeye rağmen	794	-	-		+
Her yandan	119	-	-		+
Her yıl	3038	-	-		+
Her yönüyle	267	-	-		+
Her zaman	12075	-	+		+
Her zaman için	228	-	-		+
Her zamanki gibi	1003	+	-	+	
Her zamankinden	494	-	-	+	
Heykel gibi	108	-	+	+	
Hiç de	6180	-	+		+
Hiç değilse	1277	-	+		+
Hiç gereği yokken	18	-	-	+	
Hiç hesapta yokken	13	-	-		+
Hiç kimse	2225	-	+		+
Hiç kuşkusuz	903	-	-		+
Hiç olmazsa	1385	-	-		+
Hiç şüphesiz	401	-	-		+
Hiç yokken	124	-	-		+
Hiç yoktan	105	-	+		+
Hiçbir koşulda	32	-	-		+
Hiçbir neden yokken	14	-	-	+	
Hiçbir sebep yokken	12	-	-	+	
Hiçbir surette	37	-	-		+
Hiçbir şekilde	757	-	-		+
Hiçbir şey	6883	-	-		+
Hiçbir zaman	4772	-	-		+
İkisi de	2982	-	-		+
İkisinin de	795	-	-		+
İle beraber	842	-	-	+	
İle birlikte	6337	-	-	+	
İle ilgili	11440	-	-	+	
İle ilişkili	342	-	-	+	
İlk akla gelen	184	-	-		+
İlk bakışta	907	-	+		+
İlk başta	430	-	-		+

İlk defa	3304	-	-		+
İlk günden	443	-	-		+
İlk kez	7647	-	-		+
İlk olarak	4040	-	-		+
İlk önce	1589	-	+		+
İnsan gibi	1633	-	+		+
İşte böyle	2137	-	-		+
İşte şimdi	718	-	-	+	
Kaç kez	690	-	-		+
Kadın kadına	141	-	+		+
Kağıt üzerinde	160	-	-		+
Karşı karşıya	3716	-	+		+
Kedi gibi	185	-	+		+
Kendi adına	353	-	+		+
Kendi başına	1207	-	+		+
Kendi başlarına	321	-	-		+
Kendi halinde	326	-	+		+
Kendi hesabına	237	-	-		+
Kendi kendine	2738	-	+		+
Kendine has	491	-	+		+
Kendine mahsus	57	-	+		+
Kendine özgü	1581	-	+		+
Kısa bir süre için	176	+	-		+
Kısa bir süre önce	294	+	-		+
Kısa bir süre sonra	1055	+	-		+
Kısa sürede	2783	-	-		+
Kısa süreli	692	-	-		+
Kısa süreliğine	61	-	-		+
Kısa zamanda	1830	-	-		+
Kimbilir belki de	131	-	-	+	
Kimi zaman	3709	-	+		+
Köpek gibi	147	-	+		+
Kuş gibi	284	-	+		+
Küçük küçük	1278	-	-		+
Leş gibi	134	-	+		+
Madem öyle	123	-	-		+
Mis gibi	394	-	+		+
Mümkün değil	2103	-	-		+
Mümkün olduğu kadar	540	-	-	+	
Nasıl da	4543	-	-	+	
Nasıl ki	1770	-	+		+
Nasıl olsa	1546	-	+		+
Ne biçim	661	-	+		+
Ne çok	4561	-	-		+
Ne de olsa	1156	+	+	+	
Ne denli	1718	-	+		+
Ne diye	5191	-	+		+
Ne gibi	3586	-	+		+

