

Öğrenir Derlemleri: Kapsam, Tasarım ve Uygulamalar

Yrd.Doç.Dr. Nihal Çalışkan

Yıldırım Beyazıt Üniversitesi, Türk Dili ve Edebiyatı Bölümü

caliskanihal@hotmail.com

ÖZET

Öğrenir derlemleri bir dili ikinci dil olarak edinen ya da yabancı dil olarak öğrenen dil öğrenirlerinin sözlü ve/veya yazılı dil kullanımlarını bir araya getiren veri tabanlarıdır. Bugün 100'ü aşkın dilde, çoğunluğu İngilizceyi yabancı dil olarak öğrenen bireylerin dil kullanımlarını örnekleyen öğrenir derlemi bulunmaktadır. Ancak Türkçenin yabancı dil olarak öğretiminin hız kazandığı bir dönemde henüz Türkçe Öğrenir Derlemi hazırlanabilmiş değildir. Bu çalışma ile Türkçenin yabancı dil olarak öğretimiyle ilgili pek çok çalışmaya temel oluşturacak öğrenir derlemlerini tanıtmak, bunların belli başlı örneklerini sıralamak, öğrenir derlemlerinin kapsam ve tasarımına ilişkin hususlar üzerinde durarak pedagojik uygulamalarına dikkat çekmek hedeflenmektedir. Böylece araştırmacılara yol gösterecek başlangıç noktalarına işaret etmek ve ileride hazırlanacak bir Türkçe Öğrenir Derlemi konusunda farkındalık ve motivasyon oluşmasına katkı sağlamak amaçlanmaktadır.

Anahtar Sözcükler: Öğrenir derlemleri, öğrenir derlemi tasarımı, hata etiketleme, ikinci/yabancı dil, eğitim uygulamaları

Learner Corpora: Scope, Design, and Applications

ABSTRACT

Learner corpora are collections of second/foreign language learners' spoken and written language use. There are over 100 learner corpora the majority of which the target language is English. Yet, there is no Learner Corpus of Turkish in spite of the developing character of Turkish as a Second/Foreign Language Research. This study aims to introduce the definitions and basics of learner corpora listing the prominent references of learner corpora. It calls attention to the design considerations and processing types of learner corpora. In this manner, we try to mark the starting points and hope to contribute for a future Learner Corpora of Turkish raising awareness and strengthen the motivation among Turkish language researchers.

Keywords: Learner corpora, major learner corpora, learner corpora design, error tagging, second/foreign language, pedagogical implications

1. GİRİŞ

Bilgisayar bilimlerinde kaydedilen gelişmelere bağlı olarak özellikle son 30 yıl içinde dilbilim çalışmalarına yeni bir soluk getiren derlem dilbilimine ait yaklaşım ve uygulamaların etkili olduğu alanlarından biri de dil eğitimidir. Derlem dilbiliminin geniş hacimli dil malzemesini depolama, bu malzemeyi dil üstü veya dile dair özellikler açısından etiketleme, bunlar üzerinde sıklık ve uygunluk başta olmak üzere çeşitli işlemler gerçekleştirmeyi mümkün kılan uygulamaları, dil eğitimi alanında da dikkat çekmiştir. Uygulamalı dilbilim ile derlem dilbiliminin kesiştiği bu noktada, çalışmaların özellikle yabancı dil öğretiminde derlem kullanımı ve öğrenirlerin1 dil kullanımlarını derlemeleştirme2 olmak üzere iki ana hat biçiminde ilerlediğini söylemek mümkündür. Ancak her iki alan da Türk dili araştırmalarında yeterince ele alınmış değildir. Bu çalışmada öğrenirlerin dil kullanımlarını bir araya getiren “öğrenir derlemleri” üzerinde durulacaktır. Öğrenir derlemlerinin tanıtımı yapılarak belli başlı örnekleri hakkında bilgi verilecek, öğrenir derlemlerinin tasarımı ve bunların ikinci dil edinimi ve yabancı dil öğretimi alanındaki kullanımlarından söz edilecektir.

