

İKİNCİ DİL EDİNİMİ ÇALIŞMALARINDA BİLGİSİYAR DESTEKLİ BİR TÜRK ÖĞRENİCİ İNGİLİZCESİ DERLEMİ: ICLE*’NİN BİR ALTDERLEMİ OLARAK TICLE**

Cem Can

Özet

Öğrenici derlemi birbiriyle yakından bağıntılı dört alanın kesişim noktasında yer almaktadır: derlem dilbilim, dilbilim kuramı, ikinci dil edinimi ve yabancı dil öğretimi (Granger, 2009). Uluslararası Öğrenici Derlemi (ICLE) gibi bir derlem, bu alanlara aradil gelişimini etkileyen çeşitli değişkenleri incelemeye yarayacak bol miktarda veri sağlar. Bu çalışmanın amacı, aradil alanında çalışmalara katkı sağlaması amacıyla hazırladığımız ve Belçikada Université catholique de Louvain tarafından yayınlanan ve ICLE’ bir altderlemi olarak sunduğumuz Uluslararası Türk Öğrenici Derlemine örneklerle tanıtmaktır.

Anahtar sözcükler: TICLE, Derlem Dilbilim, öğrenici derlemi, aradil, ICLE

A TURKISH LEARNER CORPUS OF ENGLISH IN SECOND LANGUAGE ACQUISITION STUDIES: TICLE AS A SUBCORPUS OF ICLE

Abstract

Learner corpus (LC) studies are situated at the crossroads of four interrelated subjects: corpus linguistics, linguistic theory, second language acquisition, and foreign language teaching (Granger, 2009). A LC, like International Corpus of Learner English (ICLE), contributes to these fields with the abundance of hands-on material enabling the researchers to examine various variables that could affect English interlanguage. This study aims at introducing Turkish International Corpus of Learner English (TICLE), which has recently been published by Université catholique de Louvain in Belgium, as a subcorpus of ICLE with samples from the actual corpus.

Key words: TICLE, Corpus Linguistics, learner corpus, interlanguage, ICLE

* International Corpus of Learner English - Uluslararası Öğrenici İngilizcesi Derlemi
<http://cecl.fltr.ucl.ac.be/Cecl-Projects/Icle/icle.htm>

** Turkish International Corpus of Learner English – Uluslararası Türk Öğrenici İngilizcesi Derlemi

1. Giriş

Tarihsel süreç içerisinde, insanın en önemli uygarlık göstergelerinden sayılan dil metinlerini incelemenin, her biri farklı önceliklere sahip üç alt dalı ortaya çıkmıştır. Bonelli ve Sinclair (2006: 206-207) bunların dil metinlerini ele alış biçimlerini şu şekilde özetlemektedir:

1. *Bilgi edinimi*: Dilbilim ve diğer bilim dallarının, metinleri içerdikleri bilgi açısından incelemesiyle ortaya çıkan çalışmalar.
2. *Dil araştırmaları*: Eski Yunanca ve Sanskritçe üzerine araştırmalarla başlayıp, ilk çevirileri ve sözcüklerin anlamları üzerine yapılmış çalışmaları da içerisine alarak gelişen ve zaman içerisinde dilbilgisi ve sözlükbilim alanlarına evrilen çalışmalar.
3. *Yazın ve düşün eserleri*: Belli başlı yazınsal, düşünsel ve dinsel içerikli metinleri öncelikle bilgi ve fikir kaynağı olarak görmekle birlikte, bunların dillerini ayrıntılı biçimde yapısal ve sözcüksel olarak inceleyerek, dizin ve tanımlı dizinlerini oluşturan çalışmalar.

Şimdilerde derlem dilbilim denilince usa ilk gelen elektronik veya bilgisayarlı derlem dilbilim olsa da, bilgisayar öncesi dönemlerde de dil derlemleri toplanmıştır. Yukarıdaki sınıflamanın ikincisine giren ve bunun önemli örneklerinden birisi olan, Murray'in 1928 tarihli *Oxford English Dictionary*, 200,000 gönüllü okuyucunun elle tuttuğu notlardan oluşan 4,000,000'dan fazla alıntıyı içermektedir (Svartvik, 2007). 1889 yılında, Lehçebilim alanında Ellis'in 800'den fazla katılımcıdan elde ettiği veri ve onların yardımları ile oluşturduğu *Existing Phonology of English Dialects* adlı çalışmasının derlemi, Jespersen'in 1954 tarihli *Modern English Grammar on Historical Principles* çalışmasında kullandığı ve sayısını Chaucer, Shakespeare, Swift ve Austin gibi İngiliz edebiyatının ileri gelen yazarlarının eserlerinden 300,000 – 400,000 alıntı olarak belirttiği dil derlemi ve Fries'in (1951) *The Structure of English* eserinin altyapısını oluşturan ve kayıtlarının elle tutulduğu 50 saatlik karşılıklı konuşma derlemi, bilgisayar öncesi dönemin önde gelen çalışmalarıdır.

Bilgisayar dönemi derlem çalışmalarının ilki 1961 yılında Francis ve Küçera tarafından toplanan ve Amerikan İngilizcesinden yaklaşık 1 milyon sözcüğü içeren Brown Derlemi'dir. Gerek İngilizce gerekse diğer dillerden toplanacak derlemler için ölçünleri belirlemiş olan Brown Derlemi, her biri 2000 sözcükten oluşan ve 15 farklı alandan toplanmış 500 metni içermektedir (Francis & Küçera, 1979). Aynı yıllarda İskoçya'nın Edinburg Üniversitesi'nde bir konuşma İngilizcesi derlemi çalışması da başlatılmıştır. Ancak, Sinclair'in de (2001) belirttiği gibi, gerektirdiği iş yükünün ağırlığından nedeniyle, bu derlemin boyutu ancak 300,000 sözcüğe kadar ulaşabilmiştir. 1978 yılında ise, Brown Derlemi'nin getirdiği ölçün izlenerek, onun İngiliz İngilizcesi eşdeğeri olan LOB¹ Derlemi oluşturuldu.

