



TÜRKİYE'DE YAPILAN İNGİLİZCE YABANCI DİL SINAVLARININ SÖZCÜK PROFİLİ: ÜDS ÖRNEĞİ*

*İhsan ÜNALDI***

Mehmet BARDAKÇI

ÖZET

Son yıllarda, yabancı dil eğitimi bağlamında, dilbilgisinden çok sözcük bilgisinin ön plana çıkmaya başlaması başka tartışmaları da beraberinde getirmiştir. Bu tartışmalarda hem anadilde, hem de yabancı bir dil öğrenirken sözcük bilgisinin dilbilgisinden önce gelişmeye başlaması ve sözcük bilgisiyle gerçekleştirilebilecek iletişim davranışlarının, dilbilgisiyle gerçekleştirilebilecek iletişim davranışlarından daha fazla olması gibi sebepler pek çok tartışmanın ve çalışmanın konusu olmuştur. Sözcük bilgisiyle ilgili tartışmalarda sık sık gündeme gelen konulardan biri de hangi sözcüklerin öğrenilmeye değer, hangilerinin göz ardı edilebilecek olduğu ile ilgili konudur. Hedef sözcükler akademik bağlamlarda öğrenilmesi gerektiğinde ise durum daha da karmaşık hale gelebilmektedir. İlgili akademik alana göre mi, yoksa ortak bir bilimsel sözcük grubuna göre mi öğrenme stratejilerinin geliştirilmesi gerektiği eğitim süreci başlamadan ortaya konulmalıdır. Doğal olarak, bu tartışmalardan yabancı dil sınavları ve bu süreçlere dahil olan bireyler de etkilenmektedir. Bu çerçevede, ülkemizde yapılmış/yapılmakta olan İngilizce dil yeterlilik sınavlarından biri olan Üniversitelerarası Kurul Dil Sınavının (ÜDS) İngilizce bölümlerinde çıkan okuma parçalarının sözcük profillerinin belirlenmesi bu çalışmanın asıl amacıdır. Bu profilleri belirleyebilmek için 2006-2012 yılları arasında ve Öğrenci Seçme ve Yerleştirme Merkezi tarafından farklı bilim alanlarına yönelik olarak yapılmış olan ÜDS sınavlarının İngilizce bölümlerinde sorulan okuma parçaları kullanılarak 44,674 sözcükten oluşan özgün bir derlem oluşturulmuştur. Bu derlem, araştırma sorusunun oluşturduğu çerçeveye uygun olarak, çevrimiçi bir veri tabanı olan Vocabprofile ve WordSmith yazılım paketi kullanılarak işlenmiştir. Sonuçlar, ilgili yazına paralel olarak, bu sınavlardaki okuma parçalarının sözcük profillerinin ortalama % 85 oranında tahmin edilebilir olduğunu göstermektedir. İlgili okuma parçalarında en sık geçen sözcükler incelendiğinde, bu örtüşmenin aynı zamanda sözcük seviyesinde de anlamlılık gösterdiği belirlenmiştir.

Anahtar Kelimeler: İngiliz dili, sözcük bilgisi, derlem, sözcük profili

*Bu makale Crosscheck sistemi tarafından taranmış ve bu sistem sonuçlarına göre orijinal bir makale olduğu tespit edilmiştir.

** Yrd. Doç. Dr. Gaziantep Üniversitesi Eğitim Fakültesi Yabancı Diller Eğitimi Bölümü, El-mek: ihsanunaldi@gmail.com



VOCABULARY PROFILES OF ENGLISH LANGUAGE PROFICIENCY EXAMS IN TURKEY: THE CASE OF ÜDS

ABSTRACT

In the context of foreign language pedagogy, the foregrounding of vocabulary in recent years, as opposed to grammar, has brought other discussions with it. In these discussions, the development of lexicon before grammar both in first language acquisition and foreign language learning, and the fact that communication acts that can be carried out merely depending on lexicon are far more than those that can be carried out merely depending on grammar have been topics of concern. Another topic of concern in these discussions is about determining the words that are worth learning and the ones that can be ignored. When the target words are to be learned in an academic context, this issue might become even more complicated. Whether a learning program should be developed by making selections according to a specific scientific domain or a more general and a common scientific lexicon is to be decided on before the beginning of any language teaching program. Naturally, not only foreign language exam processes but also the individuals involved in them are influenced from these discussions. In this context, the main aim of the current study is to determine the vocabulary profiles of the reading passages in one of the major language proficiency exams in Turkey, Interuniversity Foreign Language Examination (UDS). In order to be able to determine these profiles, a specialized corpus of 44,674 words was compiled by using the reading passages in the UDS exams between the years 2006 and 2012. In line with the research question, this corpus was processed in an online database, Vocabprofile, and through WordSmith, a software package for corpus-based studies. The results, parallel with the related literature, revealed that vocabulary profiles of the reading passages in UDS are actually predictable at about 85 %. When the most frequent lexical items in the corpus were analyzed, it was determined that this overlap was also valid at vocabulary level.

Key Words: English language, vocabulary, corpus, vocabulary profiles

Giriş

Ülkemizde uygulanmakta olan İngilizce yeterlilik sınavları orta öğretim seviyesinden (Seviye Belirleme Sınavı) başlayıp, lisansüstü seviyeye (Üniversitelerarası Kurum Seviye Tespit Sınavı, Yabancı Dil Bilgisi Seviye Tespit Sınavı) kadar yayılmış durumdadır. Uzun yıllardır yapılmakta olan bu tür İngilizce yeterlilik sınavları biçimlendirici (*formative*) değerlendirmelerden daha çok özetleyici (*summative*) değerlendirme odaklı süreçleri içermektedir. Başka bir deyişle, adaylar yılda bir-kaç kez yapılan bu sınavlarda kısa süreli değerlendirmelere tabi tutulmakta, yabancı dil gelişim evreleri ve bireysel farklılıklar gibi önemli konular göz ardı edilmektedir. Bu tür değerlendirmelerin yabancı dil eğitimi açısından ne kadar sağlıklı olduğu başka bir çalışma konusu olmakla beraber, güncel sınavların incelenmesi de ayrı bir önem taşımaktadır. Bunun yanı sıra, Türkiye’de İngilizce yeterlilik sınavlarının özellikle sözcük profilleri ile ilgili yapılan önemli bilimsel çalışmaların sayısı çok azdır.