Ne güzel	3783	-	-		+
Ne kadar	24250	+	+		+
Ne olursa olsun	2220	+	+	+	
Ne tür	987	-	-		+
Ne var ki	3380	+	+	+	
Ne yani	279	-	+		+
Ne yazık ki	3070	+	+	+	
Ne zaman	8140	-	-		+
Ne zaman ki	188	-	-	+	
Ne zamandır	530	-	-		+
Neden sonra	777	-	+		+
Neredeyse hiç	227	-	-	+	
Nesi var	144	-	+	+	
O kadar	14406	-	+		+
O zaman	15510	+	-		+
O zamana kadar	1056	-	-		+
Olacak diye	444	-	-	+	
Olduğu gibi	12005	-	-	+	
Olduğu halde	1456	-	-	+	
Olduğu için	7643	-	-	+	
Olsa bile	1285	-	-	+	
Olsa da	6357	-	-	+	
Olsa olsa	1758	-	+		+
Olsun olmasın	159	-	-	+	
Ona buna	115	-	+		+
Ona göre	2681	-	-		+
Ondan sonra	3042	-	-		+
Onun dışında	371		-	+	
Onun için	5441	-	-		+
Onun yerine	714	-	-	+	
Orada burada	107	-	+		+
Ortaya çıkan	5871	-	-		+
Ortaya koyan	577	-	-	+	
Öbür yandan	183	-	-		+
Önceki gece	441	-	-		+
Önceki gün	1441	-	-		+
Önceki yıl	617	-	-		+
Önemli derecede	359	-	-		+
Önemli oranda	270	-	-		+
Önemli ölçüde	1931	-	-		+
Öte yandan	6108	-	+		+
Öyle böyle	576	-	-		+
Öyle görünüyor ki	111	-	-	+	
Öyle ki	1953	-	-		+
Öyle olsa	131	-	-	+	
Öyle olsa da	7	-	-	+	
Öyle olsaydı	104	-	-	+	
Özellikle de	2443	-	-	+	
Pat diye	278	-	+		+
Pek az	742	-	-		+
Pek çoğu	391	-	+		+

Pek çok	9495	+	+		+
Pek çok şey	234	-	-	+	
Pek çokları	35	-	-		+
Pek de	3349	-	-		+
Pek fazla	704	-	-		+
Pek yakında	87	-	-		+
Sabaha kadar	897	-	+		+
Sadece ve sadece	296	-	-	+	
Sanıldığı gibi	166	-	-	+	
Sırası değil	78	-	-		+
Son derece	7343	+	+		+
Son günlerde	910	-	-		+
Son kez	975	-	-		+
Son olarak	3540	-	-		+
Son yıllarda	3444	-	-		+
Son zamanlarda	1612	-	-		+
Sonra da	14842	-	-	+	
Sonsuza kadar	446	-	-		+
Sonuç olarak	2410	-	-		+
Sonuna kadar	2800	-	-		+
Sorun değil	210	-	-	+	
Söylediği gibi	462	-	-	+	
Söylediğim gibi	277	-	-	+	
Söylendiği gibi	122	-	-	+	
Söz gelimi	295	-	+		+
Söz gelişi	98	-	+		+
Söz konusu	13636	-	+		+
Söz sahibi	617	-	+		+
Sözüm ona	212	-	+		+
Su gibi	1155	-	+		+
Şart değil	124	-	-	+	
Şıp diye	97	-	+		+
Şimdi bir de	93	-	-	+	
Şimdi de	4358	-	-	+	
Şimdiye kadar	2263	-	-		+
Şöyle bir	3589	-	+	+	
Şöyle böyle	604	-	+		+
Şu açıdan	40	-	+		+
Şu an	2484	-	-		+
Şu anda	5389	-	-		+
Şu anki	277	-	-		+
Şu bakımdan	33	-	+	+	
Şu sıralar	291	-	-		+
Şu şekilde	1435	-	-		+
Şu var ki	92760	-	-	+	
Tam aksine	291	-	-		+
Tam anlamıyla	1470	-	+		+
Tam manasıyla	103	-	+		+
Tam olarak	4070	-	-		+
Tam tersine	1164	-	-		+
Taş gibi	263	-	+		+
Tek şey	1130	-	-		+

Temel neden	79	-	-		+
Tüm açıklığıyla	24	-	-		+
Tüm bunlar	467	-	-	+	
Tüm bunlara karşın	27	-	-	+	
Tüm bunlara rağmen	58	-	-	+	
Tüm çıplaklığıyla	42	-	-		+
Tüm dünya	460	-	-		+
Tüm dünyada	713	-	-	+	
Tüm insanlık	88	-	-		+
Tüm zamanların	67	-	-	+	
Uzun süre	4052	-	-		+
Uzun süredir	1023	-	-		+
Uzun zamandır	925	-	-		+
Üzerinde durulması gereken	240	-	-	+	
Varsın olsun	57	-	-		+
Ve yine	1529	-	-	+	
Yakın gelecekte	467	-	-		+
Yakın zamanda	508	-	-		+
Yani aslında	112	-	-	+	
Yani ne bileyim	35	-	-	+	
Yarın bir gün	43	-	-		+
Yarın öbür gün	105	-	-		+
Yeni yeni	901	-	+		+
Yerine göre	253	-	-		+
Yerli yerine	245	-	+		+
Yeter artık	269	-	-	+	
Yeteri kadar	750	-	-	+	
Yetmezmiş gibi	146	-	-	+	
Yetti artık	44	-	-		+
Yıllar önce	1382	-	-		+
Yine de	8991	-	-	+	
Yine mi	191	-	-	+	
Yok artık	532	-	-	+	
Yok yani	713	-	-	+	
Yol boyunca	542	-	-		+
Yol üzerinde	138	-	-		+
Yolun sonunda	67	-	-		+
Yüzlerce kez	84	-	-		+
Zaman içinde	2920	-	-		+
Zerre kadar	171	-	-		+

