Ildırı, N. (2024). Özbekistan'da dil ile ilgili derlem çalışmaları ve proje tabanlı yapılan çalışmaların tanıtımı. *Uluslararası Türkçe Edebiyat Kültür Eğitim Dergisi, 13*(4), 1283-1295.

*Uluslararası **T**ürkçe **E**debiyat **K**ültür **E**ğitim Dergisi Sayı: 13/4 2024 s. 1283-1295, TÜRKİYE*

Araştırma Makalesi

ÖZBEKİSTAN'DA DİL İLE İLGİLİ DERLEM ÇALIŞMALARI VE PROJE TABANLI YAPILAN ÇALIŞMALARIN TANITIMI

Nursan ILDIRI*

Geliş Tarihi: 28 Ekim 2024 ***Kabul Tarihi: 1 Aralık 2024***

Öz

Teknolojik gelişmelerin hızlanmasıyla beraber bilgisayar destekli dil bilimi araştırmalarının ve uygulamalarının önem kazandığı görülmektedir. Dil biliminin önemli bir alanını teşkil eden derlem dil bilimi, bilgisayar destekli dil araştırmaları ve uygulamaları için oldukça elverişli bir çalışma alanı sunmaktadır. Derlem dil bilimi, diller arası benzerlikleri ve farklılıkları belirleme, doğal dil işleme uygulamaları geliştirme ve dil öğretimini destekleme gibi alanlarda önemli bir kaynak sağlamaktadır. Bu türlü çalışmalar aynı zamanda diller arasında yapı ve anlam analizi yapmak, doğal dil işleme uygulamalarıyla dil modelleri oluşturabilmek maksadıyla veri setleri oluşturmak, dil öğretiminde öğrenicilerin iki dilli ifadelerle dilleri kavrayışlarını kolaylaştırmak, edebî eserlerdeki bağlamların, anlatım tekniklerinin incelenmesine olanak sağlamak, bireylerin dil becerilerinin gelişmesine yardımcı olmak amacıyla da yapılmaktadır. Son yıllarda Türk lehçeleri için yapılan derlem tabanlı çalışmaların arttığı gözlemlenmektedir. Bu çalışmada genel olarak derlem kavramı üzerinde durulmuş, dünya dillerinde çok kullanılan derlem uygulamalarının Türk lehçelerinde var olan örnekleri, Özbekistan'da derlem çalışmalarının geçmişten günümüze ne şekilde gelişim gösterdiği izah edilmiş, Özbekistan'da gerçekleştirilen derlem tabanlı projeler ve çalışmalar tanıtılmıştır.

Anahtar Sözcükler: Özbek Türkçesi, derlem dil bilimi, ulusal derlem, paralel korpus.

CORPUS STUDIES ON LANGUAGE IN UZBEKISTAN AND CONTRIBUTIONS OF PROJECT-BASED STUDIES TO THE FIELD

Abstract

It has been seen that computer-aided linguistics researches and applications have gained importance with the acceleration of technological developments. Corpus linguistics, which constitutes an important field of linguistics, offers a very suitable field of study for computer-aided language research and applications. Corpus linguistics provides an important resource in areas such as determining similarities and differences between languages, developing natural language processing applications and supporting language teaching. Such studies also provide data to analyze the structure and meaning between languages and to create language models with natural language processing applications. It is also done to create sets, to facilitate learners' understanding of languages with bilingual expressions in language teaching, to enable the examination of contexts and expression techniques in literary

* Dr. Öğr. Üyesi; Atatürk Üniversitesi, Edebiyat Fakültesi, nursan.aslan@atauni.edu.tr

works, and to help individuals develop their language skills. It has been observed that corpus-based studies on Turkish dialects have increased in recent years. In this study, the concept of corpus is generally emphasized, examples of corpus applications that are widely used in world languages in Turkish Dialects, how corpus studies have developed in Uzbekistan from past to present are explained, and corpus-based projects and studies carried out in Uzbekistan are introduced.

Keywords: Uzbek Turkish, corpus linguistics, language corpus, national corpus, parallel corpus.

Giriş

Dil çalışmalarında dil bilimi ve dil bilgisel yöntem, gözlemlenebilen, istatistiksel ve sayısallaştırılıp sonradan veri tabanı olarak kullanılabilecek "veri" temelli uygulamalar hızla ilerlemekte ve bilgi teknolojilerindeki hıza paralel olarak gelişmektedir (Tahiroğlu, 2010, s. 185). Dil çözümlemelerinde kullanılan derlemler buna örnek olarak gösterilebilir. Belli amaçlar temelinde yapılandırılmış metinler ve konuşmalar bütünü olarak ifade edilen derlem, genel olarak bir dili temsil edebilme amacıyla belirli bir zaman aralığında, yazılı ve sözlü dil kullanım metinlerini/konuşmalarını, yazar ve konuşan özelliklerini iletişim ortamlarının alan ve türlerine dengeli ve katmanlı örnekleme yoluyla derleyip belirlediği ölçütleri kapsayan ayrıntılı veri bilgisi ve temel dil bilimsel çözümleme araçlarıyla birlikte elektronik ortamlara sunan kaynaklara denmektedir (Aksan, Aksan, Özel, Yılmazer vd. 2014, s. 723-724).

Derlem dil bilimi başlangıçta İngilizce üzerine temellenmiş, ancak kısa bir süre sonra diğer diller için de derlem çalışmaları hız kazanmıştır. 1963 yılında USA'daki Brown Üniversitesi'nde W. Francis ve G. Kucher tarafından 500 metinden oluşan ilk derlem oluşturulmuştur (Adilova, 2021, s. 526).

Derlem, bir arama programına tabi olan metinlerin koleksiyonudur. İyi tanımlanmış bir derlem, dil bilimsel araştırmanın etkililiğini sağlamada istikrarlı dilsel bir işlev görür. Yapay zekânın bir ürünü olarak derlemler; elektronik sözlük, çeviri portalı, terminolojik veri tabanı, sanal (elektronik) kütüphane, e-yayıncılık, e-ders kitapları ve kılavuzlar içermektedir (Toirova, 2023, s. 10).