Turkish Studies

International Periodical For the Languages, Literature and History of Turkish or Turkic
Volume 9/3 Winter 2014



2001-2012 yılları arasında, yılda iki kez ve son defa 2012 yılında uygulanmış olan Üniversitelerarası Kurul Yabancı Dil Sınavlarının (ÜDS) sözcük profilleri bu çalışmanın kapsamındadır. Diğer yeterlilik sınavlarından farklı olarak, ÜDS sınavlarında adaylara bilim alanlarına göre fen bilimleri, sağlık bilimleri ve sosyal bilimler olarak üç farklı test seçeneği verilmekteydi. Bu sınavla ilgili pek çok şey tartışıldı ve akla ilk gelen sorulardan biri de bu farklılığın sınavın kapsamında olan sözcük gruplarını etkileyip etkilemeyeceğiydi. Bu amaçla, 2006-2012 yılları arasında resmi olarak yayımlanmış olan testlerin İngilizce bölümleri kullanılarak 44,674 sözcükten oluşan özgün bir derlem oluşturulmuştur.

Derlem nedir?

Derlem, İngiliz diline Latin dilinden geçmiş olan *corpus* (çoğ. *corpora*) sözcüğünün Türkçe karşılığıdır. En temel tanım olarak derlem, yazılı ya da sözlü dil örneklerinin bilgisayarlar yardımıyla sayısal ortama veri tabanı mantığı ile aktarılmasıdır (Meyer, 2002: xii). Bu şekilde oluşturulan veri tabanı, çalışılan dilin doğasıyla ilgili sorgulamalar yapmak için kullanılır. Bu mantık ilk olarak 13. yüzyılda kutsal kitapları daha iyi anlamak amacıyla Hıristiyan din adamları tarafından bir dizinleme yöntemi olarak kullanılmaya başlanmıştır. 1960’lı yıllarda, 1 milyon sözcük kapasiteli Brown Derlemi ortak kullanıma açılmış, bilgi işlem sistemlerinin gelişmesiyle bu tür veri tabanları sözcük ve işlem kapasiteleri açısından devasa boyutlara ulaşmıştır. Çağdaş Amerikan İngilizcesi Derlemi (COCA – 500 milyon sözcük, Davies, 2008) ve Ulusal İngiliz Derlemi (BNC – 100 milyon sözcük) bu veri tabanlarının en güncel ve geniş örnekleridir.

Dilbilimsel bağlamda, kullanılan dil ve evrensel dil becerisi arasındaki ilişki üzerinde en çok tartışılan konulardandır. Bir grup araştırmacı, sözcük kapasitesi ne kadar olursa olsun, insanların kullandıkları dil örneklerinin dilin doğasını anlamada hiçbir işe yaramayacağını savunurken (Chomsky, 1965), diğer bazı araştırmacılar, dilin doğasında şans eseri olamayacak sıklıkta tekrarlanan yapılar olduğunu ve dilbilim çalışmalarında bunların incelenmesi gerektiğini savunur (Stubbs, 2004).

Son yıllarda, dil öğrenme sürecinde, dilbilgisinin değil de sözcük bilgisinin ön plana çıkmaya başlaması aslında şaşırtıcı bir durum değildir. Bir dilin dilbilgisel yapısını görmezden gelmek mümkün değildir, fakat dilbilgisi bilmeden sadece sözcük kullanarak da insanlar arası iletişimin yüzeysel de olsa bir şekilde mümkün olduğu da yadsınamaz. Çocuklarda anadil gelişimiyle ilgili yapılan çok sayıda çalışma bu savı destekler niteliktedir. Örneğin, çocuklarda sözcükbilgisinin (*lexicon*) ilk doğum gününden önce ve dilbilgisinin daha derin bir boyutu olan sözdiziminin (*syntax*) tüm dillerde 24 ay civarında başlaması (Guasti, 2002) sözcükbilgisinin dilbilgisinden önce gelişmeye başladığının göstergesi olarak kabul edilebilir. Sözcük bilgisiyle ilgili bu tür tartışmaların yanı sıra, sık sık gündeme gelen tartışmalardan biri de, hem öğrenen hem de öğreten açısından, hangi sözcüklerin öğrenilmeye değer olduğuyla ilgili verilmesi gereken kararlar konusudur (Coxhead, 2000). Hedef sözcükler akademik bağlamlarda öğrenilmesi gerektiğinde ise durum daha da karmaşık hale gelmektedir. İlgili bilim alanına göre mi, yoksa ortak bir bilimsel sözcük grubuna göre mi öğrenme stratejilerinin geliştirilmesi gerektiği eğitim süreci başlamadan ortaya konulmalıdır. Bununla birlikte, herhangi bir bilim dalında uzlaşmaya çalışan bir bireyin, kendi alanıyla ilgili terminolojiye hakim olduğu varsayılırsa sorunun farklı boyutlar kazanacağı çok açıktır.

Çeşit, örnekçe ve sıklık

Derlemin nasıl oluşturulduğunu anlayabilmek ve bu çalışmanın yöntembilimini ve sonuçlarını yorumlayabilmek için çalışmanın bilimsel çerçevesini oluşturan belli başlı terimlerin kapsamlarını bilmek gerekmektedir. Bu terimlerden en başta gelenleri *çeşit* ve *örnekçe*dir (*type/token*). Örnekçe bir cümle yapısı içinde birbirlerinden boşluk ya da bir noktalama işareti ile ayrılan dilbilimsel birimlerdir (Scott, 2012). Bir metinde toplamda 500 sözcük varsa bu metin 500 örnekçeden

Turkish Studies

International Periodical For the Languages, Literature and History of Turkish or Turkic
Volume 9/3 Winter 2014