Sonuç

Türkçenin söz varlığında sanılan aksine iki-birimli sözcük demetlerinin sayısının oldukça fazla olduğu görülmektedir. Sözcük demetleri eşdizimlerden farklı ölçüt ve ölçümlerle betimlenmiş ve görülmüştür ki sözcük demetleri eşdizimlerden sözcük türü açısından farklıdır. Sözcük demetleri eşdizimlerin aksine sözcüde sıklıkla edat, bağlaç,

belirtici ve zarf olarak kullanıldığı görülür. Bu çalışma sonunda derlemden çıkarılan sözcük demetlerinin birçoğunun hem alanyazında hem de Türkçe Sözlük'te yer almadığı görülmektedir. Bu çalışmada 567 sözcük demeti belirlenmiş ve sadece 136'sının Türkçe Sözlük'te yer aldığı görülmüştür. Belirlenen sözcük demetlerinin büyük bir bölümünün sözlükselleşmiş olduğu dikkate alındığında çalışmanın sonuçlarının Türkçenin söz varlığına önemli katkı sağladığı söylenebilir. Bu çalışmada sözcük demetleri kuramsal açıdan önce tanımlanmış ve derlemden yöntem bilimsel açıdan çıkarımı denenmiş ve kullanışlı bir yaklaşım olduğu görülmüştür. Ayrıca sözcük demetlerinin listede verilen sıklık bilgisi dikkate alındığında sözcük demetlerinin oluşumu ve sözlükselleşme arasında doğru orantılı bir ilişki olmadığı gözlenmemiştir. Devam niteliğindeki çalışmalarda alanyazındaki işlevsel tasnif denemelerinden hareketle iki-birimli sözcük demetlerinin söylem işlevleri ve işlevlerine göre sınıflandırılmasına çalışılacaktır. Sözcük demetlerinin çoğunlukla söylem işaretleyicisi işlevinde kullanıldığı görülmektedir. Devam niteliğinde çalışmada iki-birimli sözcük demetlerinin hangi söylem işlevleri olduğu ve söylem işaretleyicilerinin hangi sınıfına girdiği tartışılacaktır.

Kaynakça

- Aksan, M. & Aksan, Y. (2013). Multi-word Units and Pragmatic Functions in Genre Specification. *Paper presented at 13th IPrA Conference 08-13 September 2013*. New Delhi, India.
- Aksan, M. & Aksan, Y. (2015a). Multi-word Expressions in Genre Specification. *Mersin University Journal of Linguistics and Literature*, 12.
- Aksan, M. & Aksan, Y. (2015b). Multi-word in Imaginative and Informative Domains. *Ankara papers in Turkish and Turkic Linguistics*. Zeyrek, Ç. Sağın-Şimşek, U. Ataş & J. Rehbein (Eds.). Wiesbaden: Harrassowitz Verlag.
- Aksan, Y., Mersinli, Ü. & Altunay, S. (2015). Colligational analysis of Turkish multi-word units. Paper presented at CCS-2015, *Corpus-Based Word Frequency: Methods and Applications*. 19-20 February 2015. Mersin University. Türkiye.
- Aksan, Y. vd. (2016). Colligational Patterns of Turkish Multi-Word Units. *Mersin Üniversitesi Dil ve Edebiyat Dergisi*, 13(2).
- Aksan, Y. S. (2016). *Türkçe'de Çok Sözcüklü Birimler: Derlem-Temelli Yöntem ile Çıkarılmaları, Sayısal Dağılımlarının Hesaplanması, Yapı ve İşlev Özelliklerinin Saptanması, Sözlüğünün Oluşturulması*. 115K135 Numaralı Tübitak Projesi.
- Aksan, Y., Mersinli, Ü. ve Altunay, S. (2017). Türkçe Çok Kelimeli Birimlerin İlişkisel Kalıpları. *Mersin Üniversitesi Dil ve Edebiyat Dergisi*, 13(2), 71-108.
- Aksan, Y., Özel, S.A., Yılmaz, H., & Demirhan, U. (2016). The Turkish National Corpus (TNC): Comparing the Architectures of v1 and v2. *TurCling 2016*, 32-37.
- Çevrimiçi: <https://search.trdizin.gov.tr/tr/yayin/detay/615654/> [Erişim tarihi: 14.07.2023].
- Altenberg, B. (1998). On the Phraseology of Spoken English: The Evidence of Recurrent Word Combinations. *Phraseology: Theory, Analysis and Applications*. Oxford: OUP.
- Ayabakan, M. (2020). *Türkçe Sözlükte Eş Dizimli Ögelerin Sunumu ve Görünümleri*. Ankara: Türk Dil Kurumu Yay.
- Biber, D. (1996). *University Language A corpus- Based Study of Spoken and Written Registers*. Amsterdam: John Benjamins Publishing Company.