1 Lancaster – Oslo – Bergen: <http://khnt.hit.uib.no/icame/manuals/lob/INDEX.HTM>

Bilgisayarların işlemci hızlarının artması, depolama sınırlarının genişlemesi ve giderek daha fazla miktarda metnin elektronik biçimde elde edilebilirliğinin olası hale gelmesi derlemlerin boyutlarına ve çeşitliliğine de yansımaktadır. Bunun en belirgin örnekleri günümüzün en büyük yazılı dil derlemi olan ve Birmingham Üniversitesi ve Collins Yayınevi tarafından hazırlanan İngilizce Dil Bankası² ile Oxford Üniversitesi Yayınevi öncülüğünde hazırlanan İngiliz Ulusal Derlemi'dir (BNC³). İngilizce Dil Bankası derlem çalışmasına 1991 yılında başlanmış ve şu ana kadar 524 milyon⁴ sözcüğe ulaşılmıştır. İngiliz Ulusal Derlemi ise 1991 – 1994 tarihleri arasında toplanmış 100 milyon sözcükten oluşmaktadır. Bu derlem 20. yüzyılın son dönemi çağdaş İngiliz İngilizcesini hem yazılı hem de sözlü dilden örneklerle yansıtmaktadır. Teknolojik gelişmeler sözlü dil derlemlerinin boyutlarını da olumlu yönde etkilemiş ve İngilizce Dil Bankası'nın şu anda sahip olduğu yazıya dönüştürülmüş sözlü dil derlemi 50 milyon sözcüğe ulaşmıştır.

Çalışmamızın konusu olan bilgisayar destekli öğrenici derlemleri ise, ikinci dil edinimi araştırmalarında öğrenici aradilini çalışmaların odağına yeniden yerleştiren yeni kaynak olarak görülmektedir (Granger, 2008). Bu bağlamda, oluşturduğumuz Uluslararası Türk Öğrenici İngilizcesi Derlemi'nin ayrıntılı bir incelemesini sunmadan önce, bu türden derlemlerin aradil çalışmalarına katkısı ele alınacaktır.

2. Öğrenici derlemleri

İkinci dil edinimi ve yabancı dil eğitimi çalışmaları açısından bakıldığında, öğrencilerin aradil verilerinin toplanması ve çözümlenmesinde izlenecek yöntem, varılacak sonuçların geçerliliği konusunda iki açıdan önemlidir. Bunlardan ilki, ikinci dil edinim sürecinin, onu etkileyen etmenlerin ve aradilin zihinsel simgelenmesinin daha iyi anlaşılabilmesidir. İkincisi ise, öğrencilerin öğrenme sürecinde yaşadıkları güçlüklerin giderilmesi için öğretim ve öğrenim araçları tasarlayan araştırmacılara kaynak sağlamasıdır.

Aradil çözümlemesinin sağlıklı sonuç vermesi büyük ölçüde veri toplama yöntemi ile ilintilidir. İkinci dil edinimi çalışmalarında, öğrencinin dil üretimini etkileyen değişkenlere sınırlama getirmesi ve sonuçların yorumlanmasını kolaylaştırması açısından, daha çok deneysel söyletim işlemi ile veri toplanmasına ağırlık verilmektedir. Bu durumda, deneysel verinin yapay aradil örnekleri içermesi olasılığının yanısıra, sınırlı sayıda katılımcı ile gerçekleştiğinden, bulguların genelleştirilebilmesi sorunu ortaya çıkmaktadır (Granger, 2008). Konuya diğer bir açıdan yaklaşan Ellis ve Barkhuizen (2005) ise, toplanacak verinin asıl olarak öğrencilerin aradilde ileti kurgulamasına yönelik olması gerektiğini vurgulayarak, deneysel veri toplama yerine, üzerinde sınırlamanın göreceli olarak daha az olduğu yöntemlerin kullanılmasını savunmaktadır.

2 The Bank of English.

3 British National Corpus – İngiliz Ulusal Derlemi: <http://www.natcorp.ox.ac.uk/>

4 Derlem boyutu hakkında güncel bilgiye <http://www.collins.co.uk/books.aspx?group=153> adresinden ulaşılabilir.

Bilgisayar destekli öğrenici derlemleri, açık bir tasarım ölçütüne göre toplanmış, doğala yakın aradil metinlerinden oluşur. Bilgisayar ortamına aktarılmış bu derlemler, sıklık taraması ve dökümü gibi temel çözümlemelerin yanısıra, gelişmiş veri çözümlemelerini de olanaklı kılan çeşitli yazılımlar ile çözümlenerek, farklı anadil artalanlarından gelen öğrencilerin aradil özelliklerini ortaya koymak için kullanılan veri yığınlarıdır. Granger (1998) bu türden bir çözümlemenin, anadili İngilizce olanların koşt derlemleri ile öğrenici derlemlerinin Karşıtsal Aradil Çözümlemesi olduğunu ve geleneksel Karşıtsal Çözümleme ile karıştırılmamasının gerekliliğini vurgular. İlke olarak, öğrenici derlemleri İngilizce'yi hem yabancı dil hem de ikinci dil olarak öğrenenlerden toplanan metinlerden oluşabilir. Bu metinler, bağlamından arındırılmış bağımsız tümceler olmayıp, bir anlam dizgesi içerisindeki sözlü veya yazılı söylemleri içerirler (Granger, 2008).

Myles (2005), şimdiye kadar yapılan ayrıntılı olgu çalışmaları ve az sayıda katılımcı ile gerçekleştirilmiş diğer birçok çalışmanın ikinci dil edinimi alanına büyük katkı sağladığını ve şu anki varsayımları oluşturmada etkilerinin yadsınamayacağını belirtmektedir. Ancak, artık bu güncel varsayımların daha büyük ve daha iyi kurgulanmış veri kümeleriyle sınanmalarının gerekliliğini de vurgulayan Myles, bilgisayar destekli öğrenici derlemleri arasında iyi örnek olarak iki derlemi işaret etmekte ve bunların Uluslararası Öğrenici İngilizcesi Derlemi (ICLE) ile Longman Öğrenici Derlemi olduğunu belirtmektedir.

3. Uluslararası Öğrenici İngilizcesi Derlemi - ICLE

ICLE, 1990 yılında Belçika'da Katolik Louvain Üniversitesinden Granger tarafından bir araya getirilen ve 11 ülkeyi içerisine alan bir ekip tarafından oluşturulmaya başlandı. 2002 yılında, çalışmanın ilk ürünü olan öğrenici derleminin birinci sürümü ICLEv1 yayımlandığında, Bulgarca, Çekçe, Hollandaca, Fince, Fransızca, Almanca, İtalyanca, Lehçe, Rusça, İspanyolca ve İsveççe olmak üzere 11 anadil artalanından gelen İngilizce öğrencilerinden toplanmış yazılı veriyi içeriyordu.

Gelişimini sürdüren ICLE, proje ekibine Türkiye, Norveç, Çin, Japonya ve Güney Afrika'dan topladıkları altderlemlerle katılan araştırmacılarla, öğrenici aradil yelpazesini genişleterek, 2009 yılında yayınlanan ikinci sürümü ICLEv2 ile 16 altderleme sahip olmuştur (Granger, Dagneaux, Meunier ve Paquot, 2009).

