

AZƏRBAYCAN MİLLİ ELMLƏR AKADEMİYASI
NƏSİMİ ADINA DİLÇİLİK İNSTİTUTU

MƏSUD MAHMUDOV

SÜNİ İNTELLEKTİN
LİŦVİSTİK PROBLEMLƏRİ

“Elm və təhsil”
Bakı – 2024

*Azərbaycan Milli Elmlər Akademiyası Nəsimi adına
Dilçilik İnstitutu Elmi Şurasının 27 noyabr 2023-cü il tarixli
iclasının 7 №-li protokoluna əsasən çap olunur*

Ön sözün müəllifi:

ƏLİ ABBASOV

*akademik, texnika elmləri doktoru,
professor*

Elmi redaktor:

NADİR MƏMMƏDLİ

filologiya elmləri doktoru, professor

Rəyçilər:

RÜFƏT RÜSTƏMOV

filologiya elmləri doktoru, professor

İLHAM TAHİROV

filologiya elmləri doktoru, professor

KAMİLƏ VƏLİYEVƏ

filologiya elmləri doktoru, dosent

**Məsud Mahmudov. Süni intellektin linqvistik problemləri.
Bakı, “ Elm və təhsil”, 2024, 376 səh.**

*Monoqrafiyada süni intellektin dilçiliklə əlaqəli həmhüdüdü sahələri
araşdırılır. Müəllif hesab edir ki, kompüter dilçiliyi, korpus dilçiliyi, statistik lek-
sikoqrafiya, mətn dilçiliyi, koqnitiv dilçilik, neyrolinqvistika, məşin tərcüməsi,
təbii mətnin emalı sistemləri və dilçiliyin informasiya kommunikasiya texnologi-
yaları ilə bağlı digər sahələri süni intellekt problemləri kontekstində öyrənil-
məlidir. Dilçiliyin ayrı-ayrı problemlərinə yeni baxış bucağından yanaşma
daha effektiv və optimal nəticələr əldə etmək imkanı yaratmış olar.*

*Monoqrafiya dilçiliyin yeni istiqamətləri və süni intellekt problemləri ilə
maraqlananlar üçün nəzərdə tutulmuşdur.*

MÜNDƏRİCAT

ÖN SÖZ.	
SÜNİ İNTELLEKTİN LİNQVİSTİK PROBLEMLƏRİNƏ YENİ YANAŞMA	4
MÜƏLLİFDƏN.....	9
I FƏSİL SÜNİ İNTELLEKTİN YARANMASI, MAHİYYƏTİ VƏ BAŞQA ELM SAHƏLƏRİ İLƏ ƏLAQƏLƏRİ	16
II FƏSİL. DİLİN RİYAZİ-STATİSTİK METODLARLA TƏDQIQI, MÜXTƏLİF ÇEŞİDLİ STATİSTİK LÜĞƏTLƏRİN TƏRTİBİ MƏSƏLƏLƏRİ VƏ SÜNİ İNTELLEKT	42
III FƏSİL. MƏTNİN EMALI SİSTEMLƏRİ SÜNİ İNTELLEKT PROBLEMLƏRİ ÇƏRÇİVƏSİNDƏ	89
IV FƏSİL. MƏLUMAT AXTARIŞI SİSTEMLƏRİNDƏ SORĞULARIN TƏŞKİLİ MƏSƏLƏLƏRİ.....	143
V FƏSİL. MƏTNİN FORMAL LİNQVİSTİK TƏHLİLİ SİSTEMLƏRİ SÜNİ İNTELLEKTİN TƏDQIQAT OBYEKTİ KİMİ.....	186
VI FƏSİL. MAŞIN TƏRCÜMƏSİ VƏ SÜNİ İNTELLEKT	310
VII FƏSİL. MİLLİ DİL KORPUKLARI SÜNİ İNTELLEKT KONTEKSTİNDƏ	335
SON SÖZ ƏVƏZİ	351
QISALTMALAR VƏ ŞƏRTİ İŞARƏLƏR	354
İSTİFADƏ OLUNMUŞ ƏDƏBİYYAT	356

ÖN SÖZ

SÜNI INTELEKTİN LİŊVİSTİK PROBLEMLƏRİNƏ YENİ YANAŞMA

Professor Məsud Mahmudovun “Süni intellektin linqvistik problemləri” monoqrafiyasında hazırda bütün dünya ölkələrində prioritet tədqiqat obyektinə kimi araşdırılan və sürətlə inkişaf etməkdə olan süni intellektin dilçiliklə əlaqəli həmhüddud sahələri araşdırılır. Müəllif hesab edir ki, kompüter dilçiliyi, korpus dilçiliyi, statistik leksikoqrafiya, mətn dilçiliyi, koqnitiv dilçilik, neyrolinqvistika, məşın tərcüməsi, təbii mətnin emalı sistemləri və dilçiliyin informasiya kommunikasiya texnologiyaları ilə bağlı digər sahələri süni intellekt problemləri kontekstində öyrənilməlidir. Dilçiliyin ayrı-ayrı problemlərinə süni intellekt problemləri çərçivəsində yeni baxış bucağından yanaşma daha effektiv və optimal nəticələr əldə etmək imkanı yaratmış olar.

Professor Məsud Mahmudovun süni intellektin mahiyyəti ilə bağlı fikirləri də maraqlıdır. Məlum olduğu kimi, mütəxəssislər arasında belə bir fikir formalaşmış ki, süni intellekt insan düşüncəsini, intellektini təqlid edir, onun kimi düşünüb nəticə çıxarmağa cəhd göstərir. Məsud Mahmudovun fikrincə, əslində, süni intellekt insan təfəkkürünü təqlid etmir, əksinə, insan-tədqiqatçı bu və ya digər məsələni süni intellektin “qavraya biləcəyi”, “onun yerinə yetirə biləcəyi” hala gətirir, öz düşünmə qabiliyyətini, təfəkkür tərzini süni intellektə uyğunlaşdırır, “onun səviyyəsinə enir”, sırf formal kateqoriya və anlayışlarla fəaliyyət göstərir. İnsan-tədqiqatçı təfəkkürün, düşüncənin elə bir süni modelini qurur

ki, onun köməyi ilə bütün dil səviyyələrində formal linqvistik təhlil apara bilsin, mətnlərin formal prinsiplərlə emalı mümkün olsun, kompüter səsi (nitqi) mətnə, mətni səsə (nitqə) çevirə bilsin, bir dildən başqa dilə avtomatik maşın (kompüter) tərcüməsi reallaşdırılsın. Bu xüsusiyyət daha çox linqvistik texnologiyaların hazırlanması prosesində özünü büruzə verir. Müəllif belə bir nəticəyə gəlir ki, süni intellekt insan düşüncəsini təqlid etmir, əksinə insan süni intellekt səviyyəsində müəyyən əməliyyatlar aparmağa cəhd göstərir, süni intellektə xas formal anlayış və kateqoriyalardan istifadə etməklə öz düşüncə tərzini süni intellektə uyğunlaşdırır.

Monoqrafiyada süni intellektin sonrakı inkişaf mərhələlərində idarəolunmaz bir vasitəyə çevriləcəyi və insanlıq üçün təhlükə olacağı fikrinin mövcudluğu barədə məlumata da rast gəlinir. Ehtimal olunur ki, sonradan süni intellekt sistemləri toplanmış bilik və təcrübə bazasında xüsusi özüöyrənən proqramların köməyi ilə getdikcə təkmilləşərək daha ciddi, fərqli və gözlənilməz nəticələr əldə etməyə nail ola bilər. Beləliklə, süni intellekt müstəqil, qeyri-asılı sistem kimi insan intellektinin qadir olmadığı, yerinə yetirə bilmədiyi mürəkkəb proseslərin icraçısına çevrilə bilər. Məhz bu cəhəti əsas tutaraq bir çox mütəxəssislər süni intellektin bəşəriyyət üçün təhlükə törədə biləcəyi riskinin olmasını qeyd edirlər.

Professor Məsud Mahmudovun entropiya ilə bağlı fikirləri də maraqlıdır. Müşahidələr göstərir ki, elmi tədqiqat prosesinin ilkin mərhələlərində daha çox yeni faktlar toplanmış olur, yeni anlayış və terminlərdən istifadə olunur. Sonrakı mərhələlərdə tədqiqatın daha çox dərinləşməsinə və genişlənməsinə baxmayaraq yeni faktların aşkarlanması və əldə olunması kəmiyyətcə getdikcə azalmağa doğru gedir. Məlum olduğu kimi, hər hansı mətni təşkil edən cümlələrdə müxtəlif sayda hərflər işlənir və onların informasiya daşıma potensialları və imkanları fərqli olur. Tez-tez rast gəlinən, işlənmə tezliyi böyük olan hərflər az, az müşahidə

olunan, işlənmə tezliyi kiçik olan hərflər isə daha çox informasiya daşıyıcısı olur. Deməli, hərflərin daşdıqları informasiyanın miqdarı onların işlənmə tezliyi ilə tərs mütənəsb olur. Statistik lüğətlərdə də belədir. Sonradan seçmənin miqdarını yeni mətnlər əlavə etməklə nə qədər artırısaq da, tezlik lüğətlərinin sözlüyündə böyük dəyişikliklər baş vermir və ya cüzi fərqlər müşahidə olunur. Bu hadisəni çox vaxt obyektiv olaraq baş verən gözlənilməzlik, qeyri-müəyyənlik, naməlumluq, idrak prosesində qavramada baş verən nizamsızlıq kimi səciyyələndirirlər. Entropiya adlandırılan bu olay elmdə ən mürəkkəb, izahı çətin olan anlayış və eyni zamanda real həyatda müşahidə olunan ən aktual, hərtərəfli elmi şərhini gözləyən məsələ hesab olunur. Entropiya çox geniş əhatə dairəsinə malikdir, təbiət hadisələrində, insan münasibətlərində özünü büruzə verir. İnformasiya nəzəriyyəsində, mətn dilçiliyində, statistik leksikoqrafiyada entropiyanın təzahür özəllikləri süni intellekt sistemlərində nəzərə alınmalıdır.

Daha bir maraqlı yanaşma riyazi məntiq üzrə görkəmli alim Gödelin natamamlıq teoreminin linqvistik yozumu ilə bağlıdır. Həmin müddəalara əsaslanmaqla Məsud Mahmudov belə bir nəticəyə gəlir ki, təkcə fonetika, morfolojiya və ya sintaksis səviyyəsində qapanıb qalmaqla tam fonetik, morfoloji və sintaktik təhlil aparmaq mümkün deyildir. Çünki, hər hansı sistemi onun öz daxilində öz elementləri ilə tam araşdırıb dərk etmək mümkün deyildir. Elə bir mərhələ gəlir ki, mütləq digər səviyyələrə müraciət etmək zərurəti yaranır. Gödelin natamamlıq teoreminin linqvistik yozumu dil vahidlərinin təhlili və sintezi prosesinə daha dərinləndən nüfuz etməyə, formal təhlil sisteminin riyazi-məntiqi-fəlsəfi köklərini araşdırmağa imkan verir. Müəllifin fikrincə, süni intellekt sistemlərində bu amilin nəzərə alınması çox vacibdir.

Monoqrafiyada süni intellektin linqvistik problemləri və onunla bağlı məsələlər öz izahını 7 fəsilə tapmışdır. Birinci fəsilə süni intellektin yaranması, mahiyyəti, başqa elm sahələri ilə əlaqələri barədə yığcam məlumat verilmişdir. Məlum olduğu

kimi, dilçiliyin süni intellektlə əlaqəsi öz başlanğıcını dilin dəqiq elmlərlə inteqrasiyası, riyazi-statistik metodlarla tədqiqi, müxtəlif çeşidli statistik lüğətlərin tərtibi ilə bağlı aparılmış işlərdən götürmüşdür. Bu sahədə Məsud Mahmudovun böyük təcrübəsi və əməyi vardır. Onun iştirakı ilə Azərbaycan dilinin tezlik, əks lüğətləri hazırlanmış, “Kitabi-Dədə Qorqud” dastanlarının və Məhəmməd Füzulinin nəzm əsərlərinin statistik təhlili aparılmışdır. Bütün bunlar monoqrafiyanın ikinci fəslində əks olunmuşdur. Sonrakı fəsillərdə süni intellekt sistemlərində təbii dilin emalı, məlumat axtarışı və sorğu sistemləri, bütün dil səviyyələrində formal linqvistik təhlil və məşin tərcüməsi problemi şərh olunmuşdur. Sonuncu – yeddinci fəsildə milli dil korpuslarının hazırlanması məsələlərinin süni intellektlə əlaqə və münasibətləri hərtərəfli araşdırılır. Bir daha qeyd etmək istərdim ki, müəllif dilçiliyin müxtəlif aspektlərini əhatə edən bu məsələlərin süni intellekt problemləri çərçivəsində araşdırılmasını məqsədəuyğun hesab edir.

Monoqrafiyanın sonluğu (son söz əvəzi hissəsi) da böyük maraq doğurur. Müəllif göstərir ki, dialektik proses olan elmlərin diferensiasiyası və inteqrasiyası da yayın sıxılması və açılması, də nizin qabarması və çəkilməsinə bənzər bir olaydır. Vaxtilə bir fəlsəfə elmi var idi, bütün elmlər başlanğıcını oradan götürüb, bütün elm sahələri fəlsəfə deyilən bilik xəzinəsinin diferensiasiyasının məhsuludur. Yay müəyyən həddə kimi sıxıla bildiyi kimi, elmlərin diferensiasiyası da əbədi deyil, müəyyən zamandan sonra əks proses inteqrasiya baş verməlidir. Süni intellekt elmi də bu dialektik prosesin davamıdır, indiyə qədər diferensiasiyaya uğramış, mikro-sahələrə bölünmüş elmlər inteqrasiyaya meyl edir, vahid elmi konsepsiya, vahid elmi istiqamət ətrafında mərkəzləşmiş olurlar.

Dilçiliyin süni intellekt problemlərinə inteqrasiyasının mahiyyətində də bunlar durur və bu elmlərin diferensiasiyası və inteqrasiyası kimi dialektik idrak prosesinin qanunauyğun nəticəsi kimi səciyyələndirilə bilər.

Müəllif hesab edir ki, dilçilik sahəsində aparılan elmi-tədqiqat işlərinin süni intellekt problemləri çərçivəsində dəyərləndirilməsi əhəmiyyətli məsələdir. Bu həm də həmin araşdırmalara yeni baxış bucağından yanaşmaq deməkdir. Yeni, fərqli yanaşma, cəmiyyət üçün faydalı, səmərəli elmi-tədqiqatların aparılması istiqamətində işlərin genişləndirilməsi imkanı yaradacaqdır.

Professor Məsud Mahmudov dilçiliyin riyazi-statistik metodlarla tədqiqi, statistik lüğətlərin tərtibi, kompüter dilçiliyi, korpus dilçiliyi sahəsində respublikamızda və onun hüdudlarından kənar da tanınmış mütəxəssisdir. Onun “Mətnin formal təhlili sistemi”, “Kompüter dilçiliyi”, “Türk dillərinin milli korpusları” monoqrafiyaları, tezlik və əks lüğətlərin tərtibi sahəsində xidmətləri, Kitabı-Dədə Qorqud”un və Məhəmməd Füzulinin nəzm əsərlərinin statistik təhlili ilə bağlı araşdırmaları elmi ictimaiyyət tərəfindən rəğbətlə qarşılanmışdır.

“Süni intellektin linqvistik problemləri” monoqrafiyasını professor Məsud Mahmudovun növbəti elmi uğuru kimi səciyyələndirirəm.

Əli Abbasov

*Elm və Təhsil nazirliyinin İdarəetmə
Sistemləri İnstitutunun direktoru, akademik*

MÜƏLLİFDƏN

◆ Klassik məntiqin banisi Aristotelin idrak konsepsiyasına görə bir adam danışırsa, onun dediyi ya doğru, ya da yalan ola bilər. Aristotel doğru və yalan arasında aralıq dərəcələri qəbul etməyib. Klassik məntiqə görə doğrunu 1 və yalanı 0 kimi qəbul etsək, dünyanın bütün gerçəklikləri və yalanı (qeyri-gerçəklikləri) 1 və 0 qiymətlərindən ibarət hesab olunur. Lütfi Zadə tərəfindən əsası qoyulan bulanıq məntiq və ya qeyri-səlis məntiqdə isə məsələyə münasibət tamam fərqlidir. Qeyri-səlis məntiqdə isə doğru və yalan 1 və 0-dan ibarət deyil, $[0; 1]$ intervalında baş verir. Lütfi Zadənin qeyri-səlis məntiqi hesab edir ki, doğru və yalan kateqoriyaları arasında aralıq dərəcələr, mərhələlər mövcuddur və idrak prosesində bunlar hökmən nəzərə alınmalıdır. Lütfi Zadə hər şeyin müəyyən interval daxilində dərəcələndiyini deyir, yəni, dünyada qəti olaraq tam ağ və qəti olaraq tam qara deyilən şey yoxdur. Bu iki məfhum arasında minlərlə çalar dəyişikliyi – aralıq nüanslar vardır [az.wikipedia.org/wiki/qeyri-səlis_məntiq]. Onların nəzərə alınmaması idrak prosesini çətinləşdirir, yanlış nəticələrin alınmasına aparıb çıxarır. Lütfi Zadənin qeyri-səlis məntiq nəzəriyyəsinə üç anlayışa diqqət yetirilir:

- qeyri-səlis çoxluq anlayışı;
- qeyri-səlis nisbət anlayışı;
- linqvistik dəyişən anlayış.

Qeyri-səlis məntiqi dərk etmək, bir elmi fərziyyə kimi mənimsəmək üçün ənənəvi təfəkkür tərzindən qismən imtina etmək lazım gəlir.

Dünya “0” və “1”-dən ibarət qapalı sistem deyil, dünyanın mövcudluğu və reallığı $[0; 1]$ diapazonunda daha asan dərk olunur və bir çox hadisələr dəqiq və real izahını tapır.

Dünya “sevgi” və “nifrət”, “hə” və “yox” əksliklərindən (əks qütblərindən) ibarət ola bilməz və ya məsələnin bu cür qoyuluşu, belə yanaşma dünyanın real mənzərəsini, varlığını izah etmək iqtidarında deyil. “Sevgi” və “nifrət” şkalasında sevgidən doğan və getdikcə, addım-addım nifrətə yaxınlaşan və ya nifrətdən uzaqlaşib sevgiyə yaxınlaşan çoxlu sayda hissi çalar və anlayışlar var. Bunlar real dünyanın həqiqi ahəngi, rəngi, hissi çalar və anlayışları, bizi əhatə edən dünyanın, aləmin real mənzərəsi, obyektiv izahıdır. Yalnız bu reallığı qəbul etdikdən, indiyə qədər əsas götürdüyümüz ənənəvi təfəkkür tərzindən qurtulduqdan, qismən imtina etdikdən sonra dünyanı dərk edib izah etmək olar.

Axşam nə vaxt başlanır, nə vaxt bitir. Saat neçə axşam hesab olunur – saat 16:30-dan 11:30-a kimi olan müddəti axşam hesab etmək olarmı?

Səhər nə vaxt başlanır, nə vaxt başa çatır – saat 04-dən 12:30-a kimi səhər hesab olunurmu?

İstinin (eləcə də soyuğun) sərhədləri neçə dərəcədən başlanır, nə vaxt qapanır: +1 dərəcə də isti sayılır, +5 min dərəcə və daha yuxarı da. Belə qeyri-müəyyən, yaygın rəqəmli ifadə fikir dolaşığı, yanlış izahlar doğurur. Nə qədər ki, təbii dil bu mənə spektrini sözlə ifadə edə bilmir, onun da bu məsələdə problemlərinin olduğu üzə çıxır. Bu mənada təbii dilin özünün də belə reallıqları ifadə edə bilməsi üçün yeni metodologiyaya, yeni yanaşmaya ehtiyac hiss olunur.

Məhz bunları əsas götürərək Lütfi Zadə göstərirdi ki, rəqəmlə hesablamadan sözlərlə ifadəyə – linqvistik ifadəyə keçmək lazımdır. Qeyri-səlis məntiqin əsas töhfəsi sözlərlə hesablama aparılması üçün metodologiya yaradılmasıdır. Rəqəmli hesablamalardan sözlərlə hesablamalara – ölçülərlə manipulyasiyadan – hissi obrazlarla manipulyasiyalara keçid qeyri-səlis məntiqdə yeni bir mərhələnin başlanğıcı olardı [az.wikipedia.org/wiki/qeyri-səlis_məntiq; Abdulla K., Əliyev R., 2023].

Bütün bunlar dilçilik elminin tədqiqat obyektinə fərqli yanaşma tələb edir və süni intellektin linqvistik problemlərinin əhatə dairəsini bir qədər də genişləndirmiş olur.

◆ Çox vaxt belə bir fikir səslənir ki, süni intellekt insan düşüncəsini, intellektini təqlid edir, onun kimi düşünüb nəticə çıxarmağa cəhd göstərir [az.wikipedia.org/wiki/süni_intellekt; ru.wikipedia.org/wiki/искусственный_интеллект]. Burada paradoksal bir hal müşahidə olunur. Fikrimizcə, əslində, süni intellekt insan təfəkkürünü təqlid etmir. Bəzən əksinə, insan bu və ya digər məsələni süni intellektin “qavraya biləcəyi”, “onun yerinə yetirə biləcəyi” hala gətirir, öz düşünmə qabiliyyətini, təfəkkür tərzini süni intellektə uyğunlaşdırır, onun səviyyəsinə enir, sırf formal kateqoriya və anlayışlarla fəaliyyət göstərir. İnsan təfəkkürün, düşüncənin elə bir süni modelini qurur ki, ondan formal prinsiplərlə, kompüterlərdə müəyyən linqvistik və ya qeyri-linqvistik əməliyyatlar yerinə yetirmək mümkün olsun. Yəni, dilçi süni intellekt çərçivəsində və kontekstində bütün dil səviyyələrində formal linqvistik təhlil apara bilsin, mətnlərin formal prinsiplərlə emalı mümkün olsun, kompüter səsi (nitqi) mətnə, mətni səsə (nitqə) çevirə bilsin, bir dildən başqa dilə avtomatik məşın (kompüter) tərcüməsi reallaşdırılsın. Bu xüsusiyyət daha çox linqvistik texnologiyaların hazırlanması prosesində özünü büruzə verir. Təbii mətnin emalı prosesində insan süni intellektin düşünmə və qavrayıb reallaşdırma səviyyəsinə enir, onun səviyyəsində sırf formal meyarlarla həll yolları axtarır. Elə formal əlamətlər sistemi işləyib hazırlayır ki, süni intellekt onlara əsaslanıb təbii mətni emal edə bilsin, formal prinsiplər əsasında linqvistik təhlil sistemləri qura bilsin. Belə olduqda süni intellekt insan düşüncəsini təqlid etmir, əksinə insan süni intellekt səviyyəsində müəyyən əməliyyatlar aparmağa cəhd göstərir. İnsan süni intellektə xas formal anlayış və kateqoriyalardan istifadə etməklə öz düşüncə tərzini süni intellektə uyğunlaşdırır. Ehtimal olunur ki, sonra-

dan süni intellekt sistemləri əldə olunmuş bilik və təcrübə bazasında xüsusi özüöyrənən proqramların köməyi ilə getdikcə təkmilləşərək daha ciddi, fərqli və gözlənilməz nəticələr əldə etməyə nail ola bilər. Beləliklə, süni intellekt müstəqil, qeyri-asılı sistem kimi insan intellektinin qadir olmadığı, yerinə yetirə bilmədiyi mürəkkəb proseslərin icraçısına çevrilə bilər. Məhz bu cəhəti əsas tutaraq bir çox mütəxəssislər süni intellektin bəşəriyyət üçün təhlükə törədə biləcəyi riskinin olmasını qeyd edirlər.

◆ Dil qrammatika ilə, qrammatik qaydalarla tənzimlənir. Nəzərə almaq lazımdır ki, qrammatika dilin sonrakı mərhələlərində insan təfəkkürünün məhsulu olaraq yaranıb. Qrammatikanı insanlar dili daha yaxşı mənimsəmək, başa düşmək, dilin mahiyyətinə varmaq, dil hadisələrini dilin daşıyıcılarına, digər xarici dillərdə danışanlara izah etmək məqsədi ilə yaratmışlar. Yəni, dil, ünsiyyət birincidir. Qrammatika sonradan yaranma, ikincidir. Dil qrammatik qanunlarla tənzimlənir. Qrammatik qanunlar dilin mahiyyətinin, işlənmə və istifadə imkanlarının, xüsusiyyətlərinin izahına xidmət edir. “Belə demək, yazmaq olmaz, çünki qrammatikaya ziddir” ifadəsi yanlışdır. Bu və ya digər deyiliş və ya yazılış forması qrammatikaya yox, qrammatik normalara uyğun gəlməyə bilər. Qrammatika dili yaratmır, onu izah edir. Qrammatika getdikcə təkmilləşdirilməli və dilin izaholunmazlıqlarını dərk etməyə, təhlil və izah etməyə istiqamətlənməlidir.

◆ Müşahidələr göstərir ki, elmi tədqiqat prosesinin ilkin mərhələlərində daha çox yeni faktlar toplanmış olur, yeni anlayış və terminlərdən istifadə olunur. Sonrakı mərhələlərdə tədqiqatın daha çox dərinləşməsinə və genişlənməsinə baxmayaraq yeni faktların aşkarlanması və əldə olunması kəmiyyətcə getdikcə azalmağa doğru gedir.

Belə bir məsələyə diqqət yetirək. Məlum olduğu kimi, hər hansı mətni təşkil edən cümlələrdə müxtəlif sayda hərflər

işlənir və onların informasiya daşıma potensialları və imkanları fərqli olur. Tez-tez rast gəlinən, işlənmə tezliyi böyük olan hərflər az, az müşahidə olunan, işlənmə tezliyi kiçik olan hərflər isə daha çox informasiya daşıyıcısı olur. Deməli, hərflərin daşdıqları informasiyanın miqdarı onların işlənmə tezliyi ilə tərs mütənəsb olur. Statistik lüğətlərdə də belədir. Sonradan seçmənin miqdarını yeni mətnlər əlavə etməklə nə qədər artırısaq da, tezlik lüğətlərinin sözlüyündə böyük dəyişikliklər baş vermir və ya cüzi fərqlər müşahidə olunur. Bu hadisəni çox vaxt obyektiv olaraq baş verən gözlənilməzlik, qeyri-müəyyənlik, naməlumluq, nizamsızlıq kimi səciyyələndirirlər. Entropiya adlandırılan bu hadisə (xüsusiyyət) elmi tədqiqat prosesində, eləcə də real həyatda geniş müşahidə olunan və elmi şərhini gözləyən anlayışlardan hesab olunur. Entropiya çox geniş əhatə dairəsinə malikdir, təbiət hadisələrində, insan münasibətlərində də özünü büruzə verir. İnformasiya nəzəriyyəsində, mətn dilçiliyində, statistik leksikoqrafiyada entropiyanın qabarıq əksi müşahidə olunur [[ru.wikipedia.org/wiki/ энтропия](http://ru.wikipedia.org/wiki/энтропия); forbes.ru/82498].

Daha bir yanaşma süni intellekt problemləri çərçivəsində formal təhlil sistemlərinin təşkili ilə bağlı irəli sürülən prinsiplərdir:

Birinci və ən mühüm prinsip ondan ibarətdir ki, bütün səviyyələrdə dilin tam formal ifadəsi mümkündür və həmin formal ifadəyə əsaslanaraq mətnlərin kompüterlərdə müxtəlif məqsədlərlə işlənməsi (təhlili, sintezi, başqa dilə tərcüməsi, referat şəklinə salınması və s.) mümkündür.

İkinci mühüm prinsip, dilin formal ifadəsinə və formal prinsiplərə əsaslanaraq dil hadisələrinin və qanunauyğunluqlarının mahiyyətini aşkarlamağa xidmət edən modellərin yaradılması ilə əlaqədardır. Dilçilikdə modelləşdirmə yeni hadisə deyildir. Tədqiqatda istinad edilən mühəndis dilçiliyi prinsiplərinə əsaslanan linqvistik modellər real dil hadisələrini və qanu-

nauyğunluqlarını təqlid edib doğurmaq keyfiyyətinə malikdir [bax: Пиотровский, 1975, 1979].

Məlum olduğu kimi, linqvistik model bu və ya digər problemin daha dərindən araşdırılmasına, onun adi tədqiqatçı tərəfindən müşahidəsi çətin olan xüsusiyyətlərinin daha qabarıq şəkildə əks olunmasına və öyrənilməsi nə xidmət edir. Bu qəbildən olan linqvistik modelin qurulması prosesində əvvəlcə hər hansı bir linqvistik obyekt haqqında müəyyən fərziyyə irəli sürülür və bu fərziyyəyə müvafiq olaraq həmin obyektin analoqu yaradılır. R.Q.Piotrovskinin irəli sürdüyü model tipi üz quruluşuna və funksiyalarına görə fərqlidir [Пиотровский, 1979]. Bu modellərdə linqvistik obyektin analoquna əsasən onun özü yenidən bərpa olunur. Modelin elmi-nəzəri əhəmiyyəti linqvistik obyektlə onun analoqu əsasında bərpa olunmuş surəti arasında yaxınlıq dərəcəsi ilə ölçülür. Əgər model əsasında bərpa olunmuş obyektlə real linqvistik obyekt üst-üstə düşürsə, onlar arasında böyük yaxınlıq varsa, onda həmin modelin bu və ya digər dil hadisəsini aydınlaşdırmaq xüsusiyyətləri yüksək qiymətləndirilir. Bərpa olunmuş linqvistik obyekt reallıqdan uzaqdırsa, linqvistik modelin əsasını təşkil edən fərziyyə də həqiqətə uyğun hesab olunmur. Məhz bu obyektiv izahedici, aydınlaşdırıcı və həqiqəti real təsvir etmək kimi mühüm potensial imkanları nəzərə alınaraq tədqiqatda R.Q.Piotrovskinin əsaslandırdığı və sınaqdan keçirdiyi bərpaedici mühəndis linqvistik model tiplərinə üstünlük verilmişdir.

Üçüncü prinsip dilçiliyin dəqiq elmlər kateqoriyasına daxil edilməsi ilə əlaqədardır. Dilin bütün səviyyələrində ciddi bir qanunauyğunluq, tamlıq, sistem və dəqiqlik hökm sürür. Fikrimizcə, dil araşdırmalarında müşahidə olunan qeyri-dəqiqliklər, müxtəlif yozumlar dil sisteminin deyil, dilə müxtəlif baxımdan yanaşmanın nəticəsidir.

Nəhayət, dördüncü mühüm prinsip riyazi məntiq elminin inkişafında yeni bir mərhələ açmış Gödelin natamamlıq teore-

mi ilə əlaqədardır [Успенский, 1982]. Bu teorem təkcə alqoritmlər nəzəriyyəsinə deyil, dil sisteminin ayrı-ayrı qatları arasında qarşılıqlı əlaqə və nisbətlərin araşdırılması və nəzərə alınması ilə əlaqədar da uğurla tətbiq oluna bilər.

Məlum olduğu kimi, dil sistemində ayrı-ayrı yarusları təşkil edən vahidlər bir-biri ilə qarşılıqlı əlaqədə olur və bu əlaqədə formalaşır. Əgər fonemlərin bir dil vahidi kimi formalaşması morfevlərdə başa çatırsa, morfevlərin tam izahı sözlərdə açılırsa, analogi olaraq sözlərin tam və bitkin mənası (istər leksik, istər qrammatik) cümlədə məlum olur. Təsədüfi deyildir ki, izahlı lüğətlərdə sözlərin mənasını tam açmaq, bütün məna çalarlarını vermək üçün onlar cümlələrdə – sintaktik əlaqə və əhatələrdə verilir. Cümlənin tam “başla düşülməsi” üçün onu mətndə götürmək lazım gəlir. Mətndə əvvəl və sonra gələn cümlələrin informasiya tutumu və kommunikativ funksiyası cümləyə öz təsirini göstərmiş olur.

Gödel teoreminin linqvistik yozumu belə bir nəticəyə gəlməyə imkan verir ki, təkcə fonetika, morfologiya və ya sintaksis səviyyəsində qapanıb qalmaqla tam fonetik, morfoloji və sintaktik təhlil aparmaq mümkün deyildir. Çünki sistemi onun üz daxilində öz elementləri ilə tam araşdırıb dərk etmək mümkün deyildir. Elə bir mərhələ gəlir ki, mütləq digər səviyyələrə müraciət etmək zərurəti yaranır. Fonemləri tam izah etmək üçün morfologiyaya, morfevləri “anlamaq” üçün sintaksisə (və ya əksinə) müraciət etmək lazım gəlir. Gödelin natamamlıq teoreminin linqvistik yozumu dil vahidlərinin təhlili və sintezi prosesinə daha dərindən nüfuz etməyə, formal təhlil sisteminin riyazi-məntiqi-fəlsəfi köklərini araşdırmağa imkan verir [Mahmudov, 1994a, s.82-83].

Süni intellekt problemləri çərçivəsində formal təhlil sisteminin yaradılması prosesində göstərilən dörd prinsip əsas götürülmüşdür [Mahmudov, 2002, s.8-10].

I FƏSİL

SÜNİ INTELEKTİN YARANMASI, MAHIYYƏTİ VƏ BAŞQA ELM SAHƏLƏRİ İLƏ ƏLAQƏLƏRİ

Süni intellekt (Sİ, rus dilində ИИ – искусственный интеллект, ingilis dilində AI – artificial intelligence) ənənəvi olaraq insanın etməli olduğu, insana xas yaradıcı funksiyanı yerinə yetirə biləcək intellektual sistem, intellektual maşınların, xüsusilə intellektual kompüter proqramlarının yaradılması elmi və texnologiyasıdır. İntellektual sistem dedikdə, həmin sistemin yaddaşında saxlanan, ənənəvi olaraq yaradıcı proses kimi səciyyələndirilən məsələləri həll etməyə qabil olan texniki və ya proqram sistemi başa düşülür. Süni intellektin yaranma tarixi, mahiyyəti, tətbiq sahələri, əhəmiyyəti və inkişaf perspektivləri ilə bağlı bir çox araşdırmalar aparılmış, tədqiqat əsərləri yazılmışdır [Девятков, 2001, Жданов, 2020., Лорьер, 1991, Нильсон, 1973, Рассел, Норвиг, 2006, Смолин, 2004, Хант, 1978, Компьютер обретає разум, 1990 və b.].

Məlumdur ki, süni intellekt və milli dil korpuslarını yaradılması sahəsində bir çox dünya ölkələrində tədqiqatlar aparılmışdır. Aparılan işləri təhlil etməklə Azərbaycan dilinin materialında süni intellekt problemləri çərçivəsində tətbiqi və nəzəri linqvistik texnologiyaların hazırlanması, bu zəmində Azərbaycan dilinin milli korpusunun ayrı-ayrı bloklarının struktur və məzmununun müəyyənləşdirilməsi mühüm əhəmiyyət kəsb edir. Monoqrafiyanın yazılışı prosesində süni intellekt problemləri sahəsində böyük xidmətləri olan J.L.Loryer, H.Nilson, S.Rassel, E.Xant, A.Tyuring, D.V.Smolın, V.V.Devyatkov, A.A.Jdanov və b. əsərlərindən, eləcə də internet resurslarından istifadə olunmuşdur. Adları çəkilən müəlliflərin əsərlərində süni intellektin yaranması, onun inkişaf istiqamətləri, ayrı-ayrı böl-

mələri, perspektivləri, nəzəri və praktik əhəmiyyəti, tətbiq sahələri barədə ətraflı məlumat verilir.

N.Nilson “Süni intellekt” tədqiqatında süni intellektin yaranması, mahiyyəti, digər elm sahələri ilə əlaqələri, inkişaf perspektivləri barədə məlumat vermişdir [Нильсон, 1973].

J.L.Loryer “Süni intellekt sistemi” əsərində süni intellekt sisteminin ümumi məsələlərindən bəhs etmiş, bu sistem çərçivəsində hansı problemlərə önəm verildiyi ilə bağlı mülahizələrini irəli sürmüşdür. Həmin əsərin sonrakı illərdə bu sahədə aparılan tədqiqatlara təsiri böyük olmuşdur [Лорьер, 1991].

S.Rassel və P.Norviq süni intellektə müasir yanaşmaları tədqiq etmiş, onun daha mürəkkəb sistemlərə tətbiqi imkanları və perspektivləri barədə mülahizələr irəli sürmüşlər. Müasir süni intellekt sistemlərində həmin yanaşmalardan müxtəlif məqsədlərlə istifadə oluna bilər [Рассел, Норвиг, 2006].

D.V.Smolin süni intellekt fənninin ali məktəbdə tədrisi ilə bağlı mühazirələrində bu yeni və maraqlı elm sahəsi ilə bağlı dünyanın aparıcı mərkəzlərində və Rusiyada aparılmış işlər barədə yığcam məlumat verməyə müvəffəq olmuşdur [Смолин, 2004].

Əslində, süni intellekt bir çox elm sahələrinin qovuşuğunda yaranmış müstəqil elmdir. Bu elmin məqsədi insan ağılı, düşüncəsini, təfəkkürünü, insan beynini təqlid edə biləcək süni sistemlərin yaradılmasıdır. Süni intellekt elminin yaradılmasında, demək olar ki, bütün elm sahələrindən, ən yeni və özünü doğrultmuş metod, üsul və vasitələrdən istifadə olunur. Bu elmlərə riyaziyyat (nəzəri, tətbiqi, hesablama riyaziyyatı), fizika, biologiya, fiziologiya, biotexnika, kompüter elmləri, neyrofiziologiya, psixologiya, fəlsəfə, məntiq, dilçilik, iqtisadiyyat və bir sıra adını sadalamadığımız elm sahələri aid edilə bilər. Süni intellektin tətbiqində elm və mühəndislik texnologiyasından geniş şəkildə istifadə olunur. Bu sahə insanın sahib olduğu ən dəyərli mülkiyyət olan intellektin maşınlar tərəfin-

dən dəqiq şəkildə imitasiya olunması məqsədilə yaradılıb və inkişaf etdirilib. Süni intellekt yaradıldığı gündən bəri müxtəlif optimist və pessimist fikirlərlə qarşılaşsa da, hazırda ən yeni və müasir texnologiyaların yaradılmasında mühüm yer tutur.

Süni intellekt ifadəsi ilk dəfə 1956-cı ildə Con Makkarti tərəfindən istifadə olunub. O, bu ifadəni işlədərkən süni intellekti maşınları intellektual etmək elmi və mühəndisliyi kimi səciyyələndirib (СМОЛИН, 2004).

Eramızdan əvvəl 4-cü əsrdə yaşamış Aristotel süni intellekt düşüncəsinin əsasını təşkil edən sillogizmlər nəzəriyyəsinin əsasını qoymuşdur. 1275-ci ildə Roman Llull özünün “Ars Maqna” adlı məntiq aparatını ixtira etmişdir. “Ars Maqna”ya istifadəçi xristianlıq haqqında bu və ya digər sorğunu (sualı) daxil edir və qurğu insanın köməyi olmadan dərhal avtomatik olaraq bu suala cavab tapırdı. Bu aparat müsəlmanları xristianlaşdırmaq məqsədilə yaradılmışdı (az.wikipedia.org/wiki/süni_intellekt., Лорьер, 1991., Нильсон, 1973., Смолин, 2004).

Öyrədici təlim sistemləri sərbəst şəkildə süni intellekt biliklərinin və ondan istifadə proseslərinin əldə olunması əsasında qurulur. Bu süni intellekt elminin yaranmasının ilk dövrlərindən mərkəzi problem hesab olunurdu.

Robototexnika sahəsi və süni intellekt elmi bir-biri ilə sıx əlaqəlidir. Robototexnika süni intellektdən istifadə edərək intellektual robotların yaradılması kimi yeni bir elm sahəsini əhatə edir. İntellektual sistemlərdə cisimləri manipulyasiya etmək, olduğu yeri tapmaq, ətraf aləmi təhlil etmək və hərəkəti planlaşdırmaq kimi məsələlərin praktik reallaşdırılması üçün xeyli iş görülməli, yeni metod və proqramlar yaradılmalıdır.

Süni intellekt intellektual qurğular (maşınlar) yaratmağı, xüsusən yalnız insanlara məxsus olduğu düşünülmən yaradıcı fəaliyyətlə məşğul ola bilən intellektual proqram yaratmağı özünə

məqsəd seçmiş elm və texnologiyadır. Bu elm həm də insan intellektinin araşdırılması kimi də başa düşülür.

Süni intellekt bir elm kimi praktik və nəzəri hissələrdən ibarətdir. Praktik cəhətdən süni intellekt problemi bir tərəfdən informatikaya və hesablama riyaziyyatına, digər tərəfdən isə neyrofiziologiya və psixologiyaya aid edilir.

Süni intellekt haqqında ilk fikirlər mexaniki materializmin yaranması ilə eyni vaxtda yaranmış, Dekartın “Metod haqqında mühakimələr” və Tomas Qobbsun “İnsan təbiəti” əsərlərində öz əksini tapmışdır (az.wikipedia.org/wiki/süni_intellekt).

Con Makkarti hesab edirdi ki, süni intellekt bir başa insanla bağlı intellekt kimi başa düşülməməlidir. Süni intellektdə konkret problemlərin həlli prosesində insanlarla bağlı olmayan metodlardan da istifadə oluna bilər (az.wikipedia.org/wiki/süni_intellekt).

Con Makkartiyə görə məsələ ondadır ki, biz hansı hesablanma proseduralarının intellektual adlandırıldığını müəyyən edə bilmirik. Biz intellektin bəzi proseduralarını dərk edirik, digərlərini isə anlaya bilmirik. İntellekt dedikdə bu və ya digər məqsədə çatmaq imkanı verə biləcək hesablamalar nəzərdə tutulur. İstənilən hesablama, araşdırma intellektə aid edilə bilməz (az.wikipedia.org/wiki/süni_intellekt).

Digər tərəfdən intellektin yalnız bioloji fenomen olması fikri də mövcuddur.

İngilis dilində artificial “intelligence” söz birləşməsi kontekstdə “ağıllı mühakimə yürütmək bacarığı” mənasında işlənir. Bu heç də “intellekt” demək deyil. Bizim başa düşdüyümüz mənada “intellekt” termininin analoqu isə “intellect” hesab olunur (СМОЛИН, 2004).

Süni intellekti səciyyələndirən xüsusiyyətlər:

- ənənəvi olaraq intellektual hesab olunan insan fəaliyyətinin modelləşdirilməsi ilə bağlı müəyyən məsələlərin qoyuluşu və həlli;

- ənənəvi olaraq insanın preroqativi sayılan funksiyaları yerinə yetirmək üçün ağıllı sistemlər – bilikləri yaddaşında saxlaya bilən, müəyyən bir mövzu sahəsinə aid yaradıcı hesab olunan problemləri həll edə biləcək texniki və ya proqram sistemidir. Bu qəbildən olan sistemlərin quruluşuna üç əsas blok daxildir – məlumat bazası, prosessor (emal sistemi) və məlumatların daxil edilməsi üçün xüsusi proqramlar olmadan kompüterlərlə əlaqə yaratmağa imkan verən interfeys;

- kompüter elmləri və informasiya texnologiyaları sahəsindəki vəzifəsi kompüter sistemlərinin və digər süni cihazların köməyi ilə ağılabatan düşüncə və hərəkət modelinin yaradılması;

- sistemin aparıcı xüsusiyyəti xarici məlumatları düzgün şərh etmək, bu cür məlumatlardan öyrənmək və çevik uyğunlaşmadan istifadə edərək müəyyən məqsəd və vəzifələrə çatmaq üçün əldə olunan biliklərdən istifadə etmək bacarığı;

- hesablama sistemləri və digər süni qurğuların köməyi ilə əqli mühakimə və fəaliyyət sistemləri yaratmağa qabil məsələləri əhatə edən informatika və informasiya texnologiyası istiqaməti;

- süni intellektin insan və maşın üçün ümumi olan xüsusiyyətlərindən biri özüöyrənən proqramlar vasitəsilə (ilk növbədə evristik proqramların köməyi ilə) müəyyən mürəkkəblik sinfinə aid məsələlərin həlli məqsədilə proqramlar sistemi yaratmaq və onları həll etmək hesab olunur (az.wiki.pedia.org/wiki/süni_intellekt).

Paradoksal görünsə də, süni intellekt nə ilə məşğul olur sualının birmənalı cavabı yoxdur. Hər bir tədqiqatçı süni intellektə müəyyən bir aspektdən yanaşır.

Fəlsəfi cəhətdən insan intellektinin təbiəti və mənşəyi tan izahını tapmamışdır. Kompüter “şüurluluğunun” da hər hansı bir həddi yoxdur, lakin süni intellektin yaranması və yaradıl-

ması ilə bağlı bir sıra fərziyyələr məlumdur (Türinq testi və Nyuel-Saymon fərziyyəsi).

Yeni elmi istiqamət kimi süni intellektin tarixi XX əsrin ortalarından başlanır. Həmin dövrdə süni intellektin yaranması ilə bağlı ilkin şərait yaranmışdı. Filosoflar arasında insan təbiəti və dünyanın dərkilə bağlı mübahisələr gedirdi. Neyrofizioloqlar və psixoloqlar insan beyni və təfəkkürü ilə bağlı nəzəriyyə üzərində işləyirdilər. İqtisadçı və riyaziyyatçılar optimal hesablar və dünyanın formal dərkilə bağlı biliklərin təqdimi ilə məşğul olurdular. Nəhayət, həmin dövrdə hesablamaların nəzəri bazası, alqoritmlər nəzəriyyəsi və ilk kompüterlər yaranmışdı.

Yeni kompüter texnologiyası hesablama sürətinə görə insanı qat-qat üstələyir. Elmi ictimaiyyətdə kompüterlərin mümkün imkanları barəsində müxtəlif ideyalar dolaşırdı. Kompüterlər insan təfəkkürü səviyyəsində inkişafa, fəaliyyətə qabildirlərmi? 1950-ci ildə hesablama texnikası sahəsində tədqiqatların pioneri sayıla biləcək ingilis alimi Alan Türinqin “Maşın düşünə bilərmimi?” məqaləsi işıq üzünə gördü. Həmin məqalədə alim sonradan “Türinq testləri” adı almış yeni yanaşma irəli sürdü. Həmin testlə maşının insan düşüncəsi ilə müqayisə olunduğu səviyyənin təsviri verilir (Turing, 1950).

1986-cı ildə keçmiş SSRİ EA Rəyasət Heyətində “Süni intellekt” problemi üzrə Elmi Şura yaradılmışdı. Həmin Elmi Şura sonralar Rusiyada və bütün keçmiş SSRİ məkanında süni intellekt sahəsində tədqiqatların inkişafında mühüm rol oynadı (ru. wikipedia.org/wiki/искусственный_интеллект).

2019-cu ildə Rusiyada rəqəmli iqtisadiyyatın inkişafına həsr olunmuş müşavirədə süni intellekt sahəsində milli strategiya ilə bağlı qərar qəbul edilmiş və bu məqsədlə 90 milyard rubl ayrılmaqla federal proqram hazırlanması işlərinə start verilmişdir.

2019-cu ildə Rusiyada 2030-cu ilə kimi süni intellektin inkişafı ilə bağlı milli strategiya, 2020-ci ildə isə “Süni intellekt” milli layihəsi təsdiq olunmuşdur.

2020-ci ilin dekabrında Rusiyada süni intellektə dair ikinci beynəlxalq konfrans *Artifisial İntelligence Journey (AI Journey)* keçirildi. Bu konfrans dünyada analoji olaraq keçirilən konfranslar üçlüyündə hesab olunur. Konfransda (onlayn) 80 ölkədən tədqiqatçı alimlər iştirak edirdi (ru. [Wikipedia.org/wiki/искусственный_интеллект](https://ru.wikipedia.org/wiki/искусственный_интеллект)).

Süni intellektin tədqiqat obyektinə bəzə müxtəlif mülahizələr irəli sürülür. Sİ haqqında tədqiqat əsərləri yazan müəlliflər konkret tərifdən yayınır, bu sahəyə qazanılmış elmi nailiyyətlər fonunda nəzər yetirirlər.

Hesab olunur ki, fəlsəfə elmi insan intellektinin mahiyyəti (təbiəti) məsələsinin tam izahını verə bilməmişdir. Kompüter düşüncəsi, təfəkkürü məsələsinin də dəqiq meyarları məlum deyildir. Yuxarıda deyildiyi kimi, Sİ-nin yarandığı ərəfədə kompüter düşüncəsi, təfəkkürü məsələsini müəyyənləşdirməyə imkan verən bir sıra fərziyyələr – Türing testi və ya Hüeyll–Saymon fərziyyəsi təklif olunmuşdu. Sİ məsələlərinin qavranması və intellektual məlumat sistemlərinin yaradılmasına müxtəlif yanaşmalar olsa da, Sİ-nin yaradılmasında iki əsas yanaşmanı ayırmaq olar:

- alçalan, enən (ing. Top-Down AI), semiotik – çoxsəviyyəli psixi prosesləri imitasiya (təqlid edən) ekspert sistemləri, bilik bazası və məntiqi nəticə sistemlərinin yaradılması (mühakimə, nitq, emosiya, yaradıcılıq və s.);

- yüksələn, qalxan (ing Bottom-Up AI), bioloji – bioloji elementlər əsasında neyroşəbəkələr və təkamül məsələlərinin öyrənilməsi, intellektual davranışın modelləşdirilməsi, eləcə də neyrokompüter və ya biokompüterlərə müvafiq hesablama sistemlərinin yaradılması (ru. [wikipedia.org/wiki/искусственный_интеллект](https://ru.wikipedia.org/wiki/искусственный_интеллект)).

Empirik test Alan Türingin 1950-ci ildə “Mind” fəlsəfə jurnalında çap olunmuş “Hesablama maşınları və ağıl (intellekt)” məqaləsində təklif olunmuşdu. Bu testin məqsədi süni intellektinin insan təfəkkürünə yaxınlığının müəyyənəşdirilməsidir [Turing, 1950].

Bu testin standart interpretasiyası belədir: təcrübə aparan (insan) bir kompüter və bir insanla qarşılıqlı əlaqədə olur. Onun verdiyi suallara verilən cavablara əsasən təcrübə aparan, kompüter proqramı və ya insanla ünsiyyətdə olduğunu müəyyənəşdirə bilməlidir. Kompüter proqramının vəzifəsi insanı çaşdırmaq, düzgün olmayan seçimə məcbur etməkdir. Proses zamanı test iş-tirəkləri bir-birlərini görmürlər.

Ən ümumi yanaşma ehtimal edir ki, Sİ normal situasiyalarda insan davranışından seçilməyən davranış nümayiş etdirməyə qabildir. Bu yanaşma Türing testinin ən ümumiləşdirilmiş yanaşmasıdır. Həmin yanaşmaya görə insan kompüterlə ünsiyyətə girir və bu zaman o, (insan) maşınla ünsiyyət qurduğunu (ünsiyyət, danışq yazılı şəkildə aparılır) başa düşməməli, hiss etməməlidir.

Sİ tarixinin təhlili nəticəsində çox geniş bir istiqamət olan düşüncənin (fikir, mühakimə) modelləşdirilməsini xüsusi ayırmaq olar. Qeyd etmək lazımdır ki, bu istiqamət indi müasir Sİ sistemlərində ən çox inkişaf etmiş sahələrdən hesab olunur. Düşüncənin modelləşdirilməsi simvol sisteminin yaradılmasını nəzərdə tutur. Bu sistemdə girişdə müəyyən məsələ qoyulur və çıxışda onun həlli tələb olunur.

Bir qayda olaraq bu və ya digər məsələ burada artıq formallaşdırılmış, müəyyən riyazi formaya gətirilmiş olur. Alqoritm çox mürəkkəb, irihəcmli və çox zəhmət tələb edən olduğu üçün tərtibi böyük çətinlik doğurur. Düşüncənin modelləşdirilməsinə aşağıdakılar daxildir: sübut, teorem, qərarın qəbulu, oyun nəzəriyyəsi, planlaşdırma, dispəçerləşdirmə, proqnozlaşdırma.

Bilik mühəndisliyi istiqaməti sadə informasiyadan biliyin alınması, sistemləşdirilməsi və istifadəsi məsələlərini əhatə edir. Bu istiqamət hər hansı problemə aid etibarlı bilik almaq üçün ixtisaslaşdırılmış bilik bazaları ilə əlaqəlidir.

Verilənlərdən biliyin alınması onların intellektual təhlilinin əsas məsələlərindən hesab olunur. Bu problemin həlli ilə bağlı müxtəlif yanaşmalar mövcuddur. Bu yanaşmalardan biri də neyroşəbəkələrin verballaşdırılması prosedurundan istifadə edən neyroşəbəkə texnologiyasına əsaslanır.

Maşın tədrisi biliyin müstəqil olaraq intellektual sistemin işi prosesində alınması məsələlərini ehtiva edir. Sİ yaranandan bəri bu məsələ mərkəzi istiqamətlərdən biri hesab olunmuşdur. 1956-cı ildə Rey Solomonoff müəllimsiz tədris aparən ehtimali maşın haqqında hesabatla çıxış etmişdir. Müəllif bu maşını “Nəticənin induktiv maşını” adlandırmışdır [ru.wikipedia.org/wiki/искусственный_интеллект].

Maşın tədrisi alqoritmlərinin riyazi analizi nəzəri informatika bölməsinə aid edilir. Bu bölmə “hesablama tədris nəzəriyyəsi” (ing. Computational learning theory) adlanır. Obrazların tanınması ilə bağlı məsələlər sinfi də maşın tədrisi sahəsinə aid edilir. Məsələn, əlyazma mətnlərində və nitqdə simvolların tanınması, mətnlərin təhlili və s.

Bu istiqamətdə bir çox məsələlər bioloji modelləşdirmənin köməyi ilə əldə olunur. Qeyd etmək lazımdır ki, kompüter görməsi məsələsi robototexnika ilə əlaqəli məsələdir.

Con Makkartinin süni intellekt nəzəriyyəsiindən fərqli olaraq hesab olunur ki, süni sistemlər bioloji sistemlərə xas funksiyaları və strukturları eyni ilə təkrarlamağa məhkum deyillər. Bu müddəanın tərəfdarlarına görə insan davranışı fenomenləri, onun dərk etmə, mənimsəmə bacarığı və adaptasiyası bioloji struktur və onun funksional xüsusiyyətlərinin nəticəsidir (ru.wikipedia.org/wiki/искусственный_интеллект). Bura bir neçə istiqaməti aid etmək olar. Neyron şəbəkələr geometrik fi-

qurların tanınması və ya obyektlərin qeyri-dəqiq və mürəkkəb sistemlərlə təhlilində istifadə olunur.

Genetik yanaşmaya görə hər hansı alqoritm başqa alqoritmlərin yaxşı cəhətlərin əxz etdikdə daha effektiv ola bilər. Yeni yanaşmalarda, xarici mühitlə qarşılıqlı əlaqədə olan avtonom proqramların – agentlərin yaradılması məsələsində istifadə olunur. Bu yeni yanaşma agent yanaşması adlandırılır.

Robototexnika və süni intellekt problemləri bir-biri ilə sıx bağlıdır. Bu iki elm sahəsinin inteqrasiyası və intellektual robotların yaradılması süni intellektin daha bir istiqamətini təşkil edir. İntellektual robotlar obyektlərlə manipulyasiya etmək, lokallaşma problemləri ilə bağlı naviqasiya işlərini yerinə yetirmək (yerləşmə məkanını təyin etmək, yaxın obyektləri öyrənmək), hərəkəti planlaşdırmaq (nəzərdə tutulan yerə necə çatmaq) işlərində istifadə olunur.

İnsanın yaradıcı potensialı intellektin təbiətindən az öyrənilmişdir. Buna baxmayaraq süni intellekt sistemlərində bu sahə mövcuddur, kompüterin musiqi, ədəbi əsərlər (daha çox – şeir və ya nağıl), bədii yaradıcılıq məsələləri ilə məşğul olması məsələsi diqqət mərkəzindədir.

Qeyd etmək lazımdır ki, süni intellekt sisteminin texniki yaradıcılıq problemləri ayrıca öyrənilir. 1946-cı ildə Q.S.Altşuller tərəfindən təklif olunan ixtiraçılıq məsələlərinin həlli nəzəriyyəsi belə tədqiqatların başlanğıcı hesab olunur [ru. wikipedia.org/wiki/искусственный_интеллект].

Süni intellektin çoxlu sayda tətbiq sahələri vardır ki, onların da hər biri müstəqil istiqamət kimi fəaliyyət göstərir. Məsələn olaraq kompüter oyunlarında intellektin proqramlaşdırılması, qeyri-xətti idarə, informasiya təhlükəsizliyinin intellektual sistemlərini göstərmək olar.

Gələcəkdə süni intellektin inkişafının kvant kompüterləri ilə sıx əlaqəsi nəzərdə tutulur. Süni intellektin bəzi xüsusiyyətləri kvant kompüterləri ilə oxşar fəaliyyət prinsiplərinə malik-

dir. Burada bir çox elm sahələrinin kəşif, təmas nöqtələrinin olduğu müşahidə olunur. Süni intellektə ayrı-ayrı istiqamətlər arasında qarşılıqlı əlaqə daha qabarıq hiss olunur. Bu, güclü və zəif süni intellektlə bağlı fəlsəfi mübahisə ilə şərtlənir. Son zamanlar süni şüur və onun modelləşdirilməsi ideyası ətrafında söhbətlər gedir. İnsan aqlının, şüurunun modelləşdirilməsi həllini gözləyən elmi problemlərdəndir.

Hazırda süni intellektin inkişafının iki istiqamətini xüsusi ayırmaq lazımdır:

- xüsusiləşdirilmiş süni intellekt sistemlərinin imkanları ilə insanın imkanlarının yaxınlaşdırılması problemlərinin həlli və onların inteqrasiyası (intellektin gücləndirilməsi);
- insan problemlərini həll etməyə qabil olan artıq yaradılmış süni intellekt sistemlərinin vahid sistemini təmsil edən süni düşüncənin yaradılması (güclü və zəif süni intellekt).

Populyar süni intellekt sistemlərinə aşağıdakıları misal göstərmək olar:

- Pepper (robot) – Yapon şirkəti olan Aldebaran Robotics və Softbank Mobile tərəfindən 2015-ci ildən satışa çıxarılmışdır. Bu humanoid robotun digər robotlardan fərqi odur ki, onun emosiyaları var. Pepper ətraf aləmi 3D kameralar vasitəsi ilə təhlil edə bilir, boyu 1,2 metr, çəkisi isə 28 kilodur. O, insanların mimikaları ilə onların əhval-ruhiyyəsini başa düşə bilir. Həmçinin o, öz-özünü öyrədən robotdur və lazım olan cavabları axtarmaq üçün internetə çıxışı var. Pepper hətta onu itələdikdə belə öz müvazinətini saxlaya bilir, insan toxunuşlarını hiss etmək üçün onda sensor qurğular vardır;

- Watson – insanın danışığını dərk edən və çoxsaylı alqoritmlər sahəsində ehtimali axtarış aparmaq bacarığına malik robot IBM şirkətinin qabaqcıl inkişaf sistemidir. Öz bacarığını nümayiş etdirmək üçün Watson ABŞ-ın “Leopard” oyununda iştirak etmiş,

hər iki oyunda qələbə çalmağı bacarmışdır [ru. wikipedia.org/wiki/искусственный_интеллект];

- ChatMaster – virtual söhbət proqramları insan danışıqını təqlid edən, bir və daha çox insanla söhbət edən kompüter proqramıdır. İlk belə proqramlardan biri 1969-cu ildə Cozef Beyzenbaum tərəfindən yaradılan Eliza proqramı olmuşdur. Belə ehtimal olunur ki, ideal virtual söhbət proqramı Türinq testini keçə biləcək. Chatmaster özü-özünü öyrədə bilən, insanla dialoq apara bilən proqramdır. Dialoq klaviatura ilə aparılır və ekranda əks olunur. Dialoqu bütün dillərdə və hətta heroqliflərlə də aparmaq mümkündür. Proqram yalnız yazılmış sözə deyil, həm də mətnin özünə reaksiya verə bilir. Proqramın cavabını düzəltmək, istənilən cavabla dəyişdirmək olur;

- Deep Blue – IBM tərəfindən işlənib hazırlanmış şahmat proqramıdır. Şahmat üzrə dünya çempionuna qalib gəlməyə müvəffəq olmuşdur. Superkompüterin dünya çempionuna qarşı oyunu kompüter mütəxəssislərini və şahmatçıları tam qane etməmiş, sistemin bir çox qüsurları olduğu qeyd olunmuşdur;

- MYCIN – insan-həkim səviyyəsində az sayda xəstəliklərin diaqnozunu dəqiq təyin edə bilən sistem hesab olunur;

- Nitqin tanınması – ViaVoice tipli sistemlər istifadəçilərin tələbini ödəməyə qabil hesab olunurlar;

- RoboCup – hər il keçirilən sadələşdirilmiş futbol turnirində robotlar yarışır. Kompüter oyunları ixtiraçıları Sİ-dən müəyyən dərəcədə istifadə etmiş olurlar. Bu istiqamətdə “Oyun süni intellekti” anlayışı yaranmışdır. Oyunlarda Sİ-nin standart vəzifəsi ikiölçülü və ya üçölçülü məkanlarda yolun tapılması, döyüş vahidinin hərəkətinin imitasiyası, düzgün iqtisadi strategiyanın hesablanması və s. hesab oluna bilər [ru. Wikipedia.org/wiki/искусственный_интеллект].

Süni intellekt sahəsində ən iri elmi-tədqiqat mərkəzləri aşağıdakılardır:

- Amerika Birləşmiş Ştatları – Massaçuset Texnoloji İnstitutu və Maşın İntellekti Tədqiqat İnstitutu;
- Almaniya – Süni İntellekt Tədqiqatları üzrə Alman Mərkəzi;
- Yaponiya – Müasir Sənaye Elmi və Texnologiyası Milli İnstitutu;
- Rusiya – Rusiya Elmlər Akademiyasının Süni İntellekt Metodologiyası üzrə Elmi Şurası;
- Hindistan – Mədrəsdəki Hindistan Texnoloji İnstitutu [ru.wikipedia.org/wiki/искусственный_интеллект].

Azərbaycan Respublikasında Elm və Təhsil nazirliyinin İdarəetmə Sistemləri İnstitutunda Süni İntellekt laboratoriyası fəaliyyət göstərir.

Süni intellektin tətbiq sahələri müxtəlifdir:

- maliyyə təşkilatları şübhəli hadisə və hərəkətlərin qarşısını almaq üçün çoxdan neyron şəbəkələrindən istifadə edirlər. Süni intellektdən bank sistemində istifadəyə 1987-ci ildən başlanmışdır. Həmin ildə Amerika Security Pacific National Bankı dələduzluq və debet kartlarından qeyri-qanuni istifadənin qarşısını almaq üçün xüsusi qrup yaratmışdı;
- alqoritmik ticarətdə insan orqanizminin qabil olduğu sürətdən daha yüksək sürətdə ticarət şərtlərini qəbul edən mürəkkəb süni intellekt sistemlərindən istifadə ehtimal olunur. Bu gün ərzində insan müdaxiləsi olmadan milyonlarla ticarət razılaşmaları əldə etmək imkanı verir. Avtomatlaşdırılmış ticarət sistemləri bir qayda olaraq nəhəng institutsional sərmayədarlar tərəfindən istifadə olunur;
- Sİ-nin insanların öz şəxsi maliyyələrini idarə etməsinə kömək məqsədilə istifadə olunan layihələri (məhsulları) də mövcuddur. Məsələn, Digit tətbiqi insanların şəxsi vərdiş və məqsədlərinə əsasən onların xərc və əmanətlərini avtomatik olaraq tənzimləyən süni intellekt sistemidir. Həmin tətbiq aylıq

gəlir, cari balans, xərc vərdişləri və s. təhlil etməklə xüsusi qərar qəbul edə və maliyyə vəsaitini ayrıca əmanət hesabına keçirə bilər;

- Sİ-dən istifadə 2019-cu ildə “Сбербанк”a əlavə olaraq 700 milyon dollar, qazanmaq imkanı vermişdir. 2020-ci ildə qazanc məbləğinin təxminən bir milyard dollara çatdırılması planlaşdırılır[ru.wikipedia.org/wiki/искусственный_интеллект];

- Sİ hərbi sahədə döyüş sahəsinin və silahlanmanın optimal idarəsinin perspektiv sistemlərinin yaradılması işlərində böyük səmərə verə bilər.

ABŞ Müdafiə Nazirliyinin məlumatına görə Çin gələcək silahlanma sistemində Sİ-nin tətbiqi metodlarının işlənilməsi barədə qərar qəbul etmişdir. Həmin məqsədlə Hərbi Sənaye Kompleksi və xüsusi şirkətlərin birgə fəaliyyətinin birləşdirilməsi yolu ilə bu proqramın reallaşdırılması məsələsi məqsəd kimi qarşıya qoyulmuşdur.

Sİ internet – senzura işlərində də istifadə olunur. Britaniya xüsusi xidmət orqanları feyk (yalan) xəbərlərlə mübarizə məqsədilə “trol fabriki”nin fəallığını tanımağa imkan verən süni intellekt sistemlərindən istifadə edirlər. Böyük Britaniyanın Dövlət Əlaqələri Mərkəzinin məlumatına görə süni intellekt etibarlı mənbələrlə tutuşdurmaqla video və digər təsvirlərlə manipulyasiyaları aşkar etmək və şübhəli dezinformasiyaları bloklamaq imkanına malikdir.

Robotlar insanlar üçün təhlükə yaradacaq sənaye sahələrində geniş tətbiq olunmağa başlamışdır. Eləcə də, tez-tez təkrarlanan, yorucu ağır işlərdə müəyyən zaman müddətində diqqət və nəzarətin zəifləməsi nəticəsində səhvlərə və ya bədbəxt hadisələrə aparıb çıxaracaq situasiyalarda robotların istifadəsi böyük effekt verir. Həmçinin insan şəxsiyyəti üçün aşağılayıcı, alçaldıcı hesab olunan iş yerlərində süni intellekt sistemlərinin məhsulu olan robotlar geniş tətbiqini tapır.

2014-cü ildə Çin, Yaponiya, Amerika Birləşmiş Ştatları, Koreya və Almaniya dünya robot ticarətinin 70%-ni ehtiva edirdilər. Avtomobil sənayesində avtomatlaşdırma səviyyəsi daha yüksəkdir. Yaponiya sənaye robotlarının sıxlığına görə dünyada lider ölkədir: hər 10000 əməkdaşa 1414 robot düşür [ru. wikipedia.org/wiki/искусственный_интеллект].

Kompüter program təminatının Concept Processing texnologiyasında süni neyron şəbəkələr tibbi diaqnostikada qərarın qəbulunun klinik sistemi kimi istifadə olunur.

Bundan başqa, səhiyyə sistemində süni intellektdən istifadə oluna biləcək sahələrə aşağıdakıları aid etmək olar:

- tibbi təsvirlərin kompüter interpretasiyası. Belə sistemlər rəqəmli təsvirləri skanlamaq imkanı verir, məsələn, kompüter tomoqrafiyasında tipik təzahürlər və ehtimal olunan xəstəliklərdən mümkün yayınmaların seçilib ayrılması, tipik tətbiqi şişlərin aşkar edilməsi;

- ürək ritminin təhlili;

- Watson layihəsi – Sİ-nin səhiyyədə tətbiqi sahələrindən biridir. Sual-cavab sistemi həkim-onkoloqlara kömək məqsədi ilə yaradılmışdır;

- ahıllara qulluq edən robot-köməkçilər;

- daha faydalı məlumatın təqdimi üçün tibbi yazıların emalı;

- müalicə planının tərtibi;

- xəstəlik riskinin yüksəlməsinin aşkar edilməsi;

- medikamentlərin qəbulunun idarəsi də daxil olmaqla təkrarlanan məsələlərə kömək;

- məsləhətləşmələrin təqdimi;

- dərmanların yaradılması;

- klinik tədrisdə pasiyentlərin yerinə insanabənzər ma-nekenlərdən istifadə.

Hazırda səhiyyə sahəsində Sİ-dən istifadəyə əsaslanan 90-dan yuxarı startap (ingilis dilində start up – başlayan, müəyyən ideyanı həyata keçirmək cəhdi, başlanğıc fəaliyyət) işləyir (ru. wikipedia.org/wiki/искусственный_интеллект).

İnsan resurslarından istifadə və rekrutinq sahəsində də süni intellekt geniş istifadə olunur.

İnsan resurslarından istifadə və mütəxəssislərin muzzla tutulması işlərində Sİ-dən istifadənin üç üsulu vardır:

- namizədlərin xülasələrinə baxılması;
- onların kvalifikasiya səviyyələrinə uyğun olaraq rənqlərə ayrılması;
- vəzifələrin qarşılaşdırılması platformasında namizədlərin uğurlarının proqnozlaşdırılması.

Nəhayət, təkrarlanan kommunikativ vəzifələri avtomatlaşdırmağa imkan verən çat-botların yaradılmasında da Sİ işləyir.

Bir qayda olaraq, xülasəyə baxış prosesi xülasə verilənləri bazasında məlumatın axtarışı və təhlilini əhatə edir.

Pomato startapları xülasələrin yoxlanışı prosesini avtomatlaşdırmaq üçün maşın öyrədici alqoritmlərini yaradır. 2016-2017-ci illərdə tələbat malları kompaniyası Unilever bütün əməkdaşların başlanğıc səviyyəsini əks etdirmək üçün süni intellektdən istifadə etmişdi. Unilever süni intellekti neyrobiologiyaya əsaslanan oyunları, intervü yazılarını, şəxsi və nitq siqnallarını təhlil etməklə namizədlərin uğurlarını proqnozlaşdırmaq imkanına malik idi. Sİ əsasında yeni təhlil sistemi yaratmaqla bir il ərzində namizədlərin sayını 15000-dən 30000-ə yüksəltmək mümkün olmuşdu. Unilever ərizələrin emalına sərf olunan vaxtın müddətini də 4 aydan 4 həftəyə qədər azaltdı və beləliklə 50000 saatdan çox vaxta qənaət etmiş oldu.

Sİ insanın musiqi yaratmaq bacarığını təqlid etmək xüsusiyyətinə malikdir. Devid Koup Emili Xayell adı ilə alqoritmik

kompüter musiqisi sahəsində tanınmışdır. ABŞ-da Emili Xayell adı ilə patent qeydə alınmışdır.

AIVA (Artificial Intelligence Virtual Artist) alqoritmi isə klassik musiqi əsasında kinofilmlərə musiqi bəstələmək sahəsində ixtisaslaşmışdır. Bu ixtira şöhrət qazanmış və ilk virtual bəstəkar kimi professional musiqi assosiasiyası tərəfindən tanınmışdır.

Sİ tibbdə istifadə məqsədi ilə musiqi yarada bilər. Melomics stress – ağrını azaltmaq üçün kompüter musiqisindən istifadə edir.

Narrative Science kompaniyası kompüter yenilikləri, kommersiya hesabları, o cümlədən oyunların statistik verilənləri əsasında idman oyunlarına dair ingilis dilində ümumiləşdirici məlumatlar verir.

Son zamanlar avtomatlaşdırılmış jurnalistika anlayışı da yaranmış və tədricən yayılmaqdadır. Echobox kompaniyasının hazırladığı proqram təminatı ilə məqalə və məlumatlar düşünlümiş şəkildə böyük izləyicisi olan Facebook və Tvitter kimi sosial şəbəkələrdə qoyulur. Böyük həcmdə verilənlər bazasını təhlil etməklə müxtəlif auditoriyaların günün hansı vaxtlarında hansı məqalələrə diqqət yetirdiyi müəyyənləşdirilir. Sonra həmin məqalələr günün münasib saatlarında sosial şəbəkələrdə yerləşdirilir.

Sİ-nin gələcəkdə bədii əsərlər yarada bilməsi imkanları da var. 2016-cı ildə Yaponiyada Sİ vasitəsilə kiçik bir bədii əsər yazılmış və bu əsər mükafata layiq görülmüşdür.

Sİ avtomatlaşdırılmış onlayn-köməkçi kimi də istifadə olunur. Nümunə olaraq veb-səhifələrdəki çat-botları göstərmək olar. Bu müəssisələrə əməkdaşların muzzla götürülməsi və təlimatlandırılmasına çəkilən xərcləri azaltmağa imkan verir. Belə sistemlərdə əsas texnologiya dilin təbii emalıdır. Bir çox telekommunikasiya kompaniyaları öz əməkdaşlarının idarəsində evrestik axtarışdan istifadə edirlər. Məsələn, BT Group 20000

mühəndisin iş qrafikinin planlaşdırılması üçün evristik axtarış tətbiqindən istifadə edir.

6G peyk əlaqəsinin idarəsi üçün süni intellekt sistemlərindən istifadəyə böyük ümidlər bəslənir.

1990-cı illərdə təhsil və istirahətlə bağlı süni intellekt məkanlarının yaradılması işlərinə start verildi. Bu istiqamətdə robotlaşdırılmış ev heyvanları və müxtəlif oyuncaqların kütləvi istehsalına başlandı. Həmin oyuncaqlar müəyyən intellektual funksiyaları yerinə yetirməyə qabil hesab olunurlar. Mattel tipli kompaniyalar üç yaşına qədər uşaqlar üçün Sİ sistemlərindən yararlanan oyuncaqlar istehsal etməyə başladılar. Sİ-nin patentləşdirilmiş nitqin tanınması sistemlərindən istifadə etməklə bu oyuncaqlar qavramaq, intellektual cavablar və suallar verməyə qabildirlər. Bu istiqamətdə işlər getdikcə daha çox inkişaf etdirilir. Bunun nəticəsi olaraq müasir kompüter oyunları şəbəkəsi yaranmağa və getdikcə populyarlaşmağa başladı. Neyron şəbəkələrin öyrənilməsi proqramlarından bəhrələnməklə bu sistemlərdə oyunların səviyyələr üzrə təşkilinin əsas prinsipləri müəyyənləşdirilmiş və insan müdaxiləsi olmadan yeni oyun sistemləri yaradılmağa başlanmışdır.

Sİ-nin müxtəlif vasitələri təhlükəsizliyin təmini, nitqin və mətnin tanınması, verilənlərin intellektual təhlili, elektron poçtda spamların filtrlənməsi və s. sahələrdə geniş istifadə olunur. Eləcə də jestlərin tanınması (jest dilinin – işarəsinin maşın tərəfindən tanınması), səsini fərdi tanınması, səsini qlobal tanınması (səs-küylü otaqda çoxlu adamların içindən), emosiyaların və qeyri-verbal siqnalların interpretasiyası, üzün (insan üzünün) tanınması ilə bağlı yeni tətbiqlər işlənib hazırlanmışdır. Yeni tətbiqlərə – robotlaşdırılmış naviqasiya, maneələrin dəf olunması və obyektlərin tanınması da aid edilə bilər.

Göstərilən araşdırmaları və bu istiqamətdə respublikamızda aparılan digər tədqiqat işlərini süni intellekt məsələlərinə aid etmək olar.

Süni intellektin təcrübə sınaq məlumatları ilə vəhdəti nəticəsində yeni metal şüşə yaradılması prosesi 200 dəfə sürətlənmiş oldu. Bu materialın şüşə təbiətli olması onu müasir polad konstruksiyalardan daha möhkəm, yüngül və korroziyaya davamlılığı ilə nəticələnmişdir. Bu layihənin yaratıcıları Sİ-dən bəhrələnməklə bir il ərzində 20 000 mümkün variantdan birini seçməli olmuşlar.

Süni intellektin köməyi ilə müasir qırıcı təyyarələrin döyüş qabiliyyətlərinin təkmilləşdirilməsi sahəsində də uğurlu işlər aparılır. Bu layihənin reallaşmasında neyroşəbəkə alqoritmlərindən istifadə olunmuşdur.

Süni intellekt neyrofiziologiya, epistemologiya və koqnitiv psixologiya ilə birlikdə koqnitologiya adlanan ümumi elm sahəsini əmələ gətirir. Süni intellektin yaranmasında və formalaşmasında fəlsəfə mühüm rol oynayır. Eləcə də, fəlsəfə çərçivəsində bilik haqqında elm olan epistemologiya da süni intellekt ilə sıx bağlıdır. Bu problemlərlə məşğul olan filosoflar Sİ mühəndislərinin məşğul olduğu sahələrə yaxın məsələlərin həlli ilə məşğul olurlar. Bu biliyin və informasiyanın optimal təqdimi və istifadəsi ilə bağlı məsələlərdir

Kompüter elmlərində süni intellekt məsələlərinə ekspert sistemləri və bilik bazalarının layihələndirilməsi mövqeyindən baxılır. Bilik bazası dedikdə, informasiyanın dərk olunmuş emalı və məntiqi nəticəsinin alınmasına imkan verən verilənlər və nəticə qaydalarının məcmusu başa düşülür. Bütövlükdə, kompüter elmlərində süni intellekt problemlərinin tədqiqi intellektual informasiya sistemlərinin yaradılması, inkişafı və istismarına istiqamətlənmişdir. Belə sistemlərin istifadəçilərinin və yaratıcılarının hazırlanması məsələləri isə informasiya texnologiyaları sahəsində mütəxəssislər tərəfindən yerinə yetirilir.

Koqnitiv modelləşdirmə metodologiyası müəyyənləşdirilməsi çətin olan situasiyaların təhlili və qərarın qəbulu üçün nəzərdə tutulmuşdur.

Hazırda koqnitiv modelləşdirmə metodologiyası situasiyanın təhlili və modelləşdirilməsi aparatının təkmilləşdirilməsi istiqamətində inkişaf edir. Burada situasiyanın inkişafı proqnozunun modeli və əks məsələlərin həlli metodları təklif olunur.

“Süni təfəkkürün yaradılması haqqında elm” filosofların diqqətini cəlb etməyə bilməzdi. İlk intellektual sistemlərinin yaranması ilə insan və bilik, eləcə də dünya, aləm kimi fundamental məsələlər ön plana keçdi.

Süni intellektin yaradılmasının fəlsəfi problemlərini, şərti olaraq, iki qrupa ayırmaq olar: süni intellektin yaradılmasından əvvəl və sonra. Birinci qrup aşağıdakı suallara cavab verməyə çalışır: “süni intellekt nədir, onu yaratmaq mümkündürmü, əgər mümkündürsə, necə yaratmalı?”

İkinci qrup (süni intellektin etikası) bu suala cavab axtarır: “süni intellektin yaradılmasının insan cəmiyyəti, bəşəriyyət üçün nəticələri nə ola bilər?” Transhumanizm cəmiyyəti hesab edir ki, Sİ-nin yaradılması bəşəriyyətin ən mühüm vəzifələrindən biridir. Bu proqram insan təfəkkürünün, aqlının modeli deyil, əsl təfəkkür olmalıdır, insan təfəkkürü mənasında təfəkkür olmalıdır.

Baş a düşmək lazımdır ki, “təmiz süni düşüncə, ağıl, “metaağıl”, qavrayan, real problemləri həll edə bilən, eyni zamanda insan üçün səciyyəvi olan, onun fərdi mövcudluğunu şərtləndirən emosiyalardan mərhum “ağıl” mümkündürmü?

Zəif Sİ tərəfdarları hesab edirlər ki, süni intellekt bu və ya digər məsələni həll edə bilən alət rolunu oynaya bilər. Bunun üçün insanın bütün qavrama sistemi spektri tam şəkildə tələb olunmur.

Rus neyrobioloqu və filosofu N.Aseyev hesab edir ki, “güclü Sİ” heç vaxt yaradıla bilməz. Oyun alqoritmləri, tanıyan, müxtəlif məsələləri həll edən alqoritmlər isə həmişə təkmilləşəcək, yeniləri yaradılacaq. Hazırda gələcəkdə yaradılacaq insanfövqü Sİ-lərin global risk problemləri araşdırılır. Gələcə-

yin Sİ robotları insanlarla dost münasibətdə proqramlaşdırılmasa, bunun böyük fəlakətləri ola bilər [ru. wikipedia.org/wiki/искусственный_интеллект].

Amerikalı sahibkar İlon Mask müsahibələrinin birində süni intellektin böyük potensialını qeyd etməklə yanaşı onun insanlıq üçün həm də təhlükə ola bilmək ehtimalını göstərir. Məlum olduğu kimi, nüvə enerjisi də həm dinc məqsədlərlə, həm də bəşəriyyət üçün böyük faciələrə səbəb ola biləcək atom bombası kimi istifadə oluna bilər. Ümid edək ki, alimlər süni intellektin potensial imkanlarından faydalı şəkildə insanların rifahı naminə istifadə edəcəklər.

İnsan bədəninin sinir hüceyrəsi olan neyronlar sinir impulslarını hərəkətə gətirən sinir hüceyrələridir. Həmin hüceyrələr insan orqanında məlumat və stimulları emal edən sistemin əsas hissələridir.

Neyronların üç əsas hissəsi var: dendritlər (məlumatların qəbulunun baş verdiyi yer, neyronun reseptor hissəsidir); hüceyrə orqanı (məlumatların inteqrasiyasına cavabdehdir) və axons (sinir impulslarını bir neyronun digərinə və ya bir neyronun digərinə bir bez və ya əzələ lifinə nəql edir).

Neyronların şaxələnmiş ekstremətləri var (dendritlərin bir hissəsi).

Sinir impulslarının neyronlar arasında və ya bir neyronun digərinə hüceyrə tipinə keçməsi fiziki-kimyəvi reaksiya ilə baş verir.

Sinaps iki neyron arasındakı əlaqə (rabitə) yeridir. Sinapslarda sinir impulsunun ötürülməsi hadisəsi baş verir. Bunlar neyronlar tərəfindən istehsal olunan və sinaptik veziküllərdə (aksonların uclarında olan cəblər) saxlanılan biomolekullardır (kimyəvi maddələr). Yetkin bir insanın bədənində təxminən 85 milyard neyron var. Sinir sistemini öyrənən biologiyanın bu qoluna neyroşəbəkə deyilir. Sinir sisteminin pozğunluqlarını araşdıran tibb sahəsi nevrologiya kimi tanınır

“ICTnews” “vesti.ru” saytına istinadən yazır ki, “Google”-un mühəndisləri Martin Abadi və Devid Andersen tərəfindən keçirilmiş sınaqda üç neyroşəbəkə iştirak edib: “Alisa”, “Bob” və “Yeva”. Hər biri müəyyən tapşırıq yerinə yetirirdi: “Alisa” tərəfindən “Bob”a göndərilən məxfi mesajları yalnız sonuncu oxuya bilirdi, “Yeva” isə həmin mesajın tərkibini ələ keçirməli və de-şifrə etməli idi [ru. wikipedia.org /wiki/ искусственный_интеллект].

Neyroşəbəkələr öz işinə sadə mətnin mübadiləsi ilə başlayıblar: “Alisa” həmin mətni mənasız simvol yığımına konversiya edir, “Bob” isə şifrləmə açarının köməyi ilə de-şifrə etməyə çalışırdı. İlk öncə “Yeva” istənilən gizli ünsiyyətin qarşısını alırdı, ancaq 15 min cəhddən sonra “Alisa” öz kriptografiya metodunu hazırlamağı, “Bob” isə onu deşifrə etməyi öyrəndilər. Qeyd edək ki, hətta mütəxəssislər də “Alisa”nın ixtira etdiyi şifrləmə metodundan xəbərsiz olublar. Həmin texnologiyanın tətbiq sahəsi praktik cəhətdən olduqca məhduddur.

Beyin və davranış arasındakı əlaqələri araşdırmaq hipotetik, deduktiv, analitik-induktiv metodlar əsasında) aparılır.

2000-ci illərdə Sİ robotlarının yaradılmasına maraq yənidən artmağa başladı. Sİ kosmik sənayedə və məişətdə geniş tətbiq və istifadə olunmağa başladı. Ağıllı ev sistemi yaradılmağa, qabaqcıl məişət avadanlığı tətbiq olunmağa başladı. Kismet və Nomad robotları Antraktidani tədqiq etməyə başladı.

2008-ci ildən texnoloji sinqulyarlıq erası başlanır və ekspertlərin fikrincə, bu eranın 2030-cu ildə ən yüksək nöqtəyə çatacağı proqnozlaşdırılır. İnsanın hesablama texnikası ilə inteqrasiyası başlanır, insan beyninin imkanları genişlənir, biotexnologiya yaranır.

Maşın tədrisi – özüöyrənən alqoritmlər vasitəsilə süni intellektin inkişafı prinsipidir. İnsanın belə yanaşmada iştirakı maşının yaddaşına informasiyanın yüklənməsi və məsələnin

məqsədinin qoyuluşu ilə məhdudlaşır. Maşın tədrisi ilə bağlı bir neçə metodika mövcuddur:

- müəllimlə tədris – müəllim konkret məsələ qoyur, fərziyyəni yoxlamaq və ya qanunauyğunluğu təsdiqləmək istəyir;
- müəllimsiz tədris – verilənlərin intellektual emalının nəticəsi məlum deyil, kompüter müstəqil olaraq qanunauyğunluğu tapır. İnsan kimi düşünməyi öyrənir;
- dərin tədris – qarışıq üsuldur, əsas fərqi böyük massivlərin emalı və neyroşəbəkələrdən istifadədir. Dərin tədrislə bağlı bir daha qeyd etmək istərdik ki, neyroşəbəkə – canlı orqanizmin əsəb sinirlərinin quruluşu və funksiyalarını imitasiya edən riyazi modeldir. Müvafiq olaraq idealda – bu müstəqil öyrənən sistemdir. Bu prinsipi texnoloji müstəviyə keçirsək, onda deyə bilərik ki, neyroşəbəkə irihəcmli layihədə hər hansı bir məsələnin icrasını həyata keçirən prosessorlar çoxluğudur. Başqa sözlə, superkompüter çoxlu sayda adi kompüterlər şəbəkəsidir.

Dərin tədris Sİ-nin ayrıca prinsiplərinə aid edilir. Bu metod nəhəng informasiya massivlərində qanunauyğunluqların aşkar edilməsi məqsədi ilə istifadə olunur. İnsanın yerinə yetirə bilməyəcəyi işlər üçün kompüter təkmilləşdirilmiş metodlardan istifadə edir.

Koqnitiv hesablama – Sİ-nin istiqamətlərindən biri olub, insanlararası əlaqələrə bənzər insan və kompüter arasında qarşılıqlı əlaqə proseslərini tətbiq edir. Süni intellekt texnologiyasının məqsədi yüksək səviyyəli insan fəaliyyətini – nitqi, obrazlı və analitik təfəkkürü tam şəkildə təqlid etməkdir.

Sİ-nin kompüter görünüşü istiqaməti qrafik və video-təsvirələrin tanınması üçün istifadə olunur. Bu gün maşın intellekti qrafik verilənləri emal və təhlil edə bilər, ətraf mühitə (vəziyyətə) müvafiq olaraq informasiyanı şərh edə bilər.

Kompüterlər artıq insan nitqini qavramaq, təhlil etmək və doğurmaq, sintez etmək bacarığına malikdirlər. Biz artıq səs əməlləri ilə proqramları, kompüterləri və s. idarə edə bilirik.

Bundan başqa, süni intellekti verilənlərin interaktiv emalının əsası sayılan güclü qrafik prosessor olmadan təsəvvür etmək mümkün deyil. Sİ-nin müxtəlif proqram və qurğulara inteqrasiyası üçün tətbiqlərin proqram interfeysi olan xüsusi texnologiya zəruridir. Həmin texnologiyadan istifadə etməklə süni intellekt texnologiyasını asanlıqla istənilən kompüter sistemində əlavə etmək olar.

Süni intellekt adi proqram komplekslərini intellektuallaşdıraraq tədricən insan fəaliyyətinin bütün sahələrinə nüfuz edir:

Onlayn-mağazalarda pərakəndə satışla bağlı reklamlar çoxlarına məlumdur. Onların köməyi ilə istifadəçilərin maraqlarına uyğun olaraq mal və xidmətlər təklif edilir. Məsələn, Siz çimərlik paltarları internet-mağazasına gedir, hansısa modelləri gözdən keçirir, xüsusiyyətləri ilə tanış ola bilərsiniz. Mağazanı tərk etdikdən sonra çimərlik paltarlarının reklamlarını başqa saytlarda da görə bilərsiniz. İnternet – mağazalarda “oxşar mallar” bloku eyni prinsiplər əsasında işləyir. Analitik sistemlər istifadəçilərin davranış metrikasını öyrənir, onun alıcı istəklərini müəyyənləşdirir və onların rəyinə görə relevant təkliflər təqdim edirlər.

İntellektual maşınlar Barak Obamaya ikinci prezident seçkilərini udmağa kömək etmişdir. Kompüter texnologiyası sahəsində mütəxəssislər intellektual maşınların imkanlarından istifadə edərək Obamanın çıxışı üçün ən yaxşı auditoriya və uğurlu günləri müəyyənləşdirmişdilər. Mütəxəssislərin fikrincə, bu seçim 10-12% üstünlük qazanmaq imkanı verir.

Süni intellekt müxtəlif istehsal sahələrindən verilənləri təhlil edə və avadanlıqlara düşən artıq yüklənməni tənzimləyə

bilər. Bundan başqa, intellektual maşınlar sənayenin müxtəlif sahələrində tələbatı proqnozlaşdırmaq üçün istifadə oluna bilər.

Süni intellekt oyun yaradıcıları tərəfindən fəal olaraq istifadə olunur. Ağıllı maşınlar, robot texnikası bir çox dövrlərin təhsil proseslərinə təcridən tətbiq olunmağa başlayır.

Məlum olduğu kimi, indiki inkişaf mərhələsində süni intellektin imkanları hədudsuz deyil. Əsas çətinlikləri sadalayaraq:

- maşın tədrisi yalnız verilənlər massivində əsasında mümkündür. Bu o deməkdir ki, informasiyadakı istənilən qeyri-dəqiqlik son nəticədə qabarıq təzahürünü tapır;

- intellektual sistemlər konkret fəaliyyət növləri ilə məhdudlaşır. Ağıllı maşınlar vergi ödənişlərindəki fərqləri aşkar etməyə tənzimləndiyi üçün bank sferasında olan saxtakarlıqları müəyyənləşdirə bilmir. Hələ ki, dar sahəyə istiqamətlənmiş proqramlardan istifadə olunur, insanın yerinə yetirə biləcəyi çoxməsələli problemləri həll edə biləcək proqramların tərtibi yaxın gələcəkdə gözlənilir. İntellektual maşınlar avtonom deyillər. Onların fəaliyyəti üçün bütöv mütəxəssislər komandası, eləcə də böyük resurslar lazım gəlir.

Süni intellekt sahəsində ən əhəmiyyətli istiqamətlərdən biri təbii dilin emalı hesab olunur. Bu istiqamət çərçivəsində təbii mətnlərin təhlili, qavranma imkanları və sintezi tədqiq olunur. Tədqiqatın nəticəsi olaraq internetdə rast gəlinən istənilən mətnin emalı və ondan müəyyən informasiyanın əldə olunması nəzərdə tutulur. Təbii mətnin emalı sistemləri öz tətbiqini maşın tərcüməsi və məlumat axtarışı sistemlərində tapır.

Qeyd etmək lazımdır ki, süni intellektin tədqiqat obyektini hesab etdiyimiz mətnin formal təhlili, təbii mətnin emalı, mətnin səsə (nitqə), səsin (nitqin) mətnə çevrilməsi, maşın tərcüməsi, məlumatların avtomatik (formal əlamətlərinə görə) axtarışı, tarixi abidələrin və klassiklərin dilinin riyazi-statistik metodlarla tədqiqi, statistik lüğətlərin (tezlik, əks) tərtibi, elektron lüğətlərin hazırlanması sahəsində respublikamızda xeyli iş

görülmüşdür [Azərbaycan dilinin tezlik lüğəti, 2010; Azərbaycan qəzet dilinin tezlik lüğəti, 2004; Məhəmməd Füzulinin nəzm əsərlərinin əlifba-tezlik lüğəti, 2004; “Kitabi-Dədə Qorqud”un statistik təhlili, 1999; Azərbaycan dilinin əks əlifba lüğəti, 2004; Махмудов, 1991; 2001; Mahmudov, 1981; 2002; 2013; 2018; 2017].

Yuxarıda sadalanan məsələlər barədə monoqrafiyanın sonrakı fəsillərində daha geniş və ətraflı məlumat veriləcəkdir.

II FƏSİL

DİLİN RİYAZI-STATİSTİK METODLARLA TƏDQIQI, MÜXTƏLİF ÇEŞİDLİ STATİSTİK LÜĞƏTLƏRİN TƏRTİBİ MƏSƏLƏLƏRİ VƏ SÜNİ İNTELLEKT

Azərbaycan dilinin riyazi-statistik metodlar və yeni texnoloji vasitələrlə öyrənilməsi məsələləri. Azərbaycan dilinin riyazi-statistik metodlarla və yeni texnoloji vasitələrlə öyrənilməsi məsələlərinə ötən əsrin ikinci yarısından başlanılmışdır. Həmin vaxtlarda dünyanın bir çox aparıcı mərkəzlərində dilin riyazi-statistik metodlarla elektron hesablayıcı maşınlarının köməyi ilə öyrənilməsi sahəsində maraqlı tədqiqatlar aparılırdı. Keçmiş SSRİ-nin bir çox elmi mərkəzlərində, o cümlədən Moskva, Leningrad (indiki Sankt-Peterburq), Kiyev, Minsk və s. şəhərlərində də bu qəbildən olan tədqiqatlara böyük maraq var idi. Azərbaycanda dilin yeni metod və texnologiyalarla tədqiqi sahəsində çalışan ilk araşdırıcılar da adı çəkilən həmin elm mərkəzlərində aspirantura təhsili almış və ya elmi təcrübə keçmiş mütəxəssislər idi.

Müxtəlif elm sahələrinin qovuşub yeni bir elmi istiqamət yaratması və bir elmin əsasında yeni elm sahələrinin meydana çıxması müasir dövrdə elmlərin inkişaf prosesinin aparıcı cəhətlərindəndir. XX əsrin ikinci yarısında riyazi metodların bir çox elmlərə tətbiqi maraqlı elmi nəticələrə, yeni tədqiqat obyekti və istiqamətlərin yaranmasına aparıb çıxartdı. Bir çox tədqiqatçılar hesab edirdilər ki, hər bir elm yalnız riyazi metodların tətbiqi nəticəsində uğur əldə edə bilər, inkişafının ən yüksək dərəcəsinə çata bilər. Maraqlı orası idi ki, dəqiq elmlər, o cümlədən riyaziyyat humanitar elmlərə də uğurla tətbiq olunurdu. Riyaziyyatın tətbiqi bu və ya digər elm sahəsinə fərqli, bizə

qeyri-adi görünən, yeni baxış bucağı altında baxmaq imkanı yaradır. Nə qədər qərribə görünse də, riyaziyyatın uğurla tətbiq olunduğu sahələrdən biri də dilçilikdir. Bunun bir səbəbi də riyaziyyat kimi dilçiliyin də dəqiq elmlər kateqoriyasına daxil edilməsi idi. Tədqiqatçıların bir çoxu türk dillərinin, o cümlədən Azərbaycan dilinin quruluşunda, qrammatik və leksik şəkilçilərin söz kökünə birləşmə mexanizmində bir riyazi nizam, qanunauyğunluq olduğunu qeyd edirlər. Mütəxəssislər humanitar elmlər sırasında ən çox dilçiliyin riyaziyyata yaxın olduğunu göstərir və hətta dilçiliyin dəqiq elmlər kateqoriyasına daxil edilməsi təklifini irəli sürürlər. Sonralar türk dillərinin formal morfoloji analizi (təhlili) və sintezi sistemi ilə bağlı tədqiqatlar bu fərziyyənin doğruluğunu bir daha təsdiq etdi. Həqiqətən də, türk dilləri ailəsində, o cümlədən bu ailəyə aid olan Azərbaycan dilində sözdəyişdirici (qrammatik) və sözdüzəldici (leksik) şəkilçilər söz kökünə (eləcə də əsasa) müəyyən nizam və qanunauyğunluqla, riyazi dəqiqliklə birləşir. Məhz bu aparıcı xüsusiyyət təhlil və sintez prosesini formal baxımdan təsvir etmək, alqoritmləşdirmək və proqramlaşdırmaq imkanı yaradır.

Riyazi metodların dilə tətbiqinin nəticəsi olaraq sonralar riyazi dilçilik, tətbiqi dilçilik, hesablama dilçiliyi, statistik dilçilik (statistik lüğətçilik), kompüter dilçiliyi və s. istiqamətlər ayrıca bir elm sahəsi kimi inkişaf edib formalaşmağa başladı. Yeni elm sahələrinin adında olan bu müxtəlifliklərə təbii baxmaq lazımdır. Bir neçə elmin qovuşma nöqtəsində yaranan bu elm sahələri mütəxəssisin hansı ixtisas sahibi olmasından asılı olaraq müxtəlif cür izah olunub əsaslandırıla bilər. Digər tərəfdən, bu həm də qeyd olunan elm sahələrinin tam formalaşmaması, tədqiqat obyektinin və predmetinin dəqiq təyin olunmaması ilə izah oluna bilər. Azərbaycan dilçiliyində müxtəlif elm sahələrinin qovuşması və ya eyni elm sahəsinin parçalanması nəticəsində yeni elm sahələrinin yaranması, həm də dünya elmində baş verən analogi hadisələrin təsirinin qanunauyğun nəti-

cəsi hesab oluna bilər. Dediklərimizə misal olaraq akad. A.A.Axundovun 1963-cü ildə ADU-nun (indiki BDU) Elmi Əsərlərində çap olunmuş “Azərbaycan dili saiflərinin statistik təhlili təcrübəsindən” məqaləsini göstərmək olar [Axundov, 1963]. Məqalə müəllifi Azərbaycan dilçiliyində ilk dəfə olaraq statistik metodlarla Azərbaycan dili saiflərinin təhlili sahəsində apardığı tədqiqatın nəticələri ilə bağlı oxucuları məlumatlandırır. Yəni bu, tədqiqatçının Azərbaycan dilinin saiflər sisteminin tədqiqində statistik metodların tətbiqinə bir cəhdi, saifləri indiyə qədər vərdiş etmədiyimiz başqa üsulla, yeni baxış bucağından araşdırmaq cəhdi idi. Bu yanaşma uğursuz da alına bilərdi. Müəllif bu sahədə aparılan tədqiqatlarda yeni yanaşma üsulu təklif edir, bu məsələlərlə bağlı fikirlərini həmkarları ilə bölüşür, onların bu yanaşma (statistik təhlil) ilə bağlı mülahizələri ilə tanış olmaq istəyirdi. Sonralar A.A.Axundov bu istiqamətdə axtarışlarını daha da genişləndirdi və “Riyazi dilçilik” kitabı ilə dilçiliyin bu yeni sahəsinə maraq oyatmağa səy göstərdi [Axundov, 1979]. Bu kitab haqqında məqalədə ətraflı məlumat veriləcəkdir. İndi isə xronologiya ilə daha bir neçə tədqiqat əsəri barədə məlumat vermək istərdik.

Keçən əsrin 70-ci illərini Azərbaycan dilçiliyi üçün uğurlu hesab etmək olar. Həmin illərdə Moskvada V.Y.Pines “Avtomatik lüğət problemi ilə əlaqədar Azərbaycan feil formalalarının quruluşunun modelləşdirilməsi” [Пинес, 1970] və K.Vəliyeva “Azərbaycan sözlərinin sintezinin formal təsviri” [Велиева, 1971] mövzularında namizədlik dissertasiyaları müdafiə edib Bakıya qayıtdılar və Akademiyanın Nəsimi adına Dilçilik İnstitutunda işləməyə başladılar. Bu istiqamətdə aparılacaq elmi tədqiqatları mərkəzləşdirmək və genişləndirmək məqsədi ilə kiçik yaradıcı qrup yaradıldı. Sonralar M.Ə.Mahmudov (1976-cı il) və C.Ə.Rəhmanov (1978-ci il) da bu qrupa cəlb olundu. Həmin dövrdə Dilçilik İnstitutu tarixində ilk dəfə olaraq Lvov Poliqrafiya İnstitutu və “Kommunist” nəşriyyatı üçün müqavilə

əsasında Azərbaycan dilində sözlərin avtomatik olaraq sətirdən-sətərə keçirilməsi sistemi işlənib hazırlandı. Dilin tətbiqi məsələlərinin istehsalatda istifadəsi öz səmərəsini verməyə başlayırdı. Lakin, bu yaradıcı qrupun köməyə və dəstəyə ehtiyacı var idi. Nəzəriyyənin praktikaya tətbiqi bu sahəyə diqqəti və marağı artırmağa bilirdi. Riyazi dilçilik qrupu kənar təşkilatlarla müqavilə bağlayaraq praktik fəaliyyət göstərməyə başladı. Bunlar, həm də Dilçilik İnstitutunun tarixində ilk dəfə müqavilə əsasında görülmüş işlər idi. Sübut olundu ki, riyazi və tətbiqi dilçilik sırf nəzəri istiqamət deyil. Riyazi dilçilik (o vaxt Dilçilik İnstitutunda yaradılmış yaradıcı qrup belə adlanırdı) həm də nəzəriyyənin praktikaya, istehsalata tətbiqi deməkdir. Nəzəri müddəaların praktikaya tətbiqi nəinki elmi əhəmiyyət daşıyır, həm də müəyyən iqtisadi səmərə verə bilər, yeni maliyyə imkanları yarada bilər. Riyaziyyatın dilçiliyə tətbiqi nəticəsində formalaşmış yeni sahənin bu keyfiyyətləri, praktik tətbiq imkanları və potensialı üzə çıxdı. Bu yeni keyfiyyət dilçiliyə riyazi metodların tətbiqi məsələlərinə marağı bir qədər də artırdı.

