

AZƏRBAYCAN RESPUBLİKASI

Əlyazması hüququnda

**MİLLİ DİL KORPUSLARI VƏ ONLARIN
FORMALAŞMASI PRİNSİPLƏRİ
(Britaniya və Amerika ingiliscəsi materialları əsasında)**

İxtisas: 5704.01 – Dil nəzəriyyəsi
Elm sahəsi: filologiya – dilçilik

İddiaçı: **Sevinc Səlim qızı Məmmədzadə**

Fəlsəfə doktoru elmi dərəcəsi almaq üçün
təqdim edilmiş dissertasiyanın

AVTOREFERATI

Bakı – 2021

Dissertasiya işi Odlar Yurdu Universitetinin "Dillər və pedaqoji" kafedrasında yerinə yetirilmişdir.

Elmi rəhbər:

Filologiya elmləri doktoru, professor
Məsməxanım Yusif qızı Qaziyeva

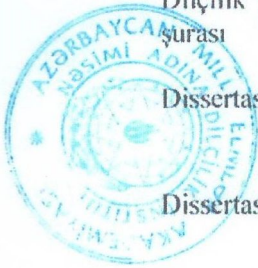
Rəsmi opponentlər:

filologiya elmləri doktoru, dosent
Kamilə Abdulla qızı Vəliyeva

filologiya üzrə fəlsəfə doktoru, dosent
Cavanşir Xankişi oğlu Muradov

filologiya üzrə fəlsəfə doktoru
Ramilə Misirxan qızı Fərəcova

Azərbaycan Respublikasının Prezidenti yanında Ali Attestasiya Komissiyasının Azərbaycan Milli Elmlər Akademiyası Nəsimi adına Dilçilik İnstitutu nəzdində fəaliyyət göstərən ED 1.06 Dissertasiya şurası



Dissertasiya şurasının sədri:

akademik

Möhsün Zellabdin oğlu Nağısoylu

Dissertasiya şurasının elmi katibi:

filologiya üzrə fəlsəfə doktoru,
dosent

Sevinc Yusif qızı Məmmədova

Elmi seminarın sədri:

filologiya elmlər doktoru,
dosent

Gülsüm İsrafil qızı Hüseynova

GİRİŞ

Mövzunun aktuallığı və işlənmə dərəcəsi. Müasir dövrdə kompüter texnologiyasının sıçrayışlı irəliləyişi, çap və nəşr işlərinin kompüterləşdirilməsi, internetin yaranması, informasiya mübadiləsinin sürətlənməsi son dərəcə böyük miqdarda dil materialının elektron daşıyıcılarında toplanıb qalması həmin mənbələrdən istifadə etmək imkanlarından bəhrələnməyi tələb edir. Dil materialının, xüsusən sözləri kartlara qeyd edərək linqvistik vasitələrin (lüğətlərin) tərtibində tətbiq etmək, bu cür kartotekaları qoruyub saxlamaq mümkünlüyü müasir dilçilikdə korpus linqvistikası istiqamətinin meydana gəlməsinə səbəb olmuşdur.

Kompüter dilçiliyinin yeni bir qolunu təşkil edən korpus linqvistikası ikili mahiyyət daşıyır. Bir tərəfdən, korpusların qurulması, onun altfondlarını təyin etmək, korpus mətnlərini zəruri annotasiyalarla təchiz etmək, korpuslararası əlaqələri qurmaq, digər tərəfdən, korpusdan istifadə etməklə linqvistik araşdırmalar aparmaq, korpusun daha hansı məsələlərin həllində yardımçı olacağını müəyyənləşdirmək lazım gəlir. Müxtəlif dillərin, o cümlədən də ingilis dilinin riyazi və statistik metodların köməyi ilə tədqiqi, dilin kəmiyyət göstəricilərini sinxron və diaxron planda öyrənmək, tezlik lüğətlərinin tərtibi, linqvistik informasiyanın avtomatik axtarış sistemlərini qurmaq, maşın tərcüməsi, avtomatik təhlil sistemlərini yaratmaqla bağlı bu günə qədər xeyli iş görülmüşdür. Lakin bu tədqiqatlar, tətbiqi araşdırmalar kifayət qədər iri həcmli dil materialları əsasında aparılmamışdır. Korpuslar material həcminə görə kifayət qədər böyük və geniş olduğundan indiyədək aparılmış tədqiqatlardan əldə edilmiş nəticələrin dəqiqliyini yoxlamaq, reprezentativ nəticələr verəcək müasir araşdırmalar aparmaq imkanları meydana çıxmışdır.

Korpus linqvistikası kompüter dilçiliyinin bir qolu sayılır. Tezlik lüğətlərinin tərtib olunması ideyasının işığında formalaşmış bu yeni linqvistik istiqamət maşın tərcüməsi, riyazi-statistik metodların dilçilikdə tətbiqi, nəhayət, təbii dillərin emalı

sistemlərinin yaradılması kimi problemlərlə əlaqəli inkişaf edir. Korpus linqvistikasının gələcək inkişafı onun bütün kompüter dilçiliyi məsələlərini özündə birləşdirə biləcəyini istisna etmir.

Korpuslardan istifadə etmək müasir dilçiliyin səciyyəvi cəhətlərindən biridir. Korpuslardan alınan müxtəlif xüsusiyyətli linqvistik məlumat və materiallar geniş çeşiddə dil problemlərinin həllində tətbiqini tapır. Bir sıra linqvistik faktların aşkara çıxarılması, belə faktların qanunauyğunluğu və təsadüfiliyinin təsdiqi olduqca böyük həcmə malik dil materialını araşdırmağı, təhlilə cəlb olunmasını tələb edir. Bu məsələnin həlli yalnız korpus linqvistikası səviyyəsində mümkündür. Korpus linqvistikası dilin istər tətbiqi, istər tədrisi, istərsə də tədqiqi problemlərinin həllində istifadə olunur. Bütün bunlar korpus linqvistikası ilə bağlı tədqiqatın aktuallığını təsdiq edən amillərdir.

Tədqiqatın obyektı və predmeti. Tədqiqatın obyektı bütövlükdə korpus linqvistikasıdır. İşdə korpus dilçiliyi ilə kompüter dilçiliyinin başqa istiqamətləri arasındakı əlaqələr də tədqiqatın obyektı səviyyəsinə qaldırılmışdır. Korpusların yaradılmasının məqsəd və vəzifələri, mövcud korpusların xüsusiyyətləri, dilin sosial və regional dialekt variantlılığı tədqiqat işinin predmeti kimi tədqiqat müstəvisinə çıxarılmışdır.

Tədqiqatın məqsəd və vəzifələri. Tədqiqatın əsas məqsədi korpusların qurulması, onların altfondlarının müəyyənləşdirilməsi, korpusların linqvistik məsələlərin həllində tətbiqi imkanlarını öyrənməkdir. Bu məqsədlə aşağıdakı vəzifələrin həlli nəzərdə tutulmuşdur:

– korpus linqvistikasının təşəkkülü və formalaşmasına təkan vermiş amilləri müəyyənləşdirmək, ilk korpusların yaranma səbəblərini öyrənmək, korpus anlayışını təyin etmək;

– əsas linqvistik korpusların təsvirini vermək, onların səciyyəvi xüsusiyyətlərini aydınlaşdırmaq, korpusa daxil edilən mətnlərin seçilmə meyarlarını mövcud korpuslar üzrə təhlil etmək;

– korpusların təsnif meyarlarını müəyyənləşdirməklə fərqli korpusların qurulması xüsusiyyətlərini öyrənmək, dil materiallarının janr, dialekt, nitq fərqliliyinə görə qruplaşdırılaraq korpusun

qurulması məsələsini təhlil etmək;

– standart dil, dialekt, dil variantı, slenq üzrə korpusların qurulma prinsiplərini öyrənmək, standart ingilis dili, ingilis dilinin variantlar və dialektləri üzrə korpusların oxşar və fərqli cəhətlərini təsvir etmək, Braun korpusu, LOB və FLOB korpusları üzrə ingilis dili variantlarının müqayisəli təhlilini vermək;

– kompüter dilçiliyində korpus linqvistikasının yerini aydınlaşdırmaq;

– korpusların annotasiyalaşdırılmasının əsas məsələlərini öyrənmək, korpusdan istifadə imkanlarını və korpusun genişləndirilməsi üsullarını araşdırmaq;

Tədqiqatın metodları; Dissertasiya işində təsviri, tarixi-müqayisəli və statistik-distributiv metodlardan istifadə edilmişdir.

Müdafiyyə çıxarılan əsas müddəalar;

1. Linqvistik korpusların yaradılması mövcud təcrübəyə əsaslanan və daim təkmilləşən prosesdir.

2. Braun korpusu və ingilis dilinin müxtəlif variantları, dialektləri üzrə yaradılan korpuslar Azərbaycan dilinin dialektləri, həmçinin türk dilləri üzrə korpusların yaradılmasının mümkünliyünü təsdiq edir.

3. Tematik elektron korpuslar tətbiqi məsələlərin həlli üçün effektiv vasitə rolunu oynayır. Korpuslar üzrə müxtəlif aspektli linqvistik məsələlərin həlli üçün geniş imkanlar açılır.

