

UOT 004,6:80

Vəliyeva K.A.

AMEA Nəsimi adına Dilçilik İnstitutu, Bakı, Azərbaycan
kamilavaliyev@gmail.com

KOMPÜTER DİLÇİLİYİNİN MÜASİR İSTİQAMƏTLƏRİ

Məqalədə kompüter dilçiliyinin əsas istiqamətləri: təbii dilin emalı, korpus dilçiliyi, kompüter vasitəsilə lüğətlərin tərtibi, dillərin kompüter vasitəsilə öyrədilməsi və tanınması, maşın tərcüməsi məsələləri şərh edilmişdir.

Açar sözlər: kompüter dilçiliyi, təbii dilin emalı, maşın tərcüməsi, korpus dilçiliyi.

Giriş

Müasir dövrdə informatika və linqvistikanın qovşağında yaranan kompüter dilçiliyi sürətlə inkişaf edən elm sahələrindəndir. O, tətbiqi dilçiliyin bir qolu olub, informatikanın linqvistik əsaslarını kompüter proqramları vasitəsilə dil və təfəkkürün modelləşdirilməsini öyrənir və bununla da, informasiyanın təqdimində ümumi dilin yaradılması və təkmilləşdirilməsi, süni intellektin əsas məsələlərindən sayılan təbii dillərin interfeysi, informasiya sistemlərinin modelləşdirilməsi kimi məsələlərin həllini qarşısına məqsəd qoyur.

Hələ XX əsrin 60-cı illərində başlanmış kompüter dilçiliyinin əsas problemləri nəzəri linqvistik modellərin hazırlanmasına yönəlmişdi. Lakin sonralar tədqiqat obyektinə daha geniş yəndə statistik metodların tətbiqi ilə maşın təliminə, mətnlərin emalına, bir sıra everestik məsələlərin həllinə istiqamətləndi [1].

Yuxarıda qeyd etdiyimiz kimi, kompüter dilçiliyi süni intellektin probleminə daxil olan insanın intellektual fəaliyyətində təbii dillərin riyazi modelini hazırlayan bir elm sahəsi olduğundan, əlbəttə, burada əsas vurğu müxtəlif təyinatlı kompüter sistemləri üçün dilin tətbiqinə və işlənməsinə düşür ki, bununla əlaqədar onun əsas istiqamətləri aşağıda sadalanan məsələlərə yönəlir:

1. təbii dilin emalı. Burada mətnin sintaksis, morfoloji, semantik təhlili nəzərdə tutulur;
2. korpus dilçiliyi (buraya mətnlərin elektron korpuslarının yaradılması və istifadəsi daxildir);
3. elektron lüğətlərin hazırlanması (tezaurus, avtomatik tərcümə lüğətləri, ensiklopediyalar, orfoqrafik, izahlı, terminoloji və sahə lüğətləri, səhvlərin avtomatik aşkarı üçün spelling lüğətləri və s.);
4. mətnlərin avtomatik tərcüməsi sistemləri (Dilmanc, *Google Translate* və s.);
5. mətnlərdən faktların məlumatların, informasiyanın çıxarılması (*fact extraction*, *text mining*);
6. avtoreferatlaşdırma. Bu funksiya *Microsoft Word*-ə daxil edilmişdir;
7. bilik (ekspert) sistemlərinin yaradılması;
8. sual-cavab (dialog) sistemlərinin yaradılması;
9. simvolların optik tanınması (*OCR*, *FineReader* proqramı);
10. nitqin avtomatik tanınması;
11. nitqin avtomatik sintezi;
12. məlumat axtarış sistemlərinin hazırlanması [2].

Təbii dillərin emalı süni intellekt və riyazi dilçiliyin bir qolu olub, hər hansı bir təbii dilin kompüter vasitəsilə analizi və sintezini öyrənir. Bir sözlə, nəzəri cəhətdən dilin emalı prosesində kompüter üçün interfeys dili (insanla kompüterin qarşılıqlı əlaqəsi) yaradılmışdır ki, bu da, o qədər asan məsələ deyil. Təbii dilin dərki (anlamı) aləm haqqında zəngin bilik tələb edir ki, məhz süni intellektin də əsas məsələlərindən biri “anlamı” kompüterə tanımaqdır. Məsələn, Azərbaycan dilindəki mətnin anlamı (dərki) söz sırasından, omonimlərdən, sinonimlərdən, durğu işarələrindən və vurğudan asılıdır. Bu baxımdan da, kompüter üçün mətnlərin anlaşılması məqsədi ilə sadalanan

meyarların formal modelləri hazırlanmalıdır.

Korpus dilçiliyi (KD) mətnlərin korpus və istifadəsinin yaradılmasını tədqiq edən tətbiqi linqvistikanın bir bölməsidir. Bu termin 1960-cı ildə, ümumiyyətlə, praktikada korpusların yaranması ilə meydana gəlmişdir. “Linqvistik korpus” dedikdə, müəyyən prinsiplərə və standartlara uyğun mətnlərin məcmusu başa düşülür. Korpusları yaratmaqda məqsəd müxtəlif səpkili linqvistik məsələlərin (məsələn, mətnlərin qrafik və leksik qrammatik təhlili) həll edilməsidir.

Dünya dilçiliyində ilk böyük korpus 1960-cı ildə yaradılmış Braunov korpusudur. Bu korpusun modelinə əsasən, Zatorina tərəfindən bir milyon sözdən ibarət rus dilinin tezlik lüğəti tərtib edilmişdir [3]. Kompüter texnologiyalarının sürətli inkişafı böyük həcmli milli korpusların hazırlanmasına təkan verdi. Bu cür korpuslara İngiltərənin Birmingen Universitetində yaradılan *British National Corpus*-u, keçmiş SSRİ-də isə A.P.Erşovun təşəbbüsü ilə hazırlanan rus dilinin Maşın fondunu nümunə göstərmək olar. Hal-hazırda rus dilinin milli korpusu 300 milyon sözləşmədən ibarətdir [2].

Azərbaycan dilinin milli korpusunun yaradılması isə hələ də həllini gözləyən məsələlərdən biri kimi qarşıda durur.