2. TANIM ve AYIRICI ÖZELLİKLER

Öğrenir derlemleri Mc Enery vd. (2006, 65) tarafından “ikinci bir dili edinen bireylerin sözlü veya yazılı dil kullanımlarının bir koleksiyonu” olarak tanımlanmaktadır. Bu tanım son derece genel olmakla birlikte devamında yer alan öğrenir derlemleri ile gelişimsel derlemlerin (developmental corpus) birbirinden farklı olduğuna dair uyarı önemlidir. Buna göre gelişimsel derlemler, dili birinci dil olarak edinen çocukların dil gelişimlerini belgelerken öğrenir derlemleri ikinci dile dair malzemeyle sınırlıdır. Öğrenir derlemleri, uzun erimli dil kullanımlarını bir araya getirdiğinde gelişimsel bir nitelik de taşıyabilir. Gelişimsel derlemlerin en bilineni Child Language Data System (CHILDES)’dir (MacWhinney 1992) (Mc Enery vd. 2006, 65-66). 3

Öğrenir derlemlerinin kapsamlı bir tanımı, bu alanın önde gelen isimlerinden Granger (2002, 7) tarafından şöyle yapılmaktadır: “Dijital öğrenir derlemleri, ikinci dil edinimi/yabancı dil öğretimine yönelik belli bir amaç için harici tasarım ölçütlerine göre bir araya getirilen ikinci dil/yabancı dile ait otantikliğinin metin verilerinin elektronik koleksiyonudur. Öğrenir derlemleri, standartlaştırılmış ve homojen bir tarzda kodlanarak köken ve kaynaklarına göre belgelenirler.”4 Granger bu tanımı Sinclair’in (1996) derlem tanıtımından yola çıkarak yaptığını belirtmekte; ancak tanımda geçen bazı ifadelerin öğrenir derlemleri söz konusu olduğunda açıklama gerektirdiğini belirtmektedir. Bunlardan biri verilerin otantikliği meselesidir. Derlemler doğal dil kullanımlarını örnekleyen veri tabanlarıdır. Ancak ikinci/yabancı dil öğretimi söz konusu olduğunda çoğu metnin sınıf ortamında oluşturulduğu aşıkardır. Bu noktada Granger, öğrenir derlemlerinin otontikliğini, günlük hayatın akışı içinde tamamıyla doğal dil kullanımlarını yansıtan durumlardan sınıf ortamında gerçekleşen etkinlikler sırasında kullanılan dili karşılayan bir çerçevede ele alınması gerektiğini belirtmektedir (2002, 8). Bir başka husus derlemlerin; metinleri, Granger’in deyimiyle “kesintisiz söylem dizileri”ni (continuous stretches of discourse) bir

¹ Burada “öğrenir” sözcüğü “learner” karşılığı bir terim olarak önerilmektedir. Bir kısım çalışmalarda “öğrenci”, “öğrenici” terimleri de kullanılmıştır. Ancak “öğrenci” sözcüğünün daha çok formal eğitim sistemi içindeki bireyleri kapsadığı, “öğrenici” sözcüğünün de ses ve türetim özellikleri bakımından uygun olmadığı düşünülerek “öğrenir” terimi tercih edilmiştir.

² Derlem dilbilimi çalışmaları, Türkçe literatür açısından yenidir. Bu nedenle bir kısım terimlerin Türkçe karşılıkları ya da “derlemeleştirme” gibi bazı kavram, durum ve olguların Türkçe ile ifadesi okuyucaya “yeni” gelebilecektir. Bunun anlayışla karşılanacağını ümit etmekteyiz.

³ “CHILDES”te temsil edilen dillerden biri de Türkçedir.

⁴ “Computer learner corpora are electronic collections of authentic FL/SL textual data assembled according to explicit design criteria for a particular SLA/FLT purpose” They are encoded in a standardised and homogeneous way and documented as to their origin and provenance.

araya getirdiğidir. Bu bakımdan ikinci/yabancı dil öğrenirlerinin dil kullanımlarını örnekleyen metinlerden seçilen cümlelerle öğrenir derlemi oluşturmak söz konusu olmamaktadır (2002, 9-10). Tanımda belirsizlik yaratan hususlardan biri de öğrenir kitlesinin “anadili konuşurları dışında herkes” gibi bir kabul içermesidir. Granger söz konusu kitlenin de kendi içinde farklılaştığını belirterek “bir dili ikinci dil olarak edinenler, yabancı dil olarak öğrenenler ve resmi dil olarak kullananlar” biçiminde bir ayırım yapmakta ve öğrenir derlemlerinin ilk iki grubun dil kullanımları ile oluşturulması gerektiğini ifade etmektedir (2002, 9).

Öğrenir derlemlerinin çerçevesini böylece çizdikten sonra, amaçları üzerinde durmakta fayda vardır. Leuven Üniversitesinde Granger ve ekibi tarafından hazırlanan İngilizce öğrenir derlemi International Corpus of Learner English (ICLE) temelde İngilizceyi yabancı dil olarak öğrenen, anadilleri birbirinden farklı ileri düzey öğrenirlerin ortak dil (interlanguage) özelliklerini betimlemek amacını taşımaktadır. Bu temel amaca bağlı olarak dil öğrenirlerinin hatalarını belirlemek, bunların evrensel ya da dile özgü nitelikler taşıyıp taşımadığını, taşıyorsa hangi değişkenlerden etkilendiğini değerlendirmek; hedef dili esas alarak öğrenir dilinin “kulağa yabancı gelen” özelliklerini ortaya çıkarmak konularına odaklanılmaktadır. ICLE bir taraftan da öğrenirlerin bizzat bu derlemden istifade etmesini amaçlamaktadır (Pravec 2002, 82-83).⁵