1977-ci ildə Azərbaycan SSR Dövlət Plan Komitəsinin Azərbaycan Elmi-Tədqiqat Elmi-Texniki Məlumat və Texniki-İqtisadi Tədqiqatlar İnstitutunda V.Y.Pines və M.Ə.Mahmudovun “Elektron hesablayıcı maşınlar dilçilikdə” adlı kitabçası çapdan çıxdı [Pines, Mahmudov, 1977]. Kitabça “Sahələrarası” seriya ilə icmal məlumatı kimi çap olunmuşdu. Məqsəd yeni sahə və istiqamətlərlə elmi ictimaiyyəti operativ olaraq yaxından tanış etmək idi. Kitabçanın annotasiya hissəsində qeyd olunurdu ki, icmal məlumatı hesablayıcı texnikanın linqvistik tədqiqatlarda tətbiqinin müxtəlif cəhətlərini əhatə edir. Müəlliflərin fikrinə görə, elektron hesablayıcı maşınlar (EHM) dilçilikdə iki əsas istiqamətdə – dil tədqiqatlarında sırf mexaniki xarakter daşıyan yorucu və ağır işlərdə dilçinin əməyini qismən yüngülləşdirmək üçün tədqiqatın bəzi elementlərinin avtomatlaşdırılmasında və dil modellərinin avtomatlaşdırılması məsələlərində

istifadə olunur. Kitabçada həmçinin EHM-lərin türkologiyada, qismən də Azərbaycan dilçiliyində tətbiqi ilə bağlı məsələlərə xüsusi bölmə ayrılmışdı. Qeyd etmək lazımdır ki, kitabça Azərbaycan SSR Elmlər Akademiyası (indiki AMEA) “Kaspi” Elmi Mərkəzi və S.M.Kirov adına ADU (indiki BDU) Azərbaycan dilçiliyi kafedrası tərəfindən bəyənilmiş və çapa məsləhət görülmüşdü. Hər iki qurum belə bir məlumatın elmi ictimaiyyətə çatdırılmasını vacibliyini vurğulamış və icmal məlumatının çapını faydalı hesab etmişdi. Kitabça giriş, 4 bölmə və istifadə olunmuş ədəbiyyat siyahısından ibarətdir. Girişdə hesablayıcı texnikanın dilçilikdə istifadə olunmasının ümumi məsələləri haqqında məlumat verilir. Bölmələr aşağıdakılardır: Linqvistik tədqiqatların avtomatlaşdırılması, Dil modellərinin avtomatlaşdırılması, Məlumat axtarışının avtomatlaşdırılması və dilçilik, EHM türkologiyada. Ədəbiyyat siyahısında rus və ingilis dillərində çap olunmuş 37 əsərin adı göstərilmişdir. Təəssüf hissi ilə qeyd etmək lazımdır ki, siyahıda Azərbaycan dilində çap olunmuş əsər adı yoxdur. Elə kitabçanın məqsədi də diqqəti bu sahəyə yönəltməkdən ibarət idi.

Qeyd etmək lazımdır ki, adı çəkilən qurum bu sahədəki xeyirxah əməllərinin davamı olaraq 1980-ci ildə “Energetika və avtomatika” seriyası ilə daha bir icmal səciyyəli informasiya vərəqi çap etmişdi. M.Ə.Mahmudov və V.Y.Pinesin imzaları ilə çap olunmuş bu məlumat vərəqi “Maşın tərcüməsi və avtomatlaşdırılmış məlumat sistemlərində Azərbaycan mətnlərinin morfoloji təhlil alqoritmi” adlanır. Yazıda ilk dəfə olaraq Azərbaycan mətnlərinin morfoloji təhlil alqoritmi və onun blok sxemi elmi ictimaiyyətə təqdim olunmuşdu [Mahmudov, Pines, 1980]. O vaxtlar artıq Azərbaycan dilçiləri dünya şöhrətli alim, prof. R.Q.Piotrovskinin rəhbərlik etdiyi Ümumittifaq “Статистика речи” qrupu ilə əməkdaşlıq etməyə başlamışdılar. Həmin vaxtlardan başlayaraq yeni texnologiyaların Azərbaycan dilinə tətbiqində kəmiyyət və keyfiyyət baxımından inkişaf dövrünün

başlanğıcı qoyuldu. Bu istiqamətin təşəkkülü və inkişafında akad. A.Axundovun 1979-cu ildə çapdan çıxmış “Riyazi dilçilik” kitabının əhəmiyyəti də az deyildi [Axundov, 1979]. Məlum olduğu kimi, 70-ci illərdə artıq keçmiş SSRİ-nin bir çox ali məktəblərində riyazi dilçilik fənni tədris olunurdu. Akademiyanın Nəsimi adına Dilçilik İnstitutunda Riyazi dilçilik qrupu fəaliyyət göstərməyə başlamışdı. Qrup tərəfindən Azərbaycan dilinin tezlik və əks lüğətləri tərtib edilmiş, elektron hesablama maşınlarının dilə tətbiqi məsələlərini əhatə edən bir çox materiallar çapa təqdim olunmuşdu. Lakin, açıqını demək lazımdır ki, respublikada, eləcə də Elmlər Akademiyasında bu sahəyə lazımı səviyyədə diqqət yetirilmir, bir çox məsələlərə barmaqarası baxılırdı. Ölkəmizdə bu istiqamətdə aparılan elmi-tədqiqatların həcmi, eləcə də mütəxəssis hazırlığı işləri heç də qənaətbəxş deyildi. A.Axundovun “Riyazi dilçilik” kitabının belə bir vaxtda çapdan çıxması bu sahəyə diqqətin artırılması baxımından əhəmiyyətli idi. Dərs vəsaiti kimi nəzərdə tutulmuş bu kitab riyazi dilçilik təlimlərinin yaranması, inkişafı və dilin tədqiqində riyazi üsulların tətbiqi məsələlərini əhatə edir.

Kitab iki bölmədən ibarətdir. Xülasələr bölməsində müəllif riyazi metodların dilə tətbiqi ilə bağlı ölkəmizdə və xaricdə aparılan tədqiqat işləri, eləcə də Azərbaycan dilində sait fonemlərin statistik təhlili sahəsində öz tədqiqatları barədə məlumat vermişdir. A.Axundov Azərbaycan dilində saitlərin meydanaçixma ehtimalı üzrə ciddi qanunauyğunluqlar olduğunu cədvəl və qrafiklər şəklində sübut edir. Müəllif belə bir nəticəyə gəlir ki, Azərbaycan dili saitləri üçün ən tipik mövqe söz ortasıdır. Dilimizdə söz əvvəli və söz ortası üçün arxa sıra, açıq, dodaqlanmayan və qısa saitlər daha səciyyəvidir. Azərbaycan dilində söz sonu üçün isə ön sıra, qapalı, dodaqlanmayan və qısa saitlər daha səciyyəvidir.

Həmin bölmədə Azərbaycan saitlərinin işlənmə tezliyinə görə sırası da verilmişdir. Kitabda riyazi dilçilik təlimləri ilə

bağlı olan təhlil üsulları, xüsusən Amerika deskriptiv dilçilik məktəbində geniş yayılmış distributiv təhlilin mahiyyəti barədə məlumat verilmişdir.

Kitabın təfriqələr bölməsində müəllif riyazi dilçiliyin əsas məsələləri barədə daha aydın təsəvvür yaratmaq üçün riyaziyyat, kibernetika, riyazi dilçilik və s. əsərlərdən hissələr vermişdir. Burada Avropa dilçilik məktəblərində, xüsusən rus dilçiliyində riyazi dilçiliklə məşğul olan görkəmli alimlərin araşdırmaları barədə qısa məlumat verilir. Eləcə də, bu bölmədə bir çox anlayış və terminlər sadə və anlaşıqlı dildə izah olunur. Həmin terminlər riyazi dilçiliklə bağlı tədqiqat aparan gənc mütəxəssislər üçün olduqca əhəmiyyətlidir və onların bu sahədəki araşdırmalarında uğurla istifadə olunmaqdadır.

Yekun olaraq demək olar ki, akad. A.A.Axundovun “Riyazi dilçilik: xülasələr və təfriqələr” kitabı dilçiliyə riyazi metodların tətbiqi, dil tədqiqatlarında yeni texnologiyaların istifadəsi məsələlərinə marağı artırdı. Elə kitabı yazmaqda müəllifin məqsədi də bu sahəyə diqqəti artırmaq, gənc tədqiqatçıları bu istiqamətə yönəltmək idi.

80-ci illərdə kompüterlərin dilə tətbiqi ilə bağlı bir çox araşdırmalar aparıldı, dissertasiya işləri müdafiə olundu. 1982-ci ildə M.Ə.Mahmudov “Türk dillərində söz-formaların formal morfoloji təhlili sisteminin işlənilib hazırlanması” mövzusunda namizədlik dissertasiyasını uğurla müdafiə etdi. Sonrakı illərdə M.H.Məmmədova (1984), C.Ə.Rəhmanov (1988) da hesablayıcı texnikanın imkanlarından yararlanmaqla namizədlik dissertasiyaları (fəlsəfə doktorluğu dissertasiyası) müdafiə etdilər [Мамедова, 1984; Рахманов, 1988].

80-ci illərdən sonra bu sahədə çoxlu namizədlik və doktorluq dissertasiyaları müdafiə olunmuşdur ki, onlar da ədəbiyyat siyahısında qeyd olunmuşdur. Onların hər biri haqqında geniş məlumat vermək lüzumsuz hesab olunmuşdur.

Məlum olduğu kimi, bu istiqamətdə olan araşdırmalar əvvəllər daha çox müxtəlif elm sahələrini əhatə edən tezlik lüğətlərinin tərtibi ətrafında mərkəzləşmişdi. Dilçiliyin, daha doğrusu, lüğətçiliyin yeni bir sahəsi – statistik lüğətçilik yaranmışdı. Statistik metodların dilçiliyə tətbiqi ilə ərsəyə gəlmiş tezlik lüğətlərinin tətbiq dairələri geniş idi. Maşın tərcüməsi sistemlərinin əsasında duran avtomatik lüğətlərin bazasının formalaşmasında da tezlik lüğətləri olmadan keçinmək mümkün deyil. Eləcə də, tədrisdə, lüğət-minimum, dərs vəsaitləri, danışiq kitabçalarının hazırlanması işlərində də tezlik lüğətləri əvəzsiz mənbədir.

Azərbaycan dilçiliyində də statistik lüğətlərin hazırlanması sahəsində zəngin təcrübə toplanmışdır. 70-ci illərdə Azərbaycan Elmlər Akademiyası Nəsimi adına Dilçilik İnstitutunda fəaliyyət göstərən Riyazi dilçilik qrupu ilk dəfə olaraq dilçitədqiqatçı fəaliyyətinin avtomatlaşdırılması sahəsində təcrübələr aparmış, C.Cabbarlının əsərləri üzrə söz-formaların tezlik, əlifba-tezlik siyahılarını hazırlamışdır. Çox təəssüf ki, həmin tədqiqatın çapı barədə heç bir məlumat yoxdur. Sonrakı illərdə qrup Azərbaycan qəzet dilinin tezlik lüğətinin tərtibi ilə məşğul olmuşdur (tərtibçilər: K.Vəliyeva, M.Mahmudov, V.Pines). Azərbaycan qəzet dilinin tezlik lüğəti üçün seçmələrin əsasını “Kommunist”, “Bakı”, “Azərbaycan gəncləri” kimi respublika qəzetlərinin 1976-cı il nömrələrindən götürülmüş mətnlər təşkil edirdi. Seçmə vahidi olaraq qəzetin hər səhifəsindən götürülmüş 800 söz-forma qəbul edilmişdi. Qəzet materialları qabaqcadan janrlara ayrılmamışdı. Qəzetin ənənəvi strukturundan çıxış edərək mətnləri şərti olaraq belə qruplaşdırmaq olar: baş məqalə, daxili həyat, sənaye, kənd təsərrüfatı, elm və texnika yenilikləri, xarici xəbərlər, incəsənət, idman və s. məlumatlar. Ümumiyyətlə, “Azərbaycan qəzet dilinin tezlik lüğəti”ndə 81778 söz-formadan ibarət seçmələrdən alınmış 6532 müxtəlif söz qeydə alınmışdı.

Azərbaycan qəzet dilinin tezlik lüğəti əlifba-tezlik və tezlik siyahılarından ibarətdir. Lüğətin belə kiçik seçmə əsasında tərtibini onun tərtib olunduğu zamanla bağlıdır. Məlum olduğu kimi, Azərbaycan qəzet dilinin tezlik lüğəti ötən əsrin 70-ci illərində köhnə “BESM-2” elektron hesablama maşınında tərtib olunmuşdur. O vaxtlar Akademiyanın Dilçilik İnstitutunda hesablama maşını yox idi. Kibernetika İnstitutunda, yaxud Bakı Dövlət Universitetinin Hesablama Mərkəzində elektron hesablama maşınından yalnız əsas iş vaxtından kənar saatlarda istifadə etmək olardı.

İnformasiyanın (qəzet materialının) maşına daxil edilməsi də çətin bir proses idi. Köhnə elektron hesablama maşınlarında informasiya perforasiya maşınlarında qabaqcadan perforlent və ya perfokartlara yazılardan sonra yaddaşa daxil edilirdi. Buna da xeyli vaxt və əmək sərf olunurdu.

1976-cı ildə tərtib olunmuş və 2004-cü ildə işıq üzü görmüş bu lüğət Azərbaycanda tezlik lüğətlərinin tərtibi sahəsində ilk təcrübə kimi əhəmiyyətlidir [Azərbaycan qəzet dilinin tezlik lüğəti, 2004]. Tezlik lüğətləri ilə bağlı mövzunun davamı olaraq qeyd etmək lazımdır ki, bu sahədə sonralar da xeyli iş görülmüş, böyük uğurlar qazanılmışdır.

2010-cu ildə üçcildlik “Azərbaycan dilinin tezlik lüğəti”-nin I cildi çapdan çıxmışdır. I cild söz köklərini əhatə edir. Lüğət Azərbaycan Respublikası Rabitə və Yüksək Texnologiyalar Nazirliyinin maliyyə dəstəyi ilə çap olunmuşdur (tərtibçilər: M.Mahmudov, Ə.Fətullayev, S.Məmmədova, R. Fətullayev, R.Hüseynli, B.Talıbov, G.Əzimova, N.Abdullayev). Azərbaycan dilinin tezlik lüğətinin tərtibi prosesində ədəbi dilin bütün üslublarının təmsil olunması təmin olunmuşdur (elmi, bədii, publisist, rəsmi, məişət və s.). İlk işlənmə prosesində təxminən 50 milyona yaxın söz-forma “saf-çürük” edilmiş, rəqəmlər, təsadüfi simvol və işarələr atıldıqdan sonra təxminən 300 minə yaxın söz-forma qalmışdır. Tərtibçilər belə hesab edirlər ki, hə-

min baza əsasında gələcəkdə müxtəlif səciyyəli tezlik lüğətləri hazırlamaq olar: əsasların, söz-formaların, söz köklərinin tezlik, əlifba-tezlik, əks-əlifba tezlik siyahıları, qrammatik, leksik şəkilçilərin, qrafem və qrafem birləşmələrinin, durğu işarələrinin tezlik səciyyələrini əks etdirən siyahılar və s.

“Dilmanc” layihəsi əməkdaşlarının tərtib etdiyi “Azərbaycan dilinin tezlik lüğəti”nin 3 cildə çapı nəzərdə tutulur. Lüğətin I cildi söz köklərinin statistik səciyyələrini əks etdirən siyahılardan ibarətdir. Sözü kökü dedikdə, sözə bitişik şəkilçilər atıldıqdan sonra qalan və müəyyən məna (leksik, əşyavi, hal, vəziyyət, mücərrəd məfhum və s.) bildirən hissə başa düşülür.

II cild əsasların statistik göstəricilərini əks etdirən siyahılardan təşkil olunmuşdur. Əsas dedikdə, söz-formadan (ünsiyyət prosesində müəyyən qrammatik formada çıxış edən nitq vahidi) sonluqlar və forma düzəldən şəkilçilər atıldıqdan sonra yerdə qalan və sözü leksik məna bildirən hissəsi başa düşülür.

Azərbaycan dilinin tezlik lüğətinin III cildi söz-formaları əhatə edir. Söz-forma dedikdə, ünsiyyət prosesində müəyyən qrammatik formada çıxış edən sözlər qrupu nəzərdə tutulur. Yuxarıda qeyd etdiyimiz kimi, lüğətin I cildi – söz kökləri hissəsi çap olunmuş, qalan iki cildin isə çapı planlaşdırılır [Azərbaycan dilinin tezlik lüğəti, 2010]. Yeri gəlmişkən, müqayisə üçün qeyd edək ki, Azərbaycan dilinin tezlik lüğətinin I cildinin çapı prosesində 50 milyondan artıq söz-forma linqvo-statistik araşdırmaya cəlb olunmuşdusa, 1977-ci ildə çapdan çıxmış rus dilinin tezlik lüğətində isə cəmi 1 milyon söz-forma (söz-işlətmə) tədqiqata cəlb olunmuşdu. Nəticədə təxminən 40 min sözdən ibarət lüğət tərtib olunmuşdu. Haqqında danışılan lüğətin tərtibində 100-ə yaxın əməkdaş iştirak etmişdi. Azərbaycan dilinin çapdan çıxmış I cildində 50 milyon (rus dilinin tezlik lüğətində tədqiqata cəlb edilmiş söz-formalardan 50 dəfə çox) söz-forma statistik baxımdan təhlil olunmuş və lüğətin tərtibi işlərində cəmi 8 nəfər iştirak etmişdir. Əlbəttə, burada informa-

siya texnologiyalarının müasir inkişaf səviyyəsi (keçən 33 il ərzindəki inkişaf səviyyəsi göz qabağındadır), yeni, daha sürətli, böyük yaddaşa malik kompüterlərin mövcudluğu, informasiyanın daha müasir üsul və vasitələrlə kompüterə daxil edilməsi amilləri də az rol oynamır.

2004-cü ildə “Nurlan” nəşriyyatında “Azərbaycan dilinin əks əlifba lüğəti” çapdan çıxdı (tərtibçilər: M.Mahmudov, Ə.Fətullayev). Qeyd etmək lazımdır ki, oxuculara təqdim olunmuş bu lüğət Azərbaycan dilçiliyində ilk təcrübədir. Ötən əsrin 80-ci illərində AMEA Nəsimi adına Dilçilik İnstitutunun Riyazi dilçilik qrupunun əməkdaşları tərəfindən belə bir lüğət tərtib olunmuş və “Elm” nəşriyyatına təqdim olunmuşdu. Lakin, həmin lüğət çap olunmamış qaldı. Haqqında söhbət gedən lüğət Azərbaycan dilinin orfoqrafiya lüğətinin artırılmış və təkmilləşdirilmiş üçüncü nəşri əsasında hazırlanmışdı [Azərbaycan dilinin orfoqrafiya lüğəti, Bakı, “Elm”, 1975]. Məlum olduğu kimi, adı çəkilən lüğətin sözlüyü qismən köhnəlmiş, müstəqillik illərində dilimizə xeyli yeni söz və ifadə keçmiş, dilimizin lüğət tərkibi zənginləşmişdir. 2001-ci ildə latın qrafikalı yeni Azərbaycan əlifbasına keçilmişdi. Bütün bunlar əsas götürülərək, Azərbaycan dilinin əks əlifba lüğəti yenidən çapa hazırlandı və bu çapın əsasında 2004-cü ildə latın qrafikalı yeni Azərbaycan əlifbasında çapdan çıxmış, 80 minə yaxın sözü əhatə edən orfoqrafiya lüğəti dururdu [Azərbaycan dilinin orfoqrafiya lüğəti, Bakı, “Lider” nəşriyyatı, 2004]. Əvvəlki lüğətdə olmayan 18 min yeni söz Azərbaycan dilinin orfoqrafiya lüğətinin 2004-cü il çapına əlavə olunmuşdu. Beləliklə, Azərbaycan dilinin əks əlifba lüğəti orfoqrafiyanın bu yeni sözlüyünə əsaslanır və təxminən 80 min sözü əhatə edirdi. Lüğətin annotasiyasında əks lüğətin adı lüğətlərdən fərqi belə izah olunur: “Adi lüğətlərdən fərqli olaraq əks lüğətlərdə sözlər başlanğıcdan sona doğru deyil, əksinə, sondan başlanğıca doğru əlifba sırasına düzülür. Adi lüğətlərdə eyni başlanğıca malik sözlər, əks lüğət-

lərdə isə eyni sonluğa malik sözlər eyni qrupda yerləşir. Azərbaycan dilinin söz əmələgətirmə tiplərini və modellərini, morfonoloji quruluşunu tədqiq etmək baxımından əks lüğətlər əhəmiyyətlidir. Azərbaycan dilçiliyində ilk təcrübə olan bu lüğətdən nəzəri dilçiliyin bir çox məsələləri ilə əlaqədar, maşın tərcüməsi və mətnlərin formal təhlili sistemlərində, tekstologiya və paleoqrafiyada, əlyazmaların oxunuşu və bərpası işlərində, naməlum yazı sistemlərinin açılması prosesində və qafiyə lüğəti kimi istifadə etmək olar” [Azərbaycan dilinin əks əlifba lüğəti, 2004].

Yeni texnoloji vasitələr qədim Azərbaycan yazılı abidələrinin öyrənilməsi ilə bağlı da uğurla tətbiq olunur. Azərbaycanda bu sahədə də bir çox işlər görülmüş və bu istiqamətdə tədqiqatlar davam etməkdə və genişlənməkdədir. Bu sahədə işlər əvvəllər tezlik və əks lüğətlərin tərtibi sahəsində toplanmış təcrübəyə əsaslanırdı. Azərbaycan Elmlər Akademiyası Nəsimi adına Dilçilik İnstitutunun bir qrup əməkdaşı tərəfindən aparılmış “Kitabi-Dədə Qorqud” dastanlarının statistik təhlili kompüterlərin Azərbaycan dili tarixinə tətbiqi sahəsində ilk təcrübə hesab oluna bilər. Bu sahədə aparılan araşdırmaların gələcəkdə daha da dərinləşdirilməsinə və genişləndirilməsinə böyük ehtiyac var. 1997-ci ildə “Kitabi-Dədə Qorqud” dastanlarının 1300 illik yubileyi ərəfəsində tədqiqatın bir hissəsi – əlifba-tezlik siyahısı çap olundu (maliyyə imkanlarına görə tədqiqatı tam şəkildə çap etmək mümkün olmamışdı). Həmin kitabçanın çapı tərtibçilər üçün uğurlu oldu. Elmi ictimaiyyət belə bir araşdırmanın aparılmasından xəbərdar oldu. “Kitabi-Dədə Qorqud” dastanlarının 1300 illiyi haqqında Azərbaycan Respublikasının Prezidenti Heydər Əliyevin 20 aprel 1997-ci il tarixli Fərmanı ilə əlaqədar “Kitabi-Dədə Qorqud”un statistik təhlili” tədqiqatı tam şəkildə çap olundu [Bakı, “Elm”, 1999].

“Kitabi-Dədə Qorqud” dastanlarının dili ilə bağlı aparılmış uğurlu linqvo-statistik araşdırmalar dil tarixinin, digər klassiklə-

rin dilinin də kompüterlərin köməyi ilə öyrənilməsinə marağı artırdı. Klassik Azərbaycan poeziyasının parlaq siması Məhəmməd Füzulinin Azərbaycan dilində olan əsərlərinin statistik baxımdan öyrənilməsinə başlandı və gərgin əməyin nəticəsi olaraq 2004-cü ildə “Məhəmməd Füzulinin nəzm əsərlərinin əlifba-tezlik sözlüyü (tərtibçilər: K.A.Vəliyeva, M.Ə.Mahmudov, C.Ə.Rəhmanov, V.S.Sultanov) “Elm” nəşriyyatında çap olundu. Statistik araşdırma akad. H.Araslının hazırladığı Füzuli əsərlərinin beşcildliyi əsasında aparılmışdır. Həmin beşcildliyin I–III cildləri 1958-ci ildə, IV cild 1961-ci ildə, V cild isə 1985-ci ildə çap olunmuşdur.

Yeni texnologiyaların dilçilikdə istifadəsi ilə bağlı internetə qoyulmuş və geniş istifadəçilər ordusu tərəfindən istifadə olunan lüğətlər haqqında da məlumat vermək yerinə düşərdi:

- Poliqlot lüğətlər sistemi istifadəçilərin dərin rəğbətini qazanmış saytlardan hesab olunur. Proqram Azərbaycan Respublikasında İnformasiya-Kommunikasiya Texnologiyalarının inkişaf etdirilməsi layihəsi çərçivəsində hazırlanmışdır. Layihədə almanca-azərbaycanca, ingiliscə-azərbaycanca, azərbaycanca-ingiliscə, azərbaycanca-almanca, fransızca-azərbaycanca, azərbaycanca-fransızca, rusca-azərbaycanca, azərbaycanca-rusca, ingiliscə-rusca lüğətlər təmsil olunmuşdur;

- “Azərbaycan dilinin orfoqrafiya lüğəti”nin elektron variantı – [Azərbaycan dilinin orfoqrafiya lüğəti”, Bakı, “Lider nəşriyyatı”, 2004, 728 s.] kitabı əsasında hazırlanmışdır (tərtibçilər: Ə.B.Fətullayev, M.Ə.Mahmudov);

- “Azərbaycan dilinin əks əlifba lüğətinin elektron variantı 2004-cü ildə “Nurlan” nəşriyyatında çap olunmuş “Azərbaycan dilinin əks əlifba lüğəti” əsasında hazırlanmışdır. “Azərbaycan dilinin əks əlifba lüğəti” orfoqrafiya lüğətinin yeni sözlüyü əsasında tərtib olunmuşdur və təxminən 80000-ə yaxın sözü əhatə edir (tərtibçilər: Ə.B.Fətullayev, M.Ə.Mahmudov).

İnternetdə azərbaycandilli lüğətlərin biblioqrafiyasında da, yuxarıda göstərilənlərdən əlavə daha bir neçə elektron lüğət barədə məlumata rast gəlmək olar (müəllif Ü.Nəsibbəyli):

- AzerDict – Azərbaycanın ən böyük pulsuz onlayn azərbaycanca-ingiliscə, ingiliscə-azərbaycanca lüğəti;

- İntelsoft – rus dilindən Azərbaycan dilinə tərcümə sistemi;

- Linqsoft – ingiliscə-azərbaycanca, azərbaycanca-ingiliscə, rusca-azərbaycanca, azərbaycanca-rusca lüğətlər sistemi;

- Google Translate – ingiliscə-azərbaycanca, azərbaycanca-ingiliscə məşin tərcüməsi;

- Azərbaycanca-türkcə sözlük;

Terminoloji lüğətlər:

- Torpaqşünaslıq və aqrokimya terminlərinin izahlı və tərcümə lüğətləri;

- Qaynaq istehsalının ümumi əsasları (anlayışlar, terminlər, qaynaq növləri, standartlar);

- Rusca-azərbaycanca fəvqəladə hallarla bağlı işlənən terminlərin izahlı lüğəti;

- Kino, televiziya, radio terminləri;

- Azərbaycan dilində fizika-riyaziyyat terminləri lüğəti;

- Elmi fəaliyyətdə istifadə olunan əsas terminlər;

- Rusca-ingiliscə-azərbaycanca baytarlıq terminləri lüğəti;

- İzahlı tarix terminləri lüğəti;

- Soyuq silahın kriminalistik ekspertizası üzrə xüsusi terminlərin izahlı sözlüyü;

- Rusca-azərbaycanca arıçılıq terminləri lüğəti.

Bunlardan əlavə onlayn və endirilə bilən lüğətlər də ayrıca siyahı ilə verilmişdir:

- Maliyyə bazarları terminlərinin izahlı lüğəti;

- Torpaqların səhrələşməsi üzrə izahlı lüğət;

- Təhsil lüğəti;
- Azərbaycanca-İngiliscə-rusca frazeoloji lüğət;
- Günəş energetikası terminləri lüğəti;
- İngiliscə-rusca-azərbaycanca aqrar istehsal və maşın-qayırmaya aid terminlər lüğəti.

Son illərdə terminoloji lüğətlərin hazırlanması və internetdə yerləşdirilməsi sahəsində işlər daha da intensivləşmiş və genişlənməmişdir. Terminoloji lüğətlərin ayrıca portalı yaradılmış və orada elmin bütün sahələrinə aid mükəmməl lüğətlər yerləşdirilmişdir.

Əlbəttə, internetə olan marağın artması, internet məkanının getdikcə genişlənməsi, istifadəçi kontingentinin artması elektron ədəbiyyat şəbəkəsinin də getdikcə genişlənməsi zərurətini doğurur. Məhz ona görə də təqdim olunan siyahının zamanın tələbləri səviyyəsində genişlənməsi ehtimalı böyükdür.

“Dilmanc” Azərbaycanda ilk rəsmi olaraq fəaliyyət göstərən, zəngin leksik bazası və mükəmməl proqram təminatı olan maşın tərcüməsi sistemidir. Azərbaycanca-İngiliscə, İngiliscə-azərbaycanca, türkcə-azərbaycanca maşın tərcüməsi sahəsində geniş istifadəçilər ordusunun dərin rəğbətini qazanmışdır.

Dilmanc layihəsi çərçivəsində Azərbaycan dili üçün yeni, zamanın və çağdaş elmi axtarışların səviyyəsinə uyğun texnologiyaların yaradılması və tətbiqi sahəsində bir çox işlər həyata keçirilmişdir. Həmin texnologiyalar dünyanın aparıcı ölkələrinin beyin mərkəzlərində, elmi laboratoriyalarında hazırlanan layihələrlə eyni səviyyədə hesab olunurdu. Bu hər şeydən əvvəl ondan irəli gəlir ki, “Dilmanc” tərcümə sistemi öz fəaliyyətində aparıcı dünya alimlərinin təcrübəsindən yaradıcı şəkildə yararlanır və həmişə yeniliklərə can atır, istifadəçilərin arzu, irad və təkliflərini təhlil edir, imkan dairəsində nəzərə almağa çalışır. Layihə çərçivəsində linqvistik baza getdikcə təkmilləşdirilir və

zənginləşdirilir, proqram təminatı daim nəzarətdə saxlanır və istifadəçiləri qane etməyən cəhətlər aradan qaldırılırdı.

Layihədəki əsas elmi tədqiqat istiqamətləri aşağıdakılardır:

- Azərbaycan dilindən digər dillərə və əksinə tərcümə sistemlərinin yaradılması;
- Azərbaycan dili üçün nitqin tanınması sisteminin yaradılması;
- Azərbaycan dili üçün mətnin səsləndirilməsi sisteminin yaradılması;
- telefon danışığını Azərbaycan dilindən digər dillərə və əksinə tərcümə edə bilən sistemin yaradılması;
- mətnin kompüter tərəfindən başa düşülməsi sistemin yaradılması.

Layihənin əsas prioritetlərindən biri də bu texnologiyaların istifadəçilər tərəfindən praktik istifadəsini təmin etmək üçün proqram təminatının hazırlanması hesab olunurdu. Başqa sözlə, “Dilmanc” layihəsində nəzəriyyə ilə praktika sıx vəhdətdə götürülürdü. İstənilən nəzəri müddəa araşdırılır, əsaslandırılır, sınaqdan keçirilir və istehsalatda öz tətbiqini tapırdı. Sonralar, praktik tətbiqdən sonra da, təkmilləşmə, yeniləşmə, dəqiqləşmə aparılırdı. Burada internet istifadəçilərinin iradə, arzu və təklifləri mühüm rol oynayırdı. “Dilmanc” layihəsinin geniş uğur qazanmasını həm də bununla izah etmək olar.

Azərbaycan dilindən digər dillərə və əksinə tərcümə prosesinin avtomatlaşdırılması məsələsi ilk dəfə 2003-cü ilin əvvəllərində akademik Ə.M.Abbasov tərəfindən qoyulmuş və onun rəhbərliyi altında maşın tərcüməsi sisteminin yaradılması istiqamətində tədqiqatlara başlanmışdır. Bu müddət ərzində Azərbaycan dilindən digər dillərə və əksinə tərcümə prosesini avtomatlaşdırmağa imkan verən xüsusi modelləşdirmə metodu (rəqəmsal modelləşdirmə metodu) yaradılmış və bu metod əsa-

sında Azərbaycan-İngilis məşin tərcüməsi sisteminin yaradılmasına başlanmışdır.

İşlərin məntiqi davamı kimi Azərbaycan Respublikası Rabitə və İnformasiya Texnologiyaları Nazirliyində 2005-ci ilin aprel ayından “Dilmanc” layihəsinin həyata keçirilməsinə başlanmışdır.

“Dilmanc” layihəsi mobil tərcüməçi vasitəsilə şifahi nitqi, mətni və ayrı-ayrı sözləri Azərbaycan dilindən digər dillərə və əksinə tərcümə etməyə imkan verir. Proqrama tərcümə nəticəsində alınmış mətnin səsləndirilməsi imkanı da əlavə edilmişdir. Tək söz daxil etdikdə, həmin sözün lüğətdəki tərcüməsi, nitq hissəsi, sinonimləri və s. də göstərilir.

Dilmanc Conversation – Söhbət rejimində əcnəbi ilə canlı söhbət üçün tərcümə aləti işə düşür. Bu zaman müvafiq düymələri basmaqla Azərbaycan dilində danışdıqda xarici dildəki tərcümə, xarici dildə danışdıqda isə Azərbaycan dilindəki tərcümə səslənir. Dilmanc İmla proqramı klaviaturadan istifadə etmədən kompüterə mətn daxil etməyə imkan verir. Mətni mikrofonla şifahi oxuduqda bu proqram nitqi mətnə çevirir, bu isə mətni klaviaturada yığmaqdan bir neçə dəfə sürətlidir. Sadəcə, proqramı işə salınır, cursor istənilən mətn redaktoruna (məs., Microsoft Word, Notepad və s.) qoyulur və mətn mikrofonla oxunur. Dilmanc İmla mətni istifadəçinin əvəzinə şifahi danışığı yazılı mətnə çevirir.

Mətnin səsləndirilməsi sistemi hər hansı daxil edilmiş mətni insan səsi ilə səsləndirir. Həmin sistem təkmilləşdirilərək başqa uyğun sahələrə də tətbiq oluna bilər.

İntellektual IVR (Interactive Voice Response) sistemi dedikdə zəng edən şəxsin sorğusunu analiz edərək ona avtomatik cavab verən sistem başa düşülür. Bu sistem əvvəlcə insan nitqini tanıyaraq onu mətnə çevirir, sonra bu mətni analiz edərək abunəçiyə hansı məlumatın lazım olduğunu “başla düşür”, daha sonra bu sorğunu məlumat bazasında axtarır, bazadakı məlu-

matdan cavab cümləsi düzəldərək onu yenidən insan nitqinə çevirərək abunəçiyə çatdırır. Sistem həmçinin daxil olan zəngin lazım olan şəxsə yönəldilməsini, görüş vaxtlarının təyin edilməsi və digər əməliyyatları tam avtomatik şəkildə, insan operatoru olmadan yerinə yetirə bilir.

2012-ci ildə Heydər Əliyev fondu və Rabitə və İnformasiya Texnologiyaları Nazirliyinin dəstəyi ilə Dilmanc layihəsi çərçivəsində dünyada ilk olaraq görmə qüsuru olan insanlar üçün səsli idarə olunan mini kompüter yaradılmışdır. Monitoru və klaviaturası olmayan bu kiçik kompüter səsli əmrləri başa düşür və istifadəçiyə səsli cavab verir. Bu kompüterdə səsli əmrlər vasitəsilə internetdə xəbər saytlarına qulaq asmaq, radio dinləmək, elektron məktub yazmaq və digər işlər görmək mümkündür. Bu kompüterdən 100 ədəd eksperimental variantda hazırlanaraq görmə qüsuru olan istifadəçilərə təqdim edilmişdir.

Azərbaycan dilinin milli korpusunun yaradılması və praktik istifadəsi sahəsində də “Dilmanc” layihəsi çərçivəsində bir çox işlər görülmüşdür. Qeyd etmək lazımdır ki, fəaliyyət göstərdiyi 10 il ərzində (2005-2015) “Dilmanc” layihəsi çərçivəsində qazanılmış böyük elmi uğurlar bu sahədə daha geniş tədqiqatlar üçün baza rolunu oynaya bilərdi. Çox təəssüf ki, 2015-ci ildə “Dilmanc” layihəsinin fəaliyyətinə xitam verildi.

Azərbaycan dilinin statistik lüğətlərinin tərtibi təcrübəsi. Ümumi dilçiliyin inkişafı bir sıra başqa elmlərlə, o cümlədən təbiət elmləri ilə sıx bağlı olmuşdur. Müasir dövrdə elmlərin inteqrasiyası isə daha çox riyaziyyatın ayrı-ayrı elm sahələrinə nüfuzu ilə müəyyənləşir. Riyaziyyatın başqa elmlərə tətbiqi nəzəriyyənin daha yüksək inkişaf pilləsinə qalxmasına kömək edir, həm də bununla eyni zamanda yeni tədqiqat metodları, yeni məntiqi üsullar, anlayışlar və konsepsiyalar meydana çıxır.

Dil bir çox mürəkkəb qarşılıqlı əlaqəli müxtəlif obyektlər sistemini üzündə əks etdirir. Dil vahidlərinin işlənməsi elə mürəkkəb amillərdən asılıdır ki, əslində onların hamısını nəzərə almaq və onların qarşılıqlı əlaqələrini izləyib inkişaf istiqamətini müəyyənləşdirmək olduqca çətinidir. Dil obyektləri haqqında dəqiq hökmlər vermək çox vaxt mümkün olmur. Buna görə də, bir çox dil hadisələrinə xüsusi statistik qanunauyğunluqlara tabe olan “təsadüfi hadisələr” kimi baxmaq lazım gəlir. Təsadüfi hadisələrin və onların qanunauyğunluqlarının öyrənilməsi ilə riyazi statistika məşğul olur.

Statistik metodlar dilçiliyin ən müxtəlif sahələrinə getdikcə daha dərindən nüfuz etməkdədir. Dilçilik elmində geniş istifadə olunan bu metodların məcmusu (lingvostatistika) nitqdə və mətndə hansı vahidlərin tez-tez və ya çox az rast gələ bilməsi hallarını müəyyən etməyə imkan verir. Dil vahidlərinin işləkliliyinin təhlili nitq ifadələrinin əmələgəlmə prosesləri üzrə müəyyən qanunauyğunluqların aşkara çıxarılmasını təmin edir.

Tezlik lüğətləri. Dil faktlarının təsvirində statistik metodların tətbiq olunması sahəsində ilk təcrübələr sözlərin işlənmə tezliyi haqqında məlumat verən tezlik lüğətlərinin tərtibi ilə əlaqədardır. İlk tezlik lüğəti 1898-ci ildə çapdan çıxmış Kedinqin alman dili lüğətidir. Bu lüğətin tərtibi üçün təxminən 11 milyon sözdən ibarət mətn araşdırılmışdı. 60-cı illərdə artıq 16 müxtəlif dil üçün 800-ə kimi tezlik lüğəti tərtib olunmuşdu [Фрумкина, 1964].

Tezlik lüğəti tədqiq olunan mətndə (seçmədə) müşahidə olunan söz, söz-forma və ya söz birləşmələrinin siyahısıdır. Lüğətdə vahidlərin mətnlərdə işlənmə tezliyini göstərən rəqəmlər qeyd olunur.

Tezlik lüğəti sözləri işlənmə tezliyinə görə müqayisə etməyə imkan verir. Bununla əlaqədar, tezlik lüğətlərinin tətbiq olunma, istifadə dairəsi çox genişdir. Tezlik lüğətləri hər hansı mətnin sözlüyə və mətndəki ayrı-ayrı sözlərə nisbəti ilə müəy-

yənləşən bir sıra xüsusiyyətləri dəqiq ölçülərlə göstərir. Tezlik lüğətləri təhlil sistemlərində və MT-də olan avtomatik lüğətlərin əsasını təşkil etdiyi üçün onun üzərində geniş dayanaq.

Sözlərin yerləşdirilməsi üsulundan, tezliklərin azalma sırasına görə düzülüşündən asılı olaraq tezlik lüğətləri müxtəlif olur.

Sözlərin əlifba sırasına görə nizamlanması istənilən sözün tezliyi haqqında məlumatı dərhal axtarıb tapmaq imkanı verir. Belə düzülmə əsasında ən işlək sözlərdən başlayaraq istənilən miqdarda sözləri seçib ayırmaq olar. Söz tezlikləri cəminin mətnin ümumi həcminə bölünməsi nəticəsində sözlərin tədqiq olunan mətnədə rolu və mətnəmələgətirmə imkanları müəyyənləşir.

Hazırda bədii ədəbiyyat nümunələri, qəzet və elmi-texniki mətnlər əsasında bir çox tezlik lüğətləri tərtib olunmuşdur.

Dilin tam tezlik lüğətlərinin tərtibi işinin bir çox çətinlikləri vardır. Bu çətinliklər, hər şeydən əvvəl, mətnlərin seçilməsi, yəni hansı mətnlərin ədəbi dili təmsil edə bilməsi məsələsinin həlli ilə əlaqədardır. Belə lüğətlərin əsasını bədii əsərlərdən, elmi-texniki və tədris ədəbiyyatından, qəzet, jurnal və digər mənbələrdən götürülmüş mətnlərin məcmusu təşkil edir. Mətnlərdəki sözlər tezlik lüğətinin məzmununda öz əksini tapdığı üçün onların seçilməsinə məsuliyyətlə yanaşmaq lazımdır. Bundan başqa, dilin lüğət tərkibini bütünlüklə tam əhatə edə biləcək tezlik lüğətlərinin tərtibi zamanı tədqiqata cəlb olunacaq mətnlərin həcmi çox böyük olmalıdır. Kiçik həcmli seçmələr əsasında tərtib olunmuş lüğətlər, təbii ki, dili tam təmsil edə bilməz və bu lüğət əsasında dilin leksikası haqqında hökm vermək mümkün deyildir.

Təcrübədə ən çox xüsusi elm sahələrini əhatə edən, elmi-texniki və qəzet materiallarından aparılan seçmələr əsasında tezlik lüğətləri tərtib olunur.

Tezlik lüğətləri onlara daxil edilən vahidlərə görə də birbirindən fərqlənirlər. Bəzən lüğətlərdə ayrıca lüğət vahidi kimi qəbul edilməyən söz-formaları (məsələn, kitab, kitabı, kitabda, kitabdan və s.) başqa lüğətlərdə “lüğət” vahidi kimi təmsil olunur. Hazırda kök-morfem, şəkilçi, habelə söz birləşmələrinin tezliyini göstərən tezlik lüğətləri də mövcuddur.

Tezlik lüğətləri əllə, məşinlərin köməyi ilə və ya qarışıq üsulla tərtib oluna bilər.

Tezlik lüğətlərinin tərtibi prosesi ümumi şəkildə aşağıdakı mərhələlərdən ibarətdir:

- məndən bütün müxtəlif sözlər (söz-formalar) götürülür;
- hər bir sözün tezliyi hesablanır;
- sözlərin əldə edilmiş məcmusu tezlik sırası ilə düzülür.

Tədqiqatın məqsədindən asılı olaraq müxtəlif siyahılar və statistik göstəricilər alınır. Məsələn, çox vaxt tezlik siyahısından əlavə sözlərin əlifba-tezlik siyahısı da verilir. Lazım gəldikdə lüğətdə hər bir sözün sıra nömrəsi, mütləq tezliyi və s. göstərilir [Алексеев, 1975; Mahmudov, Cəlilov, 1977; Частотный словарь русского языка, 1977].

Linqvistik statistika dil və nitq faktlarının kəmiyyət xüsusiyyətlərini özündə cəmləşdirir. Bir tədqiqat metodu kimi, o, dilçiliyin bütün sahələrində tətbiq oluna bilər. Linqvistik statistika, yuxarıda qeyd olunduğu kimi, dilçiliklə elm və texnikanın başqa sahələrinin qovuşduğu yerdə inkişaf edir.

Tezlik lüğətləri, leksikanın statistik modeli kimi, ümumi leksikologiyanın bir sıra problemlərini (lüğət tərkibinin nüvəsi, fəal və qeyri-fəal lüğət məsələləri və s.) həll etmək üçün material verir.

Leksikanın tərkibi haqqında statistik məlumatdan tipoloji tədqiqatlarda da geniş istifadə olunur. Ayrı-ayrı dillərin oxşar və fərqli cəhətlərini müəyyənləşdirmək məqsədi ilə dil sistemlərinin öyrənilməsi linqvistik tipologiyanın ümumi məsələlərinin

dən biri sayılır. Bu da öz növbəsində həmin sistemləri tipoloji əlamətlər əsasında təsnif etməyə imkan verir.

Linqvistik statistika dilin tarixi inkişafının öyrənilməsi ilə əlaqədar mühüm müşahidələr aparmaq üçün də zəmin yaradır. Qismən leksikanın əsas nüvəsinin dəyişmə tempini müəyyən-ləşdirməyə yönəldilmiş qlottoxronologiya da belə statistik me-todlardan biri hesab olunur.

Dilin ayrı-ayrı üslublarının – elmi, publisistik, məişət, rəsmi və s. üslubların kəmiyyətə öyrənilməsi məsələsi də bö-yük əhəmiyyət kəsb edir. Üslubi statistika fərdi üslubların kə-miyyət parametrlərini dəqiqləşdirməklə eyni zamanda ədə-biyyatşünaslıq elminə xeyir vermiş olur.

Müəllif üslubunun kəmiyyət qiyməti anonim və mü-bahisəli mətnlərin əsl müəllifinin təyin edilməsində mühüm meyar sayılır. Burada statistika ənənəvi tekstologiya və krimi-nalistikanın köməyinə gəlir.

Sözün işlənmə dərəcəsi üzərində aparılan dəqiq müşa-hidələr fərdi nitqin normadan kənaraçıxma hallarını qiymətlən-dirməyə əsas verir. Bu qiymətlər diaqnostikada və bəzi psixi xəstəliklərin müalicəsi zamanı istifadə oluna bilər.

Uşaq danışığı əsasında tərtib olunmuş tezlik lüğətləri uşaqlarda nitqin inkişaf prosesini mərhələ-mərhələ, pillə-pillə izləməyə imkan verir. Burada linqvistik statistika pedaqogika və psixologiyanın köməyinə gəlir.

Mətnin statistik quruluşu problemi ana dilinin və xarici dillərin tədrisi ilə əlaqədar daha səmərəli metodların işlənməsi məsələsini də ehtiva edir. Bu sahədə işləyən mütəxəssislərə – dərslük tərtibçilərinə tezlik lüğətlərinin böyük köməyi ola bilər. Əhəmiyyətini nəzərə alaraq, tezlik lüğətlərinin tədrisdə istifadə imkanlarını nəzərdən keçirək.

Dillərin tədrisi ilə məşğul olan müəllimlər dil materialla-rının seçilməsində statistik metodlardan geniş istifadə edə bi-lərlər. Dilin tədrisi ilə əlaqədar lüğət-minimumlar üçün sözlü-

yün tezlik siyahısı əsasında seçilməsi obyektiv metod sayılır. Bu metod leksikada sabit, ən çox müşahidə olunan, təkrarlanan vahidlərin və müəllifin iradəsindən, mövzunun xüsusiyyətlərindən asılı olub nisbətən az işlənən sözlərin seçilməsi üçün də obyektiv meyar kimi özünü doğruldur [Mahmudov, 1992].

Xarici dillərin tədrisi prosesinin optimallaşdırılması rasiional metodikadan istifadə olunması məsələsini ön plana çəkir. Bu zaman insanın linqvistik və linqvopsixoloji fəaliyyəti nəzərə alınmaqla müvafiq materialın seçilməsi, onların tədris proqramına daxil edilməsi, real həyat situasiyaları ilə bağlı materiallardan istifadə olunması məsələsi qarşıda durur. Dilin tədrisi üçün tərtib olunmuş rasiional metodikada üç mühüm məsələ – tələbənin (şagirdin) öyrəndiyi dili tətbiq edəcəyi şərait, bunun üçün zəruri olan dil vasitələri (nitq vərdişləri) və minimum vasitə və vərdişlərin müəyyənləşdirilməsi yolları nəzərə alınmalıdır [Алексеев, 1975].

Birinci məsələyə cavabı müvafiq ali məktəbin professional oriyentasiyası verir. Deməli, gələcək mütəxəssisin əldə edəcəyi peşənin də burada böyük əhəmiyyəti var. İkinci və üçüncü məsələ ilə əlaqədar, konkret professional sahədə dilin funksiyalarının öyrənilməsi böyük əhəmiyyətə malikdir. Burada nitq fəaliyyətinin ehtimali-statistik və nəzəri cəhətdən tədqiqinə diqqət yetirmək lazımdır. Bu zaman ehtimali determinist yanaşma üsullarından, müvafiq tədris modellərindən və dil modellərindən istifadə edilməsi məsələsinə diqqət yetirmək lazımdır.

Tədris prosesində qarşıda duran əsas problemlərdən biri də lazımi tədris materialının seçilməsidir. Bir çox illərdir ki, bu problem dünya metodistlərinin diqqət mərkəzində durur. Bu məqsədlə tərtib olunmuş tezlik lüğətləri ümidverici nəticələr vermişdir. Tədqiqatlar göstərmişdir ki, tezlik siyahısında birinci 100 söz mətndəki söz işlətmələrin (söz-formaların) 50 faizini əhatə edir. Birinci 1000 sözün isə mətni əhatə dərəcəsi 70-80

faizdir. Bu ümumi universal linqvostatistik qanunauyğunluq böyük əhəmiyyətə malikdir. Bir məsələni unutmaq olmaz ki, tezlik siyahısı ancaq onun tərtibi üçün istifadə olunmuş mətnləri əhatə edə bilər. Bütövlükdə dildə çox işlək sözlərin təyini üçün mühüm seçmələrin aparılması lazım gəlir. Ümumiyyətlə, dili tamamilə öyrənmək olmaz, müəyyən şəraitdə dildən istifadə olunma vərdişləri göstərilə bilər. Dili tam əhatə edən lüğət almaq üçün çox böyük həcmdə mətnlərin işlənməsi lazım gəlir (təxminən 1 mln. Standart mətn və ya 1 mlyd. söz-forma).

Lüğət-minimumların tərtibində də bir o qədər mətnin işlənməsi lazım gəlir. Məlum olduğu kimi, lüğət-minimum dildə işlənən ən mühüm sözləri əhatə edir. Deməli, belə lüğətlər dilin lüğət tərkibinin əsasını təşkil etməlidir. Onu da qeyd etmək lazımdır ki, sözün “mühümlüyü”, təkcə işlənmə tezliyi ilə deyil, həm də işlənmə dairəsi ilə müəyyən edilməlidir (həmin sözün bədii ədəbiyyatda, şifahi nitqdə, mətbuatda və s. işlənməsi).

Kompüterlərin yaranması və kibernetikanın inkişafı ilə əlaqədar dil haqqında statistik məlumatlara tələbat daha da artmışdır. Müasir informasiya sistemlərində, MT-də, referatların avtomatik hazırlanmasının, linqvistik vahidlərin statistik siyahılarının əldə edilməsində, qədim, şifrələnmiş mətnlərin oxunmasında, kriminalistikada tezlik lüğətləri ən zəruri yardımçı vasitədir. Tezlik lüğətləri maşın lüğətlərinin də əsasını təşkil edir. Onlardan informasiya axtarışı sistemlərində geniş istifadə olunur.

Lüğət tərkibi haqqında statistik məlumatların qeydə alınması qrafika və orfoqrafıyanın təkmilləşdirilməsi ilə də sıx əlaqədardır. Qrafem və hərf birləşmələrinin ehtimali-səciyyəsinin nəzərə alınması vacib şərtlərdən biridir.

Tezlik lüğətlərinin statistik parametrləri çap işinin avtomatlaşdırılması, mətnlərin avtomatik oxunması və qavranılması işlərində də müvəffəqiyyətlə istifadə oluna bilər.

Tezlik lüğətləri həm də rabitə kanallarının keçiricilik qabiliyyətini artırmaq işində, korlara oxunun öyrədilməsi və stenografiya sisteminin təkmilləşdirilməsi işlərində istifadə olunur.

Türkologiyada statistik tədqiqatların tarixi az olmasına baxmayaraq, bu sahədə müəyyən müsbət nəticələr əldə edilmişdir. Hesablayıcı texnikanın türkologiyada tətbiqinin ilk təcrübəsi də başlıca olaraq statistik leksikoqrafiya sahəsi ilə bağlıdır.

Məhz buna görə, mətn işlənərkən lüğətlərdəki sözlərin bir hesablama vahidi kimi seçilməsi prosesində mətnin istənilən seqmentinin müəyyən sintaqmatik ardıcılığa və ya paradigmatik sinfə aid olub-olmaması kimi zəruri bir məsələ ilə tez-tez qarşılaşmalı oluruq.

Azərbaycan dilinin əks lüğəti. Məlum olduğu kimi, adi lüğətlərdə sözlər əlifba sırası ilə düzülür, yəni hər hansı iki sözdən lüğətə birinci o söz daxil olur ki, başlanğıc hərfi əlifbada daha əvvəl gəlmiş olsun. Əgər ilk hərflər eynidirsə, onda ikinci hərflər müqayisə olunur və i.a. Əks lüğətlərdə də sözlər əlifba sırası ilə düzülür, lakin burada düzülüş qaydası sözün əvvəlinci hərfindən deyil, sonuncu hərfdən asılıdır. Yəni iki söz müqayisə edilərkən onların əvvəlindəki deyil, sonundakı hərflər tutuşdurulur və qarşılaşdırma sözün sonundan əvvəlinə doğru (sağdan sola) davam etdirilir. Məsələn, əks lüğətdə sınaq sözü alim sözündən əvvəl yerləşdirilir. Çünki q hərfi əlifba sırasına görə m hərfindən daha irəlidədir. Tamam sözü də alim sözündən qabaq yerləşdirilir, çünki bu sözlərin sonuncu hərfləri eyni olduğu üçün onlar sondan ikinci hərflərin (a və i) tutuşdurulması əsasında sıralandırılır. Nümunə üçün, müəyyən qrup sözlərin adi və əks lüğətdəki düzülüşünü müqayisə edək:

Cədvəl 1. Sözlərin adi lüğətlərdə və əks lüğətlərdə düzülüşü.

adi lüğət	əks lüğət
ağır	şura
çiçəkli	çiçəkli
oturmaq	təsəlli
şura	udmaq
təsəlli	oturmaq
udmaq	ağır

Beləliklə, adi lüğətlərdə əvvəli eyni hərflərdən ibarət olan sözlər bir yerdə yerləşirsə, əks lüğətdə sonu eyni hərflərdən ibarət sözlər bir sırada düzölmüş olur. Belə ki, Azərbaycan dilinin əks lüğətində bütün feillər, daha doğrusu -maq və -mək şəkilçili feil qrupları iki yerə toplanmışdır. -lıq şəkilçili bütün isimlər bir qrupda, -lı şəkilçili bütün sifətlər isə bir sırada olacaqdır. Məhz bu xüsusiyyət əks lüğətlərin praktik istifadəsi üçün mühüm amillərdən biri sayılır.

Əks lüğətlərdən daha çox morfologiya sahəsində aparılan tədqiqat əsərlərində istifadə edilir. Türk dillərində, o cümlədən Azərbaycan dilində söz tərkibindəki bütün morfoloji göstəricilər sözün kökündən sağ tərəfdə yerləşdiyinə görə əks lüğətlər bu dillərin morfoloji quruluşunun tədqiqi üçün xüsusilə əlverişlidir.

Türk dillərinin əks lüğətlərində bütün sözlər qrammatik əlamətlərinə görə (nitq hissələri, nitq hissələri daxilindəki söz qrupları, müəyyən şəkilçilər) qruplaşdırılır. Sözdüzəldici qəliblərin (məsələn, -lan, -ləş şəkilçiləri ilə isimlərdən düzələn feillərin) araşdırılmasında əks lüğətlər tədqiqatçıya bu və ya digər qəlibin bütün nümunələrini əhatə edən zəngin material verir. Bu lüğətlərdə sözdüzəldici şəkilçilərin müxtəlif əsas və ya şəkilçilərlə birləşməsi daha aydın, daha qabarıq əks olunur. Bu baxımdan növ şəkilçilərinin birləşmələri xüsusilə maraqlıdır, çünki bunların müxtəlif əsaslarla birləşmə imkanları kifayət qədər öyrənilməmişdir.

Əks lüğət fonetik hadisələrin tədqiqində də faydalı ola bilər. Xüsusilə, sözün əsasında baş verən morfonoloji proseslərin izlənməsində, söz sonunda mümkün olan bütün səs qruplarının aşkar edilməsində onun əhəmiyyəti böyükdür.

Əks lüğət morfolojiya bəhsinin tədrisində əlverişli köməkçi vasitə rolunu oynayır. Onun köməyi ilə müəyyən tip əsas və şəkilçilərə aid mükəmməl və əyani tədris materialı seçmək imkanı yaranır.

Bu lüğət vasitəsilə eyni zamanda sözdüzəldici modellərin yayılması haqqında bir çox statistik məlumatlar əldə etmək olar.

Əks lüğətlərdən avtomatik tərcümə alqoritmlərinin qurulmasında, daha doğrusu, avtomatik morfoloji təhlildə istifadə edilir. Təhlil üsullarından biri söz formalardan şəkilçiləri ayırdıqdan sonra qalan hissənin əsaslar lüğətində axtarılmasından ibarətdir. Bu prosesdə şəkilçilərin və söz formaların sonunun omonimliyi ilə əlaqədar bir çox yanlış ayırmalara yol verilə bilər, məsələn, gedək (müq.et: ged-ək), düst-ur (müqayisə et: dur-ur), vic-dan (müq.et: ayaq-dan) və s. Əks lüğətlər tədqiqatçıya belə halları nəzərə alıb asanlıqla aradan qaldırmağa kömək edir.

Əks lüğətlərdən tekstologiya və paleoqrafiya sahəsində müxtəlif məsələlərin həllində istifadə etmək olar. Tutaq ki, tekstoloq korlanmış və ya çətin oxunan əlyazmada yalnız sonu asanlıqla oxunan sözə rast gəlmişdir. Belə hallarda o, əks lüğətdən istifadə edərək həmin hərflərin birləşməsi ilə qurtaran sözlərin hamısını nəzərdən keçirib öz işini xeyli yüngülləşdirə bilər.

Eləcə də, naməlum yazı üsulunun açılmasında əks lüğət tədqiqatçının köməyinə gələ bilər. Qafiyələrin seçilməsində, yaxud xüsusi qafiyələr lüğətinin yaradılmasında da əks lüğət faydalı ola bilər.

Göründüyü kimi, əks lüğətlərin tətbiq dairəsi təkcə dilçiliklə məhdudlaşmır. Bu lüğətlər eyni zamanda müxtəlif məlumatlar mənbəyi kimi daha geniş istifadə olunma imkanlarına malikdirlər.

Abidələrin statistik təhlili və sözlüklərin tərtibi. Tezlik və əks lüğətlərin tərtibi sahəsində qazanılmış təcrübələr Azərbaycan yazılı abidələrinin öyrənilməsi, onların statistik baxımdan təhlili məsələlərinə yenidən baxılmasına imkan verir. 1988-ci ildə “Yazıçı” nəşriyyatında çap olunmuş “Kitabi-Dədə Qorqud” dastanlarının mükəmməl tənqidi mətni əsasında (tərtibçilər F.Zeynalov, S.Əlizadə) abidənin kompüterlərdə statistik təhlili ilə əlaqədar üç siyahı alınmışdır.

- tezlik;
- əlifba-tezlik;
- əks əlifba.

Tezlik siyahısı on minə yaxın (9852) söz-formanı əhatə edir. Dastanda işlənən sözlər burada işlənmə tezliklərinə görə azalma sırası ilə verilmişdir. Tezlik siyahısı abidənin dili barədə bir çox qiymətli nəticələr əldə etməyə imkan verir. Siyahıda lap yuxarıda yüksək tezlikli söz-formalar gəlir. Tezliyi 200-dən yuxarı sözlər aşağıdakı ardıcılıqla sıralanmışdır: dedi (731), aydır (714), bir (505), qara (348), Qazan (313), nə (304), gəldi (284), ağ (230), bu (224), xanım (223), ol (212), mərə (212).

İşlənmə tezliyi az olan sözlərin hərtərəfli tədqiqi və səciyyələndirilməsi də böyük maraq doğurur. Tədqiqatın məqsəd və vəzifələrindən asılı olaraq tezlik siyahısı tədqiqatçılara bir çox maraqlı məlumatlar verə bilər.

Əlifba-tezlik siyahılarında söz-formalar tezliyə görə deyil, əlifbaya görə düzülüşlər. Hər hansı bir söz-formanın tezliyini öyrənmək istədikdə, onu əlifba sırasına görə asanlıqla tapmaq olar. Əlifba-tezlik siyahısında müxtəlif tezliyə malik sözlər yanaşı işləne bilər. Məsələn, əlifba-tezlik siyahısında

tezliyi 20-yə bərabər olan av sözü tezliyi 6 olan abdəst sözündən sonra gəlir və s.

“Kitabi-Dədə Qorqud” dastanlarının statistik təhlili nəticəsində alınmış əks-əlifba siyahısında sözlər sonuna görə əlifba ilə sıralanmışlar. Məs., baba, oba, yuva sözləri bu siyahıda yanaşı gəlir. Əks-əlifba siyahısı abidənin meydana gəldiyi vaxtlarda sözdüzəltmə və söz dəyişdirmə formalarının müəyyənləşdirilməsi baxımından çox maraqlıdır. Bu və ya digər şəkilçi formasının işləkliyini və ya qeyri-ışləkliyini aşkar etmək dil tarixi ilə məşğul olan tədqiqatçılara qiymətli faktlar verə bilər.

“Kitabi-Dədə Qorqud” abidəsi ayrı-ayrı dövrlərdə tədqiqatçılar tərəfindən müxtəlif baxımdan öyrənilmişdir. İlk dəfə olaraq abidənin kompüterlərlə tədqiqi onun indiyə qədər tədqiqata cəlb olunmamış bir çox dil xüsusiyyətlərini dərinlən öyrənməyə imkan verir. Azərbaycan Elmlər Akademiyası Nəsimi adına Dilçilik İnstitutunun əməkdaşları tərəfindən aparılmış “Kitabi-Dədə Qorqud” dastanlarının statistik təhlili kompüterlərin Azərbaycan dili tarixinə tətbiqi sahəsində ilk təcrübədir. Göstərilən işin gələcəkdə daha dərinlən işlənməsinə, konkretləşdirilməsinə ehtiyac duyulur. Məsələn, abidədə işlənən söz əsaslarının, kök-sözlərin, şəkilçi formalarının statistik təhlili qiymətli nəticələr əldə etməyə imkan verərdi. Ayrı-ayrı boylar üzrə söz-formaların, əsasların, kök-sözlərin, şəkilçi formalarının statistik təhlili dastanı təşkil edən boyların müxtəlif dövrlərdə yaranması ilə əlaqədar mövcud mübahisələrə bir aydınlıq gətirə bilərdi. Tədqiqatın bu cəhətlərinin gələcəkdə daha dərinlən işlənməsinə böyük ehtiyac duyulur.

“Kitabi-Dədə Qorqud” dastanının 1300 illiyi haqqında Azərbaycan Respublikası Prezidenti Heydər Əliyevin 20 aprel 1997-ci il tarixli Fərmanı ilə əlaqədar “Kitabi-Dədə Qorqud” dilinin statistik tədqiqatı tam şəkildə çap olundu [“Kitabi-Dədə Qorqud”un statistik təhlili, 1999].

Gələcəkdə dastanın Drezden və Vatikan nüsxələrinin müqayisəli linqvostatistik təhlilinin aparılması faydalı olardı. Ayrı-ayrı əlyazma nüsxələri və boylar üzrə linqvostatistik tutuşdurmalar və ümumiləşdirmələr Vatikan və Drezden nüsxələrindəki fərqlərin elmi izahına imkan verərdi. Statistik araşdırmaların son mərhələsi “Kitabi-Dədə Qorqud” dastanlarının yarandığı dövrlərlə səsleşən digər abidələrin linqvostatistik təhlili və alınmış nəticələrin dastan mətni ilə müqayisəsi ilə yekunlaşa bilər. “Kitabi-Dədə Qorqud” dastanlarının statistik təhlili sahəsində uğurlu tədqiqatlar dil tarixinə kompüterlərin tətbiqinə stimul yaratmış oldu.

2004-cü ildə “Məhəmməd Füzulinin nəzm əsərinin əlifba-tezlik sözlüyü çapdan çıxdı [Məhəmməd Füzulinin nəzm əsərlərinin əlifba-tezlik sözlüyü, 2004].

Klassik Azərbaycan poeziyasının ən görkəmli yaradıcısı Məhəmməd Füzulinin anadilli nəzm əsərləri yüksək bədiilik nümunəsidir. Azərbaycan ədəbi dilinin təşəkkülü, formalaşması və inkişafı tarixini izləmək baxımından da şairin ölməz poeziyası olduqca əhəmiyyətlidir. Məhəmməd Füzuli söz sərrafı idi. O, doğma Azərbaycan dilinin zəngin potensial imkanlarından böyük sənətkarlıqla istifadə etmiş, hər sözə, hər ifadəyə yeni mənə, yeni məzmun verərək işləmiş, Azərbaycan ədəbi dilini yüksək inkişaf zirvəsinə çatdırmışdır.

Təqdim olunan kitabda Füzulinin anadilli nəzm əsərlərinin kompüterlər vasitəsi ilə statistik araşdırmasının nəticələri əks olunmuşdur. Onu da qeyd etmək lazımdır ki, Azərbaycan poeziyasının nəhəngi sayılan Füzulinin zəngin söz xəzinəsi indiyə qədər statistik baxımdan tədqiq olunmamışdı. Füzuli dilinin ecazkar zənginliklərinin statistik təhlili onun dil və üslub xüsusiyyətlərinin dərinədən öyrənilməsi ilə yanaşı XVI əsr Azərbaycan ədəbi dili haqqında qiymətli elmi məlumatlar əldə etməyə imkan verir. Statistik araşdırma nəticəsində iki mühüm siyahı əldə edilmişdir. Siyahılarda 23605 söz-forma ehtiva

olunmuşdur. Söz-forma dedikdə, cümlədə işlənən bütün sözlər (qrammatik şəkilçi qəbul etmiş və ya etməmiş) nəzərdə tutulur. Azərbaycan dilinin qrammatik quruluşunun, söz dəyişdirmə formalarının tədqiqi prosesində qrammatik şəkilçili sözlərin təhlili də çox faydalıdır.

Statistik təhlilin nəticələrinin əks olunduğu siyahılar quruluşlarına görə bir-birindən fərqlənir. Əlifba-tezlik sözlüyündə söz-formalar əlifba siyahısına görə sıralanmışlar. Hər bir söz-formanın qarşısında, onun işlənmə tezliyini əks etdirən statistik göstərici durur. Bu siyahı əsasında M.Füzulinin anadilli əsərlərində işlənmiş hər hansı bir sözün cəmi neçə dəfə işləndiyi barədə məlumat əldə etmək olar. Tezlik siyahısında söz-formalar işlənmə tezliyinə görə azalma siyahısı ilə sıralanmışdır. Bu siyahıda işlənmə tezliyi ən böyük olan söz birinci gəlir, qalan sözlər azalma sırası ilə düzülür.

Statistik araşdırma akad. H.Araslının hazırladığı Füzuli əsərlərinin beşcildliyi əsasında aparılmışdır. Həmin beşcildliyin I-III cildləri 1958-ci ildə, IV cild 1961-ci ildə, V cild isə 1985-ci ildə Azərbaycan Elmlər Akademiyası Nəşriyyatında çap olunmuşdur.

Tərtib olunmuş tezlik lüğətləri Azərbaycan dilinin quruluşu, qrammatik xüsusiyyətləri, leksik özəllikləri, söz dəyişdirmə və söz əmələgətirmə tipləri, inkişaf meyilləri barədə elmi cəhətdən əsaslandırılmış fikir yürütməyə imkan verir. Dilimizə son zamanlar daxil olmuş bəzi sözlərin işlənmə tezliyi çox yüksəkdir və dilin bütün üsullarına nüfuz etməkdə davam edir: administrasiya (3331), spiker (2119), eks-spiker (193), sponsor (1012), internet (11303), internet-provayder (26), internet-klub (22), internet-kafe (20), internet-sayt (21), internet-istifadə (23), departament (6276), şou (1184), şou-biznes (709), şou-proqram (74), kommersiya (5106), presedent (1074), informasiya (32242), informasiya-kommunikasiya (764), informatizasiya (36), disk (1492), displey (91), modern (1964), modifikasiya (36), disk (1492), displey (91), modern (1964), modifikasiya (36).

siya (178), modul (1061), kompüter (6873), kompüter-terminal (14) və s.

Bəzi söz və ifadələr isə paralel olaraq işlənmişdirlər. Onların statistik müşahidəsi maraqlıdır:

- informasiya (32242) – məlumat (184154);
- informasiya axtarış (19) – məlumat axtarış (35);
- informasiya sorğu (14) – məlumat-sorğu (10);
- (məlumat-soraq (172).

Həştad və səksən sayları arasında “yaşamaq”, lüğət tərkibində sabitləşmək uğrunda gedən mübarizə türk mənşəli səksən sayının xeyrinə dəyişməkdədir: səksən (373) – həştad (24); Söz-köklərinin şəkilsiz variantı ilə saitlə başlanan şəkilsiz variantlarının işlənmə tezlikləri də müxtəlifdir (söhbət sonu -q, -k şəkilsizləri ilə bitən söz-köklərindən gedir. Saitlə başlanan şəkilsiz qəbul etdikdə -q (-ğ), -k (-y) əvəzlənməsi baş verir. Eləcə də sonu -t ilə bitən söz-kökləri də saitlə başlanan şəkilsiz qəbul etdikdə -t (-d) əvəzlənməsi müşahidə olunur):

et (514874) – ed (927978)
get (40382) – ged (66290)
eşit (6177) – eşid (5751)
qayıt (9591) – qayıd (10954)
öyrət (1377) – öyrəd (1745)
kömək (22799) – köməy (6860)
ördək (129) – ördəy (28)
örnek (1544) – örnəy (178)
ürək (54) – ürəy (51)
bayraq (3222) – bayraq (1796)
bıçaq (315) – bıçağ (3734)
budaq (1032) – budağ (264)
çıraq (1256) – çıraq (305)
çiçək (3282) – çiçəy (431)
çörək (3324) – çörəy (1603)
xörək (674) – xörəy (180)

Mənsubiyyət və hal şəkilçiləri qəbul etdikdə sözün ikinci hecasında qapalı saitın düşməsi hadisəsinin statistik müşahidəsi də analogi nəticələr çıxarmaq imkanı verir. Tədqiqatçının qarşıya qoyduğu məqsədə müvafiq olaraq linqvostatistik məlumatlardan müxtəlif məqsədlərlə istifadə oluna bilər. Aşağıda “Azərbaycan dilinin tezlik lüğəti”ndə, “Azərbaycan qəzet dilinin tezlik lüğəti”ndə, “Kitabi-Dədə Qorqud”da Məhəmməd Füzulinin azərbaycandilli nəzm əsərlərində yüksək tezlikli feillər (50 feil) və yüksək tezlikli sözlər (100 söz) cədvəl şəklində verilmişdir. Göstərilən cədvəllər müasir Azərbaycan dilini (ayrıca olaraq 70-80-ci illərin qəzet dili də nəzərə alınmaqla) Dədə-Qorqud zəmanəsi və Füzuli dövrü ilə leksik baxımdan müqayisə etmək imkanı yaradır. Azərbaycan feillərinin tarixi inkişafı və bütövlükdə leksik qatın inkişaf tarixi və meylləri göstərilən lüğətlərdən götürülmüş bu cədvəllərdə öz əksini tapmışdır. Nəticə çıxarmaq dilçilərin, daha dəqiq desək, dil tarixçilərinin öhdəsinə düşür.

Azərbaycan dilinin başqa maraqlı xüsusiyyətləri, leksik-qrammatik özəllikləri, terminoloji yeniləşmə meylləri, yeni söz və ifadələrin dilimizə keçməsi məsələləri linqvostatistik fonda araşdırma obyektinə ola bilər.

Azərbaycan dilinin tezlik lüğətində yüksək tezlikli feillər (50 feil):

– ol (maq), ed/et (mək), gör (mək), ver (mək), keç (mək), de (mək), bil (mək), gəl (mək), çıx (maq), göstər (mək), art (maq), al (maq), böyü (mək), bildir (mə), apar (maq), istə (mək), seç (mək), tut (maq), yarad/yarat (maq), qur (maq), bax (maq), yaz (maq), elə (mək), düş (mək), yaşa (maq), çat (maq), qal (maq), danış (maq), qoy (maq), üz (mək), ged/get (mək), çək (mək), vur (maq), dəyiş (mək), gətir (mək), yet (mək), saxla (maq), tik (mək), qaz (maq), yaran (maq), öl (mək), çalış

(maq), öt (mək), at (maq), din (mək), kəs (mək), götür (mək), topla (maq), qazan (maq), sat (maq)

Azərbaycan qəzet dilinin tezlik lüğətində yüksək tezlikli feillər (50 feil):

– et (mək), ol (maq), ver (mək), göstər (mək), yetir (mək), edil (mək), artır (maq), olun (maq), de (mək), keçir (mək), al (maq), gör (mək), art (maq), istə (mək), gəl (mək), bil (mək), işlə (mək), keçiril (mək), yüksəlt (mək), yarat (maq), veril (mək). etdir (mək), apar (maq), görül (mək), çalış (maq), başla (maq), keç (mək), get (mək), hazırla (maq), çat (maq), qoy (maq), danış (maq), bildir (mək), çıx (maq), çatdır (maq), öyrən (mə), götür (mək), istə (mək), sal (maq), az (maq), qal (maq), genişləndir (mək), ödə (mək), tutul (maq), yaz (maq), düş (mək), qarşıla (maq), burax (maq), açıl (maq), salın (maq)

“Kitabi-Dədə Qorqud”da yüksək tezlikli feillər (50 feil):

– de (mək), ay (dır), gəl (mək), ol (maq), soylə (mə), gör (mək), al (maq), var (maq), ver (mək), söylə (mək), gir (mək), yat (maq), qoy (maq), ur (maq), baq (maq), qalq (maq), kez (mək), eylə (mək), get (mək), ayıt (maq), en (mək), gətir (mək), ağla (maq), bin (mək), düş (mək), tut (maq), dön (mək), çap (maq), tur (maq), qal (maq), dilə (mək), sor (maq), et (mək), ögün (mək), qaç (maq), çal (maq), as (maq), qaz (maq), qıy (maq), yet (mək), at (maq), dər (mək), kəs (mək), bil (mək), çıxar (maq), sıçra (maq), öldür (mək), sal (maq), bas (maq), qurtar (maq)

Məhəmməd Füzulinin azərbaycandilli nəzm əsərlərində yüksək tezlikli feillər (50 feil):

– ol (maq), eylə (mək), et (mək), qıl (maq), de (mək), ver (mək), yet (mək), gör (mək), düş (mək), gəl (mək), bil (mək), qal (maq), çək (mək), tut (maq), çıx (maq), bul (maq), dilə (mək), get (mək), ur (maq), tök (mək), görün (mək), açıl (maq), art (maq), burax (maq), eşit (mək), gəz (mək), qaz (maq), gətir (mək), keç (mək), ayıl (maq), dol (maq), götür (mək), başla

(maq), bax (maq), görün (mæk), neylə (mæk), unut (maq), əy (mæk), iç (mæk), tap (maq), yandır (maq), dön (mæk), dur (maq), oyna (maq), yaz (maq), ağla (maq), bənzət (mæk), cizgin (mæk), döndər (mæk), yor (maq)

Azərbaycan dilinin tezlik lüğətində yüksək tezlikli sözlər (100 söz):

– və, ol (maq), bu, ed/et (mæk), bir, ki, gör (mæk), ilə, il, ver (mæk), o, keç (mæk), baş, de (mæk), iş, özün, bil (mæk), öz, çox, dövlət, biç, ölkə, mən, kim, gəl (mæk), gün, isə, yer, onlar, məlumat, sonra, çıxmaq, on, nazir, ara, prezident, göstər (mæk), əsas, tərəf, art (maq), al (maq), yeni, böyü (mæk), respublika, belə, bildir (mæk), hər, qarşı, təşkil, qeyd, daha, bağ, yol, apar (maq), iki, bütün, hazır, həm, var, məsafə, həyat, söz, inkişaf, nə, istə (mæk), barə, üzrə, dünya, xalq, az (maq), zaman, rayon, əl, nəticə, haqq, sez (mæk), ad, insan, sahə, tut (maq), sir, fəaliyyət, əmək, şəhər, qədər, milli, iştirak, son, yarad/yarat (maq), hesab, əlaqə, vaxt, qur (maq), şəxs, bax (maq), yan, xidmət, da-xil, heç, həmin

Azərbaycan qəzet dilinin tezlik lüğətində yüksək tezlikli sözlər (100 söz):

– və, et (mæk), bu, bir, yaz, o, ol (maq), il, sovet, min, ki, Azərbaycan, iş, özün, doqquz, respublika, iyirmi, xalq, təsərrüfat, yer, çox, on, də, böyük, iki, da, sov.ikp, rayon, daha, məhsul, qurultay, bütün, beşillik, sahə, ver (mæk), əmək, kənd, hər, ölkə, plan, kommunist, beş, mk, baş, dövlət, gün, öz, yeni, göstər (mæk), yetir (mæk), istehsal, komitə, edil (mæk), ssri, yetmiş, ad, həyat, artır (maq), yüksək, bir, inkişaf, öz, yoldaş, beşinci, yoldaş, kolxoz, vəzifə, təşkilat, yol, qərar, olan, olun (maq), de (mæk), həmin, ittifaq, keyfiyyət, siyasi, tədbir, müəssisə, idarə, katib, kollektiv, dörd, şəhər, ton, sənaye, mübarizə, kimi, nəzər, səkkiz, keçir (mæk), yeddi, uğrunda, pambıq, proqram, vaxt, al (maq), yarış

“Kitabi-Dədə Qorqud”da yüksək tezlikli sözlər (100 söz):

– de (mæk), ay (dır), bir, qara, qazan, nə, gəl (mæk), ağ, xan (ım), bu, ol, oğul, mərə, dəxi, oğuz, yigit, kim, ol (maq), soyla (maq), mən, man (a), gör (mæk), at, ala, al (maq), oğlan, sən, kafər, bəg, üzəri (nə), var (maq), var, qız, san (a), qırq, ver (mæk), gön, qoca, üz, bura, dəli, ər, söylə (mæk), xəbər, görk, yer, qalın, yoq, adam, ilə, iki, ana, mægər, baş, və, qaba, kafir, yağı, dünya, gir (mæk), baba, əl, ya, yat (maq), saqqal (lu), xatun, gəl (mæk), ox, xoş, altmış, bin, göz, izun, ur (maq), baq (maq), qalq (maq), qatı, tənri, alp, keç (mæk), beş, qadir, eylə (mæk), neçə, sağ, hey, get (mæk), Allah, qurban, çıq (maq), biz, qan, dədə, ev, en (mæk).

Məhəmməd Füzulinin azərbaycandilli nəzm əsərlərində yüksək tezlikli sözlər (100 söz):

– kim, ki, bir, ol, bu, ilə, ey, hər, nə, ol (maq), eylə (mæk), mən, var, həm, gər, yenə, et (mæk), o, yox, tək, gül, qıl (maq), idi, can, çün (özün), çox, eşq, gün, ta, min, izrə, q (ğ)əm, dəm, yar, və, gör (mæk), göz, üzrə, könül, degil, hiç, xoş, necə, özgə, yüz, qan, nə, ah, ver (mæk), ikən, nola, mey, dəxi, yet (mæk), fələk, aşıq, dil, izhar, tut (maq), böylə, yox, müdam, əcəb, əlbəttə, imiş, dərd, guya, mægər, aləm, daim, əmma, həmişə, neçə, rəhm, zar, ərz, key, öz, əhl, iki, qələm, ləhzə, gəl (mæk), gərçi, şad, tərək, lütf, zahir, xəlq, billah, gərək, yad, ancaq, gərdun, nəzər, çək (mæk), zərt, mütləq, zaman, gecə.

Azərbaycan dilinin avtomatik lüğəti. Mətnlərin avtomatik işlənməsi sistemlərində və MT problemi ilə əlaqədar avtomatik lüğətlərin tərtibinin əhəmiyyəti və vacibliyinin sübuta ehtiyacı yoxdur. Əslində tezlik və əks lüğətləri kompüterlərdə yerləşdirilmək üçün hazırlanan avtomatik lüğətlərin bazası rolunu oynayır. Avtomatik lüğətin (AL) dili tam əhatə edə bilməsi və yığcamlığı üçün onun sözlüyü elmi cəhətdən əsaslandırılmış qaydada seçilməlidir. AL müxtəlif məqsədlərlə tərtib oluna bilər və bu məqsədlərə müvafiq olaraq onun strukturunda bir çox xüsusiyyətlər öz əksini tapmış olur. Əgər Hind-Avropa

dilləri üçün AL söz-formalar əsasında tərtib oluna bilərsə, türk dilləri üçün bu xüsusiyyət məqbul sayıla bilməz. Aqlütinativ quruluşa malik olan türk dillərində, o cümlədən Azərbaycan dilində eyni feil və ya isim kökünə nəzəri olaraq çoxlu sayda şəkilçilər zənciri birləşdirmək olar. Təbiidir ki, söz-formalar əsasında türk dilləri üçün tərtib olunmuş AL həcmcə olduqca böyük, əhatəliliyinə görə isə kiçik olar. Məhz bu baxımdan ən çox işlənən əsas və söz-formaların müəyyənləşdirilməsi böyük əhəmiyyət kəsb edir.

Azərbaycan dili üçün tərtib edilmiş AL aşağıdakı bölmələrdən ibarətdir:

- əsaslar AL;
- ən çox işlənən söz-formalar AL;
- birləşmələr AL.

Əsaslar avtomatik lüğəti. Formal təhlil sisteminin reallaşdırılması üçün linqvistik məlumatlar bazasının (bankının) olması mühüm şərtidir.

Mətnlərin avtomatik işlənməsi ilə bağlı istənilən sistemin fəaliyyət göstərməsi üçün qabaqcadan müəyyən linqvistik və ensiklopedik məlumatlar maşına daxil edilməlidir. Bu və ya digər linqvistik vahid haqqında məlumat məhz linqvistik məlumatlar bankının köməyi ilə əldə edilir. Mətnin müxtəlif məqsədlərlə işlənməsi prosesinin dəqiqliyi, etibarlılığı və tezliyi məhz bu məlumatların həcmindən, onların yerləşdirilməsi və saxlanmasının təşkilindən asılıdır [Пиотровский, 1979, 65].

Məlum olduğu kimi, türk dillərinin söz-formalar avtomatik lüğətinin tərtibi və istifadəsi mövcud mətnlərin avtomatik işlənməsi sistemlərində özünü doğrultmur, çünki bu baxımdan aqlütinativ quruluşlu dil flektiv-analitik və flektiv-sintetik dil qruplarından fərqli xüsusiyyətlərə malikdir.

Türk dillərinin quruluş xüsusiyyətləri mətnlərin formal təhlili və maşın tərcüməsi sistemlərində nəzərə alınmalıdır. Həzırda mövcud mətnlərin avtomatik işlənməsi, maşın tərcüməsi

və avtomatik referatlaşdırma sistemlərində istifadə olunan proqramlar əsasən Avropa dillərinə istiqamətlənmişdir və onların standart formaları türk dillərinə tətbiq oluna bilməz.

Türk dillərinin avtomatik işlənməsi sistemlərində avtomatik əsaslar lüğətindən istifadə olunması daha məqbul və effektivdir. Belə lüğətlər həcminə görə söz-formalar lüğətindən dəfələrlə kiçik olub, daha əhatəli hesab olunur.

Söz-forma və söz işlətmə anlayışlarının hər ikisinə ənənəvi dilçilikdə də rast gəlinir. Ənənəvi dilçilikdə – söz-forma müəyyən qrammatik formada çıxış edən söz mənasında izah olunur. Kitaba həvəs, kitabın dəyəri, kitabı sevmək, kitabdan ötrü birləşmələrində kitaba, kitabın, kitabı, kitabdan söz-formaları vardır. Halbuki söz yalnız birdir (“kitab”) [Adilov, Verdiyeva, Ağayeva, 1989, s.257].

Söz işlətmə – dedikdə isə nitqdə sözlərdən istifadə edilməsi (sözləri seçmə və birləşdirmə, əlaqələndirmə) qaydaları başa düşülür [Adilov, Verdiyeva, Ağayeva, 1989, s. 254].

Rus dilçiliyində də söz işlətmə əsasən “sözün nitqdə bu və ya digər mənada işlədilməsi”, “sözün nitqdə birləşmə imkanları və seçimi qaydaları” kimi izah olunur. “Söz işlətmə” anlayışının sırf formal baxımdan da izahı vardır ki, həmin anlamda söz işlətmə daha çox formal qrammatikalarda, formal təhlil sistemlərində işlənir. Həmin anlamda “söz işlətmə” – bir probeldən (boşluqdan) digər probele (boşluğa) qədər olan nitq vahidi (söz və ya başqa dil, işarə vahidi), iki probel arasında istənilən ardıcılıq hesab olunur. Defis və işarə də hərf kimi qəbul edilir.

Söz-forma dedikdə, müəyyən qrammatik formada çıxış edən sözlər qrupunu nəzərdə tutulur. Məsələn, “Mən dənizi sevirəm”, “Dənizə baxmaq məni sakitləşdirir”, “Dənizin suyu şordur”, “Dənizdə gəmi səyahəti maraqlı olur”, “Dənizdən danışmaqdan doymuram” cümlələrində dənizi, dənizə, dənizin, dənizdə, dənizdən söz-formaları dəniz sözündən əmələ gəl-

mişdir [Mahmudov, 2013, s.146]. Söz-forma formal təhlil sistemlərinin mərkəzi və əsas anlayışlarındanır. Bu baxımdan söz-forma mətnin formal olaraq təyin edilə bilən ən kiçik vahidi olub, ixtiyari sonlu hərflər ardıcılığından ibarətdir. Söz-formanın uzunluğu onu təşkil edən hərflərin (qrafemlərin) sayı ilə ölçülür. Ayrı-ayrı söz-formalar cümlələr çoxluğunun elementləri sayılır. Təhlil prosesində söz-formanın cümlədən ayrılması prosesində nəzərdə tutulur ki, söz-forma bir boşluqdan (probelədən) digər boşluğa (probelə) qədər məsafədə olan hərflər ardıcılığı zənciridir [Mahmudov, 2013, s.223].

Burada nəzərə alınmalıdır ki, formal təhlil sistemlərində iki probel arasında cümlənin bütün komponentləri düşə bilər və həmin komponentlərin heç də hamısı forma düzəldən şəkilçi qəbul etmir. Bu baxımdan formal olaraq cümləni təşkil edən sözlərin hamısı söz-forma kimi səciyyələndirilə bilər. Sadəcə həmin cümlə üzvləri lüğət vahidi olub-olmamasına görə təsnif edilə bilər. Lüğət vahidi olan (və ya ola bilmək ehtimalı olan) sözlər kök və ya əsas kateqoriyasına aid edilə bilər. İki boşluq (probel) arasında olub kök və ya əsas kateqoriyasına aid edilənlərdən başqa yerdə qalanlar söz-forma hesab oluna bilər. Biz söz-formalar lüğəti tərtib etsək, ora, təbiidir ki, söz kökləri və əsaslar düşməyəcək.

Söz işlətmə və söz-forma anlayışları yaxın mənəlidir. Söz işlətmə daha çox nitqə, söz-forma isə cümləyə aid anlayışlar kimi səciyyələndirilir.

Azərbaycan dili öz ön avtomatik lüğət əsaslar bazasında tərtib olunmuşdur. Lakin ən çox işlənən söz-formaların da ora daxil edilməsi məqsədəuyğun hesab olunmuşdur.

Onu təşkil edən hərflərin sayından asılı olaraq avtomatik əsaslar lüğətinin vahidləri qruplara ayrılır. Məsələn, Azərbaycan dilinin avtomatik əsaslar lüğətinin massivi 16 qrupa ayrılmışdır. Birinci qrup 16 hərfdən, ikinci qrup 15 hərfdən ibarət olan əsasları və nəhayət on altıncı qrup bir hərfdən ibarət əsas-

ları əhatə edir. Qrup daxilində sözlər əlifba sırası ilə düzülmüşdür.

Əsaslar AL Azərbaycan dilinin tezlik lüğəti bazasında yaradılmışdır. təhlil alqoritmlərində qabaqcadan qəbul edilmiş prinsiplərə müvafiq olaraq əsaslar tərkibindəki hərflərin sayına görə zonalara ayrılmış və zonalarda əlifbaya düzülüşlər. Demək lazımdır ki, türk dilləri üçün AL-i təşkil edən lüğət vahidlərinin tərkibindəki hərflərin sayına görə təsnifi özünü tam doğruldursa, başqa dillərdə bu lüzumsuz hesab oluna bilər. Ad nitq hissəsinə aid olan əsaslar adlıq halda, təkdə lüğətdə təmsil olunmuşdur. Feillər isə əmr formasında -maq, -mək məsdər şəkilçisi olmadan lüğətə daxil edilmişdir. Bu, Azərbaycan mətnlərinin avtomatik işlənməsi prosesi üçün olduqca mühüm amildir və adi lüğətlərdə olduğu kimi feilin məsdərdə verilməsi təhlil prosesində bir çox lüzumsuz anlaşılmazlıqların yaranmasına gətirib çıxarardı. Fikrimizi bir qədər aydınlaşdırmaq məqsədilə əyani misala müraciət edək.

Fərz edək ki, mətnin təhlili və ya avtomatik tərcüməsi zamanı yazıram söz-forması müşahidə olunmuşdur. təhlil alqoritmının işindən sonra söz-forma morfemlərə ayrılır: yaz-ır-am. Maşın -ır və -am şəkilçilərinin, şəkilçilər cədvəli ilə müqayisəsindən sonra müvafiq qrammatik informasiyanı əldə edəcək:

- ır – indiki zaman (feil)
- am – xəbərlik kateqoriyası, I şəxs, tək və s.

Şəkilçilərə görə məlum olur ki, yaz omonimi ad mənasında (“yaz” – “вечна”) deyil, feil kimi tərcümə olunmalıdır. Əsaslar AL-də yaz feilini tapır və onun mənasına şəkilçilərin mənalərini əlavə etməklə söz-formanın bütövlükdə qrammatik sintezini və dəqiq tərcüməsini alır. Əgər göstərilən feil AL-də məsdərdə təmsil olunsaydı, onda yaz əsasını orada birbaşa tapmaq mümkün olmazdı. təhlil alqoritmlərindən sonra yardımçı sintez alqoritmlərinə müraciət etmək zərurəti yaranardı. Sözüün əsasının (bizim misalda yaz) feil olduğunu müəyyənləşdir-

dikdən sonra sintez alqoritmi yaz sözündəki saitin (a) ahənginə uyğun olaraq -maq şəkilçisini ona birləşdirməli və onu yenidən morfoloji təhlil blokuna qaytarmalı idi. Sonra alınmış yazmaq sözü lüğətlə müqayisə oluna bilərdi. Feillərin lüğətdə əmr şəklində verilməsi prosesdəki bu artıq əməliyyatları ixtisar etməyə imkan verir.

Avtomatik əsaslar lüğətində bir sıra morfonoloji xüsusiyyətlər də nəzərə alınmışdır. Göstərmək lazımdır ki, ilkin tərtibdə bu morfonoloji dəyişmələr bir növ “primitiv” yolla qeydə alınmış və alqoritmləri mürəkkəbləşdirməmək üçün ağırlıq AL-ə keçirilmişdir. Azərbaycan dilində sonu t samiti ilə bitən feil kökləri saitlə başlanan şəkilçi qəbul etdikdə t-d cingiltiəşməsi baş verir. Məsələn,

et-edir, edək; get-gedir-gedək; eşit-eşidir, eşidək; qayıt-qayıdır, qayıdaq; unut-unudur, unudaq və s.

Bu xüsusiyyət alqoritmləşdirilmədiyi üçün belə sözlər iki variantda lüğətə daxil edilmişdir.

unut/unud – забыть, забывать; eşit/eşid – слышать, услышать; qayıt/qayıd – вернуться, возвращаться; get/ged – идти, уходить; et/ed – делать, сделать

Morfoloji proses sonu q, k samitləri ilə bitən çoxhecalı adlarda da geni müşahidə olunur:

torpaq – torpağa, torpağı; yarpaq – yarpağa, yarpağı və s.

çiçək – çiçəyə, çiçəyi; çörək – çörəyə, çörəyi və s.

Misallardan görüldüyü kimi, bu qəbildən olan sözlər də q-ğ, k-y cingiltiəşməsi baş verir. Bu sözlər əsaslar avtomatik lüğətində iki variantda verilmişdir:

torpaq / torpağ; yarpaq / yarpağ və s.

çiçək / çiçəy; çörək / çörəy və s.

Morfoloji səviyyədə söz-formaların şəkilçisiz (əsas kimi) lüğət vahidi kimi qəbul olunması bir çox çətinliklər yaradır. Bir tərəfdən, türk dilləri üçün yalnız söz-formalardan ibarət lüğət tərtib etmək aqlütinativ quruluşdan irəli gələn mühüm səbəblərə gö-

rə mümkün deyil, digər tərəfdən, söz-formaların əsaslara gətirilməsi omonim əsasların sayını çoxaldır, morfoloji təhlil alqoritmlərinin “yükünü” ağırlaşdırır.

Morfoloji səviyyədə omonimliyi aradan qaldırıla bilməyən əsaslar AL-də nitq hissələrinə potensial mənsubiyyətləri göstərilməklə verilmişdir. Omonim əsasların AL-də verilməsi: qaz – isim/feil – гусь/газ/копать; yaz – isim/feil – весна/писать

Sonrakı mərhələlərdə əsasların omonimliyi ona birləşmiş şəkildələrə görə, əgər şəkilçi yoxdursa, sintaktik-semantik blokların köməyi ilə aradan qaldırılır.

Sintaktik əhatə nəzərə alınmadan şəkilçisiz işlənən əsasların ayrılıqda, morfoloji səviyyədə morfoloji kateqoriyalarla omonimliyini və ya çoxmənalılığını aradan qaldırmaq qeyri-mümkündür.

Söz-formalar AL-nin tərtibi üçün, əvvəlcə, dilin təbii halında, heç bir dilçi müdaxiləsi olmadan götürülmüş mətnlərin əsasında hazırlanmış tezlik lüğətlərinin tərtibi vacibdir. Bunun üçün müəyyən elm sahələrini, qəzet leksikası və s. əhatə edən mətnlər kompüterə verilir. Seçmələrin həcmi böyük olduqca alınmış nəticələrin etibarlılığı artmış olur. Kompüterdə hesablandıqdan sonra yüksək tezlikli söz-formalar tapılır və onlar əsasında söz-formalar AL-i tərtib edilir.

Qeyd etmək lazımdır ki, türk dilləri üçün belə lüğətlərin tərtibi özünü doğrultmur. Türkcə-rusca məşin tərcüməsi sisteminin lüğət təminatında söz-formalar AL-nin hazırlanması və əsaslar AL-i ilə onun birgə istifadəsi nəzərdə tutulmuşdu [Бабаеров, 1981]. Lakin həmin sistemdə də əsas ağırlıq əsaslar lüğəti və morfoloji təhlil blokunun üzərinə düşür. 5400 linqvistik vahiddən ibarət söz-formalar AL-i daha çox müəyyən sahə ilə bağlı materialın rus dilinə tərcüməsi məqsədilə istifadə oluna bilər. Fikrimizcə, türk dilləri üçün ayrılıqda söz-formalar AL-i tərtib etmək məqsədəuyğun deyil, lakin mətnlərdə işlən-

mə tezliyi böyük olan söz-formalar (məsələn, mənə, səni, bunların, onlara, onları, xalqa, xalqın, illərdən, şəhərimizin, səfərdən və s.) istisna kimi əsaslar AL-nə daxil edilə bilər. Bu zaman həmin söz-formalar lazımı qrammatik informasiya ilə tam şəkildə təmin olunmalıdır.

Leninqrad (indiki Sankt-Peterburq) “Статистика речи” qrupunun əməkdaşları ilə birgə hazırlanmış türkcə-rusca AL yuxarıdakı prinsip üzrə hazırlanmışdır. AL türk dilində çıxan müxtəlif qəzetlərin materialları üzrə tərtib olunmuş əlifba-tezlik, əks tezlik, əks əlifba-tezlik, tezlik lüğətlərinin sözlükləri əsasında tərtib olunmuşdur. Türkcə-rusca AL 5 minə yaxın lüğət vahidindən ibarətdir. Əsasən əsasları əhatə edən lüğətdə ən çox işlənən və işlənmə dairəsi böyük olan söz-formalar da əlifba sırası ilə verilmişdir. Əsasların söz-formalardan ayrı verilməsi lüzumsuz hesab olunmuşdur.

Məsələn: lüğətdə “öldürmə”, “öldürücü”, “öldürülmə”, əsasları ilə yanaşı “önlənlərin” (umerevşix) söz forması da verilmiş, “politika” sözündən əlavə “politikamızın” (нашей политики) söz-forması ayrılıqda təsbit olunmuşlar və s.

Birləşmələr avtomatik lüğəti. Birləşmələr AL-nin səciyyəvi xüsusiyyətlərinin izahına keçməzdən əvvəl MT-nin və mətnlərin başqa məqsədlərlə avtomatik işlənməsinin bəzi cəhətlərinə diqqət yetirək. Məlum olduğu kimi, məşində cümlə mətndən ayrıldıqdan sonra ayrı-ayrı söz-formaların cümlədən ayrılması və təhlili, sonra tərcüməsi mərhələləri başlanır. Təcrübə göstərir ki, belə “ayrılmalar” heç də həmişə özünü doğrultmur, bəzən olduqca “qaba” və hətta yanlış tərcümələr alınması ilə nəticələnir. Məsələn, tutaq ki, cümlə baş verdi sözləri ilə bitir. Kompüterdə əvvəlcə baş sözü təhlil və tərcümə olunur. Sonra ver/mək sözü ardıcıl olaraq həmin mərhələdən keçir. Nəticədə baş ver/mək birləşməsi голова, головной – давать, отдавать kimi tərcümə olunur. Azərbaycan dilində “et”, “ol” köməkçi feilləri ilə və bir çox başqa feillərlə (məsələn, qoy,

ver, al, gör, çal, vur, gəl, çək, sal, elə, düş, olun, keçir, qazan, göstər, çıxar və s.) işlənən birləşmələrin əksəriyyətini ayrı-ayrı komponentlər üzrə tərcümə etdikdə yanlış təhlil və tərcümə nümunələri alınır. Komponentləri ayrılıqda təhlil və tərcümə olunan birləşmələr timsalında bütün bu qəbildən olan feili birləşmələr haqqında müəyyən təsəvvür əldə etmək olar.

Müşahidələr göstərir ki, formal təhlil və tərcümə prosesində Azərbaycan dilində ikinci komponenti feildən düzələn birləşmələrin bir çox xüsusiyyətləri aydınlaşır. Formal təhlil sistemi və MT-də göstərilən xüsusiyyətlərin nəzərə alınmaması hər iki sistemin fəaliyyətini və mövcudluq imkanlarını sifra endirmiş olar. Məsələn, baş sözü ikinci komponentdəki feildən asılı olaraq случиться, возникнуть, “наведываться, навеситить”, “вмешиваться, обращать внимание” kimi tərcümə olunur. Analoji xüsusiyyət ikinci komponenti yuxarıda sadaladığımız feillər ola biləcək bütün birləşmələrə xasdır. Deməli, birləşmələr AL-də sintaktik bloklara əlaqədar bir çox amillər mütləq nəzərə alınmalı, ikinci komponenti feil olan bütün birləşmələr ayrılıqda deyil, sintaktik vahid kimi tədqiq olunmalı, təhlil və tərcümə edilməlidir. Bunun üçün hər şeydən əvvəl ayrı-ayrı feil qrupları üzrə siyahılar olmalıdır. Məsələn, ikinci komponenti “ol”, “et” və s. ola biləcək sözlər əlifba, tərkiblərində hərflərin sayı və ya hər iki xüsusiyyət eyni zamanda nəzərə alınmaqla siyahı ilə verilir. Bizim tərtib etdiyimiz birləşmələr AL-də ikinci tərəfi feil olan sintaktik vahidlər iyirmi qrupa ayrılmış, hər qrupda konkret feilə (məsələn, “et”) birləşmək potensialına malik sözlər verilmişdir.

Formal təhlil və MT sistemləri üçün çox vacib olan bir xüsusiyyəti də qeyd etmək lazımdır: birləşmələr AL açıq sistemdir. İstər feil qrupları, istərsə də həmin qruplara daxil olan sözlərin siyahısı lazım gəldikdə genişləndirilə, ora müəyyən düzəlişlər və əlavələr edilə bilər. Belə dəyişikliklər nə lüğətin, nə təhlil və tərcümə alqoritmlərinin, nə də bütövlükdə formal

təhlil sisteminin işinə xələl gətirmir. Birləşmələr AL-də sintaktik vahidin birinci komponenti olan sözlər tərkiblərindəki hərf-lərin sayı nəzərə alınmaqla, azalan siyahı ilə verilir. Eyni uzunluğa malik sözlər öz aralarında əlifba siyahısı ilə düzülür. Bu prinsip təhlil və tərcümədə axtarışın optimallığını təmin edir. Komponentin birinci tərəfi olan sözlər adlıq halda, təkdə verilmişdir, çünki birləşmələrdə həmin sözlər bu şəkildə işlənir. Bir ləşmələrin ikinci komponenti olan feillər imperativdə verilir.

Azərbaycan mətnlərində müşahidə olunmuş hər hansı bir söz sintaktik əhatəsindən, ondan sonra gələn vahidin hansı nitq hissəsinə mənsubluğundan asılı olaraq müxtəlif cür təhlil və tərcümə oluna bilər. Məsələn, tanış sözdən sonra ad nitq hissəsi gələrsə (“tanış adam”) bu söz “знакомый, знакомец” kimi, “et/ed” qrupunda “ознакомить, познакомить” (“tanış et (mək)”), “ol” qrupunda “познакомиться” (“tanış ol (maq)”) kimi başa düşülür.