Məsələnin həll yolları: kompüter vasitəsi ilə lüğətlərin tərtibi

Elmi-texniki inqilab dövründə kibernetikanın bəhrəsi olan kompüterlərin sürətlə inkişafı əqli əməyin avtomatlaşdırılması üçün zəmin yaratmış və geniş perspektivlər açmışdır. Kompüter vasitəsilə aparılan tədqiqatlar dilçiləri leksikoqrafiyada mexaniki, yorucu, ağır zəhmət tələb edən, yəni materialların qruplaşdırılması, növləşdirilməsi, inventarlaşdırılması, redaktə edilməsi kimi işləri görməkdən azad etdi.

Hələ ötən əsrin 50-ci illərində Avropada elektron hesablama maşını ilə təchiz olunmuş iki böyük laboratoriya təşkil edilmişdir ki, bunlardan biri Bezansondakı fransız dilinin lüğətini öyrənən mərkəzin nəzdindəki leksikoloji analiz laboratoriyası, digəri isə İtaliyadakı filoloji analiz üzrə dilçilik mərkəzinin laboratoriyasıdır. Bunlardan başqa, Hollandiyada F.de Tollenerin rəhbərliyi altında kompüterin köməyi ilə leksikoqrafik tədqiqatlar aparılırdı [3].

Rusiya elmi mərkəzlərinin əksəriyyətində kompüterin tətbiqi ilə əlaqədar xeyli işlər görülmüşdür. Bu mərkəzlərdə hesablama texnikası vasitəsilə müxtəlif tipli lüğətlər - məndəki sözlərin işlənmə tezliyini göstərən tezlik lüğətləri, tərkibində müəyyən söz olan bütün məndəki ifadələrin siyahısını göstərən konkordanslar, sözün sonundan asılı olaraq əlifba sırası ilə sıralanan əks-lüğətlər, qafiyə lüğətləri, maşın tərcüməsi sistemi üçün hazırlanan avtomatik lüğətlər və s. tərtib edilmişdir. Adları çəkilən lüğətlər haqqında V.Y.Pines və M.Ə.Mahmudovun “Elektron hesablayıcı maşınlar dilçilikdə” adlı icmal məlumatında kifayət qədər dolğun məlumat verilmişdir [4].

Burada yalnız Azərbaycan dilçiliyində kompüter vasitəsilə tərtib olunan lüğətləri nəzərdən keçirəcəyik. Bu məsələyə toxunmadan əvvəl kompüter və onun işləmə prinsipi haqqında qısa məlumat verək.

Məlumdur ki, “insan-maşın” sistemində tədqiq olunan məsələlər aşağıdakı mərhələlər üzrə həll edilir [5].

1. məsələnin ümumi qoyuluşu;
2. məsələnin riyazi qoyuluşu;
3. kompüterin girişinə daxil olunacaq materialların (ilk verilənlərin) hazırlanması, yəni kodlaşdırılması);
4. məsələnin həlli üçün uyğun metodların seçilməsi;
5. məsələnin alqoritminin qurulması və onun blok-sxeminin hazırlanması;
6. qurulmuş alqoritmin proqramlaşdırılması;
7. məsələnin kompüterdə həlli;
8. kompüterdə alınan nəticələrin araşdırılması;

Lüğətlərin tərtibi məsələsi ilə əlaqədar bu mərhələləri qısa da olsa xarakterizə edək:

“Məsələnin riyazi qoyuluşu” deyərkən, verilənlərin (bu halda bizim qoyduğumuz məsələnin xarakterindən asılı olaraq materiallar nəzərdə tutulur) ədədi şəkllə gətirilməsi – yeni kodlaşdırılması, onun tərkibinə daxil olan hissələrin həlli nəzərdə tutulur.

Tərkib hissələrinin ardıcıl həlli qoyulmuş məsələni tamamlayır. 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. Məlumdur ki, tezlik lüğətlərində sözlər və onların mətndəki işlənmə tezliyinin göstəricisi qeyd olunur. 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.

Tezlik lüğətlərinin tətbiqinin məlumatların avtomatik işlənməsində - avtomatik tərcümədə, referatlaşdırmada, polioqrafiyada, kriminalistikada, anonim yazıların müəlliflərini müəyyənləşdirməkdə müstəsna əhəmiyyəti vardır.

Əvvəldə qeyd etdiyimiz kimi, ilk tezlik lüğəti 1997-ci ildə işıq üzü gördü [6]. 2004-cü ildə onun yenidən işlənmiş variantı nəşr olundu [7]. 2010-cu ildə isə Azərbaycan dilinin tezlik lüğəti çap edildi [8]. Bu lüğətin tərtibi üçün Azərbaycan dilinin bütün üsulları təmsil olunmaqla, təxminən 50 milyona yaxın sözdən ibarət mətnlər linqvostatistik təhlilə cəlb olunmuşdur. Lüğətdə 12000-ə yaxın söz kökü əhatə olunmuşdur. Azərbaycanın tezlik lüğəti leksikanın statistik modeli kimi ümumi leksikologiyanın, dil tədrisinin bir çox problemlərinin həllində əvəzsiz mənbə rolunu oynayır.

Bundan başqa, tarixi abidələrin öyrənilməsində statistik metodların tətbiqi təkzibolunmaz əhəmiyyət kəsb edir. Dünya dilçiliyində ilk dəfə 1927-ci ildə Hind-Avropa dillərinin qrupları arasındakı qohumluq əlaqələrinin aşkar edilməsi üçün polşa alimi Y.Çekanovski məhz bu metoddan istifadə etmişdir [9]. Amerika alimi M.Svodeş isə 1948-ci ildə leksik-statistik metoddan istifadə edərək, dillərin lüğət fondunu öyrənərkən belə qənaətə gəlmişdir ki, bütün dillərin əsas lüğət fondu sabit sürətlə, yavaş-yavaş dəyişir [9]. Eləcə də, dillərin tipoloji öyrənilməsində statistik metodların böyük rolu vardır.

Azərbaycan dilçiliyində tarixi abidələrin, o cümlədən yazılı abidələrin kompüterlərdə tədqiqi ilə bağlı “Kitabi-Dədə Qorqud”un 1988-ci ildə nəşr olunmuş tənqidi mətni üzrə (tərtibçilər: F.Zeynalov və S.Əlizadə) əlifba-tezlik və əks-əlifba siyahısı hazırlandı [10]. Burada tezlik siyahısı on minə yaxın söz formasını əhatə edir. Dastanda qeydə alınmış sözlər işlənmə tezliyinə görə azalma sırası ilə verilmişdir.