Dil araştırmalarında derlem çalışmaları, dilin söz varlığının ortaya çıkarılması, buna bağlı olarak sözlüklerin hazırlanması, bir sözcüğün ortaya çıkış zamanının belirlenmesi ve zaman içindeki kullanım sıklığının incelenmesi, dildeki yerli ve ödünç sözcüklerin oranı ve bu oranlardaki değişimlerin değerlendirilmesi, dilde meydana gelen yabancı dillerin etkisinin belirlenmesi gibi amaçlar için kullanılmaktadır (Adalı, 2022, s. 9).

Derlemin bir diğer ifadesi olan korpus teknolojisi, metodoloji tasarımı ve dilsel değişimlerin ölçülmesi açısından birçok dilsel ve niceliksel çalışmada geçerli bir araştırma aracı olarak kullanılmaktadır. Anlam farklılıklarının ve benzerliklerinin belirlenmesinde ampirik bir bileşen olmanın yanı sıra, kuralcı dil bilgisi kurallarını test etme aracı olarak da kullanılabilirler (Özbay, Gürsoy, 2023, s. 251).

Batıda 1960'lı yılların başlarından itibaren derleme dayalı araştırmalar edebiyat, sözlükbilim, ağız çalışmaları, dil öğretimi ve dil bilgisi alanlarında hızlı bir gelişim göstererek devam etmektedir. Günümüzde ise bu türlü araştırmaların sözlükbilim, dil öğretimi ve dil bilgisi alanlarında yoğunlaştığı gözlemlenmektedir (Özkan, 2020, s. 344).

Özellikle sözlükbilim uygulamalarında derlemlerden oldukça fazla yararlanılmaktadır. Sözlükbilim çalışmalarında veri elde ederken sözcük sıklığı çıkarımı (frequency) ve bağımlı dizin oluşturma (concordance), sözbirimleştirme (lemmatizing/lemmatization), sözcük türü etiketleme (part-of-speech veya tagging), cümle ayrımlama (parsing) ve eşdizim çıkarımı (collocation) gibi standartlaşmış bazı yöntemler kullanılır (Özkan, 2013, s. 157, 158).

Sözcükbilim, genel dil bilimi, bilgisayarlı dilbilim, makine çevirisi ve bilgisayar destekli çeviri, karşılaştırmalı dil bilimi, terminoloji, adli dilbilim, eleştirel dilbilim, yazın çalışmaları, ikinci dil edinimi, dil gelişimi, lehçebilim, biçimbilim, tarihsel dilbilim, psikodilbilim, toplumdilbilim ve çeviribilim gibi birçok alanın, metodolojik ya da kuramsal yaklaşımlarında derlem dil bilimi etkisiyle önemli değişiklikler ve gelişmeler gözlemlenmektedir (Barnbrook, 1996; McEnery ve Wilson, 1996; aktaran, Laviosa, 2002, s. 10; akt. Pekcoşkun Güner, 2018, s. 27).

Türk lehçelerinin karşılaştırmalı olarak incelenmesinde de derlem çalışmaları önem arz etmektedir. Türk lehçeleri için oluşturulan paralel derlemler, gelişmiş dillerin paralel derlemleri temel alınarak oluşturulmuştur. Bu türlü çalışmalar, karşılaştırmalı dil bilimi, etimoloji, diyalektoloji, folklor çalışmaları, metin çalışmaları, çeviri teorisi, edebiyat çalışmaları alanlarındaki incelemelerin zenginleşmesi için önemli bir kaynak teşkil etmektedir. Türk lehçelerinin ortak derleminin oluşturulması, Türk dilinin gelişimi sırasındaki Türk lehçelerinin özelliklerinin analiz edilmesine, Türk lehçelerinin karşılıklı tarihî ve eşzamanlı ilişkilerinin açıklığa kavuşturulmasına, fonetik-yapısal konularla ilgili sorunların çözülmesine yardımcı olmaktadır. Aynı zamanda ortak kültürel mirasın araştırılmasına ve tanıtılmasına önemli bir katkı sağlamaktadır (Musurmankulova, 2023, s. 165).

Son yıllarda birçok Türk lehçesinin ulusal derlemlerinin oluşturulduğu görülmektedir. Türkiye Türkçesi başta olmak üzere Kazak Türkçesi, Kırgız Türkçesi, Özbek Türkçesi, Tatar Türkçesi, Azerbaycan Türkçesi gibi Türk lehçelerinin web tabanlı ulusal derlemleri mevcuttur.

Türkiye Türkçesi üzerine yapılmış derlem çalışması "Türkçe Ulusal Derlem"dir. TÜBİTAK tarafından desteklenen bu proje Mersin Üniversitesi Dilbilim alanı araştırmacıları tarafından oluşturulmuştur. 2008 yılında başlanan bu çalışmanın tanıtım sürümü 2012'de kullanıma sunulmuştur. 50 milyon sözcükten oluşan, 1990-2009 yılları arasında farklı alan ve türlerde %95'i yazılı, %5'i sözlü örnekleri içeren dengeli, karışık (yazılı-sözlü), eşzamanlı ve genel bir derlemdir (URL 1) (Karaoğlu, 2014, s. 182).

Türkiye Türkçesi ile ilgili diğer dil kaynakları TS Corpus'tur. Bu veri kümesinde, Türkçe sosyal medya metinlerinden derlenmiş 491 milyon etiketli girdi (token) bulunmaktadır. trWaC – Turkish corpus from the web veri kümesi, internet üzerinden derlenmiş Türkçe metinlerden oluşur ve toplamda 32 milyon kelime içerir. ParlaMint 2.1 veri kümesi, Türkçe dâhil olmak üzere 17 farklı dil için Parlamento tartışma metinleri içermektedir (URL 2).

Kazak Türkçesinin derlemi A. Jubanov öncülüğünde Ahmet Baytursın Dil Enstitüsü tarafından oluşturulmuştur. Bu derlem içerisinde 31 milyon metin yer almaktadır (URL 3). Kazakça doğal dil işleme aracıdır. Almaty Corpus of Kazakh language (NCKL) adlı veri kümesinde toplamda 40 milyon Kazakça kelime girdisi yer almaktadır (URL 2).

Kırgız derlem Almanya Saarland Üniversitesi tarafından oluşturulmuş ve proje Dr. A. Kasiyeva tarafından yürütülmüştür (Gümüş, 2024, s. 157). (URL 4)

"Tugan Tel" adlı Tatar Ulusal Derlemi Devlet Programı çerçevesinde yürütülmektedir. 180 milyon hacimli bu derlem, farklı türlerde metinler içermekte, Tatar Türkçesine ait bir sözcüğün gramatikal açıklaması, morfolojik özellikleri hakkında da bilgi vermektedir (URL 5) (Allaberdiyeva, 2024, s. 123).

