



Prof. Dr. Mehmet Özmen
Armağanı

Editörler

Nurettin Demir - Faruk Yıldırım

Adana 2014

Prof. Dr. Mehmet Özmen Armağanı

Editörler: Nurettin Demir - Faruk Yıldırım

© Bu kitapta basılan makalelerin yayın hakları yazarlarına aittir. Makaleler, yazarlarının izni olmadan basılı veya elektronik olarak çoğaltılamaz, başka dile çevrilemez.

ISBN: 978-975-487-196-8

Kapak Tasarım: Emine Ağırbaş

Baskı: Çukurova Üniversitesi Basımevi Müdürlüğü

İsteme Adresi: Çukurova Üniversitesi

Basın ve Halkla İlişkiler Bürosu

Balcalı/ADANA

Mail: basin-halk@cu.edu.tr

Tel. : 0 322 338 71 14

Türkçenin Tarihsel Derlemi

Engin Çetin, Çukurova Üniversitesi

Tarihsel Derlem Kavramı ve Dünyadaki Örnekleri

Derlemler, amaçları ve kapsamlarına koşut olarak yazılı ya da sözlü içeriğe sahip ve belirli bir dilde dilsel çeşitliliği yansıtabilen *genel derlemler* (general corpora); bilgisayar bilimleri derlemi benzeri *özel alan derlemleri* (specialized corpora); belirli bir yazı dilini, belirli dönem ve tanıklarla içeren *yazılı derlemler* (written corpora); sözlü dili temsil için hazırlanan *sözlü derlemler* (spoken corpora); eş zamanlı dilsel veriler içeren *eş süremliler* (synchronic corpora); belirli bir dilin farklı zaman aralıklarıyla içlendiği *art süremliler* ya da *tarihsel derlemler* (diachronic or historical corpora); yine ana dili ve yabancı dil olarak belirli bir dilin öğretiminde temel verilerin derlenmesi amaçlı tasarılan *öğrenici derlemleri* (learner corpora) ve diğer derlem türlerinden daha hacimli ve ayrıntılandırılmış, kapsamlı derlemlerden olan *izlem derlemleri* (monitor corpora) olarak karşımıza çıkmaktadır (McEnery vd. 2006; Çetin-Özkan 2010: 282).

Tarihsel derlem, özetle herhangi bir dilin herhangi bir dönemine ait metinleri derlem dil biliminin yöntemleriyle bilgisayarlar aracılığıyla okunur-işlenir bir yapıyla bir araya getirilmesi ve araştırmacıların hizmetine sunulmasıdır.

Bu alana özgü farklı dünya dilleri üzerine yürütülmüş çalışmaların bazıları şunlardır:

Helsinki Derlemi olarak da bilinen Helsinki Üniversitesi İngiliz Dili Bölümü'nde 1984-1991 yılları arasında hazırlanan 1,572,800 sözcüklük *The Helsinki Corpus of English Texts* (www.helsinki.fi/varieng/CoRD/corpora/HelsinkiCorpus/generalintro.html), Eski İngilizce, Orta İngilizce ve Modern Öncesi Dönem İngilizcesi (730-1710 yılları arası) ile yazılmış 450 civarında metni içermektedir.

Zürich Corpus of English Newspapers (ZEN) (<http://es-zen.unizh.ch/>), Zürich Üniversitesi İngiliz Dili Bölümü'nce 1993-2003 yılları arasında hazırlanmıştır. 1661-1791 yılları arasındaki 349 gazetenin söz varlığından oluşan 1,6 milyon sözcüklük bir derlemdir.

The ARCHER Corpus (a corpus of British and American English from 1650-1990) (www.nau.edu/english/ling/), 1,7 milyon sözcük içerir. 1037 metin ve 10 kayıttan (drama, mektup, şiir vb.) oluşmaktadır.

Corpus Del Español (www.corpusdelespanol.org), 1200'lü yıllardan 1900'lü yıllara kadar İspanyolca yazılmış 13,926 metinden oluşan 101,311,682 sözcüklük bir tarihsel derlemidir.

Konuyla ilgili bir başka çalışma ise Cambridge Üniversitesi Dilbilim Bölümüne 2001-2004 yılları arasında hazırlanan *A Historical Corpus of the Welsh Language 1500-1850* (people.pwf.cam.ac.uk/dwew2/hcwl/menu.htm)'tir. Bu derlem, 1500 ile 1850 yılları arasında Gal dili ile yazılan 30 metni içeren yaklaşık 420.000 sözcük barındırmaktadır.

The Corpus of Historical American English (Coha) ise Amerika Birleşik Devletleri'ndeki Brigham Young University'de gerçekleştirilen 1810-2009 yılları arasındaki 107 bin metni 400 milyonu aşkın sözcüğü kapsayan bir derlemidir (Çetin-Özkan 2010: 282 vd.).

Türkçe İçin Tarihsel Derlem İhtiyacı:

Dünyada teknolojik gelişmelerin dil bilimi araştırmalarında bilgisayar kullanımını yaygınlaştırması ve bir süre sonra dil bilimi araştırmalarında bilgisayar bilimlerinin vazgeçilmez bir unsur olarak öne çıkması, derlemelerin hazırlanmasını sağlarken bilgiye daha hızlı ulaşma ve böylece daha çok ve bilgisayarlar yardımıyla ortaya koyulan, daha sağlıklı çalışmalar yapma düşüncesi Türk dili araştırmacılarının da Türkçe için tarihsel derlemin bir ihtiyaç olduğunu düşüncelerini sağlamıştır. Bununla birlikte, dünya dilleri için daha erken dönemlerde başlayan tarihsel derlem çalışmaları, Türkçe için nispeten daha geç tarihte başlamıştır. Türkçenin ilk tarihsel derlemi de 1999-2003 yılları arasındaki çalışmanın ürünü olarak Almanya'da hazırlanmıştır. *Johann Wolfgang Goethe Üniversitesi*'nde yürütülen *Vorislamische Alttürkische Texte: Elektronisches Corpus (VATEC)* [İslamiyet Öncesi Eski Türkçe Metinler: Elektronik Derlem] (<http://vatec2.fkidg1.uni-frankfurt.de/>) başlıklı bu çalışma Eski Türkçenin Orhon ve Uygur Türkçesine ait kimi metinleri içermektedir ve çözümlemeli bir derlemidir. Bu proje *Deutschen Forschungsgemeinschaft* [Alman Araştırma Vakfı] tarafından desteklenmiştir.

Türkçenin ilk tarihsel derleminin Türkiye'de değil de yurt dışında hazırlanmış olmasının nedenleri arasında, Batı'da bilgisayarlı dil bilimi çalışmalarının Türkiye'den önce başlaması ve buna bağlı olarak daha hızlı gelişmesi ile bu türden projelerin çeşitli kurumlarca zaman ve maddi olarak daha fazla desteklenmesi sayılabilir.