Əlifba-tezlik siyahısında isə söz formalar işlənmə tezliyinə görə deyil, əlifbaya görə düzülüşdür. Tərtib olunmuş bu siyahının dil tarixi ilə məşğul olan mütəxəssislər üçün böyük əhəmiyyəti vardır.

Klassik irsimizdən M.Füzulinin nəzm əsərlərinin statistik-distributiv təhlili tədqiq olunmuşdur. Əlbəttə, bunun klassik ədəbiyyat müəlliflərinin əsərlərində üslubi xüsusiyyətlərin özünəməxsusluğunu aşkar etmək, işlənmə tezliyinin göstəricilərini müqayisə etməklə naməlum müəllifli əsərlərin müəllifini müəyyənləşdirmək üçün böyük əhəmiyyəti vardır.

Bundan başqa, Azərbaycan dilçiliyində linqvistik tədqiqatlarda bir çox problemlərin həllində geniş istifadə olunan “Azərbaycan dilinin əks-lüğəti” tərtib edilmişdir [11].

Əks-lüğət adı altında nə başa düşülür və onu adi lüğətlərdən (ikidilli, izahlı və s.) fərqləndirən cəhət nədir?

Əks-lüğətlər adi lüğətlərdən quruluşuna görə fərqlənirlər. Adi lüğətlərdə sözlər başlanğıc hərflərinə görə əlifba sırası ilə düzülür. Əks-lüğətlərdə isə sözlər sonuncu hərflərinə görə əlifba sırası ilə yerləşdirilir.

“Azərbaycan dilinin əks-lüğəti”, “Azərbaycan dilinin orfoqrafiya lüğəti” əsasında tərtib olunmuşdur. Bundan başqa, “Azərbaycan dilinin orfoqrafiya lüğəti”ndə iştirak etməyən sözlər “Azərbaycanca-rusca” lüğətdən götürülərək əks-lüğətə daxil edilmişdir. “Azərbaycan dilinin əks-lüğəti” türk dilləri üzrə tərtib olunmuş lüğətlərdən həcm etibarilə ən sanballısı hesab edilə bilər.

Elektron lüğətlər - kompüterin yaddaşına daxil olan müxtəlif tipli (adı çəkilən lüğətlər qoyulmuş məqsəddən asılı olaraq tərtib edilir) lüğətlərdir. Yeri gəlmişkən, qeyd etməliyik ki, qoyulmuş məqsəddən asılı olaraq elektron lüğətlər kompüter lüğətlərindən fərqlənir. Kompüter lüğətləri istifadəçilərin kompüter proqramlarından faydalana bilməsi məqsədilə yaradılır. İnternet saytlarında müxtəlif tipli *ABBYY lingvo*, *Poliglot (Azeri dictionaries)*, *Lingvo*, tezaurus, ontoloji, izahlı, orfoqrafiya, əks-lüğətlər və s. lüğətlər daxil edilmişdir.

Maşın təlimi - süni intellektin bir bölməsi olub, öyrətmə və alqoritmlərin modellərinin qurulması məqsədini daşıyır. Süni intellektin bu istiqaməti 1958-ci ildə Dartmut konfransında R.Solomonovun müəllimsiz təlim, yəni kompüter vasitəsilə təlim haqqında məruzəsi ilə əsası qoyuldu və onun inkişafına təkan verildi. Məhz maşınla alqoritmlərin riyazi analizinin öyrənilməsi məsələsi nəzəri informatikanın bölməsi kimi kompüter təliminə zəmin yaratdı. Bu sahəyə obrazların tanınması məsələləri: simvolların, əlyazmaların, nitqin tanınması, mətnlərin analizi, robototexnika ilə əlaqədar kompüterlə görmə daxildir.

Dillərin kompüter vasitəsilə öyrədilməsi

Hal-hazırda müxtəlif sistemli dillərin kompüter vasitəsilə öyrənilməsi dillərin tədrisində böyük vüsət almışdır. Yüzlərlə elektron lüğətlər, müxtəlif sistemli dillər üçün maşın tərcüməsi sistemləri, öyrədici proqram sistemləri, ekspert sistemləri, məlumat-axtarış sistemləri istismara buraxılmışdır. Burada, qısa da olsa, aqlyutinativ dillərin kompüter vasitəsilə öyrənilməsi məsələlərini şərh edəcəyik. Nümunə üçün aqlyutinativ dillərin timsalında Azərbaycan dilinin kompüter vasitəsi ilə öyrənilməsində ən vacib mərhələlərdən biri olan mətnin linqvistik təhlili üzərində dayanacağıq.

Yeri gəlmişkən, təbii dillərin kompüter vasitəsilə öyrənilməsi zamanı gedən proseslərə aydınlıq gətirmək üçün bəzi incəlikləri nəzərdən keçirək. Adətən, dil öyrənərkən hansı vasitələrə müraciət olunur? Əlbəttə, ilk növbədə, lüğət, qrammatika və danışmaq kitabçasını əlavə etmək lazım gəlir.

Kompüterdə dilin öyrənilməsinə gəlincə isə, burada artıq mənzərə dəyişir. Lüğətin elektron versiyasını tətbiq etmək, qrammatikanı formal şəkllə salmaq, məlumat-axtarış sistemlərinin, dialoq sistemlərinin, ekspert sistemlərinin, lüğətlərin səslənməsi, sorğu sisteminin hazırlanması lazım gəlir. Bir sözlə, linqvistik təminat və informasiya texnologiyalarının məsələlərinə daxil olan problemin həlli labüdlüyü meydana çıxır ki, diqqətimizi də, əsasən, bu istiqamətə yönəldəcəyik. Qeyd etmək lazımdır ki, öyrədici sistemlər üçün tərtib olunan ikidilli lüğətlər adi ikidilli lüğətlərdən xeyli fərqlənir. Adi lüğətlərdə söz və onun mənalari verilir. Adı çəkilən sistemə daxil olan lüğətlərdə isə qoyulan məsələlərin səciyyəsiindən asılı olaraq məsələ, maşın tərcüməsi və dillərin tədrisi üçün söz və onun mənalari ilə kifayətlənməyərək, onun hansı nitq hissəsinə mənsubluğu, səs düşümü (yazıda qrafem düşümü), sözün sonuncu hərfinin q, k, t, ilə bitməsi, omonimlik sintaktik və semantik əlamətlər, sabit söz birləşməsində iştirakı, fellərin qruplaşdırılması haqqında (valentlilik və təsirli və ya təsirsiz olması) məlumat verilir. Həm də qeyd etməliyik ki, elektron lüğətlər açıq sistemli olduğundan, istənilən zaman buraya yeni sözlərin əlavə olunması imkanına malikdir.