Kırım-Tatar Türkçesi Ulusal Derlemi, 2022-2032 Kırım-Tatar Dili gelişimi stratejisinin uygulanması kapsamında Reentegrasyon Bakanlığı tarafından başlatılan bir projedir (URL 6).

Azerbaycan Türkçesi ile ilgili de çok önemli derlem projeleri yapılmıştır. https://korpus.azerbaycandili.az/ internet adresinde yer alan *Azərbaycan dilinin elektron lüğətlər korpusu* bu projelerden biridir. Bu derlem; *Azərbaycan dilinin orfoqrafiya lüğəti* (genişləndirilmiş və yenidən işlənmiş 6-cı nəşr. 110563 söz. Bakı, "Şərq- Qərb" Nəşriyyat evi, 2013, 839 s.); *Azərbaycan dilinin izahlı lüğəti. Dörd cilddə*. (I cild – 744 s.,II cild – 792 s., III cild – 672 s., IV cild – 712 s.). Bakı, "Şərq – Qərb", 2006; Paşayev A., Bəşirova A. *Azərbaycan şəxs adlarının izahlı lüğəti* (16164 ad – 8328 kişi adı, 7836 qadın adı). Bakı, Mütərcim, 2011, 340 s.; *Azərbaycan dilinin ixtisarlar lüğəti*. Bakı, "Elm", 2019, 232 s.; *Oxford Advanced Learners Dictionary. 5th edition*. "Oxford University Press". 2005, 1539 s. adlı sözlüklerden oluşmaktadır. Azerbaycan Türkçesinin diğer dijital dil kaynaklarından biri olan en-az-parallel-corpus, İngilizce-Azerice ve Azerice-İngilizce çevirilerinin yer aldığı paralel bir derlemdir. az-corpus-nlp, Azerbaycan Türkçesi için DDİ (doğal dil işleme) araçlarında kullanılmak üzere hazırlanmış bir derlemdir. azWaC: Azerbaijani corpus from the web internet üzerinden derlenmiş Azerbaycan Türkçesi metinlerinden oluşmaktadır. Toplamda 94 milyon kelime içermektedir. AZ summarization, Azerbaycan Türkçesi ile yazılmış makalelerin yer aldığı; Awesome Azeri NLP ise Azerbaycan Türkçesi için hazırlanmış birçok doğal dil işleme yazılımı ve yayın çalışmalarının listesinin yer aldığı bir web sitesidir (URL 2).

1. Özbekistan'da Derlem Çalışmaları

Özbekistan'da bilgisayarlı dilbilimin gelişmesinde en büyük rol oynayan kurumlardan biri Özbekistan Milli Üniversitesinde yer alan Bilgi Teknolojileri Enstitüsüdür. Bu enstitü, makine çevirisi, metin analizi gibi çeşitli dil tekniklerinin araştırılması ve uygulanmasında önemli rol oynamıştır. Bu kurumlar, özellikle doğal dil işleme kaynaklarının geliştirilmesine, Özbekçe derlem çalışmalarına öncülük etmişlerdir (Zokirova, 2023, s. 213).

Özbekistan'da dil istatiği ile ilgili ilk çalışmalar 1930-1940'lı yıllarda I. Kissen tarafından yürütülmüştür. 1960-1970'li yıllarda S. Rizayev, S. Muhammediyev ve S. Otamirzayeva gibi akademisyenler Özbekistan Dil ve Edebiyat Enstitüsü öncülüğünde kurulan özel bir laboratuvarda aktif bir şekilde dil bilimsel araştırmalar ile meşgul olmuşlardır. Bu laboratuvarın faaliyetleri yıllara göre azalmış, ancak bağımsızlık yıllarında yeniden ilgi görmüş ve yükseköğretim kurumlarında matematik ve bilgisayarlı dilbilim alanlarında lisansüstü düzeyde öğrencilerin yetiştirilmesi için adımlar atılmıştır. Bilgisayar programcıları ile dilbilimciler ortak çalışmalarıyla alan için önemli çalışmalar yapmışlardır (Zokirova, 2023, s. 213).

2001 yılında Özbekistan Devlet Üniversitesi Bilgisayar Teknolojileri Fakültesinde Bilgisayarlı Dil Bilimi Laboratuvarı açılmıştır. Bilgisayarlı dil bilimi çalışmalarının en önemli alanlarından biri olan derlem dil bilimi, birçok farklı teorik eğitimlerin araştırmacıları tarafından kullanılabilen çok sayıda ilgili yöntem içeren bir metodolojidir (Abjalova, Adalı, Iskandarov, 2023, s. 4).

2020-2021 yıllarındaki birkaç yıllık araştırma ve çabaların sonucunda, Taşkent Özbek Dili ve Edebiyatı Devlet Üniversitesinde Özbek Dili Eğitim Kurumu, Bilgisayar Dil Bilimi Bölümü ile Dijital Teknolojiler ve Uygulamalı Dilbilim ve Dil Eğitimi Bölümü arasındaki iş birliği çerçevesinde AM-FZ-201908172 nolu "Uzbek Language Educational Corpus" adlı bir proje gerçekleştirilmiştir. Bu proje kapsamında oluşturulan korpusta, otomatik morfolojik analiz, kelimelerin hecelere ayrılması, antonimlerin gösterimi, aranan kelimeyi içeren ifadenin göstergesi, açıklamalarıyla birlikte kelimenin paronimleri, aranan kelimenin yer aldığı ifadeler, eş anlamlılaştırıcı (synonymizer) yer almaktadır (Abjalova, Adalı ve Iskandarov, 2023, s. 4).

Özbekçe eğitim korpusunun oluşturulması, kademeli olarak yabancı deneyimlere dayalı verileri oluşturmayı amaçlamaktadır ve Özbek edebî dilinin günümüz sözcüklerini içeren elektronik bir test kitabı, Özbekçe tercüme edilmemiş leksik birimlerini içeren bir elektronik ders kitabı, ses ve video materyalleri de dâhil olmak üzere bir dizi multimedya ürünü ve aynı zamanda Özbekçeyi doğru telaffuz becerilerini geliştirme amaçlı bir mobil uygulama içerir. Eğitim korpusu, öğrencilerin Özbekçeyi devlet dili, ikinci dil ve yabancı dil olarak derinlemesine öğrenmelerini sağlar. Kullanıcılar, eğitim korpusunun ses, video, multimedya uygulamaları, telaffuz ve yazım programları ve e-öğrenme sözlüklerini içeren elektronik materyaller sayesinde Özbekçe öğrenebilirler. Bu uygulamalar, ülkenin ekonomik büyümesini ve sosyal gelişimini sağlayacak bilimsel ve teknolojik kaynakların oluşumuna katkıda bulunacaktır (Toirova, 2023, s. 4).

