

KORECE ÜZERİNE YAPILMIŞ DERLEM ÇALIŞMALARI VE YÖNTEMLERİ

Nadir Engin UZUN*
Yang-Hee LEE**

Öz: Korece, üzerinde önemli derlem çalışmaları yapılmış dillerden biridir. Bu dilin, devlet kurumlarının desteği ile Kore üniversite ve araştırma merkezleri tarafından hazırlanmış oylumlu derlem çalışmaları bulunmaktadır. Bu yazıda Korecenin genel sözcüğü özelliklerinden hareketle bugüne kadar ortaya konmuş Korece derlem çalışmaları hakkında bilgiler aktarılacaktır.

Anahtar kelimeler: Sözcüğü, derlem, Korece

* Prof. Dr., Okan Üniversitesi İnsan ve Toplum Bilimleri Fakültesi Çeviri Bölümü

** Doktora öğrencisi; Ankara Üniversitesi Dil-Tarih ve Coğrafya Fakültesi Dilbilim Bölümü

SOME STUDIES AND METHODS ABOUT KOREAN CORPUS

Nadir Engin UZUN*
Yang-Hee LEE**

Abstract

Korean is a language that has many valuable corpus issues. There are wide and detailed corpora products on Korean, supported by the several governmental organizations and achieved by the universities and research centers. In this paper, after the characteristic features of Korean vocabulary first, then some short information will be given for the some important corpora studies put forwarded till now.

Key words: Vocabulary, corpus, Korean

Giriş

“Derlem (corpus)” geniş anlamda koleksiyon demektir. Türk Dil Kurumu Türkçe Sözlüğü derlemi biraz daha dar anlamıyla, dile odaklanarak da tanımlar: “Bir dilin türlü kullanım alanlarından derlenmiş örneklerinin dil bilgisi ve kuramsal dil bilimi araştırmalarında kullanılmak üzere bilgisayar tarafından okunabilecek biçimde bir araya getirilmiş kümesi”. Bu iş, yani dile veya dillere ait yazılı ve sözlü örneklerin bir araya getirilmesi işi artık son derece gelişmiş elektronik programlar aracılığıyla ve uluslararasılaşmış standartlar çerçevesinde yürütülmektedir. Öyle ki, dilin bilimsel olarak incelenmesi ile ilgilenen dilbilim alanında “Derlem Dilbilimi” başlı başına bir disiplin olmuştur.

Derlemler dillerin kullanım örneklerinin topluca değerlendirilebileceği ürünleri olduğundan, bu ürünlerden hareketle bugün yalnızca dillerin kullanım özelliklerini değil, onların yapılarını, işleyişlerini de daha iyi anlayabilmekteyiz. Dünyanın belli başlı dillerinin çoğu için derlem çalışmaları yapıldığını görüyoruz. Sözlük ve dilbilgisi kitaplarının hazırlanmasından yabancı dil öğretim tekniklerinin geliştirilmesine, ana-dili eğitiminden çeviribilimine uzanan alanlarda başvurulan derlemler, bu dillerin en sık, en yaygın kullanım özelliklerini keşfetmemizi de sağlamaktadır.

* Prof. Dr.; Okan University Faculty of Humanities and Social Sciences Translation Studies

** Doctorate student, Ankara University Faculty of Language-History and Geography the Department of Linguistics

Korece

Derlemler yoluyla oldukça ileri tekniklerle koleksiyonlar oluşturulan dillerden biri de Korecedir. Kore Yarımadası'nda konuşulan Korece, dünyada başka hiçbir dilde bulunmayan bir özelliğe sahiptir: 1948 yılından itibaren Kuzey ve Güney Kore olarak iki devlete bölünmüş bir milletin ortaklaşa konuştuğu tek dildir. Diller hakkında en geniş ansiklopedik bilgileri sunan Ethnologue'a göre, (yarımada'nın güneyindeki Ceju Adası'nda konuşulan ve ölmekte olan dili saymazsak) Korece bir ülkede konuşulan yegane dil olma özelliğine sahip, neredeyse tek dildir (Bkz. <https://www.ethnologue.com/country/KR/languages>). Sözelimi Türkiye 36 farklı dile ev sahipliği yaparken dünyada en çok dilin konuşulduğu ülke, 840 dille Papua Yeni Gine'dir. Bu özelliği ile Kore'de yerel dillerden oluşmuş ikidillilik veya çokdillilik, toplumbilimsel bir olgu değildir.