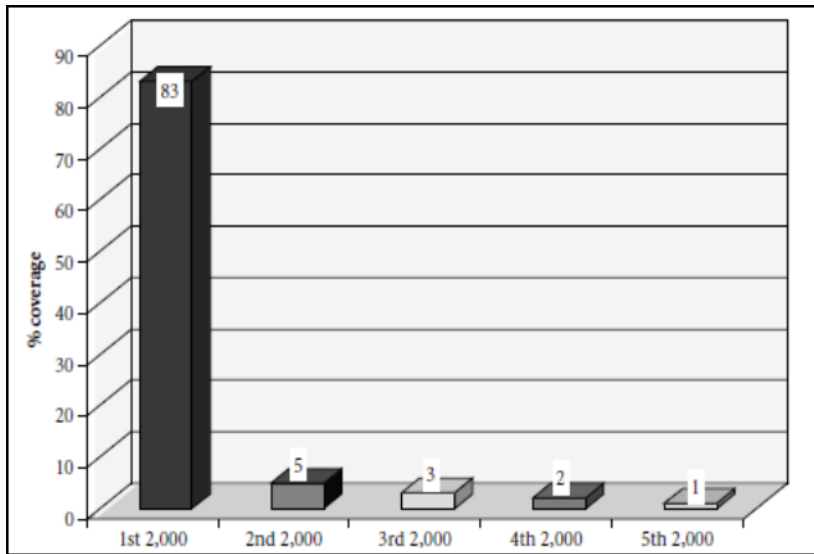
oluşmaktadır. Fakat bu metinde bazı sözcükler tekrarlanır. Birbirlerinden farklı, tekrarlanan bu sözcüklere de *çeşit* denmektedir.

Ör. The bigger they are, the faster they fall.

Yukarıdaki İngilizce tümcede toplamda sekiz sözcük bulunmaktadır. Bu sözcüklerin her biri örnekçe (*token*) olarak algılanır. Fakat tümcede geçen *the* ve *they* sözcükleri ikişer defa tekrarlanmaktadır ve bu sebeple bu tümcede altı çeşit (*type*) bulunmaktadır. Başka bir deyişle, yukarıdaki örnek tümcede sekiz sözcük vardır fakat tekrar eden sözcükleri tek sözcük olarak hesaba kattığımızda sadece altı farklı sözcükten bahsedebiliriz.

Bu çalışmanın ana konusunu oluşturan, *sıklık* (*frequency*) terimi yukarıda değinilen *çeşit* terimiyle doğrudan ilintilidir ve bir sözcüğün herhangi bir bağlamda kaç kez kullanıldığını göstermek için kullanılır (Scott, 2012). Örneğin, yukarıda verilen örnek İngilizce tümcede *the* ve *they* sözcükleri için sıklık değeri 2, diğerleri için bu değer 1'dir.

Bu mantık çerçevesinde, Nation (2001) İngilizce bir derlemi çıkış noktası olarak İngilizce yazılı ve sözlü bağlamlarının % 80'ninden fazlasının, sık kullanılan yaklaşık 2000 sözcükten oluşan ve aynı zamanda General Service List (Genel Hizmet Listesi) olarak da bilinen sözcük grubundan oluştuğunu belirtmiştir. Başka bir deyişle, İngilizce yazılmış herhangi bir metinde örnekçe sayısı ne olursa olsun bu metin kapsamındaki çeşit yaklaşık olarak % 80 oranında aynı sözcüklerden oluşur. Figür 1 incelendiğinde bu durum daha da net anlaşılacaktır.



Figür 1. İngilizce yazılı ve sözlü ifadelerden oluşturulmuş, 10 milyon örnekçe sözcük içeren bir derlemin çeşit kapsamı (O'keeffe et al., 2007)

Figür 1'de, 10 milyon sözcükten oluşan ve hem yazılı hem de sözlü ifadeler içeren bir derlemdaki sözcük dağılımları verilmiştir. İlk sütundan, bu derleminde sürekli tekrarlanan 2000 sözcük derlemin tamamının % 83'ünü oluşturmaktadır; yani bu 2000 sözcük derlem içinde sürekli yinelenmektedir. İkinci ve daha sonra gelen 2000'lik sözcük grupları toplam derlemin sırasıyla sadece % 5, 3, 2 ve 1'ini oluşturmaktadır. İlk grup sözcükten sonra görünen % 83'ten % 5'e düşüş çarpıcı olduğu kadar uygulamalı dilbilimi ve dil eğitimi açısından önem

Turkish Studies

International Periodical For the Languages, Literature and History of Turkish or Turkic
Volume 9/3 Winter 2014

taşımaktadır. Özellikle yabancı bir dil öğrenmeye çalışan bireylerin sözcük bilgilerini geliştirmeye çalışırken yaşadıkları belirsizlikler ve hedef dilin geniş sözcük kapasitesi göz önüne alındığında bu tür verilerin ne kadar kullanışlı olacağı görülecektir.

“Akademik olarak nitelendirilebilecek bir sözcük grubu var mıdır?”

Yukarıda adı geçen ve genel hizmet listesi olarak belirtilen sözcük grubunun dışında alana özel kullanılan Akademik Sözcük Listesi olup olmadığı düşüncesiyle araştırmacılar tarafından çeşitli sıklık çalışmaları yapılmıştır. Coxhead (2000) tarafından oluşturulan ve Academic Word List olarak adlandırılan ve 570 *çeşit* içeren bir grup sözcüğün akademik metinlerde sıklıkla kullanıldığı öne sürülmektedir. Bu sözcükler, Amerika Birleşik Devletleri’ndeki üniversitelerde dört akademik alanda İngilizce olarak okutulan (sanat, fen, tıp ve hukuk) üniversite ders kitaplarından oluşan bir derlem analiz edilerek oluşturulmuştur. Bu sözcük grubunun İngilizce yazılmış herhangi bir akademik metnin yaklaşık % 10’unu oluşturduğu saptanmıştır (Nation, 2005). Aynı yaklaşımla oluşturulan ve daha kapsamlı olduğu iddia edilen başka akademik sözcük listeleri de mevcuttur (ör. The Corpus of Contemporary American English, Davies, 2008). Başka bir deyişle, İngilizce yazılmış herhangi bir bilimsel metinden, sık kullanılan temel sözcükler ve o alanla ilgili teknik terimler çıkarıldığında geriye kalan ortak sözcükler akademik olarak nitelendirilebilir.