Vorislamische Alttürkische Texte: Elektronisches Corpus (VATEC) derleminin Orhon Türkçesinin ve Uygurcanın kimi metinlerini içermesi, Türkçenin tarihsel yazı dillerine ilişkin söz varlığı açısından eksik yönünü oluşturmaktadır. Oysa bizce, Türkçenin tarihsel derleminin, ilk tarihsel yazı dili dönemi olan Orhon Türkçesi metinleriyle başlaması, Uygur ve Karahanlı Türkçesi söz varlığını içermesi gerekmektedir. Tarihsel söz varlığının büyük ölçüde ortaya koyul-

ması, betimlemeli, eş zamanlı, art zamanlı, ve karşılaştırmalı çalışmaların yapılabilmesi açısından oldukça önemli bir konudur. Metinleri, içerik açısından oldukça farklı olan bu üç dönem ürünlerinin farklı söz varlığı, ancak bu biçimde büyük ölçüde ortaya koyulabilir. Bu noktadan hareketle, Türkçenin ilk yazılı dönemlerini içeren 7.-13. yüzyıl arasında farklı alfabelerle yazılmış, farklı ses ve biçim özellikleri taşıyan ve en önemlisi farklı söz varlığına sahip bu metinlerin bilgisayarlarca aranıp incelenebilir duruma getirilmesi bir ihtiyaç olarak karşımıza çıkmaktadır.

Eski Türkçe ve Karahanlı Türkçesinin Tarihsel Derlemi (7.-13. yy.) Projesi:

Eski Türkçe ve Karahanlı Türkçesinin Tarihsel Derlemi Projesi (7.-13. yy.) 15 Ekim 2010 tarihinde yürütülmeye başlamış TÜBİTAK tarafından desteklenen ulusal araştırma projesidir. 30 ay süren proje 15 Nisan 2013 tarihinde tamamlanmıştır. Proje ile ortaya koyulan Türkçenin 7.-13. yüzyıllar arasındaki metinlerinden derlenmiş söz varlığı, Mayıs 2012 tarihinden itibaren *derlem.cu.edu.tr* adresinde sınırlı sayıda metnin kullanıma açık duruma getirilmesi sonrasında deneme yayınıyla ve Nisan 2013'ten başlayarak da tüm metinlerin kullanıma açık duruma getirilmesiyle kullanıma sunulmuştur.

Eski Türkçe ve Karahanlı Türkçesinin Tarihsel Derlemi Projesi (7.-13. yy.) Çukurova Üniversitesi Fen Edebiyat Fakültesi Türk Dili ve Edebiyatı Bölümü'nde yürütülmüştür. Projenin yürütücüsü Çukurova Üniversitesi Fen Edebiyat Fakültesi Türk Dili ve Edebiyatı Bölümü Eski Türk dili anabilim öğretim üyesi Yrd. Doç. Dr. Engin Çetin'dir.

Projenin danışmanlığını aynı bölümde Eski Türk Dili anabilim dalının başkanlığını yürüten Prof. Dr. A.Deniz Abik ve Mersin Üniversitesi Eğitim Fakültesi Türkçe Eğitimi Bölümü öğretim üyesi Yrd. Doç. Dr. Bülent Özkan yapmıştır.

Çukurova Üniversitesi Fen Edebiyat Fakültesi Türk Dili ve Edebiyatı Bölümü öğretim üyesi Doç. Dr. Muna Yüceol Özezen, Yıldırım Beyazıt Üniversitesi İnsan ve Toplum Bilimleri Fakültesi Türk Dili ve Edebiyatı Bölümü öğretim üyesi Yrd. Doç. Dr. Akartürk Karahan, İstanbul Teknik Üniversitesi Bilgisayar Mühendisliği Bölümü öğretim üyesi Yrd. Doç. Dr. A. Cüneyd Tantıoğlu ve Çukurova Üniversitesi Türkoloji Araştırmaları Merkezi uzmanı Dr. Bekir Tahir Tahiroğlu projede araştırmacı / uzman olarak görev almıştır.

Projenin yazılım konusu yazılım uzmanı Sinan Yalçınkaya tarafından düzenlenmiştir.

Yöntem

Derlem dil bilimi konusunda, daha önce de değinildiği gibi, Türkiye'deki ve Batı'daki çalışmalar karşılaştırıldığında Türkiye için bu alanın henüz çok yeni olduğu görülmektedir. Türlü nedenlerle Türkçenin dilsel verilerini bütünüyle ortaya koyabilecek çalışmalar yapılabilmiş değildir. Bu nedenle, Türkçe için

temsil gücü yüksek derlemin oluşturulması, yapılması gereken en önemli çalışmalarından biri olmalıdır.

Türkçe için temsil niteliği olan bir derlemin oluşturulamamasının nedenlerinden biri, bir derlemin, bilgisayar tarafından okunabilecek yapıda işaretlenmiş olması zorunluluğudur. Bu işaretlemeler, XML (Extensible Mark-up Language), SGML (Standard Generalized Mark-up Language) ve TEI (Text Encoding Initiative) gibi işaretleme dilleri aracılığıyla yapılan ayrıntılı işaretlemelerdir. Bunun yanında kimi derlemlerin kendilerine özgü etiketleme sistemi geliştirdikleri de görülmektedir. Derlem işaretleme işlemleri zaman alıcı ve yüksek maliyetli çalışmalardır. Yukarıda sayılan ya da bunlara benzer programlarla işaretlenerek bilgisayarlarca işlenebilir ve okunabilir hale getirilen bir derlem Java, Perl, PHP vb. türlü programlama dilleriyle hazırlanan görsel arayüze sahip yazılımlarla sorgulanabilmektedir. Böylelikle dilsel verilerin görünümünün belirlenmesi, sorgulanabilmesi hızlı ve güvenilir biçimde kullanılması bir dijital ortamla mümkün olmaktadır.

Temsil gücü yüksek bir derlemin oluşturulabilmesi için yöntemle ilgili ilk aşama kuşkusuz dönemi temsil edebilecek sayıda ve nitelikte eserlerin seçilmesidir. Bu, projenin başarılı olmasının en temel koşullarından biridir. Bu amaçla, farklı türlerde ve yeterli sayıda yazılı ürünün seçilmesi dönemin genel söz varlığının yeterli derecede temsil edilmesini sağlayacaktır. Yukarıda da belirtildiği gibi, dönem ürünleri tür açısından çeşitlilik arz etmektedir.

Son yıllardaki teknolojik gelişmelere koşut olarak dil çalışmalarında, bilgisayar bilimlerindeki yeni yöntem ve uygulamaların daha hacimli dil malzemesini doğru ve daha faydacı amaçlar doğrultusunda işleme olanağı da araştırmacıların hizmetine sunulmaktadır. DDİ (Doğal Dil İşleme) olarak adlandırılan alana özgü yöntem ve uygulamalar derlemin oluşturulmasında ve araştırmacıların hizmetine sunulmasında önemli katkılar sağlamıştır.

Bir verinin metin içinde hangi türlerde kullanıldığı, farklı yazım özelliğinin olup olmadığı gibi durumlar veri işleme sırasında ortaya konulan işlemlerdir. Projede görev alan araştırmacı ve burslu öğrenciler, herhangi bir verinin yazım biçimini proje için geliştirilen yazılım aracılığıyla oluşturulan veri tabanına göndermişlerdir. Yazılım, gönderilen verilerin kim tarafından, ne zaman gönderildiğinin proje yürütücüsü tarafından görüntülenmesine olanak vermiştir. Böylelikle gönderilen veriler, yürütücü ve araştırmacılarca denetlenmiş, olası hatalara başlangıçta önlemler alınmıştır. Sistemin işleyişinin sağlıklı olması amacıyla projede görev alan yazılım uzmanından destek alınmıştır.