Öyrədici sistemlər üçün hazırlanan lüğətlərdə isə əyani vasitələrdən istifadə edilərək, sözlərin şəkillərlə təsviri və səsləndirilməsi verilir. Bazaya daxil olan şəkilçilərə gəlincə isə, onlar mənsub olduğu nitq hissələri üzrə qruplaşdırılır. Bundan başqa, sahə terminoloji lüğətləri linqvistik təminat kimi bazaya daxil edilir.

Fonologiya üzrə öyrədici proqramlarda tədrisdə təlim olunan dərslərdən fərqli olaraq, burada sait və samitlər sistemi, ahəngə uyuşma qanunauyğunluqları dialoq şəklində səsləndirilmiş nümunələrlə verilir.

Leksikologiya üzrə öyrədici proqram sisteminə, əsasən, söz yaradıcılığı daxil edilir. Burada əsas vurğu Azərbaycan dilində sözün sintezinin formal təsviri üzərinə düşür [12].

Sözün sintezinin formal təsviri üçün şəkilçilərin söz formaları ilə ahəngə uyuşması, kök və şəkilçilərin qarşılıqlı münasibəti, şəkilçilərin kökə nəzərən tutduqları mövqe, şəkilçilərin söz mövqelərində ahəngi, alqoritmi məsələlərinin sorğu-dialoq şəklə salınması lazım gəlir.

Morfologiya üzrə isə öyrədici proqram sisteminə əsas nitq hissəsi (isim, sifət, əvəzlik, say, zərf, fel) daxildir. Təbii ki, nitq hissələri haqqında məlumatlar formalizə olunmuş şəkildə kompüterin yaddaşına həkk olunur, eləcə də, köməkçi nitq hissələri: qoşma və köməkçi adlar, bağlayıcılar, ədatlar, modal sözlər, bağlama, nida, təqlidi sözlər, imperativ sözlər, uşaq sözlər, ritmik sözlər linqvistik bazaya daxil edilir [13].

Frazeologiya üzrə öyrədici proqram sisteminə frazeoloji sabit dil vahidlərinin varlığını şərtləndirən əsas amillər və frazeoloji vahidlərin təsnifatı məsələləri daxil edilir [14].

Leksik təhlildə mətnlərin abzasa, cümlələrə, sözlərə ayrılması, mətnin hansı dilə mənsub olmasının təyini (açar sözlərin və əlifbanın köməyi ilə), cümlələrin növlərinin müəyyənlişməsi, leksik ifadələrin növlərinin təyini (jarqonlar, qarğışlar və s.) ilə baxılır. Bu məsələlərin kompüterdə reallaşması o qədər də asan məsələ deyildir.

Morfoloji təhlil mətndəki sözlərin differensial hissələrə ayrılması və onların hansı nitq hissəsinə mənsubluğu, bir sözlə, əlamətlərin avtomatik təyini ilə məşğul olur.

Sintaktik təhlildə mətndəki cümlələrin sözləri arasındakı əlaqələr və onlar arasındakı sıra mövqeyi təyin edilir.

Semantik təhlil ən mürəkkəb proses olub, mətndəki cümlələrin məna və məzmununu təyin edir. Məntiqə əsaslanan bu təhlildə sözlər arasındakı məna asılılığı aşkar olunur və bununla əlaqədar qoyulan məsələ formalizə edilir. Semantikanın ənənəvi üsullarla formalaşdırılması qeyri-mümkün olduğundan, bunun üçün mənanın nə dərəcədə düzgün olub-olmamasını müəyyənlişdirən və keyfiyyətini yoxlayan xüsusi keyfiyyət ekspert sistemləri hazırlanmalıdır. Bu da süni intellekt sistemlərinin yaradılmasından asılıdır.

Elmin bu istiqamətdə inkişafı hələ də öz həllini gözləyir. Hal-hazırda mövcud olan informasiya texnologiyaları vasitəsilə axtarış sistemlərinin effektiv işlənməsi lazımı səviyyədə deyildir.

Öyrədici proqram sistemində sintaktik təhlil Azərbaycan dilindəki mətnlərin baş üzvlərinin, ikinci dərəcəli üzvlərinin, sintaktik əlaqələrin axtarışı kimi başa düşülür. Əlbəttə ki, dialoq sistemində bu cür axtarışı aparmaq üçün mütləq onları formal şəkllə salmaq lazım gəlir.

Mətnlərin avtomatik tərcüməsi

Azərbaycan dilçiliyində avtomatik tərcümə və yaxud maşın tərcüməsi (MT) haqqında geniş məlumat verildiyindən, burada yalnız Azərbaycan dilçiliyində həllini gözləyən maşın tərcüməsinin yeni istiqaməti - statistik maşın tərcüməsi üzərində dayanacağıq [2].

Ədəbiyyatda maşın tərcüməsi alqoritmlərinin qurulmasına iki müxtəlif prinsip üzrə: xüsusi qaydalarla və statistikaya əsaslanmaqla yanaşılır.

Statistik maşın tərcüməsi böyükhəcmli dil cütünün müqayisəsinə əsaslanır. Dil cütü dedikdə, bir dil mətnin cümlələrinin digər dildə ona uyğun cümlələrlə ifadəsi nəzərdə tutulur. Sanki insanın iki dildə yazdığı mətn variantlarına, eləcə də, onların tərcümələrinə bənzəyir. Dil cütləri çoxaldıqca uyğunluq əmsalı daha da dəqiqləşir və nəticə etibarilə, statistik tərcümənin keyfiyyəti də artır.

Bəzi hallarda maşın tərcüməsinin hibrit (növündən istifadə edilir. Hibrit maşın tərcüməsində MT-nin müxtəlif uyğun variantlarından istifadə olunur.

Azərbaycanda kompüter dilçiliyinə gəlincə isə, bizim dilçilikdə riyazi metodlardan istifadə edilməsi ötən əsrin 60-cı illərindən - M.Qarayevanın “Riyazi dilçilik tədrisdə” adlı kitabının nəşri ilə başladı [15].