Sözlərin sintaktik əhatəsindən asılı olaraq müxtəlif mənalar kəsb etməsi xüsusiyyəti nəzərə alınmaqla təhlil və MT alqoritmlərinin məzmununa müəyyən əlavələr etmək, sintaktik elementlərdən istifadə olunmaqla döz gün təhlil və tərcümə nümunələri almaq olar. Fərz edək ki, cümlənin xəbəri “et (mək)” köməkçi feili ilə ifadə olunmuşdur. Azərbaycan dilində xəbər cümlənin sonunda gəldiyi üçün “et (mək)” köməkçi feilini əvvəlcə cümlədən ayıraraq təhlil etməliyik (Təhlil və MT alqoritmlərinə görə söz-formalar cümlənin sonundan başlayaraq ayrılır). İkinci mərhələdə xəbərdən əvvəl gələn cümlə üzvünü ayırırıq. Əgər yeni ayırdığımız söz “et (mək)” qrupuna daxildirsə, onda həmin ifadə birləşmə kimi təhlil və ya tərcümə olunur (məsələn, cədvəldə olduğu kimi “narahat etmək” – “беспокоить, тревожить”). Xəbərdən əvvəl gələn söz “et (mək)” qrupunda təmsil olunmamışdırsa, onda hər iki söz birləşmə kimi deyil, müstəqil cümlə üzvləri kimi tərcümə olunur (məsələn, “et (mək)” – “делать, сделать” kimi, ondan əvvəl gələn söz isə

(tutaq ki, “o” şəxs əvəzliyi) – “on, ona” kimi veriləcək: “O etdi” – “Он (она) сделал”. Əgər “narahat” sözündən sonra “adam” sözü işlənərsə, onda bu söz “беспокойный, счетливый” kimi başa düşüləcək. Nümunələrdən göründüyü kimi, “ol (maq)” qrupunda bu söz tamamilə başqa məna kəsb edəcək. Bütün göstərilənləri alqoritmləşdirməklə prosesin dəqiq ifadəsini vermək olar [Mahmudov, 2013, s.176-178].

Avtomatik lüğətlərin tərtibi mərhələsi ilə formal təhlil sisteminin təşkilinin ilkin və olduqca mühüm, həlledici mərhələsi başa çatmış olur. İstər morfoloji, istərsə də sintaktik-semantik səviyyələrdə mətnlərin avtomatik işlənməsinin istiqamətləri və spesifik xüsusiyyətləri leksik blokun mövcudluğu və məzmunu ilə şərtlənir.

Tezauruslar mətnlərin işlənməsi sistemlərində.

Leksikoqrafiyanın hazırkı səviyyəsində, formal təhlil sistemləri, MT və məlumat-axtarış sistemlərində avtomatik lüğətlərlə yanaşı tezauruslardan da geniş istifadə olunur. Tezauruslar ideoqrafik leksikoqrafiya sahəsində çoxillik tədqiqatların məhsuludur. Tezauruslar ideoqrafik lüğətlərin əsas xüsusiyyətlərini özündə toplayır. Hələ vaxtilə L.V.Şerba ideoqrafik lüğətlərin böyük nəzəri əhəmiyyətini qeyd etmişdi [Шерба, 1974]. Hazırda tezauruslar təkcə MT üçün deyil, həm də nəzəri semantikanın problemləri baxımından diqqəti cəlb edir. Tezauruslar dilin bütün leksik fondunun paradigmatik baxımdan sistemləşdirilməsinə imkan verir. Bu lüğətlərdə sözlərin semantik münasibətlərinin müəyyənləşdirilməsi mühüm rol oynayıır [Чижаковский, Беляева, 1983].

Tezaurus tipli semantik lüğətlərdə sözlərin məna əlaqələrini özündə əks etdirən məlumatlar yerləşdirilir. Tezauruslar – onu təşkil edən sözlərin səciyyəsi və sahəsi (peşə leksikası), iyerarxiyanın dərinliyi və istiqaməti, istifadə olunan paradigmatik münasibətlərin tipi və kəmiyyəti, eləcə də tərtibolunma üsullarına görə bir-birindən fərqlənirlər. Tezauruslar praktik zə-

rurət nəzərə alınmaqla əlavə paradigmatik münasibətlər hesabına zənginləşdirilə bilər:

Təcrübə göstərir ki, paradigmatik münasibətlərin sayının həddindən artıq çoxaldılması və iyerarxiyanın dərinləşdirilməsi tezaurusların həcmnin böyüməsinə və ondan istifadənin çətinləşməsinə aparıb çıxarır.

Tezaurusların həcmi müxtəlif ola bilər. Hazırda 200-dən 10000 terminə kimi sözlüyü olan tezauruslar tərtib olunmuşdur. Tezauruslar təkcə bir dili deyil, bir neçə dili də əhatə edə bilər. Hazırda mətnlərin avtomatik işlənməsi və MT sistemlərinin inkişafı ideoqrafik leksikoqrafiyanın, o cümlədən tezaurusların inkişafına da geniş meydan açmışdır. Azərbaycan dilçiliyində tezaurusların tərtibi ilə bağlı müəyyən işlər görülmüş, tikinti üzrə məlumat-axtarış tezaurusu hazırlanmışdır [Мамедова, 1984; Мамедова, Пинес, 1989].

Təqdim olunan işdə mətnin formal təhlili sistemində sintaktik blokun fəaliyyəti prosesində tezaurusların istifadəsi, məsələn, sözlərin çoxmənalılığı ilə əlaqədar belə lüğətlərin əhəmiyyəti göstərilmişdir.

Dilin riyazi-statistik, kompüter dilçiliyi metodları ilə tədqiqi sahəsində 2000-ci illərdən sonra da bir çox araşdırmalar aparılmış, dissertasiya işləri müdafiə olunmuş, maraqlı elmi nəticələr əldə edilmişdir [Əmirov, 2006; Əliyev, 2012; Fətullayev, 2006; Fətullayev R., 2010; Xəlili, 2009; Məmmədzadə, 2018; Məmmədov, 2003; Иващенко, 2009; Искендерова, 2008 və b.]. Son illərdə korpus dilçiliyi, təbii dilin emalı, koqnitiv dilçilik və süni intellekt problemləri çərçivəsində digər istiqamətlərdə tədqiqatlar prioritet hesab olunur.

III FƏSİL

MƏTNİN EMALI SİSTEMLƏRİ SÜNİ İNTELLEKT PROBLEMLƏRİ ÇƏRÇİVƏSİNDƏ

Mətnin emalı sistemlərində təhlil və sintez ən mühüm komponentlərdən hesab olunur. Nəzərdə tutmaq lazımdır ki, kompüter texnologiyasının köməyi ilə reallaşdırılan mətnin emalı sistemi sırf formal səciyyə daşıyır. Başqa sözlə, emal sisteminin əsasında formal əlamətlər və prinsiplər durur. Təbiidir ki, mətnin emalı sistemlərinin optimal fəaliyyətinin təmin edilməsi üçün araşdırma obyektı olan hər hansı konkret dilin formal qrammatikası hazırlanmalıdır. Əlbəttə, bu nəzəri və praktik (təcrübi) cəhətdən çox da asan məsələ hesab oluna bilməz. İlkin mərhələdə dilin tam formal qrammatikası işlənib hazırlanmasa da, mətnin kompüterlərdə emalı ilə əlaqədar qrammatikanın zəruri bölmələri formal təhlil və sintez üçün hazırlanmalıdır. Süni intellekt problemləri çərçivəsində təbii mətnin emalı sistemlərində istifadə olunan bəzi termin və anlayışlarla tanış olmaq faydalı olardı.

Nitqin sintezi. Nitqin sintezi parametrlərinə görə nitq siqnalları formalarının bərpası deməkdir. Dar mənada – çap mətninə görə səs siqnalının formalaşdırılması da sözün sintezi hesab olunur. Nitqin sintezi insan nitqinin texniki qurğular vasitəsilə süni yolla alınması ilə bağlı məsələdir. Nitq sintezatoru proqram təminatı və ya texniki qurğuların köməyi ilə mətnin nitqə çevrilməsini təmin edən sistemdir.

Bütün hallarda əgər informasiyanı alan insandırsa, nitqin sintezi tələb oluna bilər. Nitq sintezatorunun keyfiyyəti onun insan səsinə oxşarlığı, eləcə də başadüşülən olması ilə ölçülür. Bu da görmə qabiliyyəti zəif olan adamların kompüterdən isti-

fadəsinə imkan yaradır. Nitqin sintezi yazılı mətnləri (çap variantında) kompüterlərdə səsə çevirməyə və dinləməyə imkan verir. Nitqin ən sadə sintezi, verilənlər bazasında saxlanılan nitq hissələrinin (parçalarının) birləşdirilməsi yolu ilə alınır. Bu üsulla nitqin sintezi çox geniş yayılmış və bir çox təbii mətnin emalı sistemlərində yerlərdə istifadə olunmaqdadır. Onun tətbiq dairəsi çox genişdir:

- nitqin mətnə və ya məlumat koduna görə sintezi məlumatların axtarış-sorğu sistemlərində istifadə oluna bilər. Bu həm də oxuma qabiliyyəti bu və ya digər səbəbdən zəifləmiş insanların da kompüter texnologiyasından yararlanmaqla elmi-tədqiqat işləri ilə məşğul olmasına, elmi və bədii ədəbiyyatdan yararlanmasına imkanı yaradır;

- nəqliyyat növlərinin (avtobus, qatar, təyyarə, gəmi və s.) hərəkəti (yola düşməsi və mənzil başına çatması, bu və ya digər səbəbə görə gecikməsi barədə məlumatlar) ilə bağlı elanların verilməsində bu sistemin imkanlarından istifadə olunur;

- bir çox mürəkkəb texnoloji proseslərin informasiya baxımından idarəsində (hərbi və avia-kosmik qurğuların avtomatik idarəsində, robot texnikasında, insan – kompüter akustik dialoqunun təşkilində və s.) sistemin fayda və səmərəsi böyükdür;

- məişət cihazlarının, radio-televiziya, kompüter avadanlığının, geniş istifadə dairəsinə malik akustik qurğuların tətbiqi (elektron qurğularda səs effektinin təmini) ilə bağlı da bu sistemin ayrı-ayrı komponentlərindən uğurla istifadə edilir.

Məlumat axtarışı sistemlərində sintez çox vaxt ilkin elementlərinin qabaqcadan yazıldığı sözlüklər əsasında aparılır. Sintez elementlərinin ölçüsü ən azı bir sözdən ibarət olmalıdır. Belə sistemlərdə lüğət vahidlərinin sayı təxminən bir neçə yüz sözdən ibarət olur. Kompüterin yaddaşına qənaət məqsədi ilə nitq signalının sıxılması/kodlaşdırılması metodlarından istifadə

olunur. İndi dünyada istehsal olunan bütün qurğular (hərbi təyinatlı, məişət) səsli cavab sistemləri ilə təchiz olunur.

Bu, təbii mətnin formal emalı sistemlərinin əhəmiyyətini əks etdirən mühüm amildir. Məhz ona görə də bu sistemin inkişafı və getdikcə təkmilləşdirilməsi mühüm əhəmiyyət kəsb edir [Mahmudov, Fətullayev, 2016; Fətullayev, Mahmudov, 2015].

Nitqin çap mətninə görə sintezi nitq signalının bütün parametrlərinə görə idarəsini təmin edir və öncədən məlum olmayan mətnə görə nitqi yarada (doğura) bilər. Bu halda nitq signalının təhlili zamanı alınmış parametrlər və eləcə də, səslərin söz və frazalarda birləşməsi qaydaları yaddaşda saxlanır. Sintez analoq və rəqəmli texnikanın tətbiqi ilə nitq aktının modelləşdirilməsi yolu ilə alınır. Sintez prosesində parametr göstəriciləri və fonemlərin birləşmə qaydaları müəyyən vaxt intervalı ilə ardıcıl olaraq daxil edilir. Çap mətninə görə nitqin sintezi metodu akustik və linqvistik məhdudiyyətlər barədə biliklərin proqramlaşdırılmasına əsaslanır. Bu zaman kompüter bilavasitə insan nitqinin elementlərindən istifadə etmir. Belə sintez üsuluna əsaslanan sistemlərdə iki yanaşma olur.

Birinci yanaşma insanın nitqyaratma sistemində istiqamətlənmişdir və artikulyar sintez adı ilə tanınır. İkinci yanaşma qaydalar üzrə sintezdir. Belə sintezatorların təbiiliyi və aydınlıq dərəcəsi təbii nitqin xarakteristikaları ilə müqayisə olunacaq səviyyəyə çatdırıla bilər. Belə olduqda yüksək səviyyəli sintez əldə oluna bilər.

Təbii dilin öncədən yaddaşda saxlanılmış ayrı-ayrı hissələrinin istifadəsinə əsaslanan qaydalar üzrə nitq sintezi – rəqəmli formada nitq signalının manipulyasiyası qaydaları üzrə nitq sintezinin növlərindən biridir. Sintezin giriş elementlərindən asılı olaraq onun bir neçə növü göstərilir: mikroseqmentli, allofonik, difonlu, yarımhecalı, hecalı, ixtiyari ölçülü vahidlərdən sintez.

Çox vaxt yarımhecalar – yarım samit və ona bitişmiş sait-in yarısından ibarət seqmentlərdən istifadə olunur. Bu qayda ilə qabaqcadan verilməmiş mətndən nitq sintez etmək mümkündür. Bu zaman intonasiya və vurğu səciyyələrinin nəzərə alınması çətin olur. Belə sintezin keyfiyyəti təbii nitqin keyfiyyətinə uyğun gəlmir, difonların qovuşma sərhədlərində təhriflər yaranır. Öncədən yazılmış söz-formalardan nitq kompilyasiyaları da mətnin keyfiyyətli sintezini əldə etmək imkanı vermir. Məsələ orasındadır ki, sözün akustik və prosodik səciyyələri (uzanma və intonasiya) frazanın tipindən və sözün frazadakı yerindən asılı olaraq müxtəlif olur. Söz-formaların saxlanması üçün irihəcmli yaddaşdan istifadə belə vəziyyəti dəyişmir.

Predmet – istiqamətlənmiş sintezdə tam nitq məlumatının yaradılması üçün əvvəlcədən yazılmış söz və frazalar kompilyasiya edilir. Belə sintez mətn sisteminin müxtəlifliyinin müəyyən mövzu və ya sahə ilə məhdudlaşdırılması yolu ilə əldə edilir. Daha çox avtobus, qatar, təyyarə və gəmilərin yola salınması və mənzil başına çatması, onların yenidən əks istiqamətdə hərəkəti ilə bağlı axtarış-sorğu sistemlərində istifadə olunur. Bu qəbildən olan axtarış-sorğu sistemləri həm yazılı, həm də şifahi sorğuları adekvat cavablandırmaq imkanına malik olur. Hava haqqında məlumat bürolarında da analogi üsullardan istifadə olunur. Bu texnologiyanın tətbiqi və istifadəsi sadədir və kommersiya məqsədləri ilə uzun müddətdir ki, istifadə olunur. Bu texnologiya danışan saatlar və kalkulyatorlar kimi elektron cihazların istehsalında istifadə olunur. Onların istifadə olunduğu konkret sahələrdə məhdud sayda söz və ifadələrdən istifadə olunduğu üçün bu sistemlərdə səsənmə keyfiyyəti yüksək ola bilər, çünki həmin sözlər qabaqcadan seçilir və yaddaşa daxil edilir. Qabaqcadan təyinatı məlum olan konkret elektron qurğular üçün nəzərdə tutulmuş sözlük başqa məqsədlər üçün istifadə oluna bilməz və ona görə də geniş tətbiq dairəsinə malik deyildir. Başqa sözlə, qatarların yola salınması üçün proq-

ramlaşdırılmış sözlük hava haqqında məlumat və ya danışan saatlar üçün yaramır. Hər bir qurğu və sistem üçün hər şey yenidən seçilməli və məqsədyönlü şəkildə qurulmalıdır.

Belə sistemlərin qurulması işlərinə çoxdan başlanılıb. XVIII əsrin sonlarında Rusiya Elmlər Akademiyasının həqiqi üzvü, Danimarka alimi Xristian Kratsenşteyn beş uzun sait (a, e, i, o, u) tələffüz etməyə imkan verən insan nitq aktının modelini yaratmışdı. Model hava axını ilə hərəkətə gətirilib vibrasiya edən diltiklərin köməyi ilə sait səsləri doğuran müxtəlif formalı akustik rezonatorlardan ibarət idi. 1778-ci ildə Avstriya alimi Volfanq fon Kampelen dil və dodaq modellərini əlavə etməklə Kratsenşteynin modelini tamamladı və müəyyən səs və onun kombinasiyalarını doğura bilən akustik-mexaniki danışan maşının təqdimatını keçirdi. Əl ilə idarə olunan xüsusi qurğunun köməyi ilə spesifik səslər tələffüz oluna bilirdi. 1837-ci ildə Çarlz Uitstoun saitləri və əksər samitləri tələffüz edə bilən təkmilləşdirilmiş mexaniki maşın variantını yaratdı. 1846-cı ildə Jezef Faber tək-cə nitqi deyil, mahnını da sintez edə bilən – Euphonia danışan orqanını təqdim etdi.

XIX əsrin sonlarında Aleksandr Bell konstruksiyalarına görə Uitstounun maşınına çox oxşar olan xüsusi “danışan” maşın modeli yaratdı. Bütün haqqında danışdıqlarımız mexaniki əl qurğuları kimi səciyyələndirilə bilərdi.

XX əsrdə elektrik maşınları erası başladı və alimlər səs dalğaları generatorundan istifadə etmək və onun bazasında alqoritmik model qurmaq imkanı əldə etdilər.

1950-ci illərin sonunda hesablama texnikası bazasında ilk nitqin sintezi sistemləri yaranmağa başladı, ilk “mətn-nitq” sintezatoru isə 1968-ci ildə yaradılmışdır.

Azərbaycanda da bu sahədə tədqiqatlar aparılmış və uğurlu nəticələr əldə olunmuşdur. Respublikamızda bu sahədə aparılan tədqiqatlar kompüter texnologiyasının coşqun inkişafı dövrünə təsadüf edir. Ona görə də, Azərbaycanda mexaniki əl

qurğuları ilə səs sintezi sistemlərinin yaradılmasına ehtiyac duyulmamışdır. Respublikamızda informasiya-kommunikasiya texnologiyalarının bazasında bu istiqamətdə linqvistik texnologiyaların yaradılması işlərinə rəvac verilmişdir.

Dilmanc layihəsi çərçivəsində Azərbaycanda ilk dəfə olaraq mətnlərin insan səsi ilə səsləndirilməsi məqsədi ilə kompüter proqramı (TTS – Text-To-Speech) yaradılması ilə əlaqədar Azərbaycan dili fonemlərinin akustik və anatomik-fizioloji xüsusiyyətləri ətraflı təhlil olunmuş, fono-akustik xüsusiyyətləri əks etdirən cədvəl hazırlanmışdır. Alınmış nəticələr Azərbaycan dilində mətnlərin səsləndirilməsi sisteminin yaradılmasında və təkmilləşdirilməsində istifadə edilmişdir.

Məlum olduğu kimi, dildə olan bütün vahidlər – sözlər, söz birləşmələri, cümlələr dilin ən kiçik vahidi olan fonemlərdən təşkil olunmuşdur. Bu baxımdan fonemlər dil toxumasının ən mühüm hüceyrələri hesab edilir. Lakin, fonemlər təkcə dilçilik baxımından – fonoloji baxımdan deyil, həm də akustik, fizioloji, statistik, fiziki baxımdan tədqiq olunur və olunmaqdadır. Son zamanlar səsle idarə olunan qurğuların, şifahi səsle işləyən tərcümə sistemlərinin yaranması və bu sistemlərin təkmilləşdirilməsi sahəsində aparılan intensiv işlər ən kiçik danışıq vahidi olan fonemlərin hərtərəfli öyrənilməsinə marağı artırmışdır. Ayrılıqda heç bir məna yükü daşımayan, sözlərin bir-birindən fərqlənməsinə xidmət edən, ən kiçik dil vahidi kimi daha kiçik hissələrə bölünə bilməyən fonemlər fono-akustik xüsusiyyətlərinə görə dəqiq elmlər kateqoriyasına yaxınlaşır. Təsadüfi deyil ki, fonemlərin akustik xüsusiyyətləri, səsin uzunluğu, qısalığı, gücü, ucalığı və bir çox başqa fizioloji-fiziki cəhətləri xüsusi müasir avadanlıq və cihazlarla öyrənilir. Azərbaycanda Dilmanc layihəsi çərçivəsində hazırlanmış TTS sistemi (Text-To-Speech) uğurla istifadə olunmaqdadır.

Nitqin tanınması – nitq signalının rəqəmli informasiyaya çevrilməsi prosesidir. Nitqin sintezi buna əks proses hesab olu-

nur. Yəni, əgər nitqin sintezi prosesində tədqiqatçı insan nitqini təqlid edən, doğuran bir sistem yaratmağa cəhd edirdisə, nitqin tanınması zamanı nitqi rəqəmli informasiyaya, sonradan isə mətnə çevirməyə cəhd edir.

Nitqin tanınması məqsədi ilə yaradılmış ilk qurğu 1952-ci ildə yaradılmışdır. Həmin qurğu insanın tələffüz etdiyi rəqəmləri tanıya bilirdi. 1964-cü ildə Nyu-Yorkda kompüter texnologiyaları yarmarkasında IBM Shoebox qurğusu təqdim olunmuşdur [[Распознавание речи. ru.wikipedia.org/wiki/ Распознавание_речи](http://Распознавание_речи.ru.wikipedia.org/wiki/Распознавание_речи)].

Nitqin tanınması ilə bağlı kommersiya proqramları ötən əsrin 90-cı illərinin əvvəllərində yaranmışdır. Həmin proqramlardan əlilliyi olan, əlləri zədəli olduğundan iri həcmli mətnləri çap etmək imkanından mərhum insanlar istifadə edirdilər (daha doğrusu, bu proqramlar ilk növbədə onlar üçün nəzərdə tutulmuşdu). Bu qəbildən olan proqram vasitələri (məsələn, Dragon Naturally Speaking, Voice Navigator) istifadəçinin səsini avtomatik mətnə çevirir və onu əl ilə işləməkdən (yazmaqdan) azad edirdi. Etiraf etmək lazımdır ki, bu proqramlar tələb olunan, istənilən, arzu olunan səviyyədə deyil, lakin onun keyfiyyətinin yüksəldilməsi, yaxşılaşdırılması üçün səy göstərilir.

Mobil qurğuların yaddaşının, hesablama imkanlarının artırılması, daha da gücləndirilməsi onlar üçün də nitqin tanınması funksiyasında proqramların yaradılması imkanı yaratdı (və ya kompüter üçün nəzərdə tutulmuş proqram təminatının mobil qurğulara da tətbiqi, uyğunlaşdırılması imkanı yarandı). Bu qəbildən olan proqramlara nümunə kimi səs kəməyi ilə bir çox tətbiqi əlavələrlə işləməyə imkan verən Microsoft Voice Command əlavəsini göstərmək olar. Bu proqramla pleyerdə musiqini canlandırmaq və ya yeni sənəd yaratmaq mümkündür.

Nitqin tanınması proqramı müxtəlif biznes sahələrində də öz tətbiqini tapır, məsələn, poliklinikada həkim diaqnozu səslə, nitqlə ifadə edir və həmin an diaqnoz elektron kartda öz əksini

tapır. Daha həmin diaqnozu klaviaturada yığıb kompüterə daxil etmək lazım gəlmir.

Nağıllar gerçək olur. Vaxtilə nağıllarda eşitdiyimiz “sim-sim, qapını aç!” əfsanəvi deyimi reallığa çevrilir. Səslə idarə olunan qurğular indi nağıl, əfsanə deyil, günümüzün reallığıdır. Səsin köməyi ilə müəyyən istehsalat sahəsində, hərbi məqsədlərlə texnoloji (və ya qeyri-texnoloji) proseslərin idarə olunması artıq möcüzə hesab olunmur. Heç bir toxunuş, təmas olmadan uzaqdan insan səsi ilə bu və ya digər hərəkətin icrası bir çox sahələrdə öz uğurlu tətbiqini tapmışdır. İndi insanın işığı söndürmək, suyu bağlamaq, qapını açmaq (və ya bağlamaq) tələbini (əmrini) yerinə yetirə biləcək qurğular mövcuddur. Bu sistem o qədər təkmilləşdirilmişdir ki, insan səsini tanıyır və tanımadığı, ona yad olan səslə deyilmiş əmri, xahişi icra etmir, evin qapısını yalnız ev sahibinin səsi ilə deyildiyi zaman icra edir. Bu insan nitqinin, səsinin tam dəqiqliklə tanınması proqramları ilə idarə olunan qurğuların köməyi ilə həyata keçirilir. Başqa sözlə, nağıllarda olduğu kimi hər “sim-sim, qapını aç!” deyənə qapı açılır. Əks təqdirdə, bu səslə idarə olunan sistemin əhəmiyyətini azaltmış olardı.

Son zamanlar mobil telefonlarda da nitqin avtomatik sin-tezi və tanınması sistemindən istifadə olunmağa başlanılmışdır. Qeyd olunduğu kimi, inkişaf etmiş, təkmilləşdirilmiş müasir tanınma sistemləri diktorlardan asılı deyil, belə ki, istənilən insanın səsi ayırd edilir və tanınır.

Səssiz müraciət interfeyslərinin inkişafı (silent speech interfaces, SSI) nitqin tanınması texnologiyasında növbəti addım hesab olunur. Bu qəbildən olan nitqin emalı sistemləri artikuliyasiyanın ilkin mərhələlərində nitq siqnallarının alınması və emalına əsaslanır. Nitqin tanınmasının bu inkişaf mərhələsində müasir tanınma sistemlərinin iki mühüm çatışmazlığı özünü göstərir: səsə həddindən artıq həssaslıq, eləcə də tanınma sistemə müraciət zamanı nitqin aydın və dəqiq olması zərurəti.

SSİ sisteminə əsaslanan yanaşma elə yeni sensorların olmasını tələb edir ki, akustik siqnalların emalı prosesində əlavə olaraq küyün təsirinə məruz qalmasın [[Распознавание речи. ru.wikipedia.org/wiki/ Распознавание_речи](http://ru.wikipedia.org/wiki/Распознавание_речи)].

Nitqin tanınması sistemi aşağıdakı kimi təsnif olunur:

- sözlüyün həcminə görə (məhdud və ya irihəcmli sözlük);
- diktordan asılılığına görə (diktordanasılı və diktordan asılı olmayan sistemlər);
- nitq tipinə görə (bitişik və ya ayrı nitq);
- təyinatına görə (diktə sistemi, əmr sistemi);
- istifadə olunan alqoritmə görə (neyron şəbəkələr, gizli Markov modeli, dinamik proqramlaşdırma);
- struktur vahid tipinə görə (frazalar, sözlər, fonemlər, difonlar, allofonlar);
- struktur vahidlərin ayrılması prinsipinə görə (şablona görə tanıma, leksik elementlərin ayrılması).

Nitqin avtomatik tanınması sistemləri üçün maneədən qorunma iki mexanizmin tətbiqi ilə təmin olunur:

- akustik siqnalın təhlili bazasında eyni nitq elementlərinin ayrılması üçün bir neçə paralel işləyən üsullardan istifadə;
- nitq axınında sözün seqment (fonem) kimi və tam qavranmasından paralel, qeyri-asılı istifadə [[Распознавание речи. ru.wikipedia.org/wiki/ Распознавание_речи](http://ru.wikipedia.org/wiki/Распознавание_речи)].

Nitqin qavranması modelində nitq siqnallarının emalı alqoritmı insanın indiyə qədər istifadə etdiyi, yararlandığı anlayış və münasibətlərə əsaslanır. Nitqin tanınması sistemləri mövcud tanınma sistemləri əsasında qurulur. Həmin istifadə olunan metod və alqoritmləri aşağıdakı siniflərə ayırmaq olar:

- etalonla müqayisə əsasında nitqin tanınması metodlarının təsnifi;

- dinamik proqramlaşdırma – müvəqqəti dinamik alqoritmlər (Dynamic Time Warping);
- kontekstdən asılı təsnifat – onun reallaşması zamanı nitq axınından ayrı-ayrı leksik elementlər, sonradan heca və morfemlər kimi birləşən fonem və allofonlara ayrılırlar;
- Bayeso diskriminant təhlil metodu (Bayesian discrimination);
- gizli Markov modelləri (Hidden Markov Model);
- neyron şəbəkələr (Neural networks).

Nitqin avtomatik emalı sisteminin statistik arxitekturası:

- səs-küy təmizləyən modul və faydalı siqnalın ayrılması;
- akustik model – səs səviyyəsində oxşarlıq baxımından səs seqmentinin tanınmasını qiymətləndirməyə imkan yaradılması; hər bir səs üçün qabaqcadan onun nitqdə tələffüzünü təsvir edən mürəkkəb statistik modelin qurulması;
- dil modeli – ən çox ehtimal olunan söz ardıcılığını müəyyənləməyə imkan verir. Dil modelinin mürəkkəbliyi daha çox konkret dildən asılıdır. Belə ki, ingilis dili üçün statistik modellərdən (N-qrammalar) istifadə etmək kifayətdir. Rus dilinin də aid edildiyi yüksək flektivli dillər üçün (belə dillərdə bir sözün çoxlu sayda müxtəlif formaları olur) tək-cə statistikadan istifadə olunmaqla qurulmuş model effekt vermir, sözlər arasında statistik əlaqələri etibarlı şəkildə qiymətləndirmək üçün verilənlərin sayı həddindən artıq çox olmalıdır. Ona görə də rus dilinin qaydalarını, nitq hissələri və sözlərin formaları haqqında informasiyaları və klassik statistik modeli özündə birləşdirən hibrid dil modelləri tətbiq edilir;
- dekoder – akustik və dil modellərinin tanınması gedişində onları birləşdirməklə alınan məlumatları tərkibinə alan, bitişik nitqin tanınmasının son nəticəsi olan sözlərin ən ehtimali ardıcılığını müəyyənləyən tanıma sisteminin proqram

komponentidir [Распознавание речи. ru.wikipedia. org/wiki/Распознавание_речи].

Tanıma mərhələləri:

- nitqin emalı nitq signalının keyfiyyətinin qiymətləndirilməsindən başlanır;

- qiymətləndirilmənin nəticəsi tanıma üçün zəruri olan hesablama modulunu idarə edən akustik adaptasiya moduluna daxil olur;

- signalda nitq olan sahələr ayrılır və nitq parametrlərinin qiymətləndirilməsi baş verir. Sintaktik, semantik və praqmatik təhlil üçün fonetik və prosodik ehtimali səciyyələri ayrılması baş verir (nitq hissələri, sözün forması və sözlər arasında statistik əlaqələr haqqında məlumatların qiymətləndirilməsi);

- sonra nitq parametrləri tanınma sisteminin əsas blokuna – dekoderə daxil olur. Akustik və dil modellərində saxlanan informasiya ilə daxil olan nitq axınına qarşılaşdıran bu komponent tanınmanın son nəticəsi olan sözlərin ən ehtimali ardıcılığını müəyyənləşdirir.

Rəqəm kanalları üzrə nitqin keyfiyyət parametrləri:

- nitqin heca aydınlığı;
- nitqin fraza aydınlığı;
- nitqin keyfiyyəti etalon nitq aktının keyfiyyəti ilə müqayisədə;

- real iş şəraitində nitq keyfiyyəti.

Əsas anlayışlar:

- nitqin aydınlığı – düzgün qəbul olunmuş nitq elementlərinin (səslər, hecalar, sözlər, frazalar) verilmiş elementlərin ümumi sayında faizlə ifadə olunmuş nisbi miqdarı;

- nitqin keyfiyyəti – nitqin verilməsinin sınaqdan keçirilən sistemində nitqin səslənməsinin subyektiv qiymətləndirilməsini səciyyələndirən parametr;

- nitqin normal tempi – nitqin, frazanın orta uzunluğunun 2,4 s. sürətlə tələffüzü;

- sürətləndirilmiş nitq tempi – nitqin nəzarət frazasının orta uzunluğunun 1,5-1,6 s. sürətlə tələffüzü;

- danışanın nitqinin tanınması – səsin səslənməsinin dinləyiciyə əvvəldən tanış olan konkret adamın səsi ilə qarşılaşdırılmasının mümkünlüyü;

- məna aydınlığı – nitqin məlumat məzmununun verilməsinin düzgünlük dərəcəsi;

- inteqral göstərici – qəbul edilən nitqlə bağlı dinləyicinin ümumi təəssüratını səciyyələndirən göstərici [[Распознавание речи. ru.wikipedia.org/wiki/ Распознавание_речи](http://ru.wikipedia.org/wiki/Распознавание_речи)].

Nitq əmrləri istifadəçini əmrlərin daxil edilməsi üçün sensor və digər metodlardan istifadə olunması zərurətindən azad edir:

- səsli idarə;
- səsli əmrlər;
- mətnin səsli daxil edilməsi;
- səsli axtarış.

Mobil qurğularda nitqin tanınması texnologiyasının uğurlu tətbiqinə aşağıdakıları göstərmək olar:

- ünvanların səsli “Яндекс”, “навигатор” sistemlərinə daxil edilməsi;

- Google Now-da səsli axtarış.

Mobil qurğulardan əlavə nitqin tanınması texnologiyası müxtəlif biznes sferalarında geniş tətbiq olunur:

- səsli xidmət sistemi yaratmaqla gələn və gedən zənglərin avtomatik emalını təmin etmək;

- sorğu məlumatı əldə etmək, məsləhət, xidmət və əmtəə sifarişi, mövcud xidmətlərin parametrlərinin dəyişdirilməsi, sorğu, anketləşdirmə keçirmək, məlumat toplamaq, məlumatlandırma və başqa ssenarilər;

- “Ağıllı ev” layihəsi. “Ağıllı ev” səsli interfeys sisteminin idarəsi;
- məişət texnikası və robotlar: elektron robotların səsli interfeysi; məişət texnikasının səsli idarəsi və s.;
- desktop və noutbuklar: kompüter oyunları və proqramların səsli daxil edilməsi;
- avtomobillər: avtomobil salonunda səsli idarə, məsələn, naviqasiya sistemini səsli idarə;
- məhdud imkanlı adamlar üçün sosial servislər [Распознавание речи. ru.wikipedia.org/wiki/Распознавание_речи].

Bu sahədə respublikamızda da müəyyən işlər görülmüşdür. Vaxtilə Rəhbərlik və Yüksək Texnologiyalar Nazirliyində fəaliyyət göstərmiş “Dilmanc” layihəsinin ən uğurlu işlərindən biri də şifahi nitqi mətnə çevirən “Dilmanc İmla” sistemi hesab olunurdu. “Dilmanc İmla” Azərbaycan dilində nitqi avtomatik olaraq yazıya çevirən proqramdır. Həmin sistemin yeni təkmilləşdirilmiş versiyası istifadəyə verilmişdir [<http://dilmanc.az/en/node356>]

Əvvəlki versiyadan fərqli olaraq yeni versiya daha keyfiyyətlidir və bu versiyada smartfon tipli mobil telefonu mikrofon kimi istifadə etmək imkanı var. Proqramın istifadə qaydası çox sadədir. Həm kompüterə, həm də telefona “Dilmanc İmla” proqramını yüklədikdən sonra kompüterə yazılmış proqramın təklif etdiyi birləşmə kodu telefona daxil edilərək kompüter və telefon arasında internet üzərindən əlaqə yaradılır. Bundan sonra telefon mikrofon kimi istifadə edilə bilər. Kursoru kompüterdə mətn yazmaq üçün nəzərdə tutulan pəncərəyə qoymaqla işə başlamaq olar. Telefona deyilən imla avtomatik olaraq mətnə çevrilərək kompüterə yazılır.

Proqramın ən son versiyası şifahi danışığı yazıya çevirərkən durğu işarələrini avtomatik qoyur və cümlədəki xüsusi adları böyük hərflə yazır. Bütün mövcud nitqin tanınması sistemləri şifahi danışığı yazıya çevirərkən durğu işarələrinin ha-

rada olması, nitqə qulaq asaraq hansı sözün ilk hərfinin böyük hərflə yazılmalı olduğundan məlumatsız olurlar. Problemi həll etmək üçün azərbaycandilli saytlardan mətnlər köçürülərək məşin öyrənməsi alqoritmləri ilə analiz edilmiş, durğu işarələrinin yerini və xüsusi isimləri avtomatik müəyyən edən statistik modellər yaradılmışdır.

“Dilmanc” layihəsi təbii dillərin emalı (Natural Language Processing – NLP) sahəsində ən müasir və effektiv yanaşmalardan olan Dərin Neyron Şəbəkələr (Deep Neural Networks – DNN) nəzəriyyəsini öz proqramlarında tətbiq etmişdir. Dərin Neyron Şəbəkələr yanaşması “Dilmanc” layihəsində ilk olaraq Azərbaycan dilində nitqin tanınması sisteminə tətbiq edilmişdir və bu yanaşmanın köməyi ilə hazırlanmış “Dilmanc Translator” proqramının yeni versiyasında Azərbaycan dilində şifahi nitqi daha keyfiyyətli mətnə çevirərək tərcümə etmək mümkündür.

Əslində, Neyron Şəbəkələr yanaşması yeni deyildir. Amma effektiv öyrətmə alqoritmlərinin olmaması, həmçinin hesablaşma avadanlıqlarının zəif olması səbəbindən kifayət qədər keyfiyyətli sistemlərin yaradılması mümkün olmurdu. Son illər həm yeni alqoritmlərin yaradılması, həm də çoxnüvəli qrafik prosessorların istehsalı ilə elmi tədqiqatların nəticələrinin tətbiqi üçün imkanlar yarandı. “Dilmanc” layihəsi nitqin tanınması sisteminin akustik modelinin öyrədilməsi üçün 768 nüvəli qrafik prosessordan, dil modelinin öyrədilməsi üçün isə hər birinin 8 nüvəsi olan 30 serverdən ibarət paylanmış hesablaşma klasterindən istifadə etmişdir.

“Dilmanc” layihəsi çərçivəsində hazırlanmış mobil telefonlar üçün tərcümə proqramının İOS əməliyyat sistemində çalışan versiyası da istifadəçilərə təqdim olunmuşdur.. İstifadəçilər bu versiyanı öz iPhone və iPad qurğularına App Store-da “Dilmanc” axtarmaqla yükləyə bilirlər [<http://www.dilmanc.az/az>].

Qeyd etmək lazımdır ki, mobil tərcüməçi proqramının Android əməliyyat sistemi üçün olan versiyası Azərbaycanda və ölkə xaricində çox populyar proqram məhsuludur.

“Google analytics”-ə görə bu proqramdan təkcə 2015-ci ilin yanvar ayının 01-dən 22-nə kimi 40 minə yaxın insan tərəfindən dünyanın 65 ölkəsində 360 min dəfə istifadə edilmişdir. Bu ölkələr arasında ABŞ, Almaniya, Rusiya, Ukrayna, İsveç və digər ölkələr vardır.

“Dilmanc” layihəsinin ən yeni proqram təminatlarından biri də “Dilmanc çağrı mərkəzi (Dilmanc call centre)” hesab olunur. Bu sistem Azərbaycanda insan nitqini başa düşərək ona adekvat cavab verən ilk belə sistemdir. Bu sistemin tətbiq edilməsi mövcud çağrı mərkəzlərinin məsrəflərini bir neçə dəfə azaltmağa imkan verir [Dilmanc. <http://az.wikipedia.org/wiki/dilmanc>].

Təbii dillərin emalı (NLP – Natural Language Processing) hazırda sürətlə inkişaf edən və bütün həyat sferalarına nüfuz edən informatika elminin böyük maraq doğuran mühüm sahələrindəndir. NLP həm elmi-nəzəri, həm də praktik əhəmiyyəti olan bir istiqamətdir. NLP nəzəriyyə ilə praktikanın bir vəhdət halında təzahürü kimi də səciyyələndirilə bilər. Digər elm sahələrində bəzən bu və ya digər nəzəri ideyanın, araşdırmanın praktikada tətbiqi yollarını, üsullarını axtarıqsa, informatika elmində bu məsələ fərqli şəkildə özünü büruzə verir. İnformatika elmi çərçivəsində tədqiq olunan hər hansı mövzu, ideya, əslində dövrün, cəmiyyətin sosial sifarişi kimi dəyərləndirilə bilər.

Məlum olduğu kimi, təbii dillərin emalı sisteminin yaradılması üçün bir neçə elmin sintezindən istifadə etmək lazım gəlir. İnformatikanın bu yeni sahəsində riyazi modelləşdirmə, statistik ümumiləşdirmə, alqoritmləşdirmə və proqramlaşdırma ilə yanaşı qrammatika da mühüm rol oynayır. Çünki, söhbət hər şeydən əvvəl dil materialının, daha dəqiq

ifadə etsək, təbii dil materialının emalından gedir. Məhz buna görə də, emal prosesində dil vahidləri arasında mövcud əlaqə və münasibətləri tənzimləyən qayda və qanunları, onların quruluş xüsusiyyətlərini, söz birləşməsi, cümlə əmələ gətirmək imkanlarını özündə ehtiva edən mükəmməl qrammatikasız keçinmək qeyri-mümkündür.

Məlumdur ki, ənənəvi normativ qrammatikalar təbii dillərin emalı üçün nəzərdə tutulmamışdır. Belə olduqda onlar təbii dillərin emalı sistemlərində nə dərəcədə istifadə oluna bilirlər və bu sistemləri qane etməyən boşluqları necə doldurmaq olar?

Əlbəttə, sual kifayət qədər ciddidir və onu hərtərəfli şərh etmək, qarşıya çıxan bütün problemlərə birmənalı cavab vermək çətindir. NLP sistemlərinə elə bir qrammatika lazımdır ki, həm hər hansı konkret təbii dili tam ehtiva və təmsil etsin, eyni zamanda sistemin qurulması üçün zəruri qayda və informasiyaları təqdim edə bilsin. Məlum olduğu kimi, qrammatika söz və cümlə, sözlərin müəyyən şəkilçilər (qrammatik şəkilçilər) qəbul edərək dəyişməsi, cümləni təşkil edən komponent kimi birləşməsi qaydalarını öyrənir. Başqa sözlə, qrammatikanın obyektivi söz, şəkilçi (leksik, qrammatik) və cümlədir. Dilçilik ədəbiyyatında qrammatikanın müxtəlif növləri göstərilir: assosiativ qrammatika, xarici qrammatika, daxili qrammatika, təsviri qrammatika, deskriptiv qrammatika, tarixi qrammatika, qarşılaşdırma qrammatikası, müqayisəli qrammatika, formal qrammatika və s. [Dilçilik ensiklopediyası, I cild, 2006, s.169-171]. Bu siyahıya son zamanlar geniş müzakirə olunan və inkişaf edən funksional qrammatika, maşın qrammatikası və formal qrammatikanı da əlavə etmək olar. NLP sistemlərinin yaradılması baxımından son sadalananlar daha önəmli hesab oluna bilər.

Ənənəvi qrammatikanın müasir elmi-texniki tərəqqi və inkişafı ayaqlaşma bilməməsi bir çoxları tərəfindən qeyd olun-

muş, konkret faktlar göstərilmişdir. İnformasiya-kommunikasiya texnologiyalarının sürətli inkişafı əsrində ənənəvi dilçiliyə də qeyri-ənənəvi baxış formalaşmaqdadır [Mahmudov, 2006].

Formal təhlil sistemlərində də ənənəvi normativ qrammatikalar özünü doğrultmur: “Formal təhlil sistemlərində hər hansı mətnin tam təhlili, onun ayrı-ayrı cümlələrə ayrılması, söz-formaların təhlili, izahı ənənəvi dilçilikdə insan-dilçi tərəfindən aparılan əməliyyatlardan köklü surətdə fərqlənir. Deməli, buradan belə bir nəticə çıxır ki, kompüterin əsaslandığı, istifadə etdiyi qrammatika və onun kateqoriyaları da ənənəvi qrammatikadan fərqli olmalıdır. Daha doğrusu, bu elə bir qrammatika olmalıdır ki, onun köməyi ilə sırf formal əlamətlər və prinsiplər əsasında sistem üçün zəruri məlumatı əldə etmək imkanına malik olsun” [Mahmudov, 2002, s.81].

Həmin qrammatikanı formal qrammatika, maşın qrammatikası kimi də adlandırırlar.

Formal qrammatika və ya maşın qrammatikası NLP sistemlərinin yaradılması üçün zəruri informasiyanı verə bilirmi və ya həmin sistemlər üçün kafi şərtləri ödəyirmi?

Məlum olduğu kimi, ənənəvi normativ qrammatikalar daha çox insan təfəkkürünə, düşüncə tərzinə hesablanmış qaydaları əhatə edir. Bu qrammatika çərçivəsində aparılmış təhlil daha çox semantikaya əsaslanır, sözlərin məna cəhətləri və əlaqələri ön plana çəkilir. Təbii ki, daha çox formal əlamətlərə və xüsusiyyətlərə əsaslanan təbii dillərin emalı sistemləri ənənəvi normativ qrammatikalar əsasında qurula bilməz. Bu sistemlərdə (NLP sistemlərində) formal kateqoriyalar və meyarlar ön plana çəkilməlidir.

Məlum olduğu kimi, Azərbaycan dilində üçüncü növ təyini söz birləşmələrinin hər iki tərəfinin morfoloji əlaməti olur – birləşmənin birinci tərəfi həmişə yiyəlik hal, ikinci tərəfi

isə mənsubiyyət şəkilçisini qəbul edir [Müasir Azərbaycan dili, III cild, 1981, s.118]. Aşağıdakı nümunələri nəzərdən keçirək:

- Kəndimizin şöhrəti bütün dünyaya yayılmış xalçaları muzeyləri bəzəyir.

- Xanəndənin səsi uzaqlara yayılan zənguləsi hamını heyran etmişdi.

- Səhranın qumu hər tərəfə səpələyən küləyi dayanmaq bilmirdi.

İlk baxışdan elə təsəvvür yarana bilər ki, bu nümunələrdə olan “kəndimizin şöhrəti”, “xanəndənin səsi”, “səhranın qumu” birləşmələri üçüncü növ ismi birləşmələrdir. Doğrudan da, bu söz birləşmələrinin birinci tərəfi yiyəlik hal, ikinci tərəfi isə mənsubiyyət şəkilçisi qəbul edir və beləliklə, formal olaraq üçüncü növ təyini söz birləşmələrinin tələblərinə cavab verir. Əslində isə, bu cümlələrdə üçüncü növ ismi birləşmələr bunlardır: “kəndimizin xalçaları”, “xanəndənin zənguləsi”, “səhranın küləyi”. Göstərilən misallarda “şöhrəti bütün dünyaya yayılmış”, “səsi uzaqlara yayılan”, “qumu hər tərəfə səpələyən” ifadələri isə fəli sifət tərkibi ilə ifadə olunmuş təyinlərdir. Deyilənlərdən belə bir nəticəyə gələ bilərik ki, sırf formal meyarlara və əlamətlərə görə ismi birləşmələri seçib ayırmaq və ya düzgün müəyyənləşdirmək həmişə mümkün olmur.

İkinci növ ismi birləşmələrin formal təyinin də analoji hallar müşahidə olunur:

- Dəvə yunu baha qiymətə satılır.
- Dəvə yunu tapdalayıb keçdi.

Hər iki cümlədə olan “dəvə yunu” birləşməsi formal baxımdan tam eynidir, heç bir cəhətdən fərqlənmir. Əslində isə birinci nümunədə “dəvə yunu” ikinci növ ismi birləşməsi olsa da, ikinci misalda söz birləşməsi əmələ gətirmir, fərqli cümlə üzvlərini göstərir (mübtəda və xəbər). Birinci növ təyini söz

birləşmələri ilə bağlı da bu qəbildən olan nümunələr göstərmək olar. Bütün bunlar bir daha sübut edir ki, sırf formal əlamətlər belə həmişə gözlədiyimiz nəticəni vermir. Yəni, indiyə qədər nəzərdən keçirdiyimiz qrammatikaların heç biri NLP sistemlərini qane edəcək səviyyədə hesab oluna bilməz [Mahmudov, Fəttullayev, 2016; Fəttullayev, Mahmudov, 2015].

NLP sistemlərinin yaradılması sahəsində aparılan tədqiqatlar göstərdi ki, ənənəvi (normativ) qrammatikalar bu sistemlərin yaradılması üçün zəruri olan məlumatı vermir və ya vermək iqtidarında deyil. Belə sistemlərin yaradılması tamamilə yeni bir yanaşma tələb edir. Məsələn, qrammatik qaydalara əsaslanaraq yaradılmış maşın tərcüməsi (MT) sistemləri (Rule Based Machine Translation – RBMT) əksər hallarda aşağı keyfiyyətli tərcümə verdiyindən, son vaxtlar bir çox MT sistemlərinin yaradılmasında həmin yanaşmadan istifadə olunmur.

Məsələ orasıdır ki, ənənəvi dilçilik NLP sahəsindəki yeni araşdırmalarla ayaqlaşma bilmir, qrammatikanın bu tədqiqatlara uyğunlaşdırılması və istiqamətləndirilməsi günün vacib məsələlərindəndir. Belə ki, məsələn, Azərbaycan dilinin müasir qrammatikasında yer alan əksər qaydalar kafi şərtlər şəklində deyil, zəruri şərtlər şəklində verilir [Müasir Azərbaycan dili, 1981; Seyidov, 1992].

Aşağıdakı nümunələrdə qrammatik qaydalara əsaslanmaqla ismi birləşmələrin formal baxımdan müəyyənləşdirilməsi qaydaları nəzərdən keçirilir.

“Mobil telefon mağazası” və “məşhur neft ustası” söz birləşmələrinin birincisində yanaşı gələn mobil və telefon sözləri birləşmə əmələ gətirsə də, onunla eyni qrammatik əlamətlərə malik olan ikinci söz birləşməsində məşhur və neft sözləri birləşmə əmələ gətirmir.

Başqa bir misal: “Toyuq dəni dağıldı” cümləsində toyuq və dəni sözləri 2-ci tip ismi birləşmə əmələ gətirsə də, “Toyuq

dəni dənlədi” cümləsində həmin sözlər birləşmə əmələ gətirir. Yəni, xəbərin məlum və ya məchul növdə olması birləşmənin mahiyyətə dəyişməsinə, yeni qrammatik məzmun kəsb etməsinə səbəb olur.

Daha bir misal: “Məktəbin şagirdləri gəzintiyə getdilər” cümləsində məktəbin və şagirdləri söz-formaları üçüncü növ ismi birləşmə əmələ gətirir: məktəbin şagirdləri. Lakin “Məktəbin şagirdləri gəzintiyə aparan avtobusu çox rahat idi” cümləsində həmin söz formalar birləşmə əmələ gətirmir. Burada üçüncü növ təyini söz birləşməsi funksiyasında məktəbin avtobusu ifadəsi çıxış edir. Bu söz birləşməsinin komponentləri arasında olan şagirdləri gəzintiyə aparan ifadəsi isə təyindir, avtobusu komponentini təyin edir: məktəbin hansı avtobusu – şagirdləri gəzintiyə aparan avtobusu.

Ənənəvi qrammatikalardan fərqli olaraq formal qrammatikalarda bu sadaladığımız incəliklər nəzərə alınmalıdır.

Bu misalların sayını istənilən qədər artırmaq olar.

Göründüyü kimi, formal linqvistik təhlil sistemlərində qrammatik qaydaların daha dərindən işlənilməsi zərurəti yaranır. Ənənəvi qrammatik qaydalar indiki şəkildə NLP sistemlərinin yaradılmasında istifadə edilərsə, çoxlu sayda səhv nəticələrin alınmasına səbəb olar. Lakin təəssüflə qeyd edilə bilər ki, Azərbaycan dilçiliyində bəzi istisnalar olmaqla [Axundov, 1973; Abdullayev, 1999; Mahmudov, 2002] müasir NLP sistemlərinin yaradılması sahəsində istifadə edilməsi mümkün olan ciddi tədqiqat işləri çox azdır.

Azərbaycanda NLP sistemləri yaradılmamışdan qabaq da kompüterlərin köməyi ilə bir çox statistik sözlüklər tərtib olunmuş, tezlik, əks-əlifba lüğətləri tərtib olunmuşdur. Həmin işlərin davamı olaraq sonralar yazılı abidələrin (“Kitabi-Dədə Qorqud” dastanı) və klassiklərin (Məhəmməd Füzuli) dilinin kompüterlərlə öyrənilməsi ilə bağlı işlər aparılmışdır [Mahmudov, 2013, s.82-83].

Toplanmış təcrübə əsasında sonrakı illərdə Azərbaycan dilinin formal təhlili və sintezi, maşın tərcüməsi məsələləri ilə bağlı nəzəri tədqiqatlar aparılmış, qrammatikanın ayrı-ayrı fraqmentlərinin formallaşdırılmasına cəhd göstərilmişdir. Bu tədqiqatlar sistemli aparılmamış, nəzəri müddəaların praktikada yoxlanması və tətbiqi işləri üçün linqvistik baza yaradılmamışdır. Məhz ona görə də bu işləri gələcəkdə NLP sistemlərinin yaradılması üçün hazırlıq mərhələsi hesab etmək olar.

Bu məsələ ilə bağlı istifadəçilərin rəğbətini qazanmış “Poliqlot” lüğətlər sistemini də qeyd etmək lazımdır. Proqram Azərbaycan Respublikasında informasiya-kommunikasiya texnologiyalarının inkişaf etdirilməsi çərçivəsində hazırlanmışdır. Layihədə almanca-azərbaycanca, ingiliscə-azərbaycanca, rusca-azərbaycanca, fransızca-azərbaycanca, eləcə də azərbaycanca-almanca, azərbaycanca-ingiliscə, azərbaycanca-rusca, azərbaycanca-fransızca lüğətlər təmsil olunmuşdur. Həmin proqram layihəsində Azərbaycan dilində orfoqrafiyanın yoxlanması sistemi də hazırlanıb istifadəçilərə təqdim olunmuşdur [www.poliqlot.az].

2003-cü ilin əvvəllərindən fəaliyyətə başlamış “Dilmanc” layihəsinin NLP sahəsində gördüyü işləri xüsusi qeyd etmək lazımdır (www.dilmanc.az). “Dilmanc” Azərbaycanda ilk maşın tərcüməsi sistemidir. “Dilmanc”ın fəaliyyəti ilə Azərbaycan dilinin işlək formal qrammatikasının yaradılmasına başlanılmış, ilkin dildə formal morfoloji, sintaktik və semantik təhlil, tərcümə olunmuş dildə isə cümlənin sintezi alqoritmləri işlənib hazırlanmışdır [Fətullayev, Abbasov, Fətullayev, 2008; Fətullayev, Abbasov, 2008]. Bu sistemdə böyük mətnlər də tərcümə olunur və lazım gələrsə, tərcümə olunmuş nümunələrin və ayrı-ayrı sözlərin düzgün tələffüzü də səsləndirilə bilər.

Qeyd olunduğu kimi, “Dilmanc” layihəsi təbii dillərin emalı sahəsində ən müasir və effektiv yanaşmalardan olan Dərin Neyron Şəbəkələr (Deep Neural Networks – DNN) nəzəriyyəsini öz proqramlarında tətbiq etməyə başlamışdır. Dərin Neyron Şəbəkələr yanaşması “Dilmanc” layihəsində ilk olaraq Azərbaycan dilində nitqin tanınması sisteminə tətbiq edilmişdir və hal-hazırda bu yanaşmanın köməyi ilə hazırlanmış “Dilmanc Translyator” proqramının yeni versiyasında Azərbaycan dilində şifahi nitqi daha keyfiyyətli mətnə çevirərək tərcümə etmək mümkündür.

Qeyd etmək lazımdır ki, mobil tərcüməçi proqramının Android əməliyyat sistemi üçün olan versiyası Azərbaycanda və ölkə xaricində çox populyar proqram məhsulu hesab olunur.

Türk dilləri qrupu ən geniş dil qruplarından biri hesab olunur. Bəzi mənbələrə görə 50-dən çox türk dili mövcuddur ki, bunların da 40-a qədəri hal-hazırda istifadə edilməkdədir. Lakin 15 türk dili artıq ölü dil hesab olunur [Kibrik, Tenişev və b., 1997].

Türk dilləri üçün NLP sistemlərinin yaradılması işlərinə ötən əsrin sonlarından, daha dəqiq desək, 80-ci illərdən başlanılmışdır. Həmin vaxtlar türk dillərinin məşin fondunun yaradılması ilə bağlı müzakirələr aparılır, hazırlıq işləri görülürdü. [Гузев, Пиотровский, Щербак, 1988]. İndinin özündə də “korpus” və “məşin fondu” terminlərinin yanaşı işlədilməsi halları müşahidə olunur. Milli dil korpuslarının yaradılması sahəsində çalışan tədqiqatçıların bir qrupu (Buskunbayeva, Sira-zitdinov və b.) informasiya texnologiyalarının bugünkü durumu baxımından “milli korpus” termini əvəzinə “kompüter fondu” termininin işlənməsini daha məqbul hesab edirlər. Bütün bunlar göstərilən terminin hələ də tam sabitləşməməsindən xəbər verir [Бускунбаева, Сиразитдинов, 2012, s.54-5].

Məlum olduğu kimi, 1988-ci ilin may ayında Moskvada Sovet Türkoloqları Komitəsinin XIV plenumunda keçmiş SSRİ

ərazisində yaşayan türk xalqlarının dillərini (Azərbaycan, özbək, qazax, qırğız, türkmən, başqırd, tatar, qaqaуз, yakut, çuvaş, xakas, altay, qumıq, noqay, karaim və s.) əhatə edən məşın fondunun yaradılması sahəsində işlərə başlanılması qərara alındı. İlk növbədə türk dillərinə aid leksikoqrafik, qrammatik, statistik-üslubi, tarixi-etimoloji, folklor materialı, bədii ədəbiyyat nümunələri toplanılmalı və sistemləşdirilməli idi. Daha sonra türk dilləri haqqında istənilən məlumatı dolğun və düzgün şəkildə istifadəçiyə təqdim edə biləcək qaydalar işlənilib hazırlanmalı idi. Türk dillərinin məşın fondunun yaradılması üçün mərkəz olaraq Qazaxıstan Elmlər Akademiyasının Dilçilik İnstitutu tövsiyə olunurdu [Esenbayev, Maxambetov, 2012, s.61-66].

Türk dillərinin məşın fondunun (türk dillərinin milli korpuslarının) yaradılmasının ilkin mərhələsində ulu türk dilinin rekonstruksiyası məsələsinin araşdırılması və həlli nəzərdə tutulurdu. Məhz ona görə də, yaradılacaq milli korpus üçün ilk növbədə aşağıdakı məlumatların toplanması zəruri hesab olunurdu:

- türk dillərinin birhecalı söz köklərinin müxtəlif növlərini əhatə edən struktur-fonetik məlumat;
- morfem siyahıları;
- sintaktik əlaqələri əks etdirən sxemlər;
- affikslərin qrammatik tezaurusu;
- konkret türk dillərinin fonetik, qrammatik quruluşu barədə analitik göstəricilər toplusu.

Beləliklə, ilkin mərhələdə ümumtürk və dillərarası fonetik, leksik, morfoloji, sintaktik, semantik təhlil problemlərinin öyrənilməsinə imkan verən türk dillərinin linqvistik bankının yaradılması nəzərdə tutulurdu.

Sonrakı mərhələdə çoxfunksiyalı böyük dil korpuslarının yaradılması planlaşdırılırdı. Bu korpuslar bütövlükdə dil ailələrini və ayrılıqda konkret türk dillərini modelləşdirməyə qabil

olmalı idi. Burada təbii dili ehtiva edən bütün üslublara aid dil materialı və fonetik, qrammatik qanunauyğunluqlar öz əksini tapmalı idi [Esenbayev, Maxambetov, 2012, s.189-192]. Sonralar, türk dillərinin milli korpuslarının hazırlanması prosesində bu tövsiyələr müəyyən dərəcədə nəzərə alınmışdır.

NLP sistemlərinin yaradılması sahəsində Türkiyədə bir çox işlər görülmüş və görülməkdədir [Kemal Oflazer, <https://>].

NLP ilə bağlı tədqiqatlar bir çox türkdilli xalqların milli dil korpuslarının yaradılması sahəsində işlərin genişlənməsinə imkan yaratmışdır. Belə ki, hazırda qazax, tatar və başqırd dillərinin yaradılmaqda olan milli korpusları ilə bağlı işlər də NLP sistemləri sahəsindəki tədqiqatlardan qaynaqlanır.

Qırğız və qazax mətnlərinin statistik metodlarla tədqiqi sahəsində T.Sadıkov və B.Şarşembayevin tədqiqatları böyük maraq doğurur. Həmin araşdırmalar qırğız və qazax dillərinin milli dil korpuslarının komponentləri kimi istifadə oluna bilər [Sadıkov, Şarşembayev, 2012, s.20-25]. Bu iki tədqiqatçının çoxillik zəhmətinin nəticəsi olan və 2011-ci ildə Ankarada Türk Dil Kurumu Yayınları tərəfindən çap olunmuş 1647 səhifədən ibarət “Manas destanı. Kırğızca-türkçe büyük dizin” sözlüyünü xüsusi qeyd etmək olar. Bu araşdırma türk dillərinin müştərək dil korpusunda yerləşdiriləcək material kimi önəmlidir. Bir milyondan çox misradan ibarət dastanın sözlüyünün tərtibi yeni texnoloji vasitələrin köməyi olmadan mümkün olmazdı. Dastanda işlənmiş sözlərin qrammatik formadan lüğət vahidi formasına gətirilməsi işləri də yerinə yetirilmişdir [Sadıkov, Şarşembayev, 2011]. Qazax və türk dillərində NLP sistemlərinin komponentləri kimi istifadə oluna biləcək digər tədqiqatlara da rast gəlmək mümkündür [Жубанов, 2013; Жубанов, Жанабекова, 2017; Turkish morphological analyzer, <https://>].

Türk dillərində müasir NLP sistemlərinin yaradılması işləri son zamanlar daha da genişlənməkdədir. Bunu internetdə

yer alan məlumatlardan da əyani görmək olar. Nitqin tanınması, mətnin səsləndirilməsi, maşın tərcüməsi və müstəqil axtarış sistemlərinin yaradılması ilə bağlı ayrı-ayrı türk dillərində orijinal NLP sistemləri yaradılmışdır. Bu müasir NLP sistemlərinə aşağıdakıları misal göstərmək olar:

Türk dili üçün:

- SR (nitqin tanınması) – <http://www.sestek.com/tr/konusma-tanima/>;
- TTS (mətnin səsləndirilməsi) – <http://www.linguatec.net/text-to-speech/vrs15.turk.html>, <http://www.sestek.com/tr/tts-demo/>;
- MT (maşın tərcüməsi) – <http://cevirsozluk.com/>;
- SE (axtarış sistemləri) – <http://www.geliyoo.com/>, <http://www.arabul.com/>, <http://www.arama.com/>.

Qazax dili üçün:

- SR (nitqin tanınması) – <http://uniline.kz/wordpress/?p=665>;
- TTS (mətnin səsləndirilməsi) – <http://217.15.186.68/uniline/Text-to-speech/Index.php>;
- MT (maşın tərcüməsi) – <https://sozdik.kz/ru/dictionary/translate>;
- SE (axtarış sistemi) – <http://kaz.kz/>.

Özbək dili üçün SR (nitqin tanınması) və TTS (mətnin səsləndirilməsi) ilə bağlı məlumata rast gəlinmədi.

Maşın tərcüməsi və axtarış sistemləri ilə bağlı isə özbək dili üçün müəyyən işlər görülmüşdür:

- MT (maşın tərcüməsi) – <http://solver.uz/>, <http://www.vera-lab.ru/cache/program.php?pid=293-perevodchik-russko-uzbekskiy>;
- SE (axtarış sistemləri) – <http://www.alto.uz/>, <http://www.uz/>.

Qırğız dili üçün:

SR (nitqin tanınması) və TTS (mətnin səsləndirilməsi) ilə bağlı məlumatla rast gəlinmədi.

- MT (maşın tərcüməsi) – <http://tili.kg/>, <http://tamgasoft.kg/dict/>;

- SE (axtarış sistemi) – <http://www.ai.kg/>.

Türkmən dili üçün yalnız maşın tərcüməsi ilə bağlı internetdə məlumat yerləşdirilmişdir:

- Türkmən dilindən ingilis, türk, rus dillərinə və əksinə maşın tərcüməsi sistemi hazırda uğurla istifadə olunmaqdadır (www.translatorscafe.com/cafe/member118197.htm); <http://www.ajapsozluk.com/>).

- Analoji NLP sistemləri özbək (www.freeware.Ru/proqram_proq_id_32572.html) və tatar dilləri üçün də hazırlanmışdır (russian-tatar.ru; <https://translate.yandex.ru/translator/tatar-russian>; www.linqvo-online.ru/ru/translate/tt-ru).

Digər türk dilləri ilə bağlı müəyyən işlər görülür və bu haqda internetdən geniş məlumat əldə etmək olar.

Süni intellektin tərkib hissəsi olan NLP sistemləri necə yaradılır?

İnsan yarandığı gündən ətraf mühiti dərinədən və hərtərəfli öyrənməyə, dərk edərək özü üçün daha əlverişli yaşayış şəraiti yaratmağa çalışır. Bizi əhatə edən aləm, təbiətin öyrənilməsi və mənimsənilməsi sahəsində hazırda böyük uğurlar qazanılmışdır və bu da öz növbəsində insan cəmiyyətinin təkamülü prosesinə ciddi təsir etmişdir. İnsan özü haqqında çoxlu biliklər əldə etsə də, təəssüf ki, bu biliklər onun ətraf mühitlə bağlı qazandığı biliklərlə müqayisədə cüzi görünür. Bu, xüsusilə insan beynində gedən proseslərə daha qabarıq nəzərə çarpır. İnsan beynində gedən intellektual proseslər, hələ də lazımi elmi izahını tapmamışdır. Əgər həmin prosesləri hərtərəfli və tam dərk edə bilsəydik, onda insan kimi düşünə

bilən qurğu və mexanizmlər yaratmaqla insan həyatını və əməyini daha da yüngülləşdirə bilər, aclığın, səfalətin, xəstəliklərin, müharibələrin və bu qəbildən olan digər təbii və texnogen kataklizmlərin qarşısını daha effektiv ala bilərdik.

İnsan beynində gedən prosesləri tam dərk etməsək də, beynə daxil olan (əşitdiyimiz, gördüyümüz və digər hiss üzvlərimizlə qavradığımız) informasiyanı və daxil olmuş informasiyanın emalı nəticəsində beynin verdiyi qərarları – çıxış informasiyasını (nitq, hərəkət və s.) bilməklə beyində gedən proseslərin modelini qurmaq mümkündür [Məmmədov, Əzimova, Fətullayev, 2010].

Belə hallarda riyazi statistika tədqiqatçıların köməyinə gəlir. Belə ki, insan beyninə daxil olan informasiyanı və beynin verdiyi qərarları öyrənib qarşılaşdırmaqla beyində gedən intellektual proseslərin müəyyən dəqiqliklə modelini qurmaq olar. Süni intellekt problemləri çərçivəsində yaradılan belə modellərə müasir NLP sistemlərini misal göstərmək olar.

NLP sistemlərinin yaradılması yolunda mövcud olan boşluqları doldurmaq üçün statistik metodlardan geniş istifadə olunur. Onu da qeyd etmək lazımdır ki, statistika təkcə qrammatikada olan çatışmazlıqların aradan qaldırılması üçün deyil, həm də NLP texnologiyalarının yaradılmasında mövcud olan yanaşmalardan biri kimi də geniş tətbiq olunur. Məsələn, maşın tərcüməsinə SBMT (Statistic Based Machine Translation – statistikaya əsaslanan maşın tərcüməsi) yanaşması, nitqin tanınması sistemləri və s. buna misal ola bilər.

NLP ilə bağlı aparılmış tədqiqatların və yaradılmış kompüter proqramlarının əsasında da statistik qanunauyğunluqlar durur. Belə ki, ənənəvi dilçilik (qrammatika) linqvistik alqoritmlərin yaradılması üçün zəruri olan informasiyanı verə bilmədiyi üçün bu informasiya başqa yollarla – korpuslardan istifadə edilməklə statistik metodlarla əldə olunur. Başqa sözlə, “Dilmanc” layihəsindəki korpuslar konkret NLP sistemləri yara-

dılması baxımından daha “yaxşı” statistika verən korpuslardır. Bu cəhət “Dilmanc” layihəsi çərçivəsində yaradılmış korpusların aparıcı xüsusiyyəti hesab oluna bilər [Məmmədov, Əzimova, Fətullayev, 2010].

NLP və korpuslar arasındakı əlaqələrin izahına keçməzdən əvvəl ayrı-ayrı dillər üçün NLP sistemlərinin yaradılması ilə əlaqədar bəzi ümumi məsələləri nəzərdən keçirmək lazım gəlir. Şərh olunan məsələlər eyni zamanda təkcə Azərbaycan dilinə deyil, digər qohum türk dillərinə də şamil oluna bilər.

NLP sistemlərinin yaradılmasında və dilin hərtərəfli tədqiqi ilə bağlı geniş təbiiq olunmasını nəzərə alaraq yeni yaranmaqda olan korpus dilçiliyi barədə də məlumat vermək məqsədəuyğun olardı.

Azərbaycanda NLP sistemlərinin yaradılması sahəsində tədqiqatlar “Dilmanc” layihəsi çərçivəsində [www. Dilmanc. az] hazırlanmış, dilin bütün üslublarını əhatə edən ikidilli paralel mətn korpuslarını və konkret dilləri əhatə edən irihəcmli birdilli korpusları xüsusi olaraq qeyd etmək olar. Həmin korpuslar və onların həcmi aşağıda göstərilmişdir:

- ingilis-Azərbaycan ikidilli korpusu (2 milyon cümlə);
- türk-Azərbaycan ikidilli korpusu (277 min cümlə);
- rus-Azərbaycan ikidilli korpusu (4,5 milyon cümlə);
- Azərbaycan birdilli korpusu (60 milyon cümlə);
- Türk birdilli korpusu (322 milyon cümlə).

Qeyd etmək lazımdır ki, yuxarıda qeyd olunan bu ikidilli və birdilli korpuslar mətnin formal linqvistik təhlili və məşin tərcüməsi sistemlərində uğurla istifadə olunmaqdadır. İkidilli paralel mətn korpusları Azərbaycan, türk, rus və ingilis dilləri arasında avtomatik tərcümə vasitəsi və lüğət kimi istifadə oluna bilər [Fətullayev, Mahmudov, Məmmədov, 2015].

NLP sistemləri, dilin milli korpusları və onların tərkib hissələri daim inkişafda, təkmilləşmədə, genişlənmədə olan bir

sistemdir və onların yaradılması prosesi arasıkəsilməzdir. İnformasiya-kommunikasiya texnologiyaları təkmilləşdikcə, onların imkanları və tətbiq sahələri genişləndikcə, NLP sistemlərinin də zənginləşməsi və sürətli inkişafı üçün perspektivlər yaranır. Azərbaycan dili ilə bağlı müxtəlif vaxtlarda yaradılmış tətbiqi-linqvistik texnologiyalar, müxtəlif çeşidli sözlüklər, lüğətlər (o cümlədən, statistik, maşın tərcüməsi sistemində istifadə olunan avtomatik lüğətlər) və s. dilçiliyin sonrakı inkişaf mərhələlərində müxtəlif məqsədlərlə istifadə oluna bilər.

Azərbaycan dilində tələffüzün formal təsviri. Azərbaycan dilinin çoxəsrlik qədim tarixi vardır. Eramızın III-IV əsrlərində Azərbaycan dili şifahi ünsiyyət vasitəsi kimi çox geniş bir areala malik idi. Təsadüfi deyil ki, VII əsrdə artıq “Kitabi-Dədə Qorqud” dastanları kimi möhtəşəm bir abidə Azərbaycan dilində qədim Azərbaycan ərazisində yaranmışdı. Çox sonralar yazıya alınmış bu abidə erkən dövrlər Azərbaycan dili haqqında yetərincə məlumat əldə etmək imkanı yaradır. Maraqlıdır, Azərbaycan dilinin şifahi tələffüz formalarının ilk ünsürlərinə bu dastan mətnində rast gəlmək mümkündür.

Ümumiyyətlə, şifahi dil hər bir xalqın ədəbi dilini tənzimləyən, zənginləşdirən və canlı orqanizm kimi daim inkişafda olmasını təmin edən çevik ünsiyyət kanalıdır.

Şifahi dil yazılı ədəbi dili fasiləsiz qidalandıran süzğəcdir. Şifahi danışmaq prosesində cilalanan, saf-çürük edilən sözlər bu süzğəcdə süzülərək saflaşır, durularaq ədəbi dilə keçir, dilin lüğət tərkibində vətəndaşlıq hüququ qazanır.

Danışmaq prosesində tələffüz heç də yazılı nitqin adekvat təkrarı deyil. Başqa sözlə desək, qəribə görünsə də, biz heç də yazılan kimi, kitab dilində danışmırıq. Danışmaq orqanları danışmaq aktı prosesində sabitliyə meyillidir. Əgər söz dodaq saitləri ilə başlanırsa, axıra qədər o vəziyyətdə qalmağa meyilli olan danışmaq orqanları damaq ahənginə keçmir və nəticədə

sözün tələffüzü boyu eyni formada qalırlar. Dodaq və damaq saitlərinin çox vaxt yanaşı işləndiyi ərəb alınmaları tələffüz prosesində imkan daxilində Azərbaycan dilinə uyğunlaşdırılır: müəllim-məllim, müalicə-malicə və s.

Tələffüz prosesində danışmaq orqanları asan tələffüzə meyl göstərir, müxtəlif məxrəcli sözləri eyni və ya, heç olmazsa, yaxın məxrəcə gətirir. Bu zaman alınma sözlərin türk artikulyasiyasına uyğunlaşdırılması, sadələşməyə və ixtisara meyl hiss olunur: təəccüb- təcüb, təəssürat-təsürat və s.

Məlum olduğu kimi, türk dillərində, o cümlədə Azərbaycan samitlər dörd prinsipə görə təsnif edilir: əmələgəlmə yerinə, əmələgəlmə üsuluna, səs tellərinin iştirakına və yumşaq damağın vəziyyətinə görə [Axundov, 2012, s.93].

Qohum dillərdə tələffüzdə fərqlər az olur. Nitqin yaranma mexanizmi, səsin tələffüzü prosesi, demək olar ki, eynidir. Məsələn, türk, Azərbaycan, türkmən, qaqauz və kırım tatarlarının dilində tələffüz mexanizmində ciddi fərqlər hiss olunmur.

Qohum olmayan dillərdə isə vəziyyət tamamilə fərqlidir. Əgər hər hansı bir alınma sözdə həm dodaq, həm damaq samitləri və saitləri iştirak edirsə, Azərbaycan tələffüzündə həmin variantlardan birinə üstünlük verilir. 1978-ci ildə “Elm” nəşriyyatında çapdan çıxmış “Müasir Azərbaycan dili. I cild. Fonetika, əlifba, orfoqrafiya, orfoepiya, leksikologiya” kitabında da orfoepiyanın prinsipləri, saitlərin, samitlərin, şəkilçilərin, qoşmaların tələffüzü, alınma sözlərin orfoepiyası məsələlərinə baxılmışdır [Müasir Azərbaycan dili, 1978, s.84-99].

2013-cü ildə “Elm və təhsil” nəşriyyatında çapdan çıxmış “Kompüter dilçiliyi” monoqrafiyasının “Mətnin formal morfoloji təhlilinin morfonoloji bloku” fəslində söz kökündə, söz kökü ilə sözdüzəldici, söz kökü ilə sözdəyişdirici şəkilçilər, eləcə də sözdüzəldici şəkilçi ilə sözdəyişdirici və sözdəyişdirici

şəkilçilər sərhədində baş verən morfonoloji dəyişmələr geniş şərh olunmuşdur. Həmin tədqiqat orfoepiya və tələffüzlə bağlı məsələlərin həllində böyük fayda verə bilər [Mahmudov, 2013, s.239-50].

Aşağıda Azərbaycan dilində müşahidə olunan tələffüz formalarına aid nümunələr verilir. Həmin nümunələr əsasında Azərbaycan tələffüz qaydalarını işləyib hazırlamaq, onun alqoritmini və proqramını qurmaq prosesinə başlamaq olar:

Türk dillərində, o cümlədən Azərbaycan dilində iki saitlin yanaşı işlənməsi səciyyəvi deyil. Belə hallarda tələffüzdə həmin saitlərin arasına bitişdirici samit əlavə olunur:

• saat / sa (h) at, zəif/ zə (y) if, ailə/ ai (y) lə, sait/ sa (h) it, şair/ şa- (y) ir, aid / a (y) id, ideal/ ide (y) al, radio /rad (y) o;

Söz ortasında qq–kq əvəzlənməsi baş verir:

- Saqqal/ sa (k) qal, çaqqal/ça (k) qal, baqqal/ ba (k) qal
- Haqqında – ha (k) q ında, haqq/ ha (k) q;

Söz sonunda q–k əvəzlənməsi baş verir:

- aşıq/aşik, aşıq/aşık, ittifaq/ ittifak;

Söz sonunda q–ğ-ya, bəzən isə x-yə keçir:

• arıq–arığ/arıx, varaq/varağ, çarıq–çarığ/çarıx, oxumaq/oxumağ, sarıq/ sarığ;

Söz ortasında nb–mb əvəzlənməsi gedir:

- anbar/ambar, hanbal/hambal, tənbəki/təmbəki;

Söz ortasında nm–mm əvəzlənməsi gedir:

• Qanmaz/qammaz, dinməz/dimməz, sönməz/sömməz, dönməz/dömməz;

İki sait və samitin yanaşı gəldiyi ərəb mənşəli sözlərdə ixtisar baş verir:

- təcəüb/təcüb, təəssürat/təsürat, təəssüf/təsüf;

Ərəb mənşəli sözlərdə söz ortasında iki eyni samit yanaşı gələrsə, biri ixtisar olunur:

• ədəbiyyat/ədəbiyat. səhiyyə/səhiyə, tərənnüm/ tərənnum, müvəffəqiyyət/müvəfəqiyyət;

İstisnalar:

• hissə – hisə, dərrakə – dərakə, əttar/ətar;

Xəbərlilik şəkilçisinin sonunda “ r” samiti düşür:

• tələbədir – tələbədi, pisdir – pisdidi. Yaxşıdır – yaxşıdı. səhərdir – səhərdir;

Söz ortasında nl nn əvəzlənməsi baş verir:

• onluq/onnuq, onlar/onnar;

• odunlar/odunnar, vicdanlı/vicdannı;

Söz ortasında dl – dd-yə keçir:

• odlar/oddar, qurdlar/qurddar. yadlar – yaddar;

Söz ortasında zl – zz-yə keçir:

• Qızlar/qızzar, buzlar/buzzar, sözlər/sözzər, gözlər/gözzər;

Söz ortasında zl – zd-yə keçir:

• özləri/özdəri, sözlər/sözdəri, duzlu/duzdu, sözlü/sözdü;

Söz ortasında nd – nn -ə keçir:

• ondan/onnan, bundan/bunnan, məndən/mənnən. sən-dən/sənnən, alandan/alannan, görəndən/görənnən;

Söz ortasında rl – ll-ə keçir:

• alırlar/alıllar, baxırlar/baxıllar, atırlar/atıllar, qatırlar/qatıllar;

Tələffüzdə bəzən söz ortasında rl – rr əvəzlənməsi baş verir:

• alırlar/alırrar, baxırlar/baxırrar, yazırlar/yazırrar, pozurlar – pozurrar;

Danışıqda bəzən söz ortasında tl – td-yə keçir:

• ətli/əttdi, atlı/attdı, qatlı/qattdı, çatlı/çattdı;

Tələffüzdə bəzən söz ortasında c – j-yə keçir:

əcaib/əjayib, əclaf/əjlaf, əcəl/əjəl, aclıq/ajlıq;

Söz ortasında ç – ş əvəzlənməsi baş verir:

• qaçmaq/qaşmaq, açmaq/aşmaq, uçmaq/uşmaq, saç-lar/saşlar;

Ahəng qanununun pozulduğu ərəb mənşəli sözlər mümkün qədər Azərbaycan dilinə uyğunlaşdırılır:

• müəllim/məllim, müalicə/malicə, müasir/masir.

Sonu “t” samiti ilə bitən bəzi fellər saitlə başlanan şəkilçi qəbul etdikdə t – d əvəzlənməsi baş verir:

• et – edir, get – gedir, eşit – eşidir. qayıt – qayıdır [Mahmudov, 2013, s.172].

İstisnalar: bat, at, sat, yat və s fellərdə bu əvəzlənmə baş vermir.

Sonu “k” ilə bitən sözlərə saitlə bitən şəkilçilər artırıldıqda k – y əvəzlənməsi baş verir:

• çiçək – çiçəy-im, çörək – çörəy-im, ərik – əriy-im. üzük – üzüy-üm.

İstisnalar siyahı şəklində əlavə olunur [Mahmudov, 2013, s.245-246].

Sonu “q” ilə bitən sözlərə saitlə bitən şəkilçilər artırıldıqda q – ğ əvəzlənməsi baş verir:

• torpaq – torpağ-ın, kolluq – kolluğ-un, balıq – balığ-ın. buruq – buruğ-un.

İstisnalar siyahısı əlavə olunur [Mahmudov, 2013, s. 246-247].

“y” səmindən əvvəl gələn “a” səsi “ı”-ya çevrilir:

• ataya – at ı ya, danlayır – danl ı yır, bağlayır – bağl ı yır. almayacaq – alm ı yacaq.

“y” səmindən əvvəl gələn “ə” səsi “i”-yə çevrilir:

• gəlməyəcək – gəlm i yəcək. işləyər – işl i yər, dinləyər – dinl i yər, nənəyə – nən i yə.

İkihecalı sözlər saitlə başlanan şəkilçi qəbul etdikdə qapalı hecadakı qapalı sait düşür (-ı, -i, -u, -ü):

• ağıl – ağılm. ağız – ağızım. şəkil – şəklim;

• könül – könlüm, oğul – oğlum, ömür – ömrüm [Mahmudov, 2013, s.242].

Birinci komponenti qapalı saitlə bitən mürəkkəb sözə saitlə başlanan söz əlavə olunduqda həmin sait düşür:

• Əliəkbər – Ələkbər, əmioğlu – əmoğlu, bibioğlu – biboğlu, bacioğlu – bacoğlu, dayıoğlu – dayoğlu.

Sözün ortasında və sonunda “v” səsinin düşməsi nəticəsində “o” uzun tələffüz olunur:

• dovşan – dooşan, dovğa – dooğa. cilov – ciloo, buzov – buzoo.

Sözün ortasında və sonunda “v” səsinin düşməsi nəticəsində “ö” uzun tələffüz olunur:

• növbə – nööbə, kövşən – kööşən;
• tövlə – töölə, dövlət – döölət, bülöv – bülöö, kösöv – kösöö.

Sözün ortasında “d” səmindən əvvəl “c” səsi “j” kimi tələffüz olunur:

• kərpicdə – kərpijdə, ağacda – ağajda, acdan – ajdan, gücdən – güjdən.

Söz sonunda “b” səsi bir qədər karlaşır və “p” kimi tələffüz olunur:

• kabab – kabap, qutab – qutap, hesab – hesap, alıb – alıp, gəlib – gəlip.;

Sözdüzəldici -laq şəkilçisi danışıqda -lağ/- dağ kimi tələffüz olunur:

• çaylaq – çaylağ, yaylaq – yaylağ, otlaq – otlağ/ otdağ, qışlaq – qışlağ/qışdağ.

Sözdüzəldici -sov şəkilçisinin sonunda gələn “v” samiti tələffüzdə düşür:

• dəlisov – dəliso, uzunsov – uzunso.

Sonu q – k samitləri ilə bitən şəkilçilərdə q – ğ// k – y əvəzlənməsi baş verir:

- yaxşıraq – yaxşırağ, gödərək – gödərəy, qalıq – qalığ, buruq – buruğ, hörük – hörüy, yataq – yatağ; kəsək – kəsəy, yatacaq – yatacağ, siləcək – siləcəy, tozanaq – tozanağ;

- döyənək – döyənəy, yazmaq – yazmağ;

- görmək – görməy, yayaq – yayağ, gedək – gedəy, başlayaq – başlayağ, işləyək – işləyəy.

Çıxışlıq halda -dan/ -dən yerinə tələffüzdə -nan/-nən işlənir:

- maşından – maşınnan, işimdən – işimnən, bostandan – bostannan, nənəmdən – nənəmnən.

Xəbərlilik şəkilçisi II şəxs cəmdə -sınız, -siniz, -sunuz, sünüz əvəzinə -sız, -siz, -suz, -süz işlənir:

- baxırsınız – baxırsız, gəlırsiniz – gəlırsız, durursunuz – durursuz, sürürsünüz – sürürsüz, qardaşsınız – qardaşsız, kolxozçusunuz – kolxozçusuz, tələbəsiniz – tələbəsiz, üzgüçüsünüz – üzgüçüsüz.

Arzu şəklının keçmişi bir variantda -ey kimi işlənir:

- alaydı – aleydi, gələydi – gəleydi, alaymış – aleymiş, gələymiş – gəleymiş.

Şərt şəklının keçmişi tələffüzdə bir cür -sey şəklində işlənir:

- alsaydı – alseydi, yazsaydı – yazseydi, versəydi – verseydi, alsaymış – alseymiş, yazsaymış – yazseymiş, versəymiş – verseymiş.

Feli bağlama şəkilçilərində tələffüzdə q – ğ/ k – y əvəzlənməsi baş verir:

- danışdıqca – danışdığca, dedikcə – dediyəcə, oxuduqca – oxuduğca, gördükcə – gördüyəcə. axtardıqda – axtardığda, qurduqda – qurduğda, gördükdə – gördüydə, sökdükdə – sökdüydə, axtardıqdan – axtardığıdan, qurduqdan – qurduğdan, gördükdən – gördüydən, sökdükdən – sökdüydən.

İlə (-la, -lə) qoşması danışıda -nan/-nən/ -ynan/-ynən/ -eynən şəklində işlənir:

- at ilə (atla) – atnan, ağıl ilə (ağılla) – ağılnan, ata ilə (atayla) – ateynən;

- nənə ilə (nənəylə) – nəneynən, qoyun ilə (qoyunla) – qoyunnan, sürü ilə (sürüylə) – sürüynən.

Lakin burada istisnalar ola bilər. Söz ortasında -la, -lə şəkilçisi başqa təsadüfi oxşar hissəciklərlə qarışdırılaraq düzgün olmayan formalar yarada bilər.

“Dək” qoşması danışmada “dəy” kimi işlənir:

- səhərədək – səhərədəy, indiyədək – indiyədəy, alana – alanaədək, görünədək – görünədəy.

Təkhəcalı sözlərin sonunda qoşa samitlərin biri tələffüz olunmur:

- zənn – zən, zidd – zid, sirr – sir, tibb – tib, fənn – fən, xətt – xə, hədd – hə, həll – hə, hiss – his.

Söz əvvəlində r və s samitlərindən əvvəl ahəngə uyğun qapalı saitlər əlavə olunur:

- stul – ustul, stol – ustol, rayon – irayon, Rasim – İrasim, ramazan – iramazan/ rəmazan – irəmazan, rus – urus, Rusiya – Urset, razi – irazi.

Söz sonunda -sanız, -səniz şəkilçiləri müvafiq olaraq -saz, -səz ilə əvəz olunur:

- baxsanız – baxsaz, alsanız – alsa, cəmləşsəniz – cəmləşsəz;

- tələssəniz – tələssəz, kəsdirsəniz – kəsdirsəz.

Söz ortasında bu əvəzlənmə müşahidə olunmur.

Söz sonunda tələffüz prosesində -mayırdı, -məyirdi müvafiq olaraq -mırdı, -murdu və -mirdi, mürdü ilə əvəz olunur:

- almayırdı – almırdı, çıxmayırdı – çıxmırdı, qorxutmayırdı – qorxutmırdı, tutmayırdı – tutmırdı, gülməyirdi – gülmürdü, üzməyirdi – üzürdü [Mahmudov, 2013, s.203].

Söz sonunda tələffüz prosesində -dınız, -dunuz, -diniz, -dünüz qısa formada -dız, -duz, -diz, düz kimi işlənir:

- aldınız – aldız, oxudunuz – oxuduz, gəldiniz – gəldiz, gördünüz – gördüz.

Bu xüsusiyyət həm söz sonunda, həm də söz ortasında özünü doğruldu.

Söz sonunda və ortasında -ınız (-nız), -unuz (-nuz), -iniz (-niz), -ünüz (-nüz) müvafiq olaraq -ız, -uz, -iz, -üz ilə əvəz olunur:

- adınızı – adızı, qolunuzu – qoluzu, əlinizi – əlizi, gözünüzü-gözünüzü.

Azərbaycan dilində tələffüzlə bağlı bu siyahını müxtəlif dialektoloji lüğətlərdən, folklor materiallarından gətirilmiş nümunələr hesabına bir qədər də genişləndirmək olardı. Məlum olduğu kimi, şifahi xalq ədəbiyyatında, Azərbaycan dilinin dialekt və şivələrində, canlı dildə ədəbi dildən fərqli olaraq tələffüz qaydaları özünü daha qabarıq büruzə verir. Fikrimizcə, şifahi nitqin sintezi və analizi, şifahi, canlı danışığ dilinə əsaslanan axtarış-sorğu sistemlərində, nitqin tanınması sistemlərində bu xüsusiyyətlərin ümumiləşdirilərək nəzərə alınması faydalı olardı.

Azərbaycan dilində vurğunun formal təsviri. Danışığ prosesində söz daxilində bu və ya digər heca fərqli şəkildə tələffüz olunur. Bəzən bir heca digərinə nisbətən daha qüvvətlə deyilir, hecaların birində uzanma baş verir və ya səsin tonu qalxıb-enir, musiqili ahəng yaranır.

Azərbaycan dilində vurğunun aşağıda göstərilən növləri müşahidə olunur:

- sözdə bir hecanın daha qüvvətlə deyilişi, üzərinə vurğu düşən hecanın daha qüvvətlə tələffüzü, səsləndirilməsi. Vurğunun bu növü dinamik vurğu adlanır;

- vurğulu hecada səsin tonunun qalxıb-enməsi. Vurğunun bu növü avazlı, musiqili vurğu hesab olunur;

- sözdə vurğulu heca uzanır, vurğulu hecada sait nisbətən uzun tələffüz olunur. Vurğunun bu növü kəmiyyət vurğusu kimi səciyyələndirilir;

- vurğulu və vurğusuz hecaların akustik baxımdan fərqləndirilməsi. Burada hecalardakı ayrı-ayrı səslərin akustik cəhətləri təhlil edilir və vurğunun bu növü keyfiyyət vurğusu adlanır.

Bütün bu sadalananlar vurğunun fərqləndirici xüsusiyyətləri, komponentləri hesab olunur [Adilov, Verdiyeva, Ağayeva, 1989, s.45; Müasir Azərbaycan dili, 1978, s.35-38].

Məlum olduğu kimi, sözlər birhecalı və ya çoxhecalı ola bilər. İki və ikidən çox hecası olan sözlərdə vurğunun yeri mübahisəli görünə bilər. Deyilənləri formal təsvirini verməyə cəhd edək:

Azərbaycan dilində bütün təkhecalı sözlər vurğuludur: al, bax, de, dur, sür, qaz, yaz (feillər); gül, boz, toz, söz, dil, əl, qol, baş, diş, saç, ağı, göz (adlar).

Azərbaycan dilində ikihecalı sözlərdə vurğu çox vaxt birinci hecaya düşür: `alın, `deyin, `geyin, `silin, `almır, `gəlmir, `yemir.

ikihecalı sözlərdə vurğunun yerini dəyişməklə sözün mənasını dəyişmək olar:

- `alma, `gəlmə, `demə, `baxma, `gəlin, `bəzən, `çəkmə;
- al`ma, gəl`mə, de`mə, bax`ma, gəl`in, bəz`ən, çək`mə.

Semantik-morfoloji qarşılaşdırma prosesində bu tip sözlərin fərqli xüsusiyyətləri üzə çıxır.

İkihecalı sözlərdə vurğunun ikinci hecaya düşdüyü sözlər də çoxdur. Bunlar həm alınma, həm də əsl sözlər ola bilər:

- ki`no, təh`sil, a`şiq, a`şiq, fi`nal, a`lət, si`zi, o`yun, ya`rış, ca`vab, su`al.

Vurğunun bu növü əsas nitq hissələrindən olan ikihecalı sözlərin çoxunu əhatə edir: şı`rin, al`tı, sək`kiz, doq`quz, biç`mək, ma`şın duz`lu, duz`suz.

Azərbaycan dilində vurğulu şəkilçilər hansı sözə birləşirsə, vurğu onun üzərinə düşür. Deməli, təkhecalı sözə vurğulu şəkilçi birləşərsə, vurğu onun üzərinə düşər. Belə şəkilçilərin sayı isə 70-ə yaxındır və bu vurğunun yerinin formal təyini prosesinin çətinləşdirir.

Üçhecalı sözlərdə vurğu həm birinci, həm ikinci, həm də sonuncu hecaya düşə bilər. Dördhecalı sözlərdə də vurğunun yeri sabit deyil və istənilən hecanın üzərinə düşə bilər. Alınma sözlərdə vurğu çox vaxt mənbə dildə olduğu kimi qalır: `kafedra, `spiker, `akademiya, `alim, de`kan

Azərbaycan dilində sözlərdə vurğu sərbəst olmaqla, həm də keçicidir. Müxtəlif səbəblərdən (şəkilçinin təsiri, səsdüşümü, səstərtimi, məntiqi təsir, emosional təsir, təkrar və s.) sözlərdə vurğunun yeri dəyişə bilər. Vurğunun yerinin dəyişməsi daha çox şəkilçi ilə bağlı olur. Bütün əsas nitq hissələrində şəkilçinin təsiri ilə vurğunun yerini dəyişməsi baş verir. Vurğunun yerinin dəyişməsinə təsir göstərən şəkilçiləri (vurğulu şəkilçiləri) nəzərdən keçirək:

-lıq, -lik, -luq, -lük – yaşıl (yaşıl`lıq), gözəl (gözəl`lik), `bol (bol`luq), `söz (söz`lük);

-ca, -cə (ağ`ca, göy`cə);

-dan, -dən – `bayaq (bayaq`dan), `dünən (dünən`dən)-
qoşma kimi

-çı, -çi, -çu, çü – qapı (qapı`çı), dəmir (dəmir`çi), odun (odun`çu), `gül (gül`çü);

-ıca, -ici, -ucu, -ücü – al (al`ıca), əy (əy`ici), `uç (uç`ucu),
`üz (üz`ücü);

-kı, -ki, -ku, -kü – `hazır (hazır`kı), `səhər (səhər`ki),
sən`ki, onun`ku;

-qı, -qu, -ğü, -ki, -gi, -kü, -gü - `al (al`qı), `pus (pus`qu), `səp (səp`ki), `sürt (sürt`kü), `sor (sor`ğu), `sür (sür`gü), `bil (bil`gi);

-lı, -li, -lu, lü - `yağ (yağ`lı), `ət (ət`li), `od (od`lu), `gül (gül`lü);

-sız, -suz, -siz, -süz – `ad (ad`sız), `qol (qol`suz), `ər (ər`siz), `gül (gül`süz);

-ıncı, -inci, -uncu, -üncü, -ncı, -nci, -ncu, -ncü – `qırx (qırx`inci). `beş (beş`inci), `son (son`uncu), `üç (üç`üncü), `altı (altı`nci), `iki (iki`nci);

-acaq (ğ), -əcək (y), -yacaq (ğ), -yəcək (y) – `al (al`acaq), `gör (gör`əcək), `oxu (oxu`yacaq), `bürü (bürü`yəcək);

-arkən, -ərkən, -yarkən, -yərkən, -yurkən, -yirkən, -yirkən – `bax (bax`arkən), `gör (gör`ərkən), `oxu (oxu`yarkən), `sürü (sürü`yərkən), `oxu (oxu`yurkən), `bağla (bağla`yirkən), `de (de`yirkən);

-mişkən, -muşkən, -mişkən, -müşkən – `al (al`mişkən), `ol (ol`muşkən), `gəl (gəl`mişkən), `gör (gör`müşkən);

-mışmış, -muşmuş, -mişmiş, -müşmüş – `oxu (oxumuş-muş), `yaz (yaz`mişmiş), `gəl (gəl`mişmiş), `gör (gör`müşmüş);

-maqda, -məkdə – oxu`maqda, gəl`məkdə;

-araq, -ərək, -yaraq, -yərək – bax`araq, gəl`ərək, oxu`yaraq, de`yərək;

-anda, -əndə, -yanda, -yəndə – bax`anda, gəl`əndə, oxu`yanda, sürü`yəndə;

-arkən, -ərkən, -irkən, -irkən, -yarkən, -yərkən, -yirkən, -yirkən – yaz`arkən, gəl`ərkən, bax`irkən, bil`irkən, de`yirkən;

-dıqda, -dikdə, -duqda, -dikdə – yaz`dıqda, gəl`dikdə, dur`duqda, gəl`dikdə;

-dıqca, -duqca, -dükcə, -dikcə – bax`dıqca, dur`duqca, gör`dükcə, gəl`dikcə;

-ınca, -unca, -incə, üncə, - yınca, -yunca, -yincə, yüncə – bax`ınca, qur`unca, gəl`incə, gör`üncə, oxu`yunca, baş`layınca, sürü`yüncə, de`yincə;

-andan, -əndən, -yandan, -yəndən – al`andan, gəl`əndən, quru`yandan, sürü`yəndən;

-dığı, -duğu, -diyi, -düyü – al`dığı, qur`duğu, gəl`diyi, gör`düyü;

-acağı, -əcəyi, -yacağı, -yəcəyi – al`acağı, ver`əcəyi, oxu`yacağı, de`yəcəyi;

-sınız, -sunuz, -sünüz, -siniz – yazır`sınız, qurur`unuz, sürür`sünüz, deyir`siniz;

-musan, -misən, -müsən – yaz`mısan, qur`musan, gəl`misən, gör`müsən;

-mişdi, -muşdu, -mişdi, -müşdü – al`mışdı, qur`muşdu, gəl`mişdi, gör`müşdü;

-ırmış, -urmuş, -irmiş, -ürmüş – yaz`ırmış, dur`urmuş, gəl`irmiş, gör`ürmüş;

-armış, -ırmış – yaz`armış, gəl`ərmiş;

-yardı, -yərdi, – ardi, -ərdi – oxu`yardı, de`yərdi, yaz`ardı, gəz`ərdi;

-yırdı, -yurdu, -yirdi, -yürdü – baş`layırdı, oxu`yurdu, işlə`yirdi, sürü`yürdü;

-ıbmış, -ubmuş, -ibmiş, -übmüş – yaz`ıbmış, dur`ubmuş, gəl`ibmiş, gör`übmüş;

-yası, -yəsi, -ası, -əsi – oxu`yası, de`yəsi, yaz`ası, gəl`əsi;

-malı, məli – oxu`malı, gör`məli;

-yalı, -yəli, -alı, -əli – oxu`yalı, deyəli, yaz`alı, gör`əli;

-ım -im, -um, -üm – kitab`ım, dəftər`im, qol`um, göz`üm;

-ın, -in, -un, -ün – kitab`ın, dəftər`in, odun`un, söz`ün;

-sı, -si, -su, -sü – ata`sı, əmi`si, qonşu`sı, güzgü`sü;

-ması, -məsi – yaz`ması, gör`məsi;

-dirt, -durt, -dirt, -dürt – yaz`dirt – vur`durt, gəz`dirt, sür`dürt;

- ar, -ər, -yar, -yər – yaz`ar, gəz`ər, oxu`yar, de`yər;
- ib, -ib, -ub, -üb, -yıb, -yib, -yub, -yüb – baş`layıb, al`ıb, gəl`ıb, dur`ub, gör`üb, de`yib, quru`yub, sürü`yüb;
- maz, -məz – oxu`maz, de`məz;
- caq, -cək – al`caq, gör`cək;
- yan, -yən, -an, -ən – oxu`yan, de`yən, al`an, gəl`ən;
- mış, -miş, -muş, -müş – al`mış, gəl`miş, qur`muş, gör`müş;
- ır, -ir, -ur, -ür, -yır, -yir, -yur, -yür – al`ır, gəl`ir, dur`ur, sür`ür, başla`yır, lə`yir, oxu`yur, sürü`yür;
- dan, -dən – qapı`dan, dəftər`dən (hal şəkilçisi kimi);
- mız, -miz, -muz, -müz – ata`mız, əmi`miz, qonşu`muz, sürü`müz;
- nız, -niz, -nuz, -nüz – ata`nız, əmi`niz, qonşu`nuz, sürü`nüz;
- sın, -sin, -sun, -sün – yaz`sın, gəz`sin, dur`sun, bük`sün;
- dir, -dir, -dur, -dür - yazdır, gəzdir, qurdur, sürdür (icbar fel kimi);
- sa, -sə – al`sa, gəl`sə;
- a, -ə, -ya, -yə – (gərək) al`a, bax`a, oxu`ya, de`yə, qapı`ya, gəmi`yə;
- dı, -di, -du, dü – al`dı, gəl`di, gör`dü, qur`du;
- ma, -mə – yaz`ma, gəl`mə, `burma (saç), `döymə (metal);
- da, -də – kitab`da, dəftər`də;
- aq, -ək, -yaq, -yək – al`aq, gəl`ək, oxu`yaq, de`yək;
- sı, -si, -su, -sü – ata`sı, əmi`si, qonşu`su, sürü`sü;
- ış, -iş, -uş, üş – yaz`ış, vur`uş, gir`iş, sür`üş.