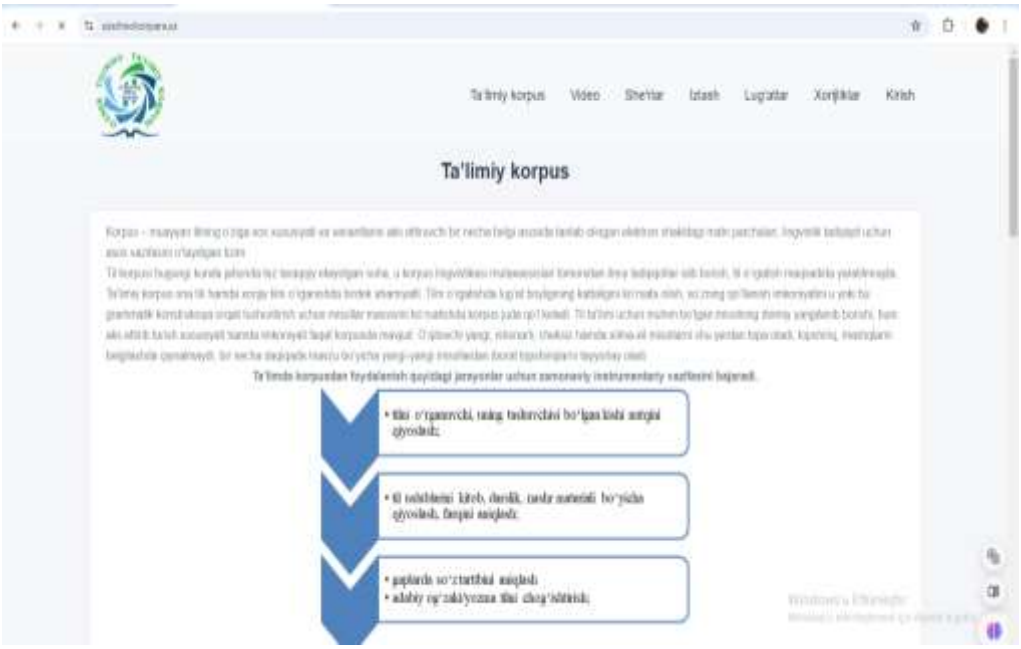
Ayrıca Özbekçeyi öğretmek isteyen bir öğreticiye korpus; öğrencilerin belirli bir seviyede dilde uzmanlaşmak için bilmeleri gereken minimum kelime sayısının belirlenmesinde, hangi kelimelerin güncelliğini kaybettiğini, hangilerinin daha çok kullanılır hâlde olduğunu belirlemede, ders kitabı yazmak veya normal bir ders için teorik materyal seçiminde, dil bilgisi kurallarının ne kadar önemli olduğu, tam olarak hangi kuralın incelenmesi gerektiğinde, kuralların istisnalarının ne sıklıkla olduğu, edebî dilin kurallarının değişip değişmediği, değişirse nasıl ve ne sebeple gerçekleşeceği, belirli bir kelimenin veya ifadenin gerçekte nasıl olduğunu göstermek için dilsel gerçekleri göstermede incelenen dil birimlerinin en sık hangi bağlamda bulunabileceği, üslup farklılıklarını ve söz dizimsel eş anlamlılığı göstermede yardımcı olmaktadır (Adilova, 2021, s. 529).

Özbekçe eğitim derlemi, Özbekçenin olanaklarını öğrenmek ve öğretmek amacıyla elektronik metinler içeren, bir web sitesi şeklinde faaliyet gösteren önemli bir derlemdir. Hem ana dil hem de yabancı bir dil olarak Özbekçenin öğretiminde kullanılan bu derlem, dilin söz varlığının zenginliğini tam olarak göstermek, gramatikal yapı aracılığıyla sözcüklerin kullanma olasılığını ortaya koymak için de oldukça yararlıdır. Derlemin içeriği incelendiğinde Özbek Türkçesine ait resmî, edebî, bilimsel vb. birçok metnin yer aldığı görülmektedir. Edebî metinlerin seçiminde *Abdulla Qahhor, Said Ahmad, Tog'ay Murod, Askad Muhtor*'ın romanları, hikâyelerinden; gazete ve dergi metinlerinin seçiminde *kun.uz, daryo.uz* adlı web sitelerinden; resmî metinlerin seçiminde *lex.uz.* adlı web sitesinden; popüler bilim tarzındaki materyaller çoğunlukla edebiyat, fizik, matematik, kimya, coğrafya, biyoloji gibi okul ders kitaplarından oluşmaktadır. Özbekçe eğitim korpusu, bir korpus arayüzü, bir arama motoru, 75 milyon kelime, bir milyondan fazla metin içeren veri tabanı, 15 elektronik sözlük kaynağından oluşmaktadır (B. Mengliyev, Sh. Khamroyeva, D. Elova, 2021, s. 4, 6-7). Bu derlem, derlem arayüzü, arama motoru ve sözlük sütunundan oluşur. "Sözlükler" sütununda, izahlı lügat,

omonim, sinonim, paronim ve antonim sözlükleri yer almaktadır. İlk sayfada derlem ve derlemi oluşturan kişiler hakkında bilgiler verilmiştir (Raupova, Elov, Abjalova, Alayev, 2021, s. 70).

(URL 8)

Görsel 1



Özbekçenin uluslararası statüsünü yükseltmek, onu iletişim dünyasında bir dil seviyesine çıkarmak, yurtdışında Özbekçe öğrenmeyi ve öğretmeyi kolaylaştırmak Özbekçe Ulusal derlem aracılığıyla gerçekleştirilebilir. Bu nedenle metin korpusunun dil bilimsel temeli ve milli korpus yazılımı oluşturma teknolojisi önem arz etmektedir (Toirova, 2023, s. 2).