Proje Aşamaları

A) Derlemde Yer Alan Metinlerin Belirlenmesi

Orhon Türkçesi döneminde çoğunluğu taş, bir ölünün ardından yazılan metinler, kimi zaman salt kendini geleceğe taşıma çabasıyla yazılmış bir tür not görünümündeyken kimi zaman da ölünün geçmişte yaptıklarını, savaşlarını,

başarılarını anlatan, gelecek kuşakların ders almasının beklendiği söylev niteliğinde ürünlerdir. Bu yazıtlar, hacimce sınırlı olmalarına karşın dil özellikleri ve söz varlığı açılarından oldukça gelişmiş ve geniş kapsamlı bir yapıya sahiptir.

Uygurca döneminde yazılan metinler ise Orhon Türkçesindeki metinlere oranla daha hacimli, dil özellikleri ve söz varlığı açılarından daha gelişmiş bir görünüme sahiptir. Bu dönemde kabul edilen dinlerin etkisiyle, bu dinleri anlamak, öğrenmek ve yaygınlaşmasını sağlamak amacıyla dini içerikli metinleri Çince, Sanskritçe, Toharca ve Sogdca gibi dillerden çeviren Uygurlar, özgün yapıtlar da meydana getirerek son derece gelişmiş bir edebiyat dili yaratmışlardır. Bu dönemde, masallar, hikâyeler, biyografi türünde eserler, dualar, tövbe duaları, fal kitabı, dini bilgiler içeren kitaplar, sutralar, ilahiler gibi dini içerikli eserler yanında vasiyetname, kira sözleşmesi, satış sözleşmesi, mektup gibi din dışı konularda düzyazı ve şiir türünde ürünler meydana getirilmiştir.

Karahanlı Türkçesi döneminde İslamiyet'in Türklerce kabulü dolayısıyla Kur'an tercümelerinin yapıldığı gözlenmektedir. Bunun yanında bu dönemin en önemli ürünleri Türk dilinin de en önemli yapıtları olan *Kutadgu Bilig* ve *Divanü Lügati't-Türk*'tür. *Atebetü'l-Hakayık* da Karahanlı Türkçesi döneminde kaleme alınmış didaktik türde bir eserdir.

Derlemde yer alan metinlerin içerik özellikleri ve türleri açısından farklılık arz etmeleri dönemin tüm söz varlığını ortaya koyabilmek açısından önemlidir. Derlem içeriğinin yalnızca dinî içerikli ya da yalnız din dışı konuları ele alan metinlerden oluşması dönemin söz varlığı açısından yanıltıcı olabilir. Bunun gibi, yalnız düz yazı ya da yalnız şiir türünde eserlerin seçilmesi de söz varlığı ve sözlerin kullanım değerleri açısından dönemi tam olarak yansıtamaz. Bu noktadan hareketle, derlemde yer alan metinlerin geniş bir yelpazeyi oluşturduğu söylenebilir.

Derlemde dinî ve din dışı konuları içeren, hukuk belgesi niteliğinde olan ve bu özelliğiyle hukuk dili terimleri içeren, biyografi niteliği taşıyan, sözlü dile özgü atasözü ve deyimleri barındıran sözleri içeren çok sayıda metin yer almıştır. Aynı zamanda bu metinlerin bir bölümünün telif, bir bölümünün çeviri olması, şiir ve düzyazı türünde olması söz varlığını oluşturacak sözcük sayısının daha da artırılması ve böylelikle dönemin daha iyi betimlenebilmesi açısından önemlidir. Bu seçim, projenin sonuçlanıp derlemin oluşmasından sonra da araştırmacıların benzer ya da farklı içerikteki metinlerin söz varlığının saptanması; düzyazı ve/veya şiir türündeki sözlerin kullanımının saptanması; telif ya da çeviri eserlerdeki söz varlığı farklılıklarının saptanması konusunda çalışmalar yapabilmesine olanak sağlaması amacıyla yapılmıştır.

Yukarıda belirtilen düşünceler doğrultusunda 7.-13. yüzyıl Türkçesinin söz varlığını büyük ölçüde temsil ettiği düşünülen aşağıdaki metinlerin derlemde yer almasına karar verilmiştir:

Dönem	Metin	Yüzyıl	Metin Türü
7-9.			
Orhon Türkçesi	Çoyr Yazıtı	7. yy. sonu (?)	Düzyazı
	Küli Çor Yazıtı	8. yy	Düzyazı
	Ongin Yazıtı	8. yy	Düzyazı
	Kül Tigin Yazıtı	8. yy	Düzyazı
	Bilge Kagan Yazıtı	8. yy	Düzyazı
	Tunyukuk Yazıtı	8. yy	Düzyazı
	Tes Yazıtı	8. yy	Düzyazı
	Taryat Yazıtı	8. yy	Düzyazı
	Şine Usu Yazıtı	8. yy	Düzyazı
	Süci Yazıtı	9.yy	Düzyazı
9.-11.			
Uygurca	Irk Bitig	9. yy.	Düzyazı
	Maniheist Metinler (uigurische Manicheismus)		Düzyazı
	Huastuanift	9.-10.yy.	Düzyazı
	Rüzgâr Tanrısı	8.-9. yy.	Düzyazı
	Uigurica I-IV	Muhtelif	Düzyazı-Şiir
	Türkische Turfan-Texte	Muhtelif	Düzyazı-Şiir
	Altun Yaruk	10. yy.	Düzyazı
	Maitrisimit Nom Bitig (Hami)	13. yy.(?)	Düzyazı
	Hsüen Tsang Biyografisi	10.-11. yy.	Düzyazı
	Hamam Sutrasi		Düzyazı
	Bahşı Ögdisi	13.-14. yy.	Düzyazı
	Et'öz-üg köñül-üg körmek at(ı)ğ nom bitig	10. yy.	Düzyazı
	Sekiz Yükmek	10., 13-14. yy.	Düzyazı
	Uigurische Sukhavativyuha-Fragmente	10.-11. yy.	Düzyazı
	İyi ve Kötü Prens	10. yy.	Düzyazı
	Abhidharma (Shogaito)	13-14. yy.	Düzyazı

	Uigurisches Totenmesse	13-14. yy.	Düzyazı
	St.Petersburg'taki Uygurca metinler (Sho-gaito)	13-14. yy.	Düzyazı
	Tövbe duası	13. yy.	Düzyazı
	Hristiyanlığa İman Metni (Zieme 1998 Nest.)	13.-14. yy.	Düzyazı
	Dindışı metinler (mektuplar, şiirler vs.)	Muhtelif	Düzyazı-Şiir
	Hukuk Belgeleri (sözleşmeler, vasiyetnameler)	13.-14. yy.	Düzyazı
	Budist Uygur şiirleri	Muhtelif 11.-13.	
Karahanlı Türkçesi	Kur'an Tercümesi Rylands Nüshası	11. yy	Düzyazı
	Kutadgu Bilig	11. yy	Şiir
	DLT'deki Türkçe parçalar	11. yy	Düzyazı-Şiir
	Atebetü'l-Hakâyık	12.-13. yy	Şiir

Tablo 1: Derlemde yer alan metinler

B) Veri İşleme Aşamaları:

Derlemde yer alacak metinlerin belirlenmesinin ardından bu metinlerin bilgisayar ortamına aktarılması işlemine geçilmiştir.