Müşahidələr göstərir ki, Azərbaycan dilində çoxhecalı sözlərdə vurğu çox vaxt sözün son hecasından əvvəlki hecaya düşür. Bu xüsusiyyətə aid də kifayət qədər nümunələr göstərmək olar. Bütün göstərilənləri ümumiləşdirərək Azərbaycan dilində vurğunun formal modelini qurmaq olar. Həmin model

nitqin tanınması və sintezi sistemlərinin yaradılması və getdikcə təkmilləşdirilməsi prosesində böyük fayda verə bilər.

SMS informasiya mübadiləsinin linqvistik özəllikləri. Təbii dilin emalı sistemləri SMS informasiya mübadiləsi prosesində də bu və ya digər dərəcədə öz təzahürünü tapır. Nəticədə sırf SMS-lərə xas, yeni bir linqvistik mübadilə vasitəsi formalaşmış olur. SMS-lər (SMS -Short Messaging Servise) informasiya mübadiləsi prosesində mühüm ünsiyyət vasitələrindən biridir. Onların yaranması və formalaşmasında linqvistik amillərlə yanaşı ekstralinqvistik amillər də mühüm rol oynayır. İnformasiya-kommunikasiya texnologiyalarının, internetin coşqun inkişaf etdiyi qloballaşma dövründə ünsiyyət vasitəsi kimi aparıcı mövqeyə keçmiş SMS-lərin universal informasiya mübadiləsi vasitəsi kimi rolu, funksiyaları, inkişaf perspektivləri təhlili, onların ünsiyyət mübadiləsi prosesində bəzi linqvistik özəllikləri böyük maraq doğurur. Məlumat axtarışı sistemlərində sorğuların təşkili prosesində də SMS-lərin linqvistik xüsusiyyətlərindən yararlanmaq olar.

Qeyd etmək lazımdır ki, SMS-lərin informasiya mübadiləsində bir ünsiyyət vasitəsi kimi işlənməsi çoxaspektli, geniş əhatə dairəsinə malik bir məsələdir. SMS-lərin ünsiyyət vasitəsi kimi funksiyalarının öyrənilməsi yalnız dilçiliklə məhdudlaşdırıla bilməz. Linqvistik texnologiyalar çərçivəsində bu məsələ ənənəvi dilçiliklə yanaşı etnokulturologiya, informatika, qeyri-səlis məntiq, ehtimal nəzəriyyəsi, linqvostatistika, psixolinqvistika və s. elm sahələrinin həmnhüdüdü tədqiqat obyektii olaraq araşdırılmalıdır.

SMS – müasir peyk rabitəsi vasitələri, mobil telefonlarla bağlı geniş istifadə olunan terminlərdəndir. SMS – kiçik mətn məlumatlarının mobil (peyk) telefonları vasitəsi ilə qəbulu və ötürülməsi texnologiyasıdır. Qeyd etmək lazımdır ki, kiçik mətn məlumatları anlayışı şərti mənada işlənir və əslində çox geniş məzmun kəsb etməkdədir. Hazırda SMS vasitəsi ilə təkcə

kiçik mətn məlumatları deyil, musiqi parçaları, şəkil və müxtəlif anlayışları əks etdirən piktoqramlar, video-kliplər və s. göndərilə bilər. Müasir mobil telefonların yaddaşında müxtəlif bayram günlərini, ad günlərini, əlamətdar hadisələri əks etdirən şəkilli, musiqili, səsli açıqcalar, standart təbrik məlumatları, eləcə də, müxtəlif hiss, həyəcan, qorxu, sevinc, təəcüb, etiraz, sevgi, nifrət və s. emosiyaları əks etdirən emotiv konstruksiyalar, illüstrasiyalar yerləşdirilmişdir. İstifadəçi istənilən hadisə və informasiyaya münasibətini dərhal və sözsüz ifadə etmək imkanına malikdir. Qeyd etmək lazımdır ki, bu qəbildən olan sözsüz (qeyri-verbal) vasitələr getdikcə zənginləşməkdə, mövzu və əhatə dairəsinə görə genişlənməkdədir. Yaddaş tutumu böyük olan mobil vasitələrdə istənilən mövzu və emosiyaya aid simvol, piktoqram, illüstrasiya tapmaq mümkündür. Burada daha mühüm bir məqama diqqət yetirmək faydalı olardı. Məlum olduğu kimi, müasir mobil telefon texnologiyasının son nailiyyətləri yaddaşına, istifadə imkanlarına görə kompüterlərdən fərqlənməyən, bəzi texniki imkanlarına görə hətta onları üstələyən mobil vasitələr servisinə malikdir. İnternetdə müxtəlif resurslardan istifadə, məlumatların qəbulu, saxlanması və ötürülməsi texnologiyası da getdikcə təkmil və universal səciyyəyə daşımağa başlamışdır. Məhz bu baxımdan müasir qloballaşma şəraitində kompüter sistemləri üçün nəzərdə tutulmuş və uğurla istifadə olunan linqvistik texnologiyalar qismən və bəzən eyni ilə mobil telefonlara da şamil oluna bilər. Elektron lüğətlər korpusuna müraciət və müxtəlif təyinatlı digər korpus materiallarından optimal istifadə və faydalanma çoxfunksiyalı və böyük yaddaşlı mobil telefonlarda uğurla həyata keçirilə bilər. Hazırda geniş istifadə üçün nəzərdə tutulan və internet məkanında yerləşdirilmək üçün tövsiyə olunan bir çox tətbiqi-linqvistik səciyyəli texnologiyaların mobil versiyaları da yaradılır. Mobil telefon texnologiyasının indiki səviyyəsində bəzən mobil

versiyaların yaradılmasına ehtiyac duyulmur. Məsələ orasındadır ki, mobil telefonlar texniki göstəricilərinə və səciyyələrinə görə kompüterlərdən heç də geri qalmır [Mahmudov, 2013; Mahmudov, 2018].

SMS abbreviaturu ingilis dilində “Short Messaging Servise” ifadəsinin ilk hərfləri əsasında yaranmışdır. Mənası “qısa məlumatlar (mesajlar) xidməti (servisi)” deməkdir. Funksiyasına görə qısa mətn məlumatlarını qəbulu və ötürülməsini təmin edən (həyata keçirən) texnologiya hesab olunur [bax: ru.wikipedia.org/wiki/sms].

Məlum olduğu kimi. İnformasiya mübadiləsi prosesində ən çox istifadə olunan ünsiyyət formalarından biri məktubdur. Məktub ən qədim ünsiyyət forması kimi böyük bir inkişaf dövrü keçirmişdir. Çox-çox əvvəllər dövlətlərarası münasibətlərin qurulması və tənzimlənməsinə xidmət edən məktublar xüsusi çaparlar tərəfindən ünvan sahibinə çatdırılırdı. Cavab məktubu da eyni qaydada hazırlanır və çaparlarla yola salınırdı. Belə yazışmalar bəzən illərlə vaxt aparır, bəzən isə məktublar yol boyu yox olur lazımı ünvana çatmırdı. Sonralar poçt xidməti yaranandan sonra məktublar dəmiryolu, gəmi, lap sonralar isə təyyarə vasitəsilə göndərilməyə başladı. Müasir informasiya-kommunikasiya texnologiyalarının coşqun inkişafı əsərində məktublar onlayn olaraq internet vasitəsilə nəzərdə tutulan ünvana göndərilir. Elektron ünvan, elektron poçt anlayışları da yaxın illərdə yaranmış və hamı tərəfindən birmənalı olaraq qəbul olunmuş anlayışlardır.

Elektron onlayn yazışma bir ünsiyyət vasitəsi kimi məktublaşma prosesində, həqiqətən də, böyük inqilabi dəyişikliklər yaratdı. Məktublar həm məzmunca, həm strukturuna, həm də çatdırılma vaxtı və cavabın alınması müddətinə görə tam fərqli şəkil aldı.

Məzmun baxımından əvvəlki üslub, bərbəzəkli, təm-təraqlı ifadələrə, ştamp başlanğıc və sonluqlara ehtiyac hiss

olunmadı. Elektron məktub yığcamlığına, lakonikliyinə, “işgüzarlığına”, konkretliyinə və hər şeydən əvvəl ünvana çatma sürətinə görə əvvəllər mövcud olmuş məktub formalarından tam fərqli xüsusiyyətlərə malik bir strukturudur.

İnformasiya-kommunikasiya texnologiyaları və geniş internet şəbəkəsi yaranmazdan poçt göndərişləri ilə yola salınan məktublarının ənənəvi yazılış formaları, məktub etiketləri formalaşmışdı. Ənənəvi məktub formaları uzun və təmtəraqlı üslubda, ən azı 3-4 səhifə həcmində olurdu. Göndərilən yazılı məlumatın çox hissəsini müraciət etiketləri, dəbdəbəli, təmtəraqlı sözlər, məktubu alan şəxsin (hökmdarın) şərinə deyilən tərifli ifadələr təşkil edirdi. Məktubun kiçik həcmli (yarım səhifə, bir səhifə) olması, haradasa “hörmətsizlik” əlaməti kimi səciyyələndirilə bilərdi.

Bu ənənə adi insan qrupları çərçivəsində də bu və ya digər şəkildə təkrar və təqlid olunurdu. Qürbət ölkədən (əsgərlikdən) göndərilən məktublarda bütün ailə üzvləri, qohum-əqrəba sadalanmalı, onların durumu, hazırkı vəziyyəti, uğur və uğursuzluqları barədə məlumat verilməsi xahiş olunardı. Bunların bir çoxu ənənəvi olaraq etik baxımdan məktubda öz əksini tapmalı idi. Hər hansı bir ailə üzvünün, qohum-qonşunun adının sadalanmaması, unudulması xoşagəlməz situasiya yarada bilərdi.

Aşağıdakı nümunəyə diqqət yetirək:

“Əziz və hörmətli qardaşım Əli, salam!

Mən yaxşıyam, sən də yaxşı olmağın o bir olan uca Tanrıdan diləyirəm. Atam necədir? Əvvəlki kimi işə gedirmi? Anam necədir? Daha xəstələnmiş ki?”

Sonra həmin qayda üzrə bütün bacı-qardaş, qohum-qonşu sadalanır. Sonda əsas məsələ barədə məlumat verilir. Məktubun sonunda onun yazılma səbəbi müəyyənləşir. Məlum olur ki, ona (məktub müəllifinə) müəyyən şeylərin alınması

göndərilməsi və ya hər hansı bir işin icrası ilə bağlı kömək edilməsi xahiş olunur və s.

Belə məktublar ənənəvi başlanğıcla başladığı kimi ənənəvi sonluqla da bitmiş olurdu. Sonda yenə bütün ailə üzvlərinə, qohum-qonşulara bir daha salam-dualar göndərilir, xoş arzular söylənirdi.

SMS məktublar isə tam fərqli məzmun və strukturda olur. Onlarda çox vaxt nəinki qrammatik qayda və üslubi normalara, hətta orfoqrafik yazılış formalarına da əməl olunmur. Sözlər heç bir norma və ənənə gözlənilmədən ixtisar formasında yazılır, süni abreviaturlar yaradılır:

“Hamıya slm (salam!). Yxşyam. Sağ olun! (Yaxşıyam. Sağ olun!);

Norm (Normaldır);

İşl. OK (İşlərim OK (yaxşıdır);

Z.o gönd (Zəhmət olmasa, göndər (in));

Tşk! (Təşəkkür edirəm!);

Sğ, tşk! (Sağ ol, təşəkkür edirəm!);

Koronav (koronavirus) bizi dustaq edib;

Koronaya (koronavirusa) görə toylar dayanıb.

Bəzən yazılmış elektron məktuba, SMS-ə cavab sözlə deyil, müəyyən işarə, piktoqram, müxtəlif emotiv hissləri əks etdirən şəkillərlə ifadə olunur. Müasir mobil telefonların yaddaşında bu və ya digər hadisəyə, baş vermiş olaya, konkret insana mənfi və ya müsbət münasibəti, emosiyanı əks etdirən çoxlu sayda şərti işarələr, qrafik vasitələr, şəkillər yerləşdirilmişdir. Bu da istifadəçiyə bu və ya digər hadisəyə dərhal münasibət bildirməyə imkan verir. Həmin sözsüz vasitələrlə istifadəçi sevgi, nifrət, təəccüb, heyranlıq, vurğunluq, qibtə, etiraz və s. kimi hiss və həyəcanını ani olaraq ifadə edə bilər. Mobil telefonların yaddaşında bu və ya digər hadisə ilə bağlı minikliplər, kiçik sükətlər də yerləşdirilmişdir.

İstənilən bayramla, ad günləri, əlamətdar hadisələr və s. ilə bağlı da mobil telefonlarda hazır süjetlər, təbrik mətnləri, açıqcalar var və istifadəçilər onlardan hazır şəkildə istifadə edə bilərlər. Həmin təbrik və açıqcalar (eyni mətnlər) qısa müddət ərzində (bir neçə dəqiqə müddətində) bütün qohum, dost və tanışlara göndərilə bilər (əvvəllər poçt vasitəsilə konvertdə yola salınan təbrik mətnləri və açıqcalardan fərqli olaraq). Eləcə də müxtəlif hiss, həyəcan, qorxu, təlaş, təhlükə, sevinc, sevgi, uğur və s. ifadəsi üçün mobil telefonun yaddaşında olan hazır emotiv vasitələrdən istifadə olunur (üzün mimikası, əl barmaqlarının vəziyyəti, gözlərin hissi ifadə görünüşləri, ağız, dil, ürək, döyünən ürək və s.).

SMS-lərin özünəməxsus linqvistik xüsusiyyətləri vardır və onlar adi linqvistik (qrammatik) qaydalardan fərqli meyarlarla tənzimlənir. Başqa sözlə, mobil texnologiya özünə uyğun dilçilik yaratmaq meylindədir. Bu dilçiliyi şərti olaraq kompüter dilçiliyinin bir sahəsi kimi “SMS dilçiliyi” adlandırmaq olar. Təcrübə göstərir ki, tədricən SMS-lərin özlərinə məxsus leksikası, morfoloqiyası, sintaksisi, semantikas, frazeologiyası və s. yaranmağa və formalaşmağa başlayır. SMS dilçiliyinin əsasında, mahiyyətində intensivlik, ani reaksiya, yığcamlıq, konkretlik, mümkün qədər az sözdən istifadə, dəqiq münasibət, bu və ya digər informasiyanın qısa vaxtda geniş yayılmasının təmini (mesajın, məlumatın paylaşılması), paylaşma ilə alınmış məlumata ani münasibət və yenidən paylaşılması və s. durur. Təbiidir ki, fərqli funksiya daşıdıqları üçün SMS-lərdə çox vaxt ənənəvi qrammatik normalara əməl olunmur, ədəbi dilin sabitləşmiş qayda və qanunları lazımi səviyyədə gözlənilmir. Başqa sözlə, SMS-lərin özlərinə məxsus qrammatikası, üslubları, kor-təbii normaları yaranmış olur.

Məlum olduğu kimi, Azərbaycan dilində zərf nitq hissəsi hərəkətin və ya əlamətin əlamətini bildirir. Sadə miqdar zərfləri az, çox, xeyli, tamam sözlərindən ibarət hesab olunur. Əslində,

az, çox, xeyli, tamam sözləri bir nitq hissəsi kimi saydır, lakin iş, hal, hərəkət bildirən sözlərlə, feillərlə əlaqələndikdə adverbiallaşmış olur [Kazımov, 2010, s. 330]. 1980-ci ildə “Elm” nəşriyyatında çap olunmuş “Müasir Azərbaycan dili” kitabında kəmiyyət zərfinin iş və ya hərəkətin qeyri-müəyyən miqdarını bildirdiyi göstərilir və bu növün az miqdarda düzəltmə və zərfləşən söz qrupunu əhatə etdiyi göstərilir (azca, çoxca, olduqca, qədərincə, lazımcıca, qədərincə) [Müasir Azərbaycan dili, 1980, s.4 08].

SMS-lərdə istifadəçilər az və çox sadə zərflərindən mesajlarda daha çox istifadə edirlər.

Bir şeyin, hissın, həyəcanın, sevginin, nifrətin həddindən artıq çox olduğu mesajlarda aşağıdakı kimi ifadə olunur:

Səni çoooooooox sevirəm! (yəni, səni lap çox, həddindən artıq, dünyalar qədər sevirəm);

Səni çox, lap çoooooooox sevirəm!

Bu mesajda “o”-ların işlənmə sayı ixtiyaridir və heç bir qayda ilə tənzimlənmir, istifadəçinin sevgi fantaziyası dərəcəsi ilə fərdi olaraq tənzimlənilir.

Təkid, qətilik, qəti etiraz, qeyri-mümkünlük, təcili tələbetmə, tam əminliyin ifadəsi və s. sözün son samitinin təkrarı (uzanması) ilə yaranır:

Gələrəmmmmmm! (hökmən, sözsüz gələrəm).

Qəti etiraz, qeyri-mümkünlüyün ifadəsi:

Yooooooooo!

Heç vaxtttttt!

Dünya dağıllllllla!

Təcili tələbetmə:

Göndərrrrrr!

Yazzzzzzz!

Soruşşşşşmaaa!

Susssssss!

Hal-əhval sorğusuna cavab, ad günü, bayram təbrik-lərində sözlər qısaldılmış formada yazılır. Bu qısaltmalar çox vaxt şəxsin fərdi fantaziyasının məhsulu olur. Burada vahid qayda və prinsiplərə əməl olunur.

Norm (vəziyyətim normaldır);

OK;

Ad günü, bayram təbrikləri:

Tbr (Təbrik edirəm!);

Adg tbr (Ad gününü təbrik edirəm!);

SMS yazışmalarında şifahi danışiq üslubuna üstünlük verilir. Feilin qəti gələcək zaman forması əksər yazışmalarda ixtisarla verilir:

Saba deyəcəm (Sabah deyəcəyəm);

Evə çtb yazacam (Evə çatıb yazacağam);

Mütləq baxacam (Mütləq baxacağam);

Mktbdə görəcəm (Məktəbdə görəcəyəm).

SMS-lərdə dialekt leksikası və tələffüz formaları müşahidə olunur:

Maa yalan atma (Mənə yalan atma/danışma);

Gozuu qırpma maa (Gözünü mənə qırpma);

Bu nümunələrdə maa, gözüü ifadələri bir çox dialektlərdə geniş işlənən danışiq dilinə xas leksik vahidlərdir. A.Vəliyev keçid şivələrində mən şəxs əvəzliyinin yiyəlik hal şəkilçisi qəbul edərkən mə//məə şəklində işləndiyini göstərir. Burada yönlük hal şəkilçisi vəzifəsini uzanan “ə” səsi öz üzərinə götürür. Göyçay şivələrində I şəxsin təki yönlük halda məə, Qazax və Şəki dialektlərində isə maa kimi tələffüz olunur [Vəliyev, 2005, s.243-244].

O da özün qağan sayır (O da özün igid, qüvvətli sayır);

Qağan sözü Lənkəran və Meğri ərəzilərində işlənir və mənası igid, qüvvətli deməkdir [Azərbaycan dialektoloji lüğəti, 1999, s.301].

Gə balama qaqqə verim (Gəl balama şirniyyat, meyvə verim);

Qaqqə sözü əsasən Azərbaycanın Ağcabədi, Kəlbəcər, Ucar rayonlarında, eləcə də bir çox başqa bölgələrdə uşaq leksikonunda hər növ şirniyyat və ya meyvə mənasında işlənir [Azərbaycan dialektoloji lüğəti, 1999, s.302].

Çər gözlərinlə yaxşı bax (Çəp gözlərinlə yaxşı bax);

Cəbrayıl, Ordubad, Qarakilsə, Şərur ərazilərində, eləcə də bir çox başqa rayonlarda çər sözü çəp mənasında işlənir [Azərbaycan dialektoloji lüğəti, 1999, s.9].

Hayıl-mayıl olub getmişən özünənən (aludə olub, meyl salıb özündən getmişən);

Hayıl-mayıl: hayıl-mayıl olmaq İsmayılı, Şəki dialektlərində meyl etmək, aludə olmaq, sevmək mənasında başa düşülür [Azərbaycan dialektoloji lüğəti, 1999, s.215].

Həs-hüs qanan deyil (him-cim qanan deyil, qanmazdır).

Lənkəran, Salyan rayonlarında həs-hüs sözü qanmaz, him-cim anlamındadır [Azərbaycan dialektoloji lüğəti, 1999, s.223].

Maraqlıdır ki, SMS yazışmaları əsasında istifadəçilərin hansı regiondan olması barədə müəyyən qənaətə gəlmək olar. Təbii ki, istifadəçilərin dilində (leksikonunda) və üslubunda yaşadıkları, boya-başa çatdıqları ərazinin dialekt və şivə xüsusiyyətləri bu və ya digər dərəcədə əks oluna bilər.

Qeyd etmək lazımdır ki, SMS informasiya mübadiləsi əsasında yazışanların yaşı, təhsili, məşğuliyyəti, fəaliyyət sahəsi, maraq dairəsi barədə müəyyən fikir yürütmək olar. Mesajlarda istifadəçinin dialekt və şivə xüsusiyyətləri ilə yanaşı peşə, sənət leksikası da öz əksini tapa bilər.

SMS-lərdə ən çox işlənən və müxtəlif mənalar verən ifadələrdən biri də okay (OK) hesab olunur.

Okay (O.K.) ingilis dilində bir çox mənalarda işlənir. İsim olaraq birinci mənası “razılıq”, “bəyənmə”, “icazə” hesab olunur.

Okay (OK) sifət kimi iki mənada işlənir: 1) doğru, düzgün, həqiqi; 2) yaxşı, qaydasında.

Okay (OK), həmçinin okayed/OK-d, okaying, Oking formasında “razılaşımaq”, “razılıq vermək”, “bəyənmək” mənalarını verir.

Nəhayət, okay (OK) dördüncü mənasında nida kimi “Yaxşı!”, “Oldu!” deməkdir [İngiliscə-azərbaycanca lüğət, 2003, s.1011].

Maraqlıdır ki, SMS-lərdə OK lüğətdə göstərilənlərdən əlavə də bir çox mənalarda işlənir. Gündəlik həyatda çox rast gəlinən bəzi SMS nümunələrini nəzərdən keçirək:

Yeni il bayramınızı təbrik edir, cansağlığı, xoş günlər, uğurlar arzulayıram!

OK

Bu mesaj mübadiləsində təbrikə cavab kimi işlənmiş OK təşəkkür mənasında işlənir və təxminən aşağıdakı kimi anlaşıla bilər:

Çox sağ olun! Sizinlə bahəm. Təşəkkür edirəm! Mən də Yeni ilinizi təbrik edir, cansağlığı, xoş günlər, uğurlar arzulayıram!

Təbrikə cavab verən istifadəçi müxtəlif səbəblərdən (vaxt azlığı, yolda və ya işdə olması, müəyyən ekstremal situasiya və s. ilə əlaqədar) qısa və yığcam olaraq OK yazmaqla kifayətlənməli olur.

Doğum günü ilə bağlı eyni məzmunlu təbrik SMS-ni nəzərdən keçirək:

Doğum gününüzü təbrik edirəm! Cansağlığı, daha böyük uğurlar arzulayıram!

OK

Buradakı cavab mesajında OK müxtəlif mənalarda anlaşıla bilər: çox sağ olun, minnətdaram, diqqətinizə görə təşəkkür edirəm və s. Ola bilsin ki, doğum günü tamam olan şəxs həmin gün dalbadal bir çox qohum, qonşu, dostlardan təbrik mesajları aldığı üçün hər birinə geniş cavab verə bilmədiyindən OK yazmaqla kifayətlənməli olmuşdur. Mesaj alınan zaman onun başqa yerdə olması, qonaqlarla başı qarışması da belə qısa cavabın səbəbi ola bilər.

2020-ci il sentyabrın 20-25 tarixlərində Belarusun Minsk şəhərində Beynəlxalq Konfrans keçiriləcək. Məruzənin tezislərini 2020-ci il avqustun 20-nə kimi təqdim etmək lazımdır. Əlavə məlumatlarla konfransın təşkilat komitəsinin saytında tanış ola bilərsiniz.

OK

Burada istifadəçi OK yazmaqla həm məlumatı göndərənə təşəkkür edir, həm də məlumatla tanış olduğunu bildirmiş olur. Bunu konfransın keçirilməsi xəbərindən məmnunluq kimi də başa düşmək olar.

Sizi iyulun 10-da oğlumun toy məclisinə dəvət edirəm!

OK

Bu mesajda OK təbrik, sevinc və toyda iştirak barədə razılıq kimi qəbul oluna bilər.

Azərbaycan şahmatçısı Beynəlxalq turnirin qalibi olub!

OK

OK cavabı sevinc, məmnunluq, qürur ifadə etməklə bərabər məlumatı göndərən adamdan razılıq, bəyənilmə mənası kimi də mənalandırıla bilər.

Nümunələrdən görüldüyü kimi, mesajlarda çox tez-tez istifadə olunan OK ifadəsi olduqca geniş istifadə dairəsinə və zəngin mənə çalarlığına malikdir.

Nəticə olaraq qeyd etmək lazımdır ki, SMS-lərin informasiya mübadiləsində bir ünsiyyət vahidi kimi işlənməsi çox-aspektli, geniş əhatə dairəsinə malik bir məsələdir. SMS-lərin

ünsiyyət vasitəsi kimi funksiyalarının öyrənilməsi sırf linqvistik çərçivə ilə məhdudlaşdırıla bilməz. Fikrimizcə, SMS-lər linqvistik texnologiyalar çərçivəsində ənənəvi dilçiliklə yanaşı etnokulturologiya, informatika, qeyri-səlis məntiq, ehtimal nəzəriyyəsi, linqvostatistika, psixolinqvistika və s. elm sahələrinin həmhüdüdü tədqiqat obyektii olaraq araşdırılmalıdır. İnformasiya-kommunikasiya texnologiyalarının sonrakı inkişafı informasiya mübadiləsi vasitəsi kimi SMS-lərin də bir çox yeni funksiyalarının, fərqli aspektlərinin yaranması və inkişafını stimullaşdıracaq.

IV FƏSİL

MƏLUMAT AXTARIŞI SİSTEMLƏRİNDƏ SORĞULARIN TƏŞKİLİ MƏSƏLƏLƏRİ

Yuxarıda qeyd edildiyi kimi, süni intellekt elminin məqsədi insan aqlını, düşüncəsini, təfəkkürünü, insan beynini, bütövlükdə insan fəaliyyətini təqlid etməyə qabil olan süni sistemlərin yaradılmasıdır.

Ənənəvi olaraq süni intellekt dedikdə, insanın etməli olduğu, insana xas yaradıcı funksiyanı yerinə yetirə biləcək intellektual sistem, intellektual kompüter proqramlarının yaradılması elmi və texnologiyası başa düşülür. Bu mənada İntellektual sistem həmin sistemin yaddaşında saxlanan, yaradıcı proses kimi səciyyələndirilən məsələləri həll etməyə qabil texnologiya və ya proqram sistemi kimi səciyyələndirilir.

Müasir həyatı internetsiz, virtual ünsiyyət kanalları olmadan təsəvvür etmək çətindir. Hazırda bank ödənişlərini etmək, alış-veriş aparmaq, kommunal xərcləri vaxtında operativ ödəmək, dünyanın istənilən nöqtəsində yaşayan insanlarla yazışmaq, görüntülü söhbət etmək internet vasitəsilə mümkün olur. Müasir tələbə, tədqiqatçı-alim kitabxanalara getmədən də internet vasitəsilə istənilən ədəbiyyatı (elmi, bədii və s.) əldə etmək, ondan yararlanmaq, faydalanmaq, müəyyən lazımi informasiyanı əldə etmək imkanına malikdir. Bu və ya digər ünvanı axtaranda, müəyyən yerə getmək istəyəndə internet naviqasiya sistemləri köməyimizə gəlir. Bütün bu nəhəng informasiya qovşağı internet şəbəkəsində pərakəndə, systemsiz şəkildə deyil, müəyyən qaydalar üzrə, sistemli şəkildə yerləşdirilmişdir. Bu saysız-hesabsız informasiya cəngəlliyində lazımi məlumatı əldə etmək kor-koranə deyil, müəyyən informasiya – axtarış sistemləri vasitəsilə həyata keçirilir.

İnternetdə məlumatlar hər an yeniləşir, zənginləşir, hər an yeni-yeni məlumat və təkmilləşmələr əlavə olunur, bir çox köhnəlmiş informasiyalar yeniləri ilə əvəzlənir və ya tamamilə sistemdən kənarlaşdırılır, silinir. Bu xidmətlər ilk növbədə İnternetdəki axtarış sistemlərin köməyi ilə əldə olunur.

İnternet şəbəkəsində informasiyanın sürətli artması, məlumatın həcmnin çoxluğu axtarış prosesini çətinləşdirir. İnformasiyanın həcmi ildə təxminən dörd dəfə artır. Hər gün milyonlarla yeni sənəd yaranır və axtarış sistemləri olmadan onların İnternetdə axtarışı qeyri – mümkündür [az.wikibooks.org/wiki/internetdə_axtarış_sistemləri/axtarış_sistemləri].

Haqqında söhbət gedən axtarış sisteminin təşkilində və reallaşmasında Mikrosoft korporasiyasının qrafik interfeys vasitəsilə idarəyə istiqamətlənmiş Windows əməliyyat sistemləri mühüm rol oynayır. Windows əməliyyat sisteminin ilk versiyası 1985-ci ildə quraşdırılmışdır. Bu əməliyyat sistemi mütəmadi olaraq təkmilləşdirilir və yenilənir. Hazırda onun son və daha təkmil versiyaları istifadəçilərə təqdim olunmuşdur [ru.wikipedia.org/wiki/windows].

Axtarış sistemləri nədir və onların işi necə təşkil olunur?

Məlum olduğu kimi, kompüterin iş prinsipləri, proqram təminatı, əməliyyat sistemləri süni intellekt problemləri çərçivəsində araşdırılır. Bu sistem insan intellektinin, insan fəaliyyətinin təqlidi əsasında qurulmuşdur.

Təsəvvür edin ki, bizim iş stolumuzun üstündə və siyirtmələrində saysız-hesabsız sənəd dağınıq, pərakəndə, systemsiz şəkildə toplanmışdır. Bu sənədlər bu və digər qurumun hesabatları (aylıq, rüblük, yarımillik, illik), ayrı-ayrı təşkilatlara müxtəlif vaxtlarda, müxtəlif məqsədlərlə göndərilmiş məktub və müraciətlər, əməkdaşların fərdi planları, beynəlxalq və respublika konfranslarına təqdim olunmuş məruzə və tezislər, elmi jurnal və məcmuələrə göndərilmiş məqalələrin müəllif nüsxələri və bu qəbildən olan digər sənədlər ola bilər. Təbiidir

ki, bu qarmaqarışmış sənəd yığımı içərisindən bu və ya digər sənədi və ya məlumatı əldə etmək olduqca çətin və əziyyətli bir prosesdir.

Prosesi sadələşdirmək, sənədlərin axtarışını asanlaşdırmaq və müəyyən əlamətlər üzrə toplanmasını təşkil etmək üçün qovluqlardan istifadə etmək məqsədəuyğun hesab olunur. Hər qovluğun üzərinə onun içindəkiləri əks etdirəcək sözlər (açar sözlər) yazılır və eynitipli, eyniməzmunlu materiallar ora yığılır. Məsələn, həmin qovluqlar təxmini olaraq belə adlandırıla bilər: “Hesabatlar”, “Məktublar”, “Fərdi planlar”, “Konfrans materialları”, “Məqalələr”, “Ərizə və müraciətlər” və s. Həmin qovluqlar illər üzrə işarələnir və şkaflara müəyyən ardıcılıq və nizamla yığılır.

Adi dəftərxana işçisinin bu fəaliyyəti Windows əməliyyat sistemində bu və ya digər şəkildə təqlid və tətbiq olunur. İş (işçi) stolunda nizamsız və pərakəndə toplanmış sənədlər aid olduqları qovluqlara yığılaraq “Sənədlər” bölümünə keçirilir, bu yolla iş (işçi) stolunda toplanmış sənədlər nisbətən azaldılır, nadir hallarda istifadə olunan sənədlər stoldan yığışdırılaraq nisbətən uzaq və kənar “Sənədlər” bölümünə göndərilir. İrihəcmli, az müraciət olunan materiallar S və D disklərində də uzunmüddətli saxlana bilər. İstənilən vaxt materiallar həmin disklərdən istifadə üçün götürülə bilər. Məlum olduğu kimi, dəftərxana işçisi də ona lazım olan bu və ya digər məlumatı bloknotlarda qeydlər şəklində saxlaya bilər, çünki onlar kiçikhəcmli və daşınabiləndir. Kompüterlərdə olan informasiya da fləş-kartlara yazılıb saxlana və daşına bilər.

Göründüyü kimi, insanın dəftərxana fəaliyyəti ofis proqramlarında bu və ya digər şəkildə təqlid olunur.

Bütün bu deyilənlər axtarış sistemlərində də öz əksini tapmışdır. Dünya informasiya resursları şəbəkəsində istifadəçiyə lazım olan vacib məlumatı axtarıb tapmağa və onlardan yararlanmağa imkan verən axtarış sistemlərinin işlənilib

hazırlanması müasir intellektual sistemlərin qarşısında duran ən vacib məsələlərdəndir. Hazırda internet məkanında bir çox axtarış sistemləri istifadə olunmaqdadır. Qeyd etmək lazımdır ki, müxtəlif istiqamətli və təyinatlı axtarış sistemlərinin, məlumat kataloqlarının və onların proqram təminatının yaradılması bu sahədə köklü dəyişikliklər yaratmışdır. İndi insanlar çox vaxt onlara lazım olan informasiyanı daha asan və tez olaraq kitabxanalardan deyil, internetdən əldə edirlər.

İnternetdə informasiya axtarışı – İnternet mühitində yığılmış sənədlər çoxluğu içində istifadəçinin əldə etməyə çalışdığı və bunun üçün daxil etdiyi xüsusi informasiya sorğusuna uyğun gələn faktlar və məlumatların axtarışı prosesidir.

İnternet mühitindəki veb-saytlar və veb-səhifələr xaotik yerləşir və yüz milyonlarla veb-saytın hamısını nəzərdən keçirmək istifadəçi üçün çətin olur. Bunun üçün əlavə köməkçi vasitələrdən istifadə etmək lazım gəlir. İnternet şəbəkəsində informasiyanın axtarışı, toplanması və sistemləşdirilməsi məqsədilə axtarış sistemi (Web search engine) və ya informasiya-axtarış sistemi (Information retrieval system) portalı yaradılmışdır. Bu sahədə işlər saytların kataloqlaşdırılmasından başlanmışdır. Kataloqlaşdırma internet resurslarına müraciətlərin təşkili üsullarından biridir. Kataloqlarda informasiya resurslarının ünvanları mövzular üzrə qruplaşdırılır. Yahoo layihəsi ilk belə kataloqlardan hesab oluna bilər. Yahoo kataloqunda saytların sayının getdikcə artması ilə əlaqədar layihə daxilində axtarış sisteminə ehtiyac yaranmış oldu. Bu axtarış sistemi internet mühitində deyil, yalnız kataloq daxilindəki informasiya resursları içində axtarış aparır.

İlk axtarış sistemi 1994-cü ildə yaradılmış WebCrawler olmuşdur.

1997-ci ildə Sergey Brin və Lari Peyc tərəfindən müasir dövrdə ən populyar axtarış sistemlərindən sayılan Google yaradılmışdır.

1997-ci ildə ilk rusdilli axtarış sistemi Yandex fəaliyyətə başlamışdır. Müasir dövrdə İnternet mühitində üç ən populyar beynəlxalq axtarış sistemi fəaliyyət göstərir: Google, Yahoo və MSN Search (Bing). Hər üç axtarış sistemi öz verilənlər bazasına və axtarış alqoritminə malikdir.

İstənilən axtarış sisteminin əsas funksiyası, üzərinə düşən vəzifə istifadəçiyə lazım olan informasiyanı tez bir zamanda və dəqiqliklə çatdırmaqdır. Axtarış sistemlərinə olan əsas tələbat dolğunluq, dəqiqlik, aktualıq, axtarış sürəti və aydınlıq hesab olunur.

Axtarış sistemləri informasiyanın İnternetdə axtarışını müxtəlif üsullarla həyata keçirirlər:

kitabdakı məlumatı tapmaq üçün istifadə olunan mündəricat üsuludur. Ənənəvi və sadə hesab edilən bu üsulda kataloqlardan istifadə edilir. Bu kataloqlarda veb-səhifələrə istinadlar mövzular üzrə toplanır;

- vacib səhifələrə keçidlər mətnə göstərilir. İstifadəçi lazım olan səhifəyə keçmək üçün uyğun mətn yazılmış hipe-ristinaddan istifadə edir;

- axtarış predmet göstəricisi və ya indeksə görə aparılır. İndeks üsulunda açar sözlər və bu sözlərin yerləşdiyi səhifələrin ünvanları bazada yığılaraq saxlanılır. İstifadəçi açar sözü göndərməklə bu sözlər yazılan veb-səhifələri əldə edə bilir [az.wikibooks.org/wiki/internetdə_axtarış_sistemləri]

Bu gün İnternetin məşhur axtarış sistemləri indeksləşmə üsulunu dəstəkləyən prinsip üzrə işləyirlər. Bunun üçün axtarış sistemlərində axtarış maşını adlanan xüsusi proqramlardan istifadə edilir. Bu proqramlar özündə xüsusi proqram və verilənlər bazasını birləşdirən alqoritm olub indekslərin təşkili və saxlanmasını həyata keçirir və bu indekslərin təyin etdikləri veb-səhifələrə keçidləri təqdim edir. Axtarış maşınlarının əsas komponentləri indeksasiya modulu, verilənlər bazası və axtarış serveridir.

Müasir axtarış sistemləri özündə bir-neçə altsistemi birləşdirir:

- veb-agentlər serverlərin axtarışını təmin edir, həmin serverlərdən lazım olan sənədləri seçərək emal sisteminə ötürür;
- emal sistemi sintaktik çeşidləmədən və standart operatorlardan istifadə etməklə alınan sənədləri indeksləşdirir;
- axtarış apararı sistemlər xidmət sistemindən sorğunu qəbul edərək emal edir, indeks fayllarında axtarış aparır, topladığı keçid ünvanlarının siyahısını hazırlayır;
- xidmət sistemi istifadəçidən axtarış sorğusunu qəbul edir, uyğun şəkildə dəyişdirir və indeks faylları ilə işləyən axtarış sisteminə göndərir, əldə olunan nəticəni istifadəçiyə qaytarır.

Veb-agentlər və axtarış sistemləri bir-birlərindən asılı olmadan işləyirlər. Veb-agentlər göndərilən sorğulardan asılı olmayaraq həmişə fəaliyyətdə olurlar. Onların əsas vəzifəsi yeni informasiya serverləri haqqında məlumatları verilənlər bazasına ötürmək hesab olunur.

İstifadəçi göndərdiyi sorğu nəticəsində həyata keçirilən əməliyyatları vizual olaraq görmür. O, yalnız gətirilən keçidlərin siyahısını əldə edir. Bunlar səhifənin başlığı, səhifənin URL ünvanı və səhifənin saxlandığı veb-sayta keçidi əhatə edir (az.wikibooks.org/wiki/internetdə_axtarış_sistemləri).

Yuxarıda qeyd olunduğu kimi, İnternetin informasiya-axtarış sistemlərində axtarışın aparılması və lazımi məlumatın tapılıb istifadəçilərə təqdim edilməsi üçün mövzu kataloqları və axtarış indekslərindən istifadə olunur.

Məşhur informasiya-axtarış sistemləri aşağıdakılar hesab olunur:

Google (<http://www.google.com>,
<http://www.google.ru>, <http://www.google.az> və s.);

Baidu (<http://www.baidu.com>);
Yahoo (<http://www.yahoo.com>);
Yandex (<http://www.yandex.ru>);
Bing (<http://www.bing.com>);
AltaVista (<http://www.altavista.com>);
Lycos (<http://www.lycos.com>);
HotBot (<http://www.hotbot.com>);
Rambler (<http://www.rambler.ru>);
Aport (<http://www.aport.ru>);
STARS (<http://www.stars.ru>);
Русский Интернет [<http://www.rosit.ru>; Qasımov, 2015,

s.12]

İnternet məkanında geniş istifadə olunan 50 ən populyar saytın 12-sini “Google” şirkətinin müxtəlif ölkələrdə fəaliyyət göstərən filialları hesab olunur. “Yahoo” axtarış sistemi isə 2 yerlə (əsas və Yaponiya filialı) təmsil olunur.

Siyahıda ABŞ saytları üstünlük təşkil edir. Çinin 6 (sohu.com, 163.com, taobao.com, sina.com.cn, qq.com, baidu.com), Rusiyanın 4 (yandex.ru, mail.ru, vkontakte.ru, odnoklassniki.ru), Yaponiyanın 1 (fc2.com), İngiltərənin 1 (bbc.co.uk) saytları 50 ən populyar sayt siyahısına düşə bilib.

Ən populyar saytlar aşağıdakılardır:

1. google.com;
2. yahoo.com;
3. facebook.com;
4. ebay.com;
5. msn.com;
6. youtube.com;
7. amazon.com;
8. live.com;
9. microsoft.com;
10. wikipedia.org;
11. wordpress.com;

12. aol.com;
13. myspace.com;
14. bing.com;
15. craigslist.org;
16. go.com;
17. blogger.com;
18. photobucket.com;
19. livejasmin.com;
20. twitter.com;
21. flickr.com;
22. imdb.com;
23. yandex.ru;
24. mail.ru;
25. linkedin.com;
26. baidu.com;
27. yahoo.co.jp;28. qq.com;
29. google.co.in;
30. google.cn;
31. sina.com.cn;
32. google.de;
33. vkontakte.ru;
34. google.co.uk;
35. taobao.com;
36. google.fr;
37. 163.com;
38. google.ru;
39. google.co.jp;
40. rapidshare.com;
41. bbc.co.uk;
42. odnoklassniki.ru;
43. google.com.br;
44. fc2.com;
45. google.es;

46. sohu.com;
47. google.com.mx;
48. google.it
49. pornhub.com;
50. doubleclick.com [news.milli.az/interest/1634.html].

Qeyd etmək lazımdır ki, saytlar müəyyən müddətdə fəaliyyət göstərdikdən sonra ləğv oluna, yenisi ilə əvəzlənə bilər. Ona görə də, internet məkanında fəaliyyət göstərən saytlara etibarlı mənbə kimi istinad etmək risklidir. Müəyyən müddətdən sonra bizim istinad etdiyimiz mənbə internet məkanında olmaya bilər. Vəziyyətdən çıxış yolu kimi istifadə olunmuş internet resursundan yararlanma tarixi göstərilməlidir. Vikipediya məlumatları da etibarlı mənbə kimi səciyyələndirilə bilməz. Orada məqalə müəlliflərinin peşəkarlıq səviyyəsi heç də həmişə qənaətbəxş olmur. Edilən redaktələr də çox vaxt yüksəkixtisaslı mütəxəssislər tərəfindən aparılır. Məhz ona görə də elmi tədqiqat əsərlərində etibarlı, nüfuzlu mənbələrdə çap olunmuş faktlara istinad edilməsi tövsiyə olunur. İnternet saytları, vikipediya resursları ensiklopedik səciyyəli ümumi məsələlərə aid məlumat mənbəyi kimi səmərə verə bilər. Digər tərəfdən, internet texnologiyası, resurslar, populyar saytlar, bir çox texnoloji yeniliklər barədə operativ məlumata yalnız vikipediya məlumatlarında rast gəlmək mümkündür.

V.Ə.Qasımov “İnformasiya axtarışı üsulları və sistemləri” dərsliyində müxtəlif mənbələrdə informasiya anlayışının izah və şərhlərinin xülasəsini vermişdir:

- informasiya – işarə və siqnallar vasitəsilə mövzu sahəsinin əks olunmasıdır;
- informasiya – xəbərlər, yeni faktlar, yeni biliklərdir;
- informasiya – məlumatın və ya xəbərin məzmunudur;
- informasiya – mövcud qeyri-müəyyənliyi aradan qaldıran məlumat və ya xəbərdir;

- informasiya – insanın ətraf aləmdə olan obyektlər və baş verən hadisələr haqqında məlumatlılıq səviyyəsini artıran xəbərlərdir;

- informasiya – fəlsəfi mənada real aləmin əksi kimi qəbul edilir.

İnformatika elmi baxımından informasiyaya aşağıdakı tərif verilir.

İnformasiya – yaranma tarixindən, təqdimat formasından və təsnifatından asılı olmayaraq şəxslər, əşyalar, faktlar, hadisələr və proseslər haqqında məlumatlar və bilgilərdir [Qasımov, 2015, s.15-16]

Obyektiv reallığın subyektiv obrazı, başqa sözlə daxili və xarici aləmin insanın şüurunda təqdimatlar, anlayışlar, mülahizələr, nəzəriyyələr formasında adekvat əksi olan biliyi süni intellekt, biliklər bazası və ekspert sistemləri baxımından V.Ə.Qasımov belə izah edir: “Süni intellekt, biliklər bazaları və ekspert sistemləri baxımından biliklər dedikdə ətraf aləm haqqında verilənlər, faktlar, xəbərlər və nəticə çıxarma qaydaları, o cümlədən obyektlərin xassələri, proses və hadisələrin qanunauyğunluqları barədə informasiyanı, qərarların qəbul edilməsi üçün bu informasiyanın istifadə olunması qaydalarını özündə saxlayan toplu başa düşülür.

Süni intellektin biliklər bazalarını və biliklərlə iş üsullarını öyrənən bölməsi biliklər mühəndisliyi adlanır [Qasımov, 2015, s.22].

Məlumat-sorğu funksiyaları dedikdə sorğu edilmiş informasiyanın dəqiqləşdirilməsi və ya konkretləşdirilməsi, habelə obyektə bağlı daha çox məlumatın əldə olunması, rekvizitlərin, xarakteristikaların və göstəricilərin dəqiqləşdirilməsi və s. başa düşülür. Web-texnologiyanın, o cümlədən HTML (HyperText Markup Language – hipermətnlərin təqdim edilməsi dili) dilinin və HTTP (HyperText Transfer Protocol) protokolunun meydana gəlməsi ilə əlaqədar internetdə ilk

olaraq informasiya axtarışı imkanları olmayan sənədli informasiya sistemləri yaranmağa başladı [Qasimov, 2015, s. 40].

İnternet – ən yeni texnologiyalara əsaslanan, geniş spektrdə informasiya və kommunikasiya ehtiyatlarına malik olan qlobal informasiya fəzasıdır. İnternet çox böyük həcmdə məlumatları özündə saxlayır.

Hazırda internet dedikdə fiziki kompüter şəbəkəsi yox, dünya hörümçək toru WWW (World Wide Web – dünya hörümçək toru) üzərində formalaşmış informasiya fəzası nəzərdə tutulur.

İnternetin yaranma tarixi 2 sentyabr 1969-cu il hesab olunur [Qasimov, 2015, s.57-58; az.wikibooks.org/wiki/internet/Yaranma_tarixi]

İnternet şəbəkəsinin əsas informasiya və telekommunikasiya xidmətləri aşağıdakılardır:

- WWW (World Wide Web) – dünya hörümçək toru;
- E-mail (Electronic mail) – elektron poçtu, məlumatların göndərilməsi siyahıları;
- FTP (File Transfer Protocol) – faylların ötürülməsi protokolu;
- Gopher – iyerarxik menyu;
- Telnet – uzaq məsafədən idarəetmə və sistemə daxilolma;
- BBS (Bulletin Board System) – elektron elan lövhələri
- Usenet – telekonfrans (xəbərlər) sistemi;
- İnternetin interaktiv informasiya xidmətləri;
- İnternet-telefon, internet-radio və internet-televiziya [az.wikibooks.org/internet/yaranma_tarixi].

Wiki layihələr Web üzərindən olduqca böyük informasiya massivlərinin formalaşmasına gətirib çıxarmışdır (məsələn, Wikipedia, WikiLeaks):

- Facebook – 1, 4 milyard;
- YouTube – 1 milyard;
- WhatsApp – 600 milyon;
- Google+ – 540 milyon;
- Twitter – 500 milyon;
- LinkedIn – 260 milyon;
- V Kontakte – 230 milyon;
- Instagram – 200 milyon;
- Odnoklasniki – 205 milyon [Qasimov, 2015, s.65-66; en.wikipedia.org/wiki/google].

İnformasiya-axtarış sistemləri – internetdə informasiya axtarışını həyata keçirməyə imkan verən web- interfeysli proqram-texniki kompleksidir.

Elektron poçtu (“e-mail” və ya “email” – ingiliscə “electronic mail” sözlərindən) – paylanmış, o cümlədən, qlobal (İnternet) kompüter şəbəkəsində elektron məlumatların göndərilməsi və alınması texnologiyası və xidmətidir. Elektron poçtu internetin ilk informasiya xidmətidir.

İnformasiya axtarışı hər hansı statistik sənədlər kolleksiyasından axtarış sorğusuna uyğun olan sənədlərin tapılmasından ibarətdir. Bura modelləşdirmə, sənədlərin təsnif edilməsi, süzğəcdən keçirilməsi və klasterləşdirilməsi, axtarış sistemlərinin arxitekturasının və istifadəçi interfeyslərinin layihələndirilməsi, sənədlərin xülasələrinin və referatlarının hazırlanması, informasiyanın çıxarılması, informasiya-axtarış dilləri və s. məsələlər aid edilir.

Ümumiyyətlə, informasiya axtarışının (bibliografik, sənədli, fakto-qrafik, analitik) həyata keçirilməsinin əsas məqsədi aşağıdakılardan biri ola bilər:

- informasiya mənbəyi haqqında zəruri məlumatların axtarılması, bibliografik informasiya və bibliografik vəsaitlərin istifadəsi;

- lazım olan informasiyanı özündə saxlayan informasiya mənbələrinin (sənədlərin və nəşrlərin) özlərinin axtarılması;
- nəşrlərdə olan faktiki (məsələn, tarixi faktlar və hadisələr, maşın, qurğu və proseslərin texniki xarakteristikaları, alimin, mühəndisin biblioqrafik verilənləri) məlumatların axtarılması;
- müxtəlif miqyaslı, ətraflı, dərin və s. tədqiqatlara əsaslanan ümumi xarakterli analitik informasiyanın axtarılması [Qasımov, 2015, s.126].

İnternetdə informasiya axtarışını həyata keçirən Web-Crawler, Excite, İfoseek, Yahoo, Tradeware Galaxy və s. axtarış sistemləri 1994-cü ildən fəaliyyətə başlamışlar.

İnternetin ilk tammətli axtarış sistemi olan WebCrawler 1994-cü ildə Amerika Online şirkəti tərəfindən yaradılmışdır. Əvvəlki axtarış sistemlərindən fərqli olaraq, WebCrawler istifadəçilərə istənilən web-səhifədə istənilən açar sözə görə axtarış aparmağa imkan verirdi.

Yahoo axtarış sistemi 1994-cü ilin əvvəlində Stendford Universitetinin aspirantları Devid Faylo və Cerri Yanq tərəfindən dünya hörümçək toru üzrə bələdçi sayt kimi yaradılmışdır. Bu sayt digər saytların kataloqunu özündə əks etdirirdi. Yahoo dünyanın ən məşhur axtarış sistemlərindən biridir.

Yahoo – 1995-ci ildə Stenford Universitetinin tələbələri Ceri Yanq (Jerry Yang) və Devid Filo (David Filo) tərəfindən ABŞ-da yaradılmış internet portalıdır. Əvvəllər axtarış sistemi funksiyasını yerinə yetirirdi. Sonralar Yahoo-nun xidmət növləri genişləndirildi. Məsələn, elektron poçt, xəbərlər, idman, birja, hava, oyunlar, alış-veriş, səyahət və s. kimi müxtəlif növ xidmətlər də Yahoo tərəfindən yerinə yetirilməyə başladı.

Ceri Yanq və Devid Filo portala ad verməkdə çətinlik çəkirdilər. Onlara müxtəlif adlar təklif olunurdu. Nəhayət, Ceri və Devid “Yahoo” adını seçdilər. Yahoo adı 18-ci əsr ingilis

yazıçısı Conatan Sviftin (Jonathan Swift) klassik əsəri “Qulliverin macəraları” əsərindən götürülmüşdür. Həmin əsərdə “yahoo” ifadəsi “kobud, inkişaf etməmiş insan” mənasında işləndirdi [<https://az.wikipedia.org/wiki/yahoo!>].

Lycos axtarış sistemi Karneqi Melon Universitetində Maykl Loren Molden tərəfindən aparılan tədqiqatların nəticəsi olaraq 1995-ci ildə meydana gəldi. Lycos internetin ilk axtarış maşınlarından biridir. Lycos sisteminin 15-dən çox Avropa dillərində interfeysi təqdim olunur.

2014-cü ilin əvvəlinə olan statistik məlumata görə bu gün dünyanın ən tanınmış və geniş istifadə olunan axtarış sistemi Google sistemidir. O, dünyada informasiya axtarışı bazarının 70,91%-ni təmin edir. İkinci yeri Çin axtarış sistemi olan Baidu (18,51%) tutur. Növbəti yerlərdə Yahoo, Yandex, Bing, AOL və Ask axtarış sistemləri gəlir Bu göstəricilər çox dinamikdir, coğrafi və demoqrafik məkanlardan asılı olaraq fərqlənə və zaman keçdikcə dəyişə bilər. Yandex – 53%.

Azərbaycanda da ən çox istifadə olunan axtarış sistemi Google sistemidir. İstifadə edənlərin sayına görə ikinci yeri Yandex axtarış sistemi tutur.

Yandex xidməti (Yandex.ru saytı) – Rusiyanın ən çox istifadə olunan veb axtarış motorudur. Yandex digər post-sovet ölkələrində də geniş istifadə olunur. 1989-cu ildə sahibkar və proqramçı Arkadi Voloj fərdi kompüterlər satan və iş yerlərinin avtomatlaşdırılması ilə məşğul olan CompTek firmasını təsis etmişdir. O, məlumatların emal alqoritmləri və dilin morfoloqiyası əsasında irihəcmli mətnlərdə informasiya axtarışını həyata keçirə biləcək proqram hazırlamaq imkanı ilə maraqlanırdı. Kompüter dilçiliyi üzrə mütəxəssis Arkadi Borkovski ilə birlikdə 1989-cu ildə “Arkadi” şirkətini təsis etmişdi. 1993-cü ildə kompüterin bərk diskində axtarış proqramı yaradıldı. Texnologiya üzrə direktor İlya Seqaloviç və şirkətin baş direktoru Arkadi Voloj proqramı “Yandex”

adlandırdılar. Seqaloviç texnologiyanın mahiyyətini təsvir edən sözlərdən fərqli törəmələri ayırdı. Nəticədə, “yandex” seçimi ortaya çıxdı – Yet ANother Indexer (“başqa bir indeksator”). “Yandex.ru” axtarış sistemi rəsmən 23 sentyabr 1997-ci ildə İlya Seqaloviç tərəfindən Softool sərgisində təqdim edilib. İnformasiya Ötürülməsi Problemləri İnstitutu ilə əməkdaşlıq edən CompTek rus dilinin morfolojiyası əsasında bu axtarış lüğətinin hazırlanmasına nail olmuşdur. 1999-cu ilin ortalarında “Yandex” Rusiyanın ən populyar internet saytlarının yeddincisi idi, Müstəqil olaraq “Yandex” 2000-ci ildən fəaliyyətə başlayıb.

“Yandeks” şirkətinin əsas ofisi Moskvada yerləşir. Şirkətin Sankt-Peterburq, Yekaterinburq, Novosibirsk, Nijni Novqorod, Odessa, Simferopol, Kiyev, Kazan, Rostov, Minsk, İstanbul, Sürix və Lüsern şəhərlərində ofisləri fəaliyyət göstərir. Şirkət 2008-ci ilin iyun ayında ABŞ-ın Kaliforniya ştatında “Yandex Labs”ın açılışını edib.

Axtarış mexanizmi hazırlaması şirkətin əsas və prioritet istiqamətidir. “Yandex” 50-dən çox xidmətə malikdir. Bunlardan ən məşhurları sırasında “Yandex. Şəkillər”, “Yandex. Poçt”, “Yandex. Xəritələr”, “Yandex. Hava”, “Yandex. Xəbərlər” və başqaları yer alır.

Alexa.com reytinginə əsasən, 2013-cü ildə məşhurluğuna görə yandex.ru saytı dünyada 18-ci və Rusiyada 1-ci yeri tutub. 10 oktyabr 2017-ci il tarixində Yandex “Alice” səs assistentini təqdim etmişdir.

“Yandex” – artıq yalnız internet-səhifə deyil, həm də brauzerdir. Belə ki, 2012-ci ildə şirkətin brauzerinin birinci versiyası təqdim edilmişdir. 2013-cü ilin əvvəlində “Yandex” 4, 84 milyard axtarış sorğusu ilə planetin dördüncü axtarış motoru olmuşdur. Bu göstəriciyə görə yalnız Google, Baidu və Yahoo!-dan geridə qalır.

Brauzer Rusiya şirkətinin internet xidmətləri – (axtarış, xəritə, poçt, tərcüməçi və s.) ilə sıx şəkildə əlaqələndirilmişdir. Digər müasir brauzerlərin əksəriyyətində olduğu kimi, “Yandex”in brauzerində də ünvan sətrinə həm saytların ünvanlarını, həm də axtarış sorğularını daxil etmək mümkündür.

Brauzer “Adobe Flash” texnologiyasını dəstəkləyir və ona, həmçinin “pdf” fayllara baxış utiliti inteqrasiya edilmişdir.

Portalın əsas xidmətlərindən biri “Yandex.Tərcümə” sözlərin, söz birləşmələrinin, bütöv mətnlərin, həmçinin veb-səhifələrin avtomatik tərcümə servisidir. Servis, “Yandex”də yaradılan maşın tərcüməsi texnologiyasını istifadə edir. Servis iki rejimdə işləyir: mətnin tərcüməsi və veb-səhifənin tərcüməsi.

Qeyd edək ki, “Yandex” şirkəti tərcümə servisinə Azərbaycan dilini də əlavə edib.

Hazırda “Yandex.Tərcümə” servisi 36 dildə tərcümə edə bilir [az.wikipedia.org/wiki/yandex].

Конец формы

Google – dünyada internetin ən məşhur, ən çox istifadə olunan və ən geniş yayılmış informasiya-axtarış sistemidir.

Bu terminin istifadəsi Google şirkətinin və onun axtarış sisteminin internetin nəhəng, praktiki olaraq, sonsuz informasiya massivində əlverişli axtarışın təşkilinə yönəlmiş məqsədini özündə əks etdirir.

Bu gün Google axtarış sistemi eyni adlı şirkət tərəfindən idarə olunur. Google şirkəti öz imkanlarını ideal axtarış serverinin işlənilib hazırlanması istiqamətində cəmləşdirən şirkətdir. Şirkətin təsisçilərinin fikrinə görə Google axtarış serveri istifadəçilərə nə lazım olduğunu başa düşür və onlara məhz istədiklərini təqdim edir. Bu məqsədlə Google şirkəti daim yeni texnologiyalar işləyib hazırlayır və tətbiq edir.

1999-cu ilin əvvəlində Palo-Alto şəhərində Avenyu Universitetində əsas ofisinə köçən, artıq 8 əməkdaşı olan və gün ərzində 500000 sorğu emal edən Google Inc. Şirkətinə maraq

günbəgün artırdı. Google 1999-cu ildə Time jurnalı tərəfindən dünyanın ən yaxşı kibernetik texnologiyalar onluğuna daxil edildi.

Peyç və Brin yeni yaratdıqları axtarış maşınına əvvəlcə "BackRub" adı vermişdilər, çünki bu saytların geri bağlantılarının yoxlanılması baxımından əhəmiyyətli idi. Sonralar "quqol" sözündə müəyyən dəyişiklik edilərək axtarış mühərriki "Google" adlandırıldı. Google adı istifadəçilərin irihəcmli bilik mənbəyinə asan giriş etmək imkanı olduqları üçün qoyulmuşdur. "Quqol" on üstü yüz (10100) deməkdir. Başlangıçda "Google" Stanford Universitetinin altfəza adı olaraq google.stanford.edu fəza adı kimi işlənmişdir [[https:// az.wikipedia.org/wiki/Google](https://az.wikipedia.org/wiki/Google)].

Google artıq 28 dildə, o cümlədən türk və ərəb dillərində axtarış aparmağa imkan verirdi. 2006-cı ildə Google şirkəti video-çarxların mübadiləsi sistemi olan Youtube saytını aldı. Bu gün milyonlarla video-çarxı özündə saxlayan Youtube saytına hər dəqiqə ərzində 60 saatlıq video yüklənir. 2007-ci ildə mobil qurğular üçün Android əməliyyat sistemi, 2008-ci ildə Google Chrome brauzeri istifadəyə verildi, 2011-ci ildə isə Google+ layihəsi həyata keçirildi.

Hal-hazırdakı Google Android əməliyyat sistemi, Youtube, Google Plus, Nexus qurğuları, axtarış və nizamlama mexanizmlərinin təkmilləşdirilməsi və digər sahələrdə müxtəlif yeniliklər və yeni texnologiyalar tətbiq etməkdə davam edir. Hazırda Google tərcüməçi sistemində dillərin ümumi sayı 80-i ötüb keçmişdir.

Bloqlar üzrə Google axtarış internetdə olan bloqların məzmunları üzrə ən sadə axtarış növüdür. Google-un web-axtarış sistemi tərəfindən idarə olunan bloqlar üzrə axtarış modulu şəbəkəyə qoşulmuş olan milyonlarla bloqlar (bloqların adları, müəllifləri, tarixləri və məlumatları) üzrə axtarış aparır və yeni, relevant nəticəni təqdim edir.

Google kitab axtarışı kitabların məzmunlarının indeksi olub, istifadəçiyə lazımı kitabı tapmağa imkan yaradır.

Google Catalogs vasitəsilə 6600-dən artıq bağlama ticarəti kataloqlarının tam mətninə giriş təqdim edilir.

Google-un baza axtarış texnologiyasının tətbiqi ilə yaradılmış və sazlanan istifadəçi axtarış sistemi axtarışın nəticəsini istifadəçi tərəfindən göstərilmiş web-saytlar və web-səhifələr üzrə nizamlayır və yaxud məhdudlaşdırır.

Google kataloqu mövzu kateqoriyaları üzrə təşkil olunmuş web-saytlara baxmağa və onlar üzrə axtarış aparmağa imkan verir.

Google – maliyyə xidməti fondlar, açıq və qapalı aksioner şirkətlər haqqında işgüzar və maliyyə məlumatlarına tez və sadə girişi təqdim edir.

Gmail Google-un pulsuz elektron poçt xidmətidir. Gmail istifadəçilərə pulsuz yaddaş və Google-un daxili axtarış texnologiyasını təklif edir. Gmail güclü süzgəc vasitəsilə daxil olan elektron poçt göndərişləri arasında spamları (arzu olunmaz poçt göndərişlərini) aşkar edir və istifadəçilər üçün yarana biləcək narahatçılığı aradan qaldırır. Eyni zamanda, Gmail antivirus qorumasını təmin edir.

Mobil qurğular üçün Google məhsulları telefonlar, planşetlər və s. üçün xüsusi olaraq elə şəkildə işlənib hazırlanmışdır ki, zəruri informasiya həmin qurğuların ekranına yerləşə bilsin. Mobil qurğular üçün Google məhsulları aşağıdakı funksiyaları təqdim edir:

- web-axtarış;
- şəkillərin axtarışı;
- Google Map;
- WAP-axtarış;
- Google SMS (Google-un qısa məlumatlar xidməti).

Google News xidməti dünyanın 10000-dən artıq xəbər mənbəyindən məlumatın toplanmasını həyata keçirir, xəbər

məlumatlarını internetdə dərc olunandan sonra bir neçə dəqiqə ərzində axtarış üçün müvafiq formatda təqdim edir.

İnternetin milyonlarla web-səhifələrinə geniş giriş imkanları ilə yanaşı Google-un web-axtarış xidmətinə şəbəkədə istifadəçilərə məhz axtarıqlarını tapmağa imkan verən xüsusi funksiyalar əlavə edilmişdir. Bunlara aşağıdakıları aid etmək olar:

- istifadəçilərə daha mürəkkəb sorğular tərtib etməyə imkan verir;
- sorğuda orfoqrafik və ya texniki səhvlərin olması ehtimal edildikdə alternativ axtarış şərti təklif olunur;
- fransız, alman, italyan, portuqal, rus, ispan və digər dillərdə nəşr olunmuş web-səhifələri ingilis və əksinə tərcümə edir.
- küçələrin xəritələrinə istinadları təqdim edir və avtomobil surnə üçün istiqamət verir;
- sadə hesablamaların aparılması, mürəkkəb riyazi məsələlərin hesablanması, ölçü vahidlərinin çevrilməsi və s.;
- sözlərin və ya ifadələrin mənalarını təqdim edir;
- küçə ünvanlarını, müəssisələrin və yaşayış evlərinin telefon nömrələrini təqdim edir;
- mobil qurğular (smartfonlar, planşetlər, netbuklar və s.) vasitəsilə internetin xidmətlərindən istifadə edilməsi üçün xüsusi funksiyalar işlənilib hazırlanmışdır;
- şəkillər üzrə və ya səsli axtarış kimi yeni axtarış üsulları reallaşdırılmışdır;

Axtarış sorğusunun və şərtlərinin düzgün tərtib edilməsi və seçilməsi axtarışın uğurla nəticəsini təmin edir. Ümumi axtarış sorğusunda bir sözdən də istifadə etmək olar. Lakin effektiv nəticənin əldə edilməsi üçün sorğuda bir neçə axtarış sözündən istifadə məqsədəuyğun hesab olunur [en.wikipedia.org/wiki/google; [ru.wikipedia.org/wii/google_ \(компания\)](http://ru.wikipedia.org/wii/google_(компания)) az.wikipedia.org/wiki/google].

SAYTLARIN SORĞU SİSTEMLƏRİ



Azərbaycan Milli Elmlər Akademiyası
Azərbaycanın ilk veb saytı (1995)

[Elm](#) [TV](#) [Foto](#) [Video](#) [EN](#) [RU](#)

[Начало формы](#)

[Конец формы](#)

[AMEA HAQQINDA](#)

[AMEA-nın PREZİDENTİ](#)

[AMEA RH APARATI](#)

[SƏNƏDLƏR](#)

[TRİBUNA](#)

[XƏBƏRLƏR](#)

[ƏLAQƏ](#)

Müəllif hüquqları © Azərbaycan Milli Elmlər Akademiyası 1995 – 2023

Azərbaycan Milli Elmlər Akademiyasının saytı Azərbaycanın 1995-ci ildən fəaliyyət göstərən ilk rəsmi veb saytıdır. Həmin saytın axtarış sistemini nəzərdən keçirək. Saytın ana

səhifəsi yeddi bilmədən ibarətdir. Həmin bölmələr yuxarıdakı şəkildə göstərilmişdir [science.gov.az/az; elm.a].

AMEA haqqında bölməsi aşağıdakı yarım bölmələrdən ibarətdir:

Yaranma tarixi;
AMEA-nın təsisçiləri;
AMEA-nın sabiq prezidentləri;
AMEA-nın strukturu;
AMEA-nın üzvləri;
Sosial xidmətlər;
Dövri nəşrlər;
Tarixi binalar.
AMEA-nın prezidenti bölməsi:
Tərcümeyi-hal;
Görüşlər;
Prezidentin qəbulu.
AMEA RH Aparatı:
Rəhbərlik;
İdarə və şöbələr;
RH-nin nəzdindəki təşkilatlar.

Sənədlər:
Elm haqqında AR Qanunu;
Rəyasət Heyətinin Qərarları.
Tribuna:
Bəyanatlar;
Müsabiqələr;
Məqalələr.
Xəbərlər:
Konfranslar, iclaslar;
Anonslar;
Əlaqə.

Bu bölmələrin hər birinin tərkibində də yarım bölmələr var. Məsələn, AMEA-nın strukturu bölməsini açsaq, onun tərkibində aşağıdakı yarım bölmələri görə bilərik:

AMEA-nın Elmi Bölmələri;
İnstitutlar;
AMEA-nın Regional Bölmələri;
AMEA-nın Regional Elmi Mərkəzləri;
Təşkilatlar.

Bunların hər birinin tərkibində də uyğun qurum və təşkilatlar var. Məsələn, AMEA-nın Elmi Bölmələri aşağıdakılardır:

Fizika-Riyaziyyat və Texnika Elmləri Bölməsi;
Kimya Elmləri Bölməsi;
Yer Elmləri Bölməsi;
Ekologiya və Tibb Elmləri Bölməsi;
Humanitar Elmlər Bölməsi;
İctimai Elmlər Bölməsi.

İnstitutlar bölməsində AMEA-nın nəzdində fəaliyyət göstərən aşağıdakı elmi-tədqiqat institutları verilmişdir:

Fizika-Riyaziyyat və Texnika Elmləri Bölməsi
Fizika İnstitutu;
İdarəetmə Sistemləri İnstitutu;
Riyaziyyat və Mexanika İnstitutu;
Radiasiya Problemləri İnstitutu;
Nəsrəddin Tusi adına Şamaxı Astrofizika Rəsədxanası;
İnformasiya Texnologiyaları İnstitutu;
Botanika İnstitutu.

Kimya Elmləri Bölməsi

Akademik Yusif Məmmədəliyev adına Neft-Kimya Problemləri İnstitutu;

Akademik Murtuza Nağıyev adına Kataliz və Qeyri-üzvi Kimya İnstitutu;

Akademik Əli Nağıyev adına Aşqarlar Kimyası İnstitutu;
Polimer Materialları İnstitutu;

Yer Elmləri Bölməsi
Neft və Qaz İnstitutu;
Akademik Həsən Əliyev adına Coğrafiya İnstitutu;
Geologiya və Geofizika İnstitutu;
Respublika Seysmoloji Xidmət Mərkəzi;
Biologiya və Tibb Elmləri Bölməsi
Molekulyar Biologiya və Biotexnologiyalar İnstitutu;
Botanika İnstitutu;
Zoologiya İnstitutu;
Akademik Abdulla Qarayev adına Fiziologiya İnstitutu;
Mikrobiologiya İnstitutu;
Mərkəzi Nəbatat Bağı;
Dendrologiya İnstitutu;
Torpaqşünaslıq və Aqrokimya İnstitutu;
Genetik Ehtiyatlar İnstitutu.
Humanitar Elmlər Bölməsi
Məhəmməd Füzuli adına Əlyazmalar İnstitutu;
Memarlıq və İncəsənət İnstitutu;
Folklor İnstitutu;
Nizami Gəncəvi adına Milli Azərbaycan Ədəbiyyatı
muzeyi;
Nizami Gəncəvi adına Ədəbiyyat İnstitutu;
Nəsimi adına Dilçilik İnstitutu;
Hüseyn Cavidin Ev Muzeyi.
İctimai Elmlər Bölməsi
Arxeologiya, Etnoqrafiya və Antropologiya İnstitutu;
Fəlsəfə və Sosiologiya İnstitutu;
İqtisadiyyat İnstitutu;
Qafqazşünaslıq İnstitutu;
Milli Azərbaycan Tarixi Muzeyi;
Abbasqulu Ağa Bakıxanov adına Tarix İnstitutu;
Akademik Ziya Bünyadov adına Şərqsünaslıq İnstitutu.

AMEA-nın regional bölmələri
Naxçıvan Bölməsi;
Gəncə Bölməsi.
AMEA-nın regional Elmi Mərkəzləri
Şəki Regional Elmi Mərkəzi;
Lənkəran Regional Elmi Mərkəzi;
Quba Regional Elm Mərkəzi.
Təşkilatlar
AMEA-nın nəzdində ictimai təşkilatlar;
AR Elmi Tədqiqatların Əlaqələndirilməsi Şurası.

AMEA-nın strukturunda olan hər bir institut və təşkilatın da müstəqil saytları fəaliyyət göstərir.



**Azərbaycan Milli Elmlər
Akademiyası Nəsimi adına
DİLÇİLİK İNSTITUTU**

AMEA-nın strukturunda olan Nəsimi adına Dilçilik İnstitutunun müstəqil veb saytı aşağıdakı bölmələrdən ibarətdir (dilcilik.az):

Ana səhifə Xəbərlər Ən son xəbərlər Elanlar
Haqqımızda;
Dissertasiya Şurası;
E-kitabxana;
Elm və təhsil;
Jurnallar;
Əlaqə.

Haqqımızda
İnstitutun tarixi;

Şöbələr;
Elmi Şura;
Gənc Alim və Mütəxəssislər Şurası;
Həmkarlar təşkilatı.
Dissertasiya Şurası
Dissertasiya Şurasının tərkibi;
Elmi seminarın tərkibi;
Dissertasiyanın tərtibi qaydası;
Avtoreferatın tərtibi qaydası;
Sənədlərin siyahısı;
Elanlar.
E- kitabxana
Haqqımızda;
Elektron kataloq;
Yeni nəşrlər;
Avtoreferatlar;
Araşdırmalar;
Elm və təhsil.
Təhsil şöbəsi
Doktorantura;
Magistratura;
Hesabatlar.
Jurnallar
Tədqiqlər;
Dilçilik araşdırmaları;
Türkologiya.
Əlaqə
Ünvan və e-mail

Haqqımızda bölümündə Şöbələr yarım bölməsində İnstitutun şöbələri barədə məlumat verilir:

Azərbaycan dilinin tarixi şöbəsi;
Azərbaycan dilinin dialektologiyası şöbəsi;

Müasir Azərbaycan dili şöbəsi;
Nəzəri dilçilik şöbəsi;
Tətbiqi dilçilik şöbəsi;
Kompüter dilçiliyi şöbəsi;
Qədim dillər və mədəniyyətlər şöbəsi;
Nitq mədəniyyəti və monitorinq şöbəsi;
Sosiolinqvistika və psixolinqvistika şöbəsi;
Türk dilləri şöbəsi;
Hind-Avropa dilləri şöbəsi;
Terminologiya şöbəsi;
İctimaiyyətlə əlaqələr şöbəsi;
Təhsil şöbəsi;
Eksperimental fonetik araşdırmalar laboratoriyası;
Toponimika şöbəsi.
Linklər
Tədqiqlər jurnalı;
Dilçilik araşdırmaları jurnalı;
Türkologiya jurnalı.

Hazırda yaşadığımız əsr informasiya texnologiyaları əsri, internet əsridir. İndi dünyanın ünsiyyət dili, ünsiyyət vasitəsi internetdir. Dünya internetlə danışır, tanış olur, öyrənir, öyrədir, əlaqə yaradır, mədəni-iqtisadi münasibətlər qurur, hesablaşır. İnternet bizim həyatımıza, məişətimizə, evimizə daxil olub, mövcudluq, yaşayış atributlarımızdan birinə çevrilib.

Bizdən asılı olmadan hamımız internet dünyasında yaşayırıq.

İndi dünya ilə ayaqlaşmaq, onu tanımaq, dərk etmək istəyiriksə, internetdən istifadə etməli, internetdən faydalanmalıyıq.

Özümüzü dünyaya tanımaq istəyiriksə, internetdə tanıtmalı, internetdə təmsil olunmağı, özümüzü internetdə təqdim etməyi bacarmalıyıq. Bunun özü də az əhəmiyyətli məsələ deyil.

Azərbaycan Respublikasının Prezidenti İlham Əliyevin “Azərbaycan dilinin qloballaşma şəraitində zamanın tələblərinə uyğun istifadəsinə və ölkədə dilçiliyin inkişafına dair Dövlət Proqramı haqqında” 2012-ci il 23 may tarixli sərəncamı çərçivəsində “Azərbaycan dilinin dövlət dili kimi ümumi informasiya bazasının yaradılması” (6.4.1) bəndi ilə bağlı www.azerbaycandili.az saytı istifadəçilərə təqdim olunmuşdur.

AZƏRBAYCAN DİLİNİN İNFORMASIYA BAZASI

[Ana Səhifə](#)

[Dil haqqında](#)

[Rəsmi sənədlər](#)

[Azərbaycan dilçiləri](#)

[Dilçilik kitabxanası](#)

[Əlaqə](#)

İnformasiya – sorğu bazası altı blokdan ibarətdir:

Ana səhifə;

Azərbaycan dili haqqında ümumi məlumat;

Qanunlar;

Azərbaycan dilçiləri;

Dilçilik kitabxanası;

Əlaqə.



“Ana səhifə”də göstərilən blokların adları və linklər göstərilmişdir. Linklər informasiya bazasının işini optimallaşdıran vasitələrdir. “Azərbaycan dili haqqında ümumi məlumat” blokunda Azərbaycan dilinin bütün sahələrini əhatə edən oçerk verilmişdir. Oçerkdə Azərbaycan dili haqqında ümumi məlumat, onun orfoqrafiyası, orfoepiyası, əlifbası, fonetikasi, leksikası, morfolojiyası, sintaksisi, frazeologiyası və s. barədə yığcam məlumat verilmişdir.

Respublikamızda ana dilinə dövlət səviyyəsində qayğı var və bu qayğı İnformasiya bazasının “Qanunlar” blokunda sənədlərdə əks olunub. “Qanunlar” blokunda Azərbaycan dili ilə bağlı müxtəlif illərə aid qanun, sərəncam, qərar və s. toplanmışdır.

Gələcəkdə bu blokda yeni sərəncam, fərman və digər bu qəbildən olan rəsmi sənədlərin toplanması nəzərdə tutulur.

“Azərbaycan dilçiləri” blokunda AMEA-nın həqiqi və müxbir üzvləri, filologiya üzrə fəlsəfə və elmlər doktorları

barədə üç dildə (Azərbaycan, rus və ingilis dillərində) yığcam anket məlumatı verilir.

“Dilçilik kitabxanası” blokunda dilçiliklə bağlı müxtəlif tipli lüğətlər, monoqrafiya və dərsliklər, eləcə də fəlsəfə və elmlər doktoru elmi dərəcəsi almaq üçün təqdim olunmuş dissertasiyaların avtoreferatlarının elektron variantı verilmişdir. Sonrakı illərdə bu işlərin mütəmadi olaraq davam etdirilməsi nəzərdə tutulur.

Qeyd etmək lazımdır ki, Rusiyada bu işlərlə Ali Attestasiya Komissiyası məşğul olur (www.dissercat.com). 2018-ci ilin sonlarında bu saytda 750 min elmi tədqiqat əsəri, 410 min dissertasiya, 340 min avtoreferat yerləşdirilmişdi. Hər il sayt yeni tədqiqat əsərləri və dissertasiyalar hesabına zənginləşir.

Bu həm də anti-plagiat sisteminin hazırlanması baxımından əhəmiyyətlidir.

“Əlaqə” blokunda Nəsimi adına Dilçilik İnstitutu ilə əlaqə koordinatları əks olunmuşdur. Xüsusi olaraq İnformasiya bazası ilə əlaqə saxlamaq üçün elektron poçt ünvanı da göstərilmişdir: azerbaycandili.info@box.az

Azərbaycan dilinin dövlət dili kimi ümumi informasiya bazasının gələcəkdə daha da zənginləşdirilməsi və genişləndirilməsi nəzərdə tutulur. Başqa elm sahələri də – riyaziyyat, fizika, kimya, biologiya, tibb və s. eyni qaydada (bəlkə, daha mükəmməl qaydada) internetdə təmsil olunsalar, gələcəkdə Azərbaycan elminin elektron fondu, internet fondu yaranmış olacaq. Beləliklə, dünya Azərbaycan elmini tanımaq, ona istinad etmək, ona münasibət bildirmək imkanı qazanacaqdır.

“Azərbaycan dilinin elektron məkanda daha geniş istifadəsinin təmin edilməsi ilə bağlı bir sıra tədbirlər haqqında” Azərbaycan Respublikası Prezidenti İlham Əliyevin 17 iyul 2018-ci il tarixli sərəncamı bu sahədə işlərin təşkili və canlandırılması ilə bağlı mühüm sənəddir. Gələcəkdə bu sahədə aşağıdakı işlərin aparılması faydalı olardı:

- ayrı-ayrı yazıçı və şairlərin əsərlərinin konkordanslarının hazırlanması;
- yazılı abidələrin və klassiklərin dilinin riyazi-statistik metodlarla tədqiqi məsələlərinə diqqətin artırılması;
- onlayn tədris kurslarının təşkili;
- Azərbaycan dilinin və xarici dillərin onlayn tədrisinin təşkili;
- İnternet məkanda yerləşdirilmiş elektron lüğətlərin əlaqələndirilməsi.

Azərbaycan Dilinin Milli Korpusunun leksikoqrafiya altkorpusunun yaradılması istiqamətində atılmış addımlardan biri “Azərbaycan dilinin lüğət tərkibinin inkişafı, dilin normalarına əməl olunmasının ictimai monitorinqi və dilin verilənlər korpusunun hazırlanması üzrə inteqrasiya olunmuş elektron sistemin və lüğətin yaradılması” adlı layihə sayıla bilər. Bu layihə Azərbaycan Respublikasının Prezidenti yanında Elmin İnkişafı Fondunun 2015-ci ilin əsas qrant müsabiqəsinə təqdim edilmiş və bu çərçivədə elmi-tədqiqat proqramlarının (EIF-KETPL-2015-1 (25) qalibi olmuşdur. Bir il (1 may 2017-ci il – 01 may 2018-ci il) müddətində yerinə yetirilməsi nəzərdə tutulmuş layihə Dilçilik İnstitutunun və AMEA İdarəetmə Sistemləri İnstitutunun əməkdaşları (Məsud Mahmudov, İlham Tahirov, Aytən Bəylərova, Rasim Heydərov, Kamil Aydayadə, Səxavət Talıbov, Firudin Əhmədov, Elşən Mustafayev) tərəfindən icra edilmişdir.

Layihə çərçivəsində aşağıdakı işlərin yerinə yetirilməsi nəzərdə tutulmuşdu:

- dilin normalarına əməl olunması və lüğətlərlə bağlı toplanılmış irad, təklif və tövsiyələrin təhlili, müvafiq tədbirlər görülməsi üçün aidiyyətli orqanlarla müzakirə və qərarların qəbulu;

- lüğətlərin elektron variantlarının hazırlanması, dilin lüğət tərkibinin zənginləşməsinə və təkmilləşdirilməsinə xidmət edən portalın yaradılması;
- sistemin mobil versiyasının hazırlanması.

Layihə çərçivəsində dünya təcrübəsində nəzəri və praktik leksikoqrafiyanın inkişaf istiqamətləri nəzərdən keçirilmişdir. Müasir dünya leksikoqrafik təcrübəsi göstərir ki, qloballaşan müasir dünyada leksikoqrafiyaya özünəməxsus industriya növü kimi yanaşılır. Bunu oxucunun söz haqqında ən müxtəlif informasiya ehtiyacının ödənilməsinə nəzərdə tutan lüğət industriyası hesab etmək olar. Həmin industriyanın inkişafı ilə bağlı aparılan araşdırmalar göstərir ki, müasir leksikoqrafiya əsasən iki istiqamətdə inkişaf edir. 1. Xüsusi sahələr üzrə ixtisaslaşdırılmış lüğətlərin yaradılması. Belə lüğətlər yalnız bir növ informasiyanı ehtiva edir (məsələn, yalnız sözün yazılışı, yalnız sözün mənşəyi, yaxud yalnız sözün birləşmə əmələ gətirmə qabiliyyəti və s. haqqında hazırlanan lüğətlər); 2. Universal, kompleks lüğətlərin yaradılması. Belə lüğətlər söz haqqında mümkün olan bütün informasiyanı özündə ehtiva edir, onlarda təkcə sözün mənasının izahı, qrammatik xarakteristikası, tələffüz və yazılış qaydaları deyil, həm də bir sözün digər sözlərlə mənə əlaqələri (sinonim, antonim), müxtəlif üslublarda işlənmə xüsusiyyətləri təsvir olunur. Bu tip kompleks lüğətlərdə bəzən sözün adlandırdığı əşya və ya hadisə ilə əlaqəli ensiklopedik məlumatın daxil edilməsi də müşahidə edilir.

Layihə çərçivəsində Azərbaycan dilinin orfoqrafiya və izahlı lüğətləri ilə bağlı kütləvi informasiya vasitələrində (qəzet, radio, televiziya, sosial şəbəkə və saytlarda) toplanılmış irad, nöqsan, təklif və tövsiyələr təhlil edilərək sistemləşdirilmiş və 14 bənddən ibarət təkliflər paketi şəklində Azərbaycan dilinin yeni orfoqrafiya qaydalarını hazırlayan Orfoqrafiya komissiyasına təqdim edilmişdir. Azərbaycan

ədəbi dili normalarının pozulma halları da layihə iştirakçıları tərəfindən qeydə alınmış və daha geniş dairədə müzakirə olunmaq üçün təhlil olunaraq sistemləşdirilmişdir. Layihə çərçivəsində yaradılmış inteqrasiyalı elektron lüğətlər sistemində Azərbaycan dilinin orfoqrafiya və izahlı lüğətləri və ədəbi dilin normalarının pozulması ilə bağlı müxtəlif mənbələrdən toplanılmış materialların mütəmadi olaraq müzakirə olunması, nəticə etibarilə sistemə daxil edilmiş lüğətlərin onlayn olaraq təzələnməsi nəzərdə tutulmuşdur.

Layihə son məhsulu kimi Azərbaycan dilinin lüğət tərkibinin inkişafı, dilin normalarına əməl olunmasının ictimai monitorinqi və dilin verilənlər korpusunun hazırlanması üzrə inteqrasiya olunmuş elektron sistem və onun əsasında “Azərbaycan dilinin elektron lüğətlər korpusu” portalı yaradılmış və internetdə yerləşdirilmişdir [<http://korpus.azerbaycandili.az>]. Portalın lüğətlər bölümündə ilkin olaraq aşağıdakı lüğətlərin mükəmməl elektron variantları yerləşdirilmişdir.

1. Azərbaycan dilinin orfoqrafiya lüğəti [Genişləndirilmiş və yenidən işlənmiş 6-cı nəşr. 110563 söz. Bakı, “Şərq-Qərb” Nəşriyyat evi, 2013, 839 s.];

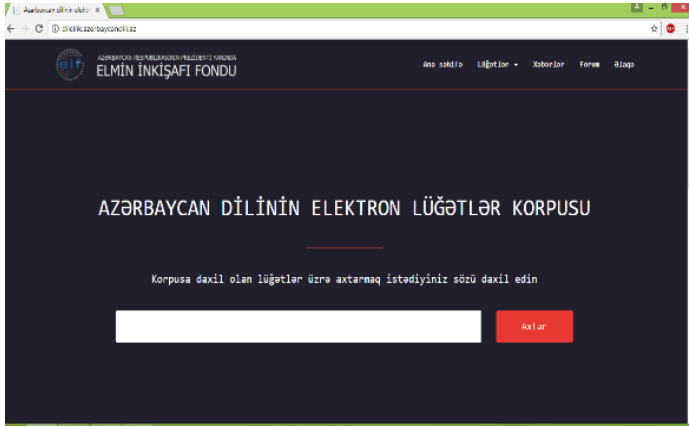
2. Azərbaycan dilinin izahlı lüğəti. Dörd cildə [I cild – 744 s., II cild – 792 s., III cild – 672 s., IV cild – 712 s. Bakı, “Şərq – Qərb”, 2006];

3. Paşayev A., Bəşirova A. Azərbaycan şəxs adlarının izahlı lüğəti [16164 ad – 8328 kişi adı, 7836 qadın adı. Bakı, Mütərcim, 2011, 340 s.] ;

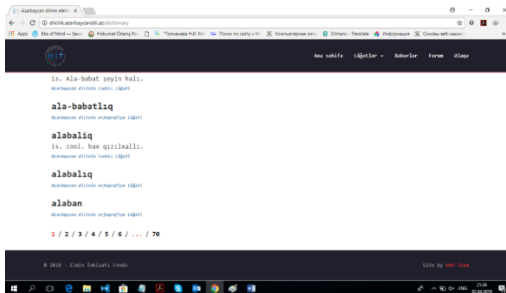
4. Azərbaycan dilinin ixtisarlar lüğəti (Azərbaycan və xarici dillərdəki ixtisarlara əhatə etməklə).

“Azərbaycan dilinin elektron lüğətlər korpusu”nun strukturunu nəzərdən keçirək.

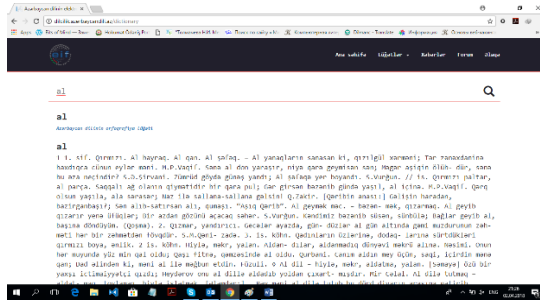
Sistemin Ana səhifəsinin ümumi görünüşü aşağıdakı kimi işlənmişdir:



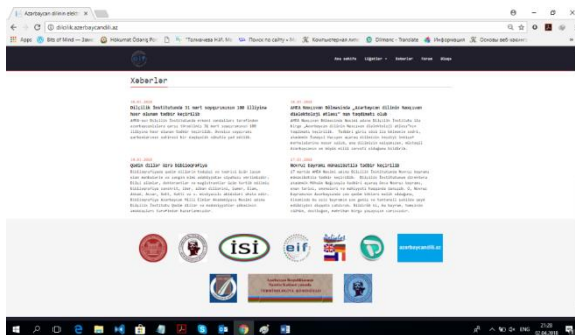
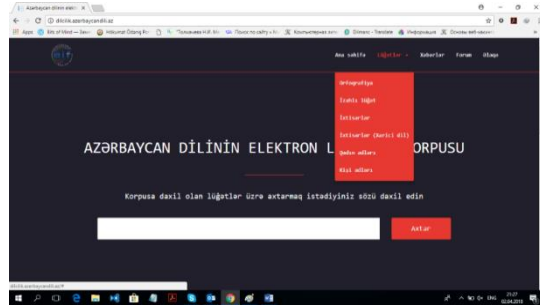
Axtarış sahəsinə istənilən frazanı daxil edib “Axtar” düyməsini kliklədikdə, elektron lüğətlər korpusunda həmin fraza ilə başlayan bütün sözlər ekranda görünəcək. Axtarış korpusa daxil edilmiş bütün lüğətlər üzrə aparılır və axtarılan ifadəyə uyğun gələn bütün sözlər müvafiq lüğətlərə istinadlarla təchiz olunur. İstinadlar müvafiq lüğətə keçməklə axtarışı seçilmiş lüğət üzrə davam etdirməyə imkan verir. Məsələn “al” frazasını axtarış sətirində daxil edib “axtar” düyməsini klikləsək, aşağıdakı nəticələri alarıq:



Ana səhifənin menyü sətirindən istifadə etməklə seçilmiş konkret lüğət üzrə axtarış etmək olar.

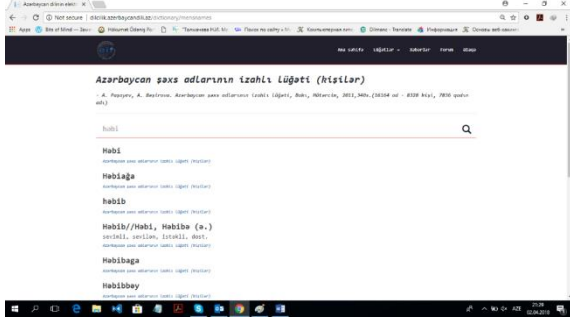


Ana səhifənin aşağı hissəsində dilçiliklə bağlı xəbərlər və partnyor təşkilatların emblemləri linklər şəklində yerləşdirilmişdir.

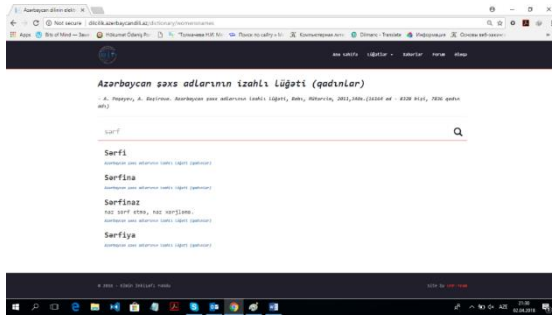


Korpusa daxil olan lüğətlər üzrə axtarış apararkən axtarış sətirinə hərflər ardıcıl daxil edildikcə, ona uyğun gələn frazaya müvafiq lüğətdəki sözlər ekranda görünür. Məsələn, kişi adları lüğətində axtarış sətirinə “Həbi” ifadəsini daxil etdikdə ekrana aşağıdakı məlumatlar çıxır:

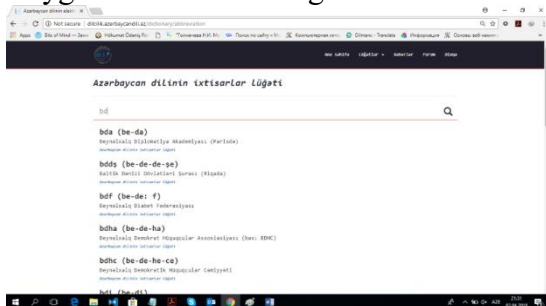
Orfoqrafiya lüğətinin elektron versiyası bütün hərflər üzrə tam şəkildə hazırlanıb başa çatdırılmışdır.



Şəxş adlar lüğəti “Kişi adları”, “Qadın adları” və “Adlar izahı ilə” kimi üç hissəyə bölünərək tam şəkildə elektron formaya keçirilmişdir. Bu lüğətdən də istifadə zamanı axtarış sətirində hərflər ardıcıl daxil edildikcə lüğətdə ona uyğun sözlər ekranda görünür.



İxtisarlar lüğəti “İxtisarlar” və “İxtisarlar (xarici dildə)” kimi iki hissəyə bölünərək tam şəkildə elektron formaya keçirilmiş və sistemə daxil edilmişdir. Bu lüğətdən də istifadə prosesində axtarış sətirində hərflər ardıcıl daxil edildikcə lüğətdə ona uyğun sözlər ekranda görünür.



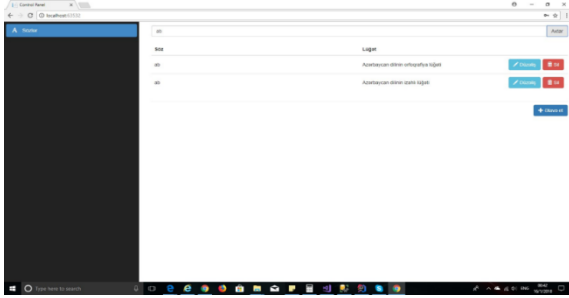
Lüğətlərdə ediləcək dəyişikliklər üçün xüsusi altsistem yaradılmışdır. Bu sistemdən ancaq administrator səlahiyyəti olan istifadəçilər yararlına bilərlər. Onlar müvafiq lüğətlərə yeni söz əlavə edə, mövcud hər hansı sözü lüğətdən silə, sözün izahı üzərində lazımı dəyişikliklər apara bilər. Bundan əlavə, moderator səlahiyyəti olan istifadəçi forumun idarə edilməsi üçün müvafiq səlahiyyətlərə malik olur.

Login

Remember me?

LOG IN

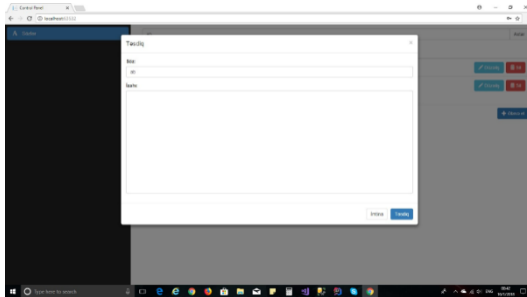
İstifadəçi səlahiyyətlərini müəyyənləşdirmək üçün xüsusi altsistem yaradılmışdır. Bu sistemə daxil olmaq üçün qeydiyyatdan keçmək lazımdır



Sistemə daxil olduqdan sonra pəncərə açılır. Açılmış pəncərədə sol tərəfdə 3 bölmə üzrə aşağıdakılar yerləşdirilmişdir:

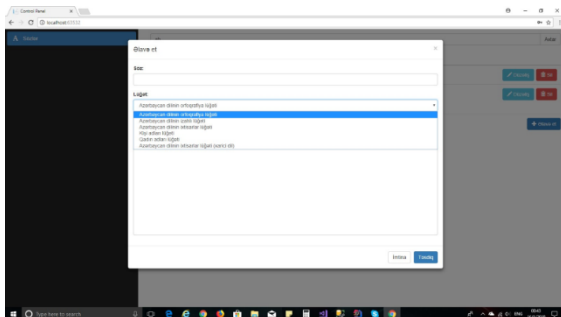
- sözlər;
- xəbərlər;
- rəy və təkliflər.

Bu bölmələrin hər hansı biri üzrə müvafiq düzəlişlər aparmaq olar.



Açılmış pəncərənin axtarış sətirinə daxil edilmiş sözlər korpusun lüğətləri üzrə axtarılır və nəticələr aşağıdakı kimi ekranda əks olunur (Məsələn, “ab” frazasını daxil etsək):

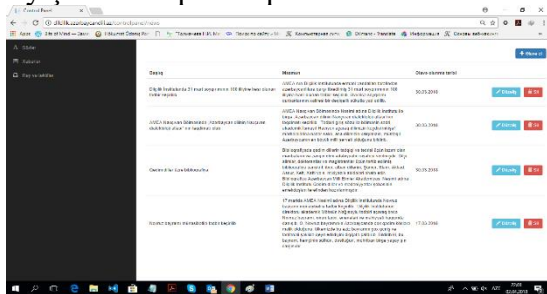
Bundan sonra seçilmiş lüğət üzrə müvafiq dəyişikliklər edilir.



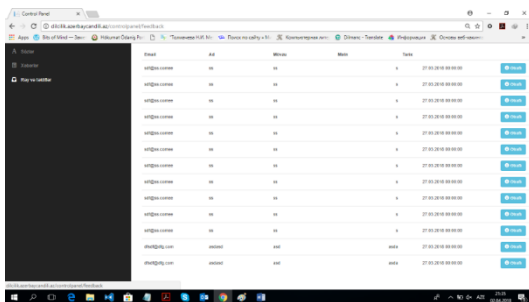
Bu dəyişikliklər konkret söz və onun haqqındakı əlavə qeyd üzərində aparıla bilər.

Hər hansı bir yeni sözü müəyyən bir lüğətə əlavə etmək və ya silmək üçün alt sistem yaradılmış və onun strukturu aşağıdakı kimi müəyyənləşdirilmişdir:

İdarəedici panel vasitəsi ilə “Xəbərlər” menyusundan istifadə etməklə yeni xəbərlər əlavə etmək və onlar üzərində müəyyən dəyişikliklər aparmaq mümkündür.



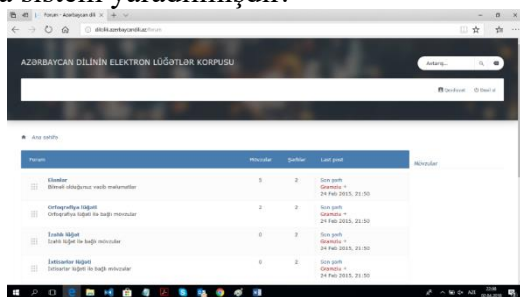
“Rəy və təkliflər” bölməsinə daxil olmaqla sayta dilçiliklə bağlı göndərilmiş elektron müraciətlərə baxmaq olar.



Dilin normalarına əməl olunması və lüğətlərlə bağlı toplanılmış irad, təklif, və tövsiyələrin təhlili və müvafiq tədbirlər görülməsi üçün aidiyyətli orqanlarla müzakirəsi, qərarların qəbulu ilə bağlı alt sistemin ümumi strukturu və iş prinsipləri də müzakirə edilmiş və müvafiq struktur və qaydalar qəbul olunmuşdur. Ayrı-ayrı lüğətlərdə edilən hər hansı bir dəyişiklik yalnız saytın moderatoru tərəfindən həyata keçirilir. Qərarlar xüsusi yaradılmış komissiya tərəfindən toplanılmış irad, təklif və tövsiyələrin təhlili nəzərə alınmaqla kollektiv qərar əsasında qəbul edilir.

Elektron lüğətlər korpusunun yaradılması prosesində Azərbaycan dilinin izahlı lüğəti (4-cildlik) elektron formata keçirilmiş, redaktə edilmişdir.

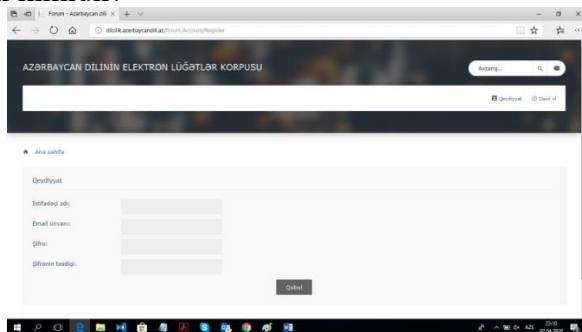
Sistemin mobil versiyada da işləməsi üçün kros platforma əsasında sistem yaradılmışdır.



Dilin normalarına əməl olunması və lüğətlərlə bağlı toplanılmış irad, təklif, və tövsiyələrin təhlili və müvafiq tədbirlər görülməsi üçün aidiyyətli orqanlarla müzakirəsi, qərarların qəbulu ilə bağlı portalın strukturu müəyyənləşmiş və onun proqram təminatı işlənib hazırlanmışdır.