Kilimci ve Can (2009)'da belirtildiği üzere, ikinci dil edinimi araştırmalarında geniş ölçekli karşılaştırmalara olanak sağlayacak olan bu derlem, öğrencilerin kullandıkları dilbilgisel kuruluşların veya sözcüklerin sıklıklarını ortaya çıkarmasının yanısıra, aradil hata çözümlemelerinde kullanılacak güvenilir bir veri yığını sunmaktadır. Ayrıca bu derlem, aradil hatalarının nereye kadar evrensellik boyutu taşıdığını, belirli bir aradile mi özgün olduğunu, öğrencilerin anadil, başka bir yabancı dil veya diller ve eğitim değişkenlerinden ne derece

etkilendiğini incelemeye yardımcı olacaktır. Zaten birinci sürüm olan ICLEv1'nin altderlemleri kullanılarak da bu doğrultuda çok sayıda araştırma yürütülmüştür (Ädel, 2006; Eriksson, 2008; Gilguin, Papp ve Díez, 2008; Granger, 1998; 2008; Granger, Hung ve Petch-Tyson, 2002; Martelli, 2007; Meunier ve Granger, 2008; Nesselhauf, 2005).

Şimdi ICLE ve yazarın da proje yürütücüsü olarak yer aldığı ekip tarafından hazırlanan ICLE'nin bir altderlemi olan Türk öğrenci İngilizcesi derleminin (TICLE) oluşturulmasında gözetilen tasarım ölçütlerini inceleyelim.

3.1. Tasarım ölçütleri

ICLEv2, 16 farklı anadil artalanından gelen, üçüncü ve dördüncü sınıf üniversite öğrencilerinden toplanmış ve bilgisayar ortamına aktarılmış tartışma içerikli denemelerden oluşmaktadır. Üniversite öğrencilerinin dil yeterlilik seviyelerinin ileri düzey olması, İngilizce'yi yabancı dil olarak öğreniyor olmaları ve her altderlemin 200,000 sözcükten oluşması derlemin toplanmasında tanımlanan önkoşullar olarak belirlenmiştir. Bunların yanısıra, derlemin toplanma aşamasında, öğrenci görünüm sormacası ile belirlenen değişkenler de derlemin çözümleme yazılımına eklenerek bu değişkenlere göre arama yapılabilmesine olanak tanınmıştır.

3.1.1. Öğrenici değişkenleri

- Toplanan altderlemlerdeki öğrenci değişkenlerine:
- anadil,
- yaş,
- cinsiyet,
- bölge (ilgili anadilin birden fazla ülkede konuşulması ve derlemin o ülkeleri içermesi durumunda),
- diğer yabancı dil(ler),
- İngilizcenin anadil olarak konuşulduğu ülkelerde kalma süresi,
- öğrenme ortamı ve
- dil yeterliliği

olarak karar verilmiştir.

Anadil ölçütü değerlendirildiğinde karşımıza yukarıda (bölüm 3.) belirtilen 16 anadil çıkmasına rağmen, veri toplama aşamasında öğrencilere evde konuşulan diğer dilleri de sorulmuş ve bunlar derleminde birinci, ikinci ve ya üçüncü dil olarak belirtilmiştir. Buna göre TICLE altderlemi Çizelge 1'deki görünümü sergilemektedir.

Çizelge 1: TICLE içerisinde anadil ve evde konuşulan diller dağılımı

Anadil	Dağılım	Yüzde (%)
Türkçe	276	98
Diğer	4	1.4
Evde konuşulan 1. dil	Dağılım	Yüzde (%)
Türkçe	271	96.8
Diğer	9	3.2
Evde konuşulan 2. dil	Dağılım	Yüzde (%)
Türkçe	8	2.9
Almanca	2	0.7
İngilizce	1	0.4
Diğer	14	5.0
Yok	255	91.1
Evde konuşulan 3. dil	Dağılım	Yüzde (%)
Yok	280	100

Altderlemlere göre yaş ve cinsiyet değişkenlerinin dağılımları Çizelge 2’de sunulmuştur.

Çizelge 2: ICLE içerisinde yaş ve cinsiyet dağılımı

Altderlem	Yaş Ortalaması	Cinsiyet	
		Kadın	Erkek
Türkçe (TICLE)	22.08	%81	%19
Bulgarca	20.55	%83	%17
Çince	20.49	%64	%36
Çekçe	22.07	%72	%28
Hollandaca	20.75	%73	%27
Fince	22.73	%85	%15
Fransızca	21.70	%88	%12
Almanca	23.39	%78	%22
İtalyanca	24.59	%92	%7
Japonca	20.06	%73	%27
Norveççe	23.94	%74	%26
Lehçe	23.39	%80	%20
Rusça	21.19	%84	%16
İspanyolca	21.72	%86	%14
İsveççe	27.73	%77	%23
Tswana	22.47	%60	%40
Genel Görünüm	22.30	%76	%24

Altderlemlerdeki öğrenici değişkenleri arasında sıraladığımız bölge değişkeni, Çin (anakara Çin ve Hong Kong), Hollanda (Belçika ve Hollanda), Almanya (Almanya, Avusturya, İsviçre) ve İsveç (İsveç, Fillandiya) tarafından o dillerin anadil konuşanlarının bulunduğu farklı ülke ve bölgelerden toplanan altderlemler için değişkenlik göstermektedir. Bizim derlemimizi oluşturan metinlerin tümü Türkiye içerisindeki üç üniversiteden (Çukurova Üniversitesi, Mersin Üniversitesi, Mustafa Kemal Üniversitesi) toplandığı için, TICLE için bu değişken geçerli değildir.

Tüm ICLE derlemi içerisinde bakıldığında İngilizcenin yanısıra öğrenilen ikinci yabancı diller açısından önceliğin Almancada (%32) sonrada Fransızcada (%27) olduğu görülmektedir. Türk öğrenciler içerisindeki dağılım da bu sıralamayı bozmamaktadır. Çizelge 3, TICLE'nin bu değişken açısından görünümü sunmaktadır.

Çizelge 3: TICLE içerisinde 2. ve 3. yabancı dillerin dağılımı

2. Yabancı dil	Dağılım	Yüzde (%)
Almanca	260	92.9
Fransızca	15	5.4
Diğer	5	1.8
3. Yabancı dil	Dağılım	Yüzde (%)
Yok	254	90.7
Fransızca	10	3.6
Almanca	5	1.8
İspanyolca	1	0.4
Hollandaca	1	0.4
Diğer	9	3.2

ICLE derleminde yazılı dil üretimleri bulunan İngilizce öğrencilerine yönelik diğer bir değişken olarak İngilizcenin anadil olarak konuşulduğu ülkelerde bulunma süresi değerlendirilebilir. Çizelge 4, buna yönelik sayısal veriyi TICLE açısından sunmaktadır.