Derlem dilbilimi konusunda daha önce de değinildiği gibi, Türkiye'deki ve Batı'daki çalışmalar karşılaştırıldığında Türkiye için bu alanın henüz çok yeni olduğu görülmektedir. Bu nedenle, Türkçe için temsil gücü yüksek derlemin oluşturulması yapılması gereken en önemli çalışmalardan biri olmalıdır.

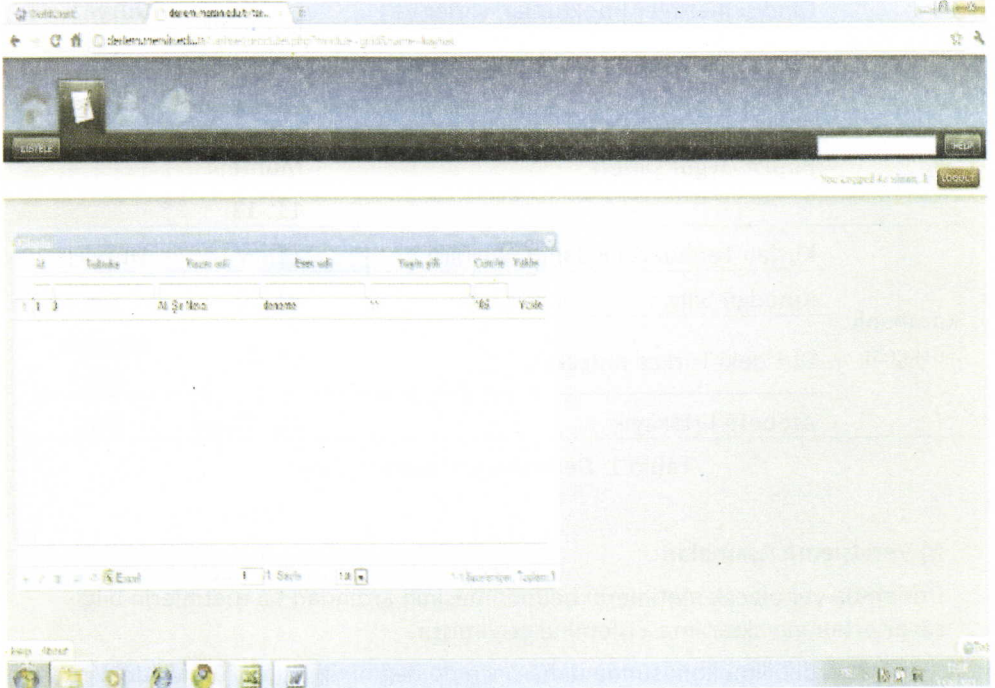
Bu çerçevede bütünleşik olarak ETKT-D'nin veri işleme ve sorgu aşamaları ise ana hatlarıyla şöyle özetlenebilir:

- 1- Derlem Oluşturma
- 2- Derlem İşaretleme (sözcükbirimsel, sözdizimsel, dönem, tür, anlam vb. işaretleme)
- 3- Derlem Sorgu Sistemi Oluşturma

B-1) Derlem Oluşturma:

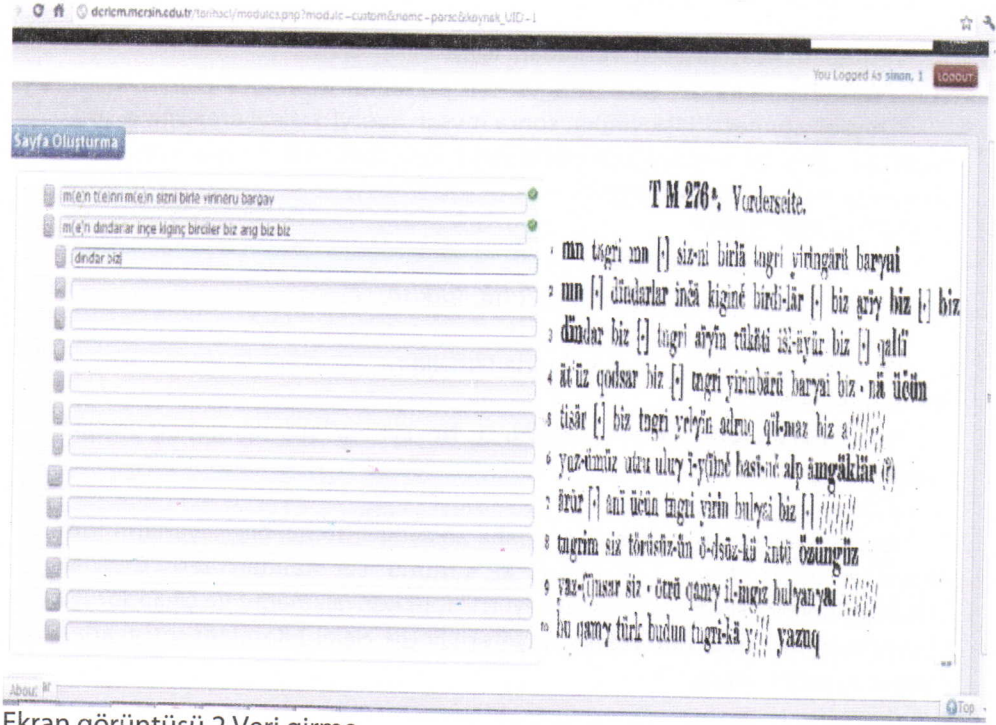
Burslu öğrenciler, sayfalarında yer alan "YÜKLE" butonunu tıkladığında kendilerinin yazması gereken metin listesi karşısına çıkmaktadır. Bu metinlerden birini seçen burslu öğrenci, metin sayfalarını ve boş satır ekranını aynı anda

görebilmekte ve böylelikle metinleri kolay bir biçimde sisteme girerek kaydetmektedir. Projenin bu aşamasında www.cu.edu.tr tabanlı bir web sayfası henüz tahsis edilmediğinden projemizde danışman olarak görev alan Yrd. Doç. Dr. Bülent ÖZKAN'ın Mersin Üniversitesi Eğitim Fakültesi Türkçe Eğitimi Bölümü'nde yürüttüğü TÜBİTAK destekli Türkiye Türkçesinde *Fiillerin Derlem Denetimi ve Derlem Tabanlı Sözlüğü* projesinin alt yapısından yararlanılmıştır.



Ekran görüntüsü 1 Sisteme giriş

Aşağıdaki resimde görülen bu sistemde ekranın sol tarafı boş satırlara sağ tarafı ise metnin ilgili sayfasına ayrılmıştır. Burslu öğrenci metnin ilgili satırını yazıp aşağı satıra geçtiğinde sistem otomatik olarak satırı kaydetmektedir. Yanlış ve eksik yazımlar ise kolaylıkla düzeltilebilmektedir.



Ekran görüntüsü 2_Veri girme

B-1-1) Derlemin denetlenmesi

Burslu öğrencilerce yazılarak oluşturulan derlem, proje yürütücüsü ve araştırmacılarca denetlenerek hataların giderilmesi böylelikle sağlam bir derlemin ortaya çıkması sağlanmıştır.

