

-sA bile: Türkçede Sözcüklerarası Örüntüler*

Mustafa Aksan, Umut Ufuk Demirhan
Mersin Üniversitesi
mustaksan@gmail.com umutufuk@gmail.com

Giriş

Birlikte kullanımları rastlantı ötesi bir sıklıkla gerçekleşen, bildirişim ortamlarında kullanıcıların düzenli olarak birleştirdikleri dil birimleri, bu birimlerin türleri ve sayısal dağılımları, derlemler üzerinden yapılan çalışmalarda ayrıntıları ile belirlenmektedir. Birlikte kullanım ile oluşan *kalıp anlatımlar* (İng. fixed expressions, formulaic language) dil incelemelerinin farklı düzlemlerinde "sözcük" kavramının yeniden yorumlanmasına neden olmuştur. *Sözcükbirim* (İng. the lexical item) kapsamlı ve büyük dil verisinde yazılımlar ile saptanan, düzenli ve sık örüntülerin yorumlanması ile yeniden tanımlanan, sayısal ölçümlere göre doğal dil kullanımının yarısından fazlasını oluşturan yapıların etkisiyle önerilen bir terim olarak araştırmalarda yer almaya başlamıştır.

Bu çalışma, Türkçe için olası kalıp anlatım biçimlerini saptamayı amaçlamaktadır. Daha dar anlamda, *Türkçe Ulusal Derlemi* (TUD) verisinde görülen, sıklıkla birlikte kullanılan dil birimlerini çıkarmak, derlem-çıkışlı yaklaşımla sayısal görünümünü belirlemek, Türkçe sözcük yapısının kalıp anlatım örüntülerinin oluşmasında etkisinin betimlenmesine dönük yeni veri sunmaktır. Türkçede sözcüklerin oluşturdukları *çok sözcüklü birimler* (İng. multiword units) ile biçimbirimlerin oluşturdukları çok ekli¹ birimler (İng. multimorpheme units) daha önceki çalışmalara konu olmuş, temel diziler ve sıklıkları, türleri ve kullanım ortamları üzerine çalışmalar yapılmıştır. Örüntü oluşu sözcüklerarası (İng. interlexical) gerçekleşen sık diziler ise henüz ayrıntısı ile incelenmemiş, yapılan kimi gözlemler ise tekil yapılar ile sınırlı kalmıştır.

Bu aşamada, çalışma örüntü çıkarma yöntemini tanıtacak, elde edilen ilk sonuçları sınırlı sayıda örnek ile paylaşacaktır. Verinin yorumu, örüntülerin biçim ve anlam çözümlemeleri ile işlevsel değerleri daha ayrıntılı incelemeleri gerektirmektedir.

2. Diziler ve Örüntüler

Birbirlerinden farklı yapısal özellikleri ve anlam değerleri olan, farklı birimler ile kurulan, çok birimli dil öğelerini inceleyen çalışmalar oldukça eskidir. Son dönemlerde ise, derlem verisindeki düzenli ve sık örüntüleri saptayan özel amaçlı yazılımlar, bu tür yapıların bulunmasında, sayısal değerlerinin hesaplanmasında araştırmacılara hem büyük kolaylık sağlamış, hem de daha önceleri görülmeyen yeni diziler ve bu tür dizilerin henüz betimlenmemiş kullanım bağlamlarını sunmuştur. İngiliz dilbilim geleneğinin derlem çalışmalarında yansımaları olarak değerlendirilebilecek bir yaklaşımla Sinclair (1998), derlem verisinde gözlenen düzenli ve sık örüntüleri tanımlamada geleneksel sözcük kavramının yeterli olmadığını belirtir, bunun yanı sıra *sözcükbirim* (İng. the lexical item) kavramını önerir. Betimleyici bir kavram olarak sözcükbirim, kendine özgü bir

¹ Yeni bir terim önerisinde bulunmadan, yalnızca anlatım kolaylığı sağlaması ve yer kazandırması amacıyla bu çalışmada "ek" sözcüğünü kullanacağız.

içyapısı olan, sözcüklerden ve öbeklerden oluşabilen, kendisini çevreleyen *eşbaşlam* (İng. co-text) birimleri ile birlikte daha büyük yapıların parçası olarak değerlendirilir.