İngilizce sözcük profilleri ile ilgili yazın ve Türkiye’de yapılmış/yapılmakta olan İngilizce yeterlilik sınavlarının sözcük profilleri ile ilgili belirsizlikler dikkate alındığında bu çalışmanın kapsamı aşağıdaki araştırma sorusu ile sınırlıdır.

Üniversitelerarası Kurul Dil Sınavı (ÜDS) İngilizce bölümünde çıkan okuma parçalarının sözcük profilleri nedir?

Yöntem

Bu çalışmanın araştırma sorusunu cevaplayabilmek için, 2006-2012 yılları arasında yapılmış ve Öğrenci Seçme ve Yerleştirme Merkezi (ÖSYM) tarafından kurumun internet sitesinde yayımlanmış olan ÜDS sınavlarının İngilizce bölümlerinde sorulan okuma parçaları kullanılarak özgün bir derlem oluşturulmuştur. Bu derlem, araştırma sorusunun oluşturduğu çerçeveye uygun olarak, çevrimiçi bir veri tabanı olan Vocabprofile ve WordSmith (Scott, 2012) adında bir yazılım paketi kullanılarak işlenmiş ve sonuçlar ilgili araştırma çerçevesi doğrultusunda yorumlanmıştır.

Sözcük profilleri ve Vocabprofile

Sözcük profili, yazılı ve/veya sözlü dil örneklerinin doğal bağlamlarından getirdikleri sözcük grupları olarak tanımlanabilir. Örneğin, İngilizce yazılmış bilimsel bir makale, doğası gereği çoğunlukla akademik sözcüklerden oluşacaktır. Anadili İngilizce olan bireylerin konuşmaları yazıya dökülüp incelendiğinde ise bu konuşmalarda geçen sözcüklerin günlük dil gereksinimi karşılamaya yetecek temel sözcüklerden oluştuğu görülecektir. Bu temel farklılık beraberinde belirsizlikleri getirmektedir ve bu durumun yabancı dil olarak İngilizce öğrenmeye çalışan bireyler için sıkıntılar doğuracağı çok açıktır. Türkiye’de 2001-2012 yılları arasında, adayların farklı bilimsel alan ve dillerdeki yeterliliklerinin belirlenmesi için yapılmış olan ÜDS sınavlarındaki İngilizce testlerinin sözcük profillerinin belirlenmesi bu çalışmanın asıl amacıdır. Çalışmamızda bu amaçla Vocabprofile (VP) veri tabanı kullanılmıştır. Bu sistem, çevrimiçi İngilizce bir veri tabanını kullanarak herhangi bir metin girdisini sözcük profillerine göre bölümlere ayırır. Bu veri tabanı Laufer ve Nation (1995) tarafından hazırlanmış olan Sözcük Sıklığı Profillerine dayanmaktadır. Veri tabanına girilen metin daha önce değinilmiş olan sözcük gruplarını temel alarak ilk 1000 (K1), ikinci 1000 (K2), akademik sözcük (AWL) ve liste-dışı (Off-list) sözcükler olarak sınıflandırılır. Bir metinde hangi sözcük gruplarının daha baskın olduğunu ya da bir okuma parçasının hedef bir okuyucu grubuna uygun olup olmadığını kontrol etmek için kullanılır. Örneğin, The New York Times gazetesinde 1 Aralık 2013

Turkish Studies

International Periodical For the Languages, Literature and History of Turkish or Turkic
Volume 9/3 Winter 2014

günü yayımlanmış olan *A commuter train accident in the Bronx kills 4 and injures dozens* (Barron ve Goodman, 2013) başlıklı haber bu veri tabanında işlendiğinde Figür 2’deki gibi bir tablo karşımıza çıkmaktadır.

WEB VP OUTPUT FOR FILE: A Commuter Train Accident				
Words recategorized by user as 1k items (proper nouns etc): NONE (total 0 tokens)				
	Families	Types	Tokens	Percent
K1 Words (1-1000):	163	186	481	75.87%
Function:	(280)	(44.16%)
Content:	(201)	(31.70%)
> Anglo-Sax	(90)	(14.20%)
=Nat. Greco-Lat/Fr Cog:	(90)	(14.20%)
K2 Words (1001-2000):	26	28	41	6.47%
> Anglo-Sax:	(15)	(2.37%)
1k+2k			...	(82.34%)
AWL Words (academic):	24	26	32	5.05%
> Anglo-Sax:	(3)	(0.47%)
Off-List Words:	?	63	80	12.62%
	213+?	303	634	100%

Current profile	
%	Cumul.
75.87	75.87
6.47	82.34
5.05	87.39
12.62	100.00

Words in text (tokens):	634
Different words (types):	303
Type-token ratio:	0.48
Tokens per type:	2.09
Lex density (content words/total)	0.56
Pertaining to onlist only	
Tokens:	554
Types:	240
Families:	213
Tokens per family:	2.60
Types per family:	1.13
Anglo-Sax Index:	70.04%
(A-Sax tokens + functors / onlist tokens)	
Greco-Lat/Fr-Cognate Index: (Inverse of above)	29.96%

Figür 2. İngilizce bir gazetede bir tren kazasını anlatan haberinin VP veri tabanına göre sözcük profili

Figür 2’de bahsedilen haberinin VP veri tabanına göre sözcük profili verilmiştir. Anadili İngilizce olan insanlar için yazılmış ve Amerika Birleşik Devletleri’nde meydana gelen bir kazadan bahseden bu haberin sözcükleri analiz edildiğinde haberde toplamda 634 örnekçenin (*token*) bulunduğu ve tekrar eden sözcüklerden dolayı çeşit (*type*) miktarının da 303 olduğu görülecektir. Sol taraftaki sütunda bulunan değerler daha dikkat çekicidir. Tablonun bu kısmında K1 ve K2 sözcüklerinden bahsedilmektedir; bunlar GSL’deki (General Service List) ilk (K1) ve ikinci (K2) 1000 sözcükleri belirtmektedir. 1k+2k ile başlayan satır incelendiğinde, anadili İngilizce olan insanlar için yazılmış olan bu haberde bulunan 634 sözcüğün % 82,34’ünün Figür 1’de gösterilen sözcük grubuna ait olduğu net olarak görülecektir. AWL (Academic Word List) ile başlayan satırda ise bu haberde geçen akademik sözcüklerin yüzdesi verilmektedir ve bu miktar tam olarak % 5.05’tir. Bu iki grup sözcük toplandığında ise % 87,39 gibi yüksek bir oran ortaya çıkmaktadır. Tablonun en altında *off-list* olarak belirtilen satır ise bu iki listede de bulunmayan sözcükleri içermektedir ve yüzde olarak da % 12,62 olarak hesaplanmıştır (virgülden sonraki basamaklar iki haneye yuvarlanmış durumdadır). Off-list olarak geçen bu sözcük grubu, özel adlar (coğrafya, insan, şirket vb.) ya da günlük hayatta çok fazla kullanılmayan sözcükler içermektedir.

