

AZƏRBAYCAN MİLLİ ELMLƏR AKADEMİYASI