Sözcük kavramının içerdiği kısıtlamalardan bağımsız olarak tanımlanan bu dil örüntüsü, sözcüklerarası kurulan ilişkinin gerçekte sözcüğün yapı ve anlam özelliklerinin betimleyicisi olduğunu, anlamın tekil sözcük ile bağlı olmadığını, aksine, birden çok sözcüğün varlığını gerektirdiğini, anlam yorumunun ise sözcüklerarası birlikte seçimlerin kurduğu örüntü ile oluştuğunu savlamaktadır.

Doğal dil kullanımı verisi ile çalışan araştırmacıların kalıp anlatım genel başlığı altında tanımladıkları, birden çok birim ile oluşturulan, genellikle bitişik olan, karşılıklı bağıntılılığın parçaların toplamından farklı bir anlatım ya da metin etkisi kurduğu dil kullanımını örnekleyen birimleri içermektedir (Wray, 2006). Farklı iç yapıları, türleri ve kullanım bağlamları olan bu tür birimler için çok sayıda terim önerilmiştir. Kapsayıcı bir terim olarak *kalıp* söz terimini kullanan çalışmalar da bulunmaktadır.

3. Örüntü Gerçekleşmeleri

Kavram tartışmalarını dışarıda bırakarak, veri yönelimli bir değerlendirme ile, Türkçe için olası üç örüntü seçeneğinden söz edebiliriz:

1. Çok Sözcüklü Birimler (İng. multiword units)

[S1 S2 S3 ... Sn] (*bir süre sonra, ne de olsa, ...*)

2. Çok Ekli Birimler (İng. multimorpheme units)

[E1 E2 E3 ... En] (*...ebil-ir-lik, laş-tır-ma-sı, ...*)

3. Sözcüklerarası Örüntüler (İng. interlexical patterns)

[...+E1+2... En S1 (S2 ... Sn)] (*... ebil-mek için*)

Burada verilen örüntülerin kapsayıcı olduğunu öne sürmek doğru bir yaklaşım olamamakla birlikte, Türkçenin yapısal özelliklerinin kurabileceği örüntü türlerinin genel bir görünümünü vermesi açısından bu aşamada yeterli olacaktır. Yukarıda (1) ve (2) de yer alan örüntüler ayrı bir çalışma konusu olarak burada incelenmeyecektir.

4. Veri ve Yöntem

Çalışmanın temel verisi *Türkçe Ulusal Derlemi* (Aksan vd., 2012) metinleridir. Derlem, 1990-2013 yıllarını kapsayan, 4,978 yazılı ve 434 sözlü metni, 9 konu alanını ve 39 dilsel türü (Örn. bloglar, forumlar, elektronik postalar, reklamlar, üniversite öncesi denemeler) içeren metin örneklerinden oluşmaktadır. Çalışmanın verisi TUD'un 49,983,288 sözcüklük yazılı metinlerinden alınmıştır.

Text-NSP (Pedersen vd., 2011) kullanılarak, 49 milyon sözcüklük TUD yazılı metinlerinden, ikili dizilerin gözlenen sıklıkları hesaplanmıştır (Örn. yap-mak için). İkili birimin ilk ögesi TUD-açımlayıcı aracılığıyla işaretlenmiş (Örn. *güldürmek* → gül,VB+caus+nzm), ilk ögenin tüm çekim ekleri ve ikinci öge temel alınarak, tekrarlayan çekim ekleri ve sözcük

örüntülerinin gözlenen sıklıkları hesaplanmıştır (Örn. *caus+nzm_k için*). İlk ögenin yalnızca bitiş eki temel alınarak, bitiş eki + sözcük dizilerinin gözlenen sıklıkları hesaplanarak (Örn. *nzm_k için*) oluşturulan ek dizilerinin içerisinde çoklu ekler seçilmiş ve diziler güncellenmiştir (Örn. *past+3s__* ve *caus+nzm_k için*).