3. TASARIM

Derlemin oluşturulma amacı, tasarımını da bütünüyle etkileyen bir özelliktir. Aynı zamanda ilk öğrenir derlemlerinden biri olan ICLE’den hareketle bu konuyu şöyle açabiliriz: ICLE’nin temel amacı, farklı dil geçmişlerinden gelen bireylerin İngilizce kullanımlarını karşılaştırmalı olarak incelemek biçiminde ifade edildiği için derlem malzemesi sözgelimi sadece anadili Flemenca ya da Fransızca olan bireylerden elde edilmemiş; aralarında Fransızca, Çince, Japonca, İbranice gibi tipolojik olarak birbirinden farklı dillerin bulunduğu 14 ayrı dilden gelen öğrenirlerin kullanımlarına yer verilmiştir. Bu noktada Granger’ın (2002, 10) “derlemlerin heterojen öğrenir verilerini rastlantısal olarak bir araya getiren koleksiyonlar olmadığı” uyarısını hatırlamakta fayda vardır. Böyle olduğunda araştırma sorularının yeterli biçimde cevaplanamaması ve verilerin bir bilgi yığına dönüşmesi kaçınılmazdır. İkinci/yabancı dil öğrenir kitlesinin kendi içinde birçok yönden farklılaştığı göz önünde bulundurulduğunda derlem tasarımı ile ilgili sıkı ölçütlere ihtiyaç duyulduğu daha rahat anlaşılacaktır. Tono’dan alıntılanan (2003, 800) aşağıdaki tablo bu tasarım kriterlerini şöyle göstermektedir:

Tablo-1: Öğrenir Derlemi Hazırlamada Kullanılacak Tasarım Kriterleri

ÖZELLİK TÜRÜ		
Dille İlgili	Görevle İlgili	Öğrenirle İlgili
Biçim (yazılı/sözlü)	Veri toplama (kesitsel/boylamsal)	İçsel-Bilişsel (yaş/bilişsel stil)
Tür (mektup/günlük/kurgu/deneme)	Çıkarım (spontan/hazırlanmış)	İçsel-Duygusal (güdü/tutum)

⁵ Pravec’in bu yazısında belli başlı öğrenir derlemlerinin amaçlarından kapsamlı biçimde söz edilmektedir.

Stil (öyküleme/tartışma)	Referans kullanımı (sözlük/kaynak metin)	L1 geçmişi L2 ortamı
Konu (genel/serbest zaman vb.)	Süre kısıtlılığı (sabit/serbest/ev ödevi)	(ikinci dil/yabancı dil) (okul düzeyi) L2 yeterliği (standart test sonucu)

4. DERLEM VERİLERİNİ İŞLEME

Derlemleri bilgisayar ortamında kaydedilen metin dosyalarından ayıran özelliklerden biri de bu verilerin dil üstü ve dile dair özellikler ile yazım/sesletim özellikleri açısından işlenmesidir. Bunları yine Tono'dan (2003: 801) alıntılan tabloda şöyle göstermek mümkündür:

Tablo-2: Öğrenir Verilerinin İşlenmesi

Metin Dışı Bilgiler	Bağlantı bilgisi (dil/görev/öğrenir değişkenleri)
Transkripsiyon Düzeyi	Yazım (+fonemik/fonetik, sözlü derlem için)
İşaretleme Düzeyi	Cümle sınırları belirleme Sözcükbirimleştirme Sözcük türü etiketleme Başsözcükleştirme Ögelere ayırma (ağaç dağılımı) Semantik etiketleme (sözcük anlamları, anlam ilişkileri ve kategoriler) Söylem etiketleme (özür/selamlaşma/nezaket/edim??/vb.) Hata etiketleme Prozodi etiketleme Gönderimsel etiketleme

Derlemler ham metin dosyalarından oluşabileceği gibi Tablo-2'de gösterildiği üzere çeşitli özellikler açısından etiketlenebilir. Derlem işaretleme (corpus annotation) adını verdiğimiz bu işlem, derlem malzemesine artı değer kazandırır (Mc Enery vd, 2006, 30). Bu sayede sözgelimi sözcük türü işaretleme yapılmış bir derlem üzerinde ad ve eylem olarak kullanılan “boya” sözcüklerini ayrı ayrı görüntülemek, “ve” bağlacının yalnızca ad türünden

sözcükleri bağlayan kullanımlarını elde etmek ya da bir eylemin yalnızca belli bir zaman veya kişi ile çekimini listelemek mümkün olabilmektedir. Aksi takdirde kullanıcı kimi durumlarda elle ayıklanması oldukça zaman alıcı bir dizi işlem gerçekleştirmek zorunda kalmaktadır. Öğrenir derlemlerinde en sık yapılan işaretleme hata etiketlemeye dayanmaktadır. Hata etiketleme daha çok elle yapılmakla birlikte sözcük türü işaretleme konusunda otomatik yöntemler daha sık kullanılmaktadır (Pravec, 2002, 97-100).