4. Korpuslarda informasiyanın optimal axtarışı korpusa daxil edilən mətnlərin markerlənməsindən asılıdır. Markerləmənin düzgün təyin edilməsi korpusun istifadə olunmaq imkanlarını artırır.

5. Linqvistik korpusların altfondları dairəsinin genişləndirilməsi mümkündür. Korpuslara xüsusi terminoloji altfondun daxil edilməsi terminoloji fondun zənginləşmə xüsusiyyətlərini təyin etmək üçün material verir. İngilis dilinin terminoloji korpusu elektron terminoloji lüğətlərin tərtibində tətbiqini tapa bilər.

6. Kompüter dilçiliyi və korpus linqvistikası bir-biri ilə sıx bağlı olan istiqamətlərdir. Kompüter dilçiliyinin həll etdiyi bütün məsələlər korpus çərçivəsində yerinə yetirilə bilər.

7. Korpus linqvistikası avtomatik təhlil və maşın tərcüməsi

məsələsinə yeni aspektdən yanaşmaq imkanı yaradır.

8. Linqvistik korpus açıq sistem olmalı və onun daim zənginləşməsi nəzərdə tutulmalıdır.

9. Müasir kompüter texnologiyasının inkişafı korpus linqvistikasının perspektivlərini artırır və linqvistik bazanın həcmnin kifayət qədər artırılması mümkündür.

Tədqiqatın elmi yeniliyi; İşdə ilk dəfə olaraq korpus dilçiliyinin tarixi, ilk korpuslar və onların ən müasir korpuslardan fərqli cəhətləri, habelə korpusun kompüter dilçiliyində tətbiqi kompleks şəkildə araşdırılmışdır. Dissertasiyada korpus linqvistikasının əsasında duran tətbiqi məsələlərdən linqvistik və kompüter dilçiliyinin nəzəri məsələlərinə keçid izlənilir, korpusun ikili mahiyyəti açılır. Tədqiqat işində kompüter dilçiliyi sahələri arasındakı əlaqələrin öyrənilməsi məsələsi ilk dəfə ortaya atılır və bu istiqamətdə lüğətlərin tərtibi, maşın tərcüməsi və korpus dilçiliyi müqayisəli planda araşdırılır. Tədqiqat işinin ən başlıca elmi yeniliyi Azərbaycan dilçiliyi sahəsində ilk dəfə olaraq Britaniya və Amerika ingiliscəsinin sosial və regional dialektləri araşdırılaraq onların korpuslarının tərtibatı və bu korpusların kompüter dilçiliyi sahəsinə və maşın tərcüməsinə tətbiqi, eyni zamanda tətbiq və istifadə zamanı qarşıya çıxan problemlər və çətinliklər təhlil edilərək araşdırılır.

Tədqiqatın nəzəri və praktik əhəmiyyəti; Tədqiqatın əsas nəzəri məsələləri korpusun annotasiyalaşdırılması problemini əhatə edir. Aparılmış araşdırmalar markerləmənin ortaya atılmış problemlə birbaşa bağlılığını sübut edir. Korpus təcrübəsinin ümumiləşdirilməsi müasir milli korpusların yaradılmasının nəzəri problemlərinə yeni baxış bucağı altında baxmağa imkan verir. İşin nəzəri əhəmiyyətinin bir aspektini Braun, LOB, FLOB və digər korpusların təcrübəsinin Azərbaycan dili materialları əsasında təhlili ilə bağlıdır.

Dissertasiya işindən əldə olunmuş nəticələrdən dilin milli korpusunun qurulmasında, bu korpusun altfondlarını yaratmaqda istifadə oluna bilər. Bundan əlavə, dissertasiya işi kompüter dilçiliyinə dair seminarların və xüsusi kursların tədrisində istifadə edilə bilər. Tədqiqatın nəticələri mətnin markerlənməsində də tətbiqini tapa bilər.

İşin aprobasiyası və tətbiqi; Dissertasiya mövzusu üzrə “Koqnitiv və tətbiqi dilçiliyin aktual problemləri” (20-21 Oktyabr 2016), “Gənc tədqiqatçıların IV beynəlxalq elmi konfransı” (29-30 aprel 2016), “Müasir təhsilin inkişafında beynəlmilləşmənin rolu” (29 Aprel 2019), “Вопросы современной науки: проблемы, тенденции и перспективы” (13 Сентябрь 2019), “Филологические и социокультурные вопросы науки и образования” (25 Октябрь 2019) adlı beynəlxalq konfranslarda, eləcə də “Doktorantların və gənc tədqiqatçıların XX respublika elmi konfransı” (24-25 May 2016) və “Birinci Türkoloji Qurultayın 90 illik yubileyinə həsr olunmuş respublika elmi konfransı”nda (28 Dekabr 2016) məruzələr edilmişdir. Dissertasiya mövzusu üzrə 8 məqalə, 5 tezis çap edilmişdir.

Dissertasiya işinin yerinə yetirildiyi təşkilatın adı. Dissertasiya Odlar Yurdu Universitetinin Dillər və pedaqoji kafedrasında yerinə yetirilmişdir.

Dissertasiyanın struktur bölmələrinin ayrılıqda həcmi qeyd olunmaqla dissertasiyanın işarə ilə ümumi həcmi. Dissertasiya giriş, üç fəsil, nəticə, istifadə edilmiş ədəbiyyat siyahısı, ixtisarların və şərti işarələrin siyahısından ibarətdir. Dissertasiyanın giriş hissəsi 6 səhifə, birinci fəslə 40 səhifə, ikinci fəslə 54 səhifə, üçüncü fəslə 33 səhifə, nəticə hissəsi 2 səhifə, ədəbiyyat hissəsi 11 səhifə, ixtisarlar və şərti işarələr siyahısı 1 səhifədən ibarətdir. Dissertasiya işi ümumilikdə 148 səhifə, 246,314 işarədən ibarətdir.

İŞİN ÜMUMİ MƏZMUNU

Dissertasiyanın **Giriş** hissəsində mövzunun aktuallığı, işlənmə dərəcəsi əsaslandırılır, obyekt və predmeti, məqsəd və vəzifələri, tədqiqatın metodları, müdafiəyə çıxarılan müddəalar müəyyənləşdirilir, tədqiqatın elmi yeniliyi, nəzəri və praktiki əhəmiyyəti, tədqiqat işinin aprobasiyası və tətbiqi, dissertasiya işinin yerinə yetirildiyi təşkilatın adı, dissertasiyanın struktur bölmələrinin ayrılıqda həcmi və işarə ilə ümumi həcmi haqqında məlumat təqdim edilir.

Dissertasiyanın *birinci* fəslə “Korpus dilçiliyi və onun

yaradılması” adlanır. Bu fəslin ilk bölməsində korpus dilçiliyinin tarixinə ekskurs edilir. Korpus dilçiliyi – kompüter dilçiliyinin bir hissəsi olaraq kompüter texnologiyasının köməyi ilə dil korpuslarının (korpus mətnlərinin) yaradılması və öyrənilməsinə yönələn bir elm sahəsidir. Ən müasir tətbiqi dilçilik sahələrindən biri olan korpus linqvistikasının qədim tarixi kökləri vardır. Bu köklər dil materialının toplanıb sistemə salınması ilə bağlıdır.

İstər kartotekaların tərtibi, istərsə də onların saxlanması müxtəlif çətinlik və problemlərlə üzləşməyə səbəb olurdu. Məsələn, dil tarixinə dair sözləri abidələrdən seçib toplamaq və sözün işlənməsinə aid nümunələri kartlara daxil etmək üçün çox sayda tədqiqatçının əməyindən istifadə olunurdu. Tədqiqatçılar bu işə illərlə vaxt sərf edirdilər. Toplanmış kartları sözlərə görə əlifba sırası ilə düzmək, nəhayət, onları yazı makinasında yazmaq lazım gəlirdi. Eyni zamanda, bu kartotekanın qorunub saxlanması üçün xüsusi dolablar hazırlatdırmaq, dolabların sayı artdıqca saxlanma otaqlarını genişləndirmək və s. bu kimi problemlər ortaya çıxırdı. Elektron maşınlardan istifadə, sonralar kompüter texnologiyasının inkişafı problemlərin həlli üçün yeni imkan və şərait yaratdı.

Elektron daşıyıcılarında ilk linqvistik korpusun qurulmasına 1960-cı ildə qərar verilmişdir. ABŞ-da Braun Universitetində müasir Amerika ingilis dilinin standart korpusu 1961-ci ildən 1964-cü il daxil olmaqla qurulmuşdur. İlk kompüterləşən korpus- Braun korpusu (The Brown Corpus) tərkibinə amerikan qəzet, jurnal və kitablarından təxminən 500 mətn daxil idi. Braun korpusuna daxil olan hər bir mətn 2000 sözdən ibarət olan 500 mətni, ümumilikdə isə, 1 milyon sözü əhatə edirdi. Korpusun müəllifləri U.Frensis və Q.Kuçera külli miqdarda material və toplulardan istifadə etmişdilər.

İlk olaraq ABŞ-da ingilis dilinin korpusunun yaradılması həm texnologiyanın ABŞ-da inkişafından, həm də korpus ideyası və onun qurulma səbəbləri, məqsədlərinin düzgün əsaslandırılmasından, eləcə də, ölkədə istifadə olunan bu dilə bütün dünyada olan maraqdan asılı olmuşdur. Braun korpusunun yaradılması bütün dünyada dilçilərin və qeyri-dilçilərin ingilis dili sahəsində araşdırmalarına xeyli kömək olmuşdur. Bu da dilçilik elminin, həmçinin ingilis dilinin inkişafına

təkan vermişdir.