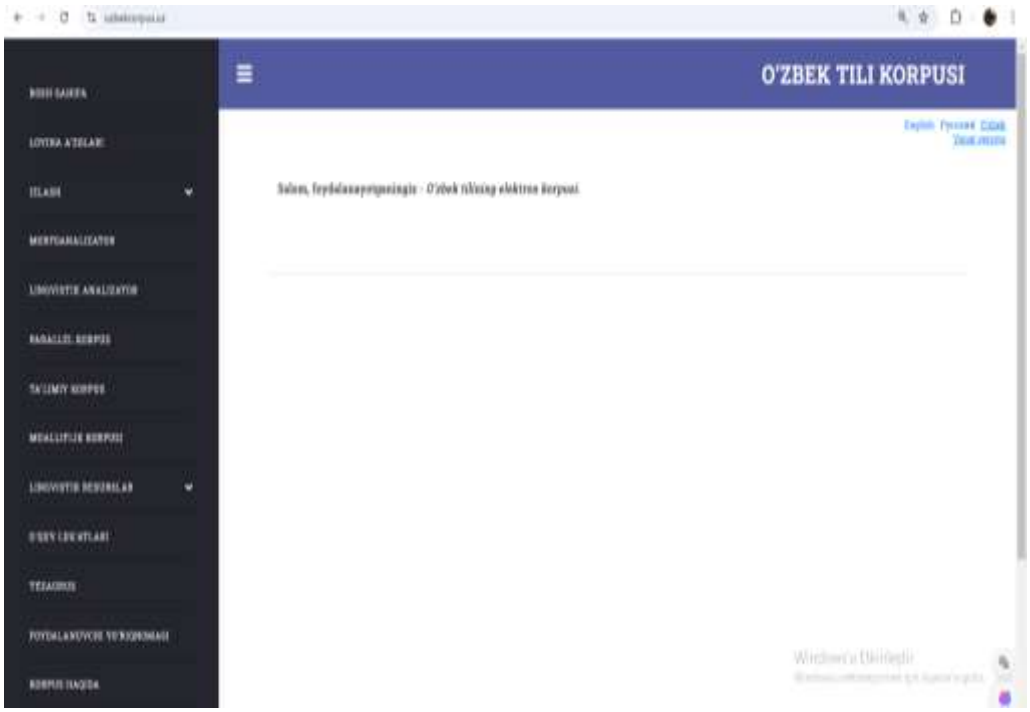
Ulusal bir korpusun oluşturulması, kaynakların listelenmesi ve metinlerin dijitalleşmesi şeklinde iki aşamada gerçekleşir. Kaynak belirleme aşamasında derlemi oluşturacak metinler listelenir. İkinci aşamada ise seçilen metinler bilgisayar tarafından okunabilir bir formata dönüştürülür. (Toirova, 2023, s. 9). Özbekçenin ulusal korpusu, Özbekçede var olan eşanlamlı kelimeler, zıt anlamlı kelimeler de dâhil olmak üzere leksik birimlerin hiyerarşik bir düzenle sıralanmasıdır; bir sözcüğün morfolojik yapısını, sözcük yapılarını, kelimelerin anlamını, morfolojik özelliklerini otomatik olarak analiz edebilir. Yani, korpus işaretlemesi sürecinde, bireysel aramalara dayalı olarak metinlerde korpusun bir parçası olan sözcükleri bulmak ve bunları özel olarak yorumlamak gereklidir. Bunun için, algoritmik, dilbilimsel modelleme çalışmaları yapılmalıdır (Toirova, 2023, s. 9). Özbekçenin ulusal korpusunun oluşturulması dilbilimsel çalışmaların gerçekleştirilmesinde araştırmacılara kolaylıklar sağlaması açısından da önem arz etmektedir. Özbekçe ulusal dil korpusunun yardımıyla, dilbilimciler, sözlükbilimciler, editörler, çevirmenler, gazeteciler, yayıncılar, bilim insanları, öğretmenler, öğrenciler ve diğer tüm dil kullanıcıları ve öğrenenler için geniş bir fırsat yelpazesinin yaratılması bu projenin ne kadar önemli olduğunun bir göstergesi olarak ifade edilebilir (Mengliyev, 2020, 54, akt. Shamsudinova, 2022, s. 693).

Aşağıdaki ekran görüntüsünde Özbekçe Ulusal Derlemi'nin eski versiyonu yer almaktadır. Bu derlem İngilizce, Rusça ve Özbekçe olmak üzere üç dilde hizmet vermektedir.

Sol tarafta; bosh sahifa, loyiha a'zolari, izlash, morfoanalizator, lingvistik analizator, paralel korpus, ta'limiy korpus, mualliflik korpusi, lingvistik resursalar, o'quv lug'atlari, tezaurus, foydalanuvchi yo'riqnomasi, korpus haqida, ilmiy nashrlar sekmeleri bulunmaktadır.

(URL 8)