5. BAŞLICA ÖĞRENİR DERLEMLERİ

Burada yer verilen öğrenir derlemleriyle ilgili bilgiler, ICLE Projesi'ne de ev sahipliği yapan Leuven Üniversitesi İngiliz Derlem Dilbilimi Merkezinin İnternet sitesinden alınmıştır. İlgili sitede 124 öğrenir derlemine ilişkin bilgiler “hedef dil, birinci dil, biçim, metin/görev tipi, yeterlik düzeyi, hacim, proje ekibi, kullanılabilirlik” başlıkları altında tablolaştırılarak verilmiştir. Aşağıda söz konusu derlemlerden 1 milyon sözcük sınırını aşanlarla ilgili bilgiler sunulmuştur. Mevcut öğrenir derlemlerinin tam listesine <http://www.uclouvain.be/en-cecl-lcworld.html> adresinden ulaşmak mümkündür.

Tablo 3. The British Academic Written English. Hiary Nesi, Sheena Gardner, Warwick, UK. Paul Thompson, University of Birmingham, UK. Paul Wickens, Oxford Brookes, UK.

Hedef Dil	L1	Biçim	Metin/Görev Tipi	Yeterlik	Hacim	Kullanılabilirlik
İngilizce	L1 ve L2 konuşurları	Yazılı	ESL makaleleri	Lisans ve lisansüstü 4 farklı düzey	6,5 milyon	Sketch Engine ile erişim

Tablo 4. The Cambridge Learner Corpus. Cambridge University Press ve Cambridge ESOL, UK.

Hedef Dil	L1	Biçim	Metin/Görev Tipi	Yeterlik	Hacim	Kullanılabilirlik
İngilizce	Çeşitli	Yazılı	Sınav kâğıtları	Çeşitli	25 milyon, genişlemeye devam ediyor.	Ticari

Tablo 5. The Longman Learners' Corpus. Longman

Hedef Dil	L1	Biçim	Metin/Görev Tipi	Yeterlik	Hacim	Kullanılabilirlik
İngilizce	Çeşitli	Yazılı	Kompozisyon, Sınav kâğıtları	Çeşitli	10 milyon	Ticari

Tablo 6. The International Corpus of Learner English. Sylviane Granger, Centre for English Corpus Linguistics, Université Catholique de Louvain, Belgium

Hedef Dil	L1	Biçim	Metin/Görev Tipi	Yeterlik	Hacim	Kullanılabilirlik
-----------	----	-------	------------------	----------	-------	-------------------

İngilizce	Çeşitli	Yazılı	Tartışmaya dayalı ve edebî kompozisyon	Orta üstü ve ileri	3 milyon	CD-ROM
-----------	---------	--------	--	--------------------	----------	--------

Tablo 7. The Michigan Corpus of Academic Spoken English (MICASE). Ute Römer, University of Michigan.

Hedef Dil	L1	Bîçim	Metin/Görev Tipi	Yeterlik	Hacim	Kullanılabilirlik
İngilizce	L1 ve L2 konuşurları	Sözlü	Akademik sözlü kullanım örnekleri	-	1,8 milyon	Çevrimiçi erişilebilir.

Tablo 8. The Michigan Corpus of Upper-Level Student Papers (MICUSP). Ute Römer, University of Michigan.

Hedef Dil	L1	Bîçim	Metin/Görev Tipi	Yeterlik	Hacim	Kullanılabilirlik
İngilizce	(Yarı dengeli) L1 ve L2 konuşurları	Yazılı	ESP makaleleri, araştırma önerileri	-	2,6 milyon	Çevrimiçi erişilebilir.

Tablo 9. The Japanese English as a Foreign Language Learner (JEFL) Corpus. Yukio Tono, Meikai University

Hedef Dil	L1	Bîçim	Metin/Görev Tipi	Yeterlik	Hacim	Kullanılabilirlik
İngilizce	Japonca	Yazılı	Kompozisyon	Başlangıçtan orta düzeye	700,000	Ücretsiz erişim hedefleniyor.