İkinci kompüter korpusunu Lankaster və Oslo Universitetləri ilə Berqen elmi mərkəzi birgə hazırlamışdır. 1970-78-ci illərdə hazırlanmış bu korpus universitetlərin və elmi mərkəzin yerləşdiyi şəhərlərin adlarının birinci hərfləri əsasında düzəldilmiş abreviaturla adlandırılmışdır: LOB (Lankaster-Oslo-Bergen). LOB linqvistik korpusunun strukturu Braun korpusunun strukturuna oxşardır. Korpusa bir milyon sözforma daxil edilmişdir. İstər Braun, istərsə də LOB korpusların mətn həcmnin bir milyon sözforma götürülməsi bu miqdarın dildəki bütün sözformaları əhatə etməsi ilə əsaslandırılmışdır. Onu da qeyd edək ki, Braun korpusu Amerika, LOB korpusu isə Britaniya ingilis dili materialını əhatə edir. Hər iki korpus bu gün də çox aktual və yararlıdır. Onlar üzərindən çoxsaylı araşdırmalar aparılır və ingilis dilinin incəlikləri bu korpusların köməyi ilə öyrənilir.

Bu korpusların yaradılmasından sonra daha sürətlə işləyən, böyük yaddaş həcminə malik kompüterlər istehsalı başlanmışdır. Bundan başqa skaynerlərin köməyi ilə mətnləri kompüterə daxil etmək işi də asanlaşmışdır. Əvvəllər bu iş klaviaturanın köməkliliyi ilə uzun zaman ərzində həyata keçirilirdi. Skaynerlər və mətni tanıma proqramları vasitəsilə əvvəl çap olunmuş mətnlərin maşın yaddaşına köçürülməsi mətnlərin toplanmasını sürətləndirdi. Nəticədə, korpusa bir milyon deyil, milyardlarla söz toplusundan ibarət mətnlər daxil etmək imkanı əldə olundu. 1990-cı ildə artıq 600 dan çox kompüter korpusu mövcud idi.

İngilis dilinin müasir korpusları arasında Britaniya milli dil korpusu (British National Corpus – BNC), müasir Amerika ingiliscəsi korpusu (Corpus of Contemporary American English – COCA) mühüm yer tutur. Korpus ingilis dilinin müəyyən bir sahəsini araşdırarkən həmin dilin variantları ilə bağlı materiallardan istifadə etmək imkanı verir. Bu istiqamətdə çətinlikləri aradan qaldırmaq üçün BNC və COCA-nın böyük əhəmiyyəti vardır.

Artıq 1990-cı ilin birinci yarısında korpus dilçiliyi ayrıca bir sahə kimi formalaşmağa başladı. C.Svartvik 1992-ci ildə korpus dilçiliyi haqqında məqaləsinin girişində bu cümləni qeyd etmişdir:

“Korpus dilçiliyi öz yetkinlik çağına çatdı.”¹ Korpus dilçiliyi kompüter dilçiliyi ilə sıx əlaqədədir, həm ondan istifadə edir, eyni zamanda onu zənginləşdirir.”²

Müasir dövrdə korpus haqqında dilçilər tam məlumatlı olmasalar belə, onlar korpusun nə qədər böyük əhəmiyyətə malik olduğunu qəbul edirlər. *“Dilin korpusu hər hansı bir dildə olan mətnlərin elektron toplusudur”³* fikrini bu cür izah edə bilərik. Normal insan mətni oxuyub hər hansısa bir məlumat almaq üçün istifadə etdiyi halda, dilçi bu mətni hissələrə bölməyə başlayır. Məsələn, qrammatik hissələrə, leksik hissələrə və s. Dilçilər mətnləri hissələrə bölür, bu şəkildə ayrılmış, artıq hissələrə bölünmüş, işlənmiş mətnlər korpusda yer alır.

Toplanması illərlə vaxt alan ədəbiyyatları və materialları müasir dövrdə qısa bir zamanda cəmləşdirmək mümkündür. İndi vaxt materialın toplanmasına deyil, araşdırmalara sərf olunur. Korpus çətin işləri xeyli yüngülləşdirir. Dilçilər üçün korpusun rolu əvəzədilməzdir. Korpusdan əldə edilən materialın miqdarı və keyfiyyəti korpusa qədərki dövrlə müqayisəedilməz dərəcədə yüksəkdir. Bir dili araşdırarkən 10000 və daha çox nümunələrdən istifadə etmək 10 nümunə istifadə etməkdən üstün sayılır. Bu 10000 nümunəni tapmaq üçün isə həmin dilin korpusundan istifadə edilir. Beləliklə, korpus linqvistikasının ötən əsrin 60-cı illərindən başlayaraq meydana çıxması, bir sıra korpusların hazırlanması son nəticədə korpus dilçiliyinin aktuallığını və vacibliyini təsdiq etmişdir.

Bu fəslin ikinci bölməsində ingilis dilinin əsas korpusları nəzərdən keçirilir. Britaniya ingiliscəsinin milli korpusu (British National Corpus, BNC) böyük korpuslardan biri hesab olunur. Tərkibinə 100 milyon söz daxil edilmişdir. Korpus Oksford universitetində Lankaster universiteti və Britaniya kitabxanasının

¹ Svartvik, J. Directions in Corpus Linguistics / J. Svartvik. – 1991. – p. 116.

² Захаров, В.П. Корпусная лингвистика / В.П. Захаров, А.С.Злыгостева. – Санкт-Петербург: – – 2013. – с. 13.

³ Кулагина, О.С. Исследования по машинному переводу / О.С. Кулагина. – Москва: Наука, – 1979. – с. 14.

iştirakı ilə yaradılmışdır. Korpusun 90%-i yazılı ingilis dilini təmsil edir, tərkibində bədii və sənəd nümunələri, qəzet və jurnal materialları, elmi əsərlər, məktəb və universitet inşaları və s. cəmləşdirir. *Korpus, həmçinin, müxtəlif üslubları da əhatə edir və mövzu etibarilə çox zəngindir. Bu korpus yalnız müasir ingilis dilində olan mətnlərdən ibarətdir. Korpusda həmçinin Britaniya ingiliscəsində işlənən alınma sözlər yer almışdır*⁴.

Britaniya milli korpusunda təqdim olunan mətnlər 3 əsas kriteriya ilə seçilib toplanmışdır: zaman, məkan və yaranma tipi. Zaman baxımından bütün mətnlər eyni dövrü əhatə edir. Bura 1975-ci illərdən sonrakı yaradılmış mətnlərdən nümunələr daxil edilmişdir. Bədii ədəbiyyat nümunələri korpusun ümumi həcmünün 25%-ni təşkil edir. Yazılı mətnlərin 75%-i elm, mədəniyyət, sosiologiya və siyasət sahəsindən götürülmüşdür.

Müasir Amerikan dilinin korpusu müxtəlif janrları özündə birləşdirən və pulsuz daxil olma imkanı olan ən böyük korpus hesab olunur. Korpus əsasən 5 janrı əhatə edir: şifahi, bədii ədəbiyyat, tanınan jurnal və qəzetlər, akademik nəşrlər. Şifahi janr materialı 85 mln sözü birləşdirir, təxminən 150 televiziya kanalından və radio verilişlərindən toplanmışdır. Ədəbi janr 81 mln sözformadan ibarətdir. Bura qısa nağıl və pyeslər, kinofilmlərin ssenariləri də daxil edilmişdir. *Jurnallardan seçmələr 86 mln sözforma həcmindədir. Korpus üçün müxtəlif mövzuda olan 100 jurnaldan material seçilmişdir. Bu jurnallar xəbərlər, idman, dini və digər mövzularda olmuşdur. Qəzetlər (81mln sözforma) ABŞ-ın 10 qəzetinin idman, xəbər, maliyyə və iqtisadiyyat bölmələrini əhatə etmişdir. Akademik nəşrlərdən seçmələr (81mln sözforma) təxminən 100 elmi jurnaldan nümunələri özündə birləşdirir*⁵.

Bu fəslin üçüncü bölməsi korpusların təsnifinə həsr olunmuşdur. Korpusların təsnifi seçilmiş meyardan asılıdır. Format meyarına görə, korpus qeyri-elektron və elektron korpus tiplərinə bölünür. Qeyri-elektron korpuslar daha əvvəlki dövrlərə aiddir.

⁴ Захаров, В.П. Göstərilən əsəri, – s. 65.

⁵ Косякова, М. Корпус современного американского английского: [элэктр. ресурс] / презентация – 2014. - <http://www.slideshare.net/>, – с. 3-6.

Hazırda yalnız elektron korpuslardan istifadə olunur. Korpus təsnifində tətbiq edilən ikinci meyar mətnlərin korpusa daxil olunma xüsusiyyətidir. Braun korpusuna müxtəlif mətnlərdən təqribən eyni həcmli mətn hissələri toplanmışdır. Ona görə də, bu korpus seçmə mətnidir. Korpuslar tam mətnli də olur. Bu halda tam mətn korpusa daxil edilir. Yazılı abidələrin korpusuna abidənin bütün mətni salınır. Ümumiyyətlə, korpusa mətnin tam şəkildə daxil edilməsi qarşıya qoyulan tədqiqat vəzifələrindən asılıdır. Korpuslar statik və yenilənən olmaqla iki qrupa bölünür.