M.Qarayeva tərəfindən təməli qoyulmuş riyazi dilçiliyin artıq bütün qolları üzrə (maşın tərcüməsi, Azərbaycan dilinin formal modellərinin hazırlanması, statistik metodla dilin tədqiqi) tədqiqat işləri aparılmağa başlandı. 1976-cı ildə Nəsimi adına Dilçilik İnstitutunda Azərbaycan dilçiliyinin yeni bir sahəsi olan tətbiqi dilçilik qrupu yarandı. V.Pinesin rəhbərliyi altında tətbiqi dilçilik məsələlərinə daxil olan problemlərlə maşın tərcüməsi (sözün sintezinin formal təsviri, morfoloji analiz, mətnlərin avtomatik sintaktik təhlili və sintezi, mətnlərin avtomatik redaktəsi və s.), formal modellərin hazırlanması (türk dilində fel şəkilləri quruluşunun modelləşdirilməsi, avtomatik sintezin modeli və s.), statistik metodla qədim abidələrin statistik təhlili ilə məşğul olmağa başladı. 1979-cu ildə A.Axundovun “Riyazi dilçilik” [16] əsəri işıq üzü gördü, bundan əlavə, onun bu yaxınlarda çapdan çıxmış ”Struktur və riyazi dilçilik metodlarının türk dillərinə tətbiq edilməsinin bəzi xüsusiyyətləri haqqında” [17] yazdığı məqalə böyük maraq doğurur.

Azərbaycan alimləri yuxarıda sadaladığımız metodlardan öz tədqiqatlarından uğurla istifadə edirlər. Belə ki, aksiomatik metodla A.Axundov, A.Vəliyev, Q.Melnikov [16-20]; modelləşdirmə metodu ilə K.Vəliyeva, V.Pines, M.Mahmudov, Z.Əmirov, Ə.Fətullayev, Ə.Xəlili, Z.Quliyeva, N.İskəndərova [12, 21-27]; nəzəri-lingvistik metodla Q.Melnikov, K.Vəliyeva, A.Vəliyev, M.Mahmudov [28-32]; çoxluqlar, avtomatlar və alqoritmlər nəzəriyyəsiindən istifadə K.Vəliyeva, M.Mahmudov [6, 7, 33, 34], statistik metodla V.Pines, K.Vəliyeva, M.Mahmudov, C.Rəhmanov, V.Sultanov, S.Məmmədova, İ.Şıxıyeva [35-40] işləri aparmışlar.

Qeyd etməliyik ki, görülən işlərlə məhdudlaşmayacağıq, gələcəkdə tətqiqata cəlb ediləcək maraqlı məsələlər: şifahi nitqin tanınması, dillərin kompüter vasitəsilə tədrisi, video-elektron lüğətlərin hazırlanması və s. öz həllini tapacaqdır.

Təbii dillərin kompüterdə tanınması

Yuxarıda qeyd etdiyimiz kimi, təbii dillərin strukturunun dərinədən öyrənilməsi ilə əlaqədar insan-maşın münasibətlərinin zəruriliyi, kommunikativ aktın bütün aspektlərini özündə əks etdirən ünsiyyət üçün ən vacib, dolğun nəzəriyyənin yaradılması, istifadəçilərin tələbatını ödəyəcək səviyyəyə çatdırılması ilə əlaqədar dilçilərin bu işə cəlb olunmasının labüdlüyü 60-cı illərin sonunda ortaya çıxdı. Təbii ki, burada insan-maşın əlaqəsində əsas ünsiyyət dialoq formasında olmalıdır. Dialoq qoyulan məqsəddən asılı olaraq üç səviyyədə: qlobal, tematik və lokal səviyyədə ola bilər. Qlobal səviyyədəki əlaqə istifadəçinin qarşıya qoyduğu məsələnin ümumi xüsusiyyətləri ilə təyin olunur, tematik səviyyədə dialoqun strukturu qoyulan məsələnin konkret alqoritmik həllindən asılıdır, lokal səviyyədə isə iştirakçıların qarşılıqlı dialoqunun ayrı-ayrı mərhələlərinə baxılır [41].

Təbii dillərin maşın vasitəsilə tanınmasına gəlinə isə, onu qeyd etməliyik ki, burada iki tanınma: mətnin və şifahi nitqin tanınması nəzərdə tutulur. Hər iki tanınma ayrı-ayrı istiqamətlər üzrə öyrənilir və geniş sahəni əhatə etdiyi üçün burada yalnız qısa məlumat verməklə kifayətlənəcəyik.

Mətnin avtomatik yolla tanınması üçün, əlbəttə, ilk növbədə, lingvistik təminat hazırlanmalıdır. Lingvistik təminata aşağıdakı məsələlər daxil edilməlidir.

1. qrafemlərin fonoloji səviyyədə avtomatik (maşın vasitəsilə) tanınması;
2. morfemlərin avtomatik (maşın vasitəsilə) tanınması;
3. sintaktik səviyyədə morfemlər zəncirinin tanınması;

Qrafem səviyyəsi dilin uyğun əlifbasının (məndə qrafemin) maşın dilinə (maşın dili əlifbası iki ədəddən ibarətdir: sıfır-0 və vahid-1) çevrilməsi, yəni kodlaşdırılmasıdır. Azərbaycan dilinin əlifbası 32 qrafemdən ibarətdir. Deməli, maşın dilinə çevirəndə 32 qrafem 0 və 1-dən ibarət rəqəmlərlə ifadə olunacaq, yəni kodlaşdırılacaqdır.

Morfemlərin avtomatik tanınmasına gəlinə isə, morfemlər, qrafemlər zəncirindən ibarət olduğundan, onların kodlaşdırılması, uyğun qrafemlərin cəmi ilə ifadə olunacaqdır. Məlumdur ki, morfemlər kök və şəkilçi morfemlərinə ayrılır. Qeyd etməliyik ki, qrafemlər kök və şəkilçi morfemləri maşın yaddaşına daxil edilməlidir.

Sintaktik səviyyədə morfemlərin tanınması üçün frazeoloji ifadələrin, sabit söz

birleşmələrinin kodlaşdırılmış sintaqmaları eyni zamanda kompüterin yaddaşına daxil edilməlidir.

Təbii dillərin kompüterdə tanınması üçün linqvistik təminatla yanaşı, riyazi təminat da hazırlanmalıdır. Əlbəttə ki, riyazi təminatda qoyulmuş məsələnin həlli alqoritminə əsasən tanınma proqramları hazırlanmalıdır.