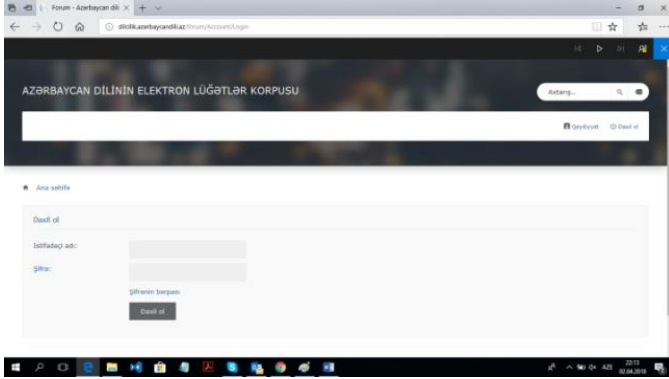
Forumda hər hansı mövzu yaratmaq üçün ilk öncə qeydiyyatdan keçmək lazımdır. Qeydiyyat proseduru aşağıdakı alqoritm əsasında aparılır. İstifadəçi yeni qeydiyyat formasını doldurduqdan və lisenziya şərtləri ilə tanış olduqdan sonra onun üçün müvəqqəti passiv hesab açılır və onun elektron ünvanına hesabı aktivləşdirmək üçün link göndərilir. İstifadəçi hesabını aktivləşdirdikdən sonra forumdan istifadə etməklə yeni mövzular açıb müzakirələrdə iştirak edə bilər. Mövzu ətrafındakı qeydlər moderator tərəfindən yoxlanarkən lisenziya şərtlərinə əməl olunmayanlar sistemdən çıxarılır, həmən istifadəçi haqqında müvafiq qeyd aparılır. Belə qeydlərin sayı 2 dəfədən çox olduqda, həmin istifadəçi sistem tərəfindən bloklanır.

Forumda qeydiyyatdan keçmək üçün səhifənin strukturu aşağıdakı kimidir:

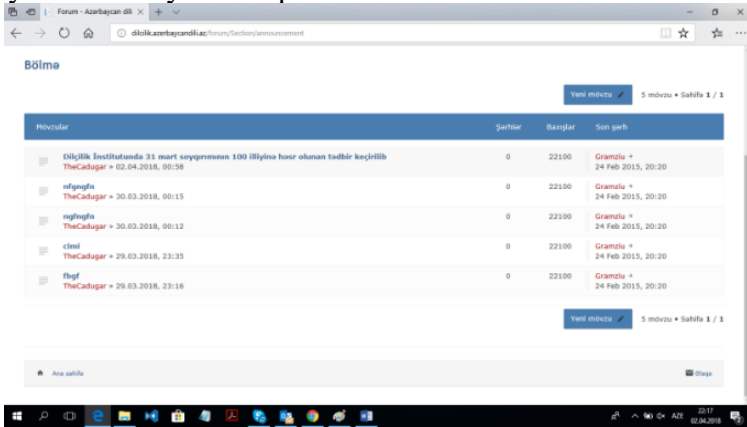


Forma doldurulduqdan sonra “qəbul” düyməsini basmaqla qeydiyyatdan keçmək olar.

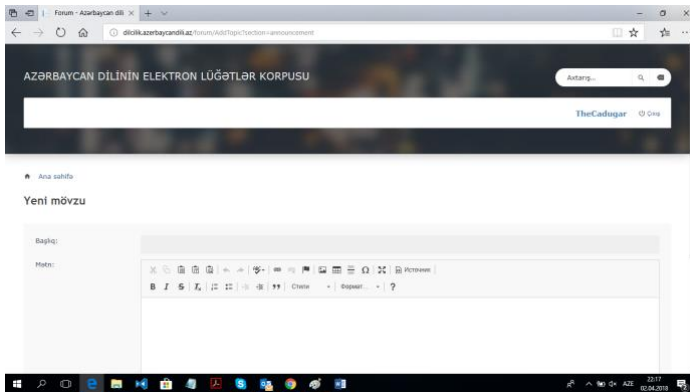
Yeni mövzu yaratmaq üçün əvvəlcə sistemə daxil olmaq lazımdır.



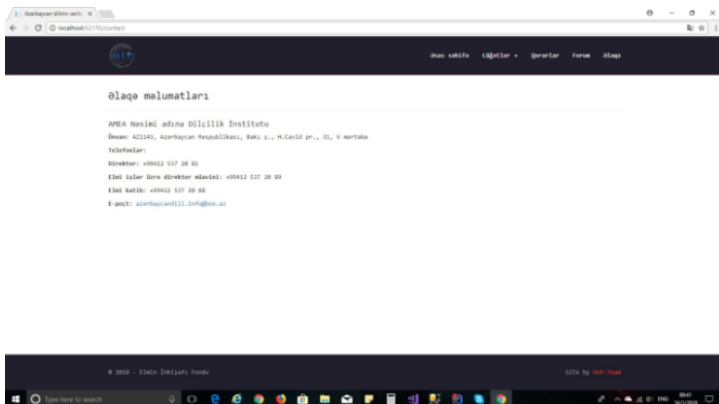
Bundan sonra seçilmiş mövzu üzərində düzəliş etmək və ya yeni mövzu yaratmaq olar.



Forumda yeni yaradılan mövzu üçün pəncərə aşağıdakı kimidir:



Əlaqə menyusunun pəncərəsi aşağıdakı strukturda görünür:



Layihənin son məhsulu olaraq ortaya çıxan inteqrasiya olunmuş elektron lüğətlər sistemi Azərbaycan dilinin milli korpusunun leksikoqrafiya altkorpusunun yaradılması və təkmilləşdirilməsi yönündə ciddi addımlardan sayıla bilər. Azərbaycan dilinin milli korpusu zəngin elmi təcrübə, yüksək ixtisas hazırlığı, böyük səriştə və zəhmət tələb edən mükəmməl məlumat axtarışı sistemi kimi səciyyələndirilə bilər [Mahmudov, Tahirov, Aydazadə, Talıbov, 2019, s.66-80; [Mahmudov, 2018, s.140-150] Bu iş Azərbaycan Respublikasının Prezidenti yanında Elmin İnkişafı fondunun maliyyə yardımı ilə yerinə yetirilmişdir [Qrant EIF-KETPL-2-2015-1 (25)-56/ 53/5].

V FƏSİL

MƏTNİN FORMAL LİŊVİSTİK TƏHLİLİ SİSTEMLƏRİ SÜNİ İNTELLEKTİN TƏDQIQAT OBYEKTİ KİMİ

Mətnin formal təhlil sistemləri süni intellekt problemləri içərisində ən mühüm və vacib komponentlərdən biri hesab olunur. Maşın tərcüməsi, məlumat axtarışı sistemləri, dil korpuslarının hazırlanması formal linqvistik təhlil olmadan qeyri-mümkündür. Həm də bu təhlil dilin bütün səviyyələrini əhatə etməlidir. Avtomatik (formal) fonetik təhlil nitqin sintezi, təhlili, nitqlə idarə olunan qurğu və sorğu sistemlərində geniş istifadə olunur. Məlumdur ki, görmə qabiliyyəti zəif olan insanlar üçün kompüterlərin hazırlanması işlərində də formal fonetik təhlilsiz keçinmək mümkün deyil. Milli dil korpuslarının da mühüm bölmələrindən biri şifahi nitq korpuslarıdır. Şifahi nitq korpuslarında nitqin sintezi və formal təhlilinin işlənmə dərəcəsi mühüm amil hesab olunur və korpusun əhəmiyyəti və istifadə olunma dərəcəsi həmin amillə sıx bağlıdır. Bu məsələ çərçivəsində tələffüzün, vurğunun, səsin intensivliyinin, tonunun, güclü və zəif mövqələrinin araşdırılması və təhlili prioritet vəzifələrdən hesab olunur.

Dilin fonetik, morfoloji, sintaktik və semantik səviyyələrdə təhlili də təbii mətnin emalı, maşın tərcüməsi, məlumat-sorğu axtarış sistemlərində, dil korpuslarının hazırlanması işlərində mərhələli qaydada istifadə olunur.

Müasir anlamda korpus hər hansı konkret dilin mətnlərinin əhatə edən kompüter bazasıdır və onun yaradılmasında xüsusi proqramlardan istifadə olunması təbiidir. Məlum olduğu kimi, bu proqramlar içərisində avtomatik işarələnmə proqramları xüsusi yer tutur. Müasir korpusların həcmi

nəzərə alsaq, onların işarələnməsi olduqca ağır və çox zəhmət tələb edən bir prosesdir. Hazırda formal morfoloji və sintaktik təhlil üçün müxtəlif proqram vasitələri mövcuddur. Həmin proqramlar – morfoloji işarələnmədə teqqlər, sintaktik səviyyədə isə parserlər adlanır (taggers və parsers). Avtomatik morfoloji təhlil proqramının işi nəticəsində hər bir leksik vahid üçün onun qrammatik səciyyələri yazılır. Qrammatik səciyyələrə aşağıdakılar daxildir:

- nitq hissələrinə mənsubluq;
- lemma (normal forması);
- qrammlər toplusu (cins, kəmiyyət, hal, canlı/cansız, təsirli/təsirsiz və s.).

Avtomatik sintaktik təhlil proqramlarının işi zamanı sözlər və söz birləşmələri arasında sintaktik əlaqələr qeydə alınır. Sintaktik kateqoriyalara müvafiq səciyyələr əlavə olunur (cümlə tipi, söz birləşmələrinin sintaktik funksiyaları və s.).

Semantik təhlil mərhələsi daha mürəkkəb və çoxşaxəlidir. Burada söz, cümlə, mətn səviyyəsində semantik mənə məsələləri hərtərəfli araşdırılır.

Azərbaycan dilində fonemlərin fono-fizioloji-akustik əlamətləri. Azərbaycan dili fonemlərinin fonetik, fizioloji, akustik xüsusiyyətlərinin, orfoepik cəhətlərinin, şifahi danışq dilində işlənmə özəlliklərinin tədqiqinin təbii mətnin emalı sistemləri üçün əhəmiyyəti böyükdür. NLP sistemlərində nitqin tanınması və yazıya (mətnə) çevrilməsi, nitqin sintezi və yazının (mətnin) nitqə çevrilməsi prosesində fonemlərin bütün funksional aspektlərinin araşdırılması mühüm mərhələ hesab olunur. Bu məsələ ilə bağlı daha geniş məlumata “Türk dillərinin milli korpusları” monoqrafiyasında rast gəlmək olar [Mahmudov, 2018, s.152-170].

Saitlərin fono-fizioloji-akustik əlamətləri. Azərbaycan dilində saitlərin səciyyəvi fono-fizioloji-akustik əlamətləri aşağıdakılardır: dilarxası-dilönü, qarışıq (neytral), dodaqlanan-

dodaqlanmayan, açıq-qapalı, yarımçıq (yarımqapalı), təkavazlı-cütavazlı (diftonq), qısa-uzun, qalın-incə [Axundov, 1973, s.6-169; Müasir Azərbaycan dili, 1978, s.16-21].

1. A (a) saiti – dilarxası, açıq, dodaqlanmayan, təkavazlı, qısa, qalın, uca.

Sözün əvvəlində, ortasında və sonunda sərbəst işlənir:

A (a) saiti şifahi nitq prosesində bəzi dəyişmələrə məruz qalır. Onların izahı və işarələnmədə nəzərə alınması da zəruridir:

- `y` səsindən sonra a-`ı` əvəzlənməsi baş verir: anaya – anıya, almaya – almiya;

- sözün sonunda gələn `y` səsindən sonra `a` bəzən düşür: qayaya – qayya, tayaya – tayya;

- -aydı, -saydı şəkilçilərində `a`-`e`-yə keçir: alaydı – aleydi, qalaydı – qaleydi, alsaydı – alseydi, qalsaydı – qalseydi;

- saitlə bitən və saitlə başlanan sözlərdə birinci sözün sonundakı `a` tələffüz olunmur – Ağəmi (Ağəmi), Balaəmi (Baləmi), xalaoğlu (xaloğlu) [Müasir Azərbaycan dili, 1978, s.85-86].

2. A (a): – (aa) – dilarxası, açıq, dodaqlanmayan, təkavazlı, uzun, qalın, uca.

İşlənmə məqamları:

- əvvəldə – ərəb və fars mənşəli sözlərdə: aaşiq, aahu, aalim, aaləm, aarif, aadi, aasan, aaləm;

- ortada – ərəb və fars mənşəli sözlərdə: Saabir, saadə, naadir, maahir, xaadim, çaarə, saakin.

- A (a): – (aa) saitinin işləndiyi alınma sözlərin siyahısı ayrıca verilir.

3. Ə (ə) saiti – dilönü, açıq, dodaqlanmayan, təkavazlı, qısa, incə, uca.

Sözün əvvəlində, ortasında və sonunda sərbəst işlənir.

Ə (ə) səsinin nitq prosesində müşahidə olunan başqa xüsusiyyətləri:

- `y` səmindən əvvəl gələn `ə` darlaşaraq `i`-yə keçir – dinləyər (dinliyər), gəlməyəcək (gəlmiyəcək), işləyər (işliyər), dişləyər (dişliyər)

- saitlə bitən və saitlə başlanan sözlərdə birinci sözün sonundakı `ə` tələffüz edilmir – Mirzəağa (Mirzağa), Mirzəəmi (Mirzəmi) [Müasir Azərbaycan dili, 1978, s.86].

4. Ə (ə): – (əə) – dilönü, açıq, dodaqlanmayan, təkavazlı, uzun, incə, uca.

İşlənmə məqamları:

- bəzi ərəb, fars mənşəli sözlərdə – təəyin, təlim, bəəzi, təərif, məəlum, bəəzən, təəsir.

Məlum olduğu kimi, Azərbaycan dilində apostroflu sözlərdə səslər uzun tələffüz olunurdu. Apostrof işarəsi həm də səsin uzun tələffüzünü bildirirdi. Latın qrafikalı yeni əlifbaya keçiddən sonra apostrof işarəsi əlifbadan çıxarıldı. Azərbaycan dilinin 2013-cü ildə çapdan çıxmış orfoqrafiya qaydalarında da apostrofla bağlı bölmə yoxdur və lüğətin tərtibi zamanı apostrofla yazılan sözlər apostrofsuz verilmişdir [Azərbaycan dilinin orfoqrafiya lüğəti, 2013].

Lakin apostroflu sözlərdə səslərin uzanması halını nəzərə alaraq həmin sözlərin siyahısı alqoritmədə nəzərə alınmışdır.

5. E (e) saiti – dilönü, yarımaçığ, dodaqlanmayan, təkavazlı, qısa, incə, orta.

Tələffüzdə əsasən sözün birinci hecasında rast gəlinir – eşit, yeşik, sel, sev, de, ye, ver.

Alınma sözlərdə iki və çoxhecalı sözlərin sonunda da müşahidə olunur:

- kafe, mövqe, vaqe, jele, rele, qane, mane, kaşne, koyne, püre, klişe, attaşə, karate, varyete.

6. E (e): – (ee) – dilönü, yarımaçığ, dodaqlanmayan, təkavazlı, uzun, incə, orta.

İşlənmiş məqamları:

- `ə` səsi ilə bitən və `e` səsi ilə başlanan sözlərin bitişik variantının tələffüzü zamanı `ə` səsi deyilmir və `e` uzun tələffüz olunur: nə edim – needim, nə etsin – neetsin.

- Bəzən ikinci sözün əvvəlində `y` səsi deyildiyindən `e` adi qaydada tələffüz olunur: nə eləyim – neyliyim.

- Ərəb və fars mənşəli sözlərdə `e` səsi bəzən uzun tələffüz edilir: eelan, eetiraz, eetibar, eecazkar, eehtiva, eehtiyac.

7. I (i) saiti – dılarxası, qapalı, dodaqlanmayan, təkavazlı, qısa, qalın, aşağı.

Sözün əvvəlində işlənmiş, ancaq sözün ortasında və axırında müşahidə olunur.

Vurğunun son heca üzərinə düşməsi nəticəsində sözün birinci hecasında `i` qısa tələffüz olunur: q (i)fil, q (i)rmızı, q (i)jiltı.

`Qijiltı` sözünün birinci hecasında `i` çox qısa, sonuncu hecada isə nisbətən uzun tələffüz olunur.

Sözün sonunda `y` səsindən sonra gələn `i` səsi yönlük hal şəkilçisi qəbul etdikdə iki `y` səsi arasında çox zəif tələffüz olunur: ayıya – ay (i)ya, dayıya – day (i)ya.

Saitlə bitən və saitle başlanan sözlərdən əmələ gəlmiş mürəkkəb sözlərdə birinci sözdə vurğu itdiyi üçün onun axırında olan `i` səsi deyilmir: dayıoğlu – dayoğlu, bacıoğlu – bacioğlu [Müasir Azərbaycan dili, 1978, s.86-87].

İ səsinin uzun variantı yoxdur. Bu səs qısalmaya, zəif tələffüzə meyillidir. Alınma sözlərdə nadir hallarda sözün əvvəlində də işlənə bilər.

8. İ (i) saiti – dilönu, qapalı, dodaqlanmayan, təkavazlı, qısa, incə, orta.

Sözün əvvəlində, ortasında və axırında müşahidə olunur:

Çoxhecalı sözlərin birinci hecasında `i` qısa tələffüz olunur: k (i)çiy, p (i)şiy, k (i)şi.

Bəzi sözlərə saitlə başlanan mənsubiyyət və hal şəkilçiləri (yiyəlik, yönlük, təsirlik) qoşulduqda vurğunun şəkilçi üzərinə keçməsi nəticəsində sözün axırıncı hecasındakı qapalı saitin düşməsi hadisəsi baş verir və `i` səsi tələffüz edilmir: şəkil – şəklə, sinif – sinfə, fikir – fikrə, nəsil – nəslə. Mənsubiyyət və hal şəkilçiləri (yiyəlik, yönlük, təsirlik) qəbul edərkən sözün ikinci hecasındakı qapalı saitin düşməsi hadisəsinə məruz qalan sözlər siyahı şəklində verilir [Mahmudov, 2013, s.242].

Saitlə bitən və saitlə başlanan sözlərdən əmələ gəlmiş mü-rəkkəb sözlərdə birinci sözün vurğusu ikinci sözün üzərinə keçdiyindən əvvəlinci sözün axırındakı `i` səsi tələffüz edilmir: Əliəkbər – Ələkbər, əmioğlu – əmoğlu, bibioğlu – biboğlu, iyirmiiki – iyirmiki [Müasir Azərbaycan dili, 1978, s.87].

9. İ (i): – (ii) – dilönü, qapalı, dodaqlanmayan, təkavazlı, uzun, incə, orta.

Bəzi ərəb mənşəli sözlərdə (ii) uzun tələffüz olunur: vəsiqə, viiran, Mədiinə, vəsiilə.

10. O (o) saiti – dilarxası, açıq, dodaqlanan, təkavazlı, qısa, qalın, orta.

Çox vaxt sözün birinci hecasında müşahidə olunur.

əvvəldə – otaq, oruc, ora, oraq, ozan, ot, od, ol, odun, on, otuz;

ortada – qohum, qoyun, boyun, toy, qol, bol, yol, dolu, sol, tor, şor;

axırda – kilo, kino, metro, pero, foto, palto, büro, piano, kazino;

11. O (o): – (oo) dilarxası, açıq, dodaqlanan, təkavazlı, uzun, qalın, orta.

Sözün ortasında və axırında `v` səsinin düşməsi nəticəsində `o` uzun tələffüz olunur: cilov – ciloo, buzov – buzoo.

12. OU diftonqu – sözün ortasında və sonunda `v` samitinin düşməsi nəticəsində yaranır: *douşan, douğa, youşan, çoğun.*

13. Ö (ö) saiti – dilönü açıq, dodaqlanan, təkavazlı, qısa, incə, orta.

Bir neçə söz istisna olmaqla ancaq sözün birinci hecasında təsadüf edilir:

əvvəldə – öz, öl (mək), öt (mək),öp (mək), ömür, ötür (mək), öhdəlik, ölç (mək), ön;

ortada – göz, söz, gör (mək), göl, göy, kömək, kətək, köpək, köpük, kötük, körpü.

Söz sonunda `ö` işlənmir. İstisna kimi bir neçə alınma və səstəqlidi söz göstərmək olar: *müsyö, tədqiq-tətəbbö, öhö, lölö.*

14. Ö (ö): – (öö) – dilönü, açıq, dodaqlanan, təkavazlı, uzun, incə, orta.

Sözün ortasında və axırında `v` səsinin düşməsi nəticəsində `ö` uzun tələffüz olunur – *kösöv – kösöö, bülöv – bülöö, dövran – dööran, bənövşə – bənööşə.*

15. ÖÜ diftonqu sözün ortasında və sonunda `v` samitinin düşməsi nəticəsində yaranır: *növbə – nöübə, kövşən – köüşən, tövlə – töülə.*

16. U (u) saiti – dilarxası, qapalı, dodaqlanan, təkavazlı, qısa, qalın, orta.

Sözün əvvəlində, ortasında və axırında işlənilir:

əvvəldə – uzun, uduş, ulu, uzaq, uşaq, utan (maq) və s.

ortada – quru, odun, bulud, duru, duman, duz, dur (maq) və s.

axırda – dolu, boru, su, bu, sorğu, durğu, burğu, turşu, arzu və s.

Sözün vurğusuz birinci hecasında `u` qısa tələffüz olunur: *budağ, bulaq, buzov, burax.*

Bəzi sonor səslərlə bitən ikihecalı sözlərə saitlə başlayan şəkilçilər artırıldıqda sözün axıncı hecasındakı `u` səsi vurğunun şəkilçi üzərinə keçməsi nəticəsində tələffüz edilmir: burun – burnu, oğul – oğlu. Bu xüsusiyyət alqoritmədə nəzərə alınmışdır.

17. U (u): – (uu) saiti – dilarxası, qapalı, dodaqlanan, təkavazlı, uzun, qalın, orta.

Bəzi ərəb mənşəli sözlərdə söz ortasında `u` səsi uzun tələffüz olunur: xüsusi – xüsuusi, Füzuli – Füzüuli, tufan – tuufan, Quran- Quuran.

18. Ü (ü) saiti – dilönü, qapalı, dodaqlanan, təkavazlı, qısa, incə, orta.

Sözün əvvəlində, ortasında və sonunda işlənir:

əvvəldə – üzüm, ürək, üfük, üz (mæk), ümid, ülgü, ün, ülvi, ürkək və s.

ortada – bütün, bürü (mæk), büt, bülbül, sünbül, gül, sür (mæk) və s.

axırda – ütü, sürü, düyü, cücü, töycü, sözçü, ölçü və s.

Sözün vurğusuz birinci hecasında `ü` səsi qısa tələffüz olunur: bütün, bürün, düyün, tüstü, tufəng, düşmən.

Sonor saitlərlə bitən sözlərə saitlə başlanan şəkilçi qoşulduqda vurğunun şəkilçi üzərinə keçməsi nəticəsində sözün axıncı hecasındakı `ü` deyilmir: ömür – ömrün, üfük – üfqı.

Samitlərin fono-fizioloji-akustik əlamətləri. Azərbaycan dilində samitlərin səciyyəvi fono-akustik, anatomik-fizioloji əlamətləri: dodaq samitləri, dilönü samitlər, dilarxası samitlər, qoşadodaq, dodaq-diş, dil-diş, dil-damaq samitləri, kar, cingiltili, burun samitləri, sonorlar, partlayanlar, titrəyənlər, kipləşənlər, affrikatlar, sürtünənlər və s. [Axundov, 1973, s.169-263; Müasir Azərbaycan dili, 1978, s.21-33].

19. H (h) samiti boğazda sürtünən, kar samitdir. Boğaz və ya farınqal da adlanır. Daha çox alınma sözlərdə sözün əvvəlində, ortasında və axırında müşahidə olunur:

əvvəldə – hesab, hörük, hörümçək, hürmək, hədə, hələ, həmişə və s.

ortada – dərhal, səhra, qəhər, səhər, şəhər, zəhər, dahi və s.

axırda – şah, gah, fərəh, tamah, gümrah, dəzgah, xeyirxah və s.

20. Q (q) samiti – dilarxası, partlayan, cingiltili.

Sözün əvvəlində, ortasında və axırında rast gəlinir:

əvvəldə – qarğa, qoz, qəzet, qədər, qəzəl, qüdrət, qala, qatar və s.

ortada – dəqiq, tənqid, intiqam, yəqin, əqidə, həqiqət və s.

axırda – arıq, çarıq, sadiq, qayıq, sayıq, layiq, qatıq, artıq və s.

Söz ortasında işlənən qoşa `q` samitindən birincisi kar tələffüz olunur: saqqal – sakqal, baqqal – bakqal, çaqqal – çakqal, doqquz – dokquz.

21. X (x) samiti dilarxası, kar samitdir. Sözün əvvəlində, ortasında və əsasən təkhecalı sözlərin sonunda rast gəlinir:

əvvəldə – xatirə, xəyal, xərc, xülasə, xələl, xaric, xalça, xalis və s.

ortada – arxa, daxma, çaxnaşma, axtarış, saxta, taxta, buxta, boxça və s.

axırda – çox, arx, bax, yax, tox, şux, çarx, şax, mıx, lax, ax, tax və s.

22. Ğ (ğ) samiti dilarxası, sürtünən, cingiltili.

Sözün əvvəlində işlənmir, sözün ortasında və sonunda müşahidə olunur:

ortada – sağlam, ağrı, sarğı, çağrı, çağırış, ağıllı, ağırtı, ağıuş və s.

axırda – sağ, ağ, yağ, bağ, biğ, lağ, çağ, dağ, tağ və s.

23. K* (k*) dilortası, partlayan, kar samitdir. Sözün əvvəlində, ortasında və sonunda rast gəlinir:

əvvəldə – kələm, kal, kitab, kağız, kürə, kisə, katib, kamil, kol və s.

ortada – tülkü, tüklü, tikinti, tikə, tikmə, təklif, təkrar, təkzib və s.

axırda – tək, tük, xərək, çörək, ürək, dirək, lək, çək, lük, şerik və s.

Çoxhecalı sözlərin sonunda k* çox vaxt `g` və ya `y` kimi tələffüz olunur: inə (y), çiçə (y), ürə (y), xərə (y) və s.

Söz sonunda k*-dan sonra saitlə başlanan şəkilçi gəldikdə, iki sait arasında k* samiti `y` kimi tələffüz olunur: çörək – çörəyi, çörəyə, çiçək – çiçəyi, çiçəyə, gəlmək – gəlməyi, gəlməyə. Bu xüsusiyyət Q samitinə də aiddir (q-ğ əvəzlənməsi): budaq – budağı, qulaq – qulağa, udlaq – udlağın və s.

Qeyd etmək lazımdır ki, saitlə başlanan şəkilçi qəbul etdikdə k-y və q-ğ əvəzlənməsinə məruz qalmayan istisna sözlərin siyahısı ayrıca verilmişdir [Mahmudov, 2013, s.242-246].

Əvəzlənməyə məruz qalan sözlər isə korpusda nəzərə alınmalıdır.

24. K** (k**) dilarxası, partlayan, kar samitdir. Bu samitə rus dilindən və rus dili vasitəsi ilə alınan sözlərdə sözün əvvəlində, sözün ortasında və sözün axıncı hecasında rast gəlmək mümkündür:

əvvəldə – kartof, kabinet, kadr, kolxoz, kombayn, kulinariya, kalor və s.

ortada – traktor, xarakter, akademiya, akasiya, makaron, makara və s.

son hecada – texnika, mexanika, leksika, dinamika, hidravlika və s.

Ərəb mənşəli sözlərdə bəzən sözün sonundakı `q` samiti `k**` kimi tələffüz olunur:

aşi (k**), ittifa (k**), şəfə (k**), üfə (k**), mara (k**), məxlu (k**), ha (k**), layi (k**).

25. G (g) samiti dilortası, partlayan, cingiltili samitdir. Bu samit sözün əvvəlində, ortasında və sonunda işlənir:

əvvəldə – gəmi, gəlir, gəlin, gümüş, gülüş, giləs, gavalı və s.

ortada – əgər, cərgə, əsgər, sərgi, hörgü, mürgü, sürgün, gərgin və s.

axırda – zəng, səhəng, pələng, əhəng, ahəng, ləng, çələng, fişəng və s.

Nümunələrdən görüldüyü kimi söz sonunda `g` samiti `n` səsindən sonra gəlir və bir qədər kar tələffüz olunur.

26. Y (y) – dilortası, sürtünən, cingiltili tələffüz olunur. Bu samitə sözün əvvəlində, ortasında və axırında rast gəlmək mümkündür:

əvvəldə – yay, yaz, yer, yat, ye, yüz, yetmiş, yemiş, yürüş, yüksək və s.

ortada – düymə, qiymə, əyinti, yeyinti, səyyah, dəyər, meyvə, eyvan və s.

axırda – yay, ay, say, çay, tay, may, lay, şey, key, dəy, səy, iy, zəy və s.

Sonu k* ilə bitən sözlər danışıda `y` kimi tələffüz edilir: çörəy, ürəy, dirəy, çiçəy, gərəy.

27. T (t) – dilönü, partlayan, kar samitdir. Bu samitə sözün əvvəlində, ortasında və axırında təsadüf edilir:

əvvəldə – top, tərənə, təmiz, tərif, təklif, təmir, tarix, təqvim, təşkil və s.

ortada – ətir, ətraf, atəş, qatar, satış, atışma, qatıq, satlıq, artıq və s.

axırda – at, ət, ot, sürət, cürət, hərəkət, cəsarət, hərərət, həqarət və s.

Sözün sonunda bəzən `t` samiti düşə bilər: dostlar – doslar, rast gəldim – ras gəldim, sərbəst danışır – sərbəs danışır.

28. D (d) – dilönü, partlayan, cingiltili samitdir. Bu samitə sözün əvvəlində, ortasında və sonunda rast gəlinir:

əvvəldə – daş, diş, divar, dava, diqqət, dəvət, dəhşət, demokratiya və s.

ortada – adaş, addım, qədəm, adil, adi, ada, sadə, ədəb, adalət, əda və s.

axırda – dad, yad, ad, şad, murad, sənəd, vəd, qədd, rədd, sədd və s.

Sözün sonunda bəzən `d` karlaşır və `t` kimi tələffüz olunur: kənddən – kən (t)dən, qəndsiz – qən (t)siz, buludlar – bulu (t)lar.

İrəli assimilyasiya nəticəsində şəkilçinin birinci səsi olan `d` səsi `n`-ə çevrilir: məndən – mənnən, səndən – sənnən, dilindən – dilinnən, əlimdən – əlimnən.

29. S (s) – dilönü, sürtünən, kar samitdir. Sözün əvvəlində, ortasında və sonunda təsadüf edilir:

əvvəldə – sərin, səhər, səhifə, səthi, süd, sərgi, satış, səmimi, səhəng və s.

ortada – həsəd, həsir, həsrət, məsləhət, miskin, məskən, məsrəf, kəsir və s.

axırda – yas, tas, iclas, ixtisas, mütəxəssis, xəsis, məclis, təcnis, həris və s.

30. Z (z) – dilönü, sürtünən, cingiltili samitdir. Bu samit sözün əvvəlində, ortasında və axırında işlənir:

əvvəldə – zəli, zərif, zəlil, zəhmət, zəhlə, zümrə, zürafə, ziyafət və s.

ortada – azad, azar, əzəl, nəzər, məzar, mizan, nazir, hazır, xəzəl və s.

axırda – az, yaz, saz, qaz, daz, naz, namaz, taraz, nasaz, uzundraz və s.

31. Ş (ş) – dilönü, sürtünən, kar samitdir. Sözün əvvəlində, ortasında və axırında işlənir:

əvvəldə – şar, şor, şit, şən, şad, şəlalə, şücaət, şimal, şərqi, şərik, şam və s.

ortada – məşəl, məşhur, məşəqqət, mişar, dəhşət, vəhşi, məşriq, nəşə və s.

axırda – daş, aş, qaş, barış, qaçış, bardaş, sirdaş, yoldaş, qardaş, həmyaş və s.

32. J (j) – dilönü, sürtünən, cingiltili samitdir. Bu samitə bir sıra alınma sözlərin əvvəlində, ortasında və axırında rast gəlmək mümkündür:

əvvəldə – jalə, jaket, jurnal, jüri, jele, jilet, jandarma, jest, jalüz, jarqon və s.

ortada – əjdaha, Əjdər, ajan, ajiotac və s.

axırda – masaj, pasaj, mesaj, qaraj, trikotaj, reportaj, dublyaj, kortej və s.

33. Ç (ç) – dilönü, partlayan, kar samitdir. Bu samitə sözün əvvəlində, ortasında və təkhecalı sözlərin axırında rast gəlinir:

əvvəldə – çay, çadır, çadra, çiçək, çəmən, çörək, çiskin, çəllək, çöp və s.

ortada – açar, qaçış, açılış, hərçənd, saçaq, parça, alça, dolça, boğça və s.

axırda – aç, qaç, saç, uç, üç, iç, keç, seç, ölç, xaç, qıç, maç, heç, peç və s.

34. C (c) – dilönü, partlayan və cingiltili samitdir. Sözün əvvəlində, ortasında və sonunda təsadüf edilir:

əvvəldə – cavan, can, calaq, cücə, cücərti, cecə, cib, corab, cavab və s.

ortada – yonca, ancaq, qoca, uca, əcəl, əlcək, hicab, şəcərə, pəncərə və s.

axırda – ac, həcc, bac, sac, uc, güc, ök, tac, qaxac, ağac, kirəc, borc və s.

Çoxhecalı sözlərin sonunda `c` səsi bir qədər kar tələffüz olunur: ağa (ç), yalava (ç), çəki (ç), kərpi (ç) və s.

Dissimilyasiya nəticəsində sözün axırında gələn `c` səsi `d` səsi qarşısında `j` ilə əvəz olunur: ağa (j)dar, bi (j)diy, gü (j)dü, kərpi (j)dən və s.

35. N (n) – dilönü, burun, sonor, cingiltili samitdir. Bu samitə söz əvvəlində, söz ortasında və söz axırında rast gəlmək mümkündür:

əvvəldə – nənə, nanə, nərd, naz, nazik, nazir, nəzir, nəfis, naşir, nalə və s.

ortada – ana, ənənə, anım, qanlı, şanlı, and, səndəl, gənclik, geniş, eniş və s.

axırda – an, san, qan, yan, aman, saman, yaman, dan, can, zaman və s.

Geri assimilyasiya nəticəsində söz ortasında və sözün axırında gələn `n` səsi `m` ilə əvəzlənir: zə (m)bil, sü (m)bül, şə (m)bə, gü (m)bəz, tə (m)bəki, qa (m)maz, di (m)məz və s.

36. L (l) – dilönü, dil-diş samitidir, dildə süzgün və cingiltili tələffüz olunur. Bu samitə sözün əvvəlində, ortasında və axırında rast gəlmək mümkündür:

əvvəldə – lalə, leylək, lil, limon, lətifə, lələzar, ləng, lərzə, lülə, lək və s.

ortada – ələk, əlləş, əlbir, əlbəttə, qələm, qələt, mələz, məlhəm, sələm və s.

axırda – əl, gəl, dəl, vəl, ol, qol, qul, qızıl, atıl, bal, şal, zal, dal, mal və s.

Tələffüzdə dəyişikliyə çox məruz qalan səslərdən biridir. `l` səsi assimilyasiya nəticəsində `t`, `d`, `s`, `z`, `ş`, `n`, `m`, `p` səslərindən sonra həmin səslərə, ya da artikulyasiyasına görə onlara yaxın olan səslərə keçir [Müasir Azərbaycan dili, 1978, s.91].

37. R (r) – dilönü, dil-diş, süzgün və cingiltili samitdir. Sözün əvvəlində, ortasında və axırında müşahidə olunur:

əvvəldə – rayon, rica, rəis, radio, ramazan, rəhmdil, rəhbər, rəndə və s.

ortada – arzu, arvad, arxa, arıq, artıq, armud, arı, arsız və s.

axırda: ar, ər, vur, bur, sür, bar, tar, yar, qar, şar, dar, nar, qur, dur və s.

38. F (f) dodaq-diş samiti, sürtünən, kar. Daha çox alınma sözlərdə söz əvvəlində, ortasında və axırında rast gəlinir:

əvvəldə – fəhlə, fikir, fağır, futbol, fizika, fəlakət, fərəh, fənn və s.

ortada – dəftər, dəfə, səfər, həftə, zəfər, nəfər, məftil, məftun və s.

axırda – dəf, zərif, lətif, hədəf, tərəf, hərf, sərf, məsrəf, əşrəf, ətraf və s.

39. V (v) – dodaq-diş samiti, sürtünən, cingiltili. Sözlərin əvvəlində, ortasında və axırında rast gəlinir:

əvvəldə – Vətən, vətər, vəzəri, vəzir, vücud, var, vur, vuruş, vida və s.

ortada – sərvət, dövlət, dəvət, mehvər, əhval, əhvalat, dava və s.

axırda – sev, ev, qov, alov, ov, div, Murov, zıncırov, çalov, dəlisov və s.

40. P (p) – qoşadodaq, partlayan, kar samitdir. Bu samitə söz əvvəlində, ortasında və daha çox təkhecalı sözlərin sonunda təsadüf edilir:

əvvəldə – papaq, palıd, pul, parıltı, pudra, peç, puç, pələng, pişik və s.

ortada – hampa, lampa, süpürgə, öpüş, apar, ürpəş (mək), çəpər, təpər və s.

axırda – qulp, ərp, çöp, qap, sap, yap, öp, qop, top, çap, lap, tap, çəp və s.

41. B (b) – qoşadodaq, partlayan, cingiltili samitdir. Sözlərin əvvəlində, ortasında və sonunda işlənir:

əvvəldə – baba, bala, balaca, baş, buket, banket, baxış, bat (maq) və s.

ortada – xəbər, araba, xaraba, zərbə, zorba, torba, mübarizə və s.

axırda – ərəb, qərb, ədəb, zərb, tələb, səlib, qürub, təb, dəb, ləb, doşab və s.

42. M (m) – qoşadodaq, burun samiti, cingiltili. Sözüün əvvəlində, ortasında və sonunda müşahidə olunur:

əvvəldə – mən, mərkəz, mədən, müzakirə, müşahidə, müsahibə və s.

ortada – hamı, həməən, mamır, məmur, dəmir, təmir, kamil və s.

axırda – tam, kam, şam, dam, salam, adam, qədəm, sələm, qələm və s.

Azərbaycan dilinin fonemlərinin səciyyəvi əlamətləri. Fonemlərin səciyyəvi fono-akustik, anatomik-fizioloji xüsusiyyətlərinin hərtərəfli tədqiqi və universal kodlaşdırma səsin tanınması sistemlərində tətbiq oluna bilər. Səsin yazıya (mətnə) və yazının (mətnin) səsə çevrilməsi sistemlərində Azərbaycan fonemlərinin fono-akustik xüsusiyyətlərindən faydalanmaq olar. Həmin əlamətləri yığcam şəkildə aşağıdakı kimi ifadə etmək olar:

Saitlər

1. A (a) – [sait – dilarxası, açıq, dodaqlanmayan, təkavazlı, qısa, qalın, uca – (+ S 3 3 – 0 0 0)].

2. A (a): – (aa) – [dilarxası, açıq, dodaqlanmayan, təkavazlı, uzun, qalın, uca – (+ 1 3 3 – 0 0 0)].

3. Ə (ə) – [sait, dilönü, açıq, dodaqlanmayan, təkavazlı, qısa, incə, uca – (+ S 3 1 – 0 0 0)].

4. Ə (ə): – (əə) – [sait, dilönü, açıq, dodaqlanmayan, təkavazlı, uzun, incə, uca – (+ 1 3 1 – 0 0 0)].

5. E (e): – [sait, dilönü, yarımçıq, dodaqlanmayan, təkavazlı, qısa, incə, orta – (+ S 2 1 – 0 0 0)].

6. E (e): – (ee) – [sait, dilönü, yarımçıq, dodaqlanmayan, təkavazlı, uzun, incə, orta – (+ 1 2 1 – 0 0 0)].

7. I (ı) – [sait, dilarxası, qapalı, dodaqlanmayan, təkavazlı, qısa, qalın, aşağı – (+ S 1 3 – 0 0 0)].

8. İ (i) – [sait, dilönü, qapalı, dodaqlanmayan, təkavazlı, qısa, incə, orta – (+ S 2 1 – 0 0 0)].

9. İ̇ (i): – (ii) – [sait, dilönü, qapalı, dodaqlanmayan, təkavazlı, uzun, incə, orta – (+ 1 2 1 – 0 0 0)].

10. O (o) – [sait, dilarxası, açıq, dodaqlanan, təkavazlı, qısa, qalın, orta – (+ S 2 3 + 0 0 0)].

11. O (o): – (oo) – [sait, dilarxası, açıq, dodaqlanan, təkavazlı, uzun, qalın, orta – (+ 1 2 3 + 0 0 0)].

12. OU – [diftonq (+ d 2 3 + 0 0 0)].

13. Ö (ö) – [sait, dilönü açıq, dodaqlanan, təkavazlı, qısa, incə, orta – (+ S 2 1 + 0 0 0)].

14. Ö̇ (ö): – (öö) – [sait, dilönü, açıq, dodaqlanan, təkavazlı, uzun, incə, orta – (+ 1 2 1 + 0 0 0)].

15. ÖÜ – [diftonq (+ d 2 1 + 0 0 0)].

16. U (u) – [sait, dilarxası, qapalı, dodaqlanan, təkavazlı, qısa, qalın, orta – (+ S 2 3 + 0 0 0)].

17. U (u): – (uu) – [sait, dilarxası, qapalı, dodaqlanan, təkavazlı, uzun, qalın, orta – (+ 1 2 3 + 0 0 0)].

18. Ü (ü) – [sait, dilönü, qapalı, dodaqlanan, təkavazlı, qısa, incə, orta – (+ S 2 1 + 0 0 0)].

Samitlər. Azərbaycan dilində samitlərin səciyyəvi fonetik, anatomik-fizioloji əlamətləri: dodaq samitləri, dilönü samitlər, dilarxası samitlər, qoşadodaq, dodaq-diş, dil-diş, dil-damaq samitləri, kar, cingiltili, burun samitləri, sonorlar, partlayanlar, titrəyənlər, kipləşənlər, affrikatlar, sürtünənlər və s.

19. H (h) – [samit, boğazda sürtünən, kar samit (boğaz və ya farinqal) – (- 0 0 0 0 f g -)].

20. Q (q) – [samit, dilarxası, partlayan, cingiltili – (- 0 0 0 0 s v +)].

21. X (x) – [samit, dilarxası, kar samit – (- 0 0 0 0 f v -)].

22. Ğ (ğ) – [samit, dilarxası, sürtünən, cingiltili – (- 0 0 0 0 f v +)].
23. K* (k*) – [dilortası, partlayan, kar samit – (- 0 0 0 0 s p -)].
24. K** (k**) – [dilarxası, partlayan, kar samit – (- 0 0 0 0 s v -)].
25. G (g) – [dilortası, partlayan, cingiltili samit – (- 0 0 0 0 s p +)].
26. Y (y) – [dodaq-diş samiti, sürtünən, cingiltili – (- 0 0 0 0 f d +)].
27. T (t) – [dilönü, partlayan, kar samit – (- 0 0 0 0 s a -)].
28. D (d) – [dilönü, partlayan, cingiltili samit – (- 0 0 0 0 s a +)].
29. S (s) – [dilönü, sürtünən, kar samit – (- 0 0 0 0 f a -)].
30. Z (z) – [dilönü, sürtünən, cingiltili samit – (- 0 0 0 0 f a -)].
31. Ş (ş) – [dilönü, sürtünən, kar samit – (- 0 0 0 0 f p -)].
32. J (j) – [dilönü, sürtünən, cingiltili samit – (- 0 0 0 0 f p +)].
33. Ç (ç) – [dilönü, partlayan, kar samit – (- 0 0 0 0 a p -)].
34. C (c) – [dilönü, partlayan və cingiltili samit – (- 0 0 0 0 a p +)].
35. N (n) – [dilönü, burun, sonor, cingiltili samit – (- 0 0 0 0 1 a +)].
36. L (l) – [dilönü, dil-diş samiti – (- 0 0 0 0 1 a +)].
37. R (r) – [dilönü, dil-diş, süzgün və cingiltili samit – (- 0 0 0 0 r a +)].
38. F (f) – [dodaq-diş samiti, sürtünən, kar – (- 0 0 0 0 f d +)].
39. V (v) – [dodaq-diş samiti, sürtünən, cingiltili – (- 0 0 0 0 f d +)].
40. P (p) – [qoşadodaq, partlayan, kar samit – (- 0 0 0 0 s 1 -)].

41. B (b) – [qoşadodaq, partlayan, cingiltili samit – (- 0 0 0 0 s 1 +)].

42. M (m) – [qoşadodaq, burun samiti, cingiltili – (- 0 0 0 0 n 1 +)].

Qeyd etmək lazımdır ki, Azərbaycan dili fonemlərinin kodlaşdırılması sistemində beynəlxalq təcrübə nəzərə alınmışdır. Bu da gələcəkdə həmin kodlaşdırmadan daha geniş dairədə istifadə olunması imkanı yaradır [Fətullayev, Mahmudov, Abbasov və b., 2015, s.33-39].

Fonetik səviyyədə avtomatik işarələnmə. Fonemlərin səciyyəvi fono-akustik, anatomik-fizioloji əlamətlərinin kodlaşdırılması (işarələnməsi) nəzəri və təcrübi baxımdan böyük əhəmiyyət kəsb edir. Fonemlərin səciyyəvi fono-akustik, anatomik-fizioloji xüsusiyyətlərinin hərtərəfli tədqiqi və universal kodlaşdırma səsin tanınması sistemlərində tətbiq oluna bilər. Səsin yazıya (mətnə) və yazının (mətnin) səsə çevrilməsi sistemlərində Azərbaycan fonemlərinin fono-akustik xüsusiyyətlərindən faydalanmaq olar.

Qeyd etmək lazımdır ki, Azərbaycan dili fonemlərinin beynəlxalq standartlara uyğun kodlaşdırılması və bütün bunların cədvəl formasında təsbit olunması fonetik səviyyədə avtomatik işarələnmə üçün də imkan yaradır. Yəni həmin texnologiya mətnlərin fonetik səviyyədə avtomatik işarələndirilməsi prosesində istifadə oluna bilər.

Fonemlərin fono-akustik, anatomik-fizioloji əlamətlərinin kodları aşağıdakılardır:

1. Sait və ya samit
(v c + -)
2. Saitin uzunluğu: qısa, uzun, diftonq, neytral
(v l n g s l d a 0)
3. Saitin ucalığı (tonu): uca (yüksək), orta, aşağı (alçaq)
(v height 1 2 3 0)
4. Saitin önlüyü (mövqeyi) : ön, orta, arxa

(v front 1 2 3 0)

5. Dodaqlanması: dodaqlanan, dodaqlanmayan

(v rnd + - 0)

6. Samit tipi: dayanıqlı, sürtünən (frikativ), affrikat, burun, lateral (dilyanı), approximant

(ctype s f a n l r 0)

7. Artikulyasiya (məxrəc) yeri: dodaq, alveylar (diş yuvaqları), damaq, dodaq-diş, diş, velyar (damaqarxası), udlaq

(cplace I a p b d v g 0)

8. Samitlərin tələffüzü: cingilti, kar

(c vox + - 0).

Azərbaycan dili fonemlərinin səciyyəvi fono-akustik, anatomik-fizioloji əlamətlərini cədvəl şəklində aşağıdakı kimi göstərmək olar [Mahmudov, 2018, s. 196-198].

Cədvəl 2. Azərbaycan dili fonemlərinin fono-akustik bloku

Nö	Fonemlər	Sa- itlər	Uzu nluq	Ton	Möv qe	Do- daq- lan ma	Sa- mit tipi	Arti- kul- ya- siya	Tə- ləf- füz
	a	+	s	3	3	-	0	0	0
	a:	+	1	3	3	-	0	0	0
	ə (aa)	+	s	3	1	-	0	0	0
	ə: (aa)	+	1	3	1	-	0	0	0
	e	+	s	2	1	-	0	0	0
	e:	+	1	2	1	-	0	0	0
	I (ii)	+	s	1	3	-	0	0	0
	i	+	s	2	1	-	0	0	0
	i:	+	1	2	1	-	0	0	0
	o	+	s	2	3	+	0	0	0

	o:	+	1	2	3	+	0	0	0
	ou	+	d	2	3	+	0	0	0
	ö (oo)	+	s	2	1	+	0	0	0
	ö: (oo)	+	1	2	1	+	0	0	0
	öü (oou u)	+	d	2	1	+	0	0	0
	u	+	s	2	3	+	0	0	0
	u:	+	1	2	3	+	0	0	0
	ü (uu)	+	s	2	1	+	0	0	0
	h (he)	-	0	0	0	0	f	g	-
	q (qe)	-	0	0	0	0	s	v	+
	x (xe)	-	0	0	0	0	f	v	-
	ğ (ğe) - (qq)	-	0	0	0	0	f	v	+
	k (ke)	-	0	0	0	0	s	p	-
	k (ka)	-	0	0	0	0	s	v	-
	g (ge)	-	0	0	0	0	s	p	+
	y (ye)	-	0	0	0	0	f	p	+
	t	-	0	0	0	0	s	a	-

Süni intellektin linqvistik problemləri

(te)									
d (de)	-	0	0	0	0	s	a	+	
s (se)	-	0	0	0	0	f	a	-	
z (ze)	-	0	0	0	0	f	a	+	
ş (şe)	-	0	0	0	0	f	p	-	
j (je)	-	0	0	0	0	f	p	+	
ç (çe)	-	0	0	0	0	a	p	-	
c (ce)	-	0	0	0	0	a	p	+	
n (en)	-	0	0	0	0	n	a	+	
l (el)	-	0	0	0	0	l	a	+	
m (em)	-	0	0	0	0	n	l	+	
r (er)	-	0	0	0	0	r	a	+	
f (fe)	-	0	0	0	0	f	d	-	
v (ve)	-	0	0	0	0	f	d	+	
p (pe)	-	0	0	0	0	s	l	-	
b (be)	-	0	0	0	0	s	l	+	

Cədvəl 2 beynəlxalq sistemdə qəbul olunmuş 8 ən mühüm fono-akustik və anatomik-fizioloji əlamət əsasında tərtib olunmuşdur [Fətullayev, Mahmudov, Abbasov və b., 2015, s.33-49]. Cədvəldə göstərilmiş kodlaşdırma sistemi korpusda fonetik səviyyədə avtomatik işarələndirilmə imkanı verir.

Morfoloji işarələnmə üçün zəruri əlamətlər toplusu. Formal təhlil sistemlərində və korpuslarda linqvistik işarələnmə bütün dil səviyyələrinə aiddir. Lakin bu sahədə dünya təcrübəsi göstərir ki, linqvistik işarələnmədə morfoloji səviyyə aparıcı rola və mövqeyə malikdir. Məsələn orasındadır ki, dilin sonrakı səviyyələrini üçün də ilkin zəruri məlumatlar morfoloji əlamətlər əsasında əldə edilir. Morfoloji işarələnmənin tamlığı, mükəmməlliyi qalan səviyyələr üçün də mühüm və zəruri amil hesab oluna bilər. Əslində, bir çox dil korpusları yalnız morfoloji səviyyədə işarələnmə ilə kifayətlənir, digər dil səviyyələrində işarələnmədən ya heç istifadə etmir, ya da qismən, fraqmentlər şəklində yararlanırlar.

Morfoloji işarələnmədə ilk olaraq mətndə müşahidə olunan söz-formanın bu və ya digər nitq hissəsinə mənsubluğu qeyd olunur. Sonra isə həmin konkret nitq hissəsinin ən mühüm morfoloji əlamətləri sadalanır. Qeyd etmək lazımdır ki, korpusun işarələnməsi prosesində bu əlamətlərin hamısı deyil, dilin təhlili, tədrisi və mənimsənilməsi baxımından əhəmiyyətli olanları əks olunur. Həmin linqvistik əlamətləri əks etdirən informasiya dilin sonrakı mərhələlərində də istifadə olunmalıdır. Bu əlamətlər aşağıdakılardır [Mahmudov, 2018, s.171-181].

İsim:

- konkret maddi varlıq bildirənlər (daş, ağac, qoyun, sərçə);
- mücərrəd mənəvi anlayış adı bildirənlər (gözəllik, cəsurluq, sevgi, arzu);

- cansız əşyaları bildirənlər (dəniz, qum, çəpər, təpə);
- canlı varlıqları bildirənlər (insan, heyvan, balıq, quş);
- xüsusi isimlər (Araz, Alabaş, Xəzər, Qafqaz);
- ümumi isimlər (meşə, dağ, su, çörək);
- sadə isimlər (daş, saç, göz, baş);
- düzəltmə isimlər (neftçi, oxucu, uçuş, çəmənlik);
- mürəkkəb isimlər (kəklükotu, qayınata, dəmirağac, əmioğlu);
- tək isimlər (alça, kitab, qələm, ana);
- cəm isimlər (kitablar, gözlər, müəllimlər, ağaclar);
- toplu isimlər (xalq, camaat, sürü, millət).

İsmin halları:

adlıq hal (müəllim, kitab, Bakı);

iyilik hal (müəllim-in, kitab-ın, Bakı-nın);

yönlük hal (müəllim-ə, kitab-a, Bakı-ya);

təsirlik hal (müəllim-i, kitab-ı, Bakı-nı);

yerlik hal (müəllim-də, kitab-da, Bakı-da);

çıxışlıq hal (müəllim-dən, kitab-dan, Bakı-dan).

İsmin mənsubiyyət kateqoriyası:

1-ci şəxs təkində: mənim kitab-ım, ata-m;

1-ci şəxs cəmində: bizim kitab-ımız, ata-mız;

2-ci şəxs təkində: sən kitab-ın, ata-n;

2-ci şəxs cəmində: sizin kitab-ınız, ata-nız;

3-cü şəxs təkində və cəmində: onun/onların kitab-ı, ata-sı.

İsmin xəbərlik kateqoriyası:

1-ci şəxs təki: müəllim-əm, tələbə-yəm; cəmi: müəllim-ik, tələbə-yik;

2-ci şəxs təki: müəllim-sən, tələbə-sən; cəmi: müəllim-siniz, tələbə-siniz;

3-cü şəxsin təki: müəllim-dir, tələbə-dir; müəllim-dirlər, tələbə-dirlər.

Sifət:

sadə sifət (sarı, ağ, şirin, turş);

düzəltmə sifət (ağıllı, duzsuz, qəndsiz çay, ətirli);

mürəkkəb sifət (uzunsaçlı, şirin-şirin, qısaboy, uzunburun);

sifətin substantivləşməsi (cavan qocaya yer verdi);

sifətin adverbiallaşması (gözəl danışmaq, yaxşı qaçmaq).

Sifətin dərəcələri:

adi dərəcə (ağ, qırmızı, şirin);

azaltma dərəcəsi (qırmızımtıl, sarımtıraq, ağımsov, sarı-şın, gödərək);

çoxaltma dərəcəsi (qıpqırmızı, gömgöy, tünd qırmızı, lap qırmızı).

Say:

miqdar sayları (bir, beş, min);

müəyyənlik bildirən (bir, iki, üç, milyon, milyard);

qeyri-müəyyənlik bildirənlər (az, çox, xeyli, lap az);

kəsr sayları (beşdə bir, ikidə bir);

sıra sayları (birinci, beşinci).

Əvəzlik:

şəxs əvəzlilikləri (mən, sən, o, biz, siz, onlar);

qayıdış əvəzlilikləri (öz, özüm, özümüz, özün, özünüz, özləri);

sual əvəzlilikləri (kim, nə, hara, neçə, necə, nə üçün, haçan);

qeyri-müəyyənlik bildirən əvəzliliklər (kimi, kimisi, kimsə, hər kim, hər nə, biri);

inkar əvəzlilikləri (heç kim, heç kəs, heç biri, kimsə);

işarə əvəzlilikləri (o, bu, elə, belə, həmin, haman);

nisbi əvəzliliklər (kim, nə, hara, nə qədər, necə, nə cür, hansı, haçan, nə zaman);

təyini əvəzliliklər (bütün, hər, hər bir, filan, bəzi).

Feil:

sadə feillər (bax, de, ye, as);
düzəltmə feillər (başla, səslən, sağal, arıqlaş);
mürəkkəb feillər (dərsə getmək, yeyə bilmək, qalxa bilmək, yaxşılıq etmək).

Növ kateqoriyası:

məlum növ (yazdı, açdı, alacaq, oxumuşdu);
məchul növ (yazdırılmaq, doğranmaq, silinmək, oxunmaq);
qayıdış növ (geyinmək, yuyunmaq, çırpınmaq, silkinmək);
qarşılıq növ (görüşmək, sevişmək, didişmək, yazışmaq);
icbar növ (yazdırmaq, açdırmaq, gözlətmək, aldırmaq);
təsdiq feillər (alacaq, baxacaq, oxudum, dedi);
inkar feillər (almayacaq, baxmayacaq, oxumadım, demədi).

Feilin şəkilləri:

əmr şəkli=feil kökü+şəxs şəkilçiləri: (1-ci şəxs, tək – açım, gəlim; cəm – açaq, gələk); (2-ci şəxs, tək – aç, gəl; cəm – açın, gəlin); (3-cü şəxs, tək – açsın, gəlsin; cəm – açsınlar, gəlsinlər);

xəbər şəkli=feil kökü və ya əsası+zaman şəkilçisi+şəxs şəkilçiləri: (1-ci şəxs tək – yaz-dı-m; cəm – yaz-dı-q); (2-ci şəxs tək – yaz-dı-n; cəm – yaz-dı-nız); (3-cü şəxs tək – yaz-dı; cəm – yaz-dı-lar);

vacib şəkli=feil kökü və ya əsası+-malı, -məli+şəxs şəkilçisi (al-malı-yam, gəl-məli-yəm, al-malı-san, gəl-məli-sən, al-malı-dır, gəl-məli-dir, al-malı-yıq, gəl-məli-yik, al-malı-sınız, gəl-məli-siniz, al-malı-dırlar, gəl-məli-dirlər);

arzu şəkli=feil kökü+-a, -ə+şəxs şəkilçiləri (gərək al-a-m, gərək al-a-san, gərək al-a, gərək al-a-q, gərək al-a-sınız, gərək al-a-lar);

lazım şəkli=feil kökü və ya əsası+ (-ası, -əsi)+şəxs şəkilçisi (al-ası-yam, gəl-əsi-yəm, al-ası-san, gəl-əsi-sən, al-ası-

dır, gəl-əsi-dir, al-ası-yıq, gəl-əsi-yik, al-ası-sınız, gəl-əsi-siniz, al-ası-dırlar, gəl-əsi-dirlər);

şərt şəkli=feil kökü və ya əsası+ (-sa, -sə)+şəxs şəkilçiləri (al-sa-m, gəl-sə-m, al-sa-n, gəl-sə-n, al-sa, gəl-sə, al-sa-q, gəl-sə-k, al-sa-nız, gəl-sə-niz, al-sa-lar, gəl-sə-lər);

feilin bacarıq tərzii=feil kökü+ (-a, -ə)+bilmək (yaz-a bilmək, gəl-ə bilmək).

Feilin şəxs kateqoriyası:

1-ci şəxs təki: -am, -əm, -m, -ım, -im, -um, -üm (al-ım, gəl-im, aldı-m, gəldi-m, almış-am, gəlmiş-əm);

1-ci şəxs cəmi: -aq, -ək, -q, -ıq, -ik, -uq, -ük (al-aq, gəl-ək, aldı-q, gəldi-k, yazmış-ıq, gəlmiş-ik, durmuş-uq, gülmüş-ük);

2-ci şəxs təki: -n, -san, -sən (aldı-n, gəldi-n, almış-san, gəlmiş-sən);

2-ci şəxs cəmi: -ın, -in, -un, -ün, -nız, -niz, -nuz, -nüz, -sınız, -sınız, sunuz, -sünüz (al-ın, gəl-in, dur-un, gül-ün, aldı-nız, gəldi-niz, durdu-nuz, güldü-nüz, alır-sınız, gəlir-siniz, durur-sunuz, gülmür-sünüz);

3-cü şəxs təki: -sın, -sin, -sun, -sün, -dır, -dir, -dur, -dür (al-sın, gəl-sin, dur-sun, gül-sün, almış-dır, gəlmiş-dir, durmuş-dur, gülmüş-dür);

3-cü şəxs cəmi: -lar, -lər, -sınlar, -sinlər, -sunlar, -sünlər, -dırlar, -dirlər (aldı-lar, gəldi-lər, al-sınlar, gəl-sinlər, dur-sunlar, gül-sünlər, yazmış-lar, gəlmiş-lər).

Feilin zamanları:

Keçmiş zaman:

şühudi keçmiş: -dı, -di, -du, -dü (gəl-di, al-dı, dur-du, gül-dü);

nəqli keçmiş: -mış, -miş, -muş, -müş (al-mış-am, gəl-miş-əm, dur-muş-am, gül-müş-əm).

İndiki zaman

-ır, -ir, -ur, -ür (al-ır-am, gəl-ir-əm, dur-ur-am, gör-ür-əm);

Gələcək zaman:

qəti gələcək zaman: -acaq, -əcək (al-acaq, gəl-əcək);

qeyri-qəti gələcək: -ar, -ər (al-ar, gəl-ər).

Feilin şəxssiz formaları. Feili sifət:

Keçmiş zaman feili sifətləri:

(-miş, -miş, -muş, -müş – sulanmış otlaq, əkilmiş sahə, qurumuş ağac, üzülmüş saplaq);

“-dıq”+ mənsubiyyət şəkilçisi:

1-ci şəxsin təkində:

-dığım, -diyim, -duğum, -düyüm (aç-dığım kitab, gəl-diyim yer, ol-duğum şəhər, sür-düyüm maşın);

1-ci şəxsin cəmində:

-dığımız, -diyimiz, -duğumuz, -düyümüz (aç-dığımız qapı, gəl-diyimiz nəticə, oxu-duğumuz məktəb, görüş-düyümüz adam);

2-ci şəxsin təkində:

-dığın, -diyin, -duğun, -düyün (al-dığın məktub, ye-diyin çörək, ol-duğun ölkə, böl-düyün vəsait);

2-ci şəxsin cəmində:

-dığınız, -diyiniz, -duğunuz, -düyünüz (bax-dığınız kino, səp-diyiniz toxum, qoş-duğunuz adam, gül-düyünüz lətifə);

3-cü şəxsin təkində:

-dığı, -diyi, -duğu, -düyü (al-dığı teleqram, de-diyi söz, qur-duğu tələ, gör-düyü mənzərə);

3-cü şəxsin cəmində:

-dıqları, -dikləri, -duqları, -dükləri (al-dıqları maaş, gəl-dikləri vaxt, qoy-duqları əmanət, bürü-dükləri paltar).

İndiki zaman feili sifətləri:

-an, -ən şəkilçisi ilə düzəlir (yaz-an şagird, gül-ən uşaq).

Gələcək zaman feili sifətləri:

- acaq, -əcək şəkilçisi az işlənəndir (atıl-acaq addım, gələcək gün);
- malı, -məli şəkilçisi (atıl-malı əşya, görül-məli iş);
- ası, -əsi şəkilçisi (yazıl-ası ərizə, deyil-əsi söz);
- ar, -ər və -maz, -məz şəkilçiləri (ax-ar su, ax-maz su, gül-ər üz, gül-məz üz).

Feili bağlama:

- ib, -ib, -ub, -üb şəkilçisi (al-ib, gəl-ib, dur-ub, de-yib);
- araq, -ərək şəkilçisi (al-araq, gəl-ərək);
- dıqda, -dikdə, -duqda, -dükdə (al-dıqda, gəl-dikdə, oturduqda, gör-dükdə);
- anda, -əndə şəkilçisi (al-anda, gəl-əndə);
- dıqca, -dikcə, -duqca, -dükcə şəkilçisi (al-dıqca, gəldikcə, ol-duqca, gül-dükcə);
- ınca, -incə, -unca, -üncə şəkilçisi (al-ınca, gəl-incə, durunca, gül-üncə);
- madan, -mədən (dayan-madan, gör-mədən);
- caq, -cək şəkilçisi (al-caq, gör-cək);
- alı, -əli şəkilçisi (al-alı, ged-əli);
- ar-maz, ər-məz şəkilçisi (gör-ər-gör-məz, bax-ar-bax-maz);
- miş, -miş, -muş, -müş (qal-mış, gəl-miş, quruma-mış, çürümə-miş, ol-muş, gül-müş);
- arkən, -ərkən, -irkən, -ürkən (yat-arkən, gəl-ərkən, bax-irkən, gör-ürkən).

Məsdər:

- maq, -mək şəkilçisi (oyna-maq, gəl-mək).

Feil kökü və ya əsasına -ma, -mə şəkilçisi artırılmaqla feili isim düzəlir (bax-ma, gəl-mə) – feilin substantiv forması olub, hərəkətin adını bildirir.

Zərf:

Hərəkətin icra tərzini, hərəkətin və əlamətin keyfiyyətini bildirən tərz zərfləri aşağıdakı şəkilçilərlə düzəlir:

- la, -lə şəkilçisi (həyəcanla, təəccüblə);
- dan, -dən (astadan, ürəkdən);
- casına, -cəsinə (amansızcasına, dəlicəsinə);
- ca, -cə (ehmalca, sakitcə);
- akı, -əki (yanakı, çəpəki);
- yana, -anə (adamyana, mərdanə);
- dıqca, -dikcə (baxdıqca, dedikcə).

Zaman zərfləri:

Sadə zaman zərfləri:

indi, bayaq, dünən, hələ, daim, həmişə, sonra, əvvəl, səhər, axşam, gecə, bildir, sabah, bəzən, tez, gec və s.

Düzəltmə zaman zərfləri:

- dan, -dən şəkilçisi (bayaq-dan, tez-dən);
- liklə, -likdə (tez-liklə, indi-likdə);
- ca, -cə (əvvəl-cə, yeni-cə).

Mürəkkəb zaman zərfləri:

səhər-səhər, tez-gec, bugün-sabah, gecə-gündüz.

Yer zərfləri:

Sadə yer zərfləri:

irəli, geri, aşağı, yuxarı, bəri və s.

Düzəltmə yer zərfləri:

lıqda, -likdə (orta-lıqda, geri-likdə).

Mürəkkəb yer zərfləri:

ora-bura, yuxarı-aşağı, aşağı-yuxarı, sağa-sola.

Miqdar zərfləri:

Sadə miqdar zərfləri:

az, çox, xeyli, tamam.

Düzəltmə yer zərfləri:

tamamilə, bütünlüklə, dəfələrlə, qədərincə, lazımcına.

Mürəkkəb miqdar zərfləri:

az-çox, çox-çox, bir az, bir xeyli, bir qədər, bu qədər, nə qədər, o qədər, bir kərə, beş dəfə.

Köməkçi nitq hissələri

Qoşma:

Yalnız ismin adlıq halında olan sözlərə qoşulan qoşmalar:

haqda, barədə, halda, içrə;

Həm adlıq, həm də yiyəlik halda olan sözlərə qoşulanlar:

tək/ təki, haqqında, barəsində, üçün;

Yönlük hal ilə işlənən qoşmalar:

görə, qarşı, tərəf, doğru, sarı, məxsus, aid, dair, nisbətən, -dək;

Adlıq, yiyəlik, yönlük halda olan sözlərə və birləşmələrə qoşulanlar:

kimi, qədər, -can, -cən;

Çıxışlıq halda olan sözlərə qoşulanlar:

əvvəl, qabaq, sonra, bəri, başqa, özgə, qeyri, savayı, ötrü, çox, artıq;

Sabit qoşmalar:

üçün, ötrü, görə, kimi, üzrə, içrə, təki, sarı, dair, sayaq, məxsus;

Qeyri-sabit qoşmalar:

tək, doğru, bəri, əlavə, əvvəl, sonra, qabaq, başqa, qeyri, tərəf, irəli, qədər, geri, əlavə və s.

Şəkilçiləşən qoşmalar:

-can, -cən, -ca, -cə, -dək.

İlə qoşması şəkilçiləşmə prosesi keçirir, həm ilə, həm də -la, -lə şəklində işlənir.

Bağlayıcı:

Və, amma, ancaq, lakin, hətta, yəni, zira, ki, ilə, çünki, isə, ona görə də, buna görə də, belə ki, ona görə ki, buna görə ki, odur ki, gah, gah da, həm, həm də, nə, nə də.

Tabesizlik bağlayıcıları:

birləşdirmə bağlayıcıları: və, ilə, həm, həm də, həmçinin, da, də, nə, nə də;

qarşılaşdırma bağlayıcıları: amma, ancaq, lakin, fəqət, halbuki, yoxsa;

bölüşdürmə bağlayıcıları: ya, ya da, yaxud da, gah, gah da, gah da ki, istər, istərsə də;

aydınlaşdırma bağlayıcıları: yəni, yəni ki.

Tabelilik bağlayıcıları:

aydınlaşdırma bağlayıcıları: ki, belə ki;

səbəb bağlayıcıları: çünki, zira, ona görə ki, buna görə ki, ondan ötrü ki, bundan ötrü ki, bunun üçün ki;

nəticə bildirən bağlayıcılar: ona görə, buna görə, ona görə də, buna görə də, onun üçün də, bunun üçün də, odur ki.

Ədat:

Sabit ədatlar:

ən, lap, daha, düm, bəs, məgər, kaş, axı, di, mağul, təki, bəli, yox, xeyr, heç;

Həm əsas nitq hissəsi, həm də ədat kimi işlənənlər:

elə, belə, düz, artıq, təkcə, bir, bircə, bax, gəl, gəlin, gör, görün, görək, qoy, qoyun və s.

Şəkilçiləşən ədatlar:

sana, -sənə; -mı, -mi, -mu, -mü;

Qüvvətləndirici ədatlar:

ən, lap, daha, daha da, bir daha, düm, ha, yenə;

Təsdiq ədatları: bəli, hə, bəli də;

İnkar ədatları: yox, xeyr, heç, heç də, əsla;

Sual ədatları: axı, bəs, məgər, yoxsa, olmaya, görəsən, yəni, -mı, -mi, -mu -mü və s.

Təkid bildirən ədatlar:

qoy, qoysana, bax, gəl, gəlin, gör, görək, görüm, di və s.

Fərqləndirici, istisnaedici ədatlar:

amma, ancaq, yalnız, fəqət, təkcə, bircə və s.

Emosionallıq bildirən ədatlar:

kaş, barı, ki, belə, məhz, görək, ha, axı, odur və s.

Modal sözlər:

Sadə modal sözlər: əlbəttə, xülasə, yaxşı, təəssüf, yəqin və s.

Düzəltmə modal sözlər: doğrusu, düzü, mənəcə, bizcə, sözsüz, şübhəsiz, əvvələn və s.

Mürəkkəb modal sözlər: sanki, doğrudan-doğruya, elə bil ki, bəlkə də, ehtimal ki, daha doğrusu, yəqin ki, tutaq ki, doğrudan da və s.

Yəqinlik bildirənlər: əlbəttə, şübhəsiz, doğrudan da, həqiqətən də və s.

Güman, ehtimal, şübhə bildirənlər: bəlkə, bəlkə də, yəqin, yəqin ki, ehtimal ki, güman ki, deyəsən və s.

Bənzətmə, müqayisə bildirənlər: elə bil, elə bil ki, sanki, deyəsən, guya ki və s.

Nəticə, sıra, ümumiləşdirmə bildirənlər: ümumiyyətlə, beləliklə, xülasə, deməli, nəhayət və s.

Mənbə və isnad bildirənlər: mənəcə, sənəcə, deyilənə görə, fikrimcə və s.

Nida:

əslı nıdalar: a, ax, eh, əh, oh, of, ay, ba, pa, uf, uy, ux, buy, bay, pah, pəh və s.

emosıonallıq bırdırənlər: ah, aha, ax, oh, oho, oy, uy, bay, buy, boy, bah, pəh, vah, tfu, xux, ura, baho, paho, ay hay, ey vah, hay-hay, paho, ey, vaxsey, heyhat, mərhəba, afərin və s.

çağırış bırdırənlər: a, a!, Çəkıl!, Ayy! və s.

Morfoloji səviyyədə avtomatik işarələnmə. Söz-formaların avtomatik lemmatizasiyası. Söz-formaların lemmatizasiyasına korpus mətnləri kontekstində baxmaq lazım gəlir. Məlum olduđu kimi, mətn müxtəlifölçülü və müxtəlifsəciyyəli vahidlərdən ibarətdir. Onların hər birinin sərhədlərinin müəyyənləşdirilməsi və mətndən ayrılması söz-formaların lemmatizasiyası baxımından əhəmiyyətlidir.

Mətndən abzasın ayrılması şərti səciyyə daşıyır. Abzaslar bir-birindən və mətndən probel (boşluq) işarəsi və ya qabaqcadan şərtləşdirilmiş hər hansı işarə ilə ayrıla bilər. Abzas bir və ya bir neçə cümlədən ibarət ola bilər.

Cümlə abzasdan kiçik, söz-formadan böyük vahid sayılır. Cümlə abzasdan onun əvvəlində və sonunda qoyulmuş durğu işarələri ilə ayrılır. İki nöqtə, nöqtə ilə sual, nida işarələri, nöqtə ilə nöqtəli vergül arasındakı məsafədə yerləşən söz-formalar çoxluğu (və ya söz-forma) cümlə hesab olunur.

Söz-formaların cümlədən ayrılması fərqli şəkildə baş verir. Bu vaxt nəzərdə tutulur ki, söz-forma bir probel işarəsindən digər probel işarəsinə qədər olan məsafədə yerləşən hərflər ardıcılığı zənciridir.

Söz-formaların hissələrinə (onu təşkil edən morfemlərə) ayrılması çox mürəkkəb və mühüm mərhələlərdəndir. Lemmatizasiya söz-formanın lüğət vahidi şəklinə salınmasıdır. Deməli, söz-formadan bütün sözdəyişdirici (qrammatik) şəkilçilər (və ya şəkilçi) ayrılmalıdır. Söz-forma ya kökdən, ya da əsasdan (kök+sözdüzəldici (leksik) şəkilçi/şəkilçilər) ibarət ola

bilər. Məlumdur ki, söz kökü və ya əsas lüğət vahidi qismində lüğətlərdə təsbit oluna bilər. Bu prosesin reallaşdırılması üçün qabaqcadan tərtib olunmuş morfem siyahılarının mövcudluğu çox vacibdir. Kompüterin yaddaşında onun üçün nəzərdə tutulmuş xüsusi sahədə lüğət və zəruri linqvistik məlumatlar şəkilçilər siyahısı, şəkilçilərin qrammatik səciyyələri əsasında söz-formaların düzgün olaraq hissələrinə (komponentlərinə) ayrılması (kökə, əsasa və ayrı-ayrı şəkilçilərə bölünməsi) baş verir.

Nəzəri olaraq türk dillərində istənilən söz kökünə n sayda şəkilçi birləşdirmək mümkündür. Gündəlik həyatda, praktikada isə vəziyyət bir qədər fərqlidir. Praktik istifadədə türk dillərində hər bir söz-formada 10-a qədər müxtəlif şəkilçi işlənə biləcəyini nəzərə alsaq, məsələnin nə dərəcədə mürəkkəb olduğu bir daha aşkar olur. Deməli, elə bir optimal üsul işlənilib hazırlanmalıdır ki, söz-formadakı morfemlər qrammatik baxımdan tam olaraq düzgün ayrılı bilsin və eyni zamanda bu zaman söz-formanın əvvəlki mənası təhrif olunmasın. Xüsusi olaraq tərtib olunmuş alqoritmin köməyi ilə söz-formanı təşkil edən morfemləri ayırmaq əvvəlcədən hazırlanmış morfem siyahıları ilə tutuşdurmaqla linqvistik baxımdan məqbul nəticələr əldə etmək mümkündür. Bir çox hallarda morfem siyahılarında morfem variantları (allomorflar) müstəqil morfemlər kimi təmsil olunur. Çünki, hətta morfem səviyyəsində belə qrammatik məna kontekstdə (qrammatik kontekstdə) məlum olur. Məsələn, Azərbaycan dilində indiki zaman şəkilçilərinin bütün allomorfları, ahəng qanunundan asılı variantları -ır (-ir, -ur, -ür, -yır, -yir, -yur, -yür) ayrı-ayrılıqda qrammatik informasiya ilə təchiz olunmuşdur. Əlbəttə, bir allomorfdan bu allomorfları təmsil edən əsas morfemə – ümumiləşdirici morfemə keçməyə imkan verən xüsusi qaydalar işləyib hazırlamaq olardı. Lakin, bu qaydalar təhlil

alqoritmlərinin daha da mürəkkəbləşməsinə və onların praktik reallaşdırılması istiqamətində çətinliklər yaratmış olardı.

Məhz ona görə də, ilk nəzərdən primitiv təsir bağıqlasa da kodlaşdırma zamanı hər bir allomorf üçün ayrılıqda qrammatik məlumatın hər dəfə təkrarlanması daha məqsədəuyğun hesab olunmuşdur. Sonralar lemmatizasiya prosesinin aprobe-siyası bu yanaşmanın düzgünlüyünü bir daha təsdiq etmişdir. Təhlildən fərqli olaraq sintez prosesi üçün bir əsas morfemdən allomorflara keçidin qanunauyğunluqlarının işlənilib hazırlanması çətinlik törətmir və təcrübədə özünü doğruldur.

Məsələn, hər hansı bir söz-formanın sintezi zamanı məlum qaydalar (şəkilçilərin sıralanması, ardıcılığı, ahəng qanunu və s.) öncədən saf-çürük edilir, yeganə düzgün variantın alınması təmin olunur. Məsələn, fərz edək ki, bizə aşağıdakı morfemlər verilmiş, onları Azərbaycan ədəbi dilinin qaydalarına müvafiq işlətməklə dilimizə zidd olmayan ifadələr almaq tələb olunur:

“baba”, “çiçək”, -lar (-lər), -ımız (-imiz, -umuz, -ümüz), -dan (-dən).

Təbiidir ki, “baba” sözündən sonra -ümüz, -dən, -lar şəkilçilərini işlətmək dilimizin qrammatikasına zidd olardı. Əvvəla, burada sözdüzəldici və sözdəyişdirici şəkilçilərin sıralanması qaydası, sonra isə ahəng qanunu pozulmuşdur. Sintez alqoritmləri “babaümüzənlər” söz-formasının düzgün olmadığını “hiss edirlər”. Çünki alqoritmlərdə şəkilçilərin sıralanması və ahəng qanunu özünün formal ifadəsini tapmışdır və formal olaraq bu proses heç bir xəyata yol verilmədən aparılır. Şəkilçilərin sıralanma ardıcılığına görə tapırıq ki, yuxarıda sadalanan bu şəkilçilər aşağıdakı ardıcılıqla düzülə bilər: əvvəlcə cəm şəkilçisi, sonra mənsubiyyət və hal şəkilçisi. Konkret nümunələrdən çıxış etsək: -lar-ımız-dan;-lər-imiz-dən. Cəm şəkilçisi olan -lar (-lər) dodaqlanan -umuz (-ümüz) şəkilçilərinə təsir edərək onları da özlərinə uyğunlaşdırır, yəni -

lar-umuz-dan, -lər-ümüz-dən deyil, -lar-ımız-dan, -lər-imiz-dən. Sonra sözün kökünə uyğun olaraq söz-forma tam şəkildə sintez edilir: “baba” sözü ahəngə görə -lar-ımız-dan, “çiçək” sözü isə -lər-imiz-dən şəkilçilərini qəbul edə bilər. Nəticədə babalarımızdan, çiçəklərimizdən söz-formaları alınır. Sintez alqoritmləri barədə bu izahat bir qədər səthi və bəsit görünə bilər, belə bir nəticəyə gəlmək olur ki, sintezdə bütün qrammatik qanunauyğunluqlar və istisnalar qabaqcadan nəzərə alınır.

Təhlildə isə proses mürəkkəb və ziddiyyətli gedir. Təhlil prosesinə cəlb olunmuş söz, onu təşkil edən söz, ona birləşmiş şəkilçilər, səsdüşümü, səstərini, kök-şəkilçi, şəkilçi-şəkilçi sərbəhlərində baş verən morfonoloji dəyişikliklər qabaqcadan məlum olmur və bu səbəblərə görə omonim və omoqraf olmuş, əslində isə müxtəlif kateqoriyaları təmsil edən morfemlər sistemdə bir qarışıqlıq, nizamsızlıq yaradır. Allomorfların bir morfemlə təmsil olunması da bu qarışıqlıq və nizamsızlığı bir qədər artırır.

Yuxarıda göstərilənləri əsas tutaraq, ümumi şəkildə ümumi morfoloji təhlil alqoritmini qurmaq olar. Həmin alqoritm söz-formanın lüğət formasına gətirilməsi – lemmatizasiyanı tam ehtiva edir. Bu alqoritmlərdə omonimlik məsələsi nəzərə alınmamışdır. Omonimliyin bu alqoritm çərçivəsində nəzərə alınması məsələni mürəkkəbləşdirmiş olardı. Belə hesab olunur ki, istər leksik-morfoloji, istərsə də sintaktik səviyyələrdə omonimliyin aradan qaldırılması, polisemiyanın ləğvi məsələləri bütövlükdə sistemin semantik blokunun üzərinə düşür.

Morfoloji təhlil alqoritmlərinin konkret izahına keçməzdən əvvəl aşağıdakı şərtləri qəbul edək:

N_i – avtomatik lüğətdə müəyyən qruplar üzrə yerləşdirilmiş əsasların uzunluğu, burada $i=1, \dots, k$;

N_1 – I qrupdakı əsasların uzunluğu, N_2 – II qrupdakı əsasların uzunluğu və nəhayət N_k – k qrupundakı əsasların uzunluğunu göstərir;

n_i – şəkilçilər siyahısında müəyyən qruplar üzrə yerləşdirilmiş şəkilçilərin uzunluğu, burada $i=1, \dots, r$.

n_1 – I qrupdakı şəkilçilərin uzunluğu, n_r – r qrupundakı şəkilçilərin uzunluğudur.

x_i – mətndə müşahidə olunan söz-formanın uzunluğu, u – həmin formadakı əsasın, z – isə şəkilçilərin (və ya şəkilçinin) uzunluğuna işarədir.

$x_i = u + z$ – mətndəki söz-formanın uzunluğu onun tərkibindəki əsasla sözdəyişdirici (qrammatik) şəkilçilərin uzunluqları cəminə bərabərdir. Bütün bu göstərilənləri nəzərə alıb, aşağıdakı alqoritmi qurmaq olar:

Söz-formanın təhlil üçün nəzərdə tutulmuş mətndən ayrılması.

Söz-formanın uzunluğunun təyini: əgər $x_i > N$, onda 5-ə, əks halda 3-ə keçid.

Söz-formanın AL-də əsaslar siyahısının müvafiq qrupu ilə müqayisəsi: əgər söz-forma tapılıbsa 4-ə, əks halda 6-ya keçid.

Tapılmış söz-formanın əlamətlərinin bu məqsədlə ayrılmış işçi sahəyə yazılması və 7-yə keçid.

$x_i = N_i - y$ bərabər uzunluqda söz-formanın ayrılması və 3-ə keçid.

Söz-formanın uzunluğunun bir hərf (simvol) azaldılması ($x_i - 1$): əgər $x = 0$ -dırsa, əks halda 3-ə keçid.

Qalığın varlığının yoxlanması (Z): əgər $Z = 0$ -dırsa 14-ə, əks halda 8-ə keçid.

Qalığın uzunluğunun müəyyənləşdirilməsi: əgər $Z > n_i$ -dirsə, 12-yə, əks halda 9-a keçid.

Qalığın müvafiq şəkilçilər qrupu ilə müqayisəsi: əgər şəkilçi tapılıbsa 10-a, əks halda 11-ə keçid.

Şəkilçinin əlamətlərinin əsasın əlamətlərindən sonra yazılması və 7-yə keçid.

Qalığın uzunluğunun bir hərf (simvol) azaldılması (Z-1): əgər $Z=0$ -dırsa, 9-a, əks halda 13-ə keçid.

ni-uzunluğunda qalığın ayrılması və 9-a keçid.

Söz-formanın naməlum sözlər siyahısına yazılması və 15-ə keçid.

Söz-formanın onun qrammatik əlamətləri ilə çapı və 15-ə keçid.

Mətnin sonu işarəsinin varlığının yoxlanması: əgər varsa, 16-ya, əks halda 1-ə keçid.

Prosesin sonu.

Söz-formaların lemmatizasiyasını tam əyani təsəvvür etmək üçün “qonşularımızdandır” söz-formasının formal təhlili prosesini nəzərdən keçirək. Söz-formanın uzunluğu 18 hərfdən ibarətdir: $x_i=18$. Əsaslar lüğətində ən uzun sözün (əsas və ya kök) uzunluğu 16 hərfdən ibarətdir ($N_i=16$). Deməli, nümunə kimi göstərdiyimiz söz yalnız iki hərf azaldıldıqdan sonra lüğətlə müqayisə oluna bilər ($x_i > N_i$ – yəni, $18 > 16$). Alqoritmə görə, bu şərt daxilində 5-ci addıma keçməliyik. 5-ci addıma görə $x_i=N_i$ -ə bərabər sayda hərf saxlamaqla söz-formanın uzunluğunu azaldırıq: $N_i=16$ olduğu üçün $x_i=16$ hərf saxlayırıq, qalıq işçi sahənin müvafiq bölməsində yaddaşa verilir: “qonşularımızdan” (“ır”-ayrılan hissə). Bu söz-formanı 16 hərfdən ibarət əsaslar yerləşən qrupla müqayisə edirik. Şərtə görə, əgər söz-forma orada tapılsaydı, 4-ə, əks təqdirdə 6-ya keçməli idik. Təbiidir ki, bu ifadə orada təsbit oluna bilməz, çünki tərkibində əsas+sözdəyişdirici şəkilçi var. Ona görə də şərtə əsasən 6-ya keçirik.

Altıncı addımın şərtinə görə, söz-formanın uzunluğunu bir hərf (bir simvol) azaldırıq. Əgər $x_i=0$ olsa 13-ə, əks halda 3-ə keçməliyik. x_i -sıfırdan fərqli olduğu üçün 3-ə keçirik. “Qonşularımızdan” ifadəsini 15 hərfdən ibarət əsaslar yerləşdirilmiş qrupla müqayisə edirik. Təbiidir ki, bu ifadə orada da

təsbit oluna bilməz. Dövri təkrarlar nəticəsində proses (hər dəfə söz-formanın sonundan bir hərf ayırmaqla onun əsaslar lüğətinin müvafiq qrupu ilə müqayisəsi) o vaxta qədər davam edir ki, “qonşu” sözü altı hərfdən ibarət sözlərin yerləşdiyi qrupda tapılır. Beləliklə, söz-formanın lemmatizasiyası başa çatır, onun lüğət formasına gətirilməsi prosesi yekunlaşır. Əslində, proses bu mərhələdə dayandırıla bilərdi. Lakin təhlil algoritmi prosesi davam etdirməyə, şəkilçilərin də bir-birindən ayrılması və onların qrammatik əlamətlərinin müəyyənləşdirilməsi imkanı yaradır. Söz-formanın lüğət vahidi ola biləcək hissəsi müəyyənləşdirildikdən sonra qalıqın (şəkilçilərin) varlığı yoxlanır.

Qalıqın varlığı aşkar olunduqdan sonra 8-ci addımın şərtləri yerinə yetirilir. $Z > ni$ -dirsə (yəni, şəkilçilər siyahısında yerləşən şəkilçilərin maksimal uzunluğundan böyükdürsə), 12-yə, əks halda 9-a keçirik. Bizim nümunədə qalıqın (larımızdandır) uzunluğu şəkilçilərin siyahıda təsbit olunmuş maksimal uzunluğundan (doqquz hərf) böyükdür. Ona görə də, 12-ci addımı yerinə yetiririk: şəkilçilərin maksimal uzunluğuna bərabər uzunluqda qalıq ayırırıq, qalan hissəni isə yadda saxlayırıq: $Z = ni = \text{larımızda}$.

Alınmış həflər zənciri müstəqil şəkilçi deyil, şəkilçilər birləşməsi olduğu üçün doqquz hərfdən ibarət olan şəkilçilərin yerləşdiyi qrupda təsbit oluna bilməz. Ona görə də, doqquzuncu addımın şərtlərinə görə 11-ə keçirik. On birinci addıma görə, həmin həflər zəncirindən bir hərf ayrılır və alınmış ifadə “larımız” ona müvafiq qrupla – səkkiz hərfdən ibarət şəkilçilər qrupu ilə müqayisə olunur. Bu qalıq orada da təmsil oluna bilməz, çünki Azərbaycan dilində belə şəkilçi yoxdur. Qalıq konkret bir şəkilçi deyil, bir neçə şəkilçinin birləşməsindən ibarət həflər zənciridir. Proses o vaxta qədər davam edir ki, “-lar” şəkilçisi üç hərfdən ibarət şəkilçilər qrupunda tapılır və onun qrammatik əlamətləri “qonşu” sözünün

qrammatik əlamətlərinə əlavə olunur. Sonra yenidən yeddinci addıma qayıdırıq və qalığın qalıb-qalmamasını yoxlayırıq. Məlum olur ki, qalıq var: “-ımızdandır”. Dövri təkrarların köməyi ilə hər dəfə qalığın sonundan bir hərf ayırıb onu şəkilçilər siyahısının müvafiq qrupu ilə müqayisə etməklə söz-formada olan bütün qrammatik şəkilçilərin səciyyələri tapılır və işçi sahəyə yazılır. Beləliklə, tam lemmatizasiya kimi səciyyələndirilə biləcək təhlil prosesində “qonşularımızdandır” söz-forması düzgün olaraq kök və şəkilçiyə bölünür, ayrı-ayrı şəkilçilər bir-birindən səhvsiz ayrılır və onların tam qrammatik səciyyələri əldə olunur. Əgər araşdırdığımız söz-formada omonim kök və ya şəkilçi müşahidə olunsaydı, onda semantik bloka müraciət etmək lazım gələrdi [Mahmudov, 2013, s.297]. Semantik blok istər morfoloji səviyyədə, istərsə də daha yuxarı dil səviyyələrində omonimliyin aradan qaldırılması və polisemiyanın optimal şəkildə həlli ilə bağlı alqoritmlər sistemi ilə təchiz olunmuşdur. Əlbəttə, omonimlik və polisemiya qismən morfoloji səviyyədə də aradan qaldırıla bilər. Bunun üçün morfoloji bloka əlavə olaraq bir sıra xüsusi alqoritmlərin əlavə edilməsi lazım gələrdi. Morfoloji səviyyədə həlli qeyri-mümkün olan problemlər semantik blokda yeni səviyyədə araşdırılır. Təcrübə göstərir ki, omonimlik və polisemiya semantik problemlər çərçivəsində daha tam, hərtərəfli və sistemli şəkildə təhlil oluna bilər.

Məlum olduğu kimi, linqvistik bazada əsaslar (söz kökləri) və şəkilçilər zəruri linqvistik məlumatla təchiz olunmuşlar. Əgər əsas və şəkilçinin yalnız bir mühüm qrammatik əlaməti varsa, birinci pozisiyada həmin əlamət qeyd olunur, qalan pozisiyalarda isə sıfır yazılır. Şəkilçilər və əsaslar siyahısında omonimlik halları fərqləndirilmir. Yəni, omonim əsasın və ya şəkilçinin bütün mümkün qrammatik əlamətləri onun qarşısında qeyd olunur. Leksik-morfoloji səviyyədə həmin əsas və ya şəkilçinin hansı qrammatik mənada işlənməsini

müəyyənləşdirmək qeyri-mümkündür. Belə situasiyalarda omonimlik və çoxmənalılıq dilin daha yuxarı səviyyələrində – sintaktik-semantik səviyyələrdə təhlil oluna bilər.

Məsələn, -acaq, -əcək (-yacaq, -yəcək) şəkilçiləri həm feilin qəti-gələcək zaman formasının, həm də feili sifətlərin əmələ gəlməsində iştirak edir. Bu qeyd olunan xüsusiyyət yalnız sintaktik səviyyədə fərqləndirilə bilər: kitab oxunacaq (qəti –gələcək zaman), oxunacaq kitab (feili sifət). Məhz bu hal nəzərə alınaraq linqvistik bazada bu qəbildən olan şəkilçilər üçün mümkün qrammatik əlamətlər göstərilir. Bəzi omonim şəkilçilər üçün isə dörd qrammatik əlamət göstərilir. Məsələn, -ın, -in, -un, -ün şəkilçisi ad nitq hissəsinə əlavə olunduqda ziyəlik halı və mənsubiyyəti (II şəxs, tək), feillərə birləşdikdə isə əmr şəklini (II şəxs, cəm) göstərir. Belə nümunələrin sayını artırmaq olardı.

Linqvistik bazada k–y, q–ğ əvəzlənmələrini nəzərə alaraq sonu -k və -q ilə bitən şəkilçilər iki variantda verilmişdir: -acaq/-acağ, -əcək/-əcəy.

Kompüterlərdə formal işlənmə sisteminə xas xüsusiyyətlər və qanunauyğunluqlar nəzərə alınaraq ünsiyyət prosesində tez-tez birləşmə şəklində, yanaşı işlənən şəkilçi birləşmələri müstəqil şəkilçi kimi qeyd olunur. Təcrübə göstərir ki, belə yanaşma formal təhlil sistemində özünü tam doğruldur. Məsələn, -sının, -sunun, -sinin, sünün, -sında, -sunda, -sində, -sündə və s.

Azərbaycan dilində söz-formaların tərkibində bəzi şəkilçilər sait və ya samit düşümü sahəsində sadələşmiş forma alır. Məsələn, -ma, -mə inkar şəkilçisi indiki zaman şəkilçisi ilə işləndikdə sait düşümü nəticəsində yeni bir forma alınır:

al+m (a+y) ır+sınız – al+mır+sınız

gəl+m (ə+y) ır+sınız – gəl+mir+siniz

Alınmış yeni şəkilçi forması (-mır, -mir, -mur, -mür) inkarlıq+indiki zaman qrammatik əlamətlərinə malikdir. Həmin

şəkilçi müşahidə olunan variantda linqvistik bazada şəkilçilər siyahısında təsbit olunmuşdur.

Nəqli keçmiş zaman şəkilçisi (-mı, -miş, -muş, -müş) son-da gələn “ş” samiti olmadan da işlənə bilər: al+mı (ş)+san – al+mı+san. Qeyd etmək lazımdır ki, bu xüsusiyyət artıq ədəbi dildə də möhkəmlənməkdədir. Digər tərəfdən, “ş” samiti olmadan işləndikdə formal təhlil sistemlərində bu şəkilçi -mı, -mi, -mu, -mü sual şəkilçiləri ilə də qarışdırıla bilər. Bu isə öz növbəsində formal təhlildə qarışıqlıqlar yarada bilər. Məhz bu xüsusiyyətləri əsas tutaraq həmin şəkilçinin birləşmə formasında verilməsi məqsədəuyğun hesab olunur: -mısan, -mısən, -musan, -müsən.

Korpusda söz-formaların lemmatizasiyası prosesində köməkçi idi, imiş feilləri linqvistik bazada ayrıca verilmişdir. Amma həmin köməkçi feillərin şəkilçilənmiş variantları da var ki, onların da şəkilçilər bölməsində yerləşdirilməsi məqsədəuyğun hesab olunmuşdur. Məsələn orasındadır ki, onlar müvafiq fonetik dəyişikliklərlə şəkilçi formasında da müşahidə olunur: almalı idi/almalıydı, gəlməli idi/gəlməliydi, almalı imiş/almalıymış, gəlməli imiş/gəlməliymiş.

Göstərilənlər ilə, üçün qoşmalarına da aiddir: atası ilə/atasıyla, əli ilə/əliylə, sənin üçün/səninçin, qardaşı üçün/qardaşıyçın.

Deyilənlərə əsaslanaraq qeyd olunan köməkçi feillərin və qoşmaların şəkilçi variantları da linqvistik bazadakı şəkilçilər bazasına daxil edilir: -ydi, -ydi, -yla, -ylə, -çın, -çin, -çun, -çün (-yçın, -yçin, -yçun, yçün), -ymış, -ymış, -ymuş, -ymüş.

Ənənəvi qrammatikalarda ayrıca şəkilçi kimi göstərilməyən, yalnız mətnlərin formal təhlil sistemlərində qeyd olunan bu şəkilçiləri məşin şəkilçiləri də adlandırırlar. Analoji olaraq yalnız formal təhlil sistemlərində müşahidə olunan məşin əsasları ifadəsinə də elmi ədəbiyyatda rast gəlinir [Махмудов, 1991; Mahmudov, 2013].

Artıq qeyd olunduğu kimi, söz kökləri və əsaslar (söz kökü+sözdüzəldici şəkilçi/şəkilçilər) ayrılıqda lüğət vahidi ola bildikləri üçün onlar linqvistik bazada ayrıca yerləşdirilir. Başqa sözlə, onların lemmatizasiyasına, lüğət vahidi formasına gətirilməsinə ehtiyac duyulmur. Məhz ona görə də əsas diqqət sözdəyişdirici şəkilçilərin (və ya şəkilçinin) söz kökündən (və ya əsasdan) düzgün ayrılmasına, onların kodlaşdırılmasına (linqvistik işarələnməsinə) verilir. Bu prosesin avtomatlaşdırılmasının korpusun linqvistik işarələnməsində misilsiz əhəmiyyəti vardır. Qeyd etmək yerinə düşərdi ki, kodlaşdırma, linqvistik işarələnmə həmişə şərti səciyyə daşıyır. Yəni, bu və ya digər səbəbə görə kodlaşdırma (linqvistik işarələnmə) dəyişdirilə, dəqiqləşdirilə bilər. Bəzən bu və ya digər dil ailəsi üçün (məsələn, türk dilləri üçün) vahid kodlaşdırma (linqvistik işarələnmə) sisteminə keçid cəhdləri ilə bağlı da kodlaşdırma sistemi tamamilə yenidən işlənib, dəyişdirilə bilər. Ona görə də, təklif olunan linqvistik işarələnmə sisteminə şərti yanaşmaq lazımdır. Kodlaşdırma zamanı mətnin formal təhlil sistemlərində istifadə olunmuş və özünü praktikada doğrultmuş kodlar əsas götürülmüşdür [Mahmudov, 2002].

Məlum olduğu kimi, Azərbaycan dili üçün işlənmiş formal təhlil sistemində sözdəyişdirici (qrammatik) şəkilçilər doqquz qrupa ayrılmışdır. Birinci qrupda doqquz hərfdən ibarət şəkilçilər, ikinci qrupda səkkiz hərfdən ibarət olanlar və nəhayət doqquzuncu qrupda yalnız bir hərfdən ibarət olan şəkilçilər yerləşdirilmişdir. Qruplarda yerləşdirilmiş şəkilçilər geniş mənada götürülmüşdür. Bura həm ənənəvi qrammatikada, məktəb qrammatikalarında təsbit olunmuş sözdəyişdirici şəkilçilər, həm də formal təhlil sistemlərində istifadə olunan şəkilçi birləşmələri, maşın şəkilçiləri daxildir. Onların bəziləri barədə yuxarıda məlumat verilmişdir.

Qruplar üzrə (doqquz qrupda) paylanmış Azərbaycan dilinin sözdəyişdirici şəkilçiləri qrammatik səciyyələri, şərti kodları və onlara aid konkret nümunələr göstərilməklə cədvəl şəklində əks olunmuşdur [Mahmudov, 2013, s. 204 – 2015].