VP yazılımı bugüne kadar çeşitli bilimsel çalışmaların çerçevesini oluşturmuştur. Örneğin, Laufer ve Nation (1995) tarafından yürütülen ve VP yazılımının geçerlilik/güvenirlilik sorunu inceleyen çalışmalarının sonuçları aynı kişi tarafından yazılmış iki metin arasında tutarlı sonuçlar verdiğini, bu veri tabanından elde edilen sayısal değerlerle bağımsız yapılan sözcük testlerinin sonuçları arasında anlamlı bir ilişkinin olduğunu ortaya çıkarmıştır. Sonuçlar ayrıca, bu veri tabanından elde edilen dil yeterlilik değerlerinin diğer yeterlilik testleriyle karşılaştırıldığında daha kapsamlı olduğunu göstermiştir.

Turkish Studies

International Periodical For the Languages, Literature and History of Turkish or Turkic
Volume 9/3 Winter 2014



Meara, Lightbown ve Halter (1997), Quebec’te bir İngilizce kursu boyunca sınıf içi etkileşimlerin kayıtlarının kullanıldığı çalışmalarında bu etkileşimleri sözcük odaklı bir yöntemle kullanarak ilgili veri tabanını (VP) kullanarak incelemişlerdir. Bu incelemeler sonunda vardıkları sonuç ise aslında sınıf içi etkileşimlerin istatistiksel açıdan anlamlı bir kısmının veri tabanında geçen 0-1000 sözcük aralığında gerçekleştiğidir. Çalışmanın ilginç bir başka sonucu ise, 1970 yılında yapılmış olan başka bir çalışmayla karşılaştırıldığında bu sözcük aralığın çok fazla değişiklik göstermemiş olmasıdır.

Meara ve Fitzpatrick (2000) daha önce İngilizce pasif sözcük dağılımını ölçmek için kullanılmış olan klasik sözcük testlerinden (Meara ve Buxton, 1987) yola çıkıp, VP veri tabanını kullanarak aktif sözcük dağılımının daha anlamlı ve ekonomik bir şekilde ölçülebilmesi için bir model geliştirmişlerdir.

Bu çalışmada kullanılan diğer bir yazılım ise WordSmith’tir. Mike Scott (2012) tarafından geliştirilmiş olan bu yazılım, sözcüklerin tümceler içinde ne tür özellikler ya da eğilimler gösterdiklerini araştırmak için kullanılan paket bir yazılımdır. Herhangi bir metinde en sık geçen sözcükleri ya da herhangi bir sözcükle beraber en sık kullanılan diğer sözcükleri (eşdizim) belirlemek için kullanılır. Toplamda 324,000 örnekten oluşan ve verisi anadili İngilizce olan üniversite öğrencilerinden toplanmış olan (Granger, 1998) bir derlem The Louvain Corpus of Native English Essays (LOCNESS) bu yazılımda sözcük sıklığı açısından incelendiğinde Figür 3’teki gibi bir tablo karşımıza çıkacaktır.

N	Word	Freq.	%	Texts	%
1	THE	21,105	6.51	14	100.0
2	TO	10,753	3.32	14	100.0
3	OF	10,728	3.31	14	100.0
4	AND	8,325	2.57	14	100.0
5	A	6,846	2.11	14	100.0
6	IN	6,367	1.97	14	100.0
7	IS	6,307	1.95	14	100.0
8	THAT	4,919	1.52	14	100.0
9	IT	3,219	0.99	14	100.0
10	BE	3,196	0.99	14	100.0
11	FOR	3,145	0.97	14	100.0
12	AS	2,832	0.87	14	100.0
13	THIS	2,806	0.87	14	100.0
14	ARE	2,555	0.79	14	100.0
15	NOT	2,403	0.74	14	100.0
16	HE	2,186	0.67	12	85.71
17	THEY	2,079	0.64	14	100.0
18	HAVE	2,046	0.63	14	100.0
19	WITH	1,909	0.59	14	100.0
20	ON	1,796	0.55	14	100.0
21	BY	1,703	0.53	14	100.0
22	#	1,689	0.52	14	100.0
23	PEOPLE	1,569	0.48	14	100.0
24	HIS	1,564	0.48	12	85.71
25	HAS	1,555	0.48	14	100.0
26	WAS	1,555	0.48	14	100.0
27	THEIR	1,540	0.48	14	100.0
28	WOULD	1,461	0.45	14	100.0
29	BUT	1,298	0.40	14	100.0
30	OR	1,286	0.40	14	100.0
31	AN	1,246	0.38	14	100.0
32	FROM	1,187	0.37	14	100.0
33	MORE	1,169	0.36	14	100.0
34	WHICH	1,137	0.35	14	100.0
35	CAN	1,116	0.34	14	100.0
36	WILL	1,116	0.34	14	100.0
37	THERE	1,080	0.33	14	100.0
38	IF	1,067	0.33	13	92.86
39	ONE	1,056	0.33	14	100.0

Figür 3. 324,000 sözcükten oluşan LOCNESS derleminde en sık tekrarlanan ilk 39 sözcük

Turkish Studies

International Periodical For the Languages, Literature and History of Turkish or Turkic
Volume 9/3 Winter 2014

Figür 3'te adı geçen derlemde bulunan yaklaşık 324,000 sözcükten en sık tekrarlanan ilk 39 sözcük verilmiştir. Figürdeki ilk sütun (Word) sözcüğü, ikinci sütun (Freq.) ilgili sözcüğün sıklığını, üçüncü sütun (%) sözcüğün tüm derlemdeki yüzdesini ve dördüncü sütun ise (Text) bu sözcüğün derlemi oluşturan metinlerin kaç tanesinde geçtiğini yansıtmaktadır. Son sütun, ilgili sözcüğün geçtiği metinlerin derleme göre yüzdesini vermektedir. Örneğin, *the* sözcüğü bu derlemdeki en sık kullanılan sözcüktür; 21,105 defa kullanılmıştır ve tüm derlemin % 6.51'ini oluşturmaktadır. Ayrıca bu sözcük derlemi oluşturan tüm metinlerde geçmektedir (% 100); başka bir deyişle, derlemde içinde bu sözcüğü barındırmayan metin bulunmamaktadır.