5. Bulgular

TUD yazılı metinleri verisinde saptanan sözcüklerarası örüntüler, gözlenen sıklık değerleri ile listelendiğinde ortaya çıkan en sık ilk 10 örüntü aşağıda verilmiştir.

Tablo 1. En sık ilk 10 sözcüklerarası örüntü

	Teksözcük (2)	Sonek (1)	Örnek	Ek (1) + Teksözcük (2) Sıklığı
1	<i>için</i>	<i>nzm_k</i>	<i>etmek için</i>	40,872
2	<i>gibi</i>	<i>pcdk+p3s</i>	<i>olduğu gibi</i>	18,020
3	<i>sonra</i>	<i>pcdk+abl</i>	<i>olduktan sonra</i>	15,825
4	<i>üzere</i>	<i>nzm_k</i>	<i>olmak üzere</i>	15,338
5	<i>için</i>	<i>gen</i>	<i>bunun için</i>	15,336
6	<i>göre</i>	<i>dat</i>	<i>buna göre</i>	15,223
7	<i>kadar</i>	<i>dat</i>	<i>bugüne kadar</i>	14,688
8	<i>sonra</i>	<i>abl</i>	<i>bundan sonra</i>	12,222
9	<i>için</i>	<i>pcdk+p3s</i>	<i>olduğu için</i>	12,069
10	<i>olarak</i>	<i>acc</i>	<i>sonucu olarak</i>	11,841

En sık ilk 10 örüntünün sözcüksel dağılım değerleri ise Tablo 2’de listelenmiştir.

Tablo 2. En sık ilk 10 örüntü toplam değerleri

	Teksözcük	Sıklık
1	<i>için</i>	136,796
2	<i>gibi</i>	56,494
3	<i>sonra</i>	38,679
4	<i>göre</i>	37,850
5	<i>kadar</i>	35,912
6	<i>olarak</i>	20,343
7	<i>üzere</i>	19,698
	Toplam	345,772

Teksözcüklere ilişkin en sık 40 örüntü ise Tablo 3’teki gibidir.

Tablo 3. En sık kullanılan 40 örüntü

	Teksözcük	Teksözcük +Sonek+ Teksözcük	Sıklık
1	<i>için</i>	<i>onlar için</i>	8734
2	<i>için</i>	<i>gerçekleştirmek için</i>	7973
3	<i>doğru</i>	<i>geriye doğru</i>	6583
4	<i>için</i>	<i>yılı için</i>	6398
5	<i>kadar</i>	<i>sonuna kadar</i>	6014
6	<i>için</i>	<i>edebilmek için</i>	5943
7	<i>göre</i>	<i>onlara göre</i>	5934
8	<i>gibi</i>	<i>görüldüğü gibi</i>	5778
9	<i>karşı</i>	<i>sabaha karşı</i>	5728
10	<i>arasında</i>	<i>bunlar arasında</i>	5301