Çizelge 4: TICLE'de İngilizce konuşulan ülkede bulunma süresi dağılımı

İngilizce konuşulan ülkede bulunma süresi	Dağılım	Yüzde (%)
Hiç	276	98.6
3 – 6 ay	2	0.7
6 – 12 ay	1	0.4
10 yıl üzeri	1	0.4

Buna göre TICLE altderleminde verisi bulunan öğrencilerin hemen hiçbiri İngilizcenin anadil olarak konuşulduğu bir ülkede kalmamışlardır. ICLE geneline baktığımızda ise öğrencilerin ancak %23'ünün İngilizce konuşulan bir ülkede üç ay veya daha fazla süreyle bulunduğunu sayısal bilgi olarak verebiliriz.

Öğrenme ortamı değişkeni incelendiğinde ise, ICLE'nin tüm altderlemlerindeki metinlerin, İngilizce'yi yabancı dil olarak sınıf ortamında öğrenenlerden toplandığını görmekteyiz.

ICLE oluşturulurken amaçlanan, ileri düzeyde İngilizce dil yeterliliğine sahip, üniversite 3. ve 4. sınıf öğrencilerinden veri toplamak olmuştur. Toplanan altderlemlerde bu hedefe ne derece ulaşıldığını görmek için 16 altderlemin her birinden 20 öğrenci denemesi gelişigüzel örneklem ile seçilmiş ve Avrupa Ortak Dil Çerçevesi yazılı anlatım değerlendirmesinde uzman bir notlandırıcıdan bunları değerlendirmesi istenmiştir (Granger, Dagneaux, Meunier, ve Paquot, 2009). Bu değerlendirme sonuçları Çizelge 5'de sunulmuştur.

Çizelge 5: ICLE'nin Avrupa Ortak Dil Çerçevesi'ne göre görünüşü

Altderlem	B2 ve altı	C1	C2	Toplam
Türkçe (TICLE)	16	4	0	20
Bulgarca	2	16	2	20
Çince	19	1	0	20
Çekçe	11	9	0	20
Hollandaca	1	11	8	20
Fince	3	8	9	20
Fransızca	3	11	6	20
Almanca	1	12	7	20
İtalyanca	10	9	1	20
Japonca	18	2	0	20
Norveççe	8	7	5	20
Lehçe	1	12	7	20
Rusça	3	15	2	20
İspanyolca	12	8	0	20
İsveççe	0	14	6	20
Tswana	18	0	2	20
Genel Görünüm	126	139	55	320

(Granger, Dagneaux, Meunier, ve Paquot, 2009:12'den uyarlanmıştır)

Buna göre, ICLE içerisinden örneklem olarak alınan denemelerin %60'nın ileri düzey, diğerlerinin ise orta ve daha alt düzey yeterlilikte olduğu gözlenmektedir.

TICLE altderlemimize bakıldığında ise, alınan 20 örnek içerisindeki denemelerin %20'sinin ileri düzey, %80'ninin ise orta ve daha alt düzeylerden olduğu anlaşılmaktadır. Derlem içerisinde dil düzeyi belirleme işlemi, Granger, Dagneaux, Meunier, ve Paquot (2009) tarafından da belirtildiği üzere, deneme amaçlı olarak, sınırlı sayıda örnekle, tek bir notlandırıcı tarafından yapılmış olup, sonuçların daha kapsamlı değerlendirme yöntemleri ile sağlama yapılarak kesinleştirilmesi gerekmektedir.

3.1.2. Metin özellikleri

Geniş bir konu çeşitliliğine sahip tartışma denemelerinden oluşan ICLE derlemi oluşturulurken, denemelerini yazmaları için öğrencilere konu seçiminde yardımcı olmak açısından birtakım önerilerde bulunulmuştur. Zorunlu olmamakla birlikte, bu konulardan birisi üzerinde yazmayı seçen farklı ülkelerdeki öğrencilerin konu seçimleri de farklılık göstermiştir. Çizelge 6, ICLE metinlerinde en çok seçilen 10 konuyu ve ülkelere göre dağılımlarını sunmaktadır.

Çizelge 6: En çok seçilen 10 konu ve altderlemlere göre dağılımları

Deneme konusu	Denemelerin sayısı	Altderlem dağılımı
Some people say that in our modern world, dominated by science, technology and industrialization, there is no longer a place for dreaming and imagination. What is your opinion?	491	%29 Bulgarca
Most university degrees are theoretical and do not prepare students for the real world. They are therefore of very little value.	249	%22 Türkçe - TICLE
Poverty is the cause of the HIV/AIDS epidemic in Africa. Marx once said that religion was the opium of the masses. If he was alive at the end of the 20th century, he would replace religion with television.	237	%19 Rusça
The prison system is outdated. No civilized country should punish its criminal: it should rehabilitate them.	176	%32 Tswana
Discuss the advantages and disadvantages of banning smoking in restaurants.	156	%100 Çince
Discuss the advantages and disadvantages of using credit cards.	149	%100 Çince
Feminists have done more harm to the cause of women than good.	149	%23 Russian

In the words of the old song “Money is the root of all evil.”	133	%22 Russian
In his novel “Animal Farm,” George Orwell wrote “All men are equal: but some are more than others.” How true is this today?	127	%39

(Granger, Dagneaux, Meunier, ve Paquot, 2009:6-7’den uyarlanmıştır)

Derlenen metinlerin yazımı sırasındaki ortama ilişkin şartlar da ICLE’nin değişkenleri arasında yer almakta ve bu değişkenler derlem yazılımı içerisinde sunulan seçenekler ile ikinci dil edinimi araştırmacılarının erekleri doğrultusunda seçim yapmalarını olanaklı kılmaktadır. Bunlar, yazım sırasında, denemelerin zaman sınırlamalı ya da sınırlamasız, bir sınavın parçası ya da bağımsız, başvuru kaynağı kullanılarak ya da kullanılmadan hazırlandığına ilişkin ortam şartlarını belirtirler. Çizelge 7, ICLE ve TICLE derlemlerinin bu değişkenler açısından görünümünü sunmaktadır.

Çizelge 7: Ortam değişkenleri açısından ICLE ve TICLE’nin görünümüleri

Değişken	ICLE		TICLE	
	ICLE	ICLE	Metin Sayısı	Dağılım
Deneme türü:				
Tartışma	5554	%100	280	%100
Zaman:				
Sınırlamasız	3662	%65.9	280	%100
Sınırlamalı	1683	%30.3	-	-
Bilinmiyor	209	%3.8	-	-
Başvuru kaynağı:				
Hayır	2846	%51.2	134	%47.9
Evet	2510	%45.2	146	%52.1
Bilinmiyor	198	%3.6	-	-
Sınav:				
Hayır	3600	%64.8	280	%100
Evet	1594	%28.7	-	-
Bilinmiyor	360	%6.5	-	-

Çizelge 7’den de anlaşılacağı üzere, TICLE içerisindeki metinlerin tümü zaman sınırlaması olmaksızın ve herhangi bir sınavın parçası olmadan, öğrenciler tarafından yazılmış tartışma denemeleridir. Bu denemelerin yazımı sırasında, öğrencilerin yarısından biraz fazlası (%52.1) bir başvuru kaynağından yararlanmışlardır.