Bulgular ve yorum

Yöntem bölümünde bahsedilen yazılımlar kullanılarak 2006-2012 yılları arasında çıkmış ve ÖSYM tarafından resmi olarak internet sitesinde yayımlanmış ÜDS soruları incelenerek, sorulan metinlerin sözcük profilleri ve kullanım sıklıkları ortaya konulmaya çalışılmıştır. Bu amaçla, ilgili dil yeterlilik sınavlarında test edilen okuma parçalarından elde edilen ve toplamda 44,674 sözcükten oluşan derlem Vocabprofile veri tabanında her bir bilim alanı için (fen, sağlık ve sosyal) ayrı ayrı işlenmiş ve sonuçlar Tablo 1'de sunulmuştur.

Tablo 1. ÜDS İngilizce bölümündeki okuma parçalarının bilimsel alanlara göre sözcük profilleri

	Toplam sözcük		K1+K2		AWL		K1+K2+AWL		Liste dışı	
	N	%	N	%	N	%	N	%	N	%
Fen	15117		11959	79,11	1002	6,63	12961	85,74	2156	14,26
Sağlık	14694		11292	76,84	1135	7,72	12427	84,56	2267	15,43
Sosyal	14863		12112	81,49	1115	7,5	13227	88,99	1636	11,01

Tablo 1'de ÜDS'de Fen Bilimleri, Sağlık Bilimleri ve Sosyal Bilimler alanında test edilmiş olan okuma parçalarının sözcük profilleri çıkarılmıştır. Sayısal verilerin gösterildiği sütunların ilkinde derlemde her bir alan için oluşturulmuş olan toplam sözcük sayısı belirtilmektedir. Derlemde Fen Bilimlerinden 15117, Sağlık Bilimlerinden 14694 ve Sosyal Bilimlerden 14863 adet sözcük bulunmaktadır. K1+K2 başlıklı sütun, her bir alanda test edilen okuma parçalarında geçen İngiliz dilinde en sık kullanılan sözcüklerin (ilk 2000) miktarlarını ve tüm derlem içindeki yüzdelik oranlarını yansıtmaktadır. AWL başlıklı sütunda, ilgili okuma parçalarında geçen akademik sözcüklerin miktarları ve bunların tüm derlemedeki yüzdeleri gösterilmektedir. Bir sonraki K1+K2+AWL başlıklı sütun, ilgili okuma parçalarında geçen sözcüklerin ilk 2000 ve AWL grubundaki sözcüklerle örtüşme miktarlarını ve oranlarını göstermektedir. Tablodaki son sütun, okuma parçalarında geçen liste dışı sözcüklerin sayısal değerlerini ve bunlara ait yüzdeleri göstermektedir. Daha önce de değinildiği gibi, bu grupta bulunan sözcükler genellikle özel adları, teknik bağlamlara ait ya da kullanım sıklığı çok düşük sözcükleri içermektedir. Bu tablonun en dikkat çekici sütunu, ilk 2000 ve AWL grubu sözcüklerin toplam miktar ve yüzdelерinin verildiği K1+K2+AWL başlıklı sütunu olduğu söylenebilir. Bu sütun incelendiğinde, ilgili sınavda çıkan okuma parçalarında geçen sözcüklerin en az % 85'lik bir bölümünün İngiliz dilinde en sık kullanılan ilk 2000 ve akademik bağlamlarda en sık kullanılan 570 kadar sözcükten meydana gelen iki gruptan oluştuğu anlaşılacaktır.

Turkish Studies

International Periodical For the Languages, Literature and History of Turkish or Turkic
Volume 9/3 Winter 2014



İlgili dil yeterlilik sınavlarında test edilen okuma parçalarında geçen sözcüklerin daha önce başka veri tabanları kullanılarak oluşturulmuş (LOCNESS) sözcük listelerinde en çok kullanılan sözcüklerle örtüşüp örtüşmediğini tespit etmek amacıyla aynı derlem WordSmith paket yazılımıyla işlenmiş ve sonuçlar Tablo 2’de sunulmuştur.