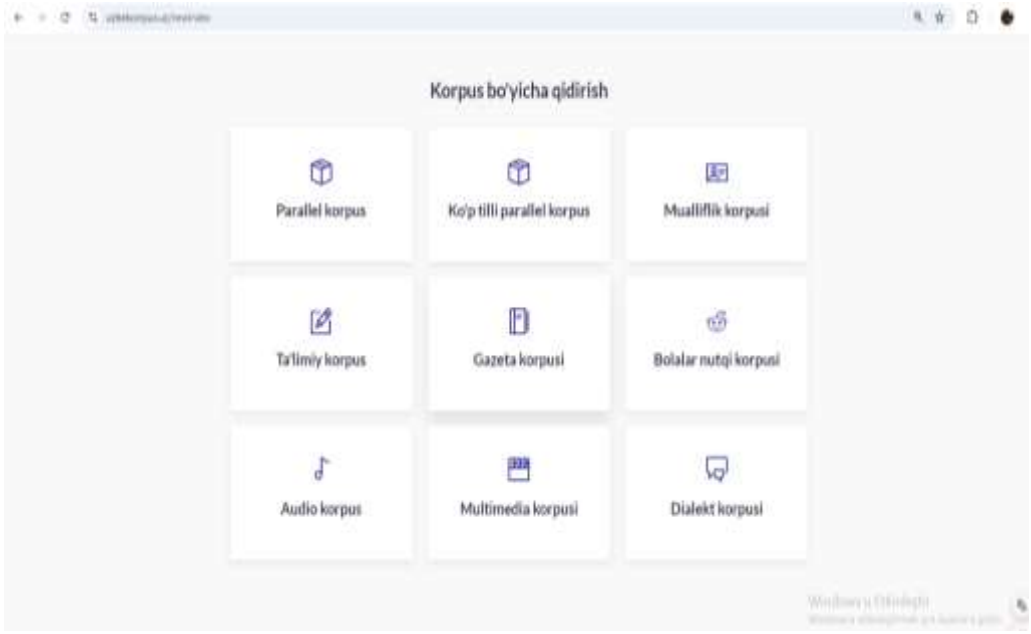
Görsel 2



(URL 9) adresinde Özbekçe Ulusal Derlemi'nin yeni versiyonu yer almaktadır. Bu derlemin ana sayfasında paralel korpus, koʻp dilli paralel korpus, mualliflik korpusi, ta'limiy korpus, gazeta korpusi, bolalar nutqi korpusi, audio korpus, multimedia korpusi, dialekt korpusi yer almaktadır.

Görsel 3

Görsel 4



Özbekçe Ulusal derlemi, "*Ziyonet kutubxonasi"* veya "*Xurshid Davron kutubxonasi"* gibi elektronik kütüphaneleri değil, dili incelemek, öğrenmek ve öğretmek amacıyla leksik, morfolojik, semantik özellikleri açısından dilbilimsel olarak incelenecek metinlerin toplamını ifade etmektedir (Fazliddin, E'zoza, 2024, s. 319).

Özbek dilbilim araştırmacılarından Sh. Hamroyeva "Özbekçe yazar korpusunun oluşturulmasının dilbilimsel temeli", A. Eshmo'minov "Özbekçe ulusal korpusunun eşanlamlı sözcük tabanının oluşturulmasının temelleri", D. Akhmedova "Adlandırma birimlerinin karşıtlığı", "Özbekçe dil korpusları için leksik-semantik etiketlemenin dilsel temelleri ve modelleri", O. Kholyirov Özbekçenin veri tabanının oluşturulması ve tek bir kaynak altında

birleştirilmesi konulu "Özbek dili eğitim korpusunun dilbilimsel temelleri" adlı çalışmalarıyla korpus çalışmalarına önemli katkılarda bulunmuşlardır (Shamsudinova, 2022, s. 693).

Türk dillerinin oluşturulan paralel derlemlerinin ilk örneklerinden biri *Uzturkfraz* olarak isimlendirilen Özbekçe-Türkçe Deyimler Derlemi (URL 10) adresinde yer almaktadır. Bu paralel derlemde Özbekçe deyimler Shavkat Rahmatullayev tarafından hazırlanan *O'zbek tilining izohli frazeologik lug'ati* adlı eser temel alınarak oluşturulmuş, Türkçe deyimler için ise Ömer Asım Aksoy'un *Atasözleri ve Deyimler Sözlüğü* adlı eseri temel alınmıştır. Bu derlem deyimlerin yapısal ve anlamsal özelliklerinin belirlenmesine yardımcı olur ve iki dilde de kullanım sıklıklarını kıyaslama imkânı sunar (Musurmankulova, 2023, s. 85).

Görsel 5

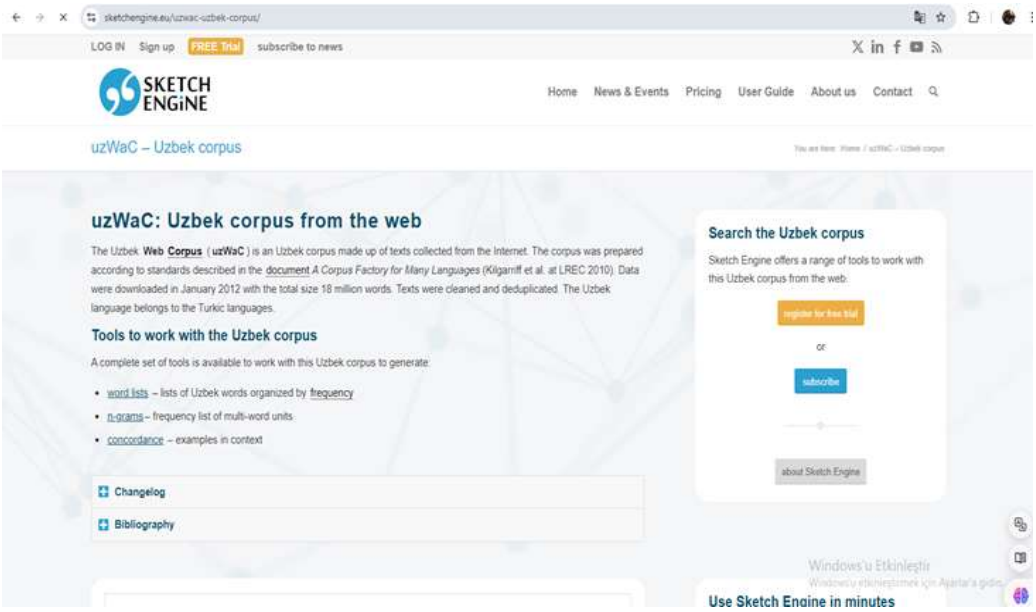
Görsel 6

Özbek Web Derlemi (uzWaC), internetten toplanan metinlerden oluşan bir Özbek derlemidir. Derlemde yer alan 18 milyon kelime 2012 yılının ocak ayında yüklenmiştir. Bu

Özbekçe derlemde; kelime listeleri (sıklığa göre düzenlenmiş), n-gramlar-çok kelimeli birimlerin frekans listesi, konkordanslar yer almaktadır (Adilova, 2021, s. 530).