Tablo 10. The NICT JLE (Japanese Learner English) Corpus. Emi Izumi, Kiyotaka Uchimoto, Hitoshi Isahara, National Institute of Information and Communications Technology, Kyoto, Japan

Hedef Dil	L1	Bîçim	Metin/Görev Tipi	Yeterlik	Hacim	Kullanılabilirlik
İngilizce	Japonca	Sözlü	Sözlü yeterlik görüşmeleri	Çeşitli	2 milyon	CD-Rom

Tablo 11. The Spoken and Written English Corpus of Cheenese Learners (SWECCCL). Wen Qiufang, Liang Maocheng, Wang Lifei

Hedef Dil	L1	Bçim	Metin/Görev Tipi	Yeterlik	Hacim	Kullanılabilirlik
İngilizce	Çince	Yazılı Sözlü	Tartışmacı ve öyküleyici kompozisyon, Ulusal Konuşma İngilizcesi Testi-uzamsal		2 milyon	Çevrimiçi erişilebilir.

Tablo 12. The City University Corpus of Academic Spoken English (CUCASE).

Hedef Dil	L1	Bçim	Metin/Görev Tipi	Yeterlik	Hacim	Kullanılabilirlik
İngilizce	Çince	Multimedya	-	-	2 milyon	

Tablo 13. The Bilingual Corpus of Chinese English Learners (BICCEL). Wen Qiufang. National Research Center for Foreign Language Education Beijing Foreign Studies University, China

Hedef Dil	L1	Bçim	Metin/Görev Tipi	Yeterlik	Hacim	Kullanılabilirlik
İngilizce	Çince	Yazılı Sözlü	Sınıf içi yazılı anlatımlar, Ulusal Sözlü İngilizce Testi	-	2 milyon	-

Tablo 14. The Chinese Learner English Corpus. Gui Shichun, Guangdong University of Foreign Studies. Yang Huizhong, Shanghai Jiatong, China

Hedef Dil	L1	Bçim	Metin/Görev Tipi	Yeterlik	Hacim	Kullanılabilirlik
İngilizce	Çince	Yazılı	-	Çeşitli	1 milyon	Kısıtlı (üniversite içi) erişilebilir.

Tablo 15. The TELEC Secondary Learner Corpus. Quentin Allan, University of Hong Kong, Hong Kong. David Yong Wey Lee, City University of Hong Kong, Hong Kong

Hedef Dil	L1	Bçim	Metin/Görev Tipi	Yeterlik	Hacim	Kullanılabilirlik

İngilizce	Çince	Yazılı			1,5 milyon	-
-----------	-------	--------	--	--	------------	---

Tablo 16. The Taiwanese Corpus of Learner English. Rebecca Hsue-Hueh Shih, Sun Yat-Sen University, Taiwan.

Hedef Dil	L1	Bçim	Metin/Görev Tipi	Yeterlik	Hacim	Kullanılabilirlik
İngilizce	Çince	Yazılı	Kompozisyon	Orta düzeyden ileriye	2 milyon	-

Tablo 17. The NUS Corpus of Learner English. Hwee Tou Ng, Siew Mei Wu, Daniel Dahlmeier, National University of Singapore, Singapore.

Hedef Dil	L1	Bçim	Metin/Görev Tipi	Yeterlik	Hacim	Kullanılabilirlik
İngilizce	Çince	Yazılı	Kompozisyon	Çeşitli	1 milyon	Erişilebilir.

Tablo 18. The AKCES/CZESL Corpus (Akvizieni korpusy èetiny-Acquisition Corpora of the Czech Language/Czech as a Second Language. Karel Sebesta. Charles University in Prague Technical University in Liberec, Czech Republic

Hedef Dil	L1	Bçim	Metin/Görev Tipi	Yeterlik	Hacim	Kullanılabilirlik
Çekçe	Çeşitli	Yazılı Sözlü	Kompozisyon Görüşme	Çeşitli	2 milyon	Geliştiriliyor.

Tablo 19. The Hong Kong University of Science and Technology (HKUST) Learner Corpus. John Milton, Hong Kong University of Science and Technology, Hong Kong.

Hedef Dil	L1	Bçim	Metin/Görev Tipi	Yeterlik	Hacim	Kullanılabilirlik
İngilizce	Çince, daha çok Kantonca	Yazılı	Süre sınırı olmadan yazılan EFL kompozisyonları, bitirme sınavları	Üniversite ve ileri düzey lise öğrencileri	25 milyon	-

Tablo 20. The PELCRA Learner English Corpus. Barbara Lewondowska-Tomaszczyk, University of Lodz, Poland.

Hedef Dil	L1	Bçim	Metin/Görev Tipi	Yeterlik	Hacim	Kullanılabilirlik
İngilizce	Lehçe	Yazılı ve Sözlü	Akademik metinler, resmi yazışmalar	Başlangıçtan ileri düzeye	3 milyon 2,8 milyon (Y) 200,000 (S)	Çevrimiçi erişilebilir.