Korecenin başka dillerde pek olmayan ikinci önemli özelliği ise, yaklaşık 500 bin sözcüklük sözcük varlığının %75'ten fazlasının yabancı sözcükler ve bu sözcüklerden türeme sözcüklerden oluşmasıdır. Ulusal Kore Dil Enstitüsü tarafından yayınlanan bir rapor göre ("Standart Korece Büyük Sözlük İstatistiksel Analiz Raporu" için bkz. Lee 2002), ödünçlenen sözcüklerin büyük çoğunlu, tarih boyunca Kore Yarımadası'nda süren Çin hakimiyetinin etkisi sonucu Çince'den gelen ve büyük çoğunluğu halen Çin karakterleri ile yazılan "Hanca" sözcüklerdir (Geniş bilgi için bkz. Altundağ, 2012). Aynı Enstitü tarafından hazırlanan "Standart Korece Büyük Sözlük"te toplam 509.076 sözcük yer almaktadır. Bu sözlük 440.594 madde başı ve 68.482 alt madde ile oluşturulmuştur (UKDE, 2002: 17). 2008 yılındaki son baskısında sözcük sayısı 2002 yılı baskısından madde başı olarak 3.600 adet daha fazladır (UKDE, 2008: 74). Koreceyi dünyanın en geniş sözcük varlığına sahip dillerden biri olduğunu göstermesinin altında, sözlüğün kendi sıradışı yapılanmasının da etkisi vardır. "Standart Korece Büyük Sözlük"te ölçünlü dilin yanı sıra Kuzey Kore Korecesine, lehçelere ve eski sözcüklere de yer verilmiştir. Sözlükte yaklaşık 190.000 adet terim ile birlikte, 70.000 adet Kuzey Kore Korecesinden, 20.000 adet lehçelerden alınma ve 12.000 adet de eski olan sözcük bulunmaktadır. Özellikle Kuzey Kore'de kullanılan sözcüklerin sözlüğe alınması yoluyla Güney ile Kuzey Kore arasındaki dil farklılaşmasının olası bir ülkeler birleşmesinden sonra yaşanabilecek dil karmaşasına yol açmasını önlemek amaçlanmıştır. Kuzey Kore'de kullanılan sözcükler, Güney Kore'de hiç bulunmayan veya Güney Kore'de kullanılan sözcüklerle aynı ancak anlamları farklı olan sözcükler arasından seçilmiştir ve bunların sayısı 72.063 olarak sözlüğün %14,1'ini oluşturmaktadır (UKDE, 2002:41-42). Bu rakam, Güney ve Kuzey Kore'de kullanılan sözcük varlığının ne boyutta farklılaştığını da göstermektedir.

Korecenin sözcük varlığına sözcük türleri açısından bakıldığında 335.057 adet (%65.82 oranında) ad bulunduğu görülmektedir. Hanca sözcüklerin oranı, ödünçlemede en çok başvurulan öğeler olarak adlar açısından %80'lere çıkmaktadır. Bununla birlikte Hancanın etkisi içerik sözcükleri düzeyinde kalmamıştır. Örneğin toplam 463 adet

adılın neredeyse yarısı Hancadan gelmez ki bu, Hancanın Korece üzerindeki açık yapısal etkisine de işaret etmektedir. Dilsel yapılardan hareketle Uzun ve Yeon, 2014 tarafından yapılan bir tipolojik ölçümlemede, Altay dili olduğu iddia edilen Korece ile açıkça farklı bir aileden olan Çincenin yapısal benzerliğinin, neredeyse, Korecenin Altay dillerine olan benzerliğine yakın oranda olduğu ortaya konmuştur.

Hanca dışındaki yabancı sözcüklerin oranı, Korecenin sözcük varlığında %5.6'dır. Bu oranın yaklaşık %90'ı İngilizceden gelen sözcüklerden oluşmaktadır. Geri kalanların %30.4'ü Almanca, %18.3'ü Fransızca, sonra da sırasıyla İtalyanca, Latince, Rusça, Yunanca, Japoncadandır. Türkçeden alınan yegane sözcük ise /pasha/ diye okunan "paşa" sözcüğüdür (Bkz. UKDE, 2002; 59).

Korece tıpkı Türkçe gibi sondan eklemeli bir dildir ve tümcede doğal sözcük sıralanışı, Özne-Nesne-Yüklem biçimindedir. Türkçede olduğu gibi pekiştirme ve ikilemeler ile yansıma sözcükler, sözcük varlığında önemli bir yer tutar. Ne var ki bolca türetim ve çekim eki olmasına karşın Korecede, Türkçedeki gibi Özne-Yüklem veya Tamlayan-Tamlanan uyumu bulunmamaktadır; yani eylem üzerinde kişi-sayı ekleri, tamlamalarda ve bileşik sözcüklerde tamlayan-tamlanan ekleri görülmez (Korecenin bu açılardan genel betimlemeleri için bkz. Lee, 1994; 85-86). Koreceyi Türkçeden ayıran en önemli özelliklerden biri de, bu dilde özellikle eylem çekimlerini etkileyen çokça saygıdili (*honorofics*) formlarının bulunmasıdır (Geniş bilgi için bkz. Cho, 1999; 268).

Korecenin Derlemleri

Çok sayıda sözlük ve dilbilgisi çalışmalarına konu olmanın yanı sıra Koreceyi çok iyi betimlenmiş diller arasına sokan bir özelliği de, bu dilin çok geniş ve ayrıntılı derlemlere (*corpora*) sahip olmasıdır.

Güney Kore'de Korece üzerine yapılmış ilk önemli biçimbilimsel derlem (*corpus*), Kore İleri Bilim ve Teknoloji Enstitüsü KAIST (*Korea Advanced Institute of Science and Technology*) tarafından 1994-1997 arasında Kore Kültür ve Turizm Bakanlığı ile Bilim ve Teknoloji Bakanlığının destekleriyle oluşturulmuş KAIST derlemidir. Bu derleminde sözcüklerin türleri de etiketlenmiştir. Bu açıdan Kore'de yapılan ilk biçimbilimsel analiz derlemi olarak önemli olduğu söylenebilir. Derlem 1998 yılında Bilim ve Teknoloji Bakanlığının desteğiyle yeniden gözden geçirilmiştir (Kim, 2002:9). Bu ilk oylumlu derlemle hemen hemen aynı tarihlerde hazırlanan bir başka derlem olan, Goryo Derlemi, 1995 yılında Kore Ulusal Araştırma Vakfı desteğiyle Goryo Üniversitesine bağlı Kore Çalışmaları Araştırma Enstitüsü tarafından yapılmış olup 1970'li ve 1990'lı yıllar arasında yayımlanan eserleri kapsayan 100 milyon sözcükten oluşmaktadır. Bunlardan 3 milyonu, sözcük türü etiketlenmesine sahiptir.

Yonsei Derlemi:

İlk oylumlu denemelerden sonra gerçekleştirilen en önemli derlemlerden biri, Güney Kore Yonsei Üniversitesi tarafından 1960'lardan 1990'lı yılların ortasına kadar Ko-

re’de yayımlanan eserleri ve sözlü dil ürünlerinden çeşitli konuları ve türleri içermek üzere hazırlanan Yonsei Derlemi’dir. Toplam 88,790,000 sözcükten oluşan, yazılı ve sözlü örnekler üzerine yapılan geniş kapsamlı bir derlem olan bu derlemde ilkököl ders kitaplarında kullanılan 810.000 sözcük ve Korece eğitimi ile ilgili 1 milyon sözcük, türlerine göre ayrı ayrı etiketlenmiştir.

Yonsei Üniversitesi Dil ve Bilgi Çalışmaları Enstitüsü tarafından, ayrıca bir “Yonsei Korece Sözlüğü” oluşturmak için 1980’li yılların sonundan hazırlanmaya başlanan derlem bilgisayar ortamında kullanıma açıktır. Sözlük oluşturulurken bu derlemden sözcüklerin gerçek kullanım örnekleri seçilmiş ve bu kullanımlara göre sözcük tanımları verilmiştir. <https://ilis.yonsei.ac.kr> internet adresinde de ulaşılabilen bilgilere göre derlemin yıllara ve alanlara göre aşamaları şöyledir:

No	Ad	İçerik	Sözcük sayısı
1	Yonsei Derlemi 1 (Dengeli derlem)	Yonsei derlemlerinden ilkidir. Bu derlem, ‘Günümüz Korelilerinin Okuma Gerçekliği’ konulu anket araştırmasından faydalanılarak yapılan dengeli ve örnek metin derlemidir.	2,9 milyon
2	Yonsei Derlemi 2 (Dengeli derlem)	Bu derlemden kitap ödünç verme sıklığına bağlı olarak dengeli derlem oluşturmak amacıyla konulara göre sözcükler seçilmiştir. Bilgi ve belge yönetimi teknikleri uygulanarak sözcüklerin kapsamlı biçimde seçilmesi hedeflenmiştir. Dewey Onlu Sınıflama sistemine dayanarak Korece metinler 10 başlık altında sınıflandırılmıştır. (Genel konular, Felsefe, Din, Toplum bilimleri, Fen bilimleri, Dilbilim, Doğa Bilimleri ve Matematik, Uygulamalı Bilimler, Sanat, Edebiyat, Tarih). Böylece yukarıda anılan her bir konudan sözcüklerin toplanması sağlanmıştır. Ayrıca kitap ödünç verme sıklığının sözcük temsilini gösterebileceğini varsayarak 1987’den 1988 yılına kadar yayımlanan kitaplar arasında ödünç verme oranı yüksek olanlardan 234 kitap örnekleme seçilmiştir. Bu verilerden, 1988-1990 yılları arasında bir derlem oluşturulmuş ve konulara göre dağılımına bakıldığında genel konular (% 7,8), felsefe (% 9,9), din (%10,7), sosyal bilimler (% 12,8), dilbilim (% 5,7), doğa bilimleri ve matematik (% 11), uygulamalı bilimler (% 11,7), sanat (% 8,1), edebiyat (% 11,2) ve tarih (% 11,3) olarak gösterilmiştir.	1,1 milyon
3	Yonsei Derlemi 3 (1980 yılı derlemi)	Sözcüklerin temsilinin insanların daha sık okuduğu metinlerden yansıdığı düşünülerek örneklemin oluşturulmasında 1980’li yılların en iyi yayın listesinden faydalanılmıştır.	5,98 milyon

4	Yonsei Derlemi 4 (1990'lar sözlü derlemi)	İçerik, gerçek sözlü dil kaydı ve tiyatro, yayın senaryosu, senaryo gibi örnekler alınarak, konuşmalar (% 26), ders anlatımları (% 24), danışma konuşmaları (% 14), senaryolar (% 13), DJ yayınları (% 13), tartışmalar (% 8) ve toplantılar (% 2) kullanılarak hazırlanmıştır. Diğer derlemlerden farklı olarak konuşmacının yaş, cinsiyet, meslek, konuşma katılım sayısı ve konuşmanın özelliği gibi bilgiler dahil olmak üzere kayıt yapan kişi hakkındaki bilgiler ve kayıt tarihi de ek olarak gösterilmiştir.	770.000
5	Yonsei Derlemi 5 (1980'ler)	İçerik, 1970'lerde yayımlanan gazeteler (% 10), roman ve denemeler (% 50), genel kitaplar (% 35), ders kitapları (% 5) kullanılarak hazırlanmıştır.	8,6 milyon
6	Yonsei Derlemi 6 (1960'lar)	Derlemdede "Kurtuluş" döneminden sonraki Koreceyi daha iyi göstermek amaçlanmış ve 1960'lı yıllara ait eserler üzerine derlem yapılmaya başlanmıştır. Derleminin çalışmaları devam etmekte olup tamamlandığında boyutunun 10 milyon sözcük olması hedeflenmiştir.	7,23 milyon
7	Yonsei Derlemi 7 (1990'lar)	1990'lı yılların ortasına kadar yayımlanan roman ve denemelerden oluşturulan derlemdir ve 1994-1995 yılları arasında yapılmıştır.	13,67 milyon
8	Yonsei Derlemi 8 (ders kitapları)	İçeriği, bütün ilkökul ders kitapları, ortaokul ve lise ders kitapları içinden Korece ve sosyal bilgiler ders kitapları kullanılarak oluşturulan derlemdir. Kore eğitim programının 5'inci ve 6'ncı dönemleri için hazırlanmış iki versiyonu bulunur.	870.000
9	Yonsei Derlemi 9 (çocuk kitapları)	Öz Korecenin kullanım sıklığı yüksek olan ve tümce açısından nispeten Korece tümce yapısının doğru olarak kullanıldığı çocuk kitapları arasından seçilen örneklerden 1996 yılında oluşturulmuştur.	1,5 milyon
10	Yonsei Derlemi 10 (1945~1965 Kitaplar)	"Yonsei Çağdaş Korece Sözlüğü" nün hazırlanması sürecinde destekleyici olarak 1945-1965 yılları arasında yayımlanan kitaplar kullanılarak oluşturulan derlemdir.	780.000
11	Yonsei Derlemi 11 (1945~1965 Ders kitapları)	"Yonsei Çağdaş Korece Sözlüğü" nün hazırlanması sürecinde destekleyici olarak 1945-1965 yılları arasında yayımlanan kitaplar kullanılarak oluşturulan derlemdir.	730.000

Yonsei derlemlerinde yaklaşık 1,81 milyon sözcük, sözcük türleri açısından etiketlenmiştir. Bunlar temel eğitim için oluşturulan, yani ilkökul ders kitaplarında kullanılan 810.000 sözcükten oluşan derlem ile 1 milyon sözcükten oluşan Korece eğitimi derlemidir. Bu derlemlerin analizinde üç düzeyde toplam 17 adet etiket yoluyla sözcükler çözümlenmiştir. Etiketlemelerin kodlaması aşağıda gösterilmiştir (Bkz. Kim, 2002:21)

Birinci ayırım	İkinci ayırım	Üçüncü ayırım	Etiket
Gövdeler	ad	genel ad	NN
		özel ad	NNX/NNF
		bağımlı ad	NX
	adıl		NP
	sayı adı		NR
Çekimleyenler	eylem		VV
	sıfat		VJ
	koşaç		VC
	yardımcı çekimleyenler		VX
Niteleyiciler	çekimlemeyen sıfat		AN
	belirteç		AV
İlişki ekleri	durum ekleri		PA
	çekim ekleri		EF
Bağımsızlar	ünlem		EX
Bileşik biçimler	kısaltma		XX
türetim ekleri	son türetim eki		SX
Noktalamalar ve diğerler			SP/SY

Sejong Derlemi:

Korecenin hiç şüphesiz en büyük ve en yeni derlemi, uzun vadeli bir proje olan “21. Yüzyıl Sejong Projesi” adı altında 1998-2007 yıllarını kapsayan ve bilgisayar ortamına taşınarak oluşturulan Sejong Derlemi’dir. Bu projenin ana amacı, başta elektronik sözlüklerin hazırlanması gibi Korece üzerine çalışma yapacaklar için bir ana kaynağın oluşturulmasıdır. Kore Kültür ve Turizm Bakanlığı’nun desteğiyle hazırlanan Sejong Derlemi’nde 2001 yılına kadar 5,5 milyon sözcük için biçimbilimsel (*morphological*) analiz yapılmıştır. Bu derlem Ulusal Kore Dil Enstitüsü tarafından “Standart Korece Büyük Sözlüğü” oluşturmak için de kullanılmıştır.

Projede Goryo Üniversitesi’ne bağlı Kore Araştırma Enstitüsü’nün elektronik metin araştırmacıları, ham derlem (*raw corpus*), biçimcik analizi (*morph analysis*) derlemi, anlam/duygu analizi (*sense analysis*) derlemi ve sözdizimsel (*syntactic*) analiz derlemi için ayrı ayrı temel veritabanları hazırlamışlardır. “21. Yüzyıl Sejong Projesi” ürünü olan Sejong Derlemi, 576 dosya ve 15 milyondan fazla sözcükten oluşan bir derlem-dir ve proje yıllarına göre oluşturulan hacimler aşağıdaki gibidir (Bkz. Choe ve Kang, 2007:338)

Oluşturma Tarihi	Hacim
1999	1.509.651
2000	2.023.351
2001	2.063.270
2002	2.530.261
2003	2.015.870
2004	2.025.386
2005	3.025.577
Toplam	15.193.366

Projenin son iki yılı olan 2006 ve 2007 yıllarında ise önceden oluşturulan derlemelerin düzeltilmesi ve geliştirilmesi yapılmıştır.

Sejong Derlemi, içerik olarak kullanılan metin türü ve dağılımı açısından Yonsei Derleminden büyük bir farklılık göstermemektedir ancak Sejong Derlemi'nde Korece eşsesli sözcüklerin ayırt edici işaretlenmesiyle derleme önemli bir yenilik getirilmiştir (Kim ve diğerleri, 2007). Ayrıca biçimbilimsel analizi biçimbirim (*morpheme*) üzerine değil, biçimcik (*morph*) üzerine yapması Sejong Derlemi'nin bir başka ayırt edici özelliğidir (Choe ve Kang, 2007:339) Bu tür analizde altbirimciklerin (*allomorphs*) çözülmesine olabildiğince özen gösterilmiştir. Analiz sırasında kullanılan ölçüt birimi ise, sözcükbirim(*lexem*)'dir, yani sözcük tanıma sözcüğün ayrı yazılmasına dayandırılmıştır. Standart etiketleme kuralları açısından Sejong Derlemi'nde Im ve Song, 1998'de önerilen 'sözcük analiz etiketleme örneği' temel alınmış ve Goryo Üniversitesi Doğal Dil İşleme Araştırma Merkezi tarafından ayrıca oluşturulan ölçüt de onunla birlikte kullanılmıştır. Böylece mevcut analiz yöntemlerindeki eksiklik tamamlanmaya çalışılmıştır. Örneğin ek ve kökün çözümlenmesi için bunların her biri ayrı birer etiket ile kodlanmıştır. Korecede kök, bağımsız bir öge değildir, dolayısıyla tek başına kullanılamaz. Ancak, kökler sözcük yapımında kullanıldığı için sözcüklerin ayrı ayrı etiketlenerek analiz edilmesi gerekmektedir.

Kim ve Kang, 2000 ve Kang ve Kim, 2004 tarafından Sejong Derlemi'nde kullanılmak üzere önerilen standart etiketler ise şöyledir:

Birinci ayırım (İşlevsel)	İkinci ayırım (Sözcük türleri)		Ayrıntılar	
Gövdeler	ad	NN	genel ad özel ad bağımlı ad	NNG NNP NNB
	adıl	NP	adıl	NP
	sayı adı	NR	sayı adı	NR
Çekimleyenler	eylem	VV	eylem	VV
	sıfat	VA	sıfat	VA

◆ Nadir Engin Uzun / Yang-Hee Lee

	yardımcı çekimleyenler	VX	yardımcı çekimleyenler	VX
	koşaç	VC	koşaç-olumlu koşaç-olumsuz	VCP VCN
Niteleyiciler	çekimlemeyen sıfat	MM	çekimlemeyen sıfat	MM
	belirteç	MA	genel belirteç bağlaç belirteç	MAG MAJ
Bağımsızlar	ünlem	IC	ünlem	IC
İlişki ekleri	durum ekleri (Josa)	JK	özne durum eki tümleç durum eki çekimlenmeyen sıfat durum eki objektif durum eki belirteç durum eki vocative durum eki alıntı (quotation) durum eki	JKS JKC JHG JKO JKB JKV JKQ
	yardımcı ekleri	JX	yardımcı ekleri	JX
	bağlaç ekleri	JC	bağlaç ekleri	JC
Bağımlı biçim	çekim ekleri	E	prefinal ending sonlandırma çekim eki bağlama çekim eki ad kaynaşma çekim eki çekilenebilir-sıfat kaynaşma çekim eki	EP EF EC ETN ETM
	ön ek	XP	gövdelerin ön ekleri	XPN
	son ek	XS	ad türetim son ekleri eylem türetim son ekleri sıfat türetim son ekleri	XSN XSV XSA
	kök	XR	Kök	XR
Noktalamalar ve diğerler	noktalama, yabancı sözcükler, hancalar, işaretler, vasayımsal ad, varsayımsal çekimleyenler, rakam, analiz dışı			

Yukarıdaki tabloda görüleceği gibi Sejong Derlemi'nde sözcükler 46 etiket ile çözümlenmiştir. Bu derlem, Yonsei derlemlerinde kullanılan etiketleme sisteminden daha detaylı bir analiz gücüne sahiptir ancak noktalamalar ve diğerleri üzerine 12 adet etiketleme yapıldığı için KAIST derlemine göre analizin daha az ayrıntılı olduğu söylenmektedir (Bkz. Kim, 2002:24). KAIST Derlemi'nde 54 adet, Goryo Derlemi'nde

ise yalın/tek sözcük türleri olarak 61 adet, bileşmiş sözcük türleri olarak 11 adet etiketleme kullanılmıştır. Ne var ki bu ilk derlem denemelerinde eksik olan, sözcük türleri ile biçimbirim ayrımının yapılmamasıdır.

Sejong Derlemi, üç aşamadan oluşan bir derlemidir. Birinci aşamada biçimbilimsel çözümleyici yazılımı “Sejong Tagger” kullanılarak analiz yapılmıştır. Projenin birinci ve ikinci yılında, önce metinlerin etiketleme işlemi yapılmış, sonra anlam bulanıklığını çözmek için etiketlenmiş veriler üzerinde tekrar düzeltme yapılmıştır. Böylece iki yılda 3,5 milyon sözcükten oluşan derlemin ilk versiyonu tamamlanmıştır. İkinci aşamada yarı otomatik etiketleme yazılımı kullanılarak analiz yapılmıştır. Bu aşamada yeni bir yöntem olan “CET Tagger” yazılımı, Cho, 2001 tarafından Sejon Projesi kapsamında tasarlanmıştır. Bu yazılımla yapılan çözümleme sonucunda, sözcüklerin etiketleri “birinci tip”, “ikinci tip” ve “varsayım” olarak ayrı ayrı gösterilmiştir. Sonraki aşamada ikinci tip ve varsayım sonuçları üzerine elle işaretleme işlemi yapılarak çözümlemede zaman kazanılmış ve verimli sonuçlar elde edilmiştir.

İkinci aşamada “CET Tagger” yazılımı kullanılmaya devam edilerek üç yıl boyunca derlem çalışmasına yoğunlaşmış ve toplam 10 milyon sözcükten oluşan derlem tamamlanmıştır. Üçüncü ve son aşama ise büyük derlem kullanılarak istatistiksel yönetime dayalı analiz yapılması olmuştur. Bu aşamada Lee, 2005 tarafından tasarlanan “KOMA Tagger” ve “N-best Tagger” yazılımları kullanılmıştır. Aşama iki yıl boyunca 5 milyon sözcük üzerine etiketleme yapılarak tamamlanmıştır.

Büyük ölçekli derlemin oluşturulması için analizde kullanılan ölçüt oldukça önemlidir. Derlemde yer alacak sözcüklerin biçimini koruyarak biçimbilimsel analiz yapılması için eksiltilmiş ya da kısaltılmış öğeler üzerine de analiz yapılmış olması gereklidir. Örneğin eksiltilmiş koşaçların, kısaltılmış adılların, çekim eklerinin de analiz edilmesi gerekmiştir. Bu durumlarda Sejong Derlemi’nde özellikle 1999 yılında Ulusal Kore Dil Enstitüsü tarafından hazırlanan “Standart Korece Büyük Sözlük” ölçüt olarak alınmıştır. Özel adlar ve bileşik adlar da bu sözlüğe dayalı olarak çözümlenmiştir.

Choe ve Kang, 2007, Sejong Derlemi’nin analiz yönteminde şimdilik birtakım eksiklikler bulunduğunu iddia etmiş ve birden fazla sözcükten (*multi-word unit*) oluşan özel adın bir sözcük olarak çözümlenmesini büyük bir sorun olarak görmüştür. Örneğin şu an Sejong Derlemi’nde Marcel Proust’un yazdığı bir roman (örneğin “*Kayıp Zamanın İzinde*”) sözcük grubu olarak arandığında, bu sözcük grubu (eylem, *kaybolmak*), (genel ad, *zaman*) ve (eylem, *izlemek*) olarak ayrı ayrı biçimde etiketleme işlemine girecek, sonra biçimbilimsel ya da biçimanlambilimsel analiz uygulanacaktır. Yazarlara göre bu durumda, özel adlar ve sözcük gruplarıyla ilgili bir sözcük listesine (sözlüğe) dayanarak çözümleme yapılması ve bunların işaret edilebilmesi için de ayrıntılı bir çalışma olması gerekmektedir (Bkz. Choe ve Kang, 2007:367).

Sejong Derlemi, sorgulama kullanımı açısından değişik seçenekler ve ayarlar sunan bir kaynaktır. Aşağıda temel düzey sorgu ekranı gösterilmiştir:

· sorgulama ayarı

metin türü - · yazılı
· sözlü

analiz türü - · ham
· biçimbilimsel
· anlamsal
· sözdicimsel

<sorgulanacak derlemler>

- sorgulama ayarı
- kullanılacak derlem dosyası seçme
- derlem veri sepeti oluşturma

kendi istediği şekilde derlemin türü, metne ait zaman kısıtını yazarak bir derlem sepetini oluşturabilir (örneğin, “1990 yıllarında yayımlanan romanlar”)

<sorgulama biçimi>

sözcük / alfabe

The screenshot shows a search interface with various filters and options. It includes a search bar, a 'Sorgulama Ayarı' (Search Settings) section with options for text type and analysis type, and a 'Sorgulanacak Derlem' (Search Corpus) section. The interface is in Korean and includes a search button at the bottom.

Sejong Derlemi’nde sözcük arama sırasında kullanıcı tarafından metinlerin türü, konusu ve zamanı da ayrı ayrı seçilebilir ve çeşitli istatistiksel sonuçlar görülebilir. Derlemi kullanmak isteyenler internet ortamında üyelik olmadan <https://ithub.korean.go.kr/user/corpus/corpusSearchManager.do> sayfasına girebilirler.

Derlem Tabanlı Çalışmalar:

Bütün Korece derlemlerinin başlıca kullanım alanları Kore dili, Korece eğitimi ve yabancı dil olarak Korece öğretimidir. Derlem çalışmalarından sonra, derlemin bu eğitim ve öğretim alanlarında çalışma yapanlara yardımcı olması beklenmektedir. Bu yöndeki alan etkisini görmek üzere Kore’de bütün yayınlar ve çalışmaların toplandığı Meclis Merkez Kütüphanesi’nin <http://dl.nanet.go.kr/SearchList.do> adresinden ‘Korece’, ‘derlem’, ‘kullanım’ anahtar kelimeleri ile arama yapılmış ve arama sonucunda derlemler üzerinden toplam 43 adet çalışma yapıldığı anlaşılmıştır. 2000-2015 arasında yapılmış bu çalışmaların dağılımına bakıldığında 29 makale, 9 tez, 6 kitap ve 2 adet elektronik materyal görülmüştür. Alanlara göre çalışmaların dağılımı aşağıda gösterilmektedir:

	Kore dili	Korece eğitimi	yabancı dil olarak Korece öğretimi	yabancı dil çalışmaları	derlem programı ve kullanımını	toplam
makale	9	6	8	2	1	26
tez	2	3	2	0	2	9
kitap	3	0	0	0	3	6
diğer	0	0	0	0	2	2
toplam	14	9	10	2	8	43
%	32.56	20.93	23.26	4.65	18.60	100.00

Kore dili çalışmalarında ele alınan konuların başlıcaları şunlardır: sözcük kullanım sıklığı, metin türüne göre tümce bitirmede kullanılan çekim ekleri, çekimleyen sözcük türleri grubunun görünümleri, belirteç kullanımları, günümüz Korecesinde kısaltma görünümleri, cinsiyete göre belirteç seçimi... Bunların yanı sıra belirli alanlarda derlem oluşturma planları ve yöntemlerine ilişkin çalışmalar da yer almaktadır. Örneğin, 21. yüzyılda yayımlanan romanlar üzerine bir derlem oluşturulmuş olması, lehçe derleminin hazırlanıp kullanım aşamasına getirilmesi, Korecenin incelenmesinde derlemden faydalanma yöntemleri ve önerileri gibi çalışmalar bunlardan bazılarıdır.

Korece eğitimi alanında ise temsili sözcüklerin listesinin oluşturulmasına, saygıdili anlatım yöntemlerine ve ilkokuldan lise seviyesine kadar Korece sözlüğünün oluşturulmasına yönelik çalışmalar bulunmaktadır. Derlem, sıklık bilgisini içermekte olduğundan, sıklığa dayalı çeşitli derlem tabanlı materyaller geliştirmek de mümkündür.

İlginç bir konu olarak, derlem tabanlı çalışmalar arasında ikinci dil olarak Korece öğretimi alanında, öğrencilerin uyruklarına göre yaptığı hatalar üzerine analizler yapıldığı da göze çarpmaktadır. Yapılan hatalar, derlemlerden hareketle kontrol edilerek saptanmaya ve bu hataların sebepleri açıklanmaya çalışılmıştır.

Sonuç

PISA testlerinde başarı sırasının hep üst sıralarında yer alan Güney Kore, son çeyrek yüzyılda yalnızca teknoloji ve sanayide değil, eğitimde de bir sıçrama yapmıştır. Bunda Korece eğitime verilen önemin de büyük yeri vardır. Bu çerçevede, oylumlu ve ayrıntılı derlemlerin yapılmış olmasının dil eğitimindeki başarıya olan katkısı da açıktır.

Kaynakça

- ALTUNDAĞ, Pınar (2012), Korecedeki Çince Kökenli Kelimelerin (Hanca) Yeri ve Korece Öğretimindeki Önemi: Dil Dergisi (156), Nisan-Mayıs-Haziran, 48-60.
- CHO, Hyun-Yong (1999), '한국어 어휘의 특징과 어휘교육 (Korece sözcük kullanımının özelliği ve sözcük öğretimi)', 한국어교육 (Kore dili eğitimi), 10-1, 265-281.
- CHO, Jin-Hyun (2001), 형태소 분석 말뭉치 구축을 위한 태깅 방안의 연구와 개발 (Biçimbilimsel analiz derleminin oluşturulması için bir etiketleme yazılım araştırması), Yüksek lisans tezi, Goryo Üniversitesi
- CHOE, Min-Woo ve Boem-Mo Kang (2007), '세종 형태(의미)분석 말뭉치 구축의 실제 (Practice in Constructing Sejong Morph(sense) Analysis Corpora)', 민족문화연구 337-372
- JUNG, Hee-Chang. ve Wun-Yung Lee (2008), '2008년 표준국어대사전 개정 (Standart Korece Büyük Sözlük düzeltme raporu 2008)', Ulusal Kore Dil Enstitüsü 2008-01-26.
- KANG, Boem-Mo ve Heung-Kyu Kim (2004) 한국어형태소 및 어휘 사용 빈도의 분석2 (Korece biçimbirlikleri ve sözcükleri üzerine kullanım sıklığı analizi 2). 고려대학교민족문화연구원 (Goryo Üniversitesi Kore Çalışmaları Araştırma Enstitüsü).

◆ Nadir Engin Uzun / Yang-Hee Lee

- KİM, Heung-Kyu ve Boem-Mo Kang (2000) 한국어형태소 및 어휘 사용 빈도의 분석1 (Korece biçimbirmleri ve sözcükleri üzerine kullanım sıklığı analizi 1). 고려대학교민족문화연구원(Goryo Üniversitesi Kore Çalışmaları Araştırma Enstitüsü).
- KİM, Tae-Hun (2002), 한국어 말모듬 구축 방안 연구: 형태 분석 말모듬을 중심으로 (Korece Derlem oluşturma yöntemi : biçimbilimsel analiz derlemi üzerine), Yüksek lisans tezi, Goryo Üniversitesi
- KİM, Heung-Kyu ve diğerler (2007), '21세기 세종계획 현대국어 기초말뭉치: 성과와 전망 (21. yüzyıl Sejong projesi Modern Korece temel derlemi: Sonuçlar ve Beklentiler)', 19. Kore alfabesi ve Kore dili bilgisayar ortamına aktarılması kurultayı bildiri kitabı.
- LEE, Chung-Woo (1994), '한국어 어휘 교육을 위한 대표어휘 설정 (Korece sözcük eğitimi için temsil edici sözcüklerin seçilmesi)' 국어교육 (Korece eğitimi) 85-86, 19-39.
- LEE, Chung-Woo (1998), '어휘 교육과 어휘의 특성 (Sözcük eğitimi ve özelliği), 국어교육 (Korece eğitimi) 95, pp. 83-108.
- LEE, Do-Gil (2005), 한국어 형태소 분석과 품사 부착을 위한 확률 모형 (Korece üzerine biçimbilimsel analiz ve sözcük türlerinin etiketlemesi için olasılık modeli), Doktora Tezi, Goryo Üniversitesi
- LEE, Wun-Yong (2002), '표준국어대사전의 통계적 분석 (Standart Korece Büyük Sözlük İstatistiksel Analiz Raporu)', Ulusal Kore Dil Enstitüsü 2002-1-10, 17.
- LEE, Yik-Seop ve Wan Chae (2011), 국어문법론강의 (Korece dilbilgisi öğretimi), Hwayeonsa Yayınevi, 46-218.
- TÜRKÖZÜ, S. Göksel. ve diğerler (2014), *Kore'yi Tanıyarak Korece Öğrenelim*, Lotus Yayınevi, 29.
- UZUN, N. Engin ve Kyoo-Seok Yeon (2014), "Genetik Sınıflamada Dil Tipolojisinin Yeri: Altay Dilleri Üzerine Bir Örneklem": 17. Uluslararası Türk Dilbilim Kurultayı, 3-5 Eylül 2014, Rouen, Fransa.