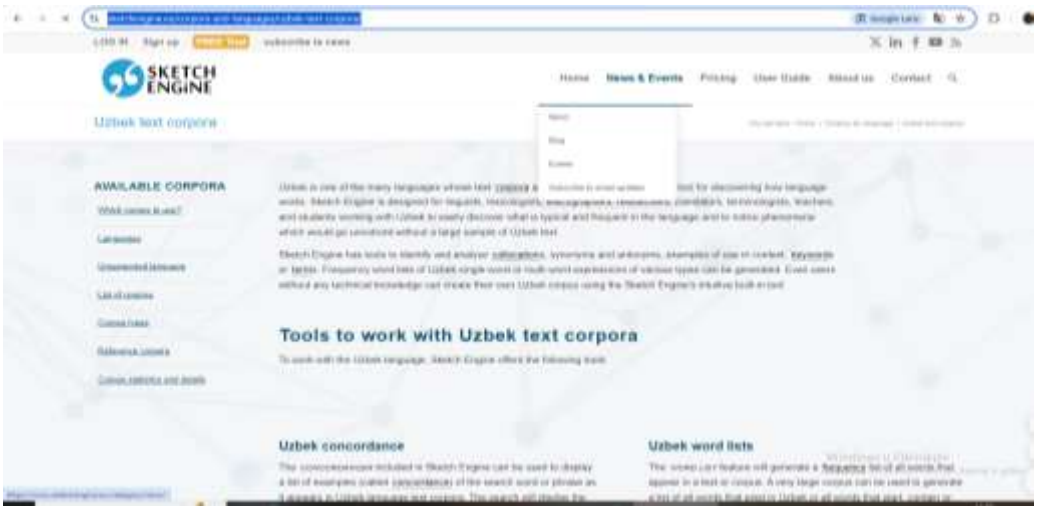
(URL 11**)**

Görsel 7



(URL 12)

Görsel 8



2. Sonuç

El ile fişleme gibi geleneksel yöntemlerle yapılan veri toplama işlemleri oldukça zaman alıcıdır. Derlem çalışmaları dilbilimsel araştırmaların hem daha hızlı hem de daha verimli bir şekilde yürütülmesine olanak sağlar. Derlemler sayesinde kullanıcılar kelime sıklığı, eşdizimlilik ve diğer dil bilgisel özellikler hakkında bilgi sahibi olurlar. Derlemler araştırmacılara diller arasında karşılaştırmalı dilbilimsel çalışmalar yapabilme imkânı sunar.

Dilin dinamik yapısını çözümlemek ve anlamlandırmak için önemli bir araç olarak kullanılan derlemler, dilbilimsel araştırmaların yanı sıra, sözlük çalışmaları, dil öğretimi, dil edinimi, çeviri gibi birçok alanda kullanılmaktadır. Türk lehçeleri arasında önemli bir yere sahip olan Özbek Türkçesi üzerine yapılan derlem çalışmaları hem yazılı hem de sözlü verileri içermektedir. Dolayısıyla bu derlemler, dilin hem günlük hem de edebî dildeki kullanımını çeşitli açılardan incelemeye olanak tanımaktadır. Derlem çalışmalarının genel olarak tüm Türk lehçelerinde yaygınlaşması Türk lehçeleri arasında karşılaştırmalı çalışmaların yapılmasını kolaylaştırmaktadır. Özbekistan'da yapılan derlem çalışmaları, Özbekçenin ikinci yabancı dil olarak öğretilmesini de amaçlamaktadır. Özbek Ulusal Derlemi elektronik ortamda erişilebilen ve dilin çeşitli alanlarını temsil eden metinleri kapsayan önemli bir araçtır. Bu derlem çalışmaları sayesinde dilin hem güncel kullanımını hem de tarihsel gelişimini takip etmek kolaylaşmaktadır ve alanla ilgilenenlere dil ile ilgili yapacakları araştırmalarda kullanacakları çok zengin bir veri tabanına erişim sağlanmaktadır.

Kaynaklar

Abjalova, M., Adalı, E. ve Iskandarov, O. (2023). Educational corpus of the Uzbek language and its Opportunities. *2023 8th International Conference on Computer Science and Engineering (UBMK)*, s. 1-5.

Adalı, E. (2022). Bilgisayarlı çeviri için koşut derlem. *O'zbek Tilining Milliy Korpusi: Muammo va vazifalar mavzusidagi xalqaro ilmiy-amaliy konfrensiya materiallari*, s. 8-26.

Adilova, A. S. (2021). Corpora and corpus-based teaching Uzbek to foreigners. *International Journal of Multicultural and Multireligious Understanding (IJMMU), 8*(4), 525-531.

Aksan, Y., Aksan, M., Özel, S. A., Yılmazer, H., Demirhan, U. U., Mersinli, Ü., Bektaş, Y. ve Altunay, S. (2014). Web tabanlı Türkçe ulusal derlemi (TUD). *Akademik Bilişim'14 - XVI. Akademik Bilişim Konferansı Bildirileri*, s. 723-730.

Gurbanmurat qizi Allaberdiyeva, D. (2024). Turkiy tillar korpusining qiyosiy tadqiqi. *Academic Research in Educational Sciences, 5*(6), 122-126.

Gümüş, M. (2024). Derlem dilbilim çalışmalarının dil öğretimine katkısı. *Journal of Turkic Civilization Studies, 5*(1), 153-170.

Karaoğlu, S. (2014). Türkçe çevirimiçi derlemler üzerine. *KMÜ Sosyal ve Ekonomik Araştırmalar Dergisi, 16*(Özel Sayı I), 181-188.

Raupova L., Elov B., Abjalova M. ve Alayev R. (2021). O'zbek tilining ta'limiy korpusi va uning imkoniyatlari. *O'zbekiston til va Madaniyat-Amaliy Filologiya*, 60-77.

Mengliev, B., Khamroeva, Sh., Elova. D. (2021). Learning corpus of Uzbek language: structure, content, opportunities. (URL 14)

Musurmankulova, Sh. (2023). Theoretical basis of the formation of Uzbek-Turkish parallel corpus, Central Asian Journal of Literature, Philosophy and Culture, (4), 163-170.

Musurmankulova, Sh. (2023). "'Uzturkfraz'" parallel korpusi ma'lumotlar bazasini shakllantirish texnologiyasi". *«Til va adabiyot – Преподавание языка и литературе – Language and literature teaching»*, 84-85.

Osipova, E. (2020). Corpus linguistic technology as a digital tool in teaching idioms' interpretation to EFL students. *IOP Conference Series: Materials Science and Engineering, 940*(1), 1-12.

Özbay, A. Ş., Gürsoy, Z. (2023). Computerized corpus as a tool for educational technology and learning in the analysis of four-word recurrent expressions. *Journal of Educational Technology & Online Learning, 6*(1), 249-272.

Özkan, B. (2020). Türkçenin söz varlığını belirlemede derlem dilbilim uygulamaları. *Turkish Studies - Language, 15*(1), 341-354. (URL 13)

Özkan, B. (2023). Yöntem ve uygulama açısından Türkiye Türkçesi söz varlığının derlem tabanlı sözlüğü. *Bilig*, (66), 149-178.