Cədvəl 3. Azərbaycan dilinin sözdəyişdirici şəkilçiləri (qrammatik səciyyələr)

Qrup	Sözdəyişdirici şəkilçilər	Qrammatik əlamət	Kod
1	2	3	4
1. 0	-yacağımız, -yəcəyimiz oxuyacağımız, dəyəcəyimiz	feili sifət, I şəxs, cəm, qəti gələcək zaman	EAJW
1. 1	-yacağınız, -yəcəyiniz oxuyacağınız, dəyəcəyiniz	feili sifət, II şəxs, cəm, qəti gələcək zaman	E+JW
1. 2	-yacaqları, -yəcəkləri oxuyacaqları, dəyəcəkləri	feili sifət, III şəxs, cəm, qəti gələcək zaman	ELJW
2. 0	-acağımız, -əcəyimiz yazacağımız, görəcəyimiz	feili sifət, I şəxs, cəm, qəti gələcək zaman	EAJW
2. 1	-acağınız, -əcəyiniz yazacağınız, görəcəyiniz	feili sifət, II şəxs, cəm, qəti gələcək zaman	E+JW
2. 2	-acaqları, -əcəkləri yazacaqları, görəcəkləri	feili sifət, III şəxs, cəm, qəti gələcək zaman	ELJW
2. 3	-yacaqmış, -yəcəkmiş oxuyacaqmış, dəyəcəkmiş	gələcək zamanın rəvayəti) X00
3. 0	-yacağım, -yəcəyim	feili sifət, I şəxs,	EALW

Süni intellektin linqvistik problemləri

	oxuyacağım, deyəcəyim	tək, qəti gələcək zaman	
3. 1	-yacağın, -yəcəyin oxuyacağın, deyəcəyin	feili sifət, II şəxs, tək, qəti gələcək zaman	E+LW
3. 2	-dıqları, -duqları, -dikləri, -dükləri yazdıqları, oxuduqları, dedikləri, gördükləri	feili sifət, III şəxs, cəm, şühudi keçmiş zaman	ELJĞ
3. 3	-yacaqdı, -yəcəkdi oxuyacaqdı, deyəcəkdidi	gələcək zamanın hekayəti) 000
3. 4	-acaqmış, -əcəkmış oxuyacaqmış, deyəcəkmış	gələcək zamanın hekayəti) X00
4. 0	-yarkən, -yərkən, -yurkən, -yürkən, -yirkən, -yirkən oxuyarkən, deyərkən, oxuyurkən, deyirkən, başlayırkən, sürüyürkən	feili bağlama, qeyri-qəti gələcək, indiki zaman	= [00
4. 1	-mişkən, -muşkən -mişkən, -müşkən almışkən, olmuşkən, gəlmişkən, görmüşkən	feili bağlama, nəqli keçmiş zaman	+Q00
4. 2	-yandan, -yəndən oxuyandan, deyəndən	feili bağlama	=000
4. 3	-dırlar, -durlar, -dirilər, -dürlər adamdırlar, ulduzdurlar, tələbədirlər, güclüdürlər	III şəxs, tək	JI000
4. 4	-acağım, -əcəyim yazacağım, görəcəyim	feili sifət, I şəxs, tək, qəti gələcək	EALW

		zaman	
4. 5	-acağın, -əcəyin yazacağın, görəcəyin	feili sifət, II şəxs, tək, qəti gələcək zaman	E+LW
4. 6	-sından, -sundan, -sindən, -sündən atasından, qonşusundan, bibisindən, sürüsündən	mənsubiyyət, III şəxs, tək, çıxışlıq hal	R JILX
4. 7	-mışmış, -muşmuş, -mişmiş, -müşmüş yazmışmış, oxumuşmuş, gəlmişmiş, görmüşmüş	nəqli keçmişin rəvayəti	BI*00
4. 8	-yırmış, -yurmuş, -yirmiş, -yürmüş başlayırmış, oxuyurmuş, deyirmiş, çürüyürmüş	indiki zamanın rəvayəti	E000
4. 9	-acaqdı, -əcəkdı yazacaqdı, görəcəkdı	qəti gələcək zamanın rəvayəti	*000
4. 10	-yarmış, -yərmiş oxuyarmış, deyərmiş	qeyri-qəti gələcək zamanın rəvayəti] X00
4. 11	-yacağı, -yəcəyi oxuyacağı, deyəcəyi	feili sifət, III şəxs, tək, qəti gələcək zaman	EJ LW
4. 12	-yıbmış- yubmuş, -yibmiş, -yübmüş başlayıbmış, oxuyubmuş, deyibmiş, sürüyübmüş	keçmiş zamanın rəvayəti	*X00
5. 0	-maqda, -məkdə	indiki davamedici	/000

Süni intellektin linqvistik problemləri

	oxumaqda, gəlməkdə	zaman	
5. 1	-yaraq, -yərək oxuyaraq, deyərək	feili bağlama	=000
5. 2	-yanda, -yəndə oxuyanda, deyəndə	feili bağlama	=000
5. 3	-arkən, -ərkən, -irkən, -urkən, -irkən, -ürkən yazarkən, gələrkən, baxırkən, dururkən, gəlirkən, görürkən	feili bağlama	=000
5. 4	-dıqda, -dikdə, -duqda, -dükdə yazdıqda, gəldikdə, oxuduqda, gördükdə	feili bağlama	=000
5. 5	-dıqca, -duqca, -dikcə, -dükcə yazdıqca, oxuduqca, gəldikcə, gördükcə	feili bağlama	=000
5. 6	-yınca, -yunca, -yincə, -yüncə başlayınca, oxuyunca, deyincə, sürüyüncə	feili bağlama	=000
5. 7	-madan, -mədən baxmadan, demədən	feili bağlama	=000
5. 8	-andan, -əndən baxandan, gələndən	feili bağlama	=000
5. 9	-dığın, -duğın, -diyın, -düyün aldığın, oxuduğın, gəldiyin, gördüyün	feili bağlama	E+LF
5. 10	-acağı, -əcəyi yazacağı, görəcəyi	feili sifət, III şəxs, tək, qəti gələcək	EJ LW

		zaman	
5. 11	-yacaq, -yəcək, -yacağ, -yəcəy oxuyacaq/oxuyacağ, deyəcək/deyəcəy	qəti gələcək zaman, feili sifət	Ю000
5. 12	-sınız, -siniz, -sunuz, -sünüz yazırsınız, deyirsiniz, oxuyursunuz, görürsünüz	xəbərlik, II şəxs, cəm	J+Ж0
5. 13	-sının, -sunun, -sinin, -sünün atasının, qonşusunun, bibisinin, sürüsünün	mənsubiyyət, III şəxs, tək, yiyəlik hal	RJLIII
5. 14	-sında, -sunda, -sində, -sündə atasında, qonşusunda, bibisində, sürüsündə	mənsubiyyət, III şəxs, tək, yerlik hal	RJL LZ
5. 15	-ından, -undan, -indən, -ündən başından, qolundan, əлиндən, gözündən	mənsubiyyət, II şəxs, çıxışlıq hal	R+LX
5. 16	-mısan, -musan, -misən, -müsən yazmısan, oxumusan, demisən, görmüsən	nəqli keçmiş zaman, II şəxs, tək	BI+Lφ
5. 17	-mışdı, -muşdu, -mişdi, -müşdü yazmışdı, oxumuşdu, gəlmışdi, görmüşdü	keçmiş zamanın hekayəti	*X00
5. 18	-irmiş, -ürmüş, -irmiş, -ürmüş	indiki zamanın rəvayəti	[X00

Süni intellektin linqvistik problemləri

	yazırmış, dururmuş, gəlirmiş, görürmüş		
5. 19	-armış, -ərmiş yazarmış, görürmüş	qeyri-qəti gələcək zamanın rəvayəti] X 00
5. 20	-yardı, -yərdi oxuyardı, deyərdi	qeyri-qəti gələcək zamanın hekayəsi] 000
5. 21	-yırdı, -yurdu, -yirdi, -yürdü başlayırdı, oxuyurdu, işləyirdi, sürüyürdü	indiki zamanın hekayəsi	[000
5. 22	-ibmiş, -ubmuş, -ibmiş, -übmüş yazıbmış, durubmuş, gəlibmiş, görübmiş	keçmiş zamanın rəvayəti	*X00
6. 0	-yası, -yəsi oxuyası, deyəsi	feilin lazım şekli/feili sifət	Д000
6. 1	-malı, -məli oxumalı, görməli	feilin vacib şekli/feili sifət	E000
6. 2	-araq, -ərək baxaraq, görərək	feili bağlama	=000
6. 3	-anda, -əndə baxanda, görəndə	feili bağlama	=000
6. 4	-ınca, -incə, -unca, -üncə alınca, gəlincə, durunca, görünəcə	feili bağlama	=000
6. 5	-yalı, -yəli oxuyalı, deyəli	feili bağlama	=000
6. 6	-acaq, -əcəк, -acağ, -əcəy baxacaq/baxacağ, görəcəк/görəcəy	feili sifət	Д000
6. 7	-ımız, -imiz, -umuz,	mənsubiyyət, I	РАЖ0

	-ümüz kitabımız, dəftərimiz, odunumuz, gülümüz	şəxs, cəm	
6. 8	-ınız, -iniz, -unuz, -ünüz kitabınız, dəftəriniz, odunuz, gülünüz	mənsubiyyət, II şəxs, cəm	R+Ж0
6. 9	-ının, -inin, -unun, -ünün kitabının, dəftərinin, odunun, gülünün	mənsubiyyət, II/III şəxs, tək, yiyəlik hal	R+LIII
6. 10	-ında, -ində, -unda, -ündə kitabında, dəftərində, odununda, gülündə	mənsubiyyət, II/III şəxs, tək, yerlik hal	R+LZ
6. 11	-sına, -sinə, -suna, -sünə atasına, əmisinə, qonşusuna, sürüsünə	mənsubiyyət, II/ III şəxs, tək, yönlük hal	RJLIII
6. 12	-sını, -sini, -sunu, -sünü atasını, əmisini, qonşusunu, sürüsünü	mənsubiyyət, II/III şəxs, tək, təsirlik hal	RJL3
6. 13	-yçın, -yçin, -yçun, -yçün atasıyçın, əmisiyçin, qonşusuyçun, sürüsüyçün	qoşma	P000
6. 14	-ymış, -ymiş atasıymış, əmisiymiş	keçmiş zamanın rəvayəti	BI000
6. 15	-ycan, -ycən atasıycan, əmisiycən	qoşma	P000
6. 16	-ması, -məsi	feili isim,	!RJL L

Süni intellektin linqvistik problemləri

	yazması, görməsi	mənsubiyyət, III şəxs, tək	
6. 17	-dırt, -dırt, -durt, -dürt yazdırt, gəzdirt, vurdurt, güldürt	icbar növ	Ф000
6. 18	-ıncı, -inci, -uncu, -üncü qırxıncı, beşinci, otuzuncu, üçüncü	sıra sayı	I 000
6. 19	-ardı, -ərđi yazardı, gələrdi	qeyri-qəti gələcəyin hekayəsi] 000
6. 20	-ırdı, -irdi, -urdu, -ürdü yazırdı, gəlirdi, dururdu, sürürdü	indiki zamanın hekayəsi	[000
7. 0	-ası, -əsi yazası, gələsi	lazım şəkli/feili sifət	Д000
7. 1	-maq, -mək, -mağ, -məy yazmaq/yazmağ, gəlmək/ gəlməy	məsdər	И000
7. 2	-yıb, -yib, -yub, -yüb başlayıb, deyib, quruyub, sürüyüb	feili bağlama/keçmiş zaman	:000
7. 3	-alı, -əli yazalı, görəli	feili bağlama	=000
7. 4	-yar, -yər oxuyar, deyər	qeyri-qəti gələcək zaman	ДW00
7. 5	-maz, -məz oxumaz, dinməz	qeyri-qəti gələcək, inkarlıq/feili sifət	Д- 00
7. 6	-caq, -cək alcaq, görcək	feili bağlama	=000
7. 7	-yan, -yən oxuyan, deyən	feili sifət/feili bağlama	Д000

7. 8	-miş, -miş, -muş, -müş yazmış, gəlmiş, oxumuş, görmüş	nəqli keçmiş zaman/feili sifət	Q000
7. 9	-yır, -yir, -yur, -yür başlayır, deyir, oxuyur, sürüyür	indiki zaman	B000
7. 10	-nın, -nin, -nun, -nün qapının, əlinin, qonşunun, sürünün	mənsubiyyət, II şəxs, tək, yiyəlik hal	R+L III
7. 11	-dan, -dən qapıdan, dəftərdən	çıxışlıq hal	X000
7. 12	-lar, -lər kitablar, dəftərlər	cəm hal	Ж000
7. 13	-yam, -yəm atayam, tələbəyəm	xəbərlik, I şəxs, tək	JAL 0
7. 14	-yım, -yim, -yum, -yüm başlayım, deyim, oxuyum, sürüyüm	əmr şəkli, I şəxs, tək	ALQ0
7. 15	-yaq, -yək oxuyaq, deyək	feilin əmr şəkli/arzu şəkli, I şəxs, cəm	AЖQ0
7. 16	-mız, -miz, -muz, -müz atımız, əmimiz, qonşumuz, sürümüz	mənsubiyyət, I şəxs, cəm	RAЖ0
7. 17	-san, -sən oğlansan, tələbəsən	xəbərlik, II şəxs, tək	+LJ 0
7. 18	-nız, -niz, -nuz, -nüz atanız, əminiz, qonşunuz, sürünüz	mənsubiyyət, II şəxs, cəm/xəbərlik	R+Ж0
7. 19	-yın, -yin, -yun, -yün bağlayın, deyın, oxuyun, sürüyün	II şəxs, cəm, əmr şəkli	Ц+Ж0

Süni intellektin linqvistik problemləri

7. 20	-sın, -sin, -sun, -sün yazsın, desin, dursun, sürüsün	III şəxs, tək, əmr şəkli	ЇЈІL 0
7. 21	-dır, -dir, -dur, -dür kitabdır, dəftərdir, odundur, sürüdür, yazdır, gəzdir, qurdur, sürdür	III şəxs, tək, xəbərlik/icbar	I ЈL 0
7. 22	-yıq, -yik, -yuq, -yük atayıq, tələbəyik, quruyuq, sürüyük	I şəxs, cəm/xəbərlik	J AЖ0
7. 23	-ını, -ini, -unu, -ünü kitabını, dəftərini, odununu, gülünü	mənsubiyyət, II/III şəxs, tək, təsirlik hal	RЈL L3
7. 24	-ına, -inə, -una, -ünə kitabına, dəftərinə, odununa, gülünə	mənsubiyyət, II/III şəxs, tək, yönlük hal	RЈLІІІ
7. 25	-mır, -mir- mur, -mür almır, gəlmir, oxumur, gülmür	indiki zamanın inkar forması	-B00
7. 26	-yla, -ylə arabayla, gəmiylə	qoşma/tərzi hərəkət zərfi	P000
7. 27	-ydı, -ydi balacaydı, tələbəydi	keçmiş zamanın hekayəsi	ІF 00
7. 28	-çın, -çin, -çun, -çün atançın, səninçin, onunçun, özünçün	qoşma	P000
7. 29	-can, -cən oracan, evəcən	qoşma	P000
7. 30	-yıl, -yil, -yul, -yül yuyulmaq, deyilmək,	məchul növ	<000
7. 31	-ncı, -nci altıncı, ikinci	sıra syı	I000

7. 32	-dığ, -diy, -duğ, -düy ayrılıda işlənmiş. Şəxs şəkilçiləri ilə birləşmə şəklində geniş müşahidə olunur: başladığım, başladığımız, başladığın, başladığınız, başladığı, başladıkları və s.	feili sifət, şühudi keçmiş zaman	EF00
8. 0	-ya, -yə oxuya, deyə, qarıya, gəmiyə	arzu şəkli/yönlük hal	2000
8. 1	-sa, -sə alsa, gəlsə	şərt şəkli	G000
8. 2	-ib, -ib, -ub, -üb alıb, gəlib, durub, görüb	feili bağlama/ keçmiş zaman	:000
8. 3	-ar, -ər alar, gələr	qeyri-qəti gələcək zaman/feili sifət	ДW00
8. 4	-an, -ən alan, gələn	feili sifət/feili bağlama	Д000
8. 5	-dı, -di, -du, -düy aldı, gəldi, durdu, gördü	şühudi keçmiş zaman	F000
8. 6	-ır, -ir, -ur, -ür alır, gəlir, durur, görür	indiki zaman	B000
8. 7	-ma, -mə yazma, gəlmə, burma (saç), döymə (metal)	inkarlıq/feili sifət	:000
8. 8	-ın, -in, -un, -ün kitabın, dəftərin, odunun, gülün, atın, gəlin, oturun, gülün	yiyyəlik hal/mənsubiyyət, II şəxs, tək/ əmr şəkli, II şəxs, cəm	9000

Süni intellektin linqvistik problemləri

8. 9	-nı, -ni, -nu, -nü qapını, gəmini, qutunu, sürünü	mənsubiyyət, II/III şəxs, tək, / təsirlik hal	3000
8. 10	-da, -də kitabda, dəftərdə	yerlik hal	Z000
8. 11	-am, -əm oğlanam, müəlliməm, baxıram, görürəm	xəbərlik, I şəxs, tək	ALJφ
8. 12	-ım, -im, -um, -üm kitabım, dəftərim, odunum, gülüm, alım, gəlim, oturum, gülüm	mənsubiyyət, I şəxs, tək/əmr şəkli, I şəxs, tək	B000
8. 13	-aq, -ik alaq, gələk	əmr şəkli, I şəxs, cəm/arzu şəkli, I şəxs, cəm	AЖИЦ0
8. 14	-sı, -si, -su, -sü atası, əmisi, qonşusu, sürüsü	mənsubiyyət, III şəxs, tək	R JIL 0
8. 15	-mı, -mi, -mu, -mü alırımı, gəlirmi, dururmu, görürmü	sual	Ə000
8. 16	-la, -lə atamla, əmimlə	qoşma	P000
8. 17	-k1, -ki, -ku, -kü Bakıdakı, bizdəki, onunku, özünkü	mənsubiyyət	R000
8. 18	-na, -nə qapına, evinə	yönlük hal, mənsubiyyət, II/III şəxs	ИЦ000
8. 19	-ıl, -il, -ul, -ül baxılmaq, dərilmək, qurulmaq, görülmək	məchul növ	<000
8. 20	-ış, -iş, -uş, -üş	qarşılıq növ	>000

	yazışmaq, girişmək, vuruşmaq, döyüşmək		
8. 21	-ıq, -ik, -uq, -ük adamıq, əsirik, qohumuq, gülük, baxırıq, gəzirik, dururuq, görürük	xəbərlilik, I şəxs, tək	JAЖ0
9. 0	-a, -ə yaza, gələ	arzu şəkli/yönlük hal	2000
9. 1	-m qapım, sürüm	mənsubiyyət, I şəxs, tək	RALφ
9. 2	-ı, -i, -u, -ü kitabı, dəftəri, odunu, gülü	təsirlilik hal/mənsubiyyət, III şəxs, tək	Я000
9. 3	-q, -k aldıq, gəldik	I şəxs, tək	AЖ00
9. 4	-n qapın, dərin	mənsubiyyət, I şəxs, tək	+L00
9. 5	-ş mələşmək	qarşılıq növ	>000

Qeyd etmək lazımdır ki, cədvəlin dördüncü sütununda verilmiş kod sistemi şərtidir. Proqramlaşdırma prosesində mətnin emalı sistemlərində həmin kodlar daha uyğun və məqbul işarələrlə əvəz oluna bilər. Cədvəldə hər bir qrammatik şəkilçi üçün dörd əlamətin göstərilməsi nəzərdə tutulmuşdur. Bu əlamətlər şəkilçini qrammatik cəhətdən səciyyələndirən ən mühüm əlamətlər hesab olunur. Əgər şəkilçinin qrammatik əlamətləri dördədən azdırsa, həmin əlamət qeyd olunur, qalan boş qalmış əlamətlərin yerinə isə 0 (sıfır) və ya sıfırlar yazılır.

Korpusda lemmatizasiya söz-formanın lüğət vahidi formasına gətirilməsi ilə səciyyələnir. Yəni, sözdəyişdirici şə-

kilçilərin kök və ya əsasdan ayrılması, sonradan hər bir şəkilçinin şəkilçilər zəncirindən ayrılması və onların qrammatik əlamətlərinin müəyyənləşdirilməsi əslində olduqca mürəkkəb bir prosesdir. Zəngin şəkilçilər sistemi ilə digər dil ailələrindən fərqlənən türk dilləri üçün bu proses daha çətin hesab olunur. Deməli, aparıcı dünya dilləri üçün hazırlanmış korpuslarda ayrı-ayrı nitq hissələri və qrammatik kateqoriyaların kodlaşdırılması daha önəmlidir.

Qeyd etmək lazımdır ki, korpusların hazırlanması prosesində nitq hissələri və qrammatik kateqoriyaların vahid prinsiplər üzrə kodlaşdırılması olduqca önəmli hesab olunur. Respublikamızda korpusların hazırlanması işləri ilə məşğul olan bütün yaradıcı kollektivlərin vahid kodlaşdırma sistemindən istifadə etmələri vacibdir. Bu, gələcəkdə həmin korpusların birləşdirilməsi və onlardan daha səmərəli istifadə olunmasına imkan yaratmış olardı. Bu və ya digər linqvistik məsələnin həlli ilə əlaqədar konkret alqoritmin tərtibi və tətbiqi zamanı vahid kodlaşdırma sisteminin mövcudluğu bir çox ziddiyyətlərin və uyğunsuzluqların aradan qaldırılmasına xidmət edir. Şübhəsiz, istənilən kod sistemi şərtidir və tədqiqatın müəyyən pilləsində təkmilləşdirilə və yenisi ilə əvəz oluna bilər. Lakin kod sisteminin qismən təzələnməsi və yenisi ilə əvəz olunması zəruri hallarda həyata keçirilməlidir və bu barədə operativ olaraq həmin kodlardan istifadə edən mütəxəssislərə məlumat verilməlidir. Bu, hər şeydən əvvəl kodların universallığı və vahid prinsiplərlə tənzimlənməsi ilə bağlıdır. Yiğcamlıq və konkretlik üçün nitq hissələrinin, ayrı-ayrı qrammatik kateqoriyaların, leksik və leksik-qrammatik informasiyaların kodları qruplar üzrə cədvəllər şəklində verilmişdir [Mahmudov, 2013, s.216-220].

Korpusların tərtibi prosesində ilkin zəruri şərt mükəmməl linqvistik bazanın mövcudluğudur. Xüsusilə, avtomatik işarələndirmədə linqvistik bazada verilənlər kodlaşdırmanın,

lingvistik işarələndirmənin əsasını təşkil edir. Lingvistik bazada yerləşdirilmiş qrammatik informasiya sistemə müxtəlif üsullarla, əvvəlcədən və ya sonradan hissə-hissə, sınaqdan keçirilməklə və təkmilləşdirilməklə daxil edilir. Kodlar qrammatik informasiyanın aid olduğu materialın qarşısında və ya cədvəl formasında verilir. Ənənəvi qrammatikalardakı qayda və qanunlar alqoritmlərlə ifadə olunur. Həmin alqoritmlər proqram təminatının əsasına qoyula bilər.

Lemmatizasiyanı bir qədər geniş mənada götürsək, mətn üzərində iş bütün dil səviyyələrində aparıla bilər. Məsələn, mətn fonemlərə, morfemlərə, kökə, əsasa, söz-formalara və s. ayrıla bilər. Tədqiqatçının qarşıya qoyduğu konkret məqsəddən asılı olaraq mətnin bu sadalanan hissələrə bölünməsi müxtəlif üsullarla aparıla bilər. Bunlardan əlavə, mətn eyni zamanda söz birləşmələrinə, cümlələrə, abzaslara və s. hissələrə ayrıla bilər. Bu proseslər də əl üsulu ilə və formal prinsiplər əsasında avtomatik aparıla bilər.

Söz-formaların şəkildə ayrılması prosesində eyni bir hal yarana bilər ki, informasiya bir elementar məlumatdan təşkil olunmuş olsun. Məsələn, yuxarıdakı cədvəllərdə də qeyd olunduğu kimi, -ın (-in, -un, -ün) şəkilçisi aşağıdakı qrammatik səciyyələrə malikdir: “yiyəlik hal” və ya “mənsubiyyət”, II şəxs, “tək” və ya “əmr” şəkli, II şəxs, “tək”. Bu şəkilçi üçün elementar qrammatik əlamətlər aşağıdakılardır: “yiyəlik hal”, “mənsubiyyət”, “tək”, “əmr şəkli”, “II şəxs”. Tam qrammatik səciyyə beş məlumatdan ibarətdir. Semantik blokun “müdaxiləsi” prosesində bu qrammatik məlumatların təhlil olunan mətnə (kontekstə) uyğun gələn variantı müəyyənləşdirilir.

Məlum olduğu kimi, lemmatizasiya prosesində ilk mərhələ mətnin təhlil vahidlərinə ayrılması hesab olunur. Fərz edək ki, “n” sayda mətnlər çoxluğu verilmişdir. Tələb olunur ki, həmin çoxluq kompüterdə təhlil olunsun və zəruri qrammatik informasiya əldə olunsun. Hər şeydən əvvəl təhlilin

həcmi müəyyənləşdirilməlidir. Kompüterin yaddaş tutumundan və tədqiqatda qarşıya qoyulmuş vəzifələrdən asılı olaraq mətn müxtəlifşəciyyəli ola bilər. Tədqiqat prosesində mətn kimi kitab, fəsil, səhifə, abzas, cümlə və s. götürülə bilər. Bir mətni digərindən ayırmaq üçün müxtəlif üsullardan (sayğacdən, mətnin əvvəlinə və sonuna qoyulmuş işarədən və s.) istifadə oluna bilər. Əgər mətn ardıcıl olaraq daxil edilirsə, qabaqcadan mənin başlanğıcı və sonu işarəsi qoyulmamışsa, sayğacdən istifadə etməklə bərabər ölçüdə mətnlər (100 və ya 1000 söz-formadan ibarət) ayırmaq olar. Məsələn, şərti olaraq qəbul etmək olar ki, hər 1000 söz-forma ayrılıqda bir mətndir.

Lemmatizasiya prosesində 1000 söz-formadan ibarət mətn (lemmatizasiya olunacaq mətn vahidi) ayrılan kimi proses dayanır və lemmatizasiya (mətnin konkret vahidlərə ayrılması və qrammatik səciyyələrin müəyyənləşdirilməsi) başlanır. Bu ayrılmış hissədə lemmatizasiya aparıldıqdan sonra prosesə cəlb olunmalı yeni mətnin olub-olmaması müəyyənləşdirilir. Əgər üzərində lemmatizasiya aparılacaq mətn (və ya mətnlər) varsa, yeni mətn vahidi (fərz edək ki, yenə 1000 söz-formadan ibarət) tədqiqata cəlb olunur. Qeyd etmək lazımdır ki, mətnlər çoxluğundan təhlil mətnlərinin ayrılması elə bir prinsipial çətinlik törətmir və sırf texniki məsələ hesab olunur. Lemmatizasiya prosesində morfoloji təhlil alqoritmlərindən də istifadə olunur [Mahmudov, 2002, s.127-134].

Lemmatizasiya prosesində morfoloji təhlilin mərhələləri aşağıdakılardır:

mətnin mətnlərdən, cümlənin mətndən və söz-formanın cümlədən ayrılması;

- söz-formanın əsas və şəkilçilərə bölünməsi, şəkilçilərin bir-birindən ayrılması;
- ayrılmış morfemlərin (kök, əsas, şəkilçi) morfemlər (əsaslar və şəkilçilər) siyahısı ilə tutuşdurulması və onların leksik-qrammatik əlamətlərinin müəyyənləşdirilməsi;

- əsas və şəkilçilərin omonimliyinin morfoloji səviyyədə aradan qaldırılması;
- girişdəki söz-formaya müvafiq leksik-qrammatik informasiyanın ayrılması.

Azərbaycan dilində cümlənin formal əlamətlərə görə təsnifi və təhlili. Təbii dillərin emalı sistemləri (Natural Language Processing – NLP) hal-hazırda süni intellekt elminin ən qabaqcıl və sürətlə inkişaf edən sahələrindəndir. Kompüterlərin insan fəaliyyətinin bütün sahələrinə getdikcə daha çox nüfuz etməsi, kompüterlərlə daha təbii və insan üçün daha asan üsullarla işləmək üçün prinsipcə yeni proqramların yaradılması zərurətini ortaya qoymuşdur. Belə kompüter proqramları sırasına ənənəvi olaraq maşın tərcüməsi, nitqin yazıya çevrilməsi, mətnlərin insan səsi ilə səsləndirilməsi, morfolojiyanın və sintaksisin yoxlanması və s. tipli geniş yayılmış proqramlar aid edilə bilər. Lakin, bütün dünya dilləri üçün yaradılmış və sınaqdan keçirilmiş bu sistemlərlə yanaşı, hazırda daha müasir olan nitqin başa düşülməsi və adekvat cavab verilməsi, interaktiv sual-cavab sistemləri sahəsində də intensiv tədqiqatlar aparılmaqdadır. Bu sistemlər bütövlükdə süni intellektin yaradılması prosesinin tərkib hissələri kimi qəbul olunur.

İnsan beyində gedən intellektual proseslərin mahiyyətini hələ də tam dərk edə bilmədiyimizdən insanın düşüncə prosesini avtomatlaşdırmağa imkan verən proqram-texniki vasitələr yaratmaq indinin özündə də həll olunmamış problem olaraq qalır. Buna baxmayaraq, insanın işini qismən yüngülləşdirən və sürətləndirən proqram vasitələri yaratmaq artıq adi hala çevrilmişdir və müasir insanın həyat və fəaliyyətini belə kompüter sistemləri olmadan təsəvvür etmək çətindir. Mətnin emalı sistemləri dilin bütün səviyyələrini əhatə edir. Monoqrafiyanın III fəslində bu məsələlər əhatə olunduğundan bu fəsildə sırf formal linqvistik təhlillə bağlı məsələlər daha

ətraflı və dərinlən araşdırılacaq. Bütün dil səviyyələrində formal linqvistik təhlil mətnin emalı sistemlərinin əsasını təşkil edir. Odur ki, bu məsələlərin tədqiqi NLP sistemlərinin də keyfiyyət baxımından daha işlək və səmərəli fəaliyyətini şərtləndirən vasitələrdən hesab olunur.

Azərbaycan dilində sadə cümlənin formal təhlili. Azərbaycan dilində cümlənin bitmiş fikir ifadə edən sintaktik vahid kimi formalaşmasında üç mühüm amil – predikativlik, modallıq və bitkinlik intonasiyası mühüm rol oynayır [Abdullayev, 1974; Abdullayev, Seyidov, Həsənov, 1972; Kazımov, 2010].

Cümlənin bu sadalanan əlamətləri sırasında predikativlik ən mühümü hesab oluna bilər. Cümlənin xəbəri zaman, şəxs şəkilçilərinin və modallığın köməyi ilə predikativliyin yaranmasına və formalaşmasına şərait yaradır. Dilçilikdə predikativliklə bağlı belə bir fikir formalaşmış ki, əslində cümləni təşkil edən hər bir sözdə predikativlik olub, sonralar xəbər cümlədə predikativliyin əsas ifadəçisinə çevrilmişdir. Beləliklə, xəbər cümlədə predikativliyin mərkəzi fiquru olmaqla, monopredikativliyin yaranmasını şərtləndirmişdir [Abdullayev, 1998; Kazımov, 2010].

Predikativlik cümlənin mahiyyətində öz əksini tapır və bütün cümlə növlərində bu və ya digər şəkildə təzahür edir. Bütün cümlə konstruksiyalarında, təktərkipli cümlələrdə, hətta üzvlənməyən cümlələrdə belə predikativlik və predikativ münasibət olur. Lakin, onların təzahür formaları və dərəcələri fərqli olur [Kazımov, Sintaksis, 2010, s.83].

Predikativliyin ifadə vasitələri – feilin forma, zaman, şəxs şəkilçiləri, ədatlar, intonasiya və modal sözlərdir. Predikativliyin cümlənin ən mühüm əlaməti və funksiyası olan kommunikativliyin əsasını təşkil etməsi xüsusiyyəti əsas götürülərək cümlənin təsnifi və təhlilində predikativlik əsas meyar kimi seçilmişdir.

Predikativlik meyarına görə Azərbaycan dili cümlələrini aşağıdakı kimi təsnif etmək olar:

formal olaraq predikatı olmayan və ya müşahidəsi mümkün olmayan cümlələr (adlıq cümlələri, söz-cümlələri və vokativ cümlələr).

Adlıq cümlələrdə əşya və hadisənin adı çəkilir, lakin onun haqqında əlavə məlumat verilmir, cümlənin xəbəri olmadığı üçün hərəkətin icrası barədə heç nə deyilmir, əşya və hadisənin mövcudluğu bildirilir. Söz-cümlələr bir sözdən və ya sözlərin sabit birləşməsindən ibarət olub sintaktik cəhətdən üzvlənmir: – Bəli. Xeyr. Baş üstə.

Əslində, bu sözlər deyilmiş müəyyən fikrə və ya sorğuya cavabdır və burada predikat olmasa da təsəvvür olunur.

Vokativ cümlələrdə də predikat olmur və belə cümlələri bəzən xitab-cümlələr də adlandırırlar [Kazımov, Sintaksis, 2010, s.214].

Predikatı olmayan və ya müşahidəsi mümkün olmayan bu tipli cümlələrdə (adlıq cümlə, söz-cümlə və vokativ cümlə) işlənən sözlər (sintaktik vahidlər) özündə predikativliyi ehtiva edir və olmayan predikat funksiyasında çıxış edir.

yalnız bir predikatdan ibarət olan cümlələr. Belə cümlələrə sadə cümlələri aid etmək olar (Maşın dayandı).

Burada predikat cümlənin xəbəri ilə üst-üstə düşür. Sadə cümlənin xəbəri ismi və feili ola bilər. Feili xəbər hərəkət, ismi xəbər isə əlamət bildirir (Əhməd gəldi feili xəbər; Əhməd tələbədir – ismi xəbər).

Cümlədə predikatların sayı iki və ya ikidən çoxdur. Predikatlarının sayı iki və ya ikidən çox olan sintaktik vahidlər mürəkkəb cümlə kateqoriyasına – tabesiz və tabeli mürəkkəb cümlələrə aid edilir.

Məlum olduğu kimi, həmciins xəbərlili sadə cümlə növündə də xəbərlərin, eləcə də predikatların sayı ikidən çox ola bilər. Əsas meyar kimi predikatların sayı götürüldüyü üçün bu

tipli cümlələrin tabesiz cümlə kateqoriyasına aid edilməsi məqsədəuyğun hesab olunur.

Formal baxımdan bizim tədqiqatımızda mürəkkəb cümlə (tabesiz və ya tabeli) sadə cümlələrdən ibarət mürəkkəb struktur vahiddir. Biz həmin mürəkkəb strukturu bütövlükdə deyil, sadə cümlələrə ayırır və sadə cümlə kimi təhlil edirik. Deməli, istər tabeli, istərsə də tabesiz mürəkkəb cümlələr bizim tədqiqatımızda iki və daha artıq sadə cümlədən ibarət sintaktik vahiddir. Lakin, nəzərdə tutmaq lazımdır ki, tabesiz mürəkkəb cümləni təşkil edən sadə cümlələrlə tabeli mürəkkəb cümləni təşkil edən sadə cümlələr qrammatik əlaqələrinə və semantikasına görə müxtəlifdir. Onlar qrammatik baxımdan bərabərhüquqlu olsalar da, müəyyən məna əlaqələri ilə (birləşdirmə, səbəb-nəticə, aydınlaşdırma, qarşılaşdırma, bölüşdürmə, qoşulma) bağlanırlar. Eləcə də, tabeli mürəkkəb cümlələr də müəyyən tabelilik əlaqəsi ilə birləşib vahid bir tam əmələ gətirirlər. Tabeli mürəkkəb cümlələrdə komponentlərdən biri müstəqil (baş cümlə), digəri isə asılı (budaq cümlə) hesab olunur. Deməli, istər tabesiz, istərsə də tabeli mürəkkəb cümləni təşkil edən sadə cümlələr tam müstəqil hesab oluna bilməz. Onlar tərkiblərində olduqları mürəkkəb cümlənin semantikasını müəyyən qədər ehtiva etmiş olurlar. Təhlil elə aparılmalıdır ki, məsələn, xəbər budaq cümləsini iki sadə cümləyə ayırıb təhlil etdikdən sonra sintez etsək (ilkin formanı bərpa etsək), sintez nəticəsində alınmış cümlə xəbər budaq cümləsi olaraq qalmalıdır (və ya xəbər cümləsi təsiri bağışlamalıdır) [Fətullayev, Mahmudov, Fətullayev, 2015].

Beləliklə, Azərbaycan dilində cümlənin formal əlamətlərə görə təhlili prosesini ümumi şəkildə belə təsvir etmək olar: əgər cümlədə predikat yoxdursa (adlıq cümlə, söz-cümlə, vokativ cümlələr), həmin cümləni təşkil edən söz (sözlər) predikatlıq məzmunu ifadə edən sintaktik vahid (predikat) funksiyasında hesab olunur və cümlə birpredikatlı sadə cümlə

kimi təhlil olunur. Əgər cümlədə predikatların sayı birdən çoxdursa, onda cümlə mürəkkəb cümlə (tabesiz, tabeli, həmcins üzvlü sadə cümlə) hesab olunur. Əgər biz qabaqcadan cümlənin, məsələn, xəbər budaq cümləsi olduğunu biliriksə, tərcümənin daha dəqiq və adekvat olması üçün tərcümə olunacaq dildə də cümlənin həmin qrammatik anlam və statusda olmasına cəhd göstərməliyik. Fikrimizcə, bu cəhətin gözlənilməsi tərcümənin keyfiyyətinə ciddi təsir göstərən amildir. Bu məsələ ilə bağlı M.N.Əsədovanın fikirləri maraqlıdır. C.D.Selincerin “The Catcher in the Rye” (“Çovdarlıqda uçurumdan qorunan”) əsərinin Azərbaycan dilinə tərcüməsi ilə bağlı statistik hesablamalar aparan tədqiqatçı aşağıdakı nəticəyə gəlmişdir: tərcümə prosesində orijinaldakı 698 tamamlıq budaq cümləsinin Azərbaycan dilinə 340-ı tamamlıq budaq cümləsi, 259-u sadə genişləndirilmiş cümlə, 44-ü mübtədə budaq cümləsi, 12-si tabesiz mürəkkəb cümlə və s. kimi tərcümə olunmuşdur.

M.F.Axundovun “Aldanmış kəvakib” əsərindəki 137 tamamlıq budaq cümləsinin 33-ü sadə genişləndirilmiş cümlə, 32-si tamamlıq budaq cümləsi, 8-i təyin budaq cümləsi və s. kimi verilmişdir [Əsədova, 2004].

Türk dillərində cümlənin formalaşması meyarları və struktur quruluşu digər dil ailələrindən (məsələn, Hind-Avropa dillərindən) köklü surətdə fərqlənir. Yəni, digər dil ailələrində cümlənin formal əlamətlərə görə təhlili alqoritmlərini eyni ilə və ya qismən türk dillərinə, o cümlədən Azərbaycan dilinə şamil etmək qeyri-mümkündür.

Formal baxımdan cümlə iki nöqtə, nöqtə ilə sual işarəsi, nöqtə ilə nida işarəsi, nöqtə ilə nöqtəli vergül, nöqtə ilə üç nöqtə arasındakı söz-forma və ya söz-formalar çoxluğu. Deməli, cümlə bir və ya bir neçə (nəzəri olaraq “n” sayda) söz-formadan ibarət ola bilər.

Söz-forma bir probel (boşluq) işarəsi ilə digər probel (boşluq) işarəsi arasında olan məsafədə olan hərflər çoxluğudur.

Deməli, formal təhlil aparmaq üçün bizim hər şeydən əvvəl söz-formaları cümlədən, cümlələri mətndən düzgün ayırıb linqvistik baxımdan təhlil etməyə qabil linqvistik bazamız və proqram təminatımız olmalıdır. Təhlilin hərtərəfli və məqbul olması linqvistik bazanın və proqram təminatının səviyyəsi və keyfiyyəti ilə ölçülür.

Azərbaycan dilində tabeli mürəkkəb cümlələrin formal meyarlara görə təsnifi və təhlili. Son illər informasiya texnologiyalarının insan fəaliyyətinin müxtəlif sahələrinə dərinlən daxil olması, bütün dünyada NLP (Natural Language Processing – təbii dillərin emalı) sistemlərinin geniş şəkildə inkişafına gətirib çıxarmışdır. Hal-hazırda NLP sistemləri dedikdə maşın tərcüməsi, orfoqrafiyanın və qrammatikanın yoxlanması, nitqin yazıya çevrilməsi və başa düşülməsi, sorğulara avtomatik cavab verilməsi, mətn üzrə insan nitqinin generasiyası və s. sistemlər başa düşülür ki, bu sistemlər də insanla kompüter arasında daha təbii interfeys yaradılmasına, insanın kompüterlərdən səmərəli şəkildə yararlanmasına və kompüterlərdən daha geniş istifadə edilməsinə şərait yaradır.

Məlum olduğu kimi, NLP sistemlərinin yaradılması dilin elementlərinin – xüsusilə söz-formaların və cümlələrin formal (kompüter vasitəsilə) təhlili sistemlərinin yaradılmasını tələb edir. NLP sistemlərinin keyfiyyəti formal təhlil (morfoloji, sintaktik, semantik və s.) alqoritmlərinin keyfiyyətindən ciddi surətdə asılıdır.

Azərbaycan dilində formal linqvistik təhlilin ilk mərhələsi morfoloji səviyyədə təhlil hesab olunur. Morfoloji təhlil ilkin pillə kimi vacibdir və formal linqvistik təhlilin sonrakı mərhələləri üçün zəruri olan linqvistik məlumatın toplanması və sistemləşdirilməsinə imkan yaradır. Sintaktik və semantik təhlil üçün zəruri olan linqvistik məlumatlar bazasının yaradıl-

ması morfoloji təhlil sayəsində mümkün olur. Morfoloji təhlil haqqında daha geniş məlumat bu sahə ilə bağlı tədqiqatlarda verilmişdir [Mahmudov, 2002].

Formal sintaktik təhlil mövzusunda Azərbaycanda barmaqda sayılası qədər tədqiqat işlərinə rast gəlmək olar. Formal sintaktik təhlildə əsas məqsədlərdən biri Azərbaycan dilinin formal sintaktik analizatorunun yaradılmasıdır ki, bu analizatorun yaradılması Azərbaycan dili üçün NLP sistemlərinin keyfiyyətinin artırılması üçün istifadə ediləcəkdir. Qeyd etmək lazımdır ki, sintaktik analizatorun ilkin versiyası artıq yaradılmışdır və hazırda istifadə edilməkdədir.

İstənilən dil üçün yaradılmış formal sintaktik analizator nisbətən sadə (həm cümləyə daxil olan söz-formaların sayının azlığı, həm də qrammatik baxımdan) cümlələr üçün daha keyfiyyətli analiz aparır. Bu baxımdan mürəkkəb (tabeli və ya tabelsiz) cümlələrin sadə cümlələrə bölünərək sintaktik təhlil olunması daha məqsədəuyğundur. Qeyd edək ki, mürəkkəb cümlənin sadə cümlələrə bölünərək təhlil olunması yeni ideya deyildir. Monoqrafiyada tabeli mürəkkəb cümlənin formal sintaktik təhlilinə dair daha bir yanaşmanı təsvir edəcəyik. Fikrimizcə, bu yanaşma formal NLP sistemləri yaradıcılarına sintaktik təhlil prosesini avtomatlaşdırmaqda dəstək olacaqdır.

Cümlənin formal əlamətlərinin müəyyənləşdirilməsi, mürəkkəb cümlənin növlərinin təyini, tabeli mürəkkəb cümlədə budaq cümlə konstruksiyalarının diferensiasiyası vacib məsələdir.

Cümlənin uzunluğu (cümlənin uzunluğu dedikdə, cümlədə olan söz-formaların sayı nəzərdə tutulur) artdıqca onun sintaktik təhlili də mürəkkəbləşir. Belə ki, cümlə üzvlərinin sayı dəyişməz olduğundan cümlədə söz-formaların sayı artdıqca hər bir cümlə üzvünün uzunluğu (cümlə üzvünün təşkil olunduğu söz-formaların sayı) da orta hesabla artır. Bu da öz növbəsində təhlil prosesini çətinləşdirir. Dilmanc layihəsi

çərçivəsində yaradılmış Azərbaycan dili korpusu (bu korpus 61 milyondan bir qədər çox cümlədən ibarətdir) üzərində apardığımız tədqiqatlara görə müasir Azərbaycan dilində cümlələrin uzunluğu 1 ilə 100 arasında dəyişir (korpusa uzunluğu 100-dən yuxarı olan cümlələr daxil edilməmişdir).

Cədvəl 4. Azərbaycan dilində cümlələrin uzunluğa görə paylanması:

Sıra №-si	Cümlənin uzunluğu (söz-forma sayı ilə)	Korpusda rast gəlmə sayı	Faizlə
1	1-10	30 702 822	50, 330%
2	11-20	21 680 233	35, 539%
3	21-30	6 246 797	10, 240%
4	31-40	1 665 406	2, 730%
5	41-50	705 980	1, 157%
6	51-100	2 153	0, 004%
	Cəmi	61 003 391	100%

Göründüyü kimi, yazıda uzunluğu 10-dan yuxarı cümlələr kifayət qədər geniş istifadə olunur (50%) və şübhəsiz ki, bu cümlələrin böyük bir hissəsini mürəkkəb cümlələr təşkil edir. Ona görə də mürəkkəb cümlələrin sadə cümlələrə formal olaraq bölünməsi alqoritmlərinin hazırlanması Təbii Dilin Emalı sistemlərinin vacib komponentlərindən biridir. Məsələn, Azərbaycan dilindən xarici dilə tərcümə edən maşın tərcüməsi sistemində daxil edilmiş cümlənin keyfiyyətli emal edilməsi üçün onun sadə cümlələrə ayrılması, hər sadə cümlədə isə cümlə üzvlərinin müəyyən edilməsi tərcümə prosesinin daha keyfiyyətli aparılması işinə kömək edir. Avtomatik sintez edilən xarici cümlədə bağlayıcıların, feil şəkilçilərinin, söz sırasının və s. düzgün seçilməsi üçün mürəkkəb cümlənin sadə

cümlələrə ayrılması ilə yanaşı, budaq cümlələrin növü də müəyyən edilməlidir [Fətullayev, Mahmudov, Fətullayev, Dərvişov, 2015]:

- mətnin səsləndirilməsi sistemlərində cümləni səsləndirərkən onun ümumi intonasiyası, cümlədaxili pauzaların qoyulması üçün cümlənin tipinin müəyyən edilməsi vacibdir;

- təbii dilin başa düşülməsi sistemləri (NLU – Natural Language Understanding) istifadəçinin şifahi əmrlərini kompüterin başa düşəcəyi əmrlər siyahısına çevirən alqoritmlər əsasında işləyir. Bu sistemlər şifahi nitqi yazılı cümlələrə çevirdikdən sonra, cümlədən niyyət və semantik slotlar adlanan informasiyanı ayıra bilməlidirlər. Məsələn, “elə avtomobillərin siyahısını göstər ki, buraxılış ili 2005-dən yuxarı olsun”. Bu cümlədə “niyyət = siyahının göstərilməsi”, semantik slotlar isə “tip = avtomobil”, “buraxılış ili > 2005” olmalıdır. Daxil edilən cümlə, məsələn, tabeli mürəkkəb cümlə olduqda əsas informasiya, yəni niyyət baş cümlədə, onun detalları isə budaq cümlədə yerləşdiyi üçün, tabeli mürəkkəb cümlənin tipi müəyyən olunmalıdır.

Formal təhlil sistemi cümləni və onu təşkil edən söz-formaları ayırmağa (cümləni mətndən, söz-formaları cümlədən) və linqvistik baxımdan formal təhlil etməyə imkan verir. Formal təhlilin ilk pillələrində söz-formalar lüğət formasına gətirilir və onun qrammatik statusu müəyyənləşdirilir. Əldə olunmuş informasiya bütövlükdə cümlənin sintaktik təhlili üçün zəruri məlumat kimi istifadə olunur.

Formal sintaktik təhlil sadə cümlə üçün daha keyfiyyətli aparıla bilər. Əgər sadə cümlənin tam sintaktik təhlili mümkündürsə, onda daha iri (geniş) sintaktik vahidləri də təhlil etmək olar. Başqa sözlə, sadə cümlənin təhlil üsulları mürəkkəb cümləyə də tətbiq oluna bilər. Mürəkkəb cümləni onu təşkil edən sadə cümlələrə ayırmaqla hər bir sadə cümləni həmin qayda ilə ayrı-ayrılıqda təhlil etmək mümkündür. Nəzərə almaq lazımdır

ki, mürəkkəb cümlə (tabesiz və ya tabeli) sadə cümlələrin birləşməsi yolu ilə düzəlir. Sadə cümlələr müəyyən bağlayıcıların (tabesizlik və ya tabelilik) və qəlib sözlərin köməyi ilə mürəkkəb cümlə əmələ gətirir. Tabeli mürəkkəb cümlələrin növlərinin müəyyənləşdirilməsi prosesində həmin bağlayıcı və qəlib sözlər fərqləndirici əlamət kimi götürülür [Fətullayev, Mahmudov, Fətullayev, Dərvişov, 2015].

Tabesiz mürəkkəb cümlələrin sadə cümlələrə bölünməsi alqoritmləri artıq müəyyən dərəcədə araşdırıldığından biz burada əsas diqqəti tabeli mürəkkəb cümlənin sadə cümlələrə bölünməsi məsələsinin həllinə yönəldəcəyik.

Məlum olduğu kimi, cümlənin sintaktik vahid kimi formalaşmasında üç mühüm amil – predikativlik, modallıq və intonasiya mühüm rol oynayır.

Formal təhlil baxımından predikativlik aydın, müşahidəsi mümkün olan, müəyyən əlaməti olan kateqoriyadır. Bu sadalanan xüsusiyyətləri modallıq və intonasiyaya aid etmək olmaz. Modallıq, danışanın söylədiyi hadisəyə, ifadə etdiyi fikrə olan münasibəti bildirir. Predikativlik cümlənin kommunikativliyini, modallıq isə emosional-ekspressiv funksiyasını ifadə edir. Predikativlikdən fərqli olaraq modallıq formal olaraq müəyyənləşdirilə bilmir. Çox vaxt intonasiyanın və modal sözlərin köməyi ilə cümlədə modallıq məzmunu yaradıla bilər. Bu baxımdan formal meyarlara əsaslanmaqla modallığı müəyyənləşdirmək, qeydə almaq və formal təhlil sistemində daxil etmək mümkün deyil.

Modallıq kimi intonasiyanı da formal cəhətdən bu və ya digər əlamət və ya semantik xüsusiyyətinə görə ayırmaq mümkün deyil. Qeyd etmək lazımdır ki, intonasiya daha çox şifahi nitqə xas olan kateqoriyadır və fonetik səviyyədə araşdırma obyekti ola bilər. Cümlənin sintaktik vahid kimi formalaşmasında müəyyən rolu olan modallıq və intonasiya formal meyarlarla təyin oluna bilmədiyi üçün formal təhlil sistemlərində nəzərə

alınmır. Bu o demək deyil ki, formal təhlil sistemlərində modallıq və intonasiya araşdırılmır, təhlil olunmur. Əvvəlki fəsillərdə fonetik təhlil, tələffüzün və vurğunun formal təsviri sahəsində aparılmış tədqiqatlar barədə məlumat verilmişdir. Şifahi, canlı danışiq, nitq sorğularını qəbul edib cavablandırان sistemlər günümüzün reallığıdır və bir çox sahələrdə uğurlu tətbiqini tapmışdır. Sintaktik səviyyədə formal təhlil sistemlərində predikativlik ön plana keçir, modallığı və intonasiyanı da ehtiva edərək cümləni formal baxımdan təhlil, təsnif və tərcümə etməyə imkan verir. Deməli, formal sintaktik təhlildə əsas yük predikativliyin üzərinə düşür.

Predikativlik cümlənin mahiyyətində öz əksini tapır və bütün cümlə növlərində bu və ya digər şəkildə təzahür edir. Tədqiqatçılara görə, bütün cümlə konstruksiyalarında, təktər-kibli cümlələrdə, hətta üzvlənməyən cümlələrdə belə predikativlik və predikativ münasibət olur. Lakin, onların təzahür formaları və dərəcələri fərqlidir.

Məhz bu cəhət nəzərə alınaraq mürəkkəb cümlənin təsnifi və təhlilində əsas meyar kimi predikativlik seçilmişdir.

Qeyd etmək lazımdır ki, tabesiz mürəkkəb cümlə və həmcins xəbərlə sadə cümlələrdən fərqli olaraq tabeli mürəkkəb cümlələrdə hər bir budaq cümlənin müəyyən fərqləndirici formal əlaməti vardır. Həmin əlamətlər əsasında hər bir budaq cümləni müəyyənləşdirmək mümkündür. Bu isə tabeli mürəkkəb cümlələrin təsnifi, təhlili və başqa dillərə tərcüməsi baxımından olduqca əhəmiyyətlidir. Bu konkret əlamətlər baş cümlədə qəlib söz funksiyasında budaq cümləni əvəz etmiş olur. Baş cümlə budaq cümləyə həmin qəlib sözlərlə, “ki” bağlayıcısı və vergüllə bağlanır.

Burada daha bir mühüm cəhət diqqətdən yayınmamalıdır: mürəkkəb cümlə ilə budaq cümlənin sərhədlərini necə müəyyənləşdirməli? Baş cümlə harada qurtarır və budaq cümlənin başlanğıcı və sonu necə təyin olunur? Formal olaraq baş cümlə

budaq cümləyə konkret qəlib sözlər, “ki” bağlayıcısı və vergüllə bağlanır. Elə həmin əlamətlərə görə də budaq cümlənin növünü müəyyənləşdirmək mümkündür. Məsələn, aşağıdakı cümləni nəzərdən keçirək:

– Arzum budur ki, müəllim olum.

Bu nümunədə “budur” işarə əvəzliyi ilə ifadə olunan qəlib söz cümlənin xəbər budaq cümləsi olduğunu göstərir. Qəlib söz, “ki” bağlayıcısı və vergül baş cümləni səciyyələndirən əlamətlərdir. Budaq cümlə (xəbər budaq cümləsi) isə “müəllim olum” hissəsidir.

Deməli, yuxarıda qeyd etdiyimiz fərqləndirici əlamətlər olduqda baş cümlə ilə budaq cümlənin sərhədinin müəyyənləşdirilməsində problem yaranmır. Problem onda yaranır ki, bu əlamətlər tam və ya qismən buraxılır. Məsələn, qarşılaşdırma budaq cümləsində baş və budaq cümlələr bir-birinə “nə qədər”, “hər nə qədər”, “hərçənd” bağlayıcı sözləri, “-sa/-sə” şəkilçiləri, “da/də”, “belə” ədatları ilə bağlana bilər:

Nə qədər (hər nə qədər) dedimsə, sözümdən dönmədi. Bu tipli cümlələrdə formal əlamətlərin bəziləri (bəzən hamısı) buraxıla bilər:

Nə qədər dedim, sözümdən dönmədi / Dedim, sözümdən dönmədi və s.

Bu qəbildən olan cümlələr, fikrimizcə, formal olaraq tabeli cümlə olmaq əlamətini itirir, tabeli mürəkkəb cümləyə çevrilmiş olur. Onları tabeli cümlə kimi təhlilə cəlb etmək olar, baxmayaraq ki, semantikasına görə bu cümləni qarşılaşdırma budaq cümləsi hesab etmək olar.

Fikrimizcə, bu ideya formal əlamətləri qismən və ya tamamilə buraxılmış bütün tabeli mürəkkəb cümlə tiplərinə aid edilə bilər. Deməli, tabeli mürəkkəb cümlənin formal əlamətləri buraxılmışsa (yoxdursa), semantika nəzərə alınmır (nəzərə alınması mümkün olmur) və həmin cümlələr ənənəvi

dilçilikdən fərqli olaraq tabesiz mürəkkəb cümlə tipinə aid edilir.

Təqdim olunan iş mətnlərinin formal təhlili sahəsində aparılan araşdırmalar çərçivəsində linqvistik yanaşma üsullarından biri kimi səciyyələndirilə bilər. Hər şeydən əvvəl, NLP sistemləri üçün sintaktik analizatorun işlənilib hazırlanması işlərində bu araşdırmalar ilkin hazırlıq, baza hesab oluna bilər.

Nəzərə almaq lazımdır ki, formal təhlil açıq sistemdir: təhlil prosesində meydana çıxan səhvlər, düzgün alınmayan nəticələr diqqətlə araşdırılır, zəruri təkmilləşdirmələr aparılır, düzəlişlər edilir və düzgün nəticə almağa imkan verən versiyalar mütəmadi olaraq sistemə daxil edilir. Məqsəd Azərbaycan dilində formal əlamətlər əsasında cümləni təhlil və təsnif etmək, alınan nəticələrin İKT sahəsində aparılan tədqiqatlara uğurlu tətbiqini təmin etməkdir.

Bu sahədə toplanmış təcrübə belə bir nəticəyə gəlməyə imkan verir ki, Azərbaycan dili üçün tərtib olunmuş sintaktik təhlil alqoritmləri digər türk dillərinə də şamil oluna bilər.

Sintaktik işarələnmə üçün zəruri əlamətlər toplusu. Dünyə təcrübəsində sintaktik-semantik səviyyələrdə işarələnmə az müşahidə olunur. Yuxarıda qeyd etdiyimiz kimi işarələnmə, bir qayda olaraq, fonetik, leksik-morfoloji səviyyələrdə aparılır. Kiçik həcmi mətn nümunələri əsasında sintaktik işarələnmə aparılması həm nəzəri, həm də təcrübi baxımdan əhəmiyyətli ola bilər. Azərbaycan dilinin milli korpusunda sintaktik səviyyədə əl ilə və avtomatik işarələnmə ilə bağlı müəyyən işlər görülmüşdür. Bu fəsildə sintaktik işarələnmə üçün zəruri əlamətlər toplusu barədə məlumat verilmiş və əl ilə sintaktik işarələnmə nümunəsi göstərilmişdir.

Söz birləşməsi və cümlədə sözlər arasında əlaqələr:

Yanaşma:

– dost ölkə, taxta qapı, yeddi bulaq, bu hadisə, gələcək gün;

Uzlaşma:

– mənim anam, sizin eviniz, bizim fikrimiz, biz gəldik, onlar getdilər;

İdarə:

– evə getmək, məktubu oxumaq, dənizdə çimmək, ondan soruşmaq;

İsmi birləşmələr:

– torpaq yol, dəmir qapı, məktəb zəngi, divar saatı, bulağın suyu, kəndin camaatı;

Feili birləşmələr:

– məktəbdə oxuyan, dərəcə gedən, yoluna baxmadan, zavodda işləyəndə, istirahət etmək, nəzər yetirmək;

Zərf birləşmələri:

– cəld qaçmaq, astadan danışmaq, xeyli sonra anlamaq, hamıdan tez dillənmək.

Cümlə:

Nəqli cümlə:

– Kənd məktəbi dağın ətəyində yerləşirdi.

Sual cümləsi:

– İclas sabah saat neçədə olacaq?

Əmr cümləsi:

– Sən onun şəxsi işlərinə qarışma!

Nida cümləsi:

– Ay camaat, köməyə gəlin!

Sadə cümlə:

– Bu il qış tez gəlmişdi.

Mübtədə – cümlədə kimin və nəyin haqqında danışıldığını bildirən, adlıq halda söz və söz birləşmələri ilə ifadə olunan, kim?, nə?, hara? suallarından birinə cavab verən baş üzvdür (Məktəb direktoru uşaqları yeni dərs ili münasibəti ilə təbrik etdi; Kənd uca dağların ətəyində yerləşirdi).

Xəbər – cümlədə mübtədəyə aid hərəkət və ya əlaməti bildirən, kimdir?, nədir?, haradır?, necədir?, neçədir?, hansıdır?

və s. suallarından birinə cavab olan baş üzvdür (Aysu hamının hörmətini qazanmış təcrübəli həkimdir; Günəş dənizdən şəhərə boylanırdı).

Tamamlıq – cümlədə üzərində iş icra olunan və ya hərəkətlə əlaqəli obyektə bildirən ikinci dərəcəli üzvdür (Uşaqlar böyük səbirsizliklə dərslərini başlanmasını, müəllimin gəlməsini gözləyirdilər; Lap kiçik yaşlarından Tunar kitabdan, dəftərdən əl çəkmirdi).

Təyin – isimlə, isimləşmiş sözlərlə ifadə olunan üzlərin əlamət, keyfiyyət, kəmiyyət və s. cəhətdən izah edən, necə?, nə cür?, hansı?, neçə?, nə qədər? suallarına cavab olan ikinci dərəcəli cümlə üzvdür (İgid və qəhrəman əsgərlərimiz torpaqlarımızı azad edəcəklər; Güllü-çiçəkli, meyvə ağacları ilə dolu həyət göz oxşayırdı).

Zərflik – cümlədə hərəkət və ya əlamətin meydana çıxma tərzini, zamanını, yerini, səbəbini, məqsədini, dərəcəsini və s. bildirir (Qış aylarında, buzlu, şaxtalı havalarda məktəbə piyada gedirdik; Səlim gülümsəyə-gülümsəyə başına gələnəri danışmağa başladı).

Əlavələr – cümlədə asılı olduğu üzvdən sonra gəlib, onu müxtəlif cəhətlərdən izah edən aydınlaşdırıcı söz və söz birləşmələrinə deyilir (Əli kişi – bütün kənd camaatının dərin hörmətini qazanmış ağsaqqal da onlara öyüd-nəsihət verdi. Samir atasını burada – heç vaxt olmadığı yerdə görəndə təəccübləndi).

Cüttərkibli cümlələr – baş üzvlərinin hər ikisinin iştirakı ilə yaranmış sadə cümlələrə deyilir (Qara Qarayev dünya şöhrətli, görkəmli Azərbaycan bəstəkarıdır). Cüttərkibli cümlələr ikinci dərəcəli üzlər hesabına genişlənə bilər.

Təktərkibli cümlələr – baş üzvlərdən birinin əsasında qurulur (Sentyabrdir. Payızdır. Hesabat vaxtıdır). Aşağıdakı növləri var:

- qeyri-müəyyən şəxslə cümlələrdə iş və ya hərəkət bir və ya bir neçə qeyri-müəyyən şəxs tərəfindən yerinə yetirilir (Mənə də Koroğlu deyirlər; Heç kəs boş yerə danışmasın);

- ümumi şəxslə cümlələrdə iş və ya hərəkət ümumən hamıya aid olur (Cücəni payızda sayarlar; Yıxılana balta çalmazlar);

- şəxssiz cümlələr – mübtədası olmayan, xəbərlə birlikdə təsəvvür olunan təktərkibli cümlələrdir (Gözəl novruz bayramı günlərindən biri idi; Axşamüstü idi);

- adlıq cümlələr – xəbəri olmayan, adlıq halda olan söz və birləşmələrlə intonasianın köməyi ilə əşya və hadisə haqqında məlumat verən təktərkibli cümlələrdir (Səhər. Küləksiz, sakit hava; Kənd evi. Yaşıl bağça. Arı pətəyi).

Bütöv cümlə. Cüttərkibli və təktərkibli cümlələrin heç bir üzvü buraxılmadıqda bütöv hesab olunur (Müəllim yeni dərs materialını çox maraqla izah edirdi).

Yarımqıq cümlə – buraxılmış üzvləri cümləni təşkil edən digər üzvlərə görə bərpa oluna bilən cümlələrə deyilir (Adın nədir? – Əli. Neçə yaşın var? – Altı).

Üzvlənməyən cümlələr:

Söz-cümlələr (bir sözdən və ya sabit söz birləşməsindən ibarət olub sintaktik cəhətdən üzvlənməyən cümlələrə deyilir (Baş üstə!; Yox; Yaxşı!, Xeyr; Əlbəttə və s.).

Vokativ cümlələr də üzvlənmir və adlıq halında olan müraciət, çağırış bildirən sözlərdən ibarət olur (Ay uşaq!, Ay camaat!; Ədə; A gədə).

Sadə cümləni genişləndirən vasitələr:

Həmcins üzvlü cümlələr – cümlədə bir üzvlə bağlı olub, bir sintaktik vəzifə daşıyan bərabərhüquqlu üzvləri olan cümlələrə deyilir. Həmcinslik həm baş, həm də ikinci dərəcəli üzvlərə aid olur (həmcins mübtədə – Əli, İslam, Sevda və Ceyran dərs əlaçılarınırlar; həmcins xəbər – Mübarizə davam etmiş, davam edir və davam edəcək; həmcins tamamlıq – Arzu

xala almaları, armudları, alçaları, xiyar və pomidorları səliqə ilə səbətlərə yığdı; həmcins təyin – İri, qara, mənəli gözləri, şirin gülüşü ilə hamını heyran eləmişdi; həmcins zərflilər: həmcins zaman zərfliləri – Hər gün, hər saat, hər dəqiqə onu düşünürdü; həmcins tərz zərfliyi – Sevinə-sevinə, gülə-gülə rəfiqəsi ilə söhbət edirdi; həmcins yer zərfliyi – Bütün rayonda, kəndlərdə, toy məclislərində, yas mərasimlərində hamı bu hadisədən danışdı; həmcins səbəb zərfliyi – Kasıblığına, imkansızlığına görə hamıdan çəkinirdi; həmcins məqsəd zərfliyi – Həyatda yüksək mövqə qazanmaq, ucalmaq, daha yaxşı yaşamaq üçün yaxşı çalışmaq, yaxşı oxumaq lazımdır və s.).

Xüsusi olmuş üzvlü cümlələr – mənə yükünü nəzərə çarpdırmaq üçün xüsusi intonasiya ilə deyilən, fasilə və vurğu ilə ayrılan üzvləri olan cümlələrə deyilir (Yay olmasına baxmayaraq, kənddə havalər çox soyuq keçirdi).

- tamamlığın xüsusi olmuşu (Səndən başqa, heç kəsə başıma gələnleri danışmamışam);
- zərfliyin xüsusi olmuşu (Bu vəzifəyə hamıdan çox layiq olmasına baxmayaraq, onu heç saya alan yox idi);
- təyinin xüsusi olmuşu (Qarsız, şaxtasız, küləksiz qış qış deyil).

Mürəkkəb cümlə:

Mürəkkəb cümlə – iki və daha artıq sadə cümlənin mənə və qrammatik cəhətdən birləşməsi nəticəsində əmələ gələn sintaktik vahiddir.

Tabesiz mürəkkəb cümlələr iki və daha artıq sadə cümlənin mənə və qrammatik cəhətdən bərabərhüquqlu vahid bir tam kimi birləşməsi yolu ilə əmələ gələn sintaktik vahiddir. Quruluş xüsusiyyətlərinə, bağlayıcıların və intonasiyanın xarakterinə görə tabesiz mürəkkəb cümlələrin tərkib hissələri arasında aşağıdakı mənə əlaqələri vardır:

- birləşdirmə əlaqəsi (Samir dinləyirdi və eşitdiklərini kağıza köçürürdü);
- səbəb-nəticə əlaqəsi (Dərslər qurtarmışdı, bütün sinif otaqları boşalmışdı);
- aydınlaşdırma əlaqəsi (Mən səndən incimirəm – həmin şəraitdə başqa çıxış yolun yox idi, sən elə etməli idin);
- qarşılaşdırma əlaqəsi (Əli çox yalvardı, lakin Fikrət fikrindən dönmədi);
- bölüşdürmə əlaqəsi (Gah çox isti olur, tərləyirsən, gah da soyuq külək əsir, soyuqlayırsan; Ya məni özünlə aparırsan, ya da mən başqasından kömək istəməli oluram);
- qoşulma əlaqəsi (Hesabata çox az vaxt qalırdı, bu isə onu narahat edirdi).

Tabeli mürəkkəb cümlə – iki və ya daha artıq cümlənin tabelilik yolu ilə birləşməsi ilə yaranan sintaktik vahiddir. Tabeli mürəkkəb cümlənin qrammatik cəhətdən müstəqil olan hissəsinə baş cümlə, baş cümlədən asılı olan, onu izah edib aydınlaşdıran hissəsinə isə budaq cümlə deyilir.

Mübtədə budaq cümləsi – baş cümlənin işarə əvəzliyi ilə ifadə olunan və ya qəlib söz şəklində təsəvvür edilən mübtədəsini izah edir (Bu da doğrudur ki, şagirdlərin biliyi çox vaxt düzgün qiymətləndirilmir; Məlumdur ki, onun bu uğurları qazanmasında çoxlarının əməyi olub).

Xəbər budaq cümləsi – baş cümlənin işarə əvəzliyi – qəlib sözlə ifadə olunan xəbərini izah edib aydınlaşdırır (Onun fikri budur ki, kənd məktəbini genişləndirsin, sinif otaqlarının sayını artırsın; Məsələ ondadır ki, başqaları da onun fikrini müdafiə edirlər).

Tamamlıq budaq cümləsi – baş cümlənin işarə əvəzliyi ilə ifadə olunan və ya qəlib söz şəklində təsəvvür edilən tamamlığı izah edib aydınlaşdırır (Onu da demək lazımdır ki, indi 30-40-cı illər deyil; Hələ də inanmaq olmur ki, belə bir hadisə baş verib);

Təyin budaq cümləsi – baş cümlədə işarə əvəzliyi ilə ifadə olunan və ya qəlib söz şəklində təsəvvür olunan təyinin mənasını izah edib aydınlaşdırır (Amma həyatda elə insanlara rast gəlmişəm ki, onların qayğısı, köməyi və himayəsi olmasaydı, mən bu uğurları qazana bilməzdim; Elə hadisələr var ki, onların mahiyyətini illər keçəndən sonra dərk edirsən).

Zərflik budaq cümlələri:

Tərz budaq cümləsi – baş cümlədəki hərəkətin icra tərzini, necə, nə şəkildə, nə vəziyyətdə icra olduğunu bildirir (Elə danışır ki, elə bil bu məsələləri ondan yaxşı bilən yoxdur; Əli elə qışqırdı ki, yanındakılar qorxuya düşdülər);

Dərəcə budaq cümləsi – baş cümlədə ifadə olunan hərəkət və ya əlamətin dərəcəsini bildirir (Onun gəlişi o qədər gözələnilməz oldu ki, hamı çaşqın vəziyyətdə qaldı; O qədər ağıllı, o qədər tədbirli, o qədər işgüzar idi ki, hamı valeh olmuşdu);

Zaman budaq cümləsi – baş cümlədəki hərəkət və ya əlamətin zamanını bildirir (O vaxt kəndə çatdılar ki, hamı yatmağa hazırlaşdı; Dərslər qurtaran kimi uşaqlar istirahətə yollandılar).

Yer budaq cümləsi – baş cümlədəki hərəkət və ya əlamətin yerini bildirir (O yerdə ki düzlük var, təmizlik var, orada inkişaf və səmimiyyət olur; Hara məsləhət görsələr, ora gedərəm);

Kəmiyyət budaq cümləsi – baş cümlədəki hərəkət və ya əlamətin kəmiyyətini bildirir (Nə qədər bacarırdımsa, o qədə də kömək edirdim; O qədər soyuq idi ki, dilimiz topuq çalırdı);

Səbəb budaq cümləsi – baş cümlədə ifadə olunan hərəkət və ya əlamətin səbəbini bildirir (Mən onu ömrüm boyu unutmaram, çünki ən çətin vaxtlarda əlimdən tutub; Ona görə demədim ki, arada inciklik yaranmasın);

Məqsəd budaq cümləsi – baş cümlədə ifadə olunan hərəkətin məqsədini bildirir (Ali məktəbdə oxumaq üçün həmin il

Bakıya gəldim; O səbəbə gəlmişəm ki, bütün məsələləri birlikdə həll edək);

Şərt budaq cümləsi – baş cümlədə ifadə olunan hərəkət və ya əlamətin şərtini bildirir (Şərtə əməl etməsən, aqibətin yaxşı olmayacaq; Belə davam edərsə, heç bir uğurdan söhbət gedə bilməz);

Qarşılaşdırma budaq cümləsində – baş və budaq cümlə qarşılaşdırılır, müqayisə edilir, fərqləndirilir (Hərçənd müəllim çox çalışdı, ancaq məsələni şagirdə başa sala bilmədi; Nə qədər gizlətməyə çalışsalar da, hamının bu işdən xəbəri oldu);

Nəticə budaq cümləsi baş cümlədən doğan nəticəni bildirir (Nə günah sahibiyəm ki, belə danlanıram; Siz haqlısınız, ona görə də Sizi mühakimə etmək olmaz);

Qoşulma budaq cümləsi baş cümləyə qoşularaq, ya baş cümlədəki fikirlə bağlı əlavə məlumat verir, yaxud həmin fikrə müəyyən münasibət bildirir (Onların dağıdıcı siyasəti ölkəyə ziyandır, bütün bunları hamıya izah edib başa salmaq lazımdır; Mənim ən çox inandığım adam o idi ki, onun da aqibətini gördünüz).

Süni intellekt problemləri çərçivəsində mətninin avtomatik işarələnməsi korpus dilçiliyinin ən çətin və mürəkkəb məsələlərindən biri hesab olunur. Məlum olduğu kimi, korpusun dili tam və əhatəli şəkildə təmsil edə bilməsi üçün onun həcmi kifayət qədər böyük olmalıdır. Yazılı dilin bütün janr və üslubları, elmi, bədii, publisistik və s. janrlarda çap olunmuş əsərlər, qəzet, jurnal, internet resursları korpus mətnlərinin əsasını təşkil etməlidir. Yazılı mətnlərlə bərabər şifahi nitq korpusları da kifayət qədər əhatəli və dili təmsil edə biləcək dərəcədə irihəcmli olmalıdır. Məhz ona görə də, kompüter texnologiyasının müasir imkanları çərçivəsində indi 50-100 milyon söz-formadan ibarət (daha böyük həcmli korpuslar da istisna olunmur) korpusların layihələndirilməsi və yaradılması işləri aparılır.

Məlum olduğu kimi, korpus həm də informasiya mənbəyi, məlumatvermə vasitəsidir. Korpusdan istənilən məlumatı ala bilmək üçün öncədən müəyyən hazırlıq işləri yerinə yetirilməlidir. Məsələn, korpusun mükəmməl və zəmanətli linqvistik məlumat verə bilməsi üçün korpus mətnində öncədən linqvistik markerlənmə, linqvistik işarələnmə aparılmalıdır. Alınacaq linqvistik informasiyanın etibarlılığı, zəmanətli olması öncədən aparılmış linqvistik işarələndirilmənin keyfiyyəti və əhatəliliyindən birbaşa asılıdır. Nisbətən kiçik həcmli korpuslarda (məsələn, 100.000 söz-formadan ibarət) linqvistik işarələndirmə nisbətən asan yolla – əl ilə işarələndirmə üsulu ilə aradan qaldırıla bilər (baxmayaraq ki, 100.000 söz-formadan ibarət mətnlər toplusunu əl ilə linqvistik məlumatla təchiz etmək bir o qədər də asan məsələ deyil).

İrihəcmli (50-100 milyon arası söz-formadan ibarət) korpusun əl ilə işarələndirilməsi həlli mümkün olmayan bir məsələyə çevrilir. Deməli, korpusların avtomatik işarələndirilməsi həlli vacib məsələ kimi qarşıya çıxır. Bunun üçün isə, dilin formal təsviri qayda və üsullarının işlənilib-hazırlanması, formal təhlil alqoritmlərinin təhlili, yeni proqramlaşdırma metodlarının layihələndirilməsi, sınaqdan keçirilməsi və praktik sistemlərə tətbiqi lazım gəlir. Azərbaycan dilçiliyində bu sahədə müəyyən işlər görülmüş, bəzi nəticələr əldə olunmuşdur [Махмудов, 1991; Mahmudov, 2002; 2013].

Formal linqvistik təhlil sahəsində Azərbaycan dilçiliyində əldə olunmuş uğurlu nəticələri korpus mətnlərinin avtomatik işarələnməsi məsələlərinə də tətbiq etmək olar. Monoqrafiyanın bu fəslində Azərbaycan mətnlərinin formal linqvistik təhlili, bu sahədə tərtib olunmuş alqoritmlər və həmin texnologiyanın korpusların avtomatik linqvistik işarələndirilməsi ilə bağlı istifadə olunması imkanları və perspektivləri barədə məlumat veriləcək.

Sintaktik səviyyədə avtomatik işarələnmə. Məlum olduğu kimi, ayrılıqda hər bir cümlə mətn daxilində həm müstəqil bir tamlıqdır, həm də eyni zamanda müəyyən bir asılılığı ifadə edir. Məhz ona görə də, cümlənin təhlili təkcə ayrılıqda bir cümlənin təhlili kimi başa düşülməməlidir, bu həm də bütövlükdə mətnin təhlili, mənimsənilməsi, qavranılması nəticəsində mümkün olur. Əgər fonemlərin tam formalaşması morfevlərdə, morfevlərin tam izahı sözlərdə baş verirsə, sözlərin də tam və bitkin mənası (istər leksik, istərsə də qrammatik) cümlədə məlum olur. Təsadüfi deyil ki, izahlı lüğətlərdə sözlərin bütün məna çalarlıqlarını vermək üçün onlar müxtəlif cümlələrdə – sintaktik əhatələrdə verilir. Müşahidələr göstərir ki, sözün həqiqi mənası yalnız kontekstdə məlum olur. Bu baxımdan cümlə də ayrılıqda tam təhlil olunub qavranılmır. Cümlənin “başa düşülməsi” üçün onu mətndə götürmək lazımdır, ondan əvvəl və sonra gələn cümlələrin informasiya tutumu və kommunikativ funksiyası cümləyə öz təsirini göstərməmiş olmur. Lakin mətnin “qavranması” cümlənin “qavranmasından” daha mürəkkəbdir, daha doğrusu, formal meyarlar burada morfoloji səviyyədə olduğu kimi böyük səmərə vermir. Mətnin formal baxımdan tam qavranılması insanın uzun illər boyu əldə etdiyi ensiklopedik məlumatın proqramlaşdırılması kimi mürəkkəb bir problemin həllini ön plana çəkir. Mətni mənimsəmə prosesində insan təkcə mətn çərçivəsində qapanıb qalmır, digər ekstralingvistik bilik və təcrübələrdən faydalanır. Sintaktik işarələnmədə nəzərə alınmalıdır ki, sintaktik təhlil təkcə cümlə konstruksiyalarının araşdırılması və sintezi üçün deyil, həm də morfoloji səviyyədə sözlərin hələ açılmamış mənalılarının tapılması, omonimiya və polisemiyanın aradan qaldırılması baxımından da faydalıdır.

Məlum olduğu kimi, formal təhlil sistemlərində morfoloji və morfonoloji blokların fəaliyyətindən sonra söz-formalar tapılır, əsas və şəkilçilər zəncirinin qrammatik əlamətləri qeydə

alınır və təhlilin sonrakı mərhələləri – sintaktik-semantik mərhələlər üçün zəruri məlumatlar əldə olunur. Avtomatik sintaktik təhlilin üzərinə cümlənin mətndən ayrılması, mübtəda və xəbərin, eləcə də digər cümlə üzvlərinin tapılması, durğu işarələrinin təhlili, söz birləşmələrinin axtarışı kimi mürəkkəb kompleksə malik problemlərin həlli düşür. Məlum olduğu kimi, mübtəda və xəbər müxtəlif nitq hissələri ilə ifadə olunur. Təhlil prosesində mübtəda və xəbərin hansı nitq hissəsi ilə ifadə olunduğu da göstərilir. Köməkçi feillərlə, feili frazeologizmlərlə ifadə olunmuş xəbərin müəyyənləşdirilməsi daha mürəkkəb bir prosesdir. Şübhəsiz, sintaktik təhlil səviyyəsində bu məsələlərin tam və qəti həllinə zəmanət verilmir. Bu problemlərin bəzilərinin həlli daha yuxarı səviyyələrdə öz həllini tapmış olur. Amma onun nəticələrindən avtomatik işarələnmədə uğurla istifadə oluna bilər.

Sintaktik işarələnmə üçün zəruri məlumatları verən təhlil prosesində mübtəda və xəbər araşdırılarkən hər şeydən əvvəl onların ifadə vasitələrini müəyyənləşdirmək lazımdır. Hər hansı bir nitq hissəsi ilə ifadə olunmuş cümlə üzvləri tapıldıqdan sonra onların morfoloji əlamətləri və sintaktik xüsusiyyətləri qeydə alınır.

Təhlil prosesində mühüm cəhətlərdən biri də cümlədə işlənmiş durğu işarələrinin təhlilidir. Bu sahədə aparılan tədqiqatlarda çox vaxt durğu işarələri nəzərə alınmır və onların araşdırılması əhəmiyyətsiz hesab olunur. Təcrübələr göstərir ki, sintaktik təhlilin mükəmməlliyi, tamlığı və dəqiqliyinin təmini durğu işarələrinin təhlilindən də asılıdır. Cümlənin sərhədlərinin müəyyənləşdirilməsi, söz birləşmələrinin aşkar olunması, cümlənin həmcins üzvlərinin tapılması durğu işarələrinin təhlili nəticəsində mümkün olur.

Xəbərin formal əlamətlərə görə təyini. Cümlənin sintaktik təhlilində xəbərin formal əlamətlərə görə təyini mühüm şərtlərdəndir. Xəbər cümləyə daxil olan cümlənin elə bir baş üzvüdür ki,

qrammatik cəhətdən yalnız mübtədadan asılı olub, mübtədanın ifadə etdiyi əşyanı geniş mənada bildirir [Abdullayev, Seyidov, Həsənov, 1972].

Azərbaycan dilində xəbər feilin dəyişən şəkilləri və ad birləşmələri ilə ifadə olunur. Ona görə də, xəbərlər ismi və feili xəbərlər kimi iki qrupa ayrılırlar. Feillər zaman, şəxs, təsirlik, növ, forma, inkarlıq və tərz kateqoriyalarına malik olmaqla daha çox predikativlik əlaqəsi yaradır və cümlədə çox vaxt xəbər olur. Azərbaycan dilində feil formalarının zənginliyi xəbərin formal əlamətlərinin də zənginliyinə gətirib çıxarır.

Göstərmək lazımdır ki, xəbər köməkçi feillər et/mək, ol/maq və feili frazeologizmlər (gözə gəlmək, ələ salmaq, dilə tutmaq, başa salmaq və s.) ilə ifadə olunur.

Feilin əmr şəkli ilə ifadə olunmuş xəbərlər. Feilin əmr şəkli ilə ifadə olunmuş xəbərlər II şəxsin təkində heç bir şəkilçi qəbul etmirlər. I və III şəxsin cəmində, eləcə də II şəxsin cəmində morfoloji üsulla ifadə olunub əmr, həmçinin xahiş və arzu olunan hərəkətin gələcək zamanda icrasını bildirirlər. İza-hın yığcamlığı üçün a, ə hərflərini v, k, g, q, ğ hərflərini x, ı, i, u, ü hərflərini isə y ilə işarə edək.

Feilin xəbər şəklinin xəbərəmələgətirmə imkanları. Mə-lum olduğu kimi, feilin xəbər şəklinin müstəqil morfoloji əla-məti yoxdur. Zaman və şəxs şəkilçiləri ilə işlənərək xəbər şəkli icra olunan və ya icra olunası işin hansı zamanda və hansı şəxs tərəfindən görüldüyünü və ya görülcəyini başqasına çatdırmağa xidmət edir. Ona görə də, xəbər şəklinin morfoloji əlamətləri kimi zaman və şəxs şəkilçiləri çıxış edir [Hüseynzadə, 1983]. Zaman şəkilçiləri feil köklərindən sonra, şəxs şəkilçilərindən əvvəl artırılır.

Feilin xəbər şəkli ilə ifadə olunmuş xəbərlərin sadəsi za-man və xəbərlilik şəkilçilərinin, mürəkkəbi isə idi, imiş hissəcik-lərinin iştirakı ilə yaranır. M.Hüseynzadə idi, imiş his-səciklərini köməkçi feil kimi səciyyələndirir [Hüseynzadə,

1973, s.160]. Bu hissəciklər dilçilik ədəbiyyatında bağlama kimi də təqdim olunur.

Feilin əmr şəkli ilə ifadə olunmuş xəbərlərin inkarı -m şəkilçisi ilə düzəlidir. Xəbər şəklinin inkarı isə şühudi keçmişdə, nəqli keçmişdə və qəti gələcək zaman formasında feil kökündən sonra inkarlıq şəkilçisinin, sonra zaman və şəxs şəkilçilərinin əlavə edilməsi ilə düzəlidir. İndiki zaman və qeyri-qəti gələcək zamanlarda -m inkar hissəciyi əlavə edilir.

Feilin arzu şəkli ilə ifadə olunan xəbərlər -v (-a/-ə/-ya/-yə) şəkilçisi ilə düzəlidir:

– Sabah gərək kitabxanaya gedəsiniz.

Feilin vacib forması ilə ifadə olunan xəbərlər – mvly (-malı/-məli) şəkilçisi ilə düzəlidir:

– Mən hər şeyi açıq deməliyəm.

Feilin lazım şəkli ilə ifadə olunmuş xəbərlər feil kökünə -vsy (-ası/-əsi/yası/yəsi) şəkilçisi artırılmaqla düzəlidir:

– Başına gələnlər danışılmalı deyilmiş.

Feilin davam şəkli ilə ifadə olunan xəbərlər -mvxdv (-maqda/-məkdə) şəkilçisi ilə əmələ gəlir:

– Hava dəyişməkdədir.

Feilin lazım şəklinin inkarı feil şəkillərindən fərqli olaraq daha çox “deyil” sözlünün köməyi ilə sintaktik yolla düzəlidir.

Məchul növü əmələ gətirən -yl (-ıl/-il/-ül/ul/-yıl/-yil/-yul/-yül) şəkilçisinin feil kökünə artırılması ilə xəbər düzəlidir:

– Rəhbərliyə bu barədə məlumat verilmişdir.

Yuxarıda göstərilənlərdən əlavə, feilin qarşılıq, qayıdış, şərt şəkilləri ilə ifadə olunmuş sadə xəbər formaları da geniş müşahidə olunur.

Qayıdış növlə düzələn xəbərlər –yn/-yl (-ın/yın –ıl/yıl – dörd variantda) şəkilçilərini qəbul edirlər:

– Məktəbli oğlan tez yuyundu.

– Gül açıldı.

Qarşılıq növ şəkilçisi (-ş/ -yş)artırılmaqla düzələn sadə xəbərlər. Qeyd etmək lazımdır ki, -ş şəkilçisi sonu saitlə bitən çoxhecalı sadə və düzəltmə feillərə, həmçinin –la/-lə şəkilçili düzəltmə feillərə artırılır. Azərbaycan dilində bir qrup feil –yş (-ış/-iş/-uş/-üş) şəkilçisi ilə düzəlib qarşılıq-birgəlik növ mənası bildirsə də tərkib hissələrinə ayrılır. Məsələ orasındadır ki, həmin feillərin kökləri ilkin mənalarnı itirmiş və müasir Azərbaycan dilində sadə feil kimi işlənilir: güləşmək, yarışmaq, barışmaq və s.

–laş/-ləş şəkilçisi də -la/-lə feil düzəldən şəkilçi ilə -ş qarşılıq –birgəlik növün şəkilçisindən ibarət olub qarşılıq-birgəlik növ əmələ gətirir: məsləhətləşmək, xudahafizləşmək və s. [Müasir Azərbaycan dili. Morfologiya, 1980, s.263].

İcbar növlü sadə xəbərlər (-dyr /-dyrt) şəkilçiləri ilə düzəliirlər:

– Evə teleqram vurdurdum.

İsmi xəbərlər. Feildən başqa digər nitq hissələri və söz birləşmələri ilə ifadə olunan xəbərlər ismi xəbərlər adlanır. İsmi xəbərlərdə xəbərlilik şəkilçisi funksiyasını -dyr şəkilçisi yerinə yetirir. Bu şəkilçidən başqa, şəxs şəkilçiləri də xəbər formaları yarada bilər. Xəbərlərin bir qismi də “var”, “yox”, “gərək” sözlərinə -dyr şəkilçisi əlavə olunmaqla düzəlir.

İsmi xəbərlərə aşağıdakıları nümunə göstərmək olar:

İsim xəbər funksiyasında:

– Aysu şagirdidir.

Əvəzlilik xəbər funksiyasında:

– Sizin fikriniz nədir?

– Mənim dostum odur.

Sifət bütün məna növləri və dərəcələri ilə xəbər funksiyasında:

– Səma mavidir.

– Dəniz gömgöydür.

Zərf xəbər funksiyasında:

– Hələ vaxta var, tezdir.

Məsdər xəbər funksiyasında:

– Mənim məqsədim işləmək və oxumaqdır.

Bundan başqa var/vardır, yox/yoxdur, deyil/deyildir, gərək/gərəkdir, deyil sözləri də xəbər funksiyasını yerinə yetirə bilirlər:

– Mənim maşınım var (vardır, yoxdur).

Deyil/ deyildir sözləri -mV şəkilçisi ilə xəbərin inkarını əmələ gətirirlər. Var/yox sözləri xəbərlik şəkilçisi qəbul etmədən də bu funksiyanı yerinə yetirə bilirlər.

– Heç nədən xəbərim yox.

İsmi xəbərlər ad nitq hissəsinə şəxs şəkilçiləri əlavə edilməklə düzəlir:

– Siz tanınmış müəllimsiniz.

– Qardaşım təcrübəli həkimdir.

Feili və ismi xəbərlərlə bağlı məlumatlar əsasında bütün mümkün xəbər formalarını və kateqoriyalarını əhatə edən formal modellər qurmaq olar. Həmin modellər hər hansı mətnə xəbər formalarının müəyyənləşdirilməsi və formal əlamətlərə görə avtomatik işarələndirilməsi imkanı yaradır. Bu məsələlər ətraflı olaraq işləndiyi üçün formal modellər və onların tətbiq dairələri üzərində dayanmağa lüzum yoxdur [Bax: Mahmudov, 2002, s.166-190].

Feili xəbər şəkilçilərini cədvəl şəklində belə ifadə etmək olar:

Cədvəl 5. Feilin xəbər şəkilçiləri

Şəxs	I	-m, -n, -am, -əm, -yam, -yəm	-q, -k, -aq, -ək, -yaq, -yək, -ıq, -uq, -ük, -ik
	II	-n, -san, -sən	-z, -nız, -niz, -nuz, -nüz, -sınız, -sunuz, -sünüz, -sınız
	III	-sın, -sun, -sün, -sin	-lar, -lər
Zaman	-ır, -ur, -ür, -ir, -yır, -yur, -yür, -yir, -dı, -di, -du, -dü, -miş, -miş, -muş, -müş, -ıb, -ib, -ub, -üb, -yıb, -yib, -yub, -yüb, -ar, -ər, -yar, -yər, -acaq, -əcək, -yacaq, -yəcək		
v	-a, -ə, -ya, -yə		
mvly	-malı, -məli		
vsy	-ası, -əsi, -yası, -yəsi		
mvxdv	-maqda, -məkdə		
yl	-ıl, -il, -ul, -ül		
yş	-ış, iş, -uş, -üş		
yn	-ın, -in, -un, -ün		
İ	-m, -ma, -mə, -maz, -məz		
U	idi, imiş, -ydi, -ymiş		
dyr/dyrt	-dır, -dir, -dur, -dür, -dırt, -dirt, -durt, -dürt		

Bu cədvəlin daha ətraflı variantı “Mətnin formal təhlili sistemi” kitabında verilmişdir [Mahmudov, 2002, s.172-175].

Mübtədanın formal əlamətlərinə görə təyini. Mübtədə cümlədə başqa üzvlərdən asılı olmur. Adlıq halda isimlərlə, əvəzliliklərlə, həmçinin substantivləşən digər nitq hissələri və söz birləşmələri ilə ifadə olunub, əlaməti (hərəkət, halvəziyyət, xassə, keyfiyyət) xəbərdə müəyyən olunan əşyanı bu

sözün geniş mənasında bildirir [Müasir Azərbaycan dili. Sintaksis., 1981, s.163].

Qeyd etmək lazımdır ki, mübtədanın formal əlamətlərinə görə müəyyənləşdirilməsi alqoritmində bir çox qayda və qanunauyğunluqların nəzərə alınması lazım gəlir. Həmin qaydaları ümumi şəkildə nəzərdən keçirək:

Mübtədə həmişə ismin adlıq halında olur:

– Məktəb evimizin yanındadır.

Lakin, bu heç də o demək deyildir ki, bütün adlıq halında olan isimlər mütləq mübtədə olmalıdırlar. Adlıq halında olan isim cümlənin başqa üzvləri yerində də ola bilər. Qeyd etmək lazımdır ki, formal cəhətdən qeyri-müəyyən təsirlik hal adlıq haldan seçilmir.

Cümlədə aralarında vergül olmayan adlıq halda işlənmiş iki isim varsa və cümlənin xəbəri feilin məchul növü ilə ifadə olunmuşsa, onda bu isimdən birincisi təyin, ikincisi mübtədə rolunda çıxış edir:

– Dəmir darvaza bağlandı.

Əgər cümlədə aralarında vergül olan adlıq halda işlənmiş iki isim varsa, onlar çox vaxt həmcins mübtədə rolunda çıxış edirlər. Bu xüsusiyyətlər durğu işarələrinin təhlili sxemləri ilə aşkar olunur:

– Aysu, Tunar məktəblidirlər.

Cümlədə adlıq halında aralarında vergül olmayan iki isim varsa və xəbər feilin məlum növü ilə ifadə olunmuşsa, onda bu iki isimdən birincisi mübtədə, ikincisi tamamlıq olur:

– Poçtalyon qəzet paylayır.

Cümlədə adlıq halında olan bir isim, bir zərf varsa və xəbər feilin məlum növü ilə ifadə olunmuşsa, onda isim mübtədə, zərf isə zərflik funksiyası daşıyır:

– Müəllim aşağı endi.

Cümlədə bir adlıq halında olan zərf varsa, o, mübtədə olur:

– Yuxarı aşağıdan istidir.

Şəxs əvəzlilikləri adlıq halında cümlədə yalnız mübtədə olur:

– O, kitabı mənə verdi.

Azərbaycan dilində feillər şəxslə olduğu üçün çox vaxt mübtədalar şəxs əvəzlilikləri ilə ifadə olunanda ixtisar oluna bilər. Belə hallarda feillə ifadə olunmuş xəbərdəki şəxs şəkilsinə əsasən mübtədanı bərpa etmək olar:

– / O/ mənə kitab verdi.

İsim, zərf və əvəzlilikdən başqa, məsdər, sifət və say da cümlədə mübtədə ola bilər:

Adlıq halında məsdər mübtədə funksiyası daşıya bilər:

– Oxumaq ən böyük arzumdur.

Əgər cümlədə eyni zamanda adlıq halında işlənən isim və məsdər varsa, onda isim mübtədə, məsdər isə tamamlıq olur:

– Sevil oxumaq istəyirdi.

Cümlədə adlıq halında işlənən sifətdən sonra isim gəlmirsə, onda sifət mübtədə olur:

– Güzəl heç kəsə məhəl qoymurdu.

Əgər sifətdən sonra isim gəlsəydi, onda sifət təyin olardı. Əslində bu qəbildən olan cümlələrdə işlənən sifətlər aid olduqları isimlərlə yanaşı təsəvvür olunurlar:

– Güzəl /qız/ heç kəsə məhəl qoymurdu.

Cümlədə adlıq halında olan say, ondan sonra isim işlənmədikdə mübtədə, əks təqdirdə təyin olur:

– Yeddinci finişə hamıdan tez çatdı.

– Üç yoldaş kəndə yollandı.

Nümunələrdən görüldüyü kimi, sifət, say və zərf nitq hissələri əşya məfhumu ifadə etdikdə mübtədə olur və bu zaman onlardan sonra işlənməli olan isim aydın təsəvvür olunur.

Mübtədə söz birləşmələri ilə də ifadə oluna bilər.

Adlıq halında olan təyini söz birləşmələri xəbərin felin məlum və ya məchul növü ilə ifadə olunub-olunmamasından asılı olmayaraq mübtədə olurlar:

- Yaşlı kənd sakini ehtiramla salam verdi.
- Qaynar şəhər həyatı ona qəribə gəlirdi.

Adlıq halında olan məsdər tərkibi mübtədə olur:

– Tələbə adına layiq görülmək mənim çoxdankı arzumdur.

Adlıq halında feili sifət tərkibləri mübtədə funksiyasında çıxış edir:

- Müharibədə başına gələnlər onu sarsıtmışdı.

Yuxarıda göstərilən qaydalara müvafiq alqoritmlər tərtib etmək olar. Həmin alqoritmlər avtomatik işarələnmə prosesində də böyük fayda verə bilər. Avtomatik işarələnmə aparılmamışdan əvvəl cümlə üzvləri formal baxımdan müəyyən-ləşdirilməli və yalnız bundan sonra kodlaşdırılmalıdır.

Cümlənin baş üzvlərinin formal təsviri uzlaşma əlaqəsi nəzərə alınmadıqda natamam təsir bağışlayır. Cümlənin təhlili və sintezi prosesində bu əlaqənin araşdırılması və alınmış nəti-cələrin formal təsviri korpus mətninin işarələndirilməsi keyfiy-yətini xeyli yüksəltmiş olur.

Uzlaşma əlaqəsinin formal təsviri. Azərbaycan dilində mübtədə və xəbəri formal əlamətlərinə görə axtaran alqoritmlərin qurulması ilə əlaqədar uzlaşma əlaqəsinin formal təsvirinin də nəzərə alınması hər şeydən əvvəl onun mahiyyətinin aydınlaşdırılmasını şərtləndirir. Məlum olduğu kimi, uzlaşma əlaqəsi şəxsə və kəmiyyətə görə olur. Tabe edən tərəf hansı şəxsə aid olursa, tabe tərəf də həmin şəxsə məxsus qrammatik əlamətlər qəbul edir, tabe edən tərəfin tək və cəmliyi də tabe tərəfdə öz inikasını tapır [Müasir Azərbaycan dili. Sintaksis, 1981, s.48]. Mübtədə xəbərə cümlədə uzlaşma əlaqəsinə görə bağlanır. Xəbər mübtədanın tələb etdiyi şəxs və kəmiyyətdə olur. Bunu belə ifadə etmək olar:

M (Şn) – X (Şn)

Burada n – I, II, III – şəxsin tək və cəmini göstərir.

Bu ümumi ifadə mübtədə ilə xəbərin uzlaşma xüsusiyyətini tam ifadə etmir və ancaq xüsusi hal kimi qəbul edilə bilər. Çox vaxt tabe edən tərəf, bəzən tabe olan tərəf, bəzən hər iki tərəf bir neçə sözdən ibarət olur. Bu halda üç hal müşahidə olunur:

Tabe edən bir neçə, tabe olan bir sözlə ifadə olunur:

tabe edən tərəfdə başqa şəxslərə aid sözlərlə yanaşı birinci şəxsə aid əvəzliliklər iştirak edərsə, tabe söz birinci şəxsin cəmi ilə uzlaşır:

M (Ş1, Ş2, Ş3) – X (Ş1c)

– Mən, sən və o (qardaşım) futbol həvəskarıyıq.

Burada Ş1–I, Ş2–II, Ş3–III şəxsin təkini, Ş1c–I şəxsin cəmini göstərir.

əgər tabe edən sözlər arasında birinci şəxs əvəzlilikləri iştirak etmərsə, həmin sözlərdən biri ikinci şəxsə aid əvəzlilik (sən, siz) olarsa, tabe söz ikinci şəxsin cəmi ilə uzlaşacaqdır:

M (Ş2, Ş3) – X (Ş2c)

– Siz və o (tələbə) tədbirə tez qatılmalısınız.

tabe edən sözlər ancaq üçüncü şəxsə aid olarsa, tabe söz həmişə üçüncü şəxsə aid olur:

M (Ş3, Ş3, Ş3) – X (Ş3c)

– Aysu, Tunar və Murad həyətdə oynayırdılar.

Tabe edən bir, tabe olan bir neçə sözlə ifadə olunur:

tabe olan sözlərin hamısı şəxsə və kəmiyyətə görə uzlaşır:

M (Ş1) – (Ş1, Ş1)

– Mən oxuyuram və yazıram.

tabe sözlərdən ancaq sonuncusu əsas sözlə şəxsə və kəmiyyətə görə uzlaşır:

M (Ş1) – X (Ş0, Ş0, Ş1)

– Mən oxuyur, işləyir və idmanla məşğul olurdum.

Tabe söz (xəbər) şəxs əvəzlilikləri ilə ifadə olunanda uzlaşma əlaqəsi pozulur. Belə birləşmələrdə əsas tərəf üçüncü şəxsə olur, tabe söz isə müxtəlif şəxs əvəzlilikləri ilə ifadə olunur və birinci tərəfin şəxs və kəmiyyətindən asılı olmayaraq müxtəlif şəkilçilər qəbul edir:

M (§3) –X (§1, §2, §3//§3c)

M (§30) –X (§10, §20, §3c)

Formal təhlil sistemlərində mübtədə və xəbəri əlamətlərinə görə axtaran sxemlərin işini konkret nümunələr əsasında nəzərdən keçirək:

– Elçin dünən anasına məktub yazdı.

Morfoloji təhlil alqoritmlərinin işindən sonra aşağıdakıları müəyyənləşdiririk:

Elçin (isim, adlıq hal, tək, xüsusi, insan adı və s.);

dünən (zərf, zaman zərfi, adlıq hal, sadə);

məktub (isim, adlıq hal, cənsiz, tək, ümumi);

anasına (isim, yönlük hal, tək, ümumi);

yazdı (feil, keçmiş zaman, III şəxs, tək, təsirli).

Morfoloji təhlil prosesində əldə etdiyimiz informasiyalar sintaktik təhlil üçün baza rolunu oynayır. Sintaktik təhlil prosesində məlum olur ki, xəbərin əlamətləri onun aşağıdakı variantı uyğun gəldiyini göstərir:

X [F/X/ - (mv) + (dy) +ş

Burada -mv (-ma, -mə) şəkilçisinin iştirak etmədiyi variantdır. Mübtədanı əlamətlərinə görə axtaran sxemin işini addım-addım izləyək.

A (İ) –X [F (m1)] – modelinə görə cümlədə adlıq halında bir isim varsa, onda o, mübtədə olmalıdır. Bizim nümunə gətirdiyimiz cümlədə adlıq halında olan iki isim vardır. Dürğu işarələrinin təhlili sxeminin işinin nəticəsində məlum olur ki, bu iki isim arasında vergül yoxdur, deməli, onların həmcins mübtədə olmaq ehtimalı aradan qalxmış olur.

İndiyə qədər araşdırdığımız əməliyyatlar nəticəsində məlum olur ki, cümlə sadədir və onun bircə xəbəri var (həmcins xəbərlili cümlə deyil). Onda sintaktik bölmədə aşağıdakı modelə müraciət etməli oluruq (formal əlamətlərə əsaslanan sintaktik modellərlə bağlı ətraflı məlumat “Mətnin formal təhlili sistemi” monoqrafiyasında verilmişdir [Mahmudov, 2002, s.164-185]:

A (İ1, İ2) –X (F) –İ1 –M–İ2 –Tm

Buradan belə nəticə çıxır ki, Elçin mübtəda, məktub isə tamamlıqdır. Məlum olduğu kimi, zərflər də cümlədə mübtəda yerində işlənə bilər. Dünən sözünün mübtəda olmadığına aşağıdakı model zəmanət verə bilər:

A (İ Z) –X (Fm2) İ–M

Yəni, cümlədə adlıq halında olan isim və zərf varsa, isim mübtəda olur.

Sintaktik təhlil blokunda köməkçi sözlər və durğu işarələrinin təhlili sxemlərinin də rolu böyükdür. Məsələn, mübtədanın müəyyən edilməsi üçün cümlədə heç bir durğu işarəsinin (xüsusən, vergülün) olmamasının özü də mühüm, həlledici əlamət sayıla bilər. Əgər təhlil etdiyimiz nümunədə omonim, çoxmənalı və naməlum sözlər müşahidə olunsaydı, onda semantik blok da işə “qarışmalı” olardı.

Semantik təhlil alqoritmlərinin işini araşdırmaq məqsədi ilə daha bir nümunəyə diqqət yetirək:

– Fikrət ağacdan alma dərdi.

Morfoloji təhlil blokundan sonra alırıq:

Fikrət (isim, adlıq hal, canlı, tək, xüsusi, insan adı, sadə);

alma (isim, adlıq hal, cansız, tək, ümumi, sadə);

alma (feil, əmr şəklinin inkarı, II şəxs, tək, təsirli);

ağacdan (isim, çıxışlıq hal, cansız, tək, ümumi);

dərdi (feil, şühudi keçmiş zaman, III şəxs, tək, təsirli);

dərdi (isim, təsirlik hal, cansız, tək, ümumi).

Cümlədə “alma” və “dərdi” sözləri omoqraf olduğu üçün semantik təhlil alqoritmləri ilə onların müvafiq variantını qəbul etməliyik.