- Biber, D. (2006). *University Language: A Corpus-Based Study of Spoken and Written Registers*. Amsterdam/Philadelphia: John Benjamins.
- Biber, D. (2009). A Corpus-driven Approach to Formulaic Language in English. *International Journal of Corpus Linguistics*, 14(3), 275–311.
- Biber, D., & Barbieri, F. (2007). Lexical Bundles in University Spoken and Written Registers. *English for Specific Purposes*, 26, 263–286.
- Biber, D., Conrad, S., & Cortes, V. (2004). “If you look at ...”: Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics*, 25(3), 371–405.
- Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Londra: Longman.
- Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 Academic Writing. *Language Learning & Technology*, 14(2), 30–49.
- Conrad, S., & Biber, D. (2005). The Frequency and Use of Lexical Bundles in Conversation and Academic Prose. *Lexicographica*, 20, 56–71.
- Cortes, V. (2004). Lexical Bundles in Published and Student Disciplinary Writing: Examples from History and Biology. *English for Specific Purposes*, 23(3), 397–423.
- Cortes, V. (2008). A Comparative Analysis of Lexical Bundles in Academic History Writing in English and Spanish. *Corpora*, 3(1), 43–58.
- Cortes, V. (2013). The purpose of This Study is to: Connecting Lexical Bundles and Moves in Research Article Introductions. *Journal of English for Academic Purposes*, 12(1), 33–43.
- Çalışkan, N. (2014a). *Yüksek Sıklıktaki Adlar: Zaman Sözcükbiçimi Üzerine Derlem Temelli Bir İnceleme (Eş dizimlilikler, Biçimbirim Dizileri, Dilbilgisel Örüntüler, Semantik Prozodi)*. Ankara: Pegem Akademi Yayınları.
- Çalışkan, N. (2014b). Çok Sözcüklü Birimler: Yazılı Türkçede Sözcüksel Setler veya ‘N’ Sayıdaki Ardeşık Sözcük Dizilerinin Yapı Özellikleri. *Türkçe Derlem Dilbilim Uygulamaları*. Adana: Karahan Kitabevi.
- Çalışkan, Nihal (2019a). Yabancı Dil Sınıflarında Tümleç Yan Cümlelerinin Öğretimi: Eş dizimlilik ve Semantik Eğilimler Temelli Bir Yaklaşım. *Uluslararası Türkçe Edebiyat Kültür Eğitim (Teke) Dergisi*, 8(3), 1709–1741.
- Çalışkan, Nihal (2019b). Yükleme Konumundaki Sıfatlar ile Fiilimsi Tamlayıcıları Arasındaki Eş Dizimlilik İlişkileri. *Dil Araştırmaları*, 24, 75–88.
- Çalışkan, N. (2019c). diye Sözcüğü Üzerine Derlem Temelli Bir Betimleme. *Türkbilig*, 38, 163–176.
- Doğan, N. (2016). *Türkiye Türkçesinde Fiillerin Eşdizimleri*. Ankara: Kitapevi
- Doğan, N. (2023). *Türkçede Eş Dizimlilik Türkiye Türkçesinde Fiillerin Eş Dizimleri*. Ankara: Türk Dil Kurumu Yayınları.
- Doğan, Nuh (2020). Türkçede Sözcük ve Dil Bilgisi İlişkisi: Dil Bilgisel Eş Dizim Kalıpları. *Anemon Muş Alparslan Üniversitesi Sosyal Bilimler Dergisi*, 8(6) 1945–1957.
- Durrant, P. (2016). To what extent is the Academic Vocabulary List relevant to university student writing? *English for Specific Purposes*, 43.