Pekcoşkun, G. S. (2018). Derlem tabanlı yaklaşımların çeviribilimdeki yeri ve önemi. Yayımlanmamış Doktora Tezi, İstanbul: İstanbul Üniversitesi Sosyal Bilimler Enstitüsü.

Shamsudinova, B. I. (2022). The problem of parameterization of auxiliary words in the database of the national corpus of the Uzbek language and providing them in the interface. *ASEAN Journal on Science & Technology for Development, 39*(4), 693-698.

Sharipov, F. ve Sharipova E. (2024). O'zbek tilshunosligida milliy korpus masalasi. *Global integratsiya sharoitida Turkiy tillar taraqqiyoti: muammo va vazifalar mavzusidagi xalqaro ilmiy-amaliy anjuman materiallari*, 318-321.

Tahiroğlu, B. T. (2010). Derlem, bilgisayar destekli sözlük bilimi, eş dizimlilik ve otomatik terim çıkarımı. *Belleten, 58*(1), 183-197.

Toirova, G. (2023). Creation and importance of language corps in Uzbekistan. *Aquaculture*, 1-12.

Zokirova, Izzatillayevna. Kh, (2023). Advancements in Computer Linguistics in Uzbekistan. *Models and Methods For Increasing The Efficiency of Innovative Research, 3*(29), 213-218.

İnternet kaynakları

URL 1 http://www.tnc.org.tr/ 05.08.2024
URL 2 https://www.ddi.itu.edu.tr/araclarkaynaklar 07.08.2024
URL 3 https://qazcorpus.kz/, https://github.com/makazhan/kaznlp 15.5.2024
URL 4 https://corpora.clarin-d.uni-saarland.de/cqpweb/kyrgyz 21.06.2014
URL 5 https://tugantel.tatar/?lang=en 12.07.2024
URL 6 https://ctcorpus.org/index.php/crh/ 10.06.2024
URL 7 https://korpus.azerbaycandili.az/ 17.09.2024
URL 8 https://uzschoolcorpara.uz/ 19.06.2024
URL 9 https://uzbekcorpus.uz/newIndex 21.09.2024
URL 10 https://uzturkfraz.uz/ 20.08.2024
URL 11 https://www.sketchengine.eu/uzwac-uzbek-corpus/ 17.09.2024
URL 12 https://www.sketchengine.eu/corpora-and-languages/uzbek-text-corpora/06.09.2024
URL 13 https://dx.doi.org/10.29228/TurkishStudies.41460 22.09.2024

URL 14 chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/http://www.turklang. 01.08.2024

Extended Abstract

The rapid advancement of technology has led to a growing significance of computer-assisted linguistics studies and applications. Corpus linguistics, which constitutes an important field of linguistics, provides an optimal environment for computer-aided language research and applications. Corpus linguistics occupies a significant position within the field of modern language studies, with the number of studies in this area increasing day by day. Corpora are created for a variety of purposes, including the revelation of similarities and differences between languages, the analysis of language structure and meaning, the creation of data sets for the development of language models with natural language processing applications, the facilitation of learners' comprehension of languages with bilingual expressions in language teaching, the analysis of contexts and expression techniques in literary works, and the advancement of individuals' language skills. Additionally, they are employed in the analysis of Turkish dialects, with the number of corpus-based project-supported studies increasing in recent years. It is of great importance to consider the contributions of corpus studies to the comparative study of Turkish dialects. In particular, parallel corpus studies serve as a valuable resource for enhancing research in a range of fields, including comparative linguistics, etymology, dialectology, folklore studies, textual studies, translation theory, and literary studies.

By leveraging the capabilities of this corpus linguistics, it is possible to examine the actual uses of language and gain insights into the grammatical structures and general language usage of Turkish dialects in comparison. In recent years, national corpora of numerous Turkish dialects have been established, contributing to this growing body of knowledge. There are web-based national corpora of Turkish dialects, including Kazakh Turkish, Kyrgyz Turkish, Uzbek Turkish, Tatar Turkish, Azerbaijani Turkish, and especially Turkey Turkish. Notable examples of corpus linguistics applications in Turkish dialects include the Turkish National Corpus, the Almaty Corpus of Kazakh language, the Tatar National Corpus named "Tugan Tel", the Crimean-Tatar Turkish National Corpus, and the Uzbek National Corpus.

Despite an increase in corpus studies on language in Uzbekistan, which is the primary focus of the present article, over the past decade, studies in computerised linguistics, which form the foundation of corpus linguistics, can be traced back to the 1940s. However, since the country gained independence, studies in this field have been intensified, and numerous corpus-based projects have been implemented. One such corpus is the Uzbek education corpus. The corpus provides students with the opportunity to gain a comprehensive understanding of Uzbek as a state language, second language and foreign language. The corpus comprises a range of digital materials, including audio, video, multimedia applications, pronunciation and spelling software, and e-learning dictionaries. This corpus, which is employed in the teaching of Uzbek as a mother tongue and as a foreign language, is also highly beneficial for illustrating the wealth of vocabulary in the language and demonstrating the potential for word usage through grammatical structure. The National Corpus of Uzbek enables the automatic analysis of the morphological structure and semantic features of a text. The creation of the National Corpus of Uzbek is also important in terms of facilitating researchers in conducting linguistic studies. This corpus is an important basic resource for linguists, lexicographers, editors, translators, journalists, publishers, scientists, teachers, students and all other language users. Notable contributions to this field have been made by Uzbek researchers such as Sh. Hamroyeva, O. Kholyirov, D. Akhmedova, and A. Eshmo'minov. In addition to the aforementioned studies, one of the earliest parallel corpora created among Turkic dialects is the Uzbek-Turkish Idioms Corpus. The Uzbek idioms included in this parallel corpus were created based on the work titled *O'zbek tilining izohli frazeologik lug'ati* prepared by Shavkat Rahmatullayev, and the Turkish idioms included in the corpus were based on *Atasözleri ve Deyimler Sözlüğü* by Ömer Asım Aksoy. The corpus allows for the comparative examination of the semantic and structural features of idioms belonging to both dialects.

The present study emphasises the concept of the corpus, presents examples of corpus applications widely used in world languages in Turkish dialects, and outlines the development of corpus studies in Uzbekistan from the past to the present. It also introduces corpus-based projects and studies carried out in Uzbekistan.