Ümumiyyətlə, korpusların təsnifi meyarları dairəsi kifayət qədər genişdir. Bu meyarlar korpusların təyinatı ilə bağlılığa malikdir. Kompüter dilçiliyi inkişaf etdikcə korpuslardan istifadənin yeni formaları axtarılıb tapılır. Korpus linqvistikasının gələcək perspektivləri və tipləri həm də korpus materiallarının annotasiyalaşdırılmasına əsaslanır. Morfoloji, sintaktik, semantik, terminoloji markerləmə ən azı bu dörd korpusu qeyd etməyin zəruriliyini təsdiqləyir.

Müxtəlif tip korpusların sayı kifayət qədərdir. V.P.Zaxarova görə: *“Qoyulmuş məqsəd və göstəricilərdən asılı olaraq müxtəlif tip korpusları fərqləndirmək olar”*⁶. Əlavə edək ki, onun klassifikasiyasını tam deyil, yalnız şərti qəbul etmək olar. Belə ki, korpuslar yaranma məqsədinə görə çoxməqsədli və xüsusiləşmiş olur. Çoxməqsədli korpuslar müxtəlif janrlarda olan mətnləri toplayır, xüsusiləşmiş korpuslar isə, yalnız bir və ya bir qrup janrı əhatə edir. Mətnlərin korpusları janrlara görə qruplaşa bilər: onlar ədəbi, folklor, dramaturji, publisist və s. qruplara ayrıla bilər. İşarələmə kriteriyası korpusu kodlaşmış (işarələnmiş) və kodlaşmamış (işarələnməmiş) qruplara bölür. Bu bölgünü başqa adla da adlandırmaq olar. Məsələn, indeksləşmiş və indeksləşməmiş. İndeksləşmiş korpusda sözlərə və ya cümlələrə işarələr (teqlər) qoyulur. (morfoloji, sintaktik, semantik və s.)

Korpusun istifadəçisi üçün vacib meyarlardan biri ondan rahat istifadədir. Azad girişi olan korpuslarda online vasitə ilə istənilən

⁶ Захаров, В.П. Göstərilən əsəri, – s. 16-17.

zaman mətnlərə tam həcmdə axtarış vermək olar. Tarixi korpuslarda isə, ondan istifadə haqqını almaq lazım gəlir. Bağlı korpuslar xüsusi məqsədlər üçün istifadəyə nəzərdə tutulmuşdur və xalq istifadəsi qadağandır. Mətnlərin həcmi kriteriyasına görə korpuslar birdilli, ikidilli və çoxdilli korpuslara bölünürlər. Birdilli korpuslarda dilin variantları, dialektləri qarşılaşdırılır. Məsələn, ingilis dilinin variativləri kimi ikidilli və çoxdilli korpusları iki əsas tipə bölmək olar: 1) çox sayda orijinal mətnləri və bu mətnlərin bir və ya bir neçə başqa dillərə olan tərcümələrini nümayiş etdirən korpuslar; 2) iki və ya bir neçə dildə yazılmasından asılı olmayaraq eyni sahəni əhatə edən mətnləri birləşdirən korpuslar.

Bu cür korpuslar tərcüməçilər tərəfindən ən çox istifadə olunan korpuslardır. Hər iki tip korpuslardan tərcümə zamanı, maşın tərcüməsində, terminoloji lüğətlərin tərtibatında, həmçinin dillərin müqayisəli araşdırılmasında geniş istifadə olunur.

Dissertasiya işinin “Regional və sosial dialektlərin korpuslarının tərtibi məsələləri” adlanan *ikinci* fəslində standart dil korpuslarının yaradılmasında sosial amillərin rolu araşdırılır. İngilis dili digər dillərə təsir edərək söz mübadiləsi edir və dünya ingilis dilinin inkişafına təkan verir. “*İngilis dilinin yayılmasının əsas göstəricilərindən biri yeni leksemlərin yaranması, dilin lüğət fondunun semantik strukturunun dəyişməsidir*”⁷. Bununla yanaşı, ingilis dilinin müxtəlif ölkələrdə rəsmi dil kimi tətbiqi bu ölkələrdə də yeni sözlərin yaranmasına səbəb olur. Nəticədə, dilin variantlarının leksik fondları bir-birindən seçilir. Belə olan halda hər bir ölkədə işlənən standart ingilis dili variantı və ümumi standart dil formalaşır. Bu dillər arasında oxşar və fərqli cəhətlərin öyrənilməsi üçün hər bir standart dilin korpusunu yaratmaq və müqayisəli tədqiqatlar aparmaq zərurəti meydana çıxır. İlk korpus olan Braun korpusunun əsas məqsədi Amerika ingiliscəsinin standart variantı əsasında tezlik lüğəti hazırlamaq, bu standart dilin leksikoqrafik korpusunu müəyyənləşdirmək olmuşdur. Sonralar, LOB, FLOB və digər korpuslar müəyyən dil variantı üzrə formalaşdırılmışdır.

⁷ Ярцева, В.Н. Развитие национального литературного английского языка / В.Н. Ярцева. – Москва: – 1969. – с. 57.

Korpusların bu şəkildə formalaşdırılması, şübhəsiz ki, standart dil, dialekt, dil variantı, slenq ilə bağlı məsələləri aydınlaşdırmağı tələb edir.

İ.V.Arnoldun fikrincə, *“Standart ingilis dili – Böyük Britaniyanın rəsmi dili, məktəblərdə, universitetlərdə danışılan, radio və televiziya işlənen, təhsilli insanların danışdıqları ədəbi ingilis dilidir. Onu təşkil edən lüğət fondu dialekt sözlərdən fərqlənir. Yerli dialektlər, ədəbi formaları olmayan, fərqli məkanlarda işlənen və anlaşılan ingilis dilləri variantlardır”*⁸.

İngilis dili müasir dünyada beynəlxalq dildir. Son 10 ildə ingilis dili beynəlxalq ünsiyyət vasitəsinə çevrilmiş və “qlobal dil” adını almışdır. İngilis dili Böyük Britaniyanın rəsmi dilidir, bu dili əksər britaniyalılar başa düşür, bu dildə danışır. Lakin Uelsin şimal və mərkəzində əhalinin bir hissəsi uels dilində, Şotlandiyanın qərbində əhali şotland dilində, şimali İrlandiyada isə irland dilində danışır. Böyük Britaniyanın 4 tərəfində danışılan ingilis dilinin öz xüsusiyyətləri və fərqli cəhətləri var. Böyük Britaniyada regional və sosial variantlara olan münasibət şimalla və cənub arasındakı aksent və dialektlərdən asılıdır. Standart ingilis dili cənub dialekti əsasında qurulub. Əsasən, London və onun ətraf bölgələrinin dialektlərini nəzərdə tutur. *“Şimal bölgələrə aid dialektlər cənublulara görə daha çətin və geridə qalmış sayılır, aşağı səviyyəli insanların dili kimi qəbul edilirdi”*⁹.

Britaniya ingiliscəsi ənənəvi olaraq leksik və qrammatik xüsusiyyətlərinə görə standart ingilis dili adlanır. London ərazisində yaşayan əhalinin əsasən yuxarı təbəqəsi standart ingilis dilində danışır. Fonetik nöqtəyi-nəzərdən Sakson elementi ilə qarışığı olan Londonun aşağı təbəqəsinin dili isə Kokni dialekti hesab edilir. Bu dialekt İngiltərənin ən təmiz sosial məhəlli dialekti sayılır.

Dilin inkişafı nəticəsində bəzi dialektlərdə slenqlər qorunub qalmışdır. Dilin ilk yaranma dövrlərində rast gəlinən bir sıra leksemlər bugünkü günə qədər bəzi dialektlərdə işlənməkdədir.

⁸ Arnold, I.V. The English Word / I.V. Arnold. – Moscow: – 1966. – p. 262.

⁹ Матюшенко, В.С. Словарь английского сленга / В.С. Матюшенко. – Москва: – 2002. – с. 214.

Slenqlər aid olduqları dialektlərin müəyyən qrammatik və fonetik xüsusiyyətlərini özlərində daşıyır. Hər hansı bir dildə tək slenqlərlə danışmaq qeyri-mümkündür. Slenq ədəbi dildə digər sözlər və ifadələrlə birlikdə işlənə bilər. Araşdırmaların nəticəsinə görə, üslubi cəhətdən neytral olan bir çox sözlər (əsasən ingilis dilində) metafor və digər üslubi fiqurlar sayəsində yaranmışdır. Bundan savayı, nitqdə neytral işlənen hər hansısa bir söz dilin erkən inkişaf mərhələsində bir neçə dəfə məna dəyişikliyinə uğramışdır. Metafor kimi qəbul olunan söz sonradan omonim kimi anlaşılmışdır¹⁰.