Genel olarak ICLE ve özel olarak TICLE içerisinde, sözcük sayısı açısından metinlerin uzunluğu ise Çizelge 8’de sunulmaktadır.

Çizelge 8: Metin uzunlukları açısından ICLE ve TICLE'nin görünümüleri

Sözcük sayısı	ICLE		TICLE	
	Metin Sayısı	Dağılım	Metin Sayısı	Dağılım
<=200	52	%0.9	-	-
200< - <=500	1837	%33.1	3	%1.1
500< - <=1000	3274	%58.9	272	%97.1
1000< - <=1500	340	%6.1	5	%1.8
1500< - <=2000	41	%0.7	-	-
>2000	10	%0.2	-	-

Buna göre TICLE metinlerinin neredeyse tamamının (%97.1), ICLE metinlerinin ise çoğunluğunun (%58.9) sözcük sayısı açısından 500 – 1000 aralığına düştüğünü görmekteyiz.

4. Biçimleme ve dilbilgisel açılımlama

Derlemi oluşturan metinlerde, veri işlenirken hata işaretlemesi yapılabilmesi için tüm öğrenici hataları olduğu gibi bırakılmıştır. Metin biçimlemesinde, tanımlaması Granger, Dagneaux, Meunier, ve Paquot, (2009:11-12) tarafından yapılan sadece dört anlamlandırma imi kullanılmıştır:

- çıkartılmış alıntı⁵: <*> veya <quote>
- çıkartılmış atıf: <R>
- anlaşılamayan sözcük: <?>
- deneme imi⁶: ICLE-TR-CUK-0001.1 (Çukurova Üniversitesinden bir denemeye ait im)

Derlemdeki metinler ANSI⁷ biçiminde ve öğrenici görünümünün bulunduğu bir veri tabanına bağlantılı durumda bulunmaktadır.

Derlem, ICLEv2 içerisinde yerleşik bir bağlamlı dizin yazılımı⁸ desteği ile sözcük biçimlerini, bunların köklerini, çoklu sözcük birimlerini, sözcük türü imlerini ve düzenli anlatımları kolaylıkla bulup çıkarmayı olası kılar. Ayrıca derlemdeki tüm denemeler CLAWS⁹ yazılımı ile biçimbirimsel olarak işaretlenmiş ve tümce kurucuları imlenmiştir¹⁰. Bunun bize sağladıklarından birkaçını örneklememiz gerekirse:

5 Metin içerisindeki alıntılar çok kısa değilse veya tümce içerisinde bütünleşik halde değilse çıkartılmıştır.

6 Deneme başlıkları ve öğrenici isimleri metinlerden çıkarılmış, bunların yerine her bir denemeye özgün bir im verilmiştir.

7 American National Standards Institute - Amerikan Ölçünler Enstitüsü.

8 Unitex adındaki bu yazılım, Paris-Est Marne-la-Vallée Üniversitesi, Gaspard-Monge Enstitüsünden Sébastien Paumier tarafından geliştirilmiştir.

9 Constituent Likelihood Automatic Word-tagging System – Kurucu Olabilirliği Özdevinimli Sözcük İmleme Dizgesi.

10 İmlmeler CLAWS7 ile Lancaster Üniversitesinden Paul Rayson tarafından yapılmıştır.

- yalın durumdaki bir sözcüğün tüm oluşumlarının sorgulanması (örneğin sorgulamada **water** <Vlex> dizimiyle *water* sözcüğünün eylem olarak oluşumlarını, veya (**he+she+it**) <Vlex> yazarak sözcüğün tüm üçüncü tekil kişi adıları ile birlikte dizinlerini);
- bir biçimbirim işaretinin (örneğin bir altderlemdeki red-light, four-wheel, gender-neutral gibi tüm tireli karmaşık nitimleri <ADJ><<->> dizimiyle) veya biçimbirim işaretlerinin oluşturduğu bağlamli dizinlerinin çıkarılması

gibi derlem dilbilimsel işlemleri sıralayabiliriz.

Yazılımın güvenilirlik sınaması, İngilizce Derlem Dilbilim Merkezi¹¹ tarafından, ICLE'deki 16 anadil artalanını gösterimleyen ve özdevinimle kurucu imlemesi yapılmış 51 öğrenici denemesinin doğruluk oranlarına bakılarak yapılmıştır. Sınama sonucu, %95 - %99.1 aralığını işaret etmektedir. Buna rağmen, Granger (1997) tarafından da önerildiği üzere, çalışmalarını derlemin kurucu imlemesi yapılmış hali üzerinde yürütecek araştırmacıların, kendi sınamalarını da yapıp, özdevinimle imlenmiş kurucuların doğrulukları için sağlama yapmaları çalışmalarının güvenilirliği açısından yarar sağlayacaktır.

5. TICLE'nin nicel özellikleri

Bilgisayar destekli öğrenici derlemleri, araştırmacıların kolaylıkla öğrencilerin hedef dildeki yazılı üretimlerine kuş bakışıyla bakabilmelerini olanaklı kılar. TICLE'nin genel nicel özelliklerini bu türlü bir bakış açısıyla inceleyelim.

5.1. Sözcük Sıklığı

Çizelge 9, WordSmith Tools (Sürüm 5.0.0.214) metin geri kazanım yazılımı aracılığı ile yaptığımız TICLE ve ICLE'de en çok kullanılan 20 sözcük taraması sonuçlarını ve bunların Granger (2008)'in sözünü ettiği İngilizce anadilde akademik yazılı derlem'inin (ACAD) sıklık sonuçları ile karşılaştırmasını sunmaktadır.

Çizelge 9: TICLE, ICLE ve ACAD derlemlerinde en sık kullanılan 20 sözcük

TICLE			ICLE		ACAD	
Sıralama	Sözcük	%	Sözcük	%	Sözcük	%
1	the	5.48	the	5.55	the	6.76
2	to	3.04	to	3.26	of	4.05
3	of	2.74	of	3.11	to	2.69
4	and	2.50	and	2.71	and	2.66
5	in	2.29	a	2.17	in	2.30

¹¹ Center for English Corpus Linguistics, Université catholique de Louvain, Belçika.