Tablo 21. The Telecollaborative Learner Corpus of English and German Telekorp. Hulie Belz, Pennsylvania State University, USA.

Hedef Dil	L1	Bçim	Metin/Görev Tipi	Yeterlik	Hacim	Kullanılabilirlik
İngilizce	Almanca	Yazılı	Anadili konuşurları ile diğerlerinin konuşmalarına ait ikidilli ve uzamsal veriler	-	1,5 milyon	Erişime açık değil.

Tablo 22. The Uppsala Student English Corpus. Ylva Berglund Prytz, Margareta Westergren Axelsson, Uppsala University, Sweden.

Hedef Dil	L1	Bçim	Metin/Görev Tipi	Yeterlik	Hacim	Kullanılabilirlik
İngilizce	İsveççe	Yazılı	Kompozisyon	Çeşitli	1,221,265	Akademik ve eğitim amaçlı kullanıma açık.

Tablo 23. The Yonsei English Learner Corpus (YELC). Seok-Chae Rhee, CK Jung, Yonsei University, Korea.

Hedef Dil	L1	Bçim	Metin/Görev Tipi	Yeterlik	Hacim	Kullanılabilirlik
İngilizce	Korece	Yazılı	Yonsei Üniversitesi İngilizce İzleme Testi	A1'den C2'ye	1,145,794	

Tablo 24. The Gachon Learner Corpus. Brian Carlstorm.

Hedef Dil	L1	Biçim	Metin/Görev Tipi	Yeterlik	Hacim	Kullanılabilirlik
İngilizce	Korece (az sayıda Çince ve Korece)	Yazılı	Köşe yazısı ödevleri	Düşük orta düzey	1,277,077	Erişilebilir.

Tablo 25. Aprescrivov (Aprendera Escribiren Lovaina). Kris Buyse, KU Leuven, Belgium.

Hedef Dil	L1	Biçim	Metin/Görev Tipi	Yeterlik	Hacim	Kullanılabilirlik
İspanyolca	Flamaca	Yazılı	Çeşitli	A1'den C1'e	1 milyon	Sınırlı çevrimiçi erişim

Tablo 26. The Estonian Interlanguage Corpus (EIC) of Tallinn University. Pille Eslon, Tallinn University, Estonia.

Hedef Dil	L1	Biçim	Metin/Görev Tipi	Yeterlik	Hacim	Kullanılabilirlik
Estonca	Rusça, Fince, İngilizce, Almanca, Litvanca, Ukraynaca, Beyaz Rusça	Yazılı	Çeşitli	A1'den C2'ye	1,145,794	Sınırlı çevrimiçi erişim

6. ÖĞRENİR DERLEMLERİNİN KULLANIM ALANLARI

Öğrenir derlemlerinin başlıca kullanım alanları ikinci dil dinimi (second language acquisition-SLA) ve yabancı dil öğretimi (foreign language teaching-FLT)dir. Granger (2002, 5-6), Mark (1998, 78; Granger 2002,6'dan) ile aynı gözlemi paylaşarak bu iki alanda öğrenirlerin dil kullanımı konusunun ihmal edildiğini; öğrenir derlemlerinin ise tam da bu noktada, dikkatleri bir süredir göz ardı edilen “öğrenir çıktıları”na çevirdiği tespitini yapmaktadır. Bu birkaç açıdan hem gerekli hem de uygulamaya dönük iyileştirmeler sağlayacak bir durumdur. Bunları maddeler hâlinde şöyle sıralamak mümkündür:

1. **Öğrenir farkındalığının artması:** Öğrenir derlemleri, dil öğrenirinin kendi problemlerini anlamalarına ve kendi dil sistemlerine ilişkin yeni bakış açıları geliştirmelerine yardımcı olmaktadır. Bunun için öğrenirlerin doğru biçimi verilerek hata etiketlemesi yapılmış derlemlere erişimi sağlanmalıdır. Böylece öğrenir, karşılaştığı hataları “Ben de bu hatayı yapıyor muyum?, Buradaki hatayı nasıl düzeltebilirim? Demek ki bu hatalı bir kullanımış.” gibi soru ve çıkarımlarla değerlendirerek kendi dil kullanımını gözden geçirecektir.

Walsh (2010, 340) gerek kendisinin gerekse başkalarının dil kullanımlarına ilişkin hataları fark etme ve düzeltmeyi öğrenir açısından bir yetkinlik göstergesi olarak değerlendirmektedir. Dil öğrenirini öğrenir derlemleriyle karşılaştırmak bu yetkinliğin artırılmasını sağlayacaktır.