Teksözcük	Teksözcük +Sonek+ Teksözcük	Sıklık
11 için	olması_ için	5249
12 olarak	parçası_ olarak	4865
13 için	benim_ için	4557
14 göre	durumuna_ göre	4438
15 göre	verilerine_ göre	4427
16 sahip	yapıya_ sahip	4309
17 arasında	yılları_ arasında	4285
18 için	sağlanması_ için	4253
19 gereken	yapılması_ gereken	4001
20 beri	günden_ beri	3964
21 önce	başlamadan_ önce	3731
22 gibi	senin_ gibi	3608
23 önce	şeyden_ önce	3593
24 birlikte	bununla_ birlikte	3583
25 rağmen	olmasına_ rağmen	3380
26 altında	adı_ altında	3346
27 üzerine	bunun_ üzerine	3130
28 gibi	der_ gibi	3000
29 içinde	süreci_ içinde	2985
30 kadar	olduğu_ kadar	2974
31 göre	maddesine_ göre	2961
32 içinde	evin_ içinde	2857
33 tarafından	insanlar_ tarafından	2838
34 için	kendisi_ için	2694
35 ilgili	konuyla_ ilgili	2628
36 için	olmadığı_ için	2610
37 rağmen	buna_ rağmen	2483
38 gibi	onlar_ gibi	2417
39 birlikte	olmakla_ birlikte	2343
40 sonra	ayrıldıktan_ sonra	2296

Tartışmalar ve ayrıntılı yorumlar için yeterli yer olmaması nedeniyle bu aşamada sunulmayacaktır. Ancak, yüzeysel bir gözlem ile, verinin neredeyse tamamının ilgeç yapılarını içerdiği, bir bölümünün ise belirteç yantümcesi kurucuları oldukları görülmektedir. Yukarıda Tablo (2) de verilen en sık bulunan dizilerin tamlayanlarının sıklıkları ise aşağıdaki tablolarda verilmiştir.

Tablo 4. İçin tamlayanları

	Biçimbirim	Açıklama	Örnek	Sıklık
1	-mAk	nzmK	etmek_ için	58,535
2	-mA	nzma+p3s	olması_ için	15,638
3	-dİk	pcdk+p3s	olduğu_ için	20,153
4	-An	pcaN+pl	olanlar_ için	844
5	-AcAk	pcck+p3s	olacağı_ için	263
6	diğer	p1p	ülkemiz_ için	41,363
Toplam				136,796

Tablo 5. Gibi tamlayanları

	Biçimbirim	Açıklama	Örnek	Sıklık
1	-dlk	pcdk+p3s	<i>olduğu gibi</i>	32,600
2	-mA	nzma+p3s	<i>olması gibi</i>	595
3	-AcAk	pasv+pcck+p3s	<i>görüleceği gibi</i>	1,705
4	-mlş	perf+3s	<i>varmış gibi</i>	3,808
5	-mAk	nzmK	<i>olmak gibi</i>	785
6	diğer	p3s	<i>eskisi gibi</i>	17,018
	Toplam			56,494

Tablo 6. Sonra tamlayanları

	Biçimbirim	Açıklama	Örnek	Sıklık
1	-dlk	pcdk+abl	<i>olduktan sonra</i>	20,158
2	-mA	nzma+p3s+abl	<i>dağılmasından sonra</i>	869
3	diğer	abl	<i>bundan sonra</i>	17,652
	Toplam			38,679

Tablo 7. Göre tamlayanları

	Biçimbirim	Açıklama	Örnek	Sıklık
1	-dlk	pcdk+p2s+dat	<i>olduğuna göre</i>	3,087
2	-mA	nzma+dat	<i>açıklamaya göre</i>	1,542
3	-An	nzma+dat	<i>açıklamaya göre</i>	118
4	-AcAk	va1+neg+pcck+p2s+dat	<i>olamayacağına göre</i>	52
5	diğer	dat	<i>buna göre</i>	33,051
	Toplam			37,850

Tablo 8. Kadar tamlayanları

	Biçimbirim	Açıklama	Örnek	Sıklık
1	-dlk	pcdk+p3s	<i>olduğu kadar</i>	3,709
2	-AcAk	futr+3s	<i>edecek kadar</i>	1,830
3	-IncA	avnce+dat	<i>gelinceye kadar</i>	1,974
4	diğer	dat	<i>bugüne kadar</i>	28,399
	Toplam			35,912

Tablo 9. Olarak tamlayanları

	Biçimbirim	Açıklama	Örnek	Sıklık
1	ad biçimleri	p3s	<i>parçası olarak</i>	20,343
	Toplam			20,343