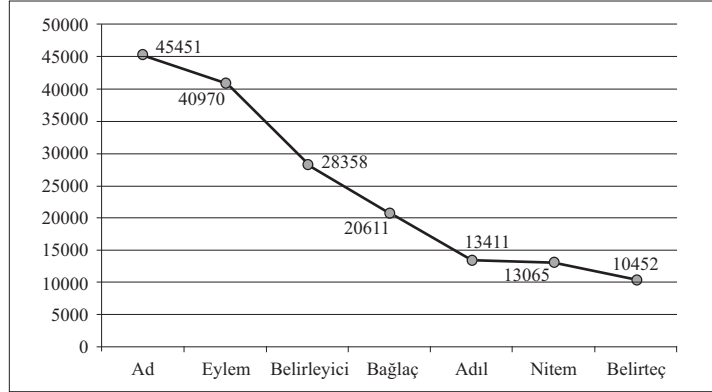
6	is	2.04	in	2.07	a	2.28
7	a	1.99	is	2.07	is	1.64
8	they	1.61	that	1.48	that	1.30
9	are	1.42	it	1.22	it	0.90
10	that	1.29	are	1.08	for	0.88
11	not	1.14	they	0.97	as	0.84
12	for	1.12	for	0.96	be	0.82
13	it	1.09	be	0.91	this	0.70
14	be	1.02	not	0.90	are	0.68
15	their	0.94	have	0.76	with	0.68
16	can	0.86	as	0.72	by	0.67
17	have	0.85	people	0.69	on	0.59
18	people	0.80	this	0.68	was	0.59
19	this	0.79	I	0.66	not	0.56
20	as	0.75	we	0.66	which	0.52

WordSmith gibi yazılımlarla sözcük veya sözcük dizilerinin bir derlemdeki sıklık değerlerinin dökümünü alabilir ve bunları diğer derlemlerle kıyaslayabiliriz. Öğrenici derlemlerinin, anadilde yazılmış karşılaştırılabilir akademik metinlerden oluşmuş derlemler ile bu türden sözlüksel karşılaştırmaları, bize öğrenici yazılı aradilinin anadil konuşanının yazılı üretiminden ne derece farklılık gösterdiği üzerine ipuçları sağlayabilmektedir. Ayrıca, bu türlü sözcük dizelgeleri yabancı dil öğretimi etkinliklerinde olası yol göstericiler olarak da görülebilir.

5.2. Sözcük türlerinin dağılımı

4. bölümde sözünü ettiğimiz CLAWS biçimbirimsel işaretleme yazılımı ile imlenen tümce kurucularını, üzerinde araştırma yapılacak sözcük türlerine göre ayrıntılı olarak dizgelemek olasıdır. Örneğin, Çizge 1., TICLE içerisinde genel bir bakışla sözcük türlerine göre dağılımı sergilemektedir. Buna göre, *ad* ve *eylem* imi taşıyan sözcükler birbirlerine yakın oranda bir dağılım sergilemekte ve derlemin çoğunluğunu oluşturmaktadırlar. Bunları, dağılımdaki oranlarına göre sırasıyla *belirleyici*, *bağlaç*, *adıl*, *nitem* ve *belirteçler* izlemektedir.

Çizge 1. TICLE derleminde sözcük türlerine göre dağılım



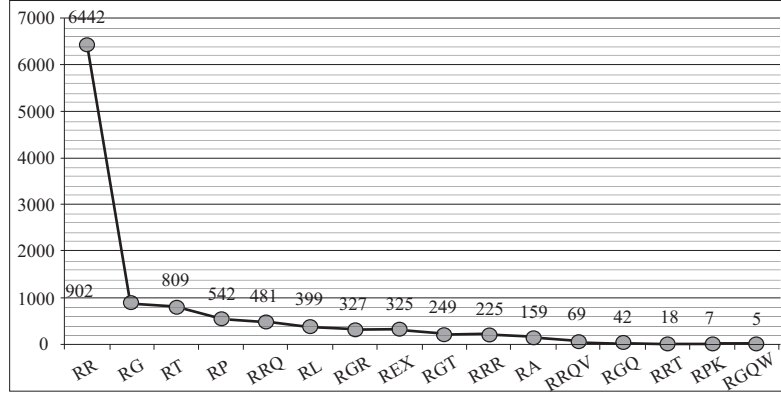
Çizge 1.'de sergilenen bu yüzeysel görünümü her sözcük türü için daha ayrıntılı olarak da irdeleyebiliriz. Örneğin, belirteç imli sözcükleri 16 ulamda inceleyebilir ve yazılı aradildeki dağılımlarını gözlemleyebiliriz. Bu dağılımı örneklemekten önce belirteç ulamlarının imlenmesinde kullanılan CLAWS7 etiketlerinin tanımlamalarını örneklerle Çizelge 10'da açıklayalım.

Çizelge 10: Belirteç ulamlarına ait CLAWS7 etiketleri

Etiket	Tanım	Örnek
RR	Genel belirteç	<i>rapidly</i>
RG	Ölçü belirteci	<i>very, so, too</i>
RT	Zaman belirteci	<i>now, today</i>
RP	Parçacık	<i>about, in</i>
RRQ	Wh- genel belirteci	<i>where, when, why, how</i>
RL	Yer belirteci	<i>alongside, forward</i>
RGR	Üstünlük derecesi belirteci	<i>more, less</i>
REX	Koşuntu yapılarını başlatan belirteç	<i>namely, e.g.</i>
RGT	Enüstünlük derecesi belirteci	<i>most, least</i>
RRR	Üstünlük genel belirteci	<i>better, longer</i>
RA	Ad kökenli baş ardından gelen belirteç	<i>else, galore</i>
RRQV	Wh-ever genel belirteci	<i>wherever, whenever</i>
RGQ	Wh- ölçü belirteci	<i>how</i>
RRT	Enüstünlük derecesi genel belirteci	<i>best, longest</i>
RPK	İlgeçli belirteç	<i>be about to içerisindeki about</i>
RGQV	Wh-ever ölçü belirteci	<i>however</i>

Bu imler kullanılarak işaretlenen belirteçlerin ulamlarına göre TICLE içerisindeki dağılımı aşağıda Çizge 2.'de sunulmaktadır.