B-2) İşaretleme

Genelde derlem oluşturma aşamasında kullanılan OKT (Optik Karakter Tanıma) yazılımları bu tip tarihsel derlemlerin oluşturulması aşamasında işe yaramamaktadır. Bunun nedeni özel karakter setlerinin kullanıma zorunluluğudur. Döneme özgü alfabenin örneğin 'k', 'h', 'ğ' vb. çeviriyazı karakterleri bu tip yazılımlarda tanınmadığı için bunlar yine birtakım yazılımlarla bilgisayarın okuyup işleyebileceği bir formata dönüştürülmüştür. Metinler (bk. Tablo 1) bu anlamda yeniden yazılmıştır.

Derlem için seçilen metinlerin ortak yazı karakterleri belirlendikten sonra bilgisayar ortamına aktarılmasıyla birlikte yukarıda söz edilen uygulamalar ve dil programları aracılığıyla metinlerin işaretlenerek bilgisayarlarca okunabilir hale getirilmesi işlemi gerçekleştirilmiştir. Çalışmanın özelliği gereği çeviriyazı karakterlerinin bilgisayarca tanınması güçlüğü aşılmıştır.

Eski Türkçe karakterler standart olarak kullanılmadıklarından, projede kullanılacak bilgisayarların bu karakter setine göre ayarlanması gereklidir. Bu işlem aşağıdaki gibi iki aşamalı olarak gerçekleşir:

I-) Sunucu Ayarı

Veritabanı Ayarı (MySQL veritabanı için)

* Veritabanı karakter seti "utf8_turkish_ci" olarak ayarlanmalıdır.

* mysql_connect() işlevinden sonra mysql_query() işlevi aracılığıyla sırasıyla şu sorgular yapılmalıdır. Böylelikle veritabanı ve programlama dili arasında uyum sağlanmış olur.

```
SET NAMES "utf8"
```

```
'SET CHARACTER SET "utf8"'
```

```
'SET COLLATION_CONNECTION="utf8_turkish_ci"'
```

Kodlama Ortamı Ayarı

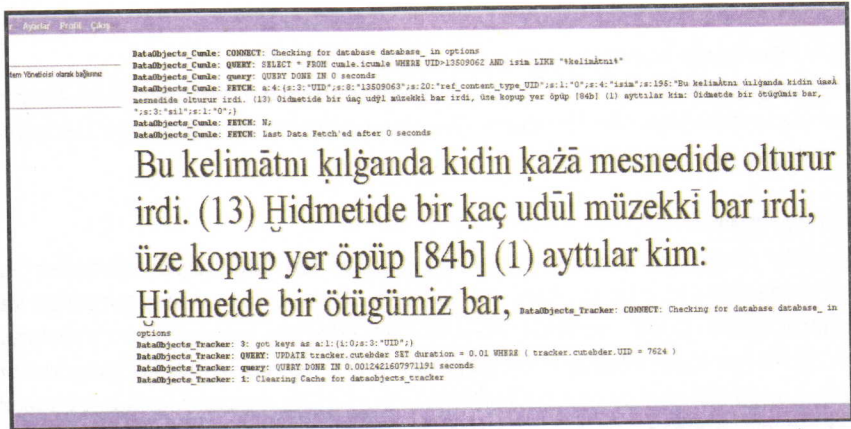
Kod dosyaları utf8 biçiminde kaydedilmelidir.

HTML Çıktısı Ayarı

Eski Türkçe yazılar komutları arasına yazılmalıdır.

II) İstemci Ayarı

İstemci bilgisayarda işlem biraz daha kolaydır. İstemci bilgisayarların font karakter setleri dizinine (Windows XP sürümü için standart yeri "C:\WINDOWS\Fonts" dizinidir) "Ali Sir Nevayi.ttf" font kopyalamamız ve çıkan soruya "Evet eklemek istiyorum" dememiz yeterlidir (bk. Şekil 1 Karakter seti ayarları yapılmış sistem çıktısı).

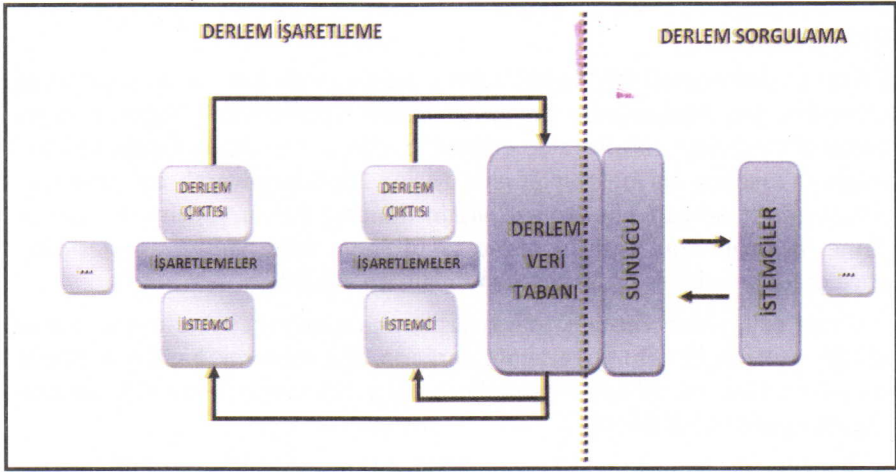


Şekil 1: Karakter seti ayarları yapılmış sistemin veri çıktısı

B-3) Derlem Sorgu Sistemi Oluşturma

Derlem veri tabanı yukarıda belirttiğimiz derlemde yer alacak metinlerin seçimi ve seçilen metinlerin sayısallaştırılması aşamalarının sonucunda oluşturulmuştur. Oluşturulan bu derlem üzerinden sözcükbirimsel ve sözdizimsel işaretlemeler ile diğer içerik işaretlemeleri (metin türü, saha, yüzyıl vb.) proje için oluşturulan bir sonucu (server) üzerinden araştırmacı ve burslu öğrencilerce yapılmıştır.

Projenin üzerinden yürütüleceği söz konusu sistem projede yer alacak olan yazılımcı ve araştırmacılarca sınanmış ve amaca uygunluğu test edilmiştir (bk. Şekil 2 Veri işaretleme ve derlem sorgulama diyagramı).



Şekil 2: Veri işaretleme ve derlem sorgulama diyagramı

Biçimbirimsel Değişkenlerde Sözcükbirimsel Sorgulamalar

Öte yandan, Eski Türkçenin Orhon Türkçesi ve Uygurca dönemi ile Karahanlı Türkçesi dönemleri ses, biçim vb. dil özellikleri açısından benzer pek çok yönü barındırmanın yanında kimi farklılıkları da barındırmaktadır. Örneğin Orhon Türkçesinin söz içi ve söz sonu *-b-*, *-b* sesleri Uygurcada *-v-*, *-v*; Karahanlı Türkçesinde *-w-*, *-w* olarak gelişmiştir. Dolayısıyla Orhon Türkçesinin *eb* “ev” sözcüğü Uygurcada *ev*; Karahanlı Türkçesinde *ew* biçimindedir. Bu durumda okuyucu derlemde *eb* sözcüğünü aradığında, sözcüğün yanında *ev* ve *ew* sözcüklerini ve bunların ilgisini görebilmelidir. Bunun gibi, Orhon Türkçesinin *anyıg* “kötü, fena” sözcüğü, Uygurcada *anıg* ve *ayıg*; Karahanlı Türkçesinde *ayıg* biçiminde görülmektedir. Okuyucu, *anyıg* sözcüğünü aradığında söz konusu farkı üç biçimi de görecektir geliştirilen yazılımla çözümlenebilmiştir. Söz konusu güçlük girişler anlam işaretlemeleriyle aşılabilmektedir.