Xəbəri əlamətlərinə görə axtaran modellərə görə iki variant alırıq:

$$X (F/\Theta) + (mv) +\S$$

$$X (F/\Theta) - (mv) + (dy) +\S$$

Bizim nümunə üçün modelin aşağıdakı variantı qəbul oluna bilər:

$$X (F/\Theta) + (mv) +0$$

$$X (F/X) - (mv) + (dy) +0$$

Semantik təhlil sxemi sintaktik təhlillə qarşılıqlı əlaqə sayəsində bu omoqraflardan birini düzgün hesab edib, bu bərdə morfoloji bloka məlumat verməlidir. Sintaktik bloka qayıdaq.

Əgər biz alma sözünün isim nitq hissəsi olduğunu hələ qəbul etməmişiksə, onda mübtədanı əlamətlərinə görə axtaran aşağıdakı model üzrə mübtədanı tapmaq imkanı yaranır:

$$A (\dot{I}) -X (F/m1/m2/\dot{I}) \dot{I} -M$$

Bu modelə görə Fikrət sözünün mübtəda olduğu aşkar olur. Mübtəda ilə xəbərin uzlaşma əlaqəsinə görə alırıq:

$$\text{Əgər } M (\S3) -X (\S1, \S2, \S3) -sə, \text{ onda } M (E1) -X (E1)$$

Bizim tədqiq etdiyimiz cümlədə xəbərin III şəxsin təkində olması tələb olunur:

$$M (E3) -X (E3)$$

$$\text{Fikrət (O) - dərdi.}$$

Deməli, xəbərin birinci modeli əsasında tapılmış alma sözü bir xəbər kimi özünü “doğrultmur”. Semantik təhlil sintaktik təhlillə qarşılıqlı əlaqədə morfoloji bloka alma sözünün birinci variantının qəbulunu məsləhət görür. Birinci variant isə bu sözün isim olduğunu göstərir:

$$\text{alma (isim, adlıq hal, cansız, tək, ümumi, sadə).}$$

Bu əlamətə görə onu mübtədə da hesab etmək olar, çünki formal baxımdan qeyri-müəyyən təsirlik halı adlıq haldan seçmək mümkün deyil:

A (İ) –X (F/m1/m2/İ) İ–M- modeli mübtədəyə görə yeganə “namizəd” olan hal üçün doğru idi. Semantik blokdan sonra iki mübtədə alındığı üçün ona uyğun modeldən istifadə olunmalıdır. Durğu işarələrinin təhlili bloku bu iki ismin həmcins mübtədə olması şərtini rədd edir:

A (İ1, İ2) –X (F/m1/m2/İ) İ1 –M

İ2 –Tm – modelinə əsasən Fikrət sözünün mübtədə, alma sözünün isə tamamlıq olduğu aşkar olur. Analoji qaydada sintaktik sıralanma və uzlaşma əlaqələrinin nəzərə alınması ilə dərdi omoqrafının isim (təsirlik halında) deyil, feil (feil, keçmiş zaman, xəbər) olduğu müəyyənləşir.

Formal əlamətlərinə görə cümlə üzvlərinin təyini. Azərbaycan dilində sintaktik əlamətlərinə görə cümlə üzvlərinin axtarışı prosesini mərhələlər üzrə aşağıdakı ardıcılıqla yerinə yetirmək olar:

Mətnin kompüterə daxil edilməsi, oxunması, cümlənin mətndən ayrılması və 2-yə keçid;

Söz-formanın cümlədən ayrılması və 3- keçid;

Söz-formanın uzunluğunun təyini (N);

$N > m$ şərtinin yoxlanması: əgər $N > m$ -dirsə 6-ya, əks halda 5-ə keçid;

Söz-formanın avtomatik lüğətin (AL) müvafiq qrupu ilə müqayisəsi: əgər söz-forma tapılıbsa 11-ə, əks halda 8-ə keçid;

Söz-formanın ayrılması ($N = m$) və 7- yə keçid;

Söz-formanın AL-in əsaslar siyahısı ilə müqayisəsi: əgər söz tapılıbsa 11-ə, əks halda 8-ə keçid;

Söz-formanın uzunluğunun bir hərf azaldılması ($N - 1$);

$N = 0$ şərtinin yoxlanması: əgər $N = 0$ -dirsə, 10-a, əks təqdirdə 5-ə keçid;

Söz-formanın naməlum sözlər siyahısına yazılması və 30-a keçid;

Əsasın omonimliyi şərtinin yoxlanması: əgər omonimdirsə 12-yə, əks təqdirdə 32-yə keçid;

Əsasların omonimliyi blokuna müraciət və 32-yə keçid;

Şəkilçilərin omonimliyi şərtinin yoxlanması: əgər omonimdirsə 14-ə, əks təqdirdə 31-ə keçid;

Şəkilçilərin omonimliyi blokuna müraciət və 31-ə keçid;

Söz-formanın feillər siyahısı ilə tutuşdurulması və 16-ya keçid;

Feilin tapılma şərtinin yoxlanması: əgər feil siyahıda tapılmışsa 18-ə, əks təqdirdə 17-yə keçid;

Sözün mənasının işçi sahəyə yazılması və 18-ə keçid;

Feildən əvvəl gələn söz-formanın cümlədən ayrılması və 19-a keçid;

Ayrılmış söz-formanın feillə müvafiq sözlər siyahısı ilə müqayisəsi və 20-yə keçid;

Sözün AL-də tapılması şərtinin yoxlanması: əgər söz tapılmışsa 21-ə, əks halda 3-ə keçid;

Birləşmənin mənasının (sintaktik əlamətinin) işçi sahəyə yazılması və 22-yə keçid;

L-hərfdən ibarət qalığın yoxlanması və 23-ə keçid;

$L=0$ şərtinin yoxlanması: əgər $L=0$ -dirsə 2-yə, əks halda 24-ə keçid;

Qalığın uzunluğunun (L) təyini və 25 şərtinə keçid;

$L>n$ şərtinin yoxlanması: əgər $L>n$ -dirsə 26-ya, əks halda 27-yə keçid;

Yeni qalığın ayrılması ($L'=n$) və 27-yə keçid;

Qalığın şəkilçilər siyahısı ilə müqayisəsi və 28-ə keçid;

Qalığın şəkilçilər siyahısında tapılması şərtinin yoxlanması: tapılmışsa 31-ə, əks halda 29-a keçid;

Qalığın uzunluğunun bir hərf azaldılması (L-1) və 30-şərtinə keçid;

L=0 şərtinin yoxlanması: əgər L=0-dırsa 10-a, əks halda 27-yə keçid;

Şəkilçinin və onun əlamətlərinin işçi sahəyə əsasdan sonra yazılışı və 33-ə keçid;

Əsas və şəkilçilərin qrammatik səciyyələri ilə işçi sahəyə yazılışı və 33-ə keçid;

Xəbərin əlamətlərinə görə axtarışı və 34-ə keçid;

Mübtədanın əlamətlərinə görə axtarışı və 35-ə keçid;

Mübtədə və xəbər qrupunun müəyyənləşdirilməsi və 36-şərtinə keçid;

Cümlə üzvlərinin sintaktik funksiyalarının tapılması şərtinin yoxlanması: əgər tapılmışsa 38-ə, əks halda 37-yə keçid;

Semantik təhlil və sintez blokuna müraciət və 38-ə keçid;

Cümlə üzvlərinin sintaktik əlamətlərinin işçi sahəyə yazılması və 39 şərtinin yoxlanması;

Mətnin sonunun yoxlanması: əgər mətn bitmişsə 40-a, əks halda 1-ə keçid;

Təhlil prosesinin sonu.

Burada N – söz-formanın uzunluğu, m – əsasların maksimal uzunluğu, L – şəkilçilər zəncirinin uzunluğu, n – şəkilçilərin maksimal uzunluğu, s/f – söz-formanı göstərir.

Yuxarıda göstərilmiş Azərbaycan dilinin təhlil alqoritmini konkret nümunələr üzərində nəzərdən keçirək. Sadəlik üçün hesab etmək birləşməsinin sintaktik təhlil prosesini izləyək. Hər şeydən əvvəl cümlənin sonundan başlayaraq söz-formanı ayıraraq onun uzunluğunu tapırıq. Söz-forma etmək beş hərfdən ibarət olduğu üçün ($N < m$ şərtinə görə) AL-in müvafiq qrupu ilə (bizim misalda beş hərfdən ibarət əsaslar qrupu) müqayisə aparılır (prosesi nisbətən sadələşdirmək üçün fərz edilir ki, feil məsdər formasında təmsil olunmuşdur. Əslində isə, məlum olduğu kimi feillər AL-də məsdər şəkilçiləri olmadan təmsil olunmuşdur). Göstərilən sözün morfoloji, sintaktik əlamətlərini bu yolla əldə edirik. Tapılmış sözün

müstəqil və ya söz birləşməsinin komponenti olub-olmaması yoxlanır. Bu məqsədlə söz feillər siyahısı ilə müqayisə olunur. Tapırıq ki, bu söz feillər siyahısında birləşmənin ikinci tərəfi kimi qeyd olunmuşdur. Sonrakı mərhələdə cümlədə həmin feildən əvvəl gələn söz-forma ayrılır və ona uyğun sözlər qrupu ilə müqayisə olunur. Hesab sözü həmin qrupda təmsil olunmuşdur. Buna görə də həmin söz ayrılıqda deyil, söz birləşməsi kimi səciyyələndirilir. Nəticədə sintaktik səviyyədə feili söz birləşməsinin qrammatik əlamətləri müəyyənləşdirilir. Həmin əlamətlərə əsaslanaraq onları avtomatik kodlaşdırmaq, işarələmək mümkündür. Bu nəzəri deyil, artıq texniki məsələ kimi öz həllini tapa bilər.

Fərz edək ki, biz sadə cümləni formal əlamətlərinə görə tam sintaktik təhlil edə bilirik. Onda tabeli və tabe mürəkkəb cümlələri formal meyarlara görə sadə cümlələrə ayırmaqla, onları da komponentlərinə görə həmin qayda ilə təhlil etmək olar. Lakin, nəzərdə tutmaq lazımdır ki, tabeli mürəkkəb cümləni təşkil edən sadə cümlələrlə tabeli mürəkkəb cümləni təşkil edən sadə cümlələr qrammatik əlaqələrinə və semantikasına görə müxtəlifdir. Məlum olduğu kimi, tabeli mürəkkəb cümləni təşkil edən cümlələr tam müstəqil deyillər. Onlar qrammatik baxımdan bərabərhüquqlu olsalar da, müəyyən məna əlaqələri ilə (birləşdirmə, səbəb-nəticə, aydınlaşdırma, qarşılaşdırma, bölüşdürmə, qoşulma) bağlanırlar. Eləcə də, tabeli mürəkkəb cümlələr də müəyyən tabelilik əlaqəsi ilə birləşib vahid bir tam əmələ gətirirlər. Tabeli mürəkkəb cümlələrdə komponentlərdən biri müstəqil (baş cümlə), digəri isə asılı (budaq cümlə) hesab olunur. Deməli, istər tabeli, istərsə də tabeli mürəkkəb cümləni təşkil edən sadə cümlələr tam müstəqil hesab oluna bilməz. Onlar tərkiblərində olduqları mürəkkəb cümlənin semantikasını müəyyən qədər ehtiva etmiş olurlar. Yəni, mürəkkəb cümlə tərkibindəki sadə cümlələri bir-birinə bağlayan məna əlaqələri, qrammatik müstəqillik və

asılılıq xüsusiyyətləri qismən komponentlərə – sadə cümlələrə də keçir. Təhlil prosesində bu cəhət imkan daxilində nəzərə alınmalıdır. Yəni, təhlil elə aparılmalıdır ki, sonradan tabesiz və ya tabeli mürəkkəb cümlə sintez olunanda əvvəlki statusunu, əvvəlki semantikasını qoruyub saxlaya bilsin. Əlbəttə, bu asan əldə olunması məsələ olmasa da, təhlilin və sintezin uzlaşdırılması baxımından olduqca vacibdir. Təhlil elə aparılmalıdır ki, məsələn, xəbər budaq cümləsini iki sadə cümləyə ayırıb təhlil etdikdən sonra sintez etsək (ilkin formanı bərpa etsək), sintez nəticəsində alınmış cümlə xəbər budaq cümləsi olaraq qalmalıdır (və ya xəbər cümləsi təsiri bağışlamalıdır).

Beləliklə, Azərbaycan dilində cümlənin formal əlamətlərə görə təhlili prosesini ümumi şəkildə belə təsvir etmək olar: əgər cümlədə predikat yoxdursa (adlıq cümlə, söz-cümlə, vokativ cümlələr), həmin cümləni təşkil edən söz (sözlər) predikatlıq məzmunu ifadə edən sintaktik vahid (predikat) funksiyasında hesab olunur və cümlə birpredikatlı sadə cümlə kimi təhlil olunur. Əgər cümlədə predikatların sayı birdən çoxdursa, onda cümlə mürəkkəb cümlə (tabesiz, tabeli, həmcins üzvlü sadə cümlə) hesab olunur. Həmin cümlələr uyğun olaraq sadə cümlələrə ayrılır və ayrı-ayrılıqda sadə cümlə kimi təhlil olunur. Təhlil olunmuş sadə cümlələr təhlildən sonra sintez edilərək yenidən əvvəlki statusuna (tabeli, tabesiz, sadə genişləndirilmiş cümlə) qaytarılır. Əgər təhlildən sonrakı sintez prosesində cümlə ilkin statusuna tam qayıda bilirsə, bu təhlilin (eləcə də tərcümənin) daha dəqiq və adekvat olmasından xəbər verir. Cümlənin formal əlamətləri dedikdə predikativlik (cümlədə olan predikatların sayı), bağlayıcılar (tabelilik və ya tabesizlik bağlayıcıları), baş və budaq cümləni bir-birinə bağlayan qəlib sözlər və digər vasitələr başa düşülür.

Bu yanaşma üsulu formal təhlil sistemlərində cümlənin formal əlamətlərə görə tədqiqi, maşın tərcüməsi sistemlərində

tərcümənin keyfiyyətinin yüksəldilməsi, daha düzgün tərcümə variantlarının alınması prosesində, eləcə də tədrisdə istifadə oluna bilər. Formal təhlil sisteminin ümumi məsələləri, iş prinsipləri və konkret tətbiq sahələri barədə ölkəmizdə və xaricdə çap olunmuş monoqrafiya və məqalələrdə ətraflı məlumat verildiyi üçün bu məsələ üzərində geniş dayanmayacağıq.

Formal təhlil sisteminin linqvistik bazası cümləni və onu təşkil edən söz-formaları ayırmağa (cümləni mətndən – cümlələr çoxluğundan, söz-formaları cümlədən) və linqvistik baxımdan təhlil etməyə imkan verir. Formal təhlilin ilk pillələrində söz-formalar lüğət formasına gətirilir (əgər söz-formalar lüğətində tapılmazsa) və onun qrammatik statusu müəyyənləşdirilir. Əldə olunmuş linqvistik informasiya bütövlükdə cümlənin linqvistik (sintaktik) təhlili üçün zəruri məlumat kimi istifadə olunur.

Deməli, formal təhlil aparmaq üçün bizim hər şeydən əvvəl söz-formaları cümlədən, cümlələri mətndən düzgün ayıraraq linqvistik baxımdan təhlil etməyə qabil linqvistik bazamız və proqram təminatımız olmalıdır. Təhlilin hərtərəfli və məqbul olması linqvistik bazanın və proqram təminatının səviyyəsi və keyfiyyəti ilə ölçülür.

Tabeli mürəkkəb cümlələrin formal təyini və işarələnməsi. Qeyd etmək lazımdır ki, tabesiz mürəkkəb cümlə və həmcins xəbərlə sadə cümlələrdən fərqli olaraq tabeli mürəkkəb cümlələrdə hər bir budaq cümlənin müəyyən fərqləndirici formal əlaməti vardır. Həmin əlamətlər əsasında hər bir budaq cümləni müəyyənləşdirmək mümkündür. Bu isə tabeli mürəkkəb cümlələrin təsnifi, təhlili və başqa dillərə tərcüməsi baxımından olduqca əhəmiyyətlidir. Bu konkret əlamətlər baş cümlədə qalib söz funksiyasında budaq cümləni əvəz etmiş olur. Baş cümlə budaq cümləyə həmin qalib sözlərlə “ki” bağlayıcısı və vergüllə bağlanır [Mahmudov, 2018, s. 255-259]:

mübtədə budaq cümləsində baş cümlədə mübtədə vəzi-
fəsində aşağıda göstərilən qəlib sözlər işlənir: o (da), bu (da),
orası (da), burası (da), eləsi, beləsi, elələri, belələri, bəziləri, bir
neçələri, o cəhət, bu cəhət, bir cəhət, o şey, bu şey, bir şey, o
məsələ, bu məsələ, kim, hər kim (ki), hər kəs (ki), nə, hər nə
(ki). Bu sözlər, bəzən, buraxıla da bilər. O zaman xəbər
funksiyasını – məlumdur ki, məlum oldu ki, aydındır ki,
bəllidir ki, bilinir ki və s. sözlər yerinə yetirir.

– Kim ki işinə ciddi yanaşmır, çıxıb gedə bilər.

xəbər budaq cümləsində baş cümlədə qəlib söz funksiyasını
işarə əvəzlilikləri yerinə yetirir: budur (ki), odur (ki), bu idi, bu
imiş, o imiş, oradadır, buradadır, orasındadır, burasındadır,
ondadır, bundadır, burada imiş, burada idi, orada imiş, orada idi,
burasında idi, burasında imiş, orasında idi, orasında imiş, ona
görədir ki, ondan ötrüdür ki, o deməkdir ki, bu deməkdir ki,
ondan ibarətdir ki, bundan ibarətdir ki, o demək idi ki, bu demək
idi, ondan ibarət idi, bundan ibarət idi, o demək imiş, ondan ibarət
imiş, o olar ki, bu olar ki, bu olub ki, o olub ki, bu oldu ki, o oldu
ki, o şeydir ki, bu şeydir ki, o kəsdir ki, bu kəsdir ki, o kəslərdir ki,
bu kəslərdir ki və s.

– Arzum budur ki, gələcəkdə həkim olum.

tamamlıq budaq cümləsində baş cümlədə qəlib kimi aşağı-
ğıdakı sözlər işlənir: ona, buna, onu, bunu, onda, bunda, ondan,
bundan, orasını, burasını, orasını da, burasını da, eləsini, beləsini,
elələrini, belələrini, elələrinə, belələrinə, elələrində, belələrində,
eləsinə, beləsinə, eləsində, beləsində, o cəhəti, bu cəhəti, o cəhətə,
bu cəhətə, o cəhətdə, bu cəhətdə, o məsələyə, bu məsələyə, o
məsələni, bu məsələni, o məsələdə, bu məsələdə, o məsələdən, bu
məsələdən, o şeyi, bu şeyi, o şeyə, bu şeyə, o şeydə, bu şeydə, o
şeydən, bu şeydən və s.

– Orasını da bilirilər ki, mən vicdanla xidmət göstər-
mişəm.

təyin budaq cümləsi baş cümlədə işarə əvəzliyi ilə ifadə olunan qəlib sözün mənasını dəqiqləşdirir. Budaq cümlənin göstəricisi olan qəlib sözlər bunlardır: elə, elə bir, belə, belə bir, o cür, bu cür, o, bu, həmin, haman, hansı, kim, nə və s. Təyin budaq cümləsinin qəlib sözləri təyin etdikləri ad nitq hissəsindən əvvəl gəlir.

– Elə müəllimlər var ki, şagirdlər dinləməkdən yorulmurlar.

tərz budaq cümləsi – baş cümlədəki hərəkətin icra tərzini, necə, nə şəkildə icra olunduğunu bildirir və baş cümlə budaq cümləyə, elə, belə, o cür, o tövr, o yolla, bu yolla, bu halda, o halda, necə, nə cür, nə təhər qəlib sözləri, “ki” bağlayıcısı ilə bağlanır.

– Elə get ki, heç kəs görməsin.

dərəcə budaq cümləsi – baş cümlədə ifadə olunan hərəkət və ya əlamətin dərəcəsini bildirir və baş cümləyə “ki” bağlayıcısı, intonasiya və aşağıdakı qəlib sözlərlə bağlanır: o dərəcədə, o qədər, elə, o yerə, bu yerə, oracan, o məqama, bir elə mərtəbəyə və s.

– O qədər yorulmuşdum ki, nə vaxt yuxuya getdiyimi bilməmişəm.

Zaman budaq cümləsi – baş cümlədəki hərəkət və ya əlamətin zamanını bildirir və baş cümləyə “ki” bağlayıcısı, intonasiya, eləcə də, o zaman, o vədə, o vaxt, elə bir vaxtda, onda, elə ki, o zaman ki, o gün ki, o vaxt ki, nə qədər, nə qədər ki, necə ki və s. bağlayıcı-qəlib sözlərlə bağlanır.

– O zaman gəldi ki, artıq gec idi.

Yer budaq cümləsi – baş cümlədəki hərəkət və ya əlamətin yerini bildirir və budaq cümləyə, o yerə, o yerdə, orada, oradan, o yerdən, həmin yerə, həmin yerdən, hansı yerə, hansı yerdən, hansı yerdə, haranın ki, haraya ki, haranı ki, harada ki, haradan ki, o yerə ki, o yerdə ki, o yerdən ki, o yerin ki, bir yerə

ki, bir yeri ki, bir yerdə ki, bir yerdən ki və s. kimi qəlib-bağlayıcı sözlərlə bağlanır.

– O yerdə ki, böyük-kiçik yoxdur, orada bərkət olmaz.

Kəmiyyət budaq cümləsi – baş cümlədəki hərəkət və ya əlamətin kəmiyyətini bildirir və budaq cümləyə nə qədər, hər nə qədər, o qədər, o kəmiyyətdə, o miqdarda qəlib sözləri ilə bağlanır.

– O qədər yazdım ki, tamam əldən düşdüm.

Səbəb budaq cümləsi – baş cümlədə ifadə olunan hərəkət və ya əlamətin səbəbini bildirir və budaq cümləyə ona görə, ondan ötrü, onun üçün, o səbəbə, bu səbəbə, buna görə, bundan ötrü, ona görə ki, ondan ötrü ki, o səbəbə ki, çünki, ki bağlayıcı və qəlib sözlərlə bağlanır.

– Sən haqlı deyilsən, o səbəbə ki, məsələdən xəbərin yoxdur.

Məqsəd budaq cümləsi – baş cümlədə ifadə olunan hərəkətin məqsədini bildirir və baş cümləyə “ki” bağlayıcısı və intonasiya, eləcə də, ona görə, ondan ötrü, o məqsədlə sözləri ilə bağlanır.

– Ondan ötrü dedim ki, xəbərin olsun.

Şərt budaq cümləsi – baş cümlədə ifadə olunan hərəkət və ya əlamətin şərtini bildirir və baş cümləyə intonasiya, “ki” bağlayıcısı, -sa, -sə (isə) köməkçi sözü və şəkilçisi, eləcə də bir şərtlə, bu şərtlə, o şərtlə, həmin şərtlə qəlib sözləri, əgər (ki), madam ki, indi ki, vaxta ki, bir halda ki bağlayıcı sözləri və ya ki ədatı ilə bağlanır.

– Madam ki, etiraf etdin, onda sənə güzəşt etmək olar.

Qarşılaşdırma budaq cümləsində baş və budaq cümlənin məzmunu bir-biri ilə nə qədər, hər nə qədər, hərçənd bağlayıcı sözləri, -sa, -sə şəkilçisi, da (də) və belə ədatları ilə qarşılaşdırılır, müqayisə edilir.

– Nə qədər çalışdım, xeyri olmadı.

Nəticə budaq cümləsi baş cümlədən doğan nəticəni bildirir və baş cümləyə intonasiya, “ki” bağlayıcısı, eləcə də odur ki, ona görə (də), buna görə (də), onun üçün (də), bunun üçün (də) bağlayıcıları ilə bağlanır. Bəzən baş cümlənin daxilində elə, elə bir, o qədər sözləri də işləne bilər.

– Ona görə inciyib ki, dərslərini yaxşı öyrənməmişən.

Qoşulma budaq cümləsi – baş cümləyə qoşularaq, ya baş cümlədəki fikirlə bağlı əlavə məlumat verir, ya da həmin fikrə münasibət bildirir. Baş cümlə budaq cümlədən əvvəl işlənir ona intonasiya və “ki” bağlayıcısı ilə bağlanır. Budaq cümlənin əvvəlində ismin müxtəlif hallarında bu, o əvəzlilikləri və da (də) adət işlənir.

– Tərlədə bircə adam var idi, o da heç bir iş görmürdü [Fətullayev, Mahmudov, Fətullayev, Dərvişov, 2015, s.3-9].

Azərbaycan dilində cümlənin formal əlamətlərinə görə təhlili alqoritmi:

Cümlənin mətndən ayrılması;

(Cümlə nöqtə ilə – nöqtə, sual işarəsi, nida işarəsi, üç nöqtə arasında yerləşən söz-formalar çoxluğu kimi təyin edilə bilər. Nəzərə almaq lazımdır ki, mətnin ilk cümləsinin qarşısında durğu işarəsi qoyulmur. Sonrakı cümlələrdə cümlənin əvvəlində durğu işarəsi – nöqtə, sual işarəsi, nida işarəsi, nöqtəli vergül, üç nöqtə olur).

Predikat mərkəzlərinin (predikatların) sayının müəyyənləşdirilməsi;

Predikat yoxdur ($P=0$), onda 5-ə keçid;

Predikat bir dənədir ($P=1$), onda yenə 5-ə keçid;

Cümlə sadədir (sadə cümlələrin ya heç predikatı olmur, ya da bir predikatı olur);

$P=0$ və $P=1$ olduqda, cümlə sadə hesab olunur və sadə cümlə kimi təhlil edilir (burada P -predikatı bildirir);

Predikat iki və ya daha çoxdur ($P>1$);

Cümlə mürəkkəbdir və mürəkkəb cümlə kimi təhlil olunur;

Cümləni səciyyələndirən xüsusi əlamət və ya qəlib söz varmı? Hə – 10-a, yox – 9-a keçid;

Cümlə tabesiz mürəkkəb cümlədir və tabesiz mürəkkəb cümlə kimi təhlil olunmalıdır;

Cümlə tabeli mürəkkəb cümlədir;

Cümlədə mübtəda budaq cümləsinin əlamətləri varmı? Hə – 12-yə, yox – 13-ə keçid;

Cümlə mübtəda budaq cümləsidir və mübtəda budaq cümləsi kimi komponentlərinə ayrılaraq təhlil olunur;

Cümlədə xəbər budaq cümləsinin əlamətləri varmı? Hə – 14-ə, yox – 15-ə keçid;

Cümlə xəbər budaq cümləsidir və xəbər budaq cümləsi kimi təhlil olunur.

Cümlədə tamamlıq budaq cümləsinin əlamətləri varmı? Hə – 16-ya, yox – 17-yə keçid;

Cümlə tamamlıq budaq cümləsidir və tamamlıq cümləsi kimi təhlil olunur;

Cümlədə təyin budaq cümləsinin əlamətləri varmı? Hə – 18-ə, yox – 19-a keçid;

Cümlə təyin budaq cümləsidir və təyin budaq cümləsi kimi təhlil olunmalıdır;

Cümlədə tərz budaq cümləsinin əlamətləri varmı? Hə – 20-yə, yox – 21-ə keçid;

Cümlə tərz budaq cümləsidir və tərz budaq cümləsi kimi təhlil olunmalıdır;

Cümlədə dərəcə budaq cümləsinin əlamətləri varmı? Hə – 22-yə, yox – 23-ə keçid;

Cümlə dərəcə budaq cümləsidir və dərəcə budaq cümləsi kimi təhlil olunmalıdır;

Cümlədə zaman budaq cümləsinin əlamətləri varmı? Hə – 24-ə, yox – 25-ə keçid;

Cümlə zaman budaq cümləsidir və zaman budaq cümləsi kimi təhlil olunmalıdır;

Cümlədə yer budaq cümləsinin əlamətləri varmı? Hə – 26-ya, yox – 27-yə keçid;

Cümlə yer budaq cümləsidir və yer budaq cümləsi kimi təhlil olunmalıdır;

Cümlədə kəmiyyət budaq cümləsinin əlamətləri varmı? Hə – 28-ə, yox – 29-a keçid;

Cümlə kəmiyyət budaq cümləsidir və kəmiyyət budaq cümləsi kimi təhlil olunmalıdır;

Cümlədə səbəb budaq cümləsinin əlamətləri varmı? Hə – 30-a, yox – 31-ə keçid;

Cümlə səbəb budaq cümləsidir və səbəb budaq cümləsi kimi təhlil olunmalıdır;

Cümlədə məqsəd budaq cümləsinin əlamətləri varmı? Hə – 32-yə, yox – 33-ə keçid;

Cümlə məqsəd budaq cümləsidir və məqsəd budaq cümləsi kimi təhlil olunmalıdır;

Cümlədə şərt budaq cümləsinin əlamətləri varmı? Hə – 34-ə, yox – 35-ə keçid;

Cümlə şərt budaq cümləsidir və şərt budaq cümləsi kimi təhlil olunmalıdır;

Cümlədə qarşılaşdırma budaq cümləsinin əlamətləri varmı? Hə – 36-ya, yox – 37-yə keçid;

Cümlə qarşılaşdırma budaq cümləsidir və qarşılaşdırma budaq cümləsi kimi təhlil olunmalıdır;

Cümlədə nəticə budaq cümləsinin əlamətləri varmı? Hə – 38-ə, yox – 39-a keçid;

Cümlə nəticə budaq cümləsidir və nəticə budaq cümləsi kimi təhlil olunmalıdır;

Cümlədə qoşulma budaq cümləsinin əlamətləri varmı? Hə – 40-a, yox – 41-ə keçid;

Cümlə qoşulma budaq cümləsidir və qoşulma budaq cümləsi kimi təhlil olunmalıdır;

Cümləni istisna hallar siyahısına yazmalı və 1-ə keçid [Mahmudov, 2018, s. 259-260].

Semantik səviyyədə avtomatik işarələnmə. Dilçilik ədəbiyyatında semantika, semantik məna anlayışlarının izahında müxtəlifliklər müşahidə olunur. Bəzən semantik məna leksik və qrammatik mənalarla qarışdırılır, onlar arasında oxşar və fərqli cəhətlər dəqiq izah olunmur. Bu mənaların eyniləşdirildiyi hallara da rast gəlinir. Məsələn, “İzahlı dilçilik terminləri lüğəti”ndə semantika belə izah olunur: 1. Sözün, ifadənin və ya qrammatik formanın mənası; 2. Dil vahidlərinin daxili (məna) cəhətinin quruluşunu, onun əsas qanunauyğunluqlarını öyrənməklə məşğul olan dilçilik sahəsi [Adilov, Verdiyeva, Ağayeva, 1989, s.231].

İkinci izah daha çox semantikanı dilçilik sahəsi kimi səciyyələndirir. Birinci izah isə semantikanın mahiyyəti ilə bağlı məsələdir. Bu izah dolaşlıq və ümumidir. Semantikanın linqvistik mahiyyəti burada açılmır. Əvvələn, sözün, ifadənin mənası leksikada öyrənilir (sözün leksik mənası), qrammatik formalar, qrammatik məna isə qrammatikada (morfologiya və sintaksis) öyrənilir. Bəs semantik məna dedikdə nə başa düşülür, aydın olmur.

“Dilçilik ensiklopediyası”nda isə semantika belə izah olunur: dildəki morfem və sözlərin təhlili və təsviri ilə məşğul olan dilçilik sahəsi [Dilçilik ensiklopediyası, I cild, 2008, s.206]. Sual olunur, əgər semantika dildəki morfem və sözlərin təhlili və təsviri ilə məşğul olursa, onda dilçiliyin morfologiya və leksika kimi sahələri nə ilə məşğul olur? Əgər leksika və morfologiya ilə yanaşı semantika da morfem və sözlərin təhlili və təsviri ilə məşğul olursa, onda onlar arasında fərq nədən ibarətdir?

Əgər heç bir fərq yoxdursa, onda ayrıca bir dilçilik sahəsi kimi semantikanın yaranmasına nə ehtiyac var idi?

Elə yenə həmin ensiklopediyada morfolojiya haqqında yazılır: dilin morfem tərkibini öyrənən elmə morfolojiya deyilir [Dilçilik ensiklopediyası, I cild, 2006, s.451].

Əvvəla, morfolojiya ayrıca elm deyil, dilçiliyin bir sahəsidir [Bax: İzahlı dilçilik terminləri, 1989, s.168]. Dildəki morfem və sözlərin təhlili və təsviri ilə dilin morfem tərkibini öyrənmək arasında nə fərq var? Əgər, doğrudan da, fərq varsa, onda bu fərq nədən ibarətdir (əslində, semantikanın izahına elə bundan başlamaq lazım idi).

Leksik məna ilə semantik məna arasındakı fərqlə bağlı Q.Kazımov yazır: “Sözün leksik mənası ilə yanaşı, semantik mənası da var. Sözün leksik və semantik mənaları nə qədər yaxın olsa da, müəyyən dərəcədə fərqlənir. Semantik məna nitq və ya mətn şəraitində sözün leksik məna əsasında qazandığı mənadır [Kazımov, Morfolojiya, 2010, s.16]. Sonra Q.Kazımov görkəmli dilçi-alim Y.Seyidova istinad edərək qeyd edir ki, semantik məna leksik mənaya tabedir, ondan doğur. Semantik məna leksik mənadan müxtəlif yol və vasitələrlə əmələ gələn məna incəlikləridir. Leksik məna məfhumun bilavasitə ifadəçisidir. Semantik məna məfhumla bağlı olsa da, onun müstəqil ifadəçisi deyil. Ona görə “leksik semantika” ifadəsini işlədirik ki, sözün leksik mənası geniş anlayışdır, semantik məna leksik mənənin içindədir... O, leksik mənədən ayrı deyildir, ondan kənarda yaşaya bilmir [Kazımov, Morfolojiya, 2010, s.16].

Leksik məna konkret olur, konkret bir əşyanın adını bildirir. Qrammatik məna ümumi, mücərrəd olur. Qrammatik məna başqa sözlərlə əlaqədə meydana çıxır və tək-tək sözlərə deyil, böyük söz qruplarına aid olan ümumiləşmiş, mücərrədləşmiş mənədir.

Məsələn, “kitab” sözünün leksik mənası belə izah olunur: üzərində hər hansı bir mətn olan vərəqlərin bir araya gətirilərək

tikilib cildlənmiş nəşriyyat növü [Azərbaycan dilinin izahlı lüğəti, II cild, 2006, s.715].

Qrammatik mənə baxımından isə bu söz ad bildirir, konkret isimdir, təkdir, adlıq haldadır, cansızdır və s.

Deməli, sözün semantikasi, tam, dəqiq mənası başqa sözlərlə əlaqə və münasibətdə məlum olur. Bu dəqiqləşmə sözün leksik və qrammatik mənalarına da aid ola bilər. Məsələn, bizdən “sarı” sözünün mənasını soruşsalar, ayrılıqda onun nə mənada işləndiyini izah edə bilmərik. Bunun üçün həmin sözün kontekstdə, cümlədə işlənməsi vacib şərtidir. Deməli, “sarı” sözü başqa sözlərlə əlaqə və münasibətdə izah oluna bilər, onun leksik və qrammatik mənası dəqiqləşdirilə bilər:

Örəyi sarı, gəl mənə sarı,

Tikanı yox eylə, yaramı sarı.

Bu Azərbaycan xalq mahnısında “sarı” sözü üç müxtəlif mənada işlənmişdir: sarı (rəng), sarı (qoşma, istiqamət bildirən), sarı (feil, “sarımaq” sözündən). Göründüyü kimi, sözün dəqiq mənası (leksik və qrammatik) digər sözlərlə əlaqə və münasibətdə, kontekstdə məlum olur.

Azərbaycan dilinin izahlı lüğətində nümunələr əsasında bu və ya digər sözün onlarla müxtəlif mənaları göstərilir. Bu mənalar (məna çalarları) göstərilmiş misallarda (kontekstlərdə), digər sözlərlə əlaqə və münasibətlərdə konkretləşir, dəqiqləşir.

Bizim tədqiqatda semantika dedikdə korpus mətnlərinin avtomatik semantik işarələnməsi prosesində sırf formal əlamətlər əsasında söz-formaların (sözlərin) mənalarının dəqiq müəyyənləşdirilməsi (leksik və qrammatik) nəzərdə tutulur.

Dil səviyyələri bir-birindən təcrid olunmuş şəkildə deyildir. Əksinə, dilin müxtəlif səviyyələri qarşılıqlı əlaqə və münasibətdədirlər. Aşağı səviyyənin çatışmazlıqları daha yuxarı səviyyələrdə tamamlanmış, dəqiqləşdirilmiş olur. Şübhəsiz, təbii dilin səviyyələrə bölünməsi şərtidir və dilin mənimsənilməsi, onun mahiyyətinin dərkinə xidmət edir. Lakin, nəzərə almaq

lazımdır ki, bütövlükdə dil sistemdir və şərti olaraq bölünmüş səviyyə çərçivəsində qapanıb qalmır.

Məlum olduğu kimi, təbii mətnin emalı və maşın tərcüməsi sistemlərində linqvistik təhlil morfoloji səviyyədən başlanır. Morfemlər haqqında (kök və şəkilçi) ətraflı məlumat əldə etdikdən sonra sintaktik təhlil mərhələsi başlanır. Çünki sintaktik təhlil üçün qabaqcadan nitq hissələri və sözlərin qrammatik kateqoriyaları məlum olmalıdır. Sintaktik təhlildə cümlə mətn çərçivəsində mətndən və digər cümlələrdən ayrılır, onu təşkil edən söz-formalar arasında sintaktik əlaqələr müəyyənləşdirilir, cümlə üzvləri, uzlaşma əlaqəsi araşdırılır. İstər morfoloji, istərsə də sintaktik səviyyələrdəki “həll olunmuş problemlər” semantik blokda araşdırılır. Sözlərin semantik təbiəti heç də həmişə morfoloji və sintaktik əlamətlərlə səciyyələnmişdir. Morfoloji və sintaktik səviyyələr əslində sözün tam mənasını açmaqda semantik bloka yardım edirlər. Maraqlıdır ki, bəzən dilin morfoloqiyasını və sintaksisini bilmədən də kontekstdəki sözlərin mənasına görə onların qrammatik əlaqələrini bərpa etmək mümkün olur. Kontekstdəki sözlərin mənalara görə cümlə konstruksiyalarını da qurmaq mümkündür. Buradan belə nəticə çıxarmaq olar ki, təbii ünsiyyət prosesində morfoloqiya sintaksisə, oradan da semantikaya deyil, əksinə semantikadan birbaşa sintaksisə keçid mümkündür. Burada insanın həyat təcrübəsi, intuisiya və bir çox başqa ekstralingvistik amillər köməyə gəlir.

Kompüterlərdə canlı insana xas olan bu keyfiyyətlər olmadığı üçün semantikanı morfoloqiya və sintaktik səviyələrsiz təsəvvür etmək qeyri-mümkündür.

Nəzərə almaq lazımdır ki, insanın təhlil qabiliyyətini tam təqlid edə bilən təhlil modelləri qurmaq mürəkkəb məsələdir. Belə model tam şəkildə semantikaya əsaslanmalı, semantik kateqoriyalarla işləməyi, nəticə çıxarmağı bacarmalıdır.

Semantik problemin həlli, hər şeydən öncə, semantik vahidlərin ayrılması meyarı, onların statusu və s. bu kimi məsələlərin aydınlaşdırılması zərurətini doğurur. Mətnlərin tam mənasını açmağa xidmət edən semantik vahidlərin müəyyənləşdirilməsi və onların mətndəki yerini və vəzifəsini araşdırmaq üçün müəyyən qaydalar işləyib hazırlamaq lazım gəlir.

Fərz edək ki, dilin lüğət fondunu kifayət qədər təmsil edən sözlük (sözlərin siyahısı) və bu sözlüyü təşkil edən ayrı-ayrı sözlərin sintaktik əhatəsi (idarə, uzlaşma və s. sintaktik əlaqələrlə bir-birinə bağlanan söz birləşmələri) məlumdur. Sintaktik birləşmələrin sayı və təsviri dil sistemi haqqında tam təsəvvür yaratmağa imkan verir.

Mətnin distributiv təhlili prosesində sintaktik birləşmələr üçün elementar semantik diferensial əlaməti tapmaq mümkündür. Konkret nümunələrə baxaq.

Əgər biz hər hansı bir sintaktik bağlılığı olan söz birləşməsi götürüb onun bir komponentini dəyişən hesab edərək yerinə müxtəlif sözlər yazsaq, müşahidə etmək olar ki, heç də bütün sözlər sintaktik baxımdan məqbul və düzgün hesab olunan birləşmə əmələ gətirmirlər. Məsələn, üçüncü növ təyini söz birləşməsi olan filmin sonu sintaktik vahidinin ikinci tərəfini sabit saxlamaqla bir çox başqa eyni əlaqə əsasında bağlanmış birləşmələr almaq olar: ömrün sonu, kitabın sonu, sevginin sonu və s. Lakin formal cəhətdən tam eyniyyət təşkil edən elə birləşmələr də tapmaq mümkündür ki, sintaktik-semantik baxımdan onlar məqbul sayıla bilməz: buğdanın sonu, dodağın sonu, boy-buxunun sonu və s. Buradan belə nəticə çıxarmaq olar ki, eyni distribusiyaya malik sözlər formal əlamətlərdən başqa, müəyyən ümumi semantik əlamətlərə də malikdirlər. Eyni distribusiyaya malik sözlər heç də həmişə eyni sintaktik-semantik əlaqələrə girə bilmirlər, baxmayaraq ki, onlar eyni elementar diferensial əlamət daşıyırlar.

Elementar diferensial əlamət bir çox sözlərin məna elementlərinin tərkibinə daxil ola bilər. Lakin hər bir sözə yalnız bir diferensial əlamət yazmaq özünü bütün hallarda doğrultmur. Bu və ya digər söz müxtəlif distibusiyalarda, müxtəlif sintaktik-semantik əlaqə formalarında olur. Bəzən bir distribusiya başqa bir distribusiya ilə izah olunur və bu məsələni bir qədər də çətinləşdirir.

Elementar diferensial əlamətin göstərilməsi bizə nə verə bilər?

Lügət vahidi kimi sözlərin səciyyələndirilməsində qeyd olunmuşdu ki, sözlərin həqiqi mənası sintaktik əlaqədə məlum olur. Bu baxımdan sözün həqiqi mənası bəzən sözün ənənəvi mənası ilə üst-üstə düşməyə bilər. Müəyyən diferensial mənaya malik ən kiçik vahid kimi monosema götürülə bilər. Monosemaya eyni məna yükünə malik morfem və ya morfem ardıcılığı kimi baxmaq olar. Morfemlərin eyni və ya müxtəlif məna kəsb etmək xüsusiyyəti diferensial əlamətlər əsasında araşdırıla bilər. Bəzən müəyyən kontekst daxilində söz birləşməsi də həmişə eyni diferensial əlamət çərçivəsində işləyə bilər. Məsələn, söz birləşməsi bəhsində “təyini söz birləşməsi” eyni məna bildirir və bu qəbildən olan sözləri avtomatik referatlaşdırma prosesində makrosema hesab etmək olar. Formal təhlil sistemlərində sözlərin feildən asılı olaraq qruplara ayrılması prosesini makrosemalara bölünmə kimi qəbul etmək olar.

Mətni məlumatların avtomatik işlənməsi sistemlərində təhlilin semantik problemlərindən bəhs edərkən semantik inventarlaşdırılma məsələləri ön plana çəkilir. Semantik inventarlaşdırma iki əsas istiqamətdə aparılır:

- leksik-semantik səviyyələrdə inventarlaşdırma;
- morfo-semantik səviyyələrdə inventarlaşdırma.

Göstərilən səviyyələrdə paradigmatik və sintaqmatik aspektlər hökmən nəzərə alınır. Semantik inventarlaşdırma prose-

sində söz və morfemlərin, onların müəyyən birləşmələrinin tezauslarının tərtibi mühüm amil hesab olunur.

Omonimliyin və çoxmənalılığın aradan qaldırılması alqoritmlərinin tərtibində sintaktik əlaqə və tezauslar sistemindən istifadə olunmuşdur.

Qeyd etmək lazımdır ki, nəinki ayrıca götürülmüş söz və ya söz birləşməsi, hətta bütövlükdə cümlənin də mənası tam şəkildə kontekstdə açılır. Ona görə də, mənlərin avtomatik işlənməsi ilə bağlı mətn semantikasi məsələlərinin hərtərəfli araşdırılması mühüm əhəmiyyət kəsb edir. Yeni nəsillə kompüterlərin getdikcə daha da inkişafı və təkmilləşdirilməsi sistemin semantik blokunun yeni baxımdan izahı, təkmilləşdirilməsi və qurulması imkanı yaradır.

Morfoloji səviyyədə əsas və şəkilçilərin omonimliyini qismən aradan qaldırmaq mümkündür. Morfoloji səviyyədə, kontekstdə götürülməmiş hər hansı bir omonim əsasən hansı nitq hissəsinə aid olduğunu göstərmək qeyri-mümkündür. Bu və ya digər nitq vahidinin hansı nitq hissəsinə mənsubluğu daha yuxarı dil səviyyələrində müxtəlif metod və üsullardan istifadə olunmaqla müəyyən edilə bilər. Gödelin natamamlıq hipotezinin linqvistik interpretasiyasını versək, belə bir nəticəyə gəlmək olar ki, hər hansı bir sistemi öz daxilində (çərçivəsində və ya səviyyəsində), onun öz elementləri ilə izah etmək qeyri-mümkündür.

Sistemin hər hansı bir elementinin tam izahı mütləq həmin sistem çərçivəsindən kənar verilməlidir. Yəni, fonetik səviyyədə öz həllini tapmayan hər hansı bir linqvistik problem morfoloji səviyyədə, morfoloji səviyyədə izah olunmayan linqvistik vahidlər isə sintaktik və semantik səviyyələrdə şərh oluna bilər. Bu baxımdan morfoloji səviyyədə aradan qaldırılmalı bilməyən omonimlik sintaktik və semantik blokların müdaxiləsindən sonra öz izahını tapa bilər. Gödelin nata-

mamlıq fərziyyəsi bu baxımdan linqvistika üçün əhəmiyyətlidir [Mahmudov, 1994].

Əsasların və şəkilçilərin omonimliyi avtomatik lüğətdə və şəkilçilər siyahısında öz əksini tapır. Məsələn, əgər əsas omonimdirsə, onda onun mümkün nitq hissələrinə mənsubluğu qeyd olunur: yaz – feil/isim. Yaz sözünün hansı nitq hissəsinə mənsubluğunu linqvistik təhlilin daha yüksək – sintaktik və semantik səviyyələrində tam aydınlaşdırmaq mümkündür, çünki omonimlik və çoxmənəlilik ilk növbədə semantik problemlər çərçivəsində həll olunmalıdır.

Bununla bərabər, nəzərə almaq lazımdır ki, əsasların omonimliyi onlara birləşdirilən şəkilçilərə görə də qismən aradan qaldırıla bilər. Şəkilçilərin qrammatik xüsusiyyətlərinə görə əsas haqqında müəyyən məlumat əldə etmək olar. Məsələn, at – şəkilçisiz işləndikdə həm feil, həm də isim kimi izah oluna bilər. Şəkilçilərlə işləndikdə at əsası aşağıdakı formaları ala bilər:

- at – atır, atdı, atacaq və s.
- at – atda, atımız, atlara və s.

Birinci halda at əsası zaman şəkilçilərini qəbul edir ki, bu da əsasın feil olduğunu göstərir.

İkinci halda at – hal və mənsubiyyət şəkilçiləri ilə işlənir. Bu şəkilçiləri isə yalnız ad nitq hissəsinə aid olan sözlər qəbul edirlər. Beləliklə, nümunələrdən görüldüyü kimi, əsaslara birləşən şəkilçilərə görə onların omonimliyini aradan qaldırmaq mümkündür.

Azərbaycan dilində elə omonim şəkilçilərə də rast gəlmək olur ki, onlar həm feillərə, həm də adlara qoşularaq müxtəlif qrammatik funksiyalar yerinə yetirirlər. Məsələn, -ım (-im, -um, -üm) şəkilçiləri feil əsaslarından sonra əmr şəklini, isimlərdən sonra isə mənsubiyyəti göstərir.

- yazım – feilin əmr şəkli, I şəxs, tək;
- kitabım – mənsubiyyət, I şəxs, tək.

Belə hallarda şəkilçilərin omonimliyi əsasa görə aradan qaldırılır. Yuxarıda göstərilənləri nəzərə alaraq omonimliyin aradan qaldırılması üçün formal qaydalar işlənib hazırlanmışdır [Mahmudov, 2002, s.202-215].

Korpusda əsas və şəkilçilərin düzgün işarələndirilməsi üçün bunun əhəmiyyəti böyükdür.

Omonimliyin formal əlamətlərə görə aradan qaldırılması prosesində iki cəhəti fərqləndirmək lazımdır:

- şəkilçilərə görə əsasların omonimliyinin aradan qaldırılması;
- əsaslar vasitəsi ilə şəkilçilərin omonimliyinin aradan qaldırılması.

Əsasların omonimliyinin aradan qaldırılması üçün aşağıdakı formal qaydaları qəbul edirik:

Əgər əsas omonim (AF), şəkilçi isə ad nitq hissəsinə aiddirsə (a), onda əsas da ad nitq hissəsinə aiddir (A).

Burada A – ad nitq hissəsini, F – feili, a – ad nitq hissəsinə, f – isə feil nitq hissəsinə aid şəkilçiləri göstərir.

Əgər əsas omonim (AF), şəkilçi isə feilə aiddirsə (f), onda əsas feil nitq hissəsinə aiddir (F).

Əgər həm əsas (AF), həm də şəkilçi (af) omonimdirsə, onda əsasın omonimliyi morfoloji səviyyədə aradan qaldırıla bilməz. Bu qayda əsasdan sonra yalnız omonim şəkilçi (şəkilçilər) gəldikdə doğrudur. Əgər omonim şəkilçidən (şəkilçilərdən) sonra yenə şəkilçi (şəkilçilər) gəlsə, onda omonimliyi aradan qaldırmaq üçün omonim şəkilçidən sonra gələn şəkilçiyə müraciət etmək lazım gəlir.

Əgər mətndə omonim əsasdan sonra yalnız bir omonim şəkilçi gəlsə, onda omonimlik aradan qalxmır və əsas üçün onun bütün mümkün qrammatik səciiyyələri qeyd olunur.

Analoji olaraq, şəkilçilərin omonimliyinin də aradan qaldırılması üçün formal qaydalar vermək olar:

Əgər affiks omonimdirsə (af), əsas ad nitq hissəsinə aiddirsə (A), onda şəkilçi də adlara (a) aiddir.

Əgər şəkilçi omonimdirsə (af), əsas isə feildirsə (F), onda şəkilçi də feil nitq hissəsinə aiddir.

Əgər həm şəkilçi (af) şəkilçinin, həm də əsas omonimdirsə (AF), onda şəkilçinin omonimliyi morfoloji səviyyədə aradan qaldırıla bilməz (bu o halda doğrudur ki, şəkilçi yeganədir, ondan sonra omonim şəkilçi gəlmir. Əgər şəkilçidən sonra yenə şəkilçi gəlsə və həmin şəkilçi omonim deyilsə, onun hansı nitq hissəsinə mənsubluğuna görə omonimlik aradan qaldırıla bilər).

Yuxarıda göstərilənləri cədvəl şəklində vermək olar.

Morfoloji səviyyədə əsas və şəkilçilərin omonimliyinə aradan qaldırılması təkcə yuxarıda göstərilən qaydalarla məhdudlaşmır. Morfoloji səviyyədə omonimliyin aradan qaldırılmasının digər istiqamətlərini nəzərdən keçirək:

omonimlik çox vaxt ünsiyyət prosesində yanaşı işlənən şəkilçi birləşmələrinin köməyi ilə də aradan qaldırıla bilər. Belə birləşmələri məşin affiksləri də hesab etmək olar.

Məsələn, feili sifət şəkilçisi -yacağımız -yacaq və -ımız şəkilçilərinin birləşməsindən ibarətdir. Birləşmənin birinci tərəfi -yacaq feilin qəti gələcək zamanı, ikinci tərəfi -ımız isə adlara (isimlərə) xas olan mənsubiyyət şəkilçisidir. Birləşmə şəklində bu şəkilçi feili sifəti göstərir.

-dığ – mənsubiyyət şəkilçisi tipli feili sifət şəkilçilərinə gəldikdə isə, bu şəkilçinin komponentlərinə ayrılmış şəkildə verilməsi daha məqsədəuyğun hesab olunur. Qrammatika kitablarında təsbit olunmasa da, -dığ (-duq, -diy, -düy) məşin şəkilçisi kimi verilmişdir. İstisna olaraq bu qəbildən olan şəkilçilərdən -dığın (-duğun, -diyın, -düyün) birləşmə şəklində verilir. Qrammatik səciyyələri: feili sifət, II şəxs, tək. Bu şəkilçinin komponentlərinə ayrılması təhlil prosesini bir qədər də mürəkkəbləşdirə bilər.

Mənsubiyyət şəkilçisi+hal şəkilçisi birləşməsi, məsələn, -sının (-sunun, -sinin, -sünün) omonimliyin aradan qaldırılması ilə bağlı qəbul olunmuşdur. Formal təhlil sistemlərində bu şəkilçilərin bir hissəsi -sın (-sun, -sin, -sün) analogi formaya malik əmr şəklinin şəkilçiləri ilə qarışdırıla bilər. Analogi olaraq qeyri-qəti gələcəklə keçmiş zaman şəkilçiləri də birləşmə şəklində verilmişdir: məsələn, -armış (-ərmiş, -yarmış, -yərmiş), -ardı (-ərdi, -yardı, -yərdi) və s.

Maşın şəkilçiləri vasitəsi ilə omonimliyin aradan qaldırılması. Əsas və şəkilçilərin omonimliyi həmçinin ənənəvi morfoloqiyalarda öz əksini tapmamış xüsusi maşın şəkilçiləri vasitəsi ilə də aradan qaldırıla bilər:

-mır (-mur, -mir, -mür). İnkər şəkilçisi -ma (-mə) ilə indiki zaman şəkilçilərinin -ır (-ir, -ur, -ür) birləşməsi nəticəsində yaranır.

-mısan (-musan, -misən, -müsən). Keçmiş zaman şəkilçisi -mış (-muş, -miş, -müş) ilə II şəxsin təkini bildirən -san (-sən) şəkilçisinin birləşməsi ilə düzəlir. Burada birinci komponentindəki -ş samitinin düşməsi nəticəsində qalan -mı (-mu, -mi, -mü) sual şəkilçisi ilə üst-üstə düşə bilər.

idi və imiş bağlamalarının və ilə qoşmasının şəkilçi formaları da müstəqil şəkilçi kimi siyahıda təmsil olunmuşdur: -ydı (-ydi), -ymış (-ymış), -yla (-ylə).

Bütün yuxarıda sadalananlar şəkilçilər siyahısında qeyd olunur.

Morfoloji səviyyədə aradan qaldırıla bilinməyən omonimlik. Gödelin natamamlıq teoreminin linqvistik interpretasiyası ilə bağlı qeyd etmişdik ki, qapalı bir dil səviyyəsində həll edilməsi qeyri-mümkün olan hər hansı bir problem daha yüksək səviyyələrdə öz həllini tapır.

Aşağıdakı hallarda şəkilçi və əsasların omonimliyi aradan qaldırıla bilməz:

əgər omonim əsas şəkilçisiz işlənirsə:

at – feil/ad;

1.3 və 2.3 hallarında – əsas və şəkilçi eyni vaxtda omonim olduqda:

məsələn, at (feil/ad) +-ım (əmr və ya mənsubiyyət, I şəxs, tək);

atım – (mənim) atım (ad) və ya atım (feil).

Bu mənaların hansının doğru olduğu yalnız kontekst daxilində, təhlilin daha yüksək – sintaktik və semantik mərhələlərində müəyyənləşdirilə bilər.

Şəkilçi siyahısında həm feili, həm də ismi əsaslara qoşula biləcək 4 omonim şəkilçi qeydə alınmışdır. Omonim şəkilçilər üçün bütün mümkün qrammatik əlamətlər qeydə alınır və dördüncü mövqedə bu əlamətlərin feili və ismi əsaslar üçün doğru olduğu göstərilir. Məsələn, -ım (-um, -im, -üm) şəkilçisi BALφ kimi kodlaşdırılmışdır. Kontekst daxilində nəzərdən keçirildikdə cədvəldə qeydə alınmış ÜALQ və RALC kodlarından birini alır. Bu kodlar göstərir ki, əgər bu şəkilçi feili əsasa birləşmişsə, onda əmr şəklinin I şəxs, tək halını (ÜALQ), ismi əsadan sonra gəlibsə, mənsubiyyət şəkilçisi, I şəxs, tək halını (RALC) göstərir. Dördüncü mövqedə gələn koda əsaslanaraq bu şəkilçinin hər bir əsasa müvafiq qrammatik səciyyələri seçilir. Beləliklə, şəkilçinin omonimliyi aradan qaldırılmış olur.

Yuxarıda göstərilənlər əsasında əsas və şəkilçilərin omonimliyinə aradan qaldırılması alqoritmi qurulur. Bu zaman nəzərə almaq lazımdır ki, omonimliyin aradan qaldırılması və düzgün variantın seçilməsinə xidmət edən alqoritmlər hər bir ayrılmış morfem (əsas və şəkilçi) üçün müəyyən qrammatik məlumat əldə edildikdən sonra alınır.

Əsasların omonimliyinə aradan qaldırılması alqoritmi:

Əsadan sonra gələn şəkilçinin mövcudluğunun yoxlanması: əgər şəkilçi varsa 2-yə, əks halda 7-yə keçməli;

Əsadan sonra gələn şəkilçini ayırmalı və 3-ə keçdi;

Yoxlamalı: əgər şəkilçi omonimdirsə 1-ə, əks təqdirdə 4-ə keçməli;

Yoxlamalı: əgər şəkilçi adlara aiddirsə 5-ə, əks halda 6-ya keçməli;

Əsas adlara aiddir, 8-ə keçməli;

Əsas feillərə aiddir, 8-ə keçməli;

Morfoloji səviyyədə əsasın omonimliyi aradan qaldırıla bilməz – sintaktik və semantik bloklara müraciət;

Əsas haqqında müvafiq qrammatik informasiyanın işçi sahəyə yazılması.

Nümunə olaraq atımdan söz-formasının təhlilini nəzərdən keçirək. Morfoloji təhlil prosesində atımdan aşağıdakı morfemlərə ayrılmış olur:

at-ım-dan. Alınmış at əsasının omonimliyinin aradan qaldırılması tələb olunur. Bu aşağıdakı kimi yerinə yetirilir:

At – əsasından sonra şəkilçi olub-olmadığını yoxlayırıq. Konkret misala görə şəkilçi gəldiyi üçün 2-yə keçirik.

At – əsasından sonra gələn şəkilçini (-ım) ayırıb 3-ə keçirik.

Yoxlayırıq: şəkilçi omonimdirmi? Bizim misalda -ım şəkilçisi omonimdir. Onda 1-ə keçirik. 1-ə görə əsasdan sonra yenə şəkilçi gəlib-gəlmədiyini müəyyənləşdiririk. Əgər -ım şəkilçisi omonim əsasdan sonra gələn yeganə şəkilçi olsaydı, onda alqoritmə görə 7-yə keçməli idik. Morfoloji səviyyədə əsasın omonimliyi aradan qaldırılmadığı üçün sintaktik və semantik bloklara müraciət etmək lazım gələrdi. Lakin omonim əsasdan sonra (at) gələn omonim şəkilçidən (-ım) sonra yenə şəkilçi gəldiyi üçün (-dan) 2-yə keçirik.

2-yə görə omonim şəkilçidən (-ım) sonra gələn şəkilçini (-dan) ayırıb yenidən 3-ə keçirik.

3-ə görə müəyyənləşdiririk: yeni ayırdığımız şəkilçi (-dan) omonimdirmi? Əgər şəkilçi yenə omonim olsaydı, 1-ə keçib

növbəti şəkilçinin mövcudluğunu yoxlamalı idik. Lakin -dan şəkilçisi omonim olmadığından şərtə əsasən 4-ə keçirik.

4-ə görə şəkilçinin adlara və feilə aid olduğunu yoxlayırıq. Şəkilçi feilə aid olsaydı, 6-ya keçməli idik. Lakin şəkilçi adlara aid olduğu üçün 5-ə keçirik.

5-in şərtinə görə əsasın adlara aid olduğu müəyyənləşdirilir və 9-a keçilir.

8-ə görə əsas haqqında müvafiq qrammatik informasiya işçi sahəyə yazılır: atımdan – söz-formasında at – əsasının qrammatik səciyyələri aşağıdakılardır:

at – isim, adlıq hal, tək, canlı və s.

Əsasların omonimliyi kimi təhlil prosesində şəkilçilərin omonimliyi halları da geniş müşahidə olunur. Formal əlamətlərə görə morfoloji səviyyə çərçivəsində şəkilçilərin omonimliyi də aradan qaldırıla bilər [Mahmudov, 2013, s.306-308].

Şəkilçilərinin omonimliyinə aradan qaldırılması alqoritmi:

Yoxlamalı: əgər şəkilçiyə birləşmiş əsas omonimdirsə, 5-ə, əks halda 2-yə keçməli;

Yoxlamalı: əgər əsas ad nitq hissəsinə aiddirsə, 3-ə, əks halda 4-ə keçməli;

Şəkilçi adlara aiddirsə, 6-ya keçməli;

Şəkilçi feilə aiddirsə, 6-ya keçməli;

Şəkilçinin omonimliyi morfoloji səviyyədə aradan qaldırılmaz;

Omonim şəkilçi haqqında qrammatik məlumatın işçi sahəyə yazılması.

Məsələn, kitabım söz-formasındakı -ım şəkilçisinin yuxarıda göstərilən alqoritmənin köməyi ilə omonimliyinə aradan qaldırılması mərhələlərini nəzərdən keçirək. Morfoloji təhlilin ilk pilləsində söz-forma əsas (kök) və şəkilçiyə ayrılır:

– kitab-ım

Alqoritmədə nəzərdə tutulduğu kimi yoxlayırıq: şəkilçiyə birləşən əsas omonimdirmi? Əgər əsas omonim olsaydı, 5-ə

keçməli idik. Əsas (kitab) omonim olmadığı üçün 2-yə keçirik. 2-nin şərtinə görə əsasın (kitab) ad nitq hissəsinə aid olub-olmaması yoxlanılır. Əsas ad nitq hissəsinə (isim) aid olduğu üçün 3-ə keçirik.

Əsas isim olduğu üçün qaydalara görə şəkilçi də isim olmalıdır. Beləliklə, -ım omonim şəkilçisinin ad nitq hissəsi ilə işləndikdə malik olduğu qrammatik səciyyələr müəyyənləşdirilir və işçi sahəyə yazılır:

-ım – mənsubiyyət şəkilçisi, I şəxs, tək və s. [Mahmudov, 2013, s.308-309].

Tezauruslar vasitəsi ilə omonimliyin və çoxmənalılığın aradan qaldırılması. Mətnlərin avtomatik işlənməsi sistemlərində çoxmənalı sözlərin və söz birləşmələrinin müəyyənləşdirilməsi, cümlədə ayrı-ayrı ifadələrin düzgün variantının seçilməsi sistemin keyfiyyətini səciyyələndirən mühüm şərtlərdəndir. Bu məsələ ilə bağlı çoxkomponentli söz birləşmələrinin çoxmənalılığının aradan qaldırılmasına xüsusi diqqət yetirilməlidir. Türkcə-rusca maşın tərcüməsi sistemində ictimai-siyasi mövzuda mətnlərin təhlili göstərir ki, bu qəbildən olan sistemlərdə mürəkkəb terminoloji söz birləşmələri geniş surətdə işlənir. Təcrübə göstərir ki, morfoloji səviyyədə çoxmənalılığın həlli bir çox hallarda qeyri-mümkün olur.

Sintaktik səviyyədə də çoxmənalılıq heç də həmişə “uğurla” aradan qaldırılmaz. Linqvistik vahidlər birləşmə tərkibində müxtəlif mənalarda işlənə bilərlər.

Tezauruslar bu müxtəlif mənalardan düzgün variantın seçilməsinə imkan verir. Misal olaraq aşağıdakı iki söz birləşməsinə nəzərdən keçirək:

Müasir istehsal vasitələri;

Beynəlxalq əmək təşkilatı.

Əgər mətdə söhbət müasir istehsal vasitələri və ya beynəlxalq əmək təşkilatlarından gedirsə, bu halda istehsal vasitələri və əmək təşkilatı korrektivlər, müasir istehsal və

beynəlxalq əmək birləşmələri isə qeyri-korrektivlərdir. Bu sonuncu birləşmələr tamamilə başqa mənada işlənilir.

Sintaktik əlaqəli elementlər çoxluğundan düzgün variantın müəyyənəşdirilməsi, semantik əlaqələri olan tərcümə ekvivalentlərinin seçilməsi formal təhlil sistemlərində tezauruslar vasitəsilə həyata keçirilir.

Hər şeydən əvvəl tezauruslar bu və ya digər elm sahəsinə aid termin vahidləri arasında həmin sahəyə aid müvafiq semantik münasibətlərin aydınlaşdırılmasına xidmət edir. Tezaurusların ikinci mühüm funksiyası – təhlil olunan elm sahəsinə aid struktur əlaqələri özündə əks etdirən semantik əlaqələrin aydın təsviridir. Bu xüsusiyyət formal təhlil və MT üçün mühüm, həlledici əhəmiyyət kəsb edir.

Tədqiq olunan elm sahəsinin semantik modeli kimi tezauruslar dil vahidlərinin (söz, söz birləşməsi) məndəki konkret mənalarının açılmasına xidmət edir.

Formal təhlil sistemində tezaurusların istifadə olunması perspektivlərini ictimai-siyasi leksika əsasında tərtib edilmiş Avtomatik lüğətin sözlüyü üzərində əyani göstərmək olar. Bunun üçün göstərilən konkret sahədə komponentlər arasında real münasibətləri səciyyələndirən semantik-statistik əlaqələri əsas götürülmüş sahə tezaurusunun yaradılması lazım gəlir. İctimai-siyasi terminologiyanın mənə əlaqələrinin formalaşması üçün termin fərqləndirilmələrindən istifadə olunmuşdur.

Tezaurus modeli kimi semantik şəbəkə götürülür. Həmin şəbəkənin zirvəsində sözlərin leksik mənaları, yanlarında isə sözlər arasındakı semantik-paradiqmatik münasibətlər qarşılaşdırılır. Semantik şəbəkənin tərtibi üçün aşağıdakı metodika əsas götürülür. Qrafın yanlarında semantik derivatlar, başqa sözlə desək, sözlər və onların vasitəsiz semantik komponentləri qarşılaşdırılır. Bu proses əksinə də götürülə bilər, yəni sözlərlə onların vasitəsiz semantik derivatlarının münasibətləri də əsas götürülə bilər [Мамедова, 1984; Мамедова, Пинес, 1989].

Belə hesab olunur ki, başqa sözün təyin olunmasında iştirak edən söz sonuncunun vasitəsiz komponenti hesab olunur. Beləliklə, əgər bu söz başqasını təyin edən sözün tərkibinə daxildirsə, onda müvafiq zirvələr birincidən ikinciyə doğru istiqamətdə birləşdirilir. Belə sadələşdirilmiş meyar semantik şəbəkələrin avtomatik qurulmasını asanlaşdırır.

Vasitəsiz və ya vasitəli semantik derivasiya əlaqələri ilə bağlı olan, eləcə də ümumi semantik komponentlərə malik sözlər semantik cəhətdən bağlı sözlər hesab olunur.

Vasitəsiz semantik derivasiya qrafı əsasında qurulmuş tezauruslar hər bir linqvistik vahid üçün həm vasitəsiz, həm də vasitəli əlaqələrin aydınlaşdırılmasına imkan verir. Vasitəsiz semantik derivasiya əlaqələri semantik əlaqələrin istənilən derivasiyasını mümkün edir. Tezaurus “cins – növ”, “tam – hissə”, istənilən iyerarxiya, assosiativ münasibətlər, ekvivalentlik münasibətləri (sinonimlik) və s. əsasında qurula bilər.

Hansı məqsəddə xidmət edəcəyi və məzmunu nəzərə alınaraq tezauruslar müxtəlif prinsiplərə müvafiq qurulur. Formal təhlil sistemlərində çoxmənəlilikin aradan qaldırılması, düzgün ekvivalentin seçilməsi ilə əlaqədar tezauruslar uğurla istifadə oluna bilər.

VI FƏSİL

MAŞIN TƏRCÜMƏSİ VƏ SÜNİ İNTELLEKT

Formal təhlil sistemlərindən danışarkən maşın tərcüməsi (MT) məsələlərinə toxunmamaq mümkün deyil. Problemin həlli ilə bağlı müxtəlif vaxtlarda böyük sıçrayışların, uğurların qazanıldığı vaxtlarda da, durğunluq və pessimist əhval-ruhiyyənin üstünlük qazandığı məqamlarda da bu məsələnin elmi əhəmiyyəti heç vaxt azalmamışdır. MT problemi ilə bağlı tədqiqatlar dilçiliyin bir çox yeni, müstəqil sahələrinin inkişafına təkan vermiş, dilçiliyin bəzi məsələlərinə yeni baxış bucağından baxılmasına gətirib çıxarmışdır. Müxtəlif dillərdə təhlil və sintez alqoritmlərinin tərtibi, milli dillərin maşın korpusunun hazırlanması, məlumat axtarışı sistemlərinin inkişafı, statistik leksikoqrafiyanın təşəkkülü də ilkin olaraq MT problemləri çərçivəsində baş vermişdir. MT probleminin müxtəlif aspektlərinə aid tədqiqatlar ədəbiyyat siyahısında öz əksini tapmışdır [Helюбин, 1991; Пиотровский, 1975; 1979; Марчук, 2002; Махмудов, Пиотровский, 2001; Махмудов, 2002; 2013 və s.]. Araşdırmamızda MT probleminin bəzi ümumi cəhətlərini nəzərdən keçirək.

Müasir dövrdə beynəlxalq əlaqələrin getdikcə intensivləşməsi, elmin sürətli inkişafı nəticəsində elmi-texniki ədəbiyyatın onlarla xarici dilə qısa müddət ərzində tərcüməsi zəruri tələbatə çevrilmişdir. Qeyd etmək lazımdır ki, elmi-texniki məlumatların avtomatik axtarışı, mətnlərin müxtəlif məna əmələgətirmə xüsusiyyətlərinin tədqiqi, annotasiyaların, referatların, icmalların və s. avtomatik alınması kimi məsələlər də MT problemləri ilə sıx əlaqədardır. Hazırda xüsusi tərcüməçi kadrların yetişdirilməsinə böyük diqqət yetirilməsi,

tərcümə büroları və mərkəzlərinin geniş fəaliyyət göstərməsi, onların beynəlxalq konqreslərinin çağırılması və xüsusi jurnallarının nəşri geniş vüsət almışdır. Bütün bunlar təsadüfi deyildir. Tərcümə ədəbiyyatının nəşri durmadan artır, lakin dünyanın müxtəlif xalqlarının dilində nəşr olunan elmi-texniki ədəbiyyatın çox cüzi bir hissəsini tərcümə etmək mümkün olur. Məhz bu səbəbdən tərcümə prosesinin avtomatlaşdırılması ideyası meydana gəlmişdir.

Maşın tərcüməsi – xüsusi kompüter proqramlarının köməyi ilə yazılı mətnlərin və şifahi nitqin bir dildən digərinə tərcüməsi prosesidir. Eyni zamanda belə sistemlərin yaradılması ilə bağlı elmi tədqiqat istiqaməti də maşın tərcüməsi adlanır.

Maşın tərcüməsi prosesində insan-maşın münasibətlərinin təşkili formaları müxtəlif olur:

- sonradan redaktə – ilkin mətn maşında emal olunur, sonradan isə insan-redaktor tərəfindən redaktə olunur;
- əvvəlcədən redaktə – insan mətni maşında emal olunmaq üçün hazırlayır (mümkün variantlardan uyğun gələnini seçir, mətni sadələşdirir və işarələyir), bundan sonra mətnin proqram üzrə emalı başlanır;
- proses zamanı redaktə – insan qarşıya çıxan çətinlikləri həll etmək üçün tərcümə sisteminin işinə müdaxilə edir;
- qarışıq sistemlər (məsələn, əvvəlcədən və sonradan redaktə aparmaqla).

Çox vaxt “maşın tərcüməsi” və “avtomatik tərcümə” terminləri eyni anlamda işlədilir. Lakin avtomatlaşdırılmış tərcümə termininin mənası tamamilə fərqlidir. Belə tərcümə prosesində proqram insanın mətni tərcümə etməsinə kömək göstərir. Tərcümə insan tərəfindən aparılır, müxtəlif təyinatlı tərcümə proqramları isə bu işdə insan-tərcüməçiyə yardımçı, köməkçi funksiyasını həyata keçirir.

Avtomatlaşdırılmış tərcümə aşağıdakı formalarda ola bilər:

- qismən avtomatlaşdırılmış tərcümə – məsələn, insan-tərcüməçi tərəfindən kompüter lüğətlərindən istifadə olunmaqla edilən tərcümə;

- bölgü sistemi ilə tərcümə – kompüter yalnız qabaqcadan verilmiş strukturda olan ifadələri tərcümə üçün nəzərdə tutulur, bu sxemə daxil olmayan qalan söz və ifadələrin tərcüməsi isə insanın öhdəsinə düşür.

İngilisdilli terminologiyada machine translation, MT (tam avtomatik tərcümə) və machine-aided və ya machine-assisted translation (MAT) (avtomatlaşdırılmış tərcümə) terminləri fərqləndirilir. Hər iki termini ümumiləşdirib bildirmək istəyəndə isə M (A)T yazılır.

Maşın tərcüməsi alqoritmləri bir-birindən kəskin fərqlənən iki yanaşma üsulu ilə tərtib olunur. Birinci yanaşma qaydalara əsaslanır (rule-based) və ənənəvi olaraq əksər maşın tərcüməsi sistemlərində istifadə olunur (PIPOMT – Rusiya,, SYSTRAN – Fransa, Linguatex – Almaniya və b.). İkinci yanaşma statistik və ya statistikaya əsaslanan sistem hesab olunur (statistical-based). Bu sistemdən Yandeks, Google və ABBYY populyar servisləri yararlanır.

Statistik maşın tərcüməsi. Statistik maşın tərcüməsi – böyük həcmərdə dil cütlərinin müqayisəsinə əsaslanan mətnlərin maşın tərcüməsi növlərindən biri hesab olunur. Dil cütləri – bir dildə yazılmış cümlələrdən ibarət mətnlər və ona digər dildə uyğun gələn cümlələrdir. Belə mətnlər iki dil daşıyıcısı olan (hər iki dili mükəmməl bilən) şəxs tərəfindən hazırlana bilər və ya insan-tərcüməçi tərəfindən hər hansı dildə yazılmış cümlələrin digər dilə tərcüməsindən ibarət cümlələr ola bilər. Beləliklə, statistik maşın tərcüməsi “özüöyrənən” xüsusiyyətə malik sistemdir. Dil cütlərinin sayı nə qədər çox olarsa və onların bir-birinə müvafiqlik dərəcəsi nə qədər yüksək olarsa, statistik maşın tərcüməsinin nəticələri də bir o qədər yüksək olar.

“Statistik maşın tərcüməsi” anlayışı altında tərcümə probleminə ümumi yanaşma nəzərdə tutulur. Belə yanaşma ikidilli mətnlər korpusundan alınmış məlumatlardan istifadə etməklə cümlənin ən çox ehtimal olunan tərcümə variantının axtarışına əsaslanır. İkidilli mətnlər toplusuna nümunə olaraq parlamentdəki debatların protokollarını əks etdirən parlament hesabatlarını göstərmək olar. İkidilli parlament hesabatları Kanadada, Honkonqda və başqa ölkələrdə çap olunur. Avropa İqtisadi Birliyinin rəsmi sənədləri 11 dildə çap olunur. Birləşmiş Millətlər Təşkilatının sənədləri bir neçə dildə çap edilir. Bu çap məhsulları statistik maşın tərcüməsi üçün qiymətli resurs rolunu oynayır.

Maşın tərcüməsinin tarixi. Elektron hesablama maşınlarının tərcümədə istifadəsi ideyası ilk dəfə 1947-ci ildə ABŞ-da ilk elektron hesablayıcı maşınlar yarandığı dövrdə irəli sürülmüşdür. Maşın tərcüməsinin ilk kütləvi nümayişi (Corcaun təcrübəsi) 1954-cü ildə baş vermişdir. Bu sistem primitiv olmasına baxmayaraq (250 sözdən ibarət lüğət, 6 qaydadan ibarət qrammatika, bir neçə sadə söz birləşməsinin tərcüməsi) bu təcrübə çox geniş rezonans doğurdu. İngiltərə, Bolqariya, Almaniya (keçmiş Almaniya Demokratik Respublikası), İtaliya, Çin, Fransa, Yaponiya və bir çox başqa ölkələrdə bu istiqamətdə tədqiqatlara başlandı. 1954-cü ildə keçmiş SSRİ-də də maşın tərcüməsi sahəsində tədqiqatlara start verildi.

1960-cı ilin ortalarında ABŞ-da praktik istifadə üçün iki rus-İngilis tərcümə sistemi təqdim olundu:

MARK (BBC ABŞ Xarici texnika departamenti);

GAT (Corcaun Universitetinin layihəsi Okric Milli Atom Enerjisi mərkəzində və İtaliyanın İspira şəhərində Avropa Atom Mərkəzində istifadə olunurdu.

Lakin belə sistemlərin qiymətləndirilməsi üçün yaradılmış ALPAC komissiyası mətnlərin maşın tərcüməsi nümu-

nələrinin aşağı keyfiyyətdə olmasını əsas götürərək ABŞ-da bu fəaliyyət növünün rentabelli olmadığı qərarına gəldi. Komissiya bu sahədə nəzəri tədqiqatların davam etdirilməsi və dərinləşdirilməsini tövsiyə etsə də, bütövlükdə onun gəldiyi nəticələr pessimizmə, maliyyələşmənin azaldılmasına, çox vaxt isə bu mövzuda işlərin tamam dayandırılmasına aparıb çıxardı.

Bütün bunlara baxmayaraq, bir sıra ölkələrdə tədqiqatlar davam etdirilirdi, hesablama texnikasının durmadan təkmilləşməsi və inkişafı da buna imkan yaradırdı. Minikompyuterlərin və fərdi kompüterlərin yaradılması, onunla yanaşı təbii dillərin emalına istiqamətlənmiş lüğət, axtarış və o qəbildən olan digər mürəkkəb sistemlərin işlənilib hazırlanması həlledici amil rolunu oynamağa başladı. Beynəlxalq əlaqələrin genişlənməsi ilə əlaqədar maşın tərcüməsinə də tələbat artmış oldu. Sürətli və keyfiyyətli tərcümə zamanın tələbinə çevrildi. Bütün bunlar 1970-ci illərin ortalarından başlayaraq bu sahənin yeni yüksəlişinə səbəb oldu. 1980-ci illərdə tərcümə sistemlərinin praktik istifadəsinə başlandı, bu sahədə kommersiya layihələri bazarı inkişaf etməyə başladı.

Maşın tərcüməsinə bəşəriyyətin bəslədiyi ümid və arzular indinin özündə də müəyyən dərəcədə arzu olaraq qalır. Geniş tematikalı keyfiyyətli tərcümə hələ də əlçatmaz hesab olunur. Amma, maşın tərcüməsi sistemindən istifadə, 1980-ci illərin qiymətləndirməsinə görə, tərcüməçinin işini beş dəfəyə qədər sürətləndirir.

Hazırda maşın tərcüməsi sahəsində bir çox kommersiya layihələri mövcuddur. Maşın tərcüməsi sahəsində pionerlərdən biri Systran kompaniyası hesab olunur. Rusiyada maşın tərcüməsi sahəsində prof. R.Q.Piotrovskinin rəhbərlik etdiyi qrupun xidmətlərini xüsusi qeyd etmək olar (A.İ.Qertsen adına Rusiya Dövlət Pedaqoji Universiteti, Sankt-Peterburq).

Dilmanc Azərbaycanda ilk maşın tərcüməsi sistemidir. Azərbaycanca-İngiliscə, İngiliscə-azərbaycanca və türkcə-azər-

baycanca maşın tərcüməsi sahəsində tanınmış maşın tərcüməsi qrupudur. Dilmanc maşın tərcüməsi sistemində alınmış tərcümə nümunələrinin və ayrı-ayrı sözlərin düzgün tələffüzü də səsləndirilə bilər. Dilmanc maşın tərcüməsi 2003-cü ildən fəaliyyətə başlamışdır (<http://www.dilmanc.az/az>).

Tərcümənin keyfiyyəti. Tərcümənin keyfiyyəti mətnin mövzu dairəsindən və üslubundan, eləcə də tərcümə olunan dillərin qrammatik, sintaktik və leksik baxımdan qohumluq dərəcəsiindən asılıdır. Bədii mətnlərin maşın tərcüməsi praktik olaraq çox vaxt keyfiyyətsiz olur. Texniki mətnlərin və sənədlərin tərcüməsi isə nisbətən fərqli səciyyə daşıyır. Texniki mətnin aid olduğu sahəyə aid terminoloji lüğət olduqda və tərcümə sisteminin tərcümə olunacaq mətnlərin xüsusiyyətlərinə uyğunlaşdırılması sonradan cüzi redaktə olunmaqla məqbul tərcümə nümunələri almaq imkanı yarada bilər. Tərcümə olunacaq mətn nə qədər çox formallaşdırılmışsa, tərcümənin keyfiyyəti bir o qədər keyfiyyətli ola bilər. Texniki və rəsmi-işgüzar üslubda yazılmış mətnlərin maşın tərcüməsi prosesində uğurlu nəticələr alınabilir [az.wikipedia.org/wiki/Maşın_tərcüməsi].

Məlum olduğu kimi, avtomatik tərcümənin mümkünlüyü ideyası ilk dəfə 1946-cı ildə Amerika alimləri A.But və U.Uiver tərəfindən söylənmişdir [Кулагина, 1979]. Onlar göstərmişlər ki, müharibə illərində düşmənin gizli kodlarının şifrəsini açmaq məqsədi ilə hazırlanmış elektron qurğular vasitəsilə bir dildən başqa dilə tərcümə mümkündür. 1949-cu ildə U.Uiverin məşhur “Tərcümə” memorandumu nəşr olunduqdan sonra bu sahədə intensiv tədqiqat işləri aparılmağa başlandı. Kompüterlər vasitəsilə ilk təcrübi tərcümə (rus dilindən ingilis dilinə) 1954-cü ildə ABŞ-da həyata keçirilmişdir. Elə həmin ildən keçmiş Sovet İttifaqında bu sahədə tədqiqatlara başlanmış, 1956-cı ildə Moskvada fransız və ingilis dillərindən rus dilinə tərcümə nümunələri alınmışdır.

Hazırda MT sahəsində dünyanın bir çox ölkələrində tədqiqat işləri aparılır. Bu məsələ ilə əlaqədar nəşr olunan elmi-tədqiqat əsərlərinin sayı sürətlə artmaqdadır. Bir dildən başqa dilə tərcümə prosesinə kompüterlərin cəlb olunması dilçilik elmi üçün təkcə təcrübi deyil, həm də nəzəri əhəmiyyətə malikdir. Bu, adi şəraitdə aşkara çıxarıla bilməyən, dilçi-tədqiqatçının aydın şəkildə, bütün detalları ilə təsvir edə bilmədiyini tərcümə prosesinin modelləşdirilməsi üçün geniş yol açır. MT sahəsində respublikamızda və keçmiş SSRİ-də aparılan tədqiqat işlərini iki mühüm qola ayırmaq olar:

– tətbiqi (MT-nin geniş miqyasda və dairədə reallaşdırılması);

– nəzəri (tədqiqat metodlarından biri kimi insanın nitq fəaliyyətinin modelləşdirilməsi, linqvistik təsvirlər üçün riyazi modellərin işlənməsi, dil obyektləri alqoritmlərinin işlənilməsi, dil hazırlanması ilə bağlı araşdırmalar, insanın beyin fəaliyyəti ilə məşin arasında münasibətin öyrənilməsi, süni intellekt istiqamətində tədqiqatlar və s.).

MT ilə bağlı tədqiqat işləri ilk günlərdən indiyə qədər elmi-texniki mətnlərin tərcüməsi ətrafında mərkəzləşmişdir. Çünki elmi-texniki ədəbiyyatın tərcüməsi xüsusi ilə zəruridir. Burada daha mühüm bir cəhət diqqəti cəlb edir: elmi-texniki mətnlər bədii mətnlərə nisbətən tərcümə üçün daha əlverişlidir (sözlərin eynicinsliliyi, idiomların demək olar ki, işlənməməsi, sintaktik quruluşun sadəliyi, lüğət tərkibinin qismən məhdudluğu, daha çox terminlərin işlənməsi və s.). Bundan başqa, elmi-texniki mətnlərdə müəllifin fərdi üslubi xüsusiyyətləri, demək olar ki, yox dərəcəsində olur. Bədii əsərlər isə texniki mətnlərə nisbətən tərcümə prosesində daha çox çətinliklər doğurur. Müxtəlif idiomatik ifadələr, frazeoloji birləşmələr belə mətnlərin tərcüməsi zamanı standart vasitələrdən istifadəni qeyri-mümkün edir. Bədii tərcümədə yazıçının fərdi üslubunun saxlanması mühüm və zəruri şərtlərdəndir. Bədii ədəbiyyat

nümunələrinin (xüsusən poeziyanın) məşınların köməyi ilə tərcüməsi haqqında söz açmaq elmin indiki inkişaf səviyyəsində məqsədəuyğun deyil.

MT bir çox əməliyyatları özündə birləşdirən alqoritmlərin tərtibi ilə reallaşdırılır. Sonradan bu alqoritmlərin köməyi ilə məşının girişinə (yaddaşına) verilmiş mətn üzərində müəyyən ardıcılıqla əməliyyatlar aparmaqla nəticədə tərcümə olunmuş mətn alınır.

MT-nin ən sadə növü sözbəsöz tərcümədir. Kompüterə iki sözlük, məsələn Azərbaycan və rus dilinin sözlüyü verilir. Hər bir rus sözünün qarşısında onun Azərbaycan dilində olan ekvivalentinin ünvanı qeyd olunur. Bu ünvanın köməyi ilə məşın hər bir rus sözünə uyğun Azərbaycan sözünü tapır və alınmış tərcüməni çapa verir.

Təcrübə göstərir ki, bir neçə qrup dillər üçün sözbəsöz tərcümə vasitəsi ilə kafi nəticə əldə etmək olar. Lakin belə “qaba” tərcümə təbii ki, müasir tərcümə qarşısında qoyulan tələblərə tamamilə cavab vermir: əvvələn, dildə özünü göstərən polisemiya, omonimlik, hər bir sözə situasiyadan asılı olaraq bir neçə sözün uyğun gəlməsi orijinaldan uzaq mənaların alınmasına aparıb çıxarır. İkincisi, sözbəsöz tərcümədə idiom və frazeologizmlər nəzərə alınmır. Bundan başqa, sözbəsöz tərcümədə tərcümə olunan cümlədəki söz sırası eyni ilə saxlanır. Bu, tipoloji cəhətdən müxtəlif söz sırasına, sintaktik quruluşa malik dillərdə (məsələn, rus, ingilis və Azərbaycan dilləri) sözbəsöz tərcüməni çətinləşdirir. Şübhəsiz, bu, MT probleminin həllinin ilk dövrləri üçün səciyyəvi idi. Sonralar problemin həlli ilə bağlı effektiv sistemlərin işlənilib hazırlanması sahəsində aparılan axtarışlar bir dil məlumatının digərinə çevrilməsi prosesinin daha da mürəkkəbləşməsinə aparıb çıxartdı. Məşına daxil olunan dil materiallarının təhlili və nəticədə digər dildən tərcümənin sintezi müstəqil bloklara

ayrılır. Öz növbəsində təhlil və sintez də ayrılıqda lüğət və qrammatika bloklarına ayrılır.

Dünyanın bir çox ölkələrində, eləcə də keçmiş SSRİ-nin elmi-tədqiqat mərkəzlərində fəaliyyət göstərmiş kollektivlərin MT sahəsində apardıqları axtarışları əsas götürərək prosesin gedişini aşağıdakı mərhələlərə bölmək olar:

- daxil edilən dildə mətnin təhlili (sözün avtomatik lüğət-də axtarışı, mətnin “qavranması” üçün modelləşdirmə, morfoloji və sintaktik təhlil);

- çevrilmə (mətnin daxil olunan dildəki quruluşundan çıxışda alınan dildəki quruluşuna keçid);

- mətnin çıxışda alınan dildə sintezi (mətnin quruluşu modelinin verilməsi, sintaktik və morfoloji formalaşma).

MT-nin hazırda fəaliyyət göstərən sistemlərində bu göstərilən mərhələlər müəyyən düzəlişlər və əlavələr edilməklə nəzərə alınmışdır.

Belə bir fikir də əsaslandırılır ki, təhlil qrammatikaları terminləri ilə ifadə olunmuş giriş mətnlərinin sintez qrammatikaları terminləri ilə ifadə olunmuş çıxış mətnlərinə çevrilməsi aralıq-keçid dilinin (vasitəçi dil) köməyi ilə yerinə yetirilə bilər. Təhlillə sintez arasında vasitəçi rolunu oynayan aralıq keçid dili bütün dil qrupları üçün yararlı olan simvolik məna yazılışına və universal sintaktik münasibətlərə malikdir.

MT probleminin həlli ilə əlaqədar təkcə ABŞ-da deyil, keçmiş SSRİ-nin bir çox elmi-tədqiqat mərkəzlərində də xeyli iş görülmüşdür. MT-nin coşqun inkişaf sürətinə baxmayaraq, əvvəllər olduqca asan və qısa müddətdə həll olunacaq bir problem kimi təsəvvür olunan tərcümə məsələsi 50-ci illərin axırında böhrana uğradı. Lakin bu sahədə tədqiqat işləri dayandırılmadı. Əksinə, MT ilə məşğul olan kollektivlərin sayı və görülən işlərin həcmi getdikcə artmağa başladı. Əlbəttə, heç kəs MT-nin maraqlı və olduqca perspektivli bir bilik sahəsi olduğunu inkar etmir. Lakin bir çox tədqiqatçılar belə bir

müddəanı əsaslandırmağa çalışırlar ki, hazırda hətta elmi-texniki ədəbiyyatın belə tamamilə avtomatikləşmiş, yüksək keyfiyyətli tərcüməsini almaq mümkün deyil. Aydın olmuşdur ki, bütün konkret dil aspektlərini əhatə edən dilin tam avtomatik modelləri olmadan MT-nin keyfiyyətinin yüksəldilməsi qeyri-mümkündür. Keyfiyyətin aşağı olmasının digər səbəbi tərcüməçiyə xas olan ekstralinqvistik biliklərin məşində olmamasıdır. Hesablayıcı texnikanın “baş çıxara” bilmədiyi sintaktik və semantik omonimlik məsələsini insan məhz belə biliklərin köməyi ilə araşdırıb müəyyənləşdirə bilər. Hazırda MT üçün ya tam avtomatlaşdırılmış aşağı keyfiyyətli tərcümə, ya da qismən avtomatlaşdırılmış yüksək keyfiyyətli tərcümə nümunələri səciyyəvidir.

MT sahəsində görkəmli mütəxəssis V.İnqve problemin həlli ilə bağlı göstərir: “Maşın tərcüməsi sahəsində aparılan işlər semantik sədd qarşısındadır... Biz belə bir müddəanı anlamağa başlamışıq ki, maşın tərcüməsində yeganə metod ondan ibarətdir ki, maşın nəyi tərcümə etdiyini “qavramalıdır”. Lakin bu, əlbəttə, çox çətin vəzifədir” [Ингве, 1965].

MT-nin böhrandan çıxması ilə əlaqədar İ.İ.Revzinin fikirləri də maraqlıdır:

- avtomatik tərcümə ilə əlaqədar bütün məsələləri təcili həll etməyə imkan verəcək qlobal konsepsiyalardan imtina etmək;

- eyni bir məsələnin həllində müxtəlif prosedurların ardıcıl tətbiqinə çalışmaq;

- semantikanın formallaşdırılması;

- dil modelləri nəzəriyyəsinin daha da inkişaf etdirilməsi və bu nəzəriyyənin avtomatik təhlillə əlaqələndirilməsi və s. [Резвин, 1968].

MT sahəsində Qrenobl (Fransa) elmi-tədqiqat mərkəzinin və digər müasir sistemlərin təhlili aşağıdakı metodoloji nəticələri çıxarmağa imkan verir:

- MT sistemində təhlil və sintez bir-birindən tamamilə ayrı olmalıdır;
- sistemdə bir neçə xüsusi semantik səviyyə, başqa sözlə desək, semantik təhlil, semantik sintez mərhələləri olmalıdır. Avtomatik təhlilin tam olması üçün kifayət qədər semantik məlumat maşına daxil edilməlidir;
- sistemin mütləq komponenti zəngin məlumatlara malik (leksik-qrammatik-semantik) avtomatik lüğətlərin olmasıdır;
- sistem bütün dil səviyyələrində müfəssəl giriş və çıxış qrammatikalarına əsaslanmalıdır;
- sistemin zəngin xidmət proqramları kitabxanasına ehtiyacı var (müxtəlif düzəlişlər aparmaq, sistemin daxili vəziyyəti haqqında məlumatların alınması və s.)

Bir çox tədqiqatçılar MT probleminin nəzəri deyil, mühəndis işi səciyyəsinə olmasa fikrini irəli sürürlər. Onlar belə hesab edirlər ki, MT üçün zəruri olan lingvistik məlumatlar nəzəri olmaqdan daha çox təcrübədirlər və konkret məsələlərin həlli zamanı alına bilərlər. Belə ki, keçmiş ümumittifaq “Nitq statistikasi” qrupunun rəhbəri R.Q.Piotrovski aşağıdakıları təklif edir:

- 100 faizli müstəqil semantik MT ideyasından əl çəkmək lazımdır və “İnsan-maşın-insan” sistemi ilə həyata keçirilən MT metodunu işləmək lazımdır. Bunun üçün kompüterlərdə ilkin mərhələ kimi əvvəlcə qaba tərcümə nümunəsi almağa imkan verən müəyyən alqoritmləri, standart, tez-tez təkrar olunan əməliyyatları yerinə yetirmək lazımdır.
- tam formallaşdırmadan imtina etmək və dilin konkret, müxtəlif növlü sistemlərinin (məs., funksional üslubları) tamamlanmış təsvirini ifadə edən dil qrupları bazasının təşkilinə diqqət yetirmək lazımdır. Belə qruplarda nəinki sistem, hətta ehtimal olunan norma da gözlənilməlidir. Bu o deməkdir

ki, yeni yaradılmış dil qrupları daha işlək və bərabər paylanmış dil vahidlərini özündə birləşdirməlidir.

Göstərilən tezislərə müvafiq olaraq Sankt-Peterburq (Leningrad) “Nıtq statistikası” qrupunun MT-nin reallaşdırılması proqramı hazırlanmışdır. Həmin proqramı aşağıdakı mərhələlərə bölmək olar:

- dil vahidlərinin qruplaşdırılması, müəyyən dil müxtəlifliklərini əhatə edən mətnlərin statistik təsviri;
- söz-forma, söz və tərkiblərin ikidilli və ya çoxdilli avtomatik lüğətlərinin tərtibi (leksik tərcümə);
- morfoloji-sintaktik MT-nin təşkili (qrammatik tərcümə);
- leksik omonimliyin və çoxmənalılığın aradan qaldırılması proqramlarının yaradılması (semantik MT);
- vahid MT sisteminin yaradılması [Пиотровский, 1975].

Qeyd etmək lazımdır ki, son zamanlar yarımavtomatik tərcümə də böyük maraq və əhəmiyyət kəsb etməyə başlamışdır. Mətnlərin avtomatik işlənməsi sistemləri ilə əlaqədar yarımavtomatik tərcümə məsələlərinin araşdırılması çox vacibdir. Yarımavtomatik tərcümədə kompüterlər köməkçi vasitə rolunu oynayır, tərcüməçini ağır, az məhsuldar və yaradıcı olmayan işlərdən azad etmək üçün maşınların köməyi ilə tərtib olunmuş xüsusi lüğətlərdən istifadə olunur. Proses aşağıdakı kimi həyata keçir: tərcüməçi mətndəki naməlum sözləri qeydə alır, həmin sözlər lüğət formasına gətirilib maşına daxil edilir. Maşında qabaqcadan yerləşdirilmiş lüğətlə müqayisə edildikdən sonra sözlərin lüğəvi mənası tapılır və nəticədə alınmış mətn ona yazılmış qlossari ilə birlikdə tərcüməçiyə verilir. Tərcüməçi sözü lüğətdə axtarmaq kimi uzun vaxt aparan bir əməliyyatdan xilas olmuş olur və az vaxt ərzində bütün naməlum sözlərin mənasını tapır. Məsələnin bir qədər də dərinlən işlənməsi ilə yaradılmış sistem sözün və onun daxil olduğu kontekstin maşın lüğətində axtarışına imkan verir. Məsələ orasıdadır ki, sözün tam, dolğun və dəqiq mənası kon-

tekstdə məlum olur və istər adi, istər yarımavtomatik tərcümədə bu amili nəzərə almamaq qeyri-dəqiq nəticələr alınmasına aparıb çıxarır. Tərcüməçi mətndə qeydə alınmış naməlum sözü deyil, onun əhatəsində olan sözləri kompüterə daxil edir, nəticədə geniş kontekstlərdə terminin mümkün tərcümə variantları alınır.

Yarımavtomatik tərcümədə maşın tərcümə etmir – tərcümə insanın özü tərəfindən yerinə yetirilir. Lakin alınmış tərcümə avtomatik sistemə daxil olduğu üçün, tədricən axtarılan sözün müxtəlif kontekstlərdə işlənmə xüsusiyyətləri təhlil olunur və lazımı məna əldə edilir. Qeyd etmək lazımdır ki, müasir sistemlərdə sorğunun qabaqcadan işlənməsinə (məsələn, söz-formanın lüğət formasına salınmasına) heç bir ehtiyac yoxdur. Sistem naməlum sözləri müstəqil olaraq lüğət vahidi formasına salan və sonradan onu lüğətdə axtaran morfoloji təhlil proqramları ilə tənzim olunur.

Süni intellekt maşın tərcüməsi sistemlərində. İntellekt elə bir mürəkkəb sistemdir ki, biz onun köməyi ilə əvvəlcədən məlum olmayan bir məsələnin həlli yollarını tapırıq. İnsan intellekti də, kibernetik sistemli süni intellekt də bu istiqamətə yönəlmişdir. Biz insan intellektini bu və ya digər dərəcədə təqlid edən süni intellekti təbii intellektə uyğun qurmağa cəhd edirik. Bəzən buna qismən nail olur, bəzən isə müəyyən nəticələr əldə etməyin yollarını yenidən axtarmalı oluruq.

Maşın tərcüməsi sistemlərində süni intellekt mətnin hər-tərəfli semantik təhlili tələbi ilə üzleşir. Bu isə formal təhlil qanunlarına əsaslanan sistemlərdə çətin maneə olaraq qalır. Problemin həllinə müxtəlif yanaşmalar var. Bəziləri, ümumiyyətlə, insan intellektinə yaxın bioloji assosiativ yaddaşa malik kompüterlərin yaranmasına qədər semantik təhlilin mümkün-lüyünü inkar edirlər.

Paradoksal haldır ki, bəzən süni intellektin qarşısına ağ-lasığmaz tələblər qoyulur. İnsan intellektinin həll edə bilmədiyi

və ya həll etməyə cəhd göstərdiyi məsələlərin optimal müddətdə həlli süni intellektə həvalə olunur. İstənilən nəticə alınmadıqda süni intellekt sistemlərinin yararsızlığı qənaətinə gəlinir. Maşın tərcüməsi sahəsində də vəziyyət belə olaraq qalmaqdadır.

Tərcümə nəyə əsaslanır? İstər insan tərəfindən edilən tərcümədə, istərsə də maşın tərcüməsində (fərqi yoxdur, yazılı və ya şifahi tərcümə), hər şeydən əvvəl, tərcümə ediləcək mətn (və ya nitq parçası) dərk olunmalı, qavranmalıdır. İnsan hiss orqanları (qulaq, göz, mimika və s.) ilə qəbul edib qavramadığı informasiyanı tərcümə edə bilməz. Maşın da mətni (nitqi) informasiya kimi texniki vasitələrin köməyi ilə qəbul etməli, “qavramalı”, sonra tərcümə mərhələsinə keçməlidir.

Qavrama prosesi insanda və maşında fərqli şəkildə baş verir. İnsan mətni assosiativ olaraq beynində əvvəlcədən mövcud olan ssenari, freym, müəyyən anlayış və biliyin köməyi ilə dərk edir. Əgər belə öncədən olan yardımçı vasitələr olmazsa, insan informasiyanı (yazılı və şifahi) dərk edə bilməz. Maşının linqvistik informasiyanı dərk edə bilməsi özündə linqvistik bazanın (avtomatik lüğət, təhlil və sintez alqoritmləri, formal tanınma sistemləri və s.) mövcudluğu mühüm şərtidir. İnsan sivilizasiyasından, cəmiyyətindən kənar bir mühitdə yaşayan bir fərd mobil telefonun nə olduğunu dərk edə bilmədiyini kimi, maşın da linqvistik baza, qabaqcadan müəyyən hazırlıq olmadan informasiyanı dərk edə bilməz.