İngilis dilində mövcud olan slenqlərin əsas leksik tərkibi uzaq keçmişləri ilə seçilir. Müasir slenqlərin əksəri ədəbi dilin yaranmasından xeyli əvvəl mövcud olmuşdur. *“Dilin digər yarusları kimi, slenq də daim inkişaf edir və yeni mənalı sözlərlə zənginləşir”*¹¹. Onu da qeyd etmək lazımdır ki, ingilis dilinə Amerikanın regional dialektlərindən keçən amerikan elementləri qeyri-sabitliyi ilə seçilir. Bir qayda olaraq onlar müəyyən vaxtdan sonra istifadədən çıxır. S.Beykerin söylədiklərinə əsasən, ingilis dilinin Avstraliya variantında slenqlərin əksəriyyəti variantlardan alınmadır. Onların yalnız 40 %-i Avstraliya slenq lüğətinə aiddir. Qalan 35%-i Kokni sözləri, 25%-i amerikanizmlərdir¹².

İngilis dilinin coğrafi-məhəlli variantları uzun müddət tarixi inkişafdən keçmişdir. Birləşmiş Krallığın özündə İngiltərə, Uels, Şotlandiya və İrlandiya ingilis dili variantları işlənməkdədir. Bundan başqa ingilis dili dünyanın başqa yerlərində yayıldığı üçün bu dilin bir çox variantları meydana gəlmişdir.

Dissertasiyanın bu fəslinin ikinci və üçüncü bölmələrində Böyük Britaniya və Amerika ingiliscəsinin korpusları təhlil edilir.

İngilis dili Böyük Britaniyada rəsmi dildir. Britaniyalılar bu dildə danışır və bu dili anlaşırlar. Böyük Britaniyanın hər 4 ölkəsində

¹⁰ Маковский, М.М. Системность и а системность в языке / М.М. Маковский. – Москва: – 1980. – с. 124-162.

¹¹ Маковский, М.М. Английские социальные диалекты / М.М. Маковский, – Москва: – 1982. – с. 23.

¹² Baker, S. The Australian Language / S. Baker. – Sydney, London: – 1936. – p. 288.

(İngiltərə, Şotlandiya, Uels, İrlandiya) danışılan ingilis dilinin özünəməxsus xüsusiyyətləri vardır. Hər bir ölkənin ingiliscəsi bir-birindən fərqlənir.

Uzun müddət qəbul edilirdi ki, RP (Received Pronunciation-qəbul olunmuş tələffüz) sosial variantdır və ingilis əhalisinin yuxarı təbəqəsinin ləhcəsidir. XIX əsrdə qəbul olunmuş (received) sözü ədəbi dil anlayışı kimi qəbul edilirdi. Əsasən aristokrat təbəqənin danışdığı dil nəzərdə tutulurdu. Sonralar bu anlam “kralların ingilis dili” kimi təbliğ olunmağa başladı.

ABŞ-da danışılan ingilis dili Amerika ingiliscəsi adını daşıyır. İ.V.Arnold görə: “*Amerika ingiliscəsi regional variant olduğundan onu dialekt adlandırmaq olmaz. Bu variant standart ingiliscədən qaynaqlanır, normalaşmış ədəbi dil forması (American National Standart) mövcuddur*”¹³.

Amerika ingiliscəsinin Amerikanın özü kimi çox maraqlı inkişaf tarixi vardır. 3 əsr yarımliq dövr Amerika ingiliscəsinin lüğət fondunda əksini tapır. “*Blue laws, sunbonnet, law cabin, forty-niner, cash house, motel, baby-sitter*” kimi sözlər və terminlər Amerikanın keçmişi və indisi haqqında məlumat verir. “*Amerika ingiliscəsi ölkənin sərhədləri daxilində qalmır, bir çox sözlər (OK, telephone) dünya dillərinə keçir*”¹⁴. Amerikada ingilis dilinin əsaslı kök salmasından əvvəl ispan və portuqal dilləri vasitəsilə çox sayda hind mənşəli sözlər (Indian words) –Azteklərin dilləri- İngiltərənin dilinə təsir etmişdir. Bir çox hind tayfalarının adları hazırda ABŞ-da onomastik vahidlərdir. Məsələn, “İowa, Kansas, Michigan” və s. (hind qəbilələrinin adları) Bu sözlərin bəziləri Avropa dillərində də kök salmışdır. Yeni dünyadan Avropaya gələn sözlər zəngin ekzotik xüsusiyyətə malik idilər. Bu da Amerikada işlənən dilin qonşu ölkələrin dilləri qarışması ilə izah olunur. Məsələn, “*potato, tomato, chocolate, cocoa, cannibal, maize, savannah*” və s. Bu sözlər hazırda adi sözlər hesab olunsalar belə, əvvəllər ingilislər, avropalılar üçün yeni sözlər hesab olunurdu. Məsələn, “*barbeque*” sözü Britaniya

¹³ Arnold, I.V. The English Word / I.V. Arnold. – Moscow: – 1966. – p. 265-266.

¹⁴ Thomas, P. Words and ways of American English / P. Thomas. – New-York: Random House, – 1952. – p. 4.

ingilis dilində yarananda artıq Amerikada az-çox istifadə olunur və tanınırdı.

Amerikanın bir çox yerlərində, həmçinin bir çox məşhur musiqilərində eşidilən variantlardan biri “Black English”dir (qara dərillərin ingiliscəsi). Bir vaxtlar bu variant dilçilər üçün o qədər də maraqlı deyildi. Lakin müasir dövrdə dilçilərin bu dialektə çox böyük maraqları var. Bu dialekt standart ingilis dilindən və ağ dərillərin ingiliscəsindən fərqlənir. R. Mc.Devidə görə: *“Ağ dərili insanları qara dərili insanlardan fərqləndirən heç bir faktor yoxdur. Lakin onların tələffüz qaydalarından hansının ağ, hansının qara dərili olduqlarını müəyyən etmək olar”*¹⁵. Şimali Amerika yazıçılarında biri 1888-ci ildə bir yazısında cənub ştatlarında yaşayan qara dərillələrin işlətdikləri bəzi sözləri vurğulamış və onların izahını vermişdir. Məsələn, *“buccra”* (ağ adam), *“brottus”* (kiçik hədiyyə) və s.

Amerika dilçisi Stüard “Black English” və “Standart English” arasındakı fərqi araşdırmaq üçün bir çox cəhdlər etmişdir. *“Goeber”* (fındıq), *“joke”* (qutu) və s. kimi sözlər ingilis dilinin lüğət fonduna daxil edilmişdir. *“Goeber”* sözü Afrika mənşəli sözdür. Haussa dilində (Qərbi Afrika dili) *“guya”* yer fındığı deməkdir¹⁶. *“U-huh (bəli) və bu kimi sözlər Birləşmiş Ştatlarda həm ağ, həm də qara dərili insanlar tərəfindən başa düşülür. Buna baxmayaraq Böyük Britaniyadan olan insanlar bu sözü nə anlayır, nə də ki, qəbul edirlər”*¹⁷.

Ümumiyyətlə, Black English çox mübahisələr doğurur. Bəziləri onu dil, bəziləri slenq, bəziləri isə dialekt adlandırır. ABŞ-ın kütləvi informasiya vasitələrində dəfələrlə qeyd edilir ki, Black English lazımsız dialektidir; bu dialekt düzəlişlərdən keçməlidir. Əslində isə, Black English ədəbi ingilis dilindən fərqlənən Amerika ingilis dilinin bir variantıdır.

¹⁵ Mc David, R. Dialects: British and American standard and nonstandard- in linguistics / R. McDavid. – New-York. – 1969.

¹⁶ Chamberlain, A.F. Negro Dialect / A.F. Chamberlain. – 1988. – p.23.

¹⁷ Steward, W. Difference between Black English and standart English / Steward W. – 1996. – p. 102.

Dissertasiyanın “Milli dil korpuslarının tətbiqi” adlanan *üçüncü* fəslinin I bölməsində kompüter dilçiliyinin tədqiqat istiqamətləri arasında əlaqələr araşdırılır. Müasir dünyada müxtəlif dil araşdırmaları apararkən kompüter dilçiliyindən geniş istifadə olunur. Kompüter dilçiliyində qarşıya çıxan ən böyük problemlərdən biri mətnlərin struktur, qrammatik və semantik cəhətdən emalıdır. Mətn formalarının, daha doğrusu, nitqin yazılı və şifahi olmasından asılı olaraq kompüter tərəfindən tanınmasıdır. Kompüter dilçiliyinin digər problemləri nitqin sintezi və analizi ilə bağlıdır. Müasir kompüter dilçiliyi sürətlə inkişaf edir və böyük elmi, eləcə də tətbiqi nailiyyətləri ilə diqqəti cəlb edir. Bütün dillərdən bir-birinə avtomatik tərcümə sistemləri yaradılmışdır. Bir sıra dillərdən bir-birinə tərcümə təkmilliyi ilə seçilir. Rusiya, ABŞ, Yaponiya kimi ölkələrdə məşin tərcüməsi üçün yaradılmış sistemlər artıq bu gün qənaətbəxş sayılır. Kompüter dilçiliyi birdilli və ikidilli məşin lüğətləri hazırlanmasında da böyük uğurlar əldə etmişdir.