Tablo 10. *Üzere* tamlayanları

	Biçimbirim	Açıklama	Örnek	Sıklık
1	-mAk	nzmK	<i>olmak üzere</i>	17,728
2	-dIk	pcdk+p3s	<i>olduğu üzere</i>	1,352
3	-AcAk	pasv+pck+p3s	<i>görüleceği üzere</i>	618
	Toplam			19,698

Tablolarda görüleceği gibi, yapısal özelliğın doğrudan yansıması olarak ilgeç tamlayanlarının çekimli adlar ya da adlaştırma tümceleridir.

Kullanım sıklığı ve örüntü kurma verisi, yukarıda verilen diğer yapıların toplamından fazla olan için yapılarının ayrıntılı dökümü ise aşağıda verilmiştir. Adlaştırıcı bitiş ekleri ve dizileri, yine yer kısıtlaması nedeniyle, burada yalnızca bir örüntü ile örneklenecektir.

Tablo 11. *İçin* ek dizileri (İng. n-morphgram) -mAk

	Biçimbirim sayısı	Sonek (Dizisi)	Örnek	Sıklık
	1 biçimbirim	nzmK	<i>etmek için</i>	40,872
		caus+nzmK	<i>gerçekleştirmek için</i>	7,973
		va1+nzmK	<i>edebilmek için</i>	5,943
		neg+nzmK	<i>olmamak için</i>	1,724
		pasv+nzmK	<i>korunmak için</i>	667
		refl+nzmK	<i>görünmek için</i>	278
	2 biçimbirimli dizi	recp+nzmK	<i>görüştirmek için</i>	255
		Toplam		16,840
		caus+va1+nzmK	<i>sürdürebilmek için</i>	493
		caus+neg+nzmK	<i>kaçırmamak için</i>	100
		recp+caus+nzmK	<i>araştırmak için</i>	87
		caus+caus+nzmK	<i>çıkartmak için</i>	84
	3 biçimbirimli dizi	recp+neg+nzmK	<i>karşılaşmamak için</i>	59
		Toplam		823

Tablo 12. *İçin* ek dizileri -mA

	Biçimbirim sayısı	Sonek Dizisi	Örnek	Sıklık
		nzma+p3s	<i>olması için</i>	5,249
	2 biçimbirimli dizi	nzma+p3p	<i>olmaları için</i>	362
		nzma+pl	<i>çalışmalar için</i>	160
		Toplam		5,771
		pasv+nzma+p3s	<i>sağlanması için</i>	4,253
		va1+nzma+p3s	<i>olabilmesi için</i>	1,622
		neg+nzma+p3s	<i>olmaması için</i>	497
	3 biçimbirimli dizi	caus+nzma+pl	<i>işlemler için</i>	192
		nzma+pl+acc	<i>çalışmaları için</i>	173
		va1+nzma+p3p	<i>olabilmeleri için</i>	137
		refl+nzma+p3s	<i>kalkınması için</i>	127
		Toplam		7,001
		caus+pasv+nzma+p3s	<i>geliştirilmesi için</i>	1,268
	4 biçimbirimli dizi	pasv+va1+nzma+p3s	<i>yapılabilmesi için</i>	937
		pasv+neg+nzma+p3s	<i>yaşanmaması için</i>	119

Biçimbirim sayısı	Sonek Dizisi	Örnek	Sıklık
	recp+pasv+nzma+p3s	anlaşılması için	91
	caus+va1+nzma+p3s	sürdürebilmesi için	90
	refl+va1+nzma+p3s	alnabilmesi için	79
	recp+caus+nzma+pl	araştırmalar için	60
	Toplam		2,644
5 biçimbirimli dizi	caus+pasv+va1+nzma+p3s	gerçekleştirilebilmesi için	131
	recp+pasv+va1+nzma+p3s	anlaşılabilmesi için	91
	Toplam		222