Hal-hazırda müxtəlif sistemli dillərin tanınma proqramları mövcuddur. Son illərin məhsulu olan skayner qurğuları vasitəsilə yazılı mətnlər kompüterə daxil edilir və tanınma proqramlarının köməyi ilə həmin mətnlər kompüterdə oxunur və onlar üzərində qoyulan məsələ ilə əlaqədar uyğun əməliyyatlar aparmaq mümkün olur.

Danışq dilinin, o cümlədən şifahi nitqin kompüter vasitəsilə tanınması, yazılı mətnlərin tanınmasından xeyli fərqlənir. Doğrudur, həll metodunda oxşar cəhətlər dilin yanaşma səviyyəsində özünü büruzə verir. Burada da artıq qrafem əvəzinə fonemləri linqvistik bazaya təqdim etmək üçün onun səslənməsini, bir tərəfdən, frazeoloji və akustik aspektdə öyrənərək tətbiq etmək lazım gəlir, digər tərəfdən də, fonemlər sisteminə aid olan bütün məlumatlar, fonemlərin əsas əlamətləri və statusu, onların perseptiv və funksional aspekti, korrelyasiyası, linqvistik bazaya daxil edilməlidir. Bundan başqa, linqvistik təminat hazırlanarkən fonemlərin allofonik variasiyası, kompleksiyası, onların tipləri, orfoepiyası, tanınma funksiyası nəzərə alınmalıdır.

Şifahi nitqin tanınmasında başlıca rol hecaların üzərinə düşür. Buna görə də hecaların strukturu dərinlən öyrənilməli və formal modelləri hazırlanmalıdır. Şifahi nitqin tanınmasında əsas amillərdən biri də vurğu problemidir. Dilçilik ədəbiyyatından məlumdur ki, müxtəlif sistemli dillərdə vurğu daşdığı funksiyadan asılıdır. Məsələn, rus dilində hər bir sözün özünəməxsus vurğusu vardır. Çünki bu dildə vurğu kuliminativ funksiya daşıyır. Türk dillərində isə vurğu delimitativ funksiya daşdığından, sözün sonuncu hecasına düşür [42]. Nitqin tanınma proqramında isə vurğu sonuncu hecanın daha yüksək tonla səslənməsi ilə təyin olunur. Bu baxımdan, şifahi nitqin tanınmasında hecaların modelləri, intonasiya, sözlərin ahəngi məsələləri ayrıca tədqiqat obyektindən, biz burada ona toxunmayacağıq.

Qeyd etməliyik ki, süni intellektin maraqlı məsələlərindən biri də qədim əlyazmaların oxunmasıdır. Buna Mayya qəbiləsinin heroqlif yazılarının oxunması üzərində aparılan tədqiqatları nümunə gətirmək olar. Amerika qitəsində yaşayan, öz yazısı olan bu xalq mədəniyyətindən məhrum edilmiş, kitabları itib-batmış, zəngin kitabxanası ispan işğalçıları tərəfindən yandırılmışdır [43]. Dövrümüzdə bu əlyazma Drezden, Madrid və Parisdə saxlanılır. Müxtəlif ölkələrin, xüsusilə ABŞ, Fransa və Almaniyanın alimləri bu əlyazmaların deşifrəsi üzərində çalışmış, lakin bir nəticə əldə edə bilməmişlər. 1951-ci ildə sovet alimi Y.Knorozov əlyazmaları açmış, yəni Mayya əlyazmalarında olan heroqlifləri eyni şəkllə gətirmişdir. Sonra alimlər heroqlifləri kodlaşdırmış, maşının yaddaşına daxil etmişlər. Bundan əlavə, Mayya əlyazmalarındakı şəkillər də kodlaşdırılmış və yaddaşa daxil edilmişdir. Burada məqsəd Mayya əlyazmaları oxunan zaman ayrı-ayrı heroqliflərin mənasının tapılması deyil, mətnin bütöv bir hissəsinin mənasının tapılması olmuşdur.

Məlumdur ki, əlyazmanın oxunmasında qrafem və hecaların tezliyinin təyini mühüm rol oynamışdır. Bir sözlə, Mayya əlyazmalarının deşifrəsi üçün müxtəlif hərflərin birləşmələrinin (qrafemlərin) rastgəlmə tezliyi müəyyən edilməli, sonra Mayya heroqliflərinin rastgəlmə tezliyi hesablanmalı və alınmış tezliklər müqayisə edilməlidir. Qurulmuş proqrama əsasən, alınma tezliklərinin müqayisəsi nəticəsində mayya əlyazmalarındakı heroqliflərin mənalari tapılmışdı. Deməli, kibernetikanın ən böyük nailiyyətlərindən biri olan kompüter vasitəsilə mayya xalqının qədim əlyazmaları dərk edildi.

Kompüterin tətbiq dairəsini nəzərdən keçirdikdə, istər-istəməz, belə bir sual yaranır ki, süni intellekt yaratmaq mümkündürmü?

Süni intellekt, yəni insan kimi “düşünə” və “başla düşülə” bilən “zəkali” maşın yaratmaq üçün əvvəlcə insanın özünü öyrənmək lazımdır. Qeyd etməliyik ki, kompüterdə ən işlənməli, təkmilləşməli hissələr daxiledici və xaricedici qurğulardır. Bu qurğular hələlik əl ilə yazılmış mətnləri birbaşa oxuya, həm də insan kimi danışa bilmirlər. Doğrudur, hal-hazırda skayner

qurğusu vasitəsilə istənilən mətni kompüterin yaddaşına daxil etmək mümkündür. Lakin tanınma proqramları məhdud olduğundan, ixtiyari mətnlərin oxunması da qeyri-mümkündür və ya *Skype* proqramı vasitəsilə İnternet şəbəkəsindən ikitərəfli əlaqələr qurmaq olur. Lakin kompüter təbii dildə idarə etmək üçün şifahi nitqin tanınması problemi hələ də həllini gözləməkdədir. Çünki maşına insanla danışmağı öyrətmək üçün dilin müxtəlif səviyyələrində fonoloji, leksikoloji, sintaktik, qrammatik, morfoloji və semantik cəhətcə geniş tədqiqatlar aparmaq lazım gəlir.