Kompüter dilçiliyi linqvistikanın bir sahəsi kimi iki yerə bölünür: 1) nəzəri və 2) tətbiqi kompüter dilçiliyi. Nəzəri sahədə dil və kompüter arasında qarşılıqlı əlaqə xüsusi yer tutur. Təbii dilin kompüterə köçürülməsi kimi başa düşülür. Təbii dil kompüterlə yenidən işlənir. Ümumiyyətlə, nəzəri kompüter dilçiliyi təbii dilin xüsusiyyətlərini öyrənir. Bu baxımdan nəzəri kompüter dilçiliyinin böyük əhəmiyyəti vardır. Nəzəri kompüter dilçiliyi sahəsində tədqiqatlar dilin formallaşdırılmasının praktik məqsədləri üçün faydalıdır.

Tətbiqi kompüter dilçiliyinin əsas problemlərindən biri təbii dilin kompüter vasitəsilə işlənməsi və süni nitqin yaradılmasıdır. Tətbiqi kompüter dilçiliyi informatika, riyaziyyat, məntiq, koqnitiv psixologiya ilə sıx bağlıdır. İlk danışmaq məşinləri insan səsinin təqlidinə xidmət edirdi. Elektron nitqin yaradılması sahəsində müasir cəhdlər nitq sintezi adlanır və bunun nəticəsində sintetik nitq alınır. Sintetik nitq sistemlərinin çevikliyi ilə bağlı problemlər mövcuddur. Məsələn, nitq istehsalı sözlərin tələffüzündən daha mürəkkəb prosesdir. Burada intonasiya, fasilə və s. amillər nəzərə alınır. Hazırda mətnlərin avtomatlaşdırılmasının bir çox məsələləri əksər

hallarda kompüterdə həll oluna bilir. Statistik göstəricilərin maşında əldə olunması üçün bir çox standart proqramlar qurulmuşdur.

Son zamanlar elmi-texniki mətnlərin, müxtəlif rəsmi sənədlərin işlənməsi, maşın tərcüməsi, referatların hazırlanmasının avtomatlaşdırılması istiqamətində böyük irəliləyişlər əldə edilmişdir. Müxtəlif dillər üçün tərcümə sistemlərinin, avtomatik lüğətlərin hazırlanması da dilçilikdə yeni bir sahənin yaradılması zərurətini doğurmuşdu. Beləliklə, XX əsrin ortalarında bir sıra dəqiq elmlərlə dilçiliyin qovuşma nöqtəsində yeni bir elm sahəsi yarandı. İlk vaxtlar ona “riyazi dilçilik”, “hesablama dilçiliyi” və s. adlar verildi. Nəhayət, “kompüter dilçiliyi” olaraq yeni bir elm sahəsi meydana gəlmişdir. *“Dünya elmində və rus dilçiliyində də “kompüter dilçiliyi” dilçiliyin yeni sahəsi kimi tanınır və bu termin artıq dünya elmində vətəndaşlıq hüququ qazanmışdır”*¹⁸.

Ümumiyyətlə, kompüter linqvistikası ilə korpus linqvistikası arasındakı münasibət məsələsi tez-tez ortaya atılır. Tədqiqatçılar onları bəzən ayırır, bəzən korpus linqvistikasını ümumiyyətlə kompüter dilçiliyinin tərkibinə daxil edirlər. Əslində “kompüter dilçiliyi” termini özü mübahisə obyektini kimi diqqəti özünə çəkir. Əvvəllər “riyazi dilçilik”, “struktur-tətbiqi dilçilik”, “mühəndis dilçiliyi”, “hesablama linqvistikası” tipli adlarla adlandırılan bu sahənin indi “kompüter dilçiliyi” ilə ifadə olunmasının əsas səbəbi bu sahədə kompüterin tətbiqi ilə bağlıdır. İlk dövrlərdə hesablama linqvistikası anlayışının istifadə edilməsinin səbəbi dil vahidlərinin hesablanması ilə birbaşa əlaqədar idi. Hesablama aparılması zərurətinin meydana çıxması, şübhəsiz ki, bu və ya digər mətnə hansı sözün neçə dəfə rast gəlməsini tapmaqla bağlı idi. Daha əvvəllər kompüterlər olmadıqda tədqiqatçı bu cür məsələnin həlli üçün mətni dəfələrlə oxuyur və orada axtarılan sözü qeyd edir, sonradan qeydə alınmış sözləri saymaqla kəmiyyət göstəricisini müəyyənləşdirirdi.

Elektron maşın sistemlərində qənaətlə işləmək, danışılanın yazıya çevrilməsi və gözlə görünməsinin təmin edilməsi və s.

¹⁸ Марчук, Ю. Н. Основы компьютерной лингвистики / Ю. Н. Марчук. Изд.-во МГУ, – 2000. – с. 38.

məsələlər dilçilərin başqa elm sahələr ilə əməkdaşlığını zəruri edir. Bu əməkdaşlıq imkan verir ki, kompüter klaviaturasında hərflərin düzülüşü onların işlənmə tezliyinə uyğun olaraq qurulsun. *“Hazırkı kompüter klaviaturası bizim dilimizin tədqiqinə əsaslanmadığından işarələrin düzümündə ciddi problemlərlə qarşılaşırıq”*¹⁹.

F.Veysəllinin qeyd etdiyi bir çox məsələlər Azərbaycan dili üçün tədqiqat səviyyəsində həllini tapmışdır. Lakin bu tədqiqatların nəticələrindən yetərincə istifadə edilməmişdir. Azərbaycan dilində olan mətnlərin kompüterdə yığılması Azərbaycan əlifbasının hərflərinin - qrafemlərin dilin kəmiyyət xarakteristikaları əsasında klaviaturada yerləşdirilməsini tələb edir.

Kompüter dilçiliyinin nailiyyəti kimi, avtomatik tərcümə real gerçəklilyə çevrilmiş və müəyyən səviyyədə insanların istifadəsinə verilmişdir. Hazırda bir çox dillərdən ingilis dilinə və əksinə tərcümə proqramları artıq internet səhifələrində yer almışdır. İnternetdə müxtəlif saytlarda avtomatik tərcümə proqramları yerləşdirilməsi davam etdirilir. Hind-Avropa dillərindən bir-birinə avtomatik tərcümənin ən mürəkkəb məsələləri artıq işlənmişdir.

Bu fəslin ikinci bölməsində milli dil korpuslarının elektron lüğətçiliyinə təsiri nəzərdən keçirilmişdir. Elektron lüğətlər maşın tərcümə sistemində olan digər lüğətlərdən çox əvvəl hazırlanmışdır. Hazırda proqram təminatı bazarında çox geniş lüğət seçimləri mövcuddur. İlk yaranan elektron lüğətlərə yalnız bir lüğət qoşmaq olurdu. İndiki zamanda isə, elektron lüğətlərə çox sayda lüğətlər daxil etmək olur. Əvvəllər elektron lüğətlərin istifadəçi tərəfindən genişləndirilməsi mümkünsüz idi, lakin, müasir LINGVO 4.6 və daha artıq versiyalarda bu mümkündür. *“Tərcümə etmə rejiminə görə elektron lüğətlərdə 2 tərcümə rejimini qeyd etmək olar: avtomatlaşmış və interaktiv rejim. Birinci rejimdə tərcümə sözbəsöz şəkildə aparılır. Bu cür lüğətlərin söz bazası çox zəif olur. Bu cür lüğətlər, təəssüf ki, tərcüməçinin köməyinə çatmır və onun işinə təsiri olmur”*²⁰. İkinci rejimdə kompüterin monitorunda görünən yad sözün üzərindən basaraq açılacaq pəncərədən tərcümə funksiyasını tapır və

¹⁹ Veysəlli, F. Dilçiliyin əsasları / F.Veysəlli. – Bakı: Mütərcim, – 2013. – s. 56.

²⁰ www.xreferat.com/31/4352-1mashinniy-perevod.html

istədiyimiz sözü tərcümə edə bilərik. Bu tərcüməçilər üçün çox effektiv bir rejimdir.

Kompüter lüğətlərinin bir sıra üstünlükləri mövcuddur; lüğətə müraciət daha sürətli və daha sadədir, eyni zamanda bir neçə lüğətə müraciət edərək sözün mənalərini müqayisə etmək olar, vaxtaşırı yenilənərək yeni mənalər əlavə olunur və dildəki yeniliklər və yeni sözlər tez bir zamanda bu lüğətlərdə öz yerlərini alır.

İngilis dilində ən çox yayılmış və geniş istifadə olunan kompüter lüğətinə “Dictionary.com”, “Dictionaries”, “Online Dictionaries” göstərmək olar. Bu lüğətlərin hər biri 200-dən çox lüğətdən ibarətdir. İngilis dilində olan “Your Dictionary” 240 dildə olan lüğətlərdən və 30-dan artıq kitablardan ibarətdir. Bu cür lüğətlərdə axtarılan sözün bu və ya digər nitq hissələrinə mənsubluğu, eləcə də həmin sözün sinonimi, antonimi və omonimi haqqında məlumat əldə etmək olar. Bu lüğətlər Avropa, eyni zamanda şərq dillərini əhatə edir.