Çizge 2. TICLE’de belirteçlerin ulamlarına göre dağılımı



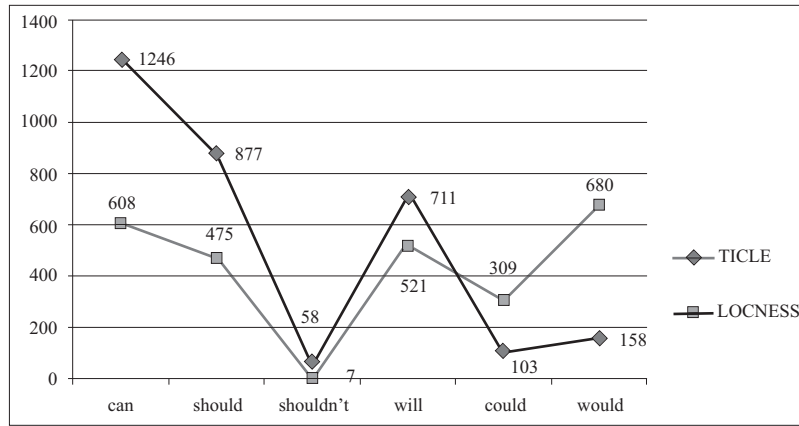
Buna göre, aradil derlemimizin İngilizce belirteçlerin kullanımına ilişkin görünüşü, bize en çok kullanılanın, belirteçlerin toplam sayısı üzerinden bakıldığında, %61.63 ile genel belirteçler olduğunu ve bunu %8.62 ile ölçü belirteçlerinin izlediğini, daha sonra sırasıyla %7.74 ile zaman belirteçlerinin, %5.18 ile parçacıkların, %4.60 ile Wh- genel belirteçlerinin, %3.81 ile yer belirteçlerinin, %3.12 ile üstünlük derecesi belirteçlerinin, %3.10 ile koşuntu yapılarını başlatan belirteçlerin, %2.38 ile en üstünlük belirteçlerinin, %2.15 ile üstünlük genel belirteçlerinin, %1.52 ile ad kökenli baş ardından gelen belirteçlerin, %0.66 ile Wh- genel belirteçlerinin, %0.40 ile Wh- ölçü belirteçlerinin, %0.17 ile en üstünlük derecesi genel belirteçlerinin, %0.06 ile ilgeçli belirteçlerin ve %0.04 ile Wh-ever ölçü belirteçlerinin geldiğini göstermektedir. ICLE içerisinde, bu sonuçları diğer anadil artalanlarından gelen İngilizce aradil kullanıcılarının belirteç kullanımları ile karşılaştırıp sonuçlarını bir aradil ortak özellikleri çerçevesinde değerlendirmek veya benzer anadil artalanlarından gelenlerin ortak paydalarına ulaşmak amacıyla yorumlar yapabilecek çalışmaları yürütmek son derece olanaklıdır.

5.3. Sözcük kullanım örgüsü

Öğrenici derlemleri kullanarak aradil özelliklerinin ortaya çıkarılmasında ve aradilde, anadil kullanıcısının üretimine ne kadar yaklaşıldığını veya ondan hangi noktalarda farklılaşıldığını anlamada kullanılan yöntemlerden birisi de, sözcük kullanım örgüsünü inceleyerek az/fazla kullanım olgularını karşılaştırmaktır. Örneğin, özellikle akademik metinlerde kipliklerin az veya fazla kullanımlarının incelenmesi, yabancı dil olarak İngilizce kullanarak yazan araştırmacıların kuşku ve

belirsizlik durumlarında kullanması gereken kiplikleri seçmede zorlandıklarını göstermektedir (Hyland ve Milton, 1997; Kennedy, 1998; Kilimci ve Can, 2009). Çizge 3., bu açıdan TICLE ve LOCNESS¹² anadil konuşanı derleminin bir karşılaştırmasını sunmaktadır.

Çizge 3. Kipliklerin kullanımları açısından TICLE ve LOCNESS



Kipliklerin TICLE ve LOCNESS derlemlerindeki az / fazla kullanımlarını incelediğimizde İngilizcenin anadil konuşucuları ile Türk aradil konuşucuları arasındaki farkları görebilmekteyiz. Buna göre her iki derlemin bütünü üzerinden oranlara baktığımızda hangi kipliğin fazla hangisinin az kullanıldığını Çizelge 11'de sunmaktadır.

Çizelge 11: Kipliklerin az / fazla kullanım oranları ($p < 0.000001$)

Kiplik	TICLE - %	LOCNESS - %
can	0.74*	0.36
should	0.54*	0.28
shouldn't	0.03*	-
will	0.52*	0.31
could	0.06**	0.18
would	0.09**	0.40

*: çok kullanım

** : az kullanım

Derlem dilbilimsel olarak *açıklama* terimini, Leech (1993: 275)'in “varolan yazılı veya sözlü bir elektronik dil derlemine bir çeşit düzenleme yoluyla eklenen dilbilgisel yorumlayıcı bilgi” tanımı çerçevesinde ele alıp TICLE incelendiğinde,

¹² Louvain Corpus of Native English Essays.

sadece bağlam içerisinde kullanılmış bir sözcük yığını değil aynı zamanda bu sözcükler hakkında bilgiler içeren öğrenici derlemi görülür. Bu da bize Çizelge 12’de *as* sözcüğü üzerinden örneklendiği üzere, öğrencilerin sözcük örüntülemeleri hakkında ayrıntılı bilgi vermekte ve diğer derlemlerle karşılaştırma olanağı sunmaktadır.

Çizelge 12: TICLE içerisinde *as* sözcüğünün örüntülenmesi

Örüntüleme	TICLE	ICLE - JP	LOCNESS
<i>as a conclusion</i>	22	1	0
<i>as far as</i>	11	9	5
<i>as far as X is concerned</i>	0	0	1
<i>as we can see / have seen / see</i>	7	4	0

Çizelge, Türk ve Japon İngilizce öğrencilerinin aradil derlemlerindeki *as* sözcüğünün bazı örüntülemelerini anadil konuşanlarının derlemindekiler ile karşılaştırmaktadır. Buna göre karşılaştırmada TICLE içerisinde 22 sıklıkla dikkat çeken *as a conclusion*’un tanıklı dizinlerini aşağıdaki derlem anlık görüntüsünden inceleyebiliriz.

Şekil 1. Bazı *as a conclusion* tümcelerinin tanıklı dizin anlık görüntüsü

decide to make euthenasia or not. **As a conclusion**, in most aspects it is right to put it into
al problems increase divorce rate. **As a conclusion**, if there are big problems in the house
sexes are in each part of the life. **As a conclusion**, people have different features and s
wer to the environment or people. **As a conclusion**, we should accept that, most of the c
and be ready for every chancing. **As a conclusion** we do not see defending people with
ll have little value in students life. **As a conclusion**, universities are not valueless instituti
an be told there is always a hope. **As a conclusion**, there are some people who are oppo
ng the rest of their lives in prison. **As a conclusion**, people who commit crime should cer
most intensely on their shoulders. **As a conclusion**, the variety and accessibility of contra
ich a concept. What I wan na say **as a conclusion** is, both of the sexes are unique, they
who fought for the whole country. **As a conclusion** men have effective role in every parts

Derlemden alacağımız bu tanıklı dizin dökümü de bize aradilde araştıracağımız yapıların kullanımına yönelik ipuçları vermekte ve anadil olarak İngilizce’ye yakınlıkları üzerine yorum yapmamıza olanak sağlamaktadır.