<Sorgu>	<Anlam Eşleşmesi>	<Çıktı>
<anıg>, <ayıg>, <anyıg>	“kötü, fena”	<anıg>, <ayıg>, <anyıg>
<ew>, <ev>, <eb>	“ev”	<ew>, <ev>, <eb>

Tablo 2 Sorgu-anlam eşleşmesi ve çıktı

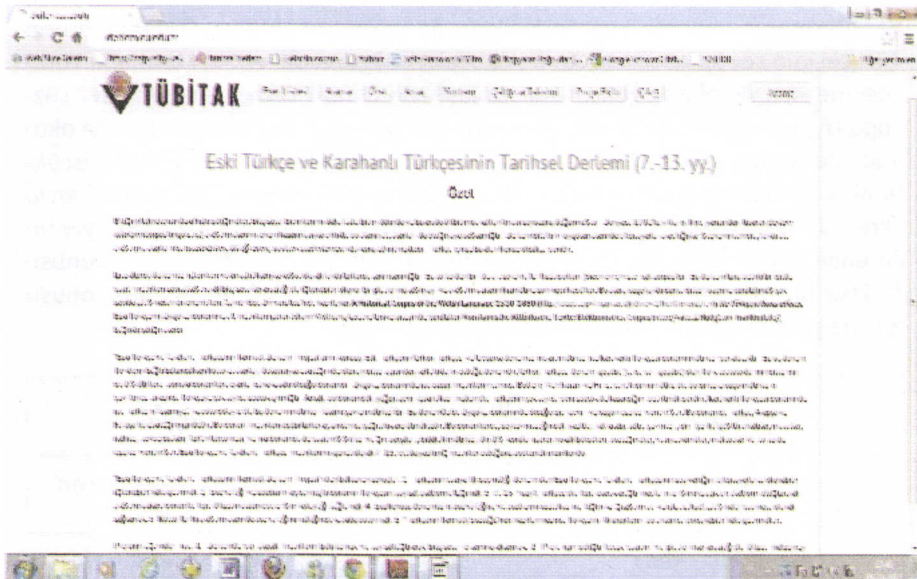
Tüm bu işlemlerin tamamlanmasının ardından yine derlemimiz için oluşturulan bir arayüzle derlem kullanıcının hizmetine sunulmuştur. Kullanıcı, aradığı sözcüğün kullanım sıklığını, kullanıldığı metinleri, yazım özelliklerini hızlı ve güvenilir bir biçimde görme olanağına sahip olmaktadır.

Deneme Yayını

Çukurova Üniversitesi Rektörlüğü'nden projenin çıktılarının ve sonuçlarının ağ üzerinden yayımlanabilmesi amacıyla *derlem.cu.edu.tr* adlı bağlantı ve bu bağlantının projeye tahsis talebi edilmiş, bu talep, Rektörlükçe uygun bulunmuştur. Yukarıda da belirtildiği gibi, Mayıs 2012 başlayarak bu yana *derlem.cu.edu.tr* adresinden projeye ilişkin deneme yayını yapılmıştır. Bu deneme yayınıyla sorunların saptanarak projenin sonuçlanma aşamasından önce çözümlenebilmesi için gereken önlemleri almaktır.

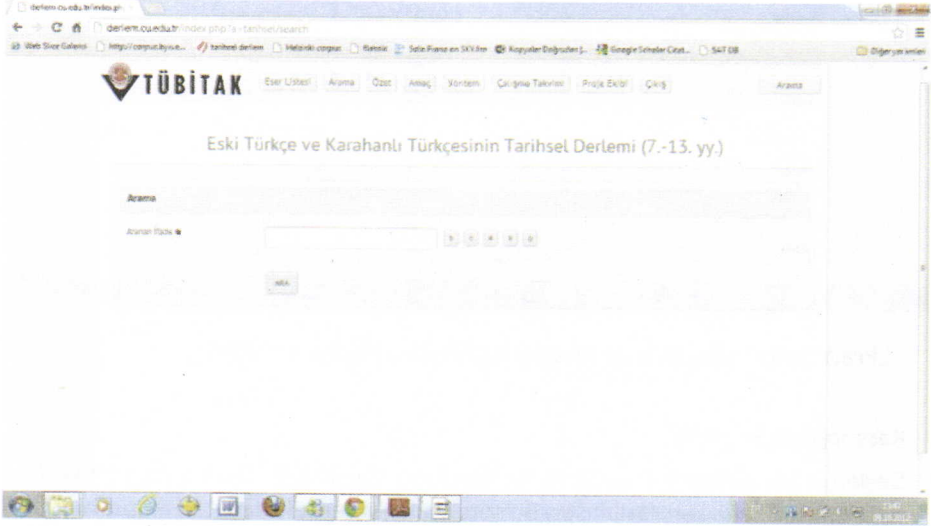
Proje çalışmaları kapsamında, belirtilen çalışma takvimine uygun olarak 18.-24. aylarda derlem işaretleme, verilerin denetlenerek hataların giderilmesi böylelikle veri tabanının tam olarak oluşturulması aşamaları sıkıntısız bir biçimde geride bırakılmıştır.

Aşağıda, projenin genel ağ üzerindeki tanıtım sayfasından görüntülere yer verilmiştir. Projenin deneme yayını aşamasında proje henüz tamamlanmadığından yalnızca bir metin üzerinde arama yapılmasına izin verilmiştir. Bu metin, Eski Türkçeden farklı ses ve söz varlığı içermesi ve bu farklılıkların yayın sırasındaki durumunun denetlenebilmesi amacıyla özellikle tercih edilmiş olan 11. yüzyıla ait manzum bir metin olan *Kutadgu Bilig* metnidir.



Ekran görüntüsü 3 Proje Tanıtım Sayfası: Özet

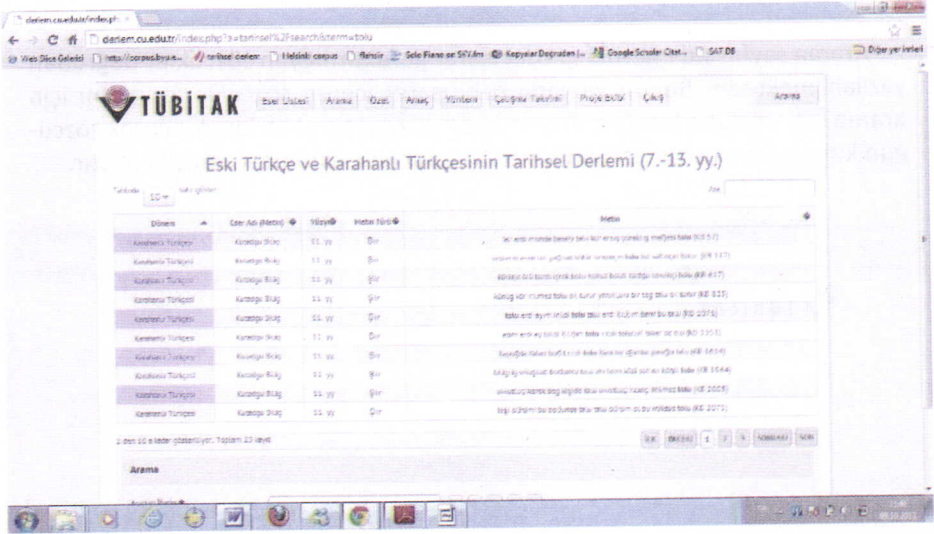
Arama sayfasında herhangi özel işaret gerektirmeyen sözcükler doğrudan yazılabilmektedir. Bunun yanında özel işaret içeren sözcüklerin yazımı için arama satırının yan tarafına *đ, ħ, ħ, ħ, ħ, vb.* özel karakterler eklenerek sözcüğün kolay yazılması ve böylelikle kesin sonucun çıkarılması amaçlanmıştır.