Korpusların yaradılması ideyası ilk əvvəl, tezlik lüğətlərinin tərtibi ilə əlaqədar olmuşdur. Mahiyyətə belə lüğətlərin tərtibi məsələnin qoyuluşundan asılı olmuşdur. “Məsələnin riyazi qoyuluşu” verilənlərin (bu halda bizim qoyduğumuz məsələnin xarakterindən asılı olaraq materiallar nəzərdə tutulur) ədədi şəkə gətirilməsi – yəni kodlaşdırılması, onun tərkibinə daxil olan hissələrin həlli nəzərdə tutulur. Əlbəttə, məsələnin həlli üçün ona uyğun metodlar araşdırılarkən qarşıya çıxan müxtəlif variantlardan optimalını seçmək lazımdır. Məsələ, tezlik lüğətinin tərtibi zamanı əlverişli metod artıq standart hala keçmiş növləşdirmə metodudur. Seçilmiş üsula əsasən isə, məsələnin alqoritmi qurulmalıdır. Dilin riyazi-statistik metodların köməyi ilə tədqiqində Azərbaycan dilçiliyində önəmli yer tutan tezlik lüğətləri tərtib olunmuşdur. Bu lüğətləri tərtib edərkən bir sıra çətinliklər qarşıya çıxır. Birincisi, mətnləri elə seçmək lazımdır ki, kafi qədər ədəbi dil normalarını özündə əks etdirmiş olsun. Bu, bir o qədər də asan iş deyil. İkincisi, daha mürəkkəb məsələ mətnlərin ümumi həcmi seçməkdən ibarətdir. Azərbaycan dilinin tezlik lüğəti üçün əsas mənbə qəzet materialları olmuşdur. İxtiyari götürülmüş mətnlərdən 100 min söz seçilmişdir.

Fəslin üçüncü bölməsi maşın tərcüməsi və linqvistik korpus məsələsinə həsr olunmuşdur. Maşın tərcüməsi – yazılı və şifahi mətnlərin bir dildən başqa dilə kompüter vasitəsilə tərcümə olunması prosesidir. Son zamanlar xarici dillərin öyrənilməsi nəinki səfərlər və qonaq qəbulu üçün lazımdır, o, həm də məşhur Hollivud filmlərini seyr etmək, xarici məhsulların üzərindəki yazıları və ya Web-səhifələri oxumaq üçün lazım olur. Belə qənaətə gəlirik ki, hətta, öz ölkəmizin hüdudlarında belə biz xarici dildən istifadə etməli oluruq.

Maşın tərcüməsinə iki yolla yanaşmaq olar: deduktiv və induktiv. Birinci metod “mətn-məna-mətn”, ikinci metod isə “mətn-mətn” modellərinin istifadəsinə əsaslanır. Birinci yanaşmada dilin semantikasından maksimal istifadə edərək yüksək səviyyəli maşın tərcüməsi almaq mümkündür. Ümumi sxemi morfoloji, sintaktik və semantik səviyyələrin mətnin mənasına keçidi kimi təsəvvür etmək olar. Sistemin əsasını, beləliklə, məndən onun mənasına doğru keçidi təşkil edir. Bu keçid daxil olan və xaric olan dillərin lüğət fondları hesabına baş verir. Əlbəttə, burada qrammatik qaydaların da böyük rolu vardır. İkinci yanaşmada mətnə çoxşaxəli bir sistem kimi baxılır.

Müxtəlif dialektlərdə işlənən sözlər də maşın tərcüməsinin mükəmməlliyinə mane olur. Məsələn, Amerikanın hər hansısa bölgəsində işlənən bir söz Britaniyada başqa məna verə bilər. Bu dialektlərdən xəbəri olan tərcüməçi tərcümə zamanı kontekstə uyğun gələn mənanı tərcümədə verdiyi halda, kompüter onu yalnız sistemə yerləşdirilən sözlə ifadə edir. Məsələn, “*kitchen*” sözü ingilis dilində mətbəx deməkdir. Amerikan ingiliscəsində isə bu sözün boyunun arxasındakı tük mənası da vardır: “*Sheila should comb her kitchen because it is looking kinda rough*”²¹. Bu cümlənin tərcüməsində maşın tərcüməsi çox güman ki, düzgün tərcümə verməyəcək. İnsan tərcüməçi isə kontekstdən yanaşaraq bu cümlənin tərcüməsini düzgün verəcək. “Şeylə görək boynun arxasındakı tükləri qırxsın, çünki, çox eybəcər görsənir”.

Görkəmli mütəxəssis V.İnqve yazır: “*Maşın tərcüməsi*

²¹ Moody, S. The Diversity of English in America // Popular linguistics, - 2011. – vol 1, - issue 2, - p. 48-50.

sahəsində işlər semantik sədd qarşısındadır. Biz belə bir müddəanı anlamağa başlamışıq ki, maşın tərcüməsində yeganə metod ondan ibarətdir ki, maşın nəyi tərcümə etdiyini qavramalıdır”²². Fikrimizcə, bu, çox doğru bir açıqlamadır. Kompüterin beyni insan beyni ilə eyni ola bilməz. Kompüterin tərcümə etdiyini qavraması çox çətin bir prosesdir. İnsan hər hansı bir mətni tərcümə etdikdə öz bədii, ədəbi biliklərini də işə salır. Dünya görüşünü ora əlavə edir, kontekstə daha uyğun olan sözü axtarıb tapır, tərcümə olunan mətnə daha çox bədii əlavə edir. Maşın tərcüməsinin bir qolu kimi yarımavtomatik tərcümə də çox məşhurdur. Yarımavtomatik tərcümədə kompüterlərdən tam istifadə olunmur. Onlar yalnız köməkçi vasitə kimi işlədilir. Bu zaman tərcüməçi maşınların köməyi ilə tərtib olunmuş xüsusi lüğətlərdən istifadə edir. Tərcüməçi sözü lüğətdə axtarmaq kimi uzun vaxt itkisindən azad olur və az vaxt ərzində bütün naməlum sözlərin mənasını tapır.

Üçüncü fəslin sonuncu bölməsində korpusda annotasiyalama və markerləmə məsələləri tədqiq edilmişdir. Korpus linqvistikası tətbiqi dilçiliyin linqvistik korpusların qurulması və istifadə edilməsinin ümumi prinsiplərini öyrənən bölməsidir. Burada dil materialının əsas mənbəyi mətnlərdir. Həmin mətnlər üzrə linqvistik tədqiqatlar aparılır. Korpus linqvistikası ənənəvi dilçilikdən fərqlənir. Korpus linqvistikası nitqi öyrənir. Ənənəvi qrammatikada isə əsas diqqət dilin tədqiqi üzərində cəmləşdirilmişdir. Korpus linqvistikasının nitqi tədqiq etməsi onun materialından irəli gəlir. Bu material yazılı, çap olunmuş və transkripsiya edilərək elektron daşıyıcılarına daxil edilmiş şifahi nitq nümunələridir. Korpusu təşkil edən mətnlər real mövcud mətnlərdir və onlar müxtəlif tipli kommunikasiya prosesinin nəticəsi olan nitq məhsullarıdır.

Korpus linqvistikasında analiz, seqmentləmə və seqment təhlili imkanları genişdir. Çünki onun obyektı hazır mətndir və bu mətni təşkil edən vahidlər üzə çıxarılır. Korpus linqvistikasında mətndə yer alan sözformalar öyrənilir. Korpusun strukturu və riyazi təminatı sözformanın yaxın ətrafını müəyyənləşdirməyə imkan yaradır.

²² Ингве, В. Значение исследования в области машинного перевода // НТН, – сер. 2, № 7, – 965. – с. 44.

Korpus hansı sözformadan sonra sözformanın gəlmək ehtimalını hesablamaq üçün material verir.

Korpus dilçiliyinin üstünlüklərindən biri odur ki, dilin kəmiyyət xüsusiyyətləri üzrə tədqiqat aparana birbaşa korpus materialından istifadə edirlər. Korpusdan istifadə etmək materialın seçilməsi, toplanması, maşına daxil edilməsi kimi ağır zəhmət tələb edən işlərdən azad olurlar. Korpus onlara lazım gələn materialı verir. Onu da qeyd etmək lazımdır ki, bir sıra məqsədlər üçün mövcud elektron kitabxanalarından istifadə etmək mümkündür. Bir çox istifadəçi proqramlarında (məsələn, Word) axtarış sistemləri vardır. Bu sistemlərdən də istifadə etməklə bəzi məsələlərin həlli mümkündür. Yazıçının, şairin dilinin tezlik lüğətlərini, eləcə də onların işlətdikləri sözlərin ümumi həcmi müəyyənləşdirmək olar.

Korpus dilçiliyində korpusun qurulması prosesində həll olunan vacib məsələlərdən biri markerlənməmiş mətdən markerlənməş mətnə keçidin təmin olunmasıdır. İngilis dilinin korpus terminologiyasında *annotation* termini istifadə olunur. Həmin termin rus dilinə də keçmişdir. Azərbaycan dilində korpus dilçiliyi yeni olduğundan müvafiq terminoloji baza yaradılmamışdır. Bununla belə, annotasiya, annotasiyalama, nişanlama, markerləmə kimi termin variantlarından istifadə etmək mümkündür.