6. Sonuç

Giriş bölümünde de vurguladığı üzere, bu çalışma, bilgisayar destekli bir Türk öğrenici İngilizcesi derlemi olan TICLE'yi tanıtmayı, oluşturulmasında izlediğimiz yolları anlatmayı ve bu türden bir öğrenici derlemi ile yapılabilecek aradil çalışmalarına örneklerle yol göstermeyi amaçlamaktadır. Granger'in (2008) de belirttiği gibi öğrenici derlemlerinden elde edilen veriler kullanılarak yabancı dil öğretimine yönelik birçok uygulamalar gerçekleştirilmektedir. Bunlar arasında, öğrencilerin aradil hatalarına yönelik açıklamaları içeren sözlükler ve dilbilgisi kitapları sayılabilir. Bu aradil hatası açıklamaları ve örneklemeleri, ileri düzey yabancı dil yeterliliğine sahip öğrencilerin sık karşılaşılan hatalarına yönelik olanları da içermektedir (accept + eylemlilik, people yerine persons, news + çoğul eylem gibi) (Granger, 2008). Bu türlü aradil notlarının yabancı dil öğretiminde hatalara dikkat çekerek ve anadil artalanını da gözönüne alarak, bilinç uyandırma noktasında ere yardımcı olacağı, özellikle yetişkinlere yabancı dil öğretimi konusunda yarar sağlayacağı düşünülmektedir. Ayrıca, öğrenici derlemlerini oluşturan metinlerin dil seviyelerinin çözümlenmesi ile ulaşılabilecek yeterlilik seviyesi görünümüleri, değerlendirme ölçekleri geliştirenlere de gerçekçi veriler sağlayacaktır. Gelecekte, yanlış yapma olasılığı en aza indirilmiş özdevinimli aradil hata imleme yazılımlarının geliştirilmesi ile, öğrenici derlemleri yabancı dil öğretimine ve edinim araştırmalarına daha büyük katkılar sağlayacak hale gelecektir.

Ayrıca, derlem dilbilim alanında çalışan araştırmacıların, farklı anadil artalanlarından gelip, Türkçe'yi yabancı dil olarak öğrenenlerden toplanacak metinlerle TICLE'ye koştur bir derlem oluşturması, karşılaştırmalı çalışmalar yapabilmeleri açısından aradil araştırmacıları için kaynak yaratacaktır.

Kaynakça

- Bonelli, E., & Sinclair, J. (2006). Corpora. In K. Brown, *Encyclopedia of Language and Linguistics* (2nd Edition ed., pp. 206-219). Oxford, United Kingdom: Elsevier.
- Eriksson, A. (2008). *Tense and Aspect in Advanced Swedish Learners' Written English*. Göteborgs: Acta Universitatis Gothoburgensis.
- Ellis, R., & Barkhuizen, G. (2005). *Analysing Learner Language*. Oxford: Oxford University Press
- Ädel, A. (2006). *Metadiscourse in L1 and L2 English*. Amsterdam: John Benjamins.
- Francis, W. N., & Küçera, H. (1979). *Manual of Information to accompany a Standard Sample of Present-day Edited American English, for use with digital computers*. Providence, R.I.: Department of Linguistics, Brown University.

- Fries, C. C. (1951). *The Structure of English*. London: Longman.
- Gilguin, G., Papp, S., & Díez-Bedmar, M. (2008). *Linking Up Contrastive and Learner Corpus*. Amsterdam: Rodopi.
- Granger, S. (1997). Automated Retrieval of Passives from Native and Learner Corpora: Precision and Recall. *Journal of English Linguistics* , 25 (4), 365-374.
- _____. (2008). Learner Corpora in Foreign Language Education. In N. Van Deusen-Scholl, & N. H. Hornberger, *Encyclopedia of Language and Education* (2 ed., Vol. 4, pp. 337-351). Philadelphia, PA, U.S.A: Springer.
- _____. (1998). The Computerized Learner Corpus: a Versatile New Source of Data for SLA Research. In S. Granger, *Learner English on Computer* (pp. 3- 18). London, United Kingdom: Longman.
- _____. (2009). The Contribution of Learner Corpora to Second Language Acquisition and Foreign language Teaching. In K. Aijmer, *Corpora and language Teaching* (pp. 13-32). Amsterdam, The Netherlands: John Benjamins Publishing Company.
- Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (2009). *International Corpus of Learner English: Version 2*. (S. Granger, E. Dagneaux, F. Meunier, & M. Paquot, Eds.) Louvain-la-Neuve, Belgium: Presses universitaires de Louvain.
- Granger, S., Hung, J., & Petch-Tyson, S. (2002). *Computer Learner Corpora, Second Language Acquisition and Foreign Language*. Amsterdam: Benjamins.
- Hyland, K. & Milton, J. (1997). Qualification and certainty in L1 and L2 students' writing. *Journal of Second Language Writing*, 16(2): 183- 205.
- Jespersen, O. (2000). *Modern English Grammar on Historical Principles*. New York, NY: Barnes & Noble. (İlk baskısı: 1954)
- Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. Essex: Addison Wesley Longman.
- Kilimci, A., Can, C. (2009). TICLE: Uluslararası Türk Öğrenci İngilizcesi Derlemi. M. Sarıca, N. Sarıca ve A. Karaca (Ed.), XXII. Ulusal Dilbilim Kurultayı Bildirileri, 1-11, Ankara: Yüzüncü Yıl Üniversitesi Yayınları.
- Leech, G. (1993). Corpus Annotation Schemes. *Literary and Linguistic Computing*, 8 (4), 275-281.
- Martelli, A. (2007). *Lexical Collocations in Learner English: A Corpus-based Approach*. Alessandria: Edizioni dell'Orso.
- Meunier, F. & Granger, S. (2008). *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins.
- Myles, F. (2005). Interlanguage Corpora and Second Language Acquisition Research. *Second Language Research* , 21 (4), 373-391.
- Sinclair, J. M. (2001). Preface to *Small corpus studies and ELT : theory and practice* . In M. Ghadessy, A. Henry, & R. L. L. Roseberry, *Small corpus studies and ELT : theory and practice* (pp. vii-xv). Amsterdam, The Netherlands: John Benjamins B.V.
- Svartvik, J. (2007). Corpus linguistics 25+ years on. In R. Facchinetti, & R. Facchinetti (Ed.), *Corpus Linguistics 25+ Years on* (1st Edition ed., pp. 11- 25). Amsterdam, The Netherlands: Rodopi.