Məlum olduğu kimi, səslər müxtəlif tezlikli dalğaların havada yayılmasıdır. Eyni bir sözü müxtəlif adamlar özünəməxsus şəkildə, müxtəlif cür səsləndirirlər. Aydın ki, hər bir fərdin səsini tonu, tembr, intonasiyası və s. fərqli olur. Bu səbəbdən də kompüter üçün adamların səsini tanımaq çox çətin olur. Lakin səs rəqslərini elektron rəqslərə çevirməklə səsi süni olaraq sintez etmək olar. Səsi tanımaq üçün o, süzgülərdən keçirilir və sonra kompüterin yaddaşında saxlanılan etalonla müqayisə edilir. Əgər bu zaman üst-üstə düşmə baş versə, onda maşın həmin səsi tanıyır, əks halda isə səsi tanıya bilmir. Başqa bir üsul da ayrı-ayrı səsləri deyil, hecaları və sözləri tanımağa əsaslanır.

Kompüterdə “görmə”yə gəlcə isə, burada da hər bir hərf kodlaşdırılıb maşının yaddaşına daxil edilir. Deməli, yalnız kompüterin yaddaşına daxil edilmiş simvollar tanıma biləcəkdir.

Beləliklə, “görən” və “danişan” kompüterin yaranması hələ də süni intellektin yaranması deyildir. Kompüter o vaxt ağıllı hesab etmək olar ki, ona təqdim edilən riyazi məsələni həll etsin və yaxud da özü müəyyən bir məsələ qoya bilsin.

Məsələ burasındadır ki, hazırda “düşünmə”nin nə olduğu haqqında ümumi bir fikir yoxdur. Adətən, “düşünmə” dedikdə, yaradıcılıq qabiliyyəti – daha doğrusu, hər hansı bir məsələni əvvəlki təcrübələrdən fərqli olana, lakin əvvəlki dünyagörüşü ilə xarici aləmin obyektləri arasında yeni qarşılıqlı əlaqə görmək yolu ilə həll etmək qabiliyyəti başa düşülür. Əslində, düşünmə daha geniş bir məfhum olub, özündə mücərrəd hiss və hissdən kənar düşünməni, öyrətmə qabiliyyətini, obrazların formalaşması, qərarların qəbulu və bir sıra başqa xüsusiyyətləri birləşdirir.

Bu sahədə çalışan mütəxəssislərin bir qrupu belə hesab edir ki, insan düşünməsinin bütün xüsusiyyəti informasiyanın emalı prosesinə gətirilə bilər. Buna görə də bu proseslər çox güclü kompüterdə proqramlaşdırılıb işləyə bilər. Beləliklə, həmin kompüter, prinsipcə, intellektual fəaliyyətin bütün sahələrində insana oxşaya bilər. Artıq süni zəkanın yaradılması üçün dərin tədqiqat işləri aparmağa başlanılmışdır.

Nəticə

Beləliklə, dilçilikdə riyazi metodların tətbiqi kor-koranə deyil, ehtiyac duyulan məsələlərin həllinə yönəlmişdir. Qoyulmuş məsələlərin avtomatik proqramlaşdırılmasını həyata keçirmək üçün beynəlxalq maşın dili yaratmaq məqsədilə universal maşın dilləri (alqoritmik dillər) yaradılmışdır. Həmin dillər sayəsində nəinki təbii dillərin qarşısında duran məsələləri, hətta humanitar elmlərin bir sıra məsələlərini, o cümlədən linqvistikada mətnlərin avtomatik təhlili, maşın tərcüməsi, lüğətlərin tərtibi kimi tədqiqatları proqramlaşdırmaq və həll etmək mümkündür.

Yeri gəlmişkən, qeyd etməliyik ki, M.Mahmudovun bu yaxınlarda işıq üzə görmüş, dilçiliyin nəzəri, təcrübə məsələlərinə həsr olunmuş “Kompüter dilçiliyi” əsəri elm aləmində böyük rezonans doğurmuşdur [42]. Monoqrafiyada Azərbaycan dilinin milli korpusu, statistik lüğətçilik, maşın tərcüməsi və formal linqvistik təhlil sistemləri hərtərəfli şərh olunmuşdur. Tədqiq edilən formal təhlil sisteminin linqvistik alqoritmləri dilin yaranması, inkişafı baxımından böyük əhəmiyyət kəsb edir və onlardan məlumat-axtarış sistemlərində, avtomatik sintez və təhlil sistemlərində geniş istifadə etmək olar. Kompüter vasitəsilə tərtib olunmuş Azərbaycan dilinin tezlik və əks-lüğətləri, tarixi abidələrin (“Kitabi-Dədə Qorqud” və M.Füzulinin dilinin) statistik təhlili linqvistik tədqiqatlarda faydalı faktlar verə bilər.

M.Mahmudov tamamilə haqlı olaraq qeyd edir: “Kompüter dilçiliyi yenidən formalaşmaqda olan, lakin böyük gələcəyi olan, bəşəriyyətə fayda gətirəcək bir elm sahəsidir. Cəmiyyətin gələcək inkişafını kompütersiz təsəvvür etmək mümkün olmadığı kimi, dilçilik elminin də gələcək inkişaf istiqamətlərini və perspektivlərini kompüter dilçiliyi olmadan təsəvvür etmək qeyri-mümkündür” [42].