Tablo 2. LOCNESS derleminde ve ÜDS okuma parçalarında en sık geçen ilk 50 sözcük

LOCNESS	Fen Bilimleri	Sağlık Bilimleri	Sosyal Bilimler
<i>the</i>	<i>the</i>	<i>the</i>	<i>the</i>
<i>to</i>	<i>of</i>	<i>of</i>	<i>of</i>
<i>of</i>	<i>to</i>	<i>to</i>	<i>and</i>
<i>and</i>	<i>and</i>	<i>and</i>	<i>to</i>
<i>a</i>	<i>a</i>	<i>in</i>	<i>in</i>
<i>in</i>	<i>in</i>	<i>a</i>	<i>a</i>
<i>is</i>	<i>is</i>	<i>is</i>	<i>that</i>
<i>that</i>	<i>that</i>	<i>that</i>	<i>#</i>
<i>it</i>	<i>#</i>	<i>for</i>	<i>is</i>
<i>be</i>	<i>as</i>	<i>#</i>	<i>for</i>
<i>for</i>	<i>are</i>	<i>as</i>	<i>it</i>
<i>as</i>	<i>for</i>	<i>are</i>	<i>was</i>
<i>this</i>	<i>by</i>	<i>be</i>	<i>as</i>
<i>are</i>	<i>from</i>	<i>from</i>	<i>by</i>
<i>not</i>	<i>it</i>	<i>it</i>	<i>are</i>
<i>he</i>	<i>on</i>	<i>with</i>	<i>with</i>
<i>they</i>	<i>be</i>	<i>or</i>	<i>be</i>
<i>have</i>	<i>have</i>	<i>by</i>	<i>have</i>
<i>with</i>	<i>this</i>	<i>an</i>	<i>on</i>
<i>on</i>	<i>with</i>	<i>can</i>	<i>this</i>
<i>by</i>	<i>which</i>	<i>this</i>	<i>his</i>
<i>#</i>	<i>but</i>	<i>which</i>	<i>from</i>
<i>people</i>	<i>or</i>	<i>have</i>	<i>not</i>
<i>his</i>	<i>has</i>	<i>on</i>	<i>but</i>
<i>has</i>	<i>they</i>	<i>people</i>	<i>their</i>
<i>was</i>	<i>at</i>	<i>but</i>	<i>an</i>
<i>their</i>	<i>an</i>	<i>has</i>	<i>has</i>
<i>would</i>	<i>their</i>	<i>blood</i>	<i>they</i>
<i>but</i>	<i>can</i>	<i>not</i>	<i>its</i>
<i>or</i>	<i>its</i>	<i>they</i>	<i>more</i>
<i>an</i>	<i>other</i>	<i>may</i>	<i>which</i>
<i>from</i>	<i>was</i>	<i>cells</i>	<i>one</i>
<i>more</i>	<i>into</i>	<i>most</i>	<i>or</i>
<i>which</i>	<i>more</i>	<i>body</i>	<i>people</i>
<i>can</i>	<i>been</i>	<i>more</i>	<i>were</i>
<i>will</i>	<i>scientists</i>	<i>brain</i>	<i>been</i>
<i>there</i>	<i>earth</i>	<i>disease</i>	<i>at</i>
<i>if</i>	<i>could</i>	<i>these</i>	<i>world</i>

Turkish Studies

International Periodical For the Languages, Literature and History of Turkish or Turkic
Volume 9/3 Winter 2014

<i>one</i>	<i>not</i>	<i>was</i>	<i>he</i>
<i>at</i>	<i>years</i>	<i>when</i>	<i>these</i>
<i>I</i>	<i>than</i>	<i>such</i>	<i>new</i>
<i>all</i>	<i>will</i>	<i>their</i>	<i>other</i>
<i>many</i>	<i>most</i>	<i>at</i>	<i>however</i>
<i>we</i>	<i>one</i>	<i>all</i>	<i>when</i>
<i>who</i>	<i>some</i>	<i>other</i>	<i>also</i>
<i>also</i>	<i>water</i>	<i>some</i>	<i>most</i>
<i>because</i>	<i>when</i>	<i>cancer</i>	<i>only</i>
<i>when</i>	<i>such</i>	<i>if</i>	<i>Europe</i>
<i>these</i>	<i>these</i>	<i>than</i>	<i>than</i>
<i>been</i>	<i>also</i>	<i>health</i>	<i>there</i>

LOCNESS veri tabanında ve üç farklı bilim alanında sorulan okuma parçalarında en sık geçen ilk 50 sözcüğü içeren Tablo 2 incelendiğinde, LOCNESS'te en sık kullanılan sözcüklerle bilim alanları arasındaki sözcük paralelliğinin dikkat çekici olduğu görülecektir. Tabloda sunulan sözcükler, içerikli sözcüklerden (content words) daha çok, ağırlıklı olarak dilbilgisi ile ilgili sözcüklerden (function words) oluşmaktadır. # işareti sayısal değerleri temsil etmektedir. Tabloda geçen *into*, *been*, *brain*, *disease*, *new*, *world* ve *water* gibi sözcüklerin dışında, bilim alanları arasında tam bir örtüşme vardır. Örtüşmeyen bu sözcüklere ise, kullanım sıklığı farklılığından dolayı, listenin devamında (ilk 50 sözcükten sonra) rastlanmaktadır.

Tabloda dikkat çeken başka bir durum ise en sık kullanılan sözcüklerin aslında İngiliz dilindeki en önemli dilbilgisi yapıları olmasıdır. Örneğin, *he*, *this*, ve *they* gibi sözcükler zamir, *in*, *on* ve *at* gibi yapılar edatlar, *a*, *an*, ve *the* gibi sözcükler ise tanımlık başlıklarını işaret etmektedir. Dil eğitimi sürecinde hangi dilbilgisi yapılarının ön plana çıkması gerektiği de bu tür listelerden çıkarılabilir.

Tartışma ve sonuç

Dil eğitimi süresince sözcük dağarcığının bir şekilde genişletilmesi gerektiği tartışmaya pek açık bir konu değildir. Aslında, *dil sözcük odaklı öğrenilir* savının geçerliliğini kabullenmek için dil eğitim uzmanı olmaya bile gerek yoktur. Kendi anadilimizi öğrenirken bile, dilbilgisi kuralları resmi olarak anadilimize tam hakim olmaya başladığımız 12-13 yaşlarımızdan sonra karşımıza çıkmaya başlar. Yani önce anadilimizi ve bu dile ait kuralları tam olarak öğrenmeye başladıktan sonra bu kurallara ad koymaya, günlük yaşamımızda onları ayırt etmeye başlarız. O zaman, nasıl oluyor da yabancı bir dil öğretmeye çalışırken bazen anadili İngilizce olan insanların bile açıklamakta zorlanacağı dilbilgisi kurallarını bu dili öğrenmeye çalışan insanların önüne koyabiliyoruz? Ya da, İngilizce yeterlilik sınavlarına girmek isteyen ve bu amaçla yabancı dil öğrenmeye çalışan insanlara altından kalkılmayacak sözcük listeleri nasıl tavsiye edilebiliyor? Konuyla ilgili uzmanlık gerektiren bölüm bu tür sorulara tutarlı cevaplar vermeye çalışırken başlıyor. Sözcük odaklı dil öğrenme yaklaşımında akla ilk gelecek soru büyük olasılıkla '*Hangi sözcükler?*' sorusudur. Bu sorunun cevabının da '*Herhangi bir sözcük*' olmayacağı açıktır. İngiliz dilinde, bazı kaynaklara göre yaklaşık 250,000 (Oxford Dictionaries, 2013), başka kaynaklara göre de (Crystal, 2004: 119) en az bir milyon ana sözcük vardır. Böyle bir dili sınav ya da herhangi başka bir amaç için öğrenen insanlar için bu miktarda sözcüğün öğrenilmesi gereken ve göz ardı edilebilir şeklinde sınıflara ayrılması zorunluluğu çok açıktır. İngiliz dilinde yapılan sınavlar için 10 binlerce sözcük hedefi koymak gerçekçilikten uzaktır. Kendi anadilinde ortalama en fazla beş bin sözcüğü tanıyabilen ve kullanabilen bir üniversite mezununa yabancı bir dilde en fazla kaç sözcük öğretilir? Kaldı ki,