2. **Hata analizlerinin çıktıları:** Öğrenir derlemlerinin verileri anadili konuşurlarından elde edilen malzemelerle bir karşılaştırma yapma imkânı sunarak en sık hata yapılan alanları, hata yapma sebeplerini açığa çıkarmaktadır (Walsh, 2010, 341). Aslında öğrenir derlemlerinin pek çok uygulaması hata analizlerinden beslenmektedir.
3. **Müfredat geliştirme:** Öğrenir derlemlerinin verileri “öğrenme içeriğinin belirlenmesi ve yapılandırılması” açısından önemlidir. Yabancı dil öğretiminde kazandırılacak söz varlığının seçimi konusunda çoğu zaman sıklık ölçütüne başvurulmaktadır. Ancak öğrenirin hangi sözcüksel birimleri kazanmakta güçlük geçtiğinin belirlenmesi de sıklık yanında yol gösterici olacaktır. Aynı durum dil bilgisi öğretiminde de geçerlidir (Granger, 2002, 23-24).
4. **Materyal geliştirme:** Dil öğretimiyle ilgili materyaller denince ilk akla gelenler ders kitapları ve sözlüklerdir. Ders kitaplarında yer alan söz varlığı ve dil bilgisi etkinliklerinin belirlenmesinde öğrenir verileri önemlidir. Krashen (1983), öğrenir uygun materyalle karşılaştığında ikinci dil edinim sürecinin daha hızlı işlediğini belirtmektedir. Öğrenir derlemlerinin, öğrencinin ihtiyacı ve bulunduğu düzeye uygun dil malzemesinin belirlenmesinde kullanımı, Krashen’in dikkat çektiği konuda bir boşluğun doldurulmasına katkıda bulunacaktır. Sözlük alanında öğrenir derlemleri verisini kullanarak farklı bir tasarıma giden Longman Essential Activator Dictionary’den de burada söz etmekte fayda vardır. Bu sözlükte, sözcüğe ilişkin açıklamaların yanı sıra o sözcüğün kullanımında en sık yapılan hatalara ilişkin uyarılar yer almaktadır. Örneğin söz konusu sözcük “information (bilgi)” ise “!Don’t say ‘informations’. Say information. (“Bilgiler” demeyin. Bilgi deyin.) !Don’t say an information. Say a piece of information or some information. (“Bir bilgi” demeyin. Bir parça veya biraz bilgi deyin) !Don’t say an important problem. Say a serious problem or a big problem. (“Önemli bir problem” demeyin. Ciddi veya büyük bir problem deyin.)” gibi uyarılar yapılmaktadır (Granger, 2002, 25).
5. **Veri yönetimli öğrenme:** Veri yönetimli öğrenme (data driven learning-DDL), pedagojik amaçlar doğrultusunda derlem dilbiliminin yöntem ve araçlarını kullanmaya dayanmaktadır. Burada esas olan öğrencilerin bir öğrenir derlemi ve onu işleyecek yazılımı kullanarak dil hatalarına ilişkin keşifler yapması, hatalı dil kullanımları ile anadili konuşurlarının kullanımlarını karşılaştırarak farkındalığını artırmasıdır (Gilquin ve Granger, 2010).

7. SONUÇ

Bu çalışmada öğrenir derlemlerinin tanımı yapılarak öğrenir derlemleri hazırlanırken göz önünde bulundurulması gereken hususlara dikkat çekilmiştir. Belli başlı öğrenir derlemleri ve bunlara ilişkin bilgiler verilerek konuyla ilgilenen araştırmacılar için başlangıç noktalarına işaret edilmiştir. Dil öğretiminde derlem kullanımı gibi geniş bir çalışma sahasında öğrenir derlemlerinin gelecek vadeden uygulamalarına değinilmiştir. Bu yolla, ileride hazırlanacak bir Türkçe Öğrenir Derlemi konusunda farkındalık ve motivasyon oluşmasına katkı sağlanmaya çalışılmıştır.

KAYNAKLAR

Gilquin, G., Granger S., (2010). How can data-driven learning be used in language teaching?, In A. O’Keeffe and M. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*, London, Routledge, p. 359-370.

Granger, S. (2002). A Bird's-eye View of Computer Learner Corpus Research. In Granger, S., Hung, J. and Petch-Tyson, S. (eds) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam and Philadelphia: Benjamins.

Krashen, S. (1983). *The Input Hypothesis*. London: Longman.

Mark, K.L. 1998. “The Significance of Learner Corpus Data in Relation to the Problems of Language Teaching”. *Bulletin of General Education* 312: 77-90.

McWhinney, B. (1992). *The CHILDES Database* (2nd edition). Dublin,OH: discovery Systems.