Ədəbiyyat

1. Толдова С.Ю., Ляшевская О.Н. Современные проблемы и тенденции компьютерной лингвистики. В зеркале 24-ой международной конференции по компьютерной лингвистике COLING 2012, Мумбаи // Вопросы языкознания, 2014, № 1, с.120-145.
2. <http://www.wikipedia.org>
3. Штиндлова Й. Применение методов механизации и автоматизации в лексикологической работе за рубежом. Автоматизации в лингвистике. М., 1966, 240 с.
4. Pines V.Y., Mahmudov M.Ə. Elektron hesablayıcı maşınlar, dilçilikdə, B., 1977, 63 s.
5. Зубов А.В. Переработка текста естественного языка в системе «человек-машина». В сб. Статистика речи и автоматический анализ текста, М., 1971, с. 286-434.
6. Vəliyeva K.A., Mahmudov M.Ə., Pines V.Y., Rəhmanov S. Azərbaycan qəzet dilinin tezlik lüğəti, B., 1997, 212 s.
7. Vəliyeva K.A., Mahmudov M.Ə., Rəhmanov S. Azərbaycan qəzet dilinin tezlik lüğəti, B., 2004, 264 s.
8. Məmmədov M.Ə., Fətullayev Ə., Məmmədova S. və b. Azərbaycan dilinin tezlik lüğəti (söz kökləri) B., 2010, I c, 464 s.
9. Методы математической статистики и моделирование в сравнительной истории языкознании, <http://www.yazikoznanie.ru/content/view/127/215>.
10. Vəliyeva K.A., Mahmudov M.Ə., Pines V.Y və b. “Kitabi-Dədə Qorqud” dastanlarının statistik təhlili, B., 1999, 248 s.
11. Mahmudov M.Ə., Fətullayev Ə. Azərbaycan dilinin əks lüğəti, B., 2004, 258 s.
12. Vəliyeva K.A. Mətnin avtomatik təhlili və sintezi, B., 1996, 158 s.
13. Müasir Azərbaycan dili. B.,1981, 443 s.
14. Mirzəliyeva M.M. Türk dillərinin frazeologiyası. B., 2009, 240 s.
15. Караева М.С. Математическая лингвистика в обучении. (К обоснованию проблемы обучение иностранным языком по моделям). Баку, 1964, 55 с.
16. Axundov A.A. Riyazi dilçilik. B., 1979, 79 s.
17. Axundov A.A. Seçilmiş əsərləri. II c., B., 2012, 464 s.
18. Vəliyev A.H. Azərbaycan dilinin keçid şivələri. Doktorluq dissertasiyası. B., 1974, 400 s.
19. Vəliyev A.A. Azərbaycan dilinin keçid şivələri. B., 2005, 334 s.
20. Мельников Г.П. Некоторые способы описание и анализа гармонии гласных в современных тюркских языках // ВЯ, 1962, 6, с.31-58.
21. Пинес В.Я. Моделирование структуры азербайджанских глагольных форм в связи с проблемой автоматического словаря. АКД, 1970, 19 с.
22. Махмудов М.А. Разработка системы формально-морфологического анализа тюркской словоформы (на материале азербайджанского языка) АКД, Баку, «ЭЛМ», 1982, 26 с.
23. Əmirov Z.M. Azərbaycan dilinin formal modellərinin yaradılması və onların əsasında linqvistik prosessorun qurulması. Texniki elmləri namizədi alimlik dərəcəsi almaq üçün təqdim edilmiş dissertasiyanın avtoreferatı, B., 2006, 19 s.
24. Fətullayev Ə.B. Azərbaycan-ingilis maşın tərcüməsi sistemi üçün rəqəm-modelləşdirmə metodunun işlənilib hazırlanması və tətbiqi. NDA, B., 2006, 19 s.
25. Xəlili Ə.M. Deduktiv maşının bilik bazasının tərkib hissəsi kimi “Məhdud Azərbaycan dili”nin formal qrammatikasının işlənilib-hazırlanması, NDA, B., 2009, 34 s.
26. Кулиева З.Ю. Определение оптимальной структуры автоматического словаря и системы машинного перевода. АКД, B., 2011, 46 с.
27. Искендерова Н.А. Программно-частотный и качественно-семантический анализ адекватности переводов и его применение к переводам на различных языках. АКД, 2008, 19 с.
28. Мельников Г.П. Системология и языковые аспекты кибернетики. М., 1978, 368 с.
29. Велиева К.А. Формальное описание синтеза слова АКД, М., 1971, 20 с.

30. Велиев А.Г. Переходные говоры Азербайджанского языка. АДД. Б., 1975, 65 с.
31. Mahmudov M.Ə. Azərbaycan mətnlərinin avtomatik işlənməsi sistemi, DDA, B., 1994, 64 s.
32. Vəliyeva K.A. Mirzəliyeva M.M. Azərbaycanca-İngiliscə tərcümə sistemi (nəzəri problemlər) Əlyazma, 2015, 245 s.
33. Vəliyeva K.A., Məmmədova M.H. Mətnlərin avtomatik redaktəsi. B., 2003, 80 s.
34. Vəliyeva K.A., Mahmudov M.Ə. Linqvostatistika: Fərziyyə və reallıq. B., Tədqiqatlar 1, 2000, s. 56-63.
35. Vəliyeva K.A., Mahmudov M.Ə. Pines V.İ. və b. "Kitabi-Dədə Qorqud"un statistik təhlili (ilkin nəticələr), B., 1999, 248 s.
36. Vəliyeva K.A., Mahmudov M.Ə., Sultanov V.S. M.Füzulinin nəzm əsərlərinin əlifba-tezlik sözlüyü. B., 2004, 548 s.
37. Рахманов Дж. Ф. Статистико-дистрибутивный анализ азербайджанского текста (на уровне графем и фонем). Автореферат канд. дисс. Б., 1988, 23 с.
38. Мамедова М.Г. Автоматизированный отбор лексики в информационно - поисковый тезаурус на основе анализа терминологических словарей. Авт. дисс. канд. тех.наук М., 1984, 20 с.
39. Мамедова М.Г. О создании банка терминологических данных азербайджанского языка // Советская тюркология, 1990, №2, с.84-89.
40. Şixiyeva İ.X. Kitabi-Dədə Qorqud dastanın linqvostatistik xüsusiyyətləri. Filologiya elm. nam., Avtoreferat, B., 1995, 12 s.
41. Попов Э.В. Искусственный интеллект. Кн., 1, Система общения экспертные системы. М., 1990, 464 с.
42. Вейселли Ф.Я. Элементы общей и частной лингвистики. I часть. В., 2011, 357 с.
43. Большая советская энциклопедия, Том 26, второе издание, 1954, 652 с.

УДК 004,6:80

Велиева Камиля А.

Институт Языкознания им. Насими, Баку, Азербайджан

kamilavaliyev@gmail.com

Направления компьютерной лингвистики

В статье рассматриваются основные направления компьютерной лингвистики – обработка естественного языка, корпусная лингвистика, составление словарей посредством компьютера, вопросы машинного перевода.

Ключевые слова: компьютерная лингвистика, обработка естественного языка, машинный перевод.

Kamila A. Veliyeva

Institute of Linguistics named after Nasimi of ANAS, Baku, Azerbaijan

kamilavaliyev@gmail.com

Direction of computational linguistics

The main trends of computer linguistics- natural language processing, corpus linguistics, dictionary compiling, learning and recognition of languages through a computer, matters of machine translation have been analyzed in this article.

Key words: computational linguistics, natural language processing, machine translation.