Tablo 13. İçin biçimbirim dizileri -dlk

Biçimbirim sayısı	Sonek Dizisi	Örnek	Sıklık
2 biçimbirimli dizi	pcdk+p3s	olduğu için	12,069
	pcdk+p1s	olduğum için	1,757
	pcdk+p3p	oldukları için	1,497
	pcdk+p1p	olduğumuz için	237
	pcdk+p2p	olduğunuz için	138
	pcdk+p2s	olduğun için	121
	Toplam		15,819
3 biçimbirimli dizi	neg+pcdk+p3s	olmadığı için	2,610
	pasv+pcdk+p3s	bulunduğu için	802
	neg+pcdk+p1s	olmadığım için	222
	neg+pcdk+p3p	olmadıkları için	121
	refl+pcdk+p3s	alındığı için	75
	caus+pcdk+p3s	gerektirdiği için	60
Toplam		3,890	
4 biçimbirimli dizi	pasv+neg+pcdk+p3s	bulunmadığı için	218
	va1+neg+pcdk+p3s	bulamadığı için	169
	va1+neg+pcdk+p3p	bulamadıkları için	57
	Toplam		444

Tablo 14. *İçin* dizileri

Biçimbirim	Ek dizisi	Örnek	Sıklık
- An	pcan+pl	olanlar_ için	787
	neg+pcan+pl	olmayanlar_ için	57
	Toplam		844
-AcAk	pcck+p3s	olacağı_ için	263
	futr+3s	gelecek_ için	251
	Toplam		514
diğer	gen	bunun_ için	15,336
	pl	onlar_ için	8,734
	poss	yılı_ için	6,398
	c1s	benim_ için	4,557
	p3s	kendisi_ için	2,694
	p3p	çocukları_ için	1,640
	pl	amaçlar_ için	571
	p1s	kendim_ için	305
	p1p	ülkemiz_ için	280
	Toplam		40,991

6. Sonuç

Türkçenin biçimbilim yapısı, sık kullanılan ek dizilerinin sözcük yapısında ayrıntılı incelemeleri gerektiren zengin gerçekleştirmeler sunmaktadır. Daha kapsamlı bir betimleyici olarak sözcükbirim kavramı ile bu yapıları baktığımızda, en az üç temel tür belirleyebiliyoruz. Bu çalışma, olası sık kullanım örüntülerinde sözcüklerarası gerçekleşen dizileri bulmayı amaçlamış, ilk gözlemleri ve sayısal dağılımı araştırmacılara sunmuştur. Sınırlı sayıdaki dilde görülebilen bu tür örüntülerin tartışılması, Türkçenin tipolojik özelliklerinin diğer diller ile bu boyutta karşılaştırılması ile daha da belirginleşebilecektir.

* Bu bildiri TÜBİTAK 115K135 projesi çerçevesinde hazırlanmıştır. Katkılarından dolayı TÜBİTAK'a teşekkür ederiz.

Kaynakça

- Aksan, Y. vd. (2012). Construction of the Turkish National Corpus (TNC). Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC 2012), (3223-3227) İstanbul, Türkiye.
- Pedersen, T., Banerjee, S., McInnes, B., Kohli, S., Joshi, M. & Liu, Y. (2011). The Ngram Statistics Package (Text:NSP) - A Flexible Tool for Identifying Ngrams, Collocations, and Word Associations. Proceedings of Multiword Expressions: from Parsing and generation to the Real World (MWE 2011), an ACL HLT 2011 Workshop, (131-122), Portland.
- Sinclair, J. (1998). The lexical item. E. Weigand (Haz.), Contrastive Lexical Semantics (Current Issues in Linguistic Theory 171) (pp. 1-24). Amsterdam: John Benjamins.
- Wray, A. (2006). Formulaic language. Concise encyclopedia of pragmatics, Elsevier, 590-597.