Turkish Studies

International Periodical For the Languages, Literature and History of Turkish or Turkic
Volume 9/3 Winter 2014



anadili İngilizce olan üniversite mezunları üzerine yapılan bir araştırma, katılımcıların pasif sözcük dağarcığının 20 bin civarında olduğunu belirlemiştir (Nation ve Waring 1997). Sözcük öğretimi ile ilgili yapılan çalışmalarda da önceliğin sıkça kullanılan sözcüklere verilmesi gerektiğinden, bu sözcüklere dikkat çekilmesi gerektiğinden ve tekrar tekrar öğrencilerin karşılaşabileceği metinler verilmesi gerektiğinden, buna karşın kullanım sıklığı düşük olan sözcüklerin ise göz ardı edilebileceğinden bahsedilmektedir (Nation, 2005). Bu ve buna benzer çalışmalardan çıkan sonuçlar bu tür belirsizlikleri ortadan kaldıracak ve yabancı dil eğitim süreçlerinin planlanmasına katkı sağlayacaktır.

Öneriler

Elde edilen veriler ışığında Genel Hizmet Listesi ve Akademik Sözcük Listesi öğrenildiğinde bu sınavlarda sorulan metinlerin %85 gibi önemli bir kısmının anlaşılabilirliği söylenebilir. O halde, bu çalışmanın sonuçlarından çıkan sözcük profilinin tamamı ya da bir kısmı kullanılarak oluşturulacak bir yabancı dil eğitim programı deneysel bir çalışmanın bilimsel çerçevesini oluşturabilir. Bu şekilde, ilk olarak sözcük odaklı bir programın alışılmış dilbilgisi odaklı bir programdan anlamlı bir şekilde farklılaşıp farklılaşmadığını net olarak görebiliriz. Ayrıca bu tür bir çalışmayla, öğrenim sürecinin en önemli değişkenlerinden biri olan motivasyonun böyle bir yaklaşımdan olumlu ya da olumsuz etkilenip etkilenmeyeceğini ortaya çıkarabiliriz.

Bu çalışmada oluşturulmuş olan çerçeve, aynı zamanda, ülkemizde yapılmakta olan diğer dil yeterlilik sınavlarının sözcük profillerini belirlemek için de kullanılıp daha sonra tüm bu analizlerin sonuçları bir meta-analize tabi tutulabilir ve bu şekilde ülkemizdeki dil yeterlilik sınavlarıyla ilgili, en azından sözcük bilgisi açısından öğrenilmesi gereken hedef sözcükler belirlenerek bahsedilen belirsizliklerin belli bir oranda ortadan kalkmasına imkân sağlanabilir.

KAYNAKÇA

- BARRON, J. & GOODMAN, J. D. (2013). A commuter train accident in the Bronx kills 4 and injures dozens. *The New York Times*. Retrieved [December 21, 2013] from <http://www.nytimes.com>
- CHOMSKY, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- COXHEAD, A. (2000). A new academic word list. *TESOL Quarterly*. 34(2), 213–38.
- CRYSTAL, D. (2004). *The Cambridge encyclopedia of the English language*. Cambridge: Cambridge University Press.
- DAVIES, M. (2008). The Corpus of Contemporary American English: 450 million words, 1990-present.
- GRANGER, S. (1998). *Learner English on Computer*. London & New York: Addison Wesley Longman.
- GUASTI, M. T. (2002). *Language Acquisition: The growth of Grammar*. London: The MIT Press
- LAUFER, B. & P. Nation (1995). Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics* 16 (3): 307-322.
- MEARA, P. & BUXTON, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing*. 4, 142-154.
- MEARA, P., LIGHTBOWN, P. M., & HALTER, R. (1997). Classrooms as lexical environments. *Language Teaching Research*. 1(1) 28-46.

Turkish Studies

International Periodical For the Languages, Literature and History of Turkish or Turkic
Volume 9/3 Winter 2014



-
- MEARA, P. & FITZPATRICK, T. (2000). Lex30: An improved method of assessing productive vocabulary in an L2. *System*. 28 (1), 19-30.
- MEYER, C.F. (2002). *English Corpus Linguistics: An Introduction*, Cambridge: Cambridge University Press.
- NATION, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- NATION, I. S. P. (2005). Teaching and learning vocabulary. In E. Hinkel (Ed.), *Handbook of research on second language teaching and learning* (pp. 581–596). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- NATION, P. (2005). Teaching Vocabulary. *Asian EFL Journal*. 7(3), 253-259.
- NATION, P. & WARING, R. (1997). Vocabulary size, text coverage, and word lists. In N. Schmitt & M. McCarthy (Eds). *Vocabulary: Description, Acquisition, Pedagogy* (pp. 6-19). New York: Cambridge University Press.
- O'KEEFFE, A., MCCARTHY, M. J. and CARTER, R. A. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge: Cambridge University Press.
- Oxford Dictionaries, <http://www.oxforddictionaries.com/words/how-many-words-are-there-in-the-english-language>. Retrieved [December 4, 2013]
- SCOTT, M. (2012). *WordSmith Tools version 6*. Liverpool: Lexical Analysis Software.
- STUBBS, M. (2004). Language corpora. In A. Davies & C. Elder (Eds.), *The handbook of applied linguistics* (pp. 106–132). London: Blackwell Publishing.

Turkish Studies

International Periodical For the Languages, Literature and History of Turkish or Turkic
Volume 9/3 Winter 2014