- Durrant, Philip. (2013). Formulaicity in an Agglutinating Language: The case of Turkish. *Corpus Linguistics and Linguistic Theory*, 9(1), 1-38.
- Ellis, N. C. (2008). Temporal Cognition and Temporal Language the First and Second Times Around. *Language Learning*, 58(Supplement 1), 115-121.
- Gezegin-Bal, B. (2019). Lexical Bundles in Published Research Articles: A corpus-based Study. *Journal of Language and Linguistic Studies*, 15(2), 520-534.
- Gökdayı, H. (2011). *Türkçede Kalıp Sözcükler*. İstanbul: Kriter Yayınları.
- Howarth, P. A. (1996). *Phraseology in English Academic Writing*. Tübingen: De Gruyter.
- Howarth, P. A. (1996). *Phraseology in English Academic Writing*. Tübingen: De Gruyter.
- Kopaczky, J. (2013). *The Legal Language of Scottish Burghs*. New York: Oxford University Press.
- Kumova-Metin, S. & Karaoğlan B., (2010). Collocation Extraction in Turkish texts using Statistical Methods. *Paper Presented at 7th International Conference on Natural Language Processing (LNCS-ISI) IceTAL 2010, Reykjavik, Iceland*.
- Manning Cd, S. H. (1999). *Foundations of Statistical Natural Language Processing*, MIT Press, Cambridge, MA
- Mardale, A. (2012). Prepositions as a semilexical category. *Bucharest Working Papers in Linguistics, University of Bucharest*, 13(2), 35-50.
- Mersinli, Ü. (2015). Associative Measures and Multi-word Unit Extraction in Turkish. *Dil ve Edebiyat Dergisi*, 12(1), 43-61.
- Mersinli, Ü. & Demirhan, U. (2012). Çok Sözcüklü Kullanımlar ve İlköğretim Türkçe Ders Kitapları. *Türkçe Öğretiminde Güncel Çalışmalar*, Mersin: Mersin Üniversitesi, 113-122.
- Moon, Rosamund (1998). *Fixed Expressions and Idioms in English: A Corpus-Based Approach*, Oxford: Clarendon Press.
- Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- Oflazer, K., Çetinoğlu, Ö. & Say, B. (2004). Integrating morphology with multi-word expression processing in Turkish. *Proceedings of the Workshop on Multiword Expressions: Integrating Processing (MWE '04)*. Association for Computational Linguistics, (pp.64-71). Stroudsburg, PA, USA.
- Özkan, B. (2020). *Türkiye Türkçesinde Ön Adlarla Adların Birliktelik Kullanımı ve Eş Dizimi Sözlüğü*. Ankara: Türk Dil Kurumu Yay.
- Özkan, B. (2007). *Türkiye Türkçesinde Belirteçlerin Fiillerle Birliktelik Kullanımları ve Eş Dizimliliği*. Doktora Tezi. Adana: Çukurova Üniversitesi.
- Özel, S. A., Bektaş, Y., & Yılmaz, H., (2016). Formulaicity in Turkish: Evidence from the Turkish National Corpus. *Mersin Üniversitesi Dil ve Edebiyat Dergisi*, 13(2), 1-33.
- Pawley, A. & Syder F. H. (1983). Two Puzzles for Linguistic Theory: Nativelike Selection and Nativelike Fluency. *Language and Communication*, London and New York: Longman, 191-226.
- Perkins, M. R. (1999). Productivity and Formulaicity in Language Development. *Issues in Normal and Disordered Child Language: From Phonology to Narrative*, 51-67.

- Salazar, D. (2014). *Lexical Bundles in Native and Non-native Scientific Writing*. Amsterdam: John Benjamins Publishing Company.
- Schmitt, N. and Carter, R. (2004). *Formulaic Sequences in Action: An Introduction. Formulaic Sequences: Acquisition, Processing and Use*. Amsterdam: John Benjamin Publishing Company.
- Simpson-Vlach, R., & Ellis, N. C. (2010). An Academic Formulas list: New Methods in Phraseology Research. *Applied Linguistics*, 31(4), 487-512.
- Sinclair, John (1991). *Corpus, Concordance, Collocation*, Oxford: Oxford University Press.
- Stubbs, M. (2001). *Words and Phrases, Corpus Studies of Lexical Semantics*. Oxford: Oxford University Press.
- Türkçe Sözlük* (2023) [Kurul]. Ankara: Türk Dil Kurumu Yayınları
- Wood, David (2010). *Perspectives on formulaic language: Acquisition and Communication*. New York/London: Continuum.
- Wood, D. (2015). *Fundamentals of Formulaic Language: An Introduction*. London and New York, NY: Bloomsbury.
- Wray, A. (2002). *Formulaic Language and The Lexicon*. Cambridge: CUP.
- Wray, A., & Perkins, R. M. (2000). The Functions of Formulaic Language: An Integrated Model. *Language and Communication*, 20(1).