Məsələnin başqa tərəfini müzakirə edək. Hər hansı xarici dili bilən insan istənilən mətni tərcümə edə bilərmidi? Hətta tərcümə etsə belə, tərcümə nə dərəcədə keyfiyyətli olacaq? Hər hansı xarici dili (fərz edək ki, ingilis dilini) bilən şəxsin riyazi biliyi və savadı yoxdursa, riyazi analizə aid mətni qavrayıb tərcümə edə bilərmidi? Əsla yox. Çünki onun beynində riyaziyyata aid hazır ssenarilər, freymilər, anlayışlar, müəyyən bilik yoxdur.

Süni intellekt sistemlərinə əsaslanan maşın tərcüməsində də vəziyyət eynidir. Riyaziyyata (və ya hər hansı konkret elm sahəsinə) aid linqvistik bazası, terminoloji lüğəti, müəyyən formul və işarələr bazası olmayan tərcümə sistemi həmin sahədə tərcümə apara bilməz. Bəzən maşın tərcüməsi sistemlərindən istənilən sahəyə aid dəqiq tərcümə tələb olunur.

İnsan-tərcüməçi müəyyən ixtisas sahəsində ixtisaslaşdığı kimi, maşın tərcüməsi sistemləri də konkret elm sahələri üzrə ixtisaslaşmış olur. Əgər tərcüməçi-insan bütün elm sahələrinə aid mətnləri eyni səviyyədə tərcümə etmək iqtidarında deyilsə, bunu maşın tərcüməsi sistemlərindən niyə tələb etməliyik?!

Maşın tərcüməsi sistemləri insan tərcüməçini təqlid edir, onun kimi “qavramağa”, təhlil aparmağa cəhd edirdi. Məlum olduğu kimi, əvvəllər sözbəsöz tərcümədən istifadə olunurdu. Mətnə iki boşluq (probel) arasındakı hərflər (qrafemlər) ardıcılığı avtomatik lüğətlə (maşın lüğəti) tutuşdurulur və sözün (söz-formanın) başqa dilə tərcüməsi bu qayda ilə alınmış hesab olunurdu. Məsələn: “tanış etmək”, “tanış” – «знакомый», “etmək” – «делать», «сделать» və s. Sonra məlum oldu ki, insan-tərcüməçi bu tipli ifadələri ayrı-ayrılıqda deyil, söz birləşməsi şəklində tərcümə etməyə cəhd göstərir. Yalnız belə olduqda “tanış etmək” və ya “tanış olmaq” ifadəsi düzgün tərcümə oluna bilər. Belə situasiyalarda anlaşılmaz tərcümə nümunəsi alındıqda və ya tərcümə zəif alındanda, başa düşülməyəndə yenə insan-tərcüməçiyə müraciət edilir, tərcümənin uğursuz alınmasının səbəbləri araşdırılır. Bir vaxtlar, maşın tərcüməsinin təşəkkül dövründə sözbəsöz tərcümə daha geniş yayılmışdı. Yuxarıda qeyd olunduğu kimi, sözbəsöz tərcümədə söz birləşməsinin hər bir komponenti ayrılıqda başqa dilə tərcümə olunurdu. Təbiidir ki, belə tərcümə məqbul hesab oluna bilməzdi. Müasir tərcümə sistemlərində söz birləşmələri (sabit və sərbəst söz birləşmələri, frazeoloji birləşmələr) bütövlükdə, söz birləşməsi kimi tərcümə

olunur: “tanış etmək” – «познакомить», «ознакомить», “tanış olmaq” – «познакомиться» və s. Tərcümə prosesi bu qayda ilə tədricən təkmilləşir, dəqiqləşir.

Son vaxtlar mətn dilçiliyi sürətlə inkişaf etməkdədir. Mətn dilçiliyi, mətnin qavranması və təhlili süni intellekt sistemlərinə də öz təsirini göstərməmiş deyil. Mətn dilçiliyi mətnin (və ya nitqin) tərcüməsi məsələlərinə yeni baxış bucağı altında baxmaq zərurəti yaradır.

Məhz bu baxımdan mətni anlama modellərinin qurulması, onların maşın tərcüməsi sistemlərində aprobeşiyası mühüm əhəmiyyət kəsb edir və tərcümənin sərhədləri bir qədər də genişlənməmiş olur. Sözü mənəsi cümlədə tam aydınlaşdığı kimi, cümlədə mətn səviyyəsində daha dəqiq, daha düzgün anlaşılır.

Türk dillərinin maşın tərcüməsi sistemi. Türk dillərinin materialları əsasında maşın tərcüməsinin aktual problemlərinə, mətnlərin kompüterlərdə müxtəlif məqsədlərlə işlənməsi məsələlərinə, leksik, morfoloji, sintaktik-semantik səviyyələrdə türk mətnlərinin işlənməsinə aid çoxlu tədqiqatlar aparılmışdır [Вах: Ахундов, 1979; Бабанаров, 1981; Пинес, 1970; Рахманов, 1988; Вəliyeva, 1999; Махмудов, 1991; Махмудов, 2002; 2013; Жубанов, 2013 və b.]

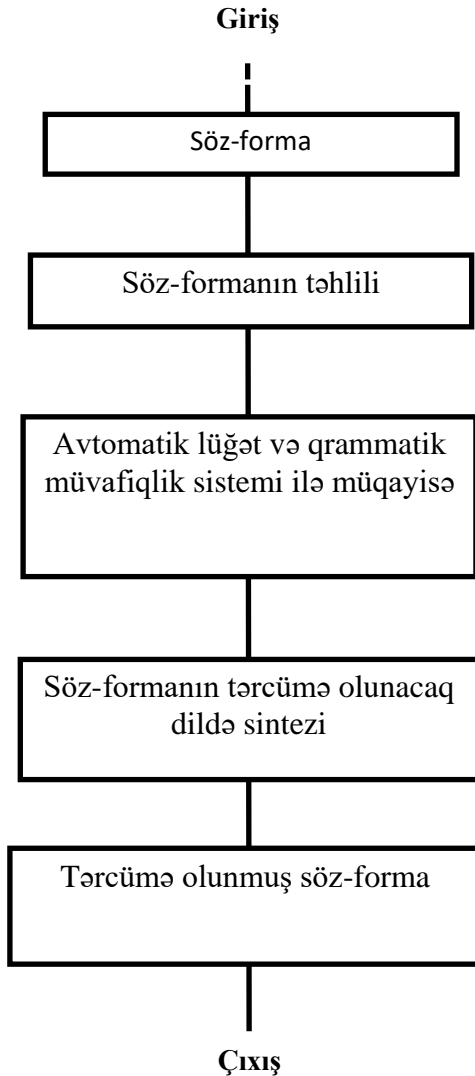
Yuxarıda qeyd edildi ki, istər avtomatik lüğətlər, istər kompüterlərdə istifadə olunmuş digər lüğətlər müxtəlif məqsədlərə müvafiq olaraq fərqli qaydalar üzrə tərtib olunur. Leksik bazanın təhlil sistemlərində müstəsna əhəmiyyətini nəzərə alaraq Azərbaycan dilindən başqa dillərə tərcümənin ən ümumi prinsipləri ilə tanış olaq.

Tərcümə olunmaq üçün nəzərdə tutulmuş hər hansı mətn (daha çox elmi-texniki mətnlər tədqiqata cəlb edilir) kompüterə daxil edilir. İşin sonrakı mərhələsi həmin mətnin maşında “qavranması” ilə davam edir. Ayrı-ayrı cümlələr mətndən və bir-birindən ayrılır. Formal cəhətdən cümlə dedikdə, iki nöqtə,

nöqtə və nöqtəli vergül, nöqtə və nida işarəsi, nöqtə və sual işarəsi arasında yerləşən söz-formalar zənciri başa düşülür. Göründüyü kimi, ilkin mərhələdə cümlənin digər mühüm cəhətləri (kommunikativ xüsusiyyətləri, cümlə üzvləri, onların leksik-semantik mənaları və s.) nəzərə alınmır. Cümlələr bir-birindən ayrıldıqdan sonra hər bir cümlə daha kiçik struktur vahidlərə bölünür. Bu ayırmada da əsas meyar kimi formal əlamətlər götürülür. Nəticədə cümlə söz-formalara ayrılır. Mövcud mətnlərin avtomatik işlənmə sistemlərində söz-forma əsas anlayışlardan biridir.

Söz-forma cümlənin (bütövlükdə mətnin) formal olaraq müəyyənləşdirilən ən kiçik vahidi sayılır. Söz-forma yazılı mətnlərdə iki boşluq (probel) arasında olan ixtiyari hərflər zənciri (ardıcılığı) hesab olunur. Sonrakı mərhələlərdə alqoritmlərin köməyi ilə söz-formaların daha kiçik hissələrinə ayrılması və onların təhlili prosesi baş verir. Söz-forma əsas və şəkildələrə ayrılır.

Əsas dedikdə, kök və ya kök birləşməsi ilə sözdüzəldici şəkildənin birləşməsi başa düşülür. Mövcud tədqiqatlarda MT və təhlil sistemlərinin tələblərinə uyğun olaraq sözün əsası daha kiçik mənalı vahidlərə ayrılır, çünki əsas ayrılıqda leksik vahid kimi bütün lüğətlərdə təmsil olunur. Lakin bu o demək deyildir ki, əsasın daha kiçik vahidlərə ayrılması qeyri-mümkündür və ya elmi maraq doğurmur. Sözdüzəldici şəkildələrin və söz-köklərinin özləri də avtomatik işlənmə sistemlərinin tədqiqat obyektinə kimi dərindən araşdırıla bilər. Başqa dillərdən fərqli olaraq türk dilləri üçün MT və mətnlərin müxtəlif işlənmə sistemləri işlənilib hazırlanarkən buna lüzum olmamışdır.



Sxem 1. Leksik-morfoloji səviyyədə söz-formaların tərcüməsinin ümumi sxemi

Məsələn, tutaq ki, mətndən cümləni, cümlədən isə “məktəblilərin” söz-formasını ayırmışıq. MT sistemi çərçivəsində həmin söz-formanın təhlili tələb olunur.

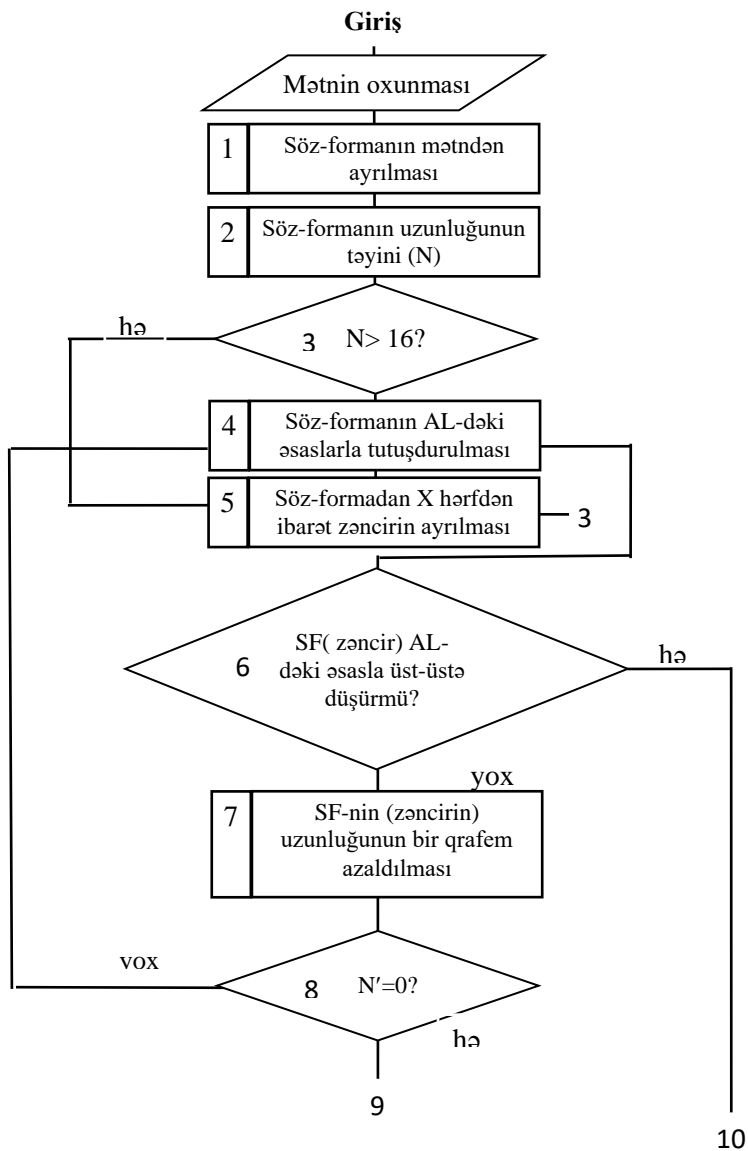
Təhlil alqoritmlərinin köməylə həmin söz-forma “məktəbli”, “lər” və “in” hissələrinə ayrılır. “Məktəbli” sözü lüğətdə (avtomatik lüğətdə) təmsil olunduğu üçün onun mənası tapılır və üzərinə şəkilçilər cədvəlində təsbit olunmuş “lər” və “in” şəkilçilərinin qrammatik əlamətləri əlavə olunur. MT çərçivəsində “məktəbli” sözünün “məktəb” – “li” hissələrinə ayrılması, sonradan “məktəb” kökünə “li” sözdüzəldici şəkilçisinin əlamətlərinin, daha sonra bu göstərilənlərə “lər” və “in” sözdəyişdirici şəkilçilərinin qrammatik əlamətlərinin əlavə edilməsi ilə həmin söz-formanın məşının “qavrama” səviyyəsi və imkanları həddində sintezi məsələni bir qədər də mürəkkəbləşdirmiş olardı. MT prosesinin növbəti pilləsində söz-formanın tərcümə olunacaq dildə qarşılığı tapılır. Ayrılmış ayrı-ayrı vahidlərin – əsas (və ya kök) və şəkilçilərin qrammatik əlamətləri tərcümə olunacaq dildəki ekvivalenti ilə əvəz olunur. Lazım gəldikdə daha dəqiq ekvivalentin tapılması məqsədilə sintaktik-semantik təhlil və sintez blokları “işə qarışmalı” olur. Bütün bu dəqiqləşdirmə və araşdırmalardan sonra söz-forma tərcümə olunacaq dildə sintez olunur. Bu zaman söz-formanın hər iki dildə – sintaktik əhatədə kontekstdə olan məna xüsusiyyətləri də mütləq nəzərə alınır. Bu morfoloji səviyyədə əsas və şəkilçilərin çoxmənalılığı və omonimliyinin aradan qaldırılmasına imkan yaradır. Son mərhələdə tərcümə olunacaq dildə mətnin sintezi alınır. Sintaktik-semantik təhlil və sintez bloklarının “himayəsi” ilə sintaktik quruluşu (söz sırası, cümlə üzvlərinin mövqeyi) dəqiqləşdirməklə tərcümə mətni çap olunur. Əlbəttə, bütöv bir sistemin ən ümumi şəkildə təsvirində qarşıya çıxacaq biləcək çətinlik və maneələrin hamısını əhatə etmək mümkün deyil. Ayrı-ayrı blokların işi ilə əlaqədar bu problemlərə bir daha

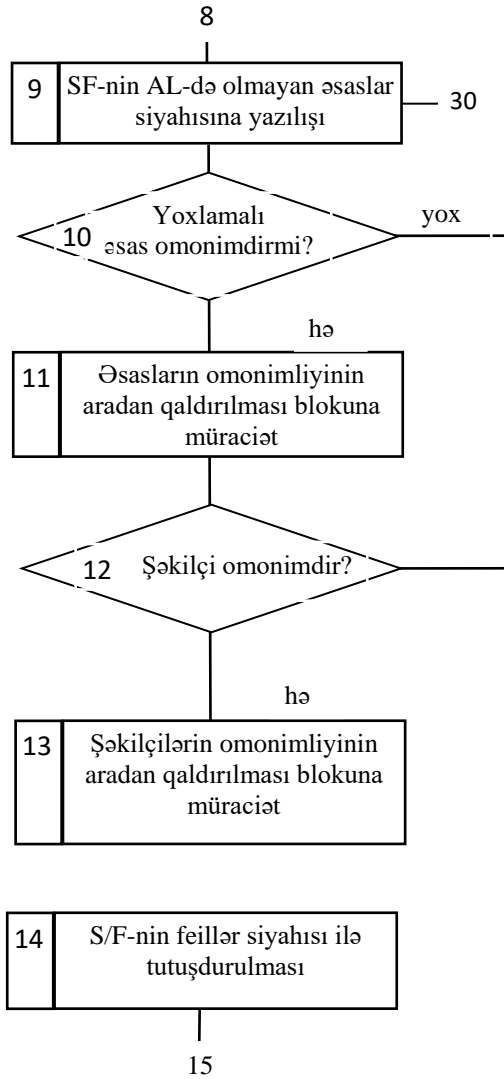
toxunulacaq. Məqsəd leksik, morfonoloji, morfoloji, sintaktik-semantik səviyyələrdə təhlil sistemləri işlənib hazırlanarkən hansı amillərin nəzərə alınıb-alınmaması məsələləri barədə ümumi təsəvvür yaratmaqdır.

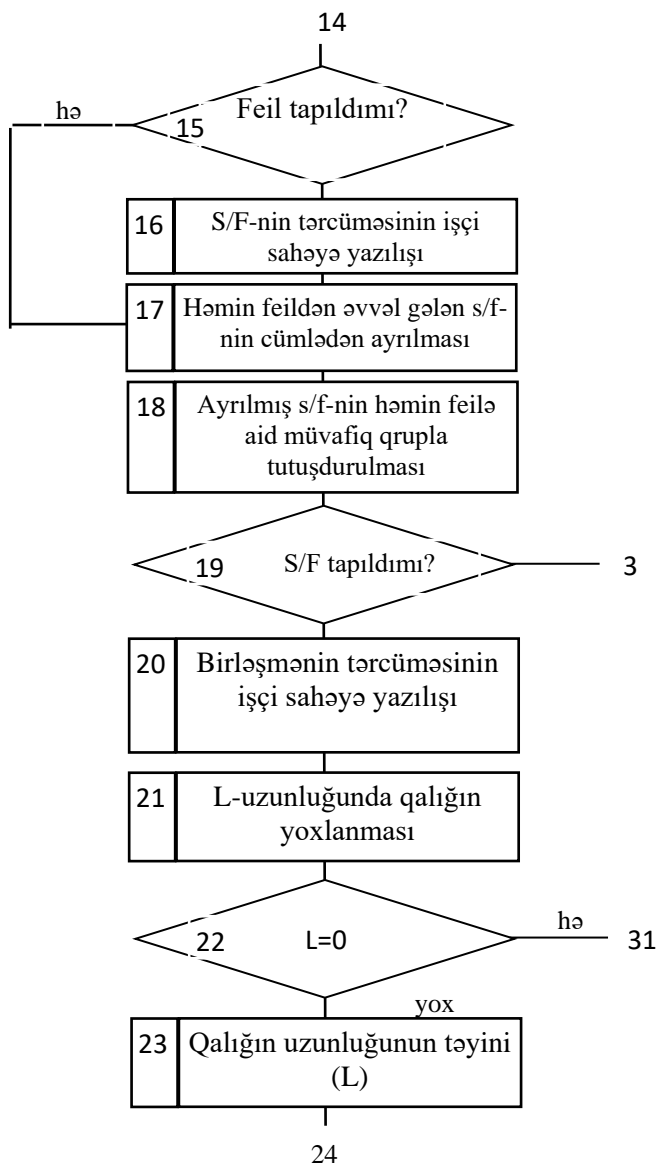
MT problemi indi daha aktual məzmun kəsb etmişdir. Əvvəlki illərdə (1983, 1985, 1989-cu illərdə) və sonralar müasir maşın tərcüməsi sistemlərinin müxtəlif məsələlərinə həsr olunmuş beynəlxalq seminar və konfransların keçirilməsi dediklərimizi bir daha sübut edir. Bu elmi məclislərdə və 70-ci illərdən bəri keçirilmiş digər beynəlxalq, ümumittifaq və respublika məclislərində MT probleminin ayrı-ayrı istiqamətləri barədə respublikamızın nümayəndələri də fəal iştirak etmiş və türkdilli xalqlar arasında ilk dəfə olaraq əldə etdikləri bir çox nəzəri və təcrübi elmi nailiyyətlər barədə məlumat vermişlər.

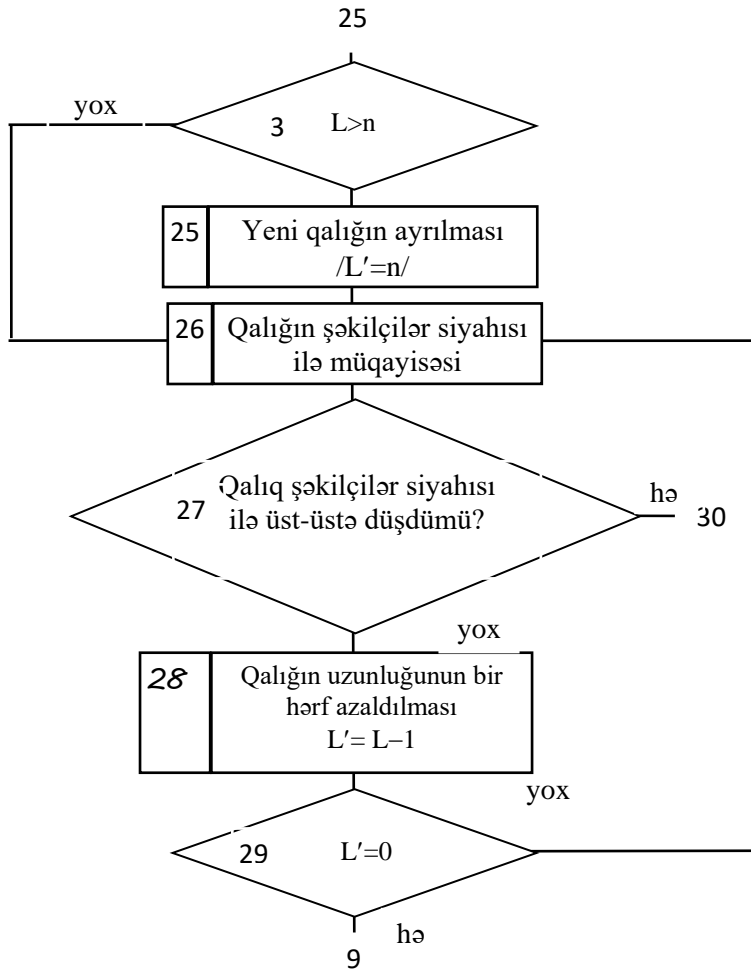
Maşın tərcüməsi sistemlərində ən mühüm bloklardan biri leksik-leksikoqrafik blok hesab olunur. Tərcümənin keyfiyyəti daha çox sistemdə yerləşdirilmiş maşın lüğətlərinin keyfiyyət və həcmindən asılı olur. Maşına daxil edilən lüğətlərin riyazi modelləri ideal lüğət strukturlarına əsaslanır. Elə bir lüğət modeli seçilməlidir ki, kifayət qədər mobil və eyni zamanda əhatəli olsun.

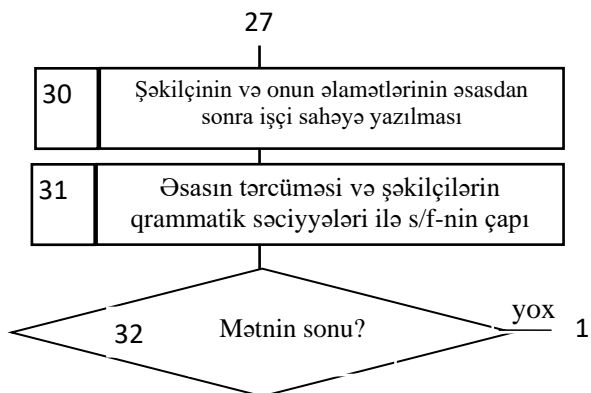
Aşağıdakı sxemdə leksik-morfoloji səviyyədə azərbaycanca-rusca maşın tərcüməsinin ümumi blok sxemi göstərilmişdir. Sxemdə N – əsasların, n – şəkilçilərin, X – söz-formaların, L – qalığın uzunluğunu göstərir.











Sxem 2. Leksik-morfoloji səviyyədə azərbaycanca-rusca məşin tərcüməsinin ümumi blok sxemi

VII FƏSİL

MİLLİ DİL KORPUSLARI SÜNİ İNTELLEKT KONTEKSTİNDƏ

Azərbaycan dilinin milli korpusunun yaradılması sahəsində aparılan işlər də süni intellekt sistemləri kontekstində araşdırılmalıdır. Bu, dilçiliyin bir çox problemlərinə yeni baxış bucağından baxmaq, bu istiqamətdə aparılan elmi-tədqiqat işlərinin səmərəsini artırmaq imkanı yarada bilər. Təbii mətnin emalı, mətnin formal linqvistik təhlili, dilin riyazi-statistik metodlarla öyrənilməsi və bu səpkidə olan digər linqvistik texnologiyalar ayrı-ayrılıqda, bir-birindən təcrid olunmuş halda deyil, süni intellekt problemləri kontekstində tədqiq olunmalıdır. Belə olduqda, daha dəqiq və optimal nəticələr əldə etmək imkanı yaranır. Milli dil korpuslarının süni intellekt sistemləri ilə qarşılıqlı əlaqə və münasibətlərini izah etməzdən əvvəl bəzi terminlərlə tanış olaq.

Korpus dilçiliyi hər hansı konkret dili təmsil edən mətnlərin qabaqcadan müəyyənləşdirilmiş qaydalar üzrə sistemli şəkildə elektron formatda toplanmasına istiqamətlənmiş məlumat axtarışı sistemlərinin yaradılması və onların tədqiqat prosesində istifadəsi məsələlərini öyrənir. Məlum olduğu kimi, “korpus” bir çox mənalarda işlənir: “hərbi birləşmə, korpus”, “tədris korpusu”, “gəminin korpusu”, “şrift dəsti” və s. Korpus dilçiliyində “korpus” elektron formatda toplanmış mətnlər toplusu, elektron fond mənasında işlənir.

Milli dil korpusları. Konkret dilin milli korpusunda müəyyən zaman kəsiyində həmin dil tam şəkildə bütün janrları, üslubları, ədəbi dili, dialektləri, qrammatikası, lüğətləri və s. ilə təmsil olunur. Başqa sözlə, milli dil korpusu elektron daşıyıcıda saxlanan müxtəlif dil hadisələri və faktları ilə bağlı tədqiqatçılara

operativ və dəqiq məlumat verən, müəyyən nizamla düzülmüş təbii mətnlər çoxluğudur. Burada “milli” sözü də müxtəlif anlamlarda izah olunur. Qeyd etməliyik ki, Britaniya milli korpusundan gələn ənənə kimi işlənməkdədir.

Təbii dilin emalı. NLP (Natural Language Processing) – təbii dilin emalı sistemləri dedikdə, maşın tərcüməsi, orfoqrafiyanın və qrammatikanın yoxlanması, nitqin yazıya (mətnə) çevrilməsi və başa düşülməsi (qavranması), sorğulara avtomatik cavab verilməsi, mətnin nitqə çevrilməsi (mətn üzrə insan nitqinin generasiyası) və s. sistemlər başa düşülür. Bu sistemlər insanla kompüter arasında daha təbii interfeys yaradılmasına, insanın kompüterlərdən daha səmərəli yararlanmasına və kompüterlərdən daha geniş istifadə edilməsinə şərait yaradır.

Elektron lüğətlər. Bəzən bu tipli lüğətlər kompüter lüğətləri də adlandırılır. Elektron formatda tərtib olunmuş və daha çox internetdə istifadə üçün nəzərdə tutulmuş elektron lüğətlər sistemidir. Lüğətlərin kitab (kağız) variantından fərqli olaraq daha geniş yayılma və sürətli istifadə imkanlarına malikdir. Hazırda kitab variantında olan qiymətli və nadir lüğətlərin də elektron variantlarının yaradılması sahəsində intensiv işlər aparır. Konkordansları da elektron lüğətlərə aid etmək olar. Milli dil korpuslarının mühüm bloklarından biri də elektron lüğətlər altkorpusu hesab olunur.

Qeyd etmək lazımdır ki, mətnin formal təhlili, təbii mətnin emalı, mətnin səsə (nitqə), səsin (nitqin) mətnə çevrilməsi, maşın tərcüməsi, məlumatların avtomatik (formal əlamətlərinə görə) axtarışı, tarixi abidələrin və klassiklərin dilinin riyazi-statistik metodlarla tədqiqi, statistik lüğətlərin (tezlik, əks) tərtibi, elektron lüğətlərin hazırlanması sahəsində Azərbaycanda xeyli iş görülmüşdür [Mahmudov, Fətullayev, 2004; Махмудов, 1991; 2001; Mahmudov, 2002; 2013; 2017; 2018; Fətullayev,

Mahmudov, 2015; Mahmudov, Fətullayev, 2016; Mahmudov, Tahirov, 2019].

Bu araşdırmaları və analogi istiqamətdə Azərbaycan Respublikasında aparılan digər tədqiqat işlərini süni intellekt sistemlərinə aid tədqiqatlar kimi səciyyələndirmək olar.

Korpus mətnlərin elektron formatda toplanmasına istiqamətlənmiş axtarış sistemi olaraq ona hesablanmışdır ki, gələcəkdə dilin bu və ya digər məsələləri ilə maraqlanan tədqiqatçı həmin mənbələrdən lazımi linqvistik məlumatı əldə edə bilsin. Korpusu təmsil edən mətnlər dili kifayət qədər təmsil etməlidirlər. Yəni dilin bütün üslubları və janr müxtəliflikləri korpusda öz əksini tapmalıdır. Yalnız bu şərtlər daxilində korpus faydalı və səmərəli axtarış funksiyasını yerinə yetirə bilər [Mahmudov, 2018].

Korpusun ayrı-ayrı bloklarının təşkili və əlaqələndirilməsi prosesində də süni intellektlə bağlı tədqiqatların elementlərindən istifadə olunur.

Məlumdur ki, ayrı-ayrı türk dilləri üçün məşin fondunun yaradılması ideyası keçmiş SSRİ məkanında XX əsrin 80-ci illərinə təsadüf edir. Sonrakı illərdə Türk Dillərinin Məşin Fondu əsasında keçmiş SSRİ məkanında yaşayan bir çox türkdilli xalqlar üçün milli dil korpusları yaradılmağa başladı. Hazırda Azərbaycan, türk, qazax, özbək, başqırd, tatar və bir sıra başqa türkdilli xalqlar üçün korpus yaradıcılığı sahəsində müəyyən işlər görülmüş və bu sahədə araşdırmalar davam etməkdədir.

Müasir Azərbaycan ədəbi dilinin ayrı-ayrı üslublarının, Azərbaycan klassiklərinin və yazılı tarixi dil abidələrin riyazi-statistik metodlarla öyrənilməsi sahəsində aparılmış tədqiqatlar da milli korpusun yaradılması prosesində böyük fayda verə bilər [“Kitabi-Dədə Qorqud”un statistik təhlili, 1999; Məhəmməd Füzulinin nəzm əsərlərinin əlifba-tezlik sözlüyü, 2004; Azərbaycan dilinin tezlik lüğəti, 2010; Azərbaycan

dilinin əks əlifba lüğəti, 2004 və s.]. Məlum olduğu kimi, Azərbaycanda “Dilmanc” layihəsi çərçivəsində NLP (Natural Language Processing) təbii dilin (mətnin) emalı sahəsində bir çox işlər görülmüşdür [Mahmudov, Fətullayev və b., 2016; Fətullayev, Mahmudov, Məmmədov, 2015; Fətullayev, Mahmudov, Fətullayev, 2015]. Azərbaycan dilinin tezlik və əks lüğətlərinin tərtibi sahəsində də həmin yaradıcı kollektivin xidmətləri danılmazdır.

Yuxarıda qeyd olunduğu kimi, təbii mətnin emalı, məlumatların formal əlamətə görə axtarışı, maşın tərcüməsi, elektron lüğətlərin tərtibi, dil materialının riyazi-statistik metodlarla tədqiqi məsələləri korpus dilçiliyinin, eləcə də süni intellekt sistemlərinin tədqiqat obyektinə hesab olunur. Yeni linqvistik texnologiyaların yaradılması və tətbiqi süni intellekt sistemlərinin əsasını təşkil edir. Robot sənayesini bunlarsız təsəvvür etmək qeyri-mümkündür. Nitqi qavrayıb müəyyən suallara cavab verə bilən, şifahi və yazılı əmrləri icra etmək qabiliyyətinə malik robotlar təbii mətnin emalı texnologiyalarından istifadə etmiş olurlar. Verilənlərdən biliyin alınması texnologiyası intellektual təhlilin əsas problemlərindən biridir. Bu problemin həlli ilə bağlı müxtəlif yanaşmalar mövcuddur. Bu yanaşmalardan biri də neyroşəbəkə texnologiyasına əsaslanır. Neyroşəbəkə – canlı orqanizmin əsəb sinirlərinin quruluşu və funksiyalarını imitasiya edən riyazi modeldir [az.wikipedia.org/wiki/süni_intellekt].

Süni intellektlə neyrofiziologiya, epistemologiya və koqnitiv psixologiyanın qovuşuğunda koqnitologiya adlanan yeni elm sahəsi yaranmışdır. Koqnitologiya dilçilik elminin də tədqiq obyektlərindən biri kimi səciyyələndirilir.

Britaniya Milli Korpusunun (BNC) yaradılması sahəsində əsaslı işlərə 1991-ci ildən başlanıb. 1994-cü ildə ilkin versiya hazırlanıb. 2001-ci və 2007-ci illərdə bir çox düzəlişlər

və struktur dəyişmələri aparılıb və yeni versiya istifadəçilərə təqdim olunub.

BNC-nin yaradılması layihəsində üç nəşriyyat – Oksford Universiteti Nəşriyyatı (aparıcı həmmüəllif kimi), eləcə də Longman və W. & R. Chambers), iki universitet (Oksford və Lankaster) və Britaniya kitabxanası iştirak etmişdir.

29 aprel 2004-cü ildə istifadəçilərə təqdim olunmuş Rus Dilinin Milli Korpusunun yaradılmasında aşağıdakı təşkilatlardan olan mütəxəssislər iştirak etmişlər:

Rusiya Elmlər Akademiyasının V.V.Vinoqradov adına Rus Dili İnstitutu (ИРЯ РАН);

Rusiya Elmlər Akademiyasının Dilçilik İnstitutu (ИЯ РАН);

Rusiya Elmlər Akademiyasının A.A.Xarkeviç adına İnformasiyanın Ötürülməsi Problemləri İnstitutu (ИППИИ РАН);

Rusiya Elmlər Akademiyasının Sankt-Peterburqdakı Linqvistik Tədqiqatlar İnstitutu (ИЛИ РАН) – Sankt-Peterburq Dövlət Universiteti (СПГУ) ilə birlikdə;

Voronej Dövlət Universiteti.

Türk Dilinin Milli Korpusunun yaradılmasında iştirak edən araşdırmaçılar (yaradıcı qrup):

Prof.Dr. Yeşim Aksan (Mersin Universiteti); Prof.Dr. Mustafa Aksan (Mersin Universiteti); Prof.Dr. Selma Ayşə Özel (Çukurova Universiteti); Bilg.Yüks.Müh. Hakan Yılmaz (Çukurova Universiteti); Ögr.Gör. Yasin Bektaş (Mersin Universiteti); Ögr.Gör. Ümit Mersinli (Mersin Universiteti); Dr. Gülsüm Atasoy (Mersin Universiteti); Arş.Gör. Umut Ufuk Demirhan (Mersin Universiteti) [Aksan, Mersinli, Yaldır, 2011; Aksan, Mersinli, Yaldır, Demirhan, 2011].

Özbək Dilinin Milli Korpusunun yaradılması işlərində Əlişir Nəvai adına Daşkənd Dövlət Özbək Dili və Ədəbiyyatı Universiteti, Fərqanə Dövlət Universiteti, Namanqan Dövlət Universiteti, Özbəkistan Milli Universiteti və başqa təşki-

latlarından ibarət yaradıcılıq qrupu iştirak edir (<http://uzbekcorpus.uz>).

Azərbaycan dilinin milli korpusunun yaradılması sahəsində artıq zəmin yaradıldığından hazırda daha geniş miqyasda işlərə başlamaq olar. Bunun üçün isə daha böyük mətnlər korpusuna ehtiyac duyulur. Belə ki, korpusa daxil edilən mətnlər Azərbaycan dilinin bütün üslublarını eyni səviyyədə və həcmdə ehtiva etməlidir. Məlum olduğu kimi, korpusun həcmi nə qədər böyük olsa, bir o qədər etibarlı və əhatəli hesab olunur. Azərbaycan dilinin milli korpusuna aşağıdakı mətnlərin daxil edilməsi məqsədəuyğun hesab edilir: Əsas korpus (təxminən 50 milyon söz-forma) –

1. Bədii mətnlər (nəsr əsərləri, nəzm əsərləri, dramaturgiya mətnləri, folklor mətnləri, klassik Azərbaycan ədəbiyyatı, müasir Azərbaycan ədəbiyyatı), publisist mətnlər, rəsmi və işgüzar üsluba aid mətnlər, elmi mətnlər, elmi-texniki mətnlər, humanitar elmlərə aid mətnlər, dini və fəlsəfi mətnlər, dialekt mətnləri;

2. Azərbaycan dilinin elektron lüğətlər altkorpusu (Azərbaycan dilinin izahlı, orfoqrafiya, terminoloji, omonimlər, sinonimlər, antonimlər, dialektoloji, tezlik, əks əlifba, ikidilli, çoxdilli, frazeologiya, orfoepiya, statistik və s.);

3. Şifahi mətn altkorpusları (məişət-danışiq dili, rəsmi çıxışlar, dialekt mətnləri və s.

4. Paralel altkorpuslar (ikidilli, çoxdilli, ingilis-Azərbaycan, Azərbaycan-ingilis, rus-Azərbaycan, Azərbaycan-rus, alman-Azərbaycan, Azərbaycan-alman və s., eləcə də analogi olaraq çoxdilli);

5. Çoxdilli və birdilli mətnlər;

6. Linqvistik analizatorlar – dilin müxtəlif səviyyələri (fonetik, morfoloji, sintaktik, sözdüzəltmə, semantik) üçün nəzərdə tutulmuş təhlil və axtarış proqramları;

7. Yazılı mətnin səsə (nitqə), səsin (nitqin) mətnə çevrilməsini təmin edən proqram sistemləri;

8. Tədris altkorpusları;

9. Konkordanslar;

10. Korpusun istifadəsi üçün mükəmməl proqram təminatı (axtarış proqramları).

Azərbaycan dilinin milli dil korpusunun yaradılması prosesində qabaqcıl dünya təcrübəsi nəzərə alınır. Burada korpuslar üçün çox önəmli olan mətnlərin formal meyarlara görə təhlili ön plana çəkilir. Hazırda “Azərbaycan dilinin tezlik lüğəti”nin birinci cildi çap edilmiş, Azərbaycan dilinin qloballaşma şəraitində zamanın tələblərinə uyğun istifadəsinə və ölkədə dilçiliyin inkişafına dair Dövlət proqramına uyğun “Azərbaycan dilinin ümumi informasiya bazası” [<http://azerbaycandili.az/>] yaradılmışdır. “Elmin inkişafı fondu”nun dəstəyi ilə hazırlanan “Elektron lüğətlər korpusu”nda (<http://korpus.azerbaycandili.az/>) isə bir sıra lüğətlərin elektron variantları qoyulmuşdur. Bütün bunlar Azərbaycan dilinin milli korpusunun yaradılması üçün ilkin baza rolunu oynaya bilər [Mahmudov, Tahirov v. B., 2019; Mahmudov, 2017; Махмудов, 2012; 2017].

Yuxarıda qeyd etdiyimiz kimi, Azərbaycan ədəbi dilinin ayrı-ayrı üslublarının, Azərbaycan klassiklərinin və yazılı tarixi dil abidələrin riyazi-statistik metodlarla öyrənilməsi sahəsində aparılmış tədqiqatlar da milli korpusun yaradılması prosesində ilkin mərhələ rolunu oynayır və böyük fayda verə bilər [Mahmudov, 2018; Mahmudov, Tahirov və b., 2019;].

Məlum olduğu kimi, “Dilmanc” layihəsi çərçivəsində Azərbaycanda NLP (Natural Language Processing) təbii dilin (mətnin) emalı sahəsində bir çox işlər görülmüşdür. Layihə çərçivəsində hazırlanmış, dilin bütün üslublarını əhatə edən ikidilli paralel mətn korpuslarını (ingilis-Azərbaycan ikidilli korpusu – 2 milyon cümlə, türk-Azərbayca ikidilli korpusu – 277 min cümlə, rus-Azərbaycan ikidilli korpusu – 4, 5 milyon

cümlə) və konkret dilləri əhatə edən irihəcimli birdilli korpusları (Azərbaycan birdilli korpusu – 60 milyon cümlə, türk birdilli korpusu – 322 milyon cümlə) xüsusi olaraq qeyd etmək olar [Mahmudov, Fətullayev və b., 2016].

Süni intellektin ən əhəmiyyətli istiqamətlərdən biri təbii dilin emalı hesab olunur. Mətnin formal təhlili, təbii mətnin emalı, maşın tərcüməsi, məlumatların formal əlamətlərə görə axtarışı, tarixi abidələrin və klassiklərin dilinin riyazi-statistik metodlarla tədqiqi, statistik lüğətlərin tərtibi, elektron lüğətlərin hazırlanması sahəsində aparılan tədqiqatlar və görülən işlər də korpus dilçiliyinin tədqiqat obyektini hesab olunur. Korpus dilçiliyi konkret dili təmsil edən mətnlərin elektron formatda toplanmasına istiqamətlənmiş məlumat axtarışı sistemlərinin yaradılması və onların tədqiqat prosesində istifadəsi məsələlərini öyrəndiyi üçün milli dil korpuslarının mühüm komponentlərindən biri hesab olunan təbii dilin emalı prioritet istiqamət imi səciyyələndirilir. Bu istiqamət çərçivəsində təbii mətnlərin təhlili, qavranma imkanları və sintezi tədqiq olunur. Tədqiqatın nəticəsi olaraq korpusdan yararlanan istifadəçi tərəfindən internetdə rast gəlinən istənilən mətnin emalı və ondan müəyyən informasiyanın əldə etməsi nəzərdə tutulur. Təbii mətnin emalı sistemləri korpuslarla yanaşı öz tətbiqini maşın tərcüməsi və məlumat axtarışı sistemlərində tapır [Mahmudov: 2002; Fətullayev, Mahmudov: 2015; Mahmudov, Fətullayev: 2016; Mahmudov: 2017; Mahmudov, Tahirov: 2019].

Qeyd etmək lazımdır ki, təbii mətnin emalı, məlumatların formal əlamətə görə axtarışı, maşın tərcüməsi, elektron lüğətlərin tərtibi, dil materialının riyazi-statistik metodlarla tədqiqi məsələləri korpus dilçiliyinin, eləcə də süni intellekt sistemlərinin tədqiqat obyektini hesab olunur. Yeni linqvistik texnologiyaların yaradılması və tətbiqi süni intellekt sistemlərinin əsasını təşkil edir. Robot sənayesini bunlarsız təsəvvür

etmək qeyri-mümkündür. Nitqi qavrayıb müəyyən suallara cavab verə bilən, şifahi və yazılı əmrləri icra etmək qabiliyyətinə malik robotlar təbii mətnin emalı texnologiyalarından istifadə etmiş olurlar.

Verilənlərdən biliyin alınması texnologiyası intellektual təhlilin əsas problemlərindən biridir. Bu problemin həlli ilə bağlı müxtəlif yanaşmalar mövcuddur. Bu yanaşmalardan biri də neyroşəbəkələrin verballaşdırılması prosedurundan istifadə edən neyroşəbəkə texnologiyasına əsaslanır. Neyron şəbəkələr geometrik fiqurların tanınması və ya obyektlərin qeyri-dəqiq və mürəkkəb sistemlərlə təhlilində istifadə olunur. Neyroşəbəkə – canlı orqanizmin əsəb sinirlərinin quruluşu və funksiyalarını imitasiya edən riyazi modeldir. Müvafiq olaraq idealda – bu müstəqil öyrənən sistemdir. Bu prinsipi texnoloji müstəviyə keçirsək, onda deyə bilərik ki, neyroşəbəkə irihəcmli layihədə hər hansı bir məsələnin icrasını həyata keçirən proqramlar çoxluğudur. Başqa sözlə, superkompüter, çoxlu sayda adi kompüterlər şəbəkəsidir.

SI-nin tətbiq sahələri və imkanları çox geniş və çoxşaxəlidir. Maraqlı orasıdır ki, bu intellektual sistemlərin və insan-robotların yaradılmasında dilçilikdən, onun müasir sahələrindən, metodlarından da geniş istifadə olunur. İnsan nitqini qavrayan, başa düşən və müvafiq cavablar verə bilən robotların yaradılması prosesində təbii dilin emalı, yazılı mətni nitqə (səsə), nitqi (səsi) mətnə çevirən linqvistik texnologiyalardan istifadə olunur.

Qeyd etmək lazımdır ki, mətnin formal təhlili, təbii mətnin emalı, mətnin səsə (nitqə), səsin (nitqin) mətnə çevrilməsi, maşın tərcüməsi, məlumatların avtomatik (formal əlamətlərinə görə) axtarışı, tarixi abidələrin və klassiklərin dilinin riyazi-statistik metodlarla tədqiqi, statistik lüğətlərin (tezlik, əks) tərtibi, elektron lüğətlərin hazırlanması sahəsində respublikamızda xeyli iş görülmüşdür [Mahmudov, Fətullayev, 2004;

Махмудов, 1991; 2001; Mahmudov, 2002; 2012; 2013; 2017; 2018; Fətullayev, Mahmudov, 2015; Mahmudov, Fətullayev, 2016; Mahmudov, Tahirov, 2019].

Bu araşdırmaları və analogi istiqamətdə respublikamızda aparılan digər tədqiqat işlərini süni intellekt sistemlərinə aid tədqiqatlar kimi səciyyələndirmək olar.

1943-45-ci illərdə neyron şəbəkələrinin başa düşülməsi, dərki, mənimsənilməsi və yaradılmasının əsası qoyulmuş, 1950-ci ildə isə Alan Türinq elmi jurnalda intellektual şahmat oyununun təhlilini vermişdir. 1958-ci ildə süni intellektin ilk proqramlaşdırılma dili – Lisp yaranmışdır. 1960-1970-ci illərdə alimlər sübut etdilər ki, kompüterlər təbii dili qavramağa, təhlil və sintez etməyə qabildir [ru. wikipedia.org/wiki/искусственный_интеллект].

Yuxarıda göstərilənlər əsasında belə nəticəyə gəlmək olar ki, korpus dilçiliyi çərçivəsində istifadə olunan linqvistik texnologiyalar, onların tətbiqi və istifadəsi məsələləri mahiyyətinə görə süni intellekt məsələləri ilə üzvi surətdə bağlıdır.

Məlum olduğu kimi, milli korpus hər hansı bir konkret dildə mətnlərin elektron formada toplanmasına istiqamətlənmiş məlumat axtarış sistemidir. Konkret dilin milli korpusunda həmin dil tam şəkildə, bütün səviyyələrdə təmsil olunur. Korpusu təşkil edən mətnlər və həmin mətnlərə müraciət formaları müəyyən qaydalar üzrə nizamlanır. İstifadəçi onu maraqlandıran istənilən məsələ ilə bağlı korpusdan lazımi məlumat əldə etmək imkanına malik olmalıdır. Korpusda dilin bütün üslubları, janrları inkişaf mərhələləri, özəllikləri, fonetikas, qrammatikası, frazeologiyası, lüğətləri (ikidilli, çoxdilli tərcümə lüğətləri, terminoloji, tarixi, etimoloji, dialektoloji, statistik, antonimlər, sinonimlər, omonimlər, müxtəlif sorğu lüğətləri, onomastik, ensiklopedik və s.) ehtiva olunur. Bundan başqa, milli korpusda müxtəlif istiqamətli və janrlı bədii mətnlər, jurnal, qəzet materialları, elmi, elmi-populyar,

tədris mətnləri, dini-fəlsəfi mətnlər, texniki mətnlər, rəsmi-ışğuzar və hüquqi mətnlər, məişət mətnləri və s. geniş əks olunur. Çox geniş əhatə dairəsinə malik olan korpus təkcə dilçilik elminin obyektini hesab oluna bilməz. Elmi ədəbiyyatda milli korpusun yaradılması çox vaxt dilçiliklə əlaqələndirilir. Təsadüfi deyil ki, “milli korpus” maşın fondu, “kompüter fondu” terminləri ilə yanaşı bir çox tədqiqatçılar “korpus dilçiliyi” terminini təklif edir və milli korpusla bağlı məsələləri dilçiliyin yeni bir sahəsi olan “korpus dilçiliyi”nin predmeti və obyektini hesab edirlər.

Korpus dilçiliyi sahəsində ilk tədqiqatlar 1956-cı ildən başlayaraq ABŞ-da aparılmışdır. Həmin dövrdə çoxfunksional məlumat bazalarının və lüğətlərin hazırlanmasına başlanılmışdır. Bu işləri müasir milli dil korpuslarının ilk nümunələri hesab etmək olar. Hazırda dünyanın bir çox ölkələrində inkişaf etmiş aparıcı dillər üçün milli korpuslar işlənilib hazırlanmışdır. Əlbəttə, bu korpuslar elmi və texniki göstəricilərinə görə eyni deyil.

Dünyanın ən çox sayılan və örnək kimi götürülən korpusu Britaniya Milli Korpusu hesab olunur. Slavyan dillərindən rus və çex dillərinin milli korpusları xüsusilə seçilir.

Türk dillərinin maşın fondunun yaradılması ideyası 1988-ci ildə Moskvada Sovet Türkoloqlarının XIV plenumunda irəli sürülmüş və onun əsas istiqamətləri müəyyənləşdirilmişdir. Hazırda keçmiş SSRİ ərazisində yaşayan bir çox türkdilli xalqların milli dil korpuslarının yaradılması sahəsində işlər aparılmaqdadır. Qazax, tatar və başqırd dilləri üçün bu sahədə xeyli iş görülmüşdür [Гузев, Пиотровский, 1988, с.92-101].

Məlum olduğu kimi, hazırda respublikamızda elektron hökumət və elektron kitabxana xidmətləri fəaliyyət göstərməkdədir. Milli kitabxanada, Azərbaycan Milli Elmlər Akademiyasının kitabxanasında, bir çox sahə kitabxanalarında, elmi-

tədqiqat müəssisələrində, ali məktəblərdə nəşr olunmuş və olunan ədəbiyyatın elektron variantı yaradılır və internetdə yerləşdirilir. Bütün bu resurslar milli korpusda mətn massivi kimi istifadə oluna bilər. Lakin, kitabların elektron variantları internetdə çox vaxt ilkin işlənməsiz, çap olunduğu kimi (kağız variantda olduğu kimi) qoyulur. Bu da onların müxtəlif məqsədlərlə istifadəsini (onlardan müxtəlif məlumatların alınmasını) çətinləşdirir. Milli korpus üçün nəzərdə tutulmuş mətnlər isə birbaşa deyil, müəyyən işlənmə prosesindən sonra (lingvistik işlənmə, markerlənmə və s.) sistemə daxil edilə bilər. İnternetdə yerləşdirilmiş bu ədəbiyyatın korpusda istifadə üçün ilkin işlənmə prosesinin forma və üsulları barədə düşünməyə dəyər. Mövcud potensialdan səmərəli şəkildə yararlanmaqla Azərbaycan dilinin milli korpusunun yaradılması sahəsində işləri intensivləşdirmək olar.

Milli korpus açıq sistemdir və onun strukturunda və funksiyalarında istənilən vaxtda zərurət yarandıqda dəyişiklik aparıla bilər [Mahmudov, 2013, s.61-63].

Milli korpusun altkorpuslarının hamısı eyni vaxtda fəaliyyət göstərmir. Altkorpuslar formalaşdıqca, onların təminat bazası yarandıqca fəaliyyətləri reallaşır. İlk vaxtlarda korpusda bir neçə altkorpus fəaliyyət göstərir. Onların qarşılıqlı əlaqələri, tələbatı yeni altkorpusların yaranması üçün stimula ola bilər.

Məlum olduğu kimi, altkorpus konkret mətn massivinə birbaşa müraciəti formalaşdırır, oradan informasiya almaq imkanı yaradır.

Məsələn, lingvistik təhlil altkorpusunun funksiyası multi-mediya və ya paralel mətnlər altkorpuslarından fərqlidir. Onlara müraciət və məlumat alınması fərqli prinsiplər əsasında reallaşdırılır. Hər bir altkorpus quruluşuna, təyinatına, vəzifəsinə, informasiyavericilik qabiliyyətinə görə özəl xüsusiyyətlərə malikdir. Başqa sözlə, hər bir altkorpus konkret

müraciət forması tələb edir. Milli dil korpusları çox vaxt internetdə tam şəkildə qoyulmur. Ana səhifədə göstərilən bir çox bölmə və blokların sonralar hazır olduqca əlavə olunması nəzərdə tutulur. Bu ondan irəli gəlir ki, milli dil korpuslarının hazırlanması çox vaxt tələb edən mürəkkəb texnologiyadır. Onun bəzi bölmə və bloklarının hazırlanması işləri başa çatdırıldıqdan sonra korpusa əlavə oluna bilər. Əslində, internetdə yerləşdirilmiş digər bloklara da sonradan əlavələr edilə və təkmilləşdirmələr aparıla bilər.

Fikrimizcə, ilk növbədə leksikoqrafiya – lüğət altkorpusunun funksiyasını, fəaliyyətini canlandırmaq olardı. Qısa zaman müddətində respublikamızda nəşr olunmuş və olunan bütün lüğətlərin elektron variantını yaratmaq lazımdır. Bu sahədə ilk növbədə “Dilmanc” və “Poliqlot” sistemlərinin mövcud lüğət potensialından yararlanmaq faydalı olardı. Bunlardan əlavə, Azərbaycan dilinin orfoqrafiya və əks lüğətlərinin elektron variantları da hazırlanmış və internetdə yerləşdirilmişdir. Onlar da hazır material kimi milli korpusa daxil edilə bilər.

Leksikoqrafiya altkorpusunda respublikamızda olan bütün lüğətlərin elektron variantları arasında əlaqə yaradılmalıdır. Korpusda qısa zaman kəsiyində hər hansı sözün izahlı lüğətdəki mənalarını, transkripsiyasını, tələffüzünü (lazım gələrsə, gələcəkdə hətta həmin sözün ifadə etdiyi əşyanın şəklini), dünya dillərinə tərcüməsini əldə etmək imkanına malik olmalıyıq. Leksikoqrafik altkorpus kitabxanaya getmədən, onlarla lüğəti vərəqləmədən hər hansı bir söz haqqında tam linqvistik (leksikoqrafik) məlumat əldə etməyi təmin etməlidir. Tam linqvistik məlumata sözün izahlı lüğətdəki mənaları, tələffüzü, düzgün yazılışı, transkripsiyası, etimologiyası, dialektoloji lüğətlərdəki məna variantları, frazeoloji vahidlərdə işlənmə məqamları, başqa dillərə (ingilis, rus, alman, fransız,

ispan, italyan, ərəb, fars, çin, yapon və s.) tərcüməsi və s. aid edilə bilər.

Qloballaşan dünyada hər hansı bir milli dil öz varlığını sübut etmək, mövcud olmaq, nüfuz qazanmaq istəyirsə internet məkanında təmsil olunmalıdır. İndi dünya bir bilik mənbəyi, örnək kimi daha çox internetdən istifadə edir, ondan faydalanır.

Dövlət dili olan Azərbaycan dili də öz varlığını, mövcudluğunu, nüfuzunu internetdə geniş və hərəkətli təmsil olunmaqla sübut etməlidir. İlk növbədə Azərbaycan dilinin milli korpusu – lüğətləri (izahlı, orfoqrafiya, orfoepiya, dialektoloji, tarixi, etimoloji, terminoloji, statistik, ikidilli, çoxdilli və s.), Azərbaycan dilinin məşin tərcüməsi sistemi, elmi, bədii ədəbiyyatı, publisistikası, qrammatikası, təhlil alqoritmləri və s. internetdə, yerləşdirilməlidir. Qeyd etmək lazımdır ki, məşin tərcüməsi və məlumat axtarışı sistemlərində geniş istifadə olunan mətnin formal təhlili və sintezi, həmçinin səsə tanınması (səsə mətnə, mətnə səsə çevrilməsi) sistemləri də Azərbaycan dilinin milli korpusunun tərkib hissələri, komponentləri hesab olunur. Azərbaycan dilinin milli korpusu elə qurulmalıdır ki, axtarış sistemində “Azərbaycan” yazmaqla onun dili, ədəbiyyatı, tarixi, incəsənəti, mentaliteti, coğrafiyası, təbiəti və s. haqqında hərtərəfli və dolğun məlumat əldə oluna bilsin. Azərbaycan dilinin milli korpusunun yaradılması bu baxımdan ən vacib, əhəmiyyətli addım olardı.

Qeyd etmək lazımdır ki, yuxarıda qeyd olunan bu ikidilli və birdilli altkorpuslar “Dilmanc” layihəsi çərçivəsində fəaliyyət göstərən mətnin formal linqvistik təhlili və məşin tərcüməsi sistemlərində uğurla istifadə olunmaqdadır. İkidilli paralel mətn altkorpusları Azərbaycan, türk, rus və ingilis dilləri arasında avtomatik tərcümə vasitəsi və lüğət kimi istifadəçilərə təqdim olunmuşdur.

Birdilli mətn altkorpusları həm də tərcümə olunmuş cümlələrin düzgünlüyünün, ünsiyyət prosesində müşahidə

olunma ehtimalının yoxlanması baxımından əhəmiyyətlidir. Nitqin tanınması prosesində yanlış, qüsurlu tələffüz nəticəsində düzgün tanınmamış sözlərin dəqiqləşdirilməsində birdilli altkorpuslar dil modeli kimi böyük fayda verə bilər. Nəzərdə tutmaq lazımdır ki, düzgün tələffüz olunmamış sözlərin işlənmə tezliyi birdilli altkorpusda təsbit olunmuş sözlərin düzgün variantından az olacaq. Məsələn, əgər tələffüz prosesində “Günəş işıq saçır” cümləsində “günəş” sözü səhvən “cünəş”, “düneş”, “küneş” kimi deyilmişsə, birdilli korpusun köməyi ilə bu sözün düzgün variantı (günəş) tapılır. Birdilli altkorpuslar tərcümə olunmuş cümlələrdə sintaktik vahidlərin yanaşıgəlmə və birləşmə variantlarını da düzgün müəyyənləşdirməyə imkan verir. Cümlədə (mətnədə) ümumi və xüsusi isimlərin dəqiq müəyyənləşdirilib izahı olunması və durğu işarələrinin düzgün qoyuluşu prosesində birdilli mətnlər altkorpusu əvəzsiz mənbədir.

Korpus materialı süni intellekt məsələlərinin həlli prosesində əvəzsiz mənbə kimi uğurla tətbiq oluna bilər. Məhz ona görə də korpus dilçiliyi məsələlərinin və bu sahədə yaradılmış linqvistik texnologiyaların süni intellekt problemləri kontekstində araşdırılması, istifadəsi və dəyərləndirilməsi faydalı olardı.

Süni intellekt – elə bir texnologiya, daha dəqiq desək, müasir elmi istiqamətdir ki, insan kimi düşünə bilən öyrədici kompüter, robotlaşdırılmış texnika, analitik sistem yaratmaq üsullarını öyrənir. Beləliklə, süni intellekt müxtəlif mənbə və elm sahələrindən – informatika, riyaziyyat, dilçilik, psixologiya, biologiya, fiziologiya, biotexnologiya və s. məlumat toplamaqla öyrədici maşın texnologiyasının köməyi ilə verilənlər massivi əsasında insan intellektini təqlid etməyə qabil intellektual sistemlər yaratmağa cəhd göstərir. Bu prosesdə son zamanlar təbii dilin emalı sistemlərində uğurla istifadə olunan neyron şəbəkələrdən, maşın tərcüməsi, mətnin səsə (nitqə),

nitqin (səsin) mətnə çevrilməsi proqramlarından, təbii dili tam ehtiva edən korpus dilçiliyindən, statistik və elektron lüğətlərdən faydalanır. Bu sadalananlar robototexnikada, intellektual robotların yaradılmasında mühüm vasitələrdən hesab olunur. Fikrimizcə, ilk əvvəllər riyazi dilçilik, hesablama dilçiliyi, statistik dilçilik, kompüter dilçiliyi, korpus dilçiliyi kimi dilçiliyin yeni sahələrinin tədqiqat obyektinə daxil olan bir çox məsələlər süni intellekt sistemləri kontekstində araşdırılmalıdır. Bu, dilçiliyin bir çox problemlərinə yeni baxış bucağından baxmaq, bu istiqamətdə aparılan elmi-tədqiqat işlərinin səmərəsini artırmaq imkanı yarada bilər.

SON SÖZ ƏVƏZİ

Bizi əhatə edən aləm, təbiət hadisələri və ictimai həyatda baş verənlər insanlar tərəfindən tam dərk olunmayıb və bəlkə də heç vaxt olunmayacaq. İnsan həmişə dünyanı anlamağa, dərk etməyə can atacaq və həmişə də hiss edəcək ki, hələ onun anlamadığı, ağına sığışdıra bilmədiyi sirlər, hikmətlər, məfhümlər, anlayışlar çoxdur, sonsuzluğun özü qədər çoxdur və insan oğlu heç vaxt idrakın son nöqtəsinə çata bilməyəcək, həmişə buna can atacaq, əbədi olaraq bu yolda irəliləyəcək, hərdən elə bil fasilə edəcək, geri dönəcək, sonra unuduqlarını yenidən xatırlayıb irəli gedəcək. Və bu prosesin, idrak prosesinin sonu yoxdur, heç vaxt da olmayacaq, əvvəli və ortası var, bir də sona doğru irəliləmək cəhdi, aləmi, ətrafı dərk etmək cəhdi, anlamaq arzusu...

Yayı sıxırıq, müəyyən həddə kimi, sonra yay güclü müqavimət göstərir, açılmaq istəyir və açılır. Yay açmaq istəyirik, dartırıq var gücümüzzlə, müəyyən həddə kimi, sonra güclü müqavimət və sıxılma baş verir. Qalaktikaların hərəkət trayektoriyaları da analogi qanunauyğunluqla tənzimlənir, açılma, yayılma, genişlənmə, sonra əks proses – yığılma, sıxılma.

Dənizin qabarması, məcrasından çıxması və yenidən məcrasına qayıtması, çəkilməsi, min illərlə, sonsuzluğa qədər davam edən bir proses. Dəniz varsa, mövcuddursa, qabaracaq, dalğalar sahilə çırpılacaq, sonra əvvəlki yerinə, məcrasına qayıdacaq. Qabarma və çəkilmə yoxdursa, dəniz də yoxdur, onun da varlığı, mövcudluğu elə qabarıq çəkilməsindədir. Allahın bizə tam açmadığı, izah olunmaz, dərk olunmaz sirləri çoxdur.

Dialektik proses olan elmlərin diferensiyası və inteqrasiyası da yayın sıxılması və açılması, dənizin qabarması və

çəkilməsinə bənzər bir olaydır. Vaxtilə bir fəlsəfə elmi var idi, bütün elmlər başlanğıcını oradan götürüb, bütün elm sahələri fəlsəfə deyilən bilik xəzinəsinin diferensiyasının məhsuludur. Yay müəyyən həddə kimi sıxıla bildiyi kimi, elmlərin diferensiyası da əbədi deyil, müəyyən zamandan sonra əks proses inteqrasiya baş verməlidir. Bu, dialektik vəhdətdə olan bir prosesin, qanunauyğunluğun nəticəsi kimi səciyyələndirilə bilər.

Süni intellekt elmi də bu dialektik prosesin davamıdır, indiyə qədər diferensiasiyaya uğramış, mikrosahələrə bölünmüş elmlər inteqrasiyaya meyl edir, vahid elmi konsepsiya, vahid elmi istiqamət ətrafında mərkəzləşmiş olurlar.

Dilçiliyin süni intellekt problemlərinə inteqrasiyasının mahiyyətində də bunlar durur və bu, elmlərin diferensiasiyası və inteqrasiyası kimi dialektik idrak prosesinin qanunauyğun nəticəsi kimi səciyyələndirilə bilər.

Dilçilik sahəsində aparılan elmi-tədqiqat işlərinin süni intellekt problemləri çərçivəsində dəyərləndirilməsi əhəmiyyətli məsələdir. Bu həm də dilçilik sahəsində araşdırmalara yeni baxış bucağından yanaşmaq deməkdir.

Süni intellekt problemləri kontekstində yaxın gələcəkdə dilçilik elminin qarşısında duran vəzifələr çoxdur: Azərbaycan dilinin mükəmməl korpusunun yaradılması sahəsində işlərin yekunlaşdırılması və internet məkanında yerləşdirilməsi, təbii dilin emalı sistemlərinin təkmilləşdirilməsi, formal təhlil sistemləri çərçivəsində bütün dil səviyyələrində linqvistik təhlil analizatorlarının yaradılması, maşın tərcüməsi sahəsində dünya standartlarına uyğunlaşma, mətni səsə (nitqə), səsi (nitqi) mətnə çevirən mükəmməl sistemlərin yaradılması və onların məlumat axtarışı sorğu sistemlərində tətbiqi perspektivlərinin nəzərdən keçirilməsi, irihəcmli, əhatəli statistik lüğətlərin, konkordansların, tezaurusların tərtibi və həmin lüğətlərin digər elektron lüğətlərlə əlaqələndirilməsi sahəsində dönüş

yaradılması, Azərbaycan dilinin dünya internet məkanında təmsili məsələlərinin monitorinqi, təhlili və təşkili ilə bağlı işlərin intensivləşdirilməsi və s.

Dilçiliyin bir çox məsələlərinin süni intellekt problemləri kontekstində dəyərləndirilməsi və öyrənilməsi cəmiyyət üçün faydalı, səmərəli elmi-tədqiqatların aparılması və bu istiqamətdə işlərin genişləndirilməsi imkanı yaradacaqdır.

QISALTMALAR VƏ ŞƏRTİ İŞARƏLƏR

- ADMK – Azərbaycan Dilinin Milli Korpusu
ADMF – Azərbaycan Dilinin Maşın Fondu
AL – Avtomatik lüğət
AMEA – Azərbaycan Milli Elmlər Akademiyası
ADİL – Azərbaycan dilinin izahlı lüğəti
DDA – Doktorluq dissertasiyasının avtoreferatı
EİF – Elmin İnkişafı Fondu
NDA – Namizədlük dissertasiyasının avtoreferatı
MT – Maşın tərcüməsi
RDMK – Rus Dilinin Milli Korpusu
BNC – British National Corpus
BMK – Britaniya Milli Korpusu
DDİ – Doğal Dil İşləməsi
İCE – International Corpus of English
KİV – kütləvi informasiya vasitələri
UK – Upsal Korpusu
BC – Brown Korpusu
Sİ – Süni intellekt
TDMF – Türk Dillərinin Maşın Fondu
TUD – Türkçe Ulusal Derlem
TÜBİTAK – Türkiyə Bilimsel ve Teknolojik Araştırma Kurumu
TNC – Turkish National Corpus
MAİK – Müasir Amerika İngiliscəsi Korpusu
CCAE – Corpus of Contemporaru American English
F (a) – feilin arzu şəkli
F (v) – feilin vacib şəkli
F (ə) – feilin əmr şəkli
F (x) – feilin xəbər şəkli
F (l) – feilin lazım şəkli
F (d) – feilin davam şəkli

F (m1) – feilin məlum növü
F (m2) – feilin məchul növü
F (m0) – məsdər
F (q) – feilin qarşılıq növü
F (i) – feilin icbar növü
F (qy) – feilin qayıdış növü
F (e) – feilin şərt şəkli
İ – isim
S – sifət
Z – zərf
TSB – təyini söz birləşməsi
F (m0t) – məsdər tərkibi
FST – feili sifət tərkibi
ŞŞ – şəxs şəkilçisi
AQ – ad qrupu
FQ – feil qrupu
MQ – mübtəda qrupu
N – təyin
M – mübtəda
XQ – xəbər qrupu
ZS – zərflik
Tm/Zs – tamamlıq/zərflik
XB – xəbər
P – qoşma
B – bağlayıcı
Ə – ədat
MS – modal sözlər
B1 – bağlamalar
Sa – say
Əv – əvəzlik
V – a, ə
X – k, y, q, ğ
Y – ı, i, u, ü

İSTİFADƏ OLUNMUŞ ƏDƏBİYYAT

Azərbaycan dilində:

1. Abdullayev Ə. Müasir Azərbaycan dilində tabeli mürək-kəb cümlələr. Bakı, “Maarif”, 1974, 390 s.
2. Abdullayev Ə., Seyidov Y., Həsənov A. Müasir Azərba-ycan dili (sintaksis), Bakı, “Maarif”, 1972, 464 s.
3. Abdullayev K. Azərbaycan dili sintaksisinin nəzəri prob-lemləri. Yenidən işlənmiş və əlavələr edilmiş ikinci nəşri. Bakı: MTM-İnnovation, 2016, 360 s.
4. Abdulla K., Əliyev R. Kitabı-Dədə Qorqud və qeyri-səlis məntiq. Bakı, TEAS Press Nəşriyyat Evi, 2023, 180 s.
5. Abdullayev Ə. Aktual üzvlənmə və mətn. Bakı, Xəzər Universiteti nəşriyyatı, 1998, –190 s.
6. Abdullayev Ə. Mətni anlama modelləri. Bakı, “Səda” nəş-riyyatı, 1999, –344 s.
7. Adilov M.İ., Verdiyeva Z.N., Ağayeva F.M. İzahlı dilçilik terminləri (sorgu lüğəti), Bakı, “Maarif”, 1989, 364 s.
8. Axundov A.A. Riyazi dilçilik: Xülasələr və təfriqələr. Dərs vəsaiti. Bakı, ADU nəşriyyatı, 1979, 79 s.
9. Axundov A.A. Seçilmiş əsərləri. II cild. Bakı, “Elm və təhsil”, 2012, 464s.
10. Axundov A.A. Ümumi dilçilik. Bakı, “Maarif”, 1979, – 209 s.
11. Axundov A. Azərbaycan dilinin fonemlər sistemi (fizio-loji, akustik, statistik, fonoloji tədqiqat təcrübəsi). Bakı, “Maarif”, 1973, 304 s.
12. Axundov A.A. Azərbaycan dili saıtlərinin statistik təhlili təcrübəsindən. “ADU-nun elmi əsərləri (Dil və Ədəbiyyat seriyası)”, №5, 1963, s.63-67

13. Azərbaycan qəzet dilinin tezlik lüğəti (tərtibçilər: K.Vəliyeva, M.Mahmudov, V.Pines). Bakı, “Nurlan”, 2004, 264 s.
14. Azərbaycan dilinin izahlı lüğəti. Dörd cildə. Bakı, “Şərq–Qərb” nəşriyyatı, 2006, (I cild – 744 s., II cild – 792 s., III cild – 672 s., IV cild – 712 s.).
15. Azərbaycan dilinin orfoqrafiya lüğəti. 6-cı nəşr, Bakı, “Şərq–Qərb” Nəşriyyat evi, 2013, 840 s.
16. Azərbaycan dilinin əks əlifba lüğəti (tərtibçilər: M.Ə.Mahmudov, Ə.B.Fətullayev), Bakı, “Nurlan”, 2004, 524 s.
17. Azərbaycan dilinin tezlik lüğəti (söz kökləri), I cild (tərtibçilər: M.Ə.Mahmudov, Ə.B.Fətullayev və b.), Bakı, “Elm”, 2010, 464 s.
18. Azərbaycan dialektoloji lüğəti. I cild (A-L)/redaktorlar: M.Ş.Şirəliyev, M.İ.İslamov.- Ankara: Türk Dil Kurumu, 1999, 374 s.
19. “Azərbaycan Respublikasında dövlət dili haqqında” Azərbaycan Respublikasının Qanunu (http://files.preslib.az/projects/remz/pdf/atr_dil.pdf)
20. Azərbaycan dilinin qloballaşma şəraitində zamanın tələblərinə uyğun istifadəsinə və ölkədə dilçiliyin inkişafına dair Dövlət Proqramı haqqında Azərbaycan Respublikası Prezidentinin Sərəncamı (<http://president.az/articles/4934>)
21. Aksan, Y., Mersinli, Ü., Yaldır, Y. (2011). İlköğretim Türkçe Ders Kitapları Derlemi ve Türkçe Ulusal Dil Derlemi örneklemindeki sözcük sıklıkları. Türkçe öğretimi üzerine çalışmalar. İzmir. Dokuz Eylül Üniversitesi Yayınları, ss.397-408.
22. Aksan, Y. Mersinli, Ü. Yaldır, Y., Demirhan, U.U. (2011). Türkçe Ulusal Dil Derlemi projesi biçimbirim çalışmalarında belirsizliklerin sınıflandırılması ve dağılımı. Ulusal Dilbilim Kurultayı bildiri kitabı. Adana. Çukurova Üniversitesi, ss.219-226.

23. Dilçilik ensiklopediyası (dərs vəsaiti). Prof. F.Veysəllinin redaktəsi ilə – I cild, Bakı, Mütərcim, 2006, 516 s., II cild, Bakı, “Mütərcim”, 2008, 528 s.
24. Əmirov Z.M. Azərbaycan dilinin formal mətnlərinin yaradılması və onların əsasında linqvistik prosessorun qurulması. Nam. diss. Avtoreferatı. Bakı, AMEA akad. Ə.Hüseynov adına Kibernetika İnstitutu, 2006, 19 s.
25. Əliquliyev R., Şükürlü S., Kazımova S. Elmi fəaliyyətdə istifadə olunan əsas terminlər. Bakı, “İnformasiya Texnologiyaları”, 2009, 201 s.
26. Əsədova M.N. İngilis və Azərbaycan dillərində tamamlıq budaq cümləsi. Nam. diss. avtoreferatı, Bakı, 2004, 25 s.
27. Əliyev Ə.A. İngiliscə-azərbaycanca məşin tərcüməsi sistemində feilin çoxmənalılığının alqoritmik həlli. Nam. disser.avtoreferatı. Bakı, Azərbaycan Dövlət Universiteti, 2012, –20 s.
28. Fətullayev Ə.B. Azərbaycan – ingilis məşin tərcüməsi sistemi üçün rəqəmsal modelləşdirmə metodunun işlənilib hazırlanması və tətbiqi. Nam. diss. Avtoreferatı, Bakı, AMEA akad.Ə.Hüseynov adına Kibernetika İnstitutu, 2006, 19 s.
29. Fətullayev Ə., Mahmudov M., Məmmədov R. və b. Mətnin səsləndirilməsi sistemi və Azərbaycan dilinin fonemlərinin fono-akustik bloku. Terminologiya məsələləri. Bakı, Elm nəşriyyatı, 2015, №1, s.33-49.
30. Fətullayev Ə., Mahmudov M., Fətullayev R., Dərvişov R. NLP sistemlərində Azərbaycan dilində tabeli mürəkkəb cümlələrin formal meyarlara görə təsnifi və təhlili. Filologiya məsələləri, № 7, Bakı, 2015, “Elm və təhsil” nəşriyyatı, s.3-9.
31. Fətullayev R.Ə. Məşin tərcüməsi sistemində Azərbaycan dilində cümlənin formal sintezi alqoritmlərinin yaradılması

- və tətbiqi. Nam. diss. avtoreferatı, Bakı, AMEA akad. Ə.Hüseynov adına Kibernetika İnstitutu, 2010, 19 s.
32. Hüseynzadə M. Müasir Azərbaycan dili. Morfologiya. III hissə. Ali məktəblər üçün dərslik. Bakı, “Maarif”, 1973, 358 s., 1983, 319 s.
33. Xəlili Ə.M. Deduktiv məşının biliklər bazasının tərkib hissəsi kimi “Məhdud Azərbaycan dili”nin formal qrammatikasının işlənilib hazırlanması. Nam. diss. avtoreferatı, Bakı, AMEA Nəsimi adına Dilçilik İnstitutu, 2009, 33s.
34. İngiliscə-azərbaycanca lüğət. Prof. O.Musayevin təşəbbüsü, rəhbərliyi və redaktorluğu ilə “Qismət” nəşriyyatı, Bakı, 2003, 1696 s.
35. Kazımov Q.Ş. Müasir Azərbaycan dili. Morfologiya. Bakı, “Elm və təhsil”, 2010, 400 s.
36. Kazımov Q.Ş. Müasir Azərbaycan dili. Sintaksis. Bakı, “Elm və təhsil”, 2010, 500 s.
37. Kazımov Q. Mətn dilçiliyi problemləri. Ə.Abdullayevun “Aktual üzvlənmə və mətn” kitabına ön söz. Bakı, Xəzər Universiteti nəşriyyatı, 1998, s. 5-21.
38. “Kitabi-Dədə Qorqud”un statistik təhlili (tərtibçilər: Vəliyeva K., Mahmudov M., Pines V., Rəhmanov C., Sultanov V.), Bakı, “Elm”, 1999, 248 s.
39. Qasımov V.Ə. İnformasiya axtarış üsulları və sistemləri. Dərslik. Bakı: MTN-in Maddi-texniki Təminat Baş İdarəsinin Nəşriyyat – Poliqrafiya Mərkəzi, 2015, 288 s.
40. Mahmudov M. Azərbaycan dilinin milli korpusu süni intellekt problemləri kontekstində. Ümummillə lider Heydər Əliyevin 100 illik yubileyinə və Beynəlxalq Ana dili gününə həsr olunmuş “ İKT-nin Azərbaycan dili problemləri, Azərbaycan dilinin İKT problemləri” adlı respublika elmi-praktiki konfransının materialları, Bakı, 21-22 fevral 2023-cü il, s. 44-49

41. Mahmudov M. Kompüterlər linqvistik araşdırmalarda: problemlər, perspektivlər. Dil dərgisi. Türkoloji. Ocak, 1995, 27, Ankara, s.10-13
42. Mahmudov M.Ə. Maşın tərcüməsi sistemində avtomatik morfoloji analizin yeri. Azərbaycan EA-nın xəbərləri. Ədəbiyyat, dil və incəsənət seriyası, №1, 1981, s.68-72
43. Mahmudov M. Kompüter dilçiliyi. Bakı, “Elm və təhsil”, 2013, 356 s.
44. Mahmudov M. Mətnin formal təhlili sistemi. Bakı, “Elm”, 2002, 244 s.
45. Mahmudov M.Ə. Gödelin natamamlıq teoreminin linqvistik interpretasiyası. “Dil, etnos və etnonimiya”. Bakı, “Örnək”, 1994 a, s. 82-83.
46. Mahmudov M.Ə. Tezlik lüğətləri dil tədrisində. Azərbaycan EA-nın xəbərləri, 1992, №3-4, s.99-102
47. Mahmudov M.Ə. Azərbaycan mətnlərinin avtomatik işlənməsi sistemi. Doktorluq dissertasiyasının avtoreferatı, Bakı, 1994 b, 64 s.
48. Mahmudov M.Ə., Sultanov V.S. Dilimizin tarixi və kompüter dilçiliyi: problemlər, perspektivlər.” Tarixi və tətbiqi dilçilik”, Bakı, 1993, s.3-4
49. Mahmudov M., Cəlilov F. Tezlik lüğətlərinin tərtib olunma qaydaları haqqında. ADU-nun elmi əsərləri (Dil və Ədəbiyyat seriyası), №3, 1977, s.49-53
50. Mahmudov M.Ə., Pines V.Y. Maşın tərcüməsi və avtomatlaşdırılmış məlumat sistemlərində Azərbaycan mətnlərinin morfoloji təhlil alqoritmi. İnformasiya vərəqi.” Energetika və avtomatika” seriyası, № 2, Bakı, AzETETİİ, 1980
51. Mahmudov M.Ə. Süni intellekt maşın tərcüməsi sistemlərində. Tətbiqi dilçiliyin müasir problemləri. Beynəlxalq elmi konfrans (Azərbaycan, Bakı, 27-28 noyabr, 2012-ci il), Bakı, “Mütərcim”, 2013, s.61-63.

52. Mahmudov M. Ənənəvi dilçiliyə qeyri-ənənəvi baxış. AMEA-nın xəbərləri. Humanitar Elmlər Seriyası, 2006, №2-4, s.101-103.
53. Mahmudov M., Fətullayev Ə., Abbasov S., Fətullayev R., Abdullayev N. Azərbaycan dili üçün NLP sistemləri və milli korpusun yaradılmasının nəzəri və tətbiqi məsələləri. Bakı, Türkologiya, 2016, №4, s.15-28.
54. Mahmudov M. Türk dillərinin milli korpusları. Bakı, "Elm və təhsil", 2018, 392 s.
55. Mahmudov M.Ə. Azərbaycan dilinin milli korpusunun yaradılmasının ilkin şərtləri və optimal strukturu. Время собирать камни... Professor İlyas Həmidovun 70-illiyinə həsr olunmuş məqalələr toplusu. Bakı, "Mütərcim", 2017, s. 155- 172.
56. Mahmudov M., Tahirov İ. və b. İntegrasiya olunmuş elektron lüğətlər sistemi Azərbaycan dilinin milli korpusunun yaradılmasında bir mərhələ kimi/ Türkologiya №1, Bakı, 2019, s. 66-80.
57. Məhəmməd Füzulinin nəzm əsərlərinin əlifba-tezlik sözlüyü (tərtibçilər: K.A.Vəliyeva, M.Ə.Mahmudov, C.Ə.Rəhmanov, V.S.Sultanov), Bakı, "Elm", 2004, 548 s.
58. Məmmədzadə S.S. Milli dil korpusları və onların formalaşması prinsipləri. Filologiya üzrə fəlsəfə doktoru elmi dərəcəsi almaq üçün təqdim edilmiş dissertasiya. Bakı, 2018, 149 s.
59. Məmmədov A. Mətn yaranmasında formal əlaqə vasitələrinin sistemi. Dokt. disser.avtoreferatı, Bakı, 2003, –54 s.
60. Müasir Azərbaycan dili. I cild. Fonetika, əlifba, orfoqrafiya, orfoepiya, leksikologiya. Bakı, "Elm", 1978, 324 s.
61. Müasir Azərbaycan dili. II cild. Morfologiya. Bakı, "Elm", 1980, 510 s.
62. Müasir Azərbaycan dili. III cild, Sintaksis. Bakı, "Elm", 1981, 444 s.

63. Novruzova N. Mətn sintaksisi. Bakı, “Təhsil” nəşriyyatı, 2002, –184 s.
64. Vəliyev A. Azərbaycan dilinin keçid şivələri, Bakı, “Elm”, 2005, 335s.
65. Vəliyeva K. A. Mətnin avtomatik təhlili və sintezi. Bakı, “Örnək” 1996, 158 s.
66. Vəliyeva K.A., Məmmədova M. Mətnlərin avtomatik redaktəsi. Bakı, “Azərbaycan Universiteti” nəşriyyatı, 2003, –89 s.
67. Pines V.Y., Mahmudov M.Ə. Elektron hesablayıcı maşınlar dilçilikdə. İcmal məlumatı. Bakı, AZETETMİ, 1977, 36s.
68. Rusca-azərbaycanca lüğət. Üç cilddə, Bakı, «Elm» nəşriyyatı, I cild, 1971, 606 s., II cild, 1976, 712 s., III cild, 1979, 556 s.
69. Seyidov Y. Azərbaycan dilində söz birləşmələri. Bakı Universiteti nəşriyyatı, 1992, 408 s.
70. Sadıkov T., Şarşembayev B. Manas destanı. Kırgızca-türkçe büyük dizin. Ankara, Türk Dil Kurumu Yayınları, 2011, 1647 s.
71. Sultanov V. Cümlənin semantik özəyi. Bakı, “Elm”, 1997.

İngilis dilində:

72. Aksan, Y. et al., "Construction of the Turkish National Corpus (TNC)", Proceeding of the Eight International Conference on Language Resources and Evaluation (LREC 2012), İstanbul, (2012) Aksan, M., Aksan, Y. (2015). Multi-word Units in Genre Specification. Dil ve Edebiyat Dergisi, 12 (1), 1-42.
73. Fatullayev R., Abbasov A., Fatullayev A. “Dilmanc” is the 1st MT system for Azerbaijani/Proceedings of SLTC 2008: Swedish Language Technology Conference, November 20 & 21, 2008, Stockholm, Sweden. pp.63-64.

74. Fatullayev R., Abbasov A., Fatullayev A. 2008 a. Set of active suffix chains and its role in development of MT system for Azerbaijani. In: Proc. of IMCSIT-08, Wisla, Poland, pp.363-368.
75. Kemal Oflazers Web Page. [https:// www. andrew. cmu. edu/user/ko/](https://www.andrew.cmu.edu/user/ko/)
76. Kibrik A.A., Tenishev E.R., Pocoluevskiy E.A. and Kormushin İ.V., 1997 Jazyki mira: Tjurkskie jazyki [Languages of the world: Turkic languages]. Moscow: Indrik. 542 p.
77. Mammadova S., Azimova G., Fatullayev A. Text corpora and its role in development of the linguistic technologies for the Azerbaijani language. The third International Conference “Problems of Cybernetics and Informatics”. September 6-8, 2010, Baku, Azerbaijan.
78. Say, B., Zeyrek, D., Oflazer, K. ve Özge, U., “Development of a corpus and a treebank for present-day written Turkish”, Current research in Turkish linguistics: proceedings of the 11th International Conference of Turkish Linguistics, 183-192, (2002).
79. Turkish morphological analyzer. <https://user.clarin.eu/resources/turkish-morphological-analyzer>.
80. Turing Alan (October 1950). Computing Machinery and Intelligence. *Mind*. LIX (236): 433-460. ISSN 0026-4423, doi:10. 1093mind/ LIX.236.433.

Rus dilində:

81. Алексеев П.М. Статистическая лексикография. Учебн. пособие. Ленинград, ленинградский пед. Институт им.А.И.Герцена. 1975, -120 с.
82. Бабанаров А. Разработка принципов построения словарного обеспечения турецко-русского машинного перевода, АКД, Ленинград, 1981, -17 с.

83. Большой энциклопедический словарь [http: www.edic.ru/res/](http://www.edic.ru/res/)
84. Бускунбаева Л.А., Сиразитдинов З.А. О проблемах национального корпуса башкирского языка. Материалы. «Современное казахское языкознание: актуальные вопросы прикладной лингвистики». Алматы 2012, с.54-55.
85. Велиева К.А. Формальное описание синтеза азербайджанского слово. АКД, Москва, 1971, 20 с
86. Гузев В.Г., Пиотровский Р.Г., Щербак А.М. О создании машинного фонда тюркских языков. «Советская тюркология», 1988, №2, с.92-101.
87. Девятков В.В. Системы искусственного интеллекта / Гл. ред. И.Б.Фёдоров. – М.: Изд-во МГТУ им. Н.Э.Баумана, 2001. – 352 с. – (Информатика в техническом университете). – bookwork.ru/book/devyatkov_sii
88. Доклады научной конференции «Корпусная лингвистика и лингвистика базы данных» / Под ред. А.С.Герда. СПб., 2002.
89. Есенбаев Ж., Махамбетов О., Карабалева М. Текстовый корпус казахского языка. Материалы. «Современное казахское языкознание: актуальные вопросы прикладной лингвистики». Международная научно-теоретическая конференция посвященная 75-летию юбилею профессора Жубанова Аскара Кудайбергеноулы, Алматы, 2012, с.61-66.
90. Жубанов А.К. Компьютерлик лингвистикага кіріспе, Алматы, Ал-Фараби атындагы Казак Улттык Университети, 2013, 204 бер.
91. Жубанов А.К. Казахское языкознание: прикладная лингвистика. Алматы, «КИЕ», 2012, 696 с.
92. Жубанов А., Жанабекова А. Корпустык лингвистика. – Алматы: «Казак тили» баспасы, 2017, 336 б.

93. Жданов А.А. Автономный искусственный интеллект. – М.: БИНОМ. 5-е изд. электрон. Лаборатория знаний, 2020. – 362 с. – avidreaders.ru/read-book/avtonominyu-iskusstvennyu-intellekt.html
94. Захаров В.П. Корпусная лингвистика. Учебно-методическое пособие. Санкт-Петербург, Санкт-Петербургский государственный университет, 2005, 48 с.
95. Иващенко В.М. Язык и стиль научно-технического текста (на материалах английского, азербайджанского и русского языков). Автореф. канд. дисс., Баку, 2009, – 42 с.
96. Ингве В. Значение исследования в области машинного перевода. НТИ.сер.2, №7, 1965, с. 42-47
97. Искендерова Н.А. Программно-частотный и качественно – семантический анализ адекватности переводов и его применение к переводам на различных языках. АКД, Баку, И-т Языкознания им. Насими, 2008, 19с.
98. Компьютер обретает разум = Artificial Intelligence Computer Images / под ред. В.Л. Стефанюка. – Москва: Мир, 1990. – 240с. – soft – munea.ucoz.ru/news/kompyuter_obretaet_razum/2012-02-28-6655
99. Кулагина О.С. Исследования по машинному переводу. Москва, «Наука», 1979, -320 с.
100. Леонтьева Н.Н. Автоматическое понимание текстов: Системы, модели, ресурсы. Москва, Академия, 2006, 304 с.
101. Лорьер Ж.Л. Системы искусственного интеллекта. – М.: Мир, 1991. – 568 с. – log-in.ru/books/sistemy-iskusstvennogo-intellekta-1-zh-1-loror-programming
102. Мамедова М.Г. Автоматизированный отбор лексики в информационно-поисковый тезаурус на основе анализа терминологических словарей. АКД, Москва, 1984, 25 с.

103. Мамедова М.Г., Мамедова З.Ю. Машинный перевод: эволюция и основные аспекты моделирования. Баку, изд. «İnformasiya texnologiyaları», 2006, –158 с.
104. Мамедова М.Г., Пинес В.Я. Принципы построения информационно-поискового тезауруса по строительству (Тез. докл. семинара – совещ. «Республиканская система научно-технической информации и опыт организации информационного обеспечения народного хозяйства». Баку, 1989, с. 39-41.
105. Машинные фонды языков народов СССР: Материалы рабочего совещание (Тбилиси, 15–22 сентября 1987 г.), Тбилиси, 1988.
106. Машинные фонды языков народов СССР: Материалы рабочего совещание (Таллин, 19–22 декабря 1988 г.), Таллин: Инст. Яз. и литературы, АН Эст. ССР, 1988.
107. Марчук Ю.Н. Основы компьютерной лингвистики. Москва, изд.-во МГУ, 2002, –234 с.
108. Машинный фонд русского языка: идеи и суждения. М., Наука, 1986, –240 с.
109. Махмудов М.А. Система автоматической переработки тюркского текста на лексико-морфологическом уровне. Баку, «Елм», 1991, 143 с.
110. Махмудов М.А. Создание интегрированной электронной системы в связи с упорядочением словарного состава азербайджанского языка. – Материалы международной научно-теоретической конференции «Цифровые технологии в языкознании: Создание национального корпуса казахского языка» посвященной 80-летию профессора Жубанова Аскара Кудайбергеноулы. Алматы, «Ғылым Ордасы» баспасы, 2017, с.137-142.
111. Махмудов М.А. Лингвостатистика: проблемы, перспективы. Современное казахское языкознание: Актуальные вопросы прикладной лингвистики. Между-

- народная научно-теоретическая конференция посвященная 75-летию юбилею профессора Жубанова Аскара Кудайбергенулы, Алматы, 2012, с. 10-14.
112. Махмудов М.А. Разработка системы формального морфологического анализа тюркской словоформы. АКД, Баку, 1982, 26с.
113. Махмудов М.А. Принципы инженерно-лингвистического моделирования тюркского текста (R.Piotrovski, V.Pines, T.Sadikovla birlikdə). *Türkologiya* №1-4, Bakı, 2001, s. 53-69.
114. Михайлов М.Н. Контекстно-свободная лемматизация как временное решение насущных проблем//Алфавит. Смоленск, СПГУ, 2002, с.197-210.
115. ФИЗМАТЛИТ.2004. – 208 с. – studmed.ru>smolin-dv-vvedenie-v-is...>
116. Нелюбин Л.Л. Компьютерная лингвистика и машинный перевод. М., ВЦП, 1991, 154 с.
117. Нильсон Н. Искусственный интеллект. – М.: Мир, 1973. – 274 с.- log-in.ru/books/nilson-iskusstvennyiy-intellekt-nilson-n-programming
118. Пинес В.Я. Моделирование структуры азербайджанских глагольных форм в связи с проблемой автоматического словаря, АКД, Москва, 1970, 19 с.
119. Пиотровский Р.Г. Инженерная лингвистика и теория языка. Ленинград. « Наука», 1979, 112 с.
120. Пиотровский Р.Г. Текст, машина, человек. Ленинград. «Наука», 1975, 327с.
121. Рассел С., Норвиг П. Искусственный интеллект: современный подход = Artificial Intelligence: a Modern Approach / Пер. с англ. и ред. К. А. Птицына. – 2-е изд. – М.: Вильямс, 2006. – 1408 с. – studizba.com>files/show/djvu/3615-...>

- 122.Рахманов Дж.А. Статистико-дистрибутивный анализ азербайджанского текста (на уровне графем и фонем), АКД, Баку, 1988, 23с.
- 123.Ревзин И.И. Пути преодоления кризиса в вычислительной лингвистике.НТИ.сук.2, №2, 1968
- 124.Садыков Т., Шаршембаев Б. Кыргыз-казак текстеринин салыштырма статистикалык айрым онуттору. Материалы. “Современное казахское языкознание: актуальные вопросы прикладной лингвистики”. Международная научно-теоретическая конференция посвященная 75-летнему юбилею профессора Жубанова Аскара Кудайбергенулы, Алматы, 2012, с.20-25
- 125.Селегей В. Электронные словари и компьютерная лексикография //http://www.linevoda.ru/ – www.tech-tr.ru
- 126.Сиразитдинов З.А., Ишмухаметова А.Ш., Ибрагимова А.Д. Информационное обеспечение башкирского языка: опыт и перспективы. Современное казахское языкознание: актуальные вопросы прикладной лингвистики. Материалы межд. научн. конф. Алматы, 2012, с.54 - 5.
- 127.Современное казахское языкознание: актуальные вопросы прикладной лингвистики. Материалы международной научно-теоретической конференции, посвященн. 75-летнему юбилею Жубанова А.К., Алматы, И-т Языкознания им.А.Байтурсынова, 2012, –488 с.
- 128.Смолин Д.В. Введение в искусственный интеллект: конспект лекций. – М.: ФИЗМАТЛИТ.2004. – 208 с. – studmed.ru>smolin-dv-vvedenie-v-is...>
- 129.Труды Международной научной конференции «Корпусная лингвистика» – 2004», /Под ред. А.С.Герда. СПб., 2004.
- 130.Успенский В.А. Теорема Геделя о неполноте. Москва, « Наука», 1982, 111с.

131. Фрумкина Р.М. Статические методы изучения лексики. Москва, «Наука», 1964, 110 с.
132. Хант Э. Искусственный интеллект /Artificial intelligence / Под ред. В.Л. Стефанюка.-Москва: Мир, 1978.-562 с. – [log.in ru/books/khant-e-iskusstvennyiy-intellekt-khant- I – informasiya](http://log.in.ru/books/khant-e-iskusstvennyiy-intellekt-khant-I-informasiya)
133. Частотный словарь русского языка. Москва, изд. «Русский язык. 1977.-935 с.
134. Черкасова Г.А. Русский ассоциативный тезаурус: компьютерная создания и издания // Этнокультурная специфика языкового сознания/ Под ред. Н.В. Уфимцевой. М., 1998, с. 129-134.
135. Чижиковский В.А., Беляева Л.Н. Тезаурус в системах автоматической переработки текста. Кишинев, «Штиинца, 1983, -163 с.
136. Шаров С.А. Представительный корпус русского языка в контексте мирового опыта// НТИ Сер. 2. -2003, № 6, с.9-17.
137. Шаров С.А. Корпусная лингвистика. [http: //www. dialoq-21. ru](http://www.dialoq-21.ru)
138. Шаров С.А. Частотный словарь русского языка [http: www.artint.ru/projects/frylist.asp](http://www.artint.ru/projects/frylist.asp).
139. Шайкевич А.Я. Дифференциальные частотные словари и изучение языка Достоевского (на примере романа «Идиот») //Слово Достоевского, Москва: ИКЯ РАН, 1996, с.195-253. Также [http: //irlras-cfrl.rema.ru:8101/publications/ryano.htm](http://irlras-cfrl.rema.ru:8101/publications/ryano.htm)
140. Щерба Л.В. Опыт общей теории лексикографии.-В кн. «Языковая система и речевая деятельность», Ленинград, 1974.

İnternetə istinadlar:

141. İB: İngilis Bankı <http://titania.cobuild.collins.co.uk/>

- 142.BK: Braun Korpusu <http://www.hd.uib.no/icame> /brown/bcm.html
- 143.BMK: Britaniya Milli Korpusu <http://sara.natcorp.ox.ac.uk/lookup.html>
- 144.İCE: İngilis dilinin Beynəlmiləl Korpusu <http://www.ucl.ac.uk/english-usage/ice/>
- 145.Penn Threebank: <http://www.cis.upenn.edu/-treebank/>
- 146.RDMK: Rus dilinin Milli Korpusu <http://www.ruscorpora.ru/>
- 147.RDBK: Rus dilinin Böyük Korpusu <http://bokrcorpora.narod.ru/>
- 148.Nişanlanmış etalon korpus <http://corpora.yandex.ru/>
- 149.UK: Upsal Korpusu (Tyubingen Universiteti) <http://www.sfb441.uni-tuebingen.de/bl/en/korpora.html>
- 150.RDMF Rus dilinin Maşın Fondu <http://irlras-cfml.rema.ru/>
- 151.Ümumi və kompüter leksikologiyası və leksikoqrafiyası laboratoriyası (rəhbər – A.A.Polikarpov) <http://www.philol.msu.ru/~lex/main.htm>
- 152.Dialing: morfoloji analizator <http://www.aot.ru/>
- 153.Müasir rus dilinin tezlik lüğəti <http://bokrcorpora.narod.ru/frqllist/frqllist.html>
- 154.Bolqar dili: <http://www.lml.bas.bg/>
- 155.İtalyan dili: <http://www.cilta.unibo.it>
- 156.Çin dili: <http://www.sinica.edu.tw/ftms-bin/ftmsw/3>
- 157.Alman dili: <http://corpora.ids-mannheim.de/-cosmas/>
- 158.NEGRA – alman dilinin sintaktik annotasiyalı korpusu <http://www.coli.uni-sb.de/sfb378/negra-corpus/>
- 159.Portuqal dili <http://acdc.linguatca.pt/>
- 160.Xorvat dili <http://www.hnk.ffzg.hr/>
- 161.Çex dili <http://ucnk.ff.cuni.cz/>
- 162.PDT – çex dilinin sintaktik annotasiyalı korpusu <http://ufal.mff.cuni.cz/pdt/pdt.html>
- 163.Eston dili <http://psych.ut.ee/gling/en/corpusb/>

164. https://www.tnc.org.tr/wp-content/uploads/2018/05/AksanMersinli_30UDK.pdf
165. Windows. ru.wikipedia.org/wiki/windows
166. Национальный корпус русского языка. <http://www.ruscorpora.ru/corpora-info>
167. Dilmanc. <http://www.dilmanc.az/az>
168. Dilmanc. <http://az.wikipedia.org/wiki/dilmanc>
169. Poliqlot. www.poliqlot.az
170. Linqvo Online linqvo.ru
171. Britannica Online <http://www.eb.com>
172. Google. en.wikipedia.org/wiki/google
173. Искусств. интеллект. ru. wikipedia.org/wiki/искусственный_интеллект
174. Энтропия. ru.wikipedia.org/wiki/энтропия
175. Süni intellekt. az.wikipedia.org/wiki/süni_intellekt
176. Axtarış sistemləri. az.wikibooks.org/wiki/internetdə_axtarış_sistemləri/Axtarış_sistemləri
177. Inform. axtarışı. az.wikipedia.org/wiki/informasiya_axtarışı
178. az.wikibooks.org/wiki/internet/Yaranma_tarixi
179. Qeyri-səlis məntiq. az.wikipedia.org/wiki/qeyri-səlis_məntiq
180. Yahoo! Search. az.wikipedia.org/wiki/yahoo!
181. Google. az.wikipedia.org/wiki/google
182. Yandex. az.wikipedia.org/wiki/yandex
183. Национальный корпус казахского языка. <http://qazcorpus.kz>
184. Национальный корпус узбекского языка. <http://uzbekcorpus.uz>
185. Azərbaycan dilinin informasiya bazası. <http://azerbaycandili.az>
186. Azərbaycan Milli Elmlər Akademiyası. science.qov.az/az
187. AMEA Nəsimi adına Dilçilik İnstitutu. dilcilik.az

188. Azərbaycan dilinin elektron lüğətlər korpusu.
korpus.azerbaycandili.az
189. SMS. ru.wikipedia.org/wiki/sms
190. Google (компания). ru.wikipedia.org/wiki/google_
(компания)
191. news.milli.az/interest/1634.html
192. Распознавание речи. ru.wikipedia.org/wiki/
Распознавание_речи.

MƏSUD MAHMUDOV

**SÜNİ İNTELLEKTİN
LİNQVİSTİK PROBLEMLƏRİ**

“Elm və təhsil” nəşriyyatının direktoru:

İNAL MƏMMƏDLİ

Dizayner: Zahid Məmmədov

Texniki redaktor: Rövşanə Nizamiqızı

Çapa imzalanmışdır: 15.07.2024

Format: 70x100

Həcmi: 1/16

Sifariş:

Tiraj: 300

Kitab “Elm və təhsil” nəşriyyat-poliqrafiya
müəssisəsində nəşr olunmuşdur.
E-mail: elm.ve.tehsil@mail.ru
Tel: 497-16-32; 050-311-41-89
Ünvan: Bakı, İçərişəhər, 3-cü Maqomayev 8 /4