Annotasiyalama və ya markerləmə anlayışı korpusa daxil edilən linqvistik informasiyanı nəzərdə tutur. Korpus yalnız mətnlərin seçilməsi, kontekstlərin təsadüfi ədədlər qaydası ilə müəyyənləşdirilib korpusa daxil edilməsi ilə tamamlanmır. Bu halda korpus mənasını itirir və elektron kitabxana ilə müqayisədə olduqca əhəmiyyətsiz bir topluya çevrilir. Korpusun mühüm təyinedici amillərindən biri ona daxil edilən mətnlərin annotasiyalanmasıdır. Annotasiyalama ilkin dövrdə yalnız linqvistik informasiyanı əhatə etmişdir. Korpus dilçiliyi inkişaf etdikcə mətnlərin əlavə informasiyalarla təchiz edilməsinin də zəruriliyi aşkara çıxmışdır. Korpusa onun yaradılması, mətnin janrı, müəllifi, yazılma dövrü, mətnin seçildiyi əsərin adı, nəşr haqqında dəqiq məlumat, kontekstin çap versiyasında yerləşdiyi səhifənin nömrəsi haqqında informasiyalar daxil edilir. Bunlar isə linqvistik məlumatlar deyildir.

Linqvistik informasiya sintaktik strukturla bağlı olduqda cümləyə, leksem və qrammatik xüsusiyyətlər barədə olduqda sözə aiddir.

Qeyd olunan məsələlərin hər birinin həlli müəyyən çətinliklər yaradır. İstifadəçi müəllif haqqında informasiya blokuna birbaşa daxil olduqda o, korpusdan bu müəllifin hansı əsərlərindən istifadə olunması haqqında məlumat alır və buradan da birbaşa mətn korpusuna daxil olaraq korpusları əldə edir. Digər halda istifadəçi korpusdan konteksti seçir və sonra onun kimə aid olması, hansı tarixdə qələmə alınması, hansı mənbədən seçilməsi haqqında məlumatları əldə edir. Korpusun müəllifi kontekstin çap variantı barədə məlumatlara fərqli istiqamətlərdən daxil olmaq kimi vəzifələr qarşısında qalır. Ona görə də markerləmə, annotasiya materiallarına istinadların korpusda yerləşdirilməsi, onlara fərqli nöqtələrdən müraciətlərin mümkünlüyü korpus dilçiliyində nəzərə alınır.

Dissertasiyanın “Nəticə” hissəsində tədqiqat zamanı əldə edilmiş elmi qənaətlər ümumiləşdirilmişdir:

1. Korpus linqvistikası böyük həcmdə dil materialını toplamaq, saxlamaq və müxtəlif linqvistik problemləri həll etmək üçün bu materialdan istifadə etmək məqsədi ilə bağlı təşəkkül tapmışdır.

2. Dilin milli korpusunu yaradarkən ora toplanan mətnlərin seçilmə meyarlarını nəzərə almaq vacibdir. Korpus mətnindən istifadə zamanı mətndə işlənən sözləri, söz birləşmələrini, qrammatik kateqoriyaları təhlil etmək imkanı əldə olunmalıdır.

3. Korpusların təsnifi meyarları dairəsi kifayət qədər genişdir. Bu meyarlar korpusların təyinatı ilə bağlılığa malikdir. Kompüter dilçiliyi inkişaf etdikcə korpuslardan istifadənin yeni formaları axtarılıb tapılır. Korpus linqvistikasının gələcək perspektivləri və tipləri həm də korpus materiallarının annotasiyalaşdırılmasına əsaslanır. Morfoloji, sintaktik, semantik, terminoloji markerləmə ən azı bu dörd korpusu qeyd etməyin zəruriliyini təsdiqləyir.

4. Korpus linqvistikasında standart dil, sosial dialektlər, slenqlər, dilin variantları korpusları yaradır. Müasir dövrdə ingilis dili qlobal dil hesab edilir və hər bir ölkənin ingiliscəsi bir-birindən fərqlənir. Böyük Britaniya dialektləri Amerika dialektlərindən sayca daha çoxdur. Bu baxımdan dilin korpusunu yaradarkən ora daxil

edilən mətn fraqmentlərinin dilin bütün dialekt formalarını əhatə etməsi vacibliyi əsas məsələlərdəndir. Dil variantları, dialekt və ləhcələr, slenqlər əsasında müxtəlif korpuslar yaratmağın məqsədəuyğunluğu Braun, LOB, FLOB kimi mövcud korpuslar təcrübəsində təsdiqini tapır.

5. Korpus linqvistikası ənənəvi dilçilikdən fərqlənir. Korpus linqvistikası nitqi öyrənir. Ənənəvi dilçilikdə isə əsas diqqət dilin tədqiqi üzərində cəmləşdirilmişdir. Korpus linqvistikasının nitqi tədqiq etməsi onun materialından irəli gəlir. Bu material yazılı, çap olunmuş və transkripsiya edilərək elektron daşıyıcılarına daxil edilmiş şifahi nitq nümunələridir. Korpusu təşkil edən mətnlər real mövcud mətnlərdir və onlar müxtəlif tipli kommunikasiya prosesinin nəticəsi olan nitq məhsullarıdır.

6. Annotasiyalama və ya markerləmə anlayışı korpusa daxil edilən linqvistik informasiyanı nəzərdə tutur. Korpus yalnız mətnlərin seçilməsi, kontekstlərin təsadüfi ədədlər qaydası ilə müəyyənləşdirilib korpusa daxil edilməsi ilə tamamlanmır. Korpus dilçiliyi inkişaf etdikcə mətnlərin əlavə informasiyalarla təchiz edilməsinin də zəruriliyi aşkara çıxmışdır.

Dissertasiyanın əsas məzmunu və elmi müddəaları müəllifin aşağıdakı məqalə və tezislərində öz əksini tapmışdır:

1. Korpus dilçiliyində məqsədi və əsas istiqamətləri. // – Bakı: Filologiya məsələləri, – 2015. №1, – s. 258-261

2. Korpus dilçiliyi kompüter dilçiliyinin əsası kimi // – Bakı: Filologiya məsələləri, – 2015. №1, – s. 103-108.

3. American variety of English language // – Toronto (Canada): International Journal of English Linguistics, – 2015. – Vol 5, – p.159-163.

4. Sosial and Regional Variations of English Language // - USA: Global Journals of Human-Social Sciences, – 2015. – Vol 15, – s. 13-15.

5. Different dialects in corpus linguistics. (Based on bnc and coca.) // – Odlar Yurdu Universitetinin Elmi və Pedaqoji Xəbərləri, – Bakı: – № 41 – 2015. – s. 152-158

6. Scientific Interest to American English // – Gənc tədqiqatçıların IV beynəlxalq elmi konfransı, filologiya elmləri, – Bakı: – 29-30 May, – 2016. – s. 1281-1282.

7. İngilis dilinin standart dil variativləri // Koqnitiv və tətbiqi dilçiliyin aktual problemləri adlı beynəlxalq elmi konfrans, – Bakı: – 20 Oktyabr, – 2016. – s. 320-322.

8. Amerikan ingiliscəsinin “Black English” dialekti // Doktorantların və Gənc tədqiqatçıların XX Respublika elmi konfransının materialları, – Bakı: – 24-24 May, – 2016. – s. 182-184.

9. Müasir ingilis dilində slenqlər // Birinci Türkoloji Qurultayın 90 illik yubileyinə həsr olunmuş Respublika Elmi Konfransının materialları, – Bakı: – 28 Dekabr, – 2016. – s. 235-236.

10. Erkən amerikanizmlər və onların ingilis dilində təzahürü // – Bakı: ADU-nun Elmi xəbərləri, – Cild 2, №3, – 2017. – s. 49-52.

11. Qloballaşma və onun dünya dillərinə təsiri // – Bakı: Filologiya məsələləri, №4, – 2018. – s. 188-195.

12. Amerika ingiliscəsinin inkişaf yolu və Noah Vebster // – Bakı: Slavyan Universitetinin “Elmi əsərləri”, – 2019. – s.183-188.

13. Maşın tərcüməsi və linqvistik korpus // “Müasir təhsilin inkişafında beynəlmilləşmənin rolu” mövzusunda Beynəlxalq elmi-praktiki konfransın materialları, – Bakı: – 29 Aprel, – 2019. – s.49-51

14. Using national corpora in translation // “Вопросы современной науки: проблемы, тенденции и перспективы”, международная научная конференция, – Выпуск 9, – Москва: – 13 Сентябрь, – 2019. – с. 44-46.

15. Theoretical basis of machine translation // Филологические социокультурные вопросы науки и образования» IV международная научно-практическая очно-заочная конференция, – Краснодар: 25 Октябрь, – 2019. – с. 486-492.

Dissertasiyanın müdafiəsi 20 aprel 2021-ci il tarixində saat 14⁰⁰ Azərbaycan Milli Elmlər Akademiyası Nəsimi adına Dilçilik İnstitutu nəzdində fəaliyyət göstərən ED 1.06 Dissertasiya şurasının iclasında keçiriləcək.

Ünvan: Bakı, AZ 1143, H.Cavid prospekti 115, V mərtəbə, AMEA Nəsimi adına Dilçilik İnstitutu.

Dissertasiya ilə Azərbaycan Milli Elmlər Akademiyası Nəsimi adına Dilçilik İnstitutunun kitabxanasında tanış olmaq mümkündür.

Dissertasiya və avtoreferatın elektron versiyaları Azərbaycan Milli Elmlər Akademiyası Nəsimi adına Dilçilik İnstitutunun rəsmi internet saytında yerləşdirilmişdir.

Avtoreferat 19 mart 2021-ci il tarixində zəruri ünvanlara göndərilmişdir.

Çapa imzalanıb: 15.03.2021
Kağız formatı: 60x84 16¹
Həcm: 40 948 işarə
Tiraj: 100