Pravec, N. (2002). Survey of learner corpora. *ICAME Journal* 26: 81-114.

Sinclair, J. 1996. *EAGLES. Preliminary recommendations on Corpus Typology*. <http://www.ilc.pi.it/EAGLES96/corpusstyp/corpusstyp.html> (10 Aralık 2013)

Tono, Y. 2003. Learner corpora: design, development and applications. In D. Archer, P. Rayson, A. Wilson and McEnery (eds) *Proceedings of Corpus Linguistics*, pp. 800-809. Lancaster University.

Tony M., Richard, X., and Yukio, T.. *Corpus-based language studies: An advanced resource book*. London and New York: Routledge, 2006. 408 pp. ISBN: 978-0-415-28623-7.

Walsh, S. (2010) What features of spoken and written corpora can be exploited in creating language teaching materials and syllabuses? In A. O’Keeffe and M. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*. Abingdon: Routledge, pp.333-344.

SUMMARY

Corpus linguistics is a widening branch of linguistics which effects nearly all branch of linguistics in one way or other. One of the research fields it has been effective is language pedagogy. There are two main research streams at the intersection of applied linguistics and corpus linguistics: using corpus in language learning/teaching and compiling learner corpora. This study aims to introduce the concept of learner corpora.

Learner corpora are collections of second/foreign language learners’ spoken and written language use Mc Enery etc. (2006, 65) . Learner corpora are different from developmental corpora in the sense of they do not necessarily collect the longitudinal data. Granger (2002, 7) suggests the following definition which is based on Sinclair’s definition of corpora: “Computer learner corpora are electronic collections of authentic FL/SL textual data assembled according to explicit design criteria for a particular SLA/FLT purpose. They are encoded in a standardized and homogeneous way and documented as to their origin and provenance.” Here, she calls attention

to the different sense of “authenticity” when the learner corpora are in question, and point out the textual properties of corpora in general (2002, 8-9).

Learner corpora can be designed in many ways, but which determines the design features is the purpose of linguistic research. Table-1 shows some design considerations for building learner corpora (Tono, 2003, 800).

TYPES OF FEATURE		
Language-related	Task-related	Learner-related
mode	data collection	internal-cognitive
written/spoken	(cross-sectional/longitudinal)	(age/cognitive style)
genre	elicitation	internal-affective
(letter/diary/fiction/essay)	(spontaneous/prepared)	(motivation/attitude)
style	use of references	L1 background
(narration/argumentation)	(dictionary/source text)	L2 environment
topic	time limitation	(ESL/EFL)
(general/leisure/etc)	(fixed/free/homework)	(level of school
		L2 proficiency
		(standard test score)

Learner corpora are processed in some ways which is shown in Table-2:

Extra-textual information	Header information (learner/ language/ task variables)
Level of transcription	Orthographic (+ phonemic/ phonetic for spoken corpora)
Level of annotation	Sentence-boundary disambiguation Tokenisation POS tagging Lemmatisation Parsing (Treebanking) Semantic tagging (word senses/ semantic relationships and categories) Discourse tagging (apologies/greetings/politeness/?? moves/acts??/etc.) Error tagging Prosody annotation Anaphoric annotation

There are over 100 learner corpora the majority of which the target language is English. The major corpora are as follows: The British Academic Written English (BAWE), The Cambridge Learner Corpus (CLC), The

Longman Learners' Corpus. (LLC), The International Corpus of Learner English (ICLE), The Michigan Corpus of Academic Spoken English (MICASE), The Michigan Corpus of Upper-Level Student Papers (MICUSP), The Japanese English as a Foreign Language Learner (JEFL) Corpus, The NICT JLE (Japanese Learner English) Corpus, The Spoken and Written English Corpus of Chinese Learners (SWECCCL), The City University Corpus of Academic Spoken English (CUCASE), The Bilingual Corpus of Chinese English Learners (BICCEL), The Chinese Learner English Corpus (CLEC), The TELEC Secondary Learner Corpus, The Taiwanese Corpus of Learner English, The NUS Corpus of Learner English, The AKCES/CZESL Corpus (Akvizieni korpusy èetiny-Acquisition Corpora of the Czech Language/Czech as a Second Language, The Hong Kong University of Science and Technology (HKUST) Learner Corpus, The PELCRA Learner English Corpus, The Telecollaborative Learner Corpus of English and German TELEKORP, The Uppsala Student English Corpus, The Yonsei English Learner Corpus (YELC), The Gachon Learner Corpus, Aprescrilov (Aprendera Escribiren Lovaina), The Estonian Interlanguage Corpus (EIC) of Tallinn University.

Learner corpus data and implications can increase learner awareness of his or her language development. It can be used in curriculum design, material development, data-driven learning.