Ekran görüntüsü 4 Proje Tanıtım Sayfası: Arama

İlgili sözcük yazıldıktan sonra *Ara* tuşuna basılır. Sonuç, ekranda belirir. Her sayfada toplam 10 satır yer almaktadır. Ancak kullanıcı, satır sayısını artırarak tercihe göre 25, 50, 100 yapabilir. Her sayfada 10 satır yer aldığında 10'dan fazla kayıt diğer sayfalarda yer almaktadır. Sözcüğe ilişkin toplam kayıt sayısı sayfa altında belirtilmektedir.

Ayrıca, sözcüğün saptandığı dönem, eser, eserin yazıldığı tarih (yüzyıl olarak), eserin türü, sözcüğün saptandığı metin bölümünün solunda yer alan bölümlerde gösterilmekte, sözcük, metin içinde koyu biçimde yer alarak kullanıcının dikkatine sunulmaktadır:



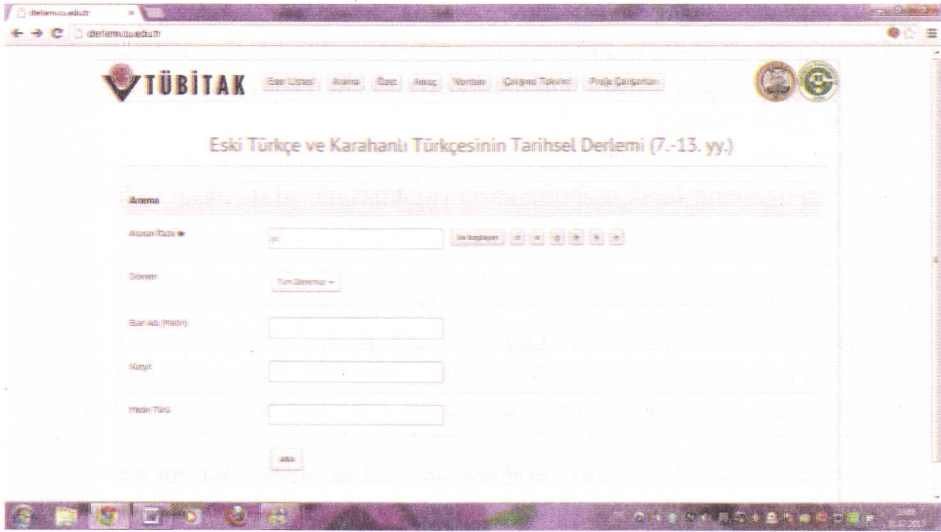
Ekran görüntüsü 6 *tolu* sözcüğüne ilişkin arama sonuçları

Karşılaşılan Sorunlar

Derlemin oluşturulması sırasında her projede olabileceği gibi, bu çalışmada da birtakım sorunlarla karşılaşmıştır. Bu sorunların en temel nedeni kuşkusuz metinlerin tarihî yazı dillerinden kalmış olmasıdır. Söz konusu dönemlerde yazımda bir standartlaşmanın olmayışı, dönemlerin farklı ses, biçim ve söz varlığı özelliklerine sahip olması, metinlerin farklı alfabelerle yazılmış olması ve bu alfabelerin o dönem Türkçesinin ses yapısını günümüze ulaştırmakta yetersiz kalması metinlerin bilgisayar ortamına aktarılması sırasında birçok sorunla karşılaşılmasına neden olmuştur. Yukarıda sözü edilen nedenlerle ve özellikle de alfabenin Türkçenin ses değerlerini yansıtmadaki yetersizliği nedeniyle araştırmacıların çok sayıda sözcüğü farklı biçimde okumaları aşılması gereken sorunlar olarak karşımıza çıkmıştır. Bu durumda gerek yazılım uzmanı Sinan Yalçınkaya, araştırmacı Yrd. Doç. Dr. A. Cüneyd Tantıuş'un teknik konulardaki deneyimi, gerekse danışman ve yürütücü görüşleriyle, derlem oluşturma sırasında ortak bir yazımın belirlenmesi ancak web sayfasındaki yayında bir bağlantı eklenerek farklı okumaların burada gösterilmesi düşüncesi benimsenmiştir.

Derlemin Yayına Başlaması

Yaklaşık 10 ay süren deneme yayını sırasında sorgulamalarda ortaya çıkan sorunlar saptanmış ve bu sorunlar için çözüm yolları bulunmuştur. Bu sırada deneme yayını sırasında olmayan pek çok filtreleme satırı eklenmiş, böylelikle kullanıcının aramayı genel ya da özel alanlarda yapması sağlanmıştır. Bu aşamaların ardından derlem, yukarıda da belirtildiği gibi, Nisan 2013 tarihinde asıl yayına geçmiştir.



Ekran görüntüsü 7derlemin asıl yayını ile getirilen yenilikler

Derlem ana sayfası olan *derlem.cu.edu.tr* adresinde derlemin asıl yayınında Arama sayfası ana sayfa yapılmış, bu sayfada kullanıcının işini kolaylaştırıcı ve daha ayrıntılı aramalar yapmasını sağlayıcı yollar aranmıştır. Bu amaçla, arama satırının altına dönem, eser adı, yüzyıl ve metin türü satırları da eklenerek kullanıcı için farklı filtreler oluşturulmuş, ayrıca derlemin kullanımını kolaylaştırıcı açıklamalar eklenmiştir.

Sonuç

Türk dilinin en eski yazılı belgelerini kapsayan Eski Türk dili araştırmaları, başta Almanya olmak üzere, Rusya, Japonya, Macaristan, İsveç, İngiltere, Fransa, Çin Halk Cumhuriyeti gibi çok sayıda ülkede 1900'li yılların başlarından itibaren yapılmaktadır. Adı geçen ülkelerde yer alan üniversiteler ve enstitüler Eski Türk dilinin türlü konularına ilişkin çalışmalar yapılmaktadır. Türkiye'de Eski Türk dili araştırmaları, Türkoloji açısından büyük önem taşımaktadır. Türkiye'deki Eski Türk dili çalışmalarının 1940'lı yıllardan sonra canlandığı gözlenmektedir. Eski Türk dili araştırmaları, Türkiye'de ve dünyada, öncelikle metin yayınlarına dayanmaktaydı. İlk olarak Orhon yazıtlarının bulunuşu, çözümüne, okunuşu, ardından Uygurlardan kalma yazıt ve binlerce sayfalık metinlerin bulunuşu, okunuşu Eski Türkçe çalışmalarına büyük canlılık getirmiştir. Metinlerin okunması ve yayınlanması çalışmalarının kimi zaman yoğun olarak kimi zaman da yavaş bir biçimde 1980'li yıllara kadar devam ettiği görülür. Bu çalışmalar yanında, 1940'lı yıllardan başlayarak 1970'lerden itibaren yoğunluğu artmış bir biçimde Eski Türk dilinin

gramer konularına, anlam bilgisine, söz varlığına, kavram alanına ilişkin çalışmaların da yapıldığı gözlenir. Bugün Eski Türk dili araştırmaları, az da olsa metin yayınının yanında dil malzemesinin değerlendirilmesi ve böylelikle eski Türk topluluklarının yaşam biçimleri, ekonomik düzeyleri vb. kültürel boyutuyla ele alınması yönünde seyretmektedir.

Derlem dil bilimi, metinlerin bilgisayar ortamında işlenerek, söz konusu dil malzemesi üzerinde farklı araştırmaların yapılmasına, istatistiksel sonuçların ortaya konmasına yardımcı olur.

Eski Türkçe ve Karahanlı Türkçesinin Tarihsel Derlemi (7.-13. yy.) Projesi, derlem dil biliminin bugün öne çıkan alt alanlarından olan tarihsel derlemin Türkiye'deki ilk önemli çalışması olmuştur. *Eski Türkçe ve Karahanlı Türkçesinin Tarihsel Derlemi (7.-13. yy.) Projesi*'nin Türkiye'de ilk olmasının yanı sıra Türkçeye ilgili de ilk çalışmalardan biri olması projenin nasıl bir boşluğu doldurduğu konusunda önemli bilgiler vermektedir. Eski Türkçe dönemi içinde yer alan çok sayıda metnin söz varlığının ilk kez bilgisayar ortamında erişime açık hale gelmesi yanında Karahanlı Türkçesi metinlerinin söz varlığının da ilk kez bir derleminde yer alması projenin önemini ve bilim dünyasına katkısını göstermektedir. Derlemin başarıyla sonuçlanmasının ardından Türkçenin tarihsel derleminin dünyada başka diller için yapılan çalışmalar arasında da önemli bir yerinin olacağı ve benzer çalışmalara örnek oluşturacağı düşünülmektedir.

Projenin tamamlanmasıyla, 7.-13. yüzyıllarda oluşan Türkçenin ilk yazılı dönemlerinin söz varlığı ortaya koyulmuştur. Böylece, dönemin söz varlığına ilişkin önemli ve güvenilir yeni bilgiler elde etmek mümkün olacaktır. Proje, günlük kullanıcının ve araştırmacıların yararlanmasına açık olacağından günlük kullanıcının merakının giderilmesi, araştırmacıların yeni araştırmalar yapabilmesi sağlayacak bir ortam olacaktır. Böylelikle projenin, diller için son derece önemli olan tarihsel sözlüğün güvenilir biçimde hazırlanması konusunda araştırmacılara önemli katkılar sağlayacağı düşünülmektedir. Türkçenin tarihsel sözlüğünün hazırlanması konusunda Türk Dil Kurumu bir süredir çalışma yapmaktadır. Projenin tamamlanmasının tarihsel sözlüğün oluşumu açısından da yararlı olması beklenmektedir. Bu anlamda projenin, Türkiye'de ve dünyadaki Türk dili araştırmaları için sıkça başvurulacak bir kaynak olması beklenmektedir. Yalnız Türk dili araştırmaları için değil, karşılaştırmalı dil araştırmaları için de projenin yararlı olacağını söylemek mümkündür.

Eski Türkçe ve Karahanlı Türkçesinin Tarihsel Derlemi (7.-13. yy.) ile;

- 1- Türkçenin yazıya ilk geçirildiği dönem olan Eski Türkçe ve Karahanlı Türkçesinin söz varlığını bilgisayarlarca okunabilir-işlenebilir hale getirilmiş olmakta;
- 2- Söz varlığı ve sözdizimi açısından ilgili dönemin Türkçesini sayısal platforma taşınmış olmakta;

3- 7.-13. Yüzyıl Türkçesiyle ilgili olarak araştırmacıların erişimine açık bir platform oluşturarak araştırmacılara dönemle ilgili dilsel malzemeye erişim kolaylığı sağlanmakta;

4- Söz konusu dönemlerin söz varlığına ve sözdizimi özelliklerine ilişkin karşılaştırmalı ve istatistiksel çalışmalar yapmaya olanak sağlanmakta;

5- Kültür tarihi araştırmalarında söz varlığının izleğinde katkıda bulunulmakta;

6- “Türkçenin Tarihsel Sözlüğü”nün hazırlanmasında Türkçenin ilk ürünleri bu anlamda incelenebilir duruma getirilmiş olmaktadır.

Bu türden projeler dünyada uzun zaman dilimlerinde geniş bir uzman kadrosuyla ve büyük bir bütçeyle yapılmaktadır. Dileğimiz ve önerimiz Türkiye’de Türkçenin bu derlemde yer almayan tarihî dönem metinlerinin de derlenerek kullanıcıya sunulması, bundan sonra yapılacak derlem çalışmaları için daha büyük bir bütçenin ayrılması ve sonraki çalışmaların bu projenin sunduklarının ötesinde yararlar ve sonuçlar ortaya koymasıdır.

Kaynaklar:

Çetin, Engin, Bülent Özkan (2011). “Tarihsel derlem kavramı ve Eski Türkçe ve Karahanlı Türkçesinin Tarihsel Derlemi Projesi”. *III. Uluslararası Dünya Dili Türkçe Sempozyumu (16-18 Aralık 2010 İzmir) Bildiri Kitabı*. İzmir. 282-289.

McEnery, Tony, Richard Xiao, Yukio Tono (2006). *Corpus-Based Language Studies, An Advanced Resource Book*. Newyork: Routledge.

Oflazer, Kemal (1994). “Two-level description of Turkish morphology”, *Literary and Linguistic Computing* 9 No. 2. 472.

Oflazer, Kemal vd. (1994). “Tagging and Morphological Disambiguation of Turkish Text”, *Submitted to 4. Conference on Applied Natural Language Processing*. 144-149.

Say, Bilge vd. (2002) “Development of a Corpus and a Treebank for Present-day Written Turkish”, *Current Research in Turkish Linguistics, Proceedings of 11th International Conference on Turkish Linguistics. Eastern Mediterranean University, Eds. Kamile İmer and Gürkan Doğan*. Gazimağusa: EMU Press. 183-192.

<http://derlem.mersin.edu.tr/derlem516/>

<http://es-zen.unizh.ch/>

<http://people.pwf.cam.ac.uk/dwew2/hcwl/menu.htm>

<http://turkcederlem.mersin.edu.tr/>

<http://vatec2.fkidg1.uni-frankfurt.de/>

www.corpusdelespanol.org

www.helsinki.fi/varieng/CoRD/corpora/HelsinkiCorpus/generalintro.html

www.nau.edu/english/ling/

www.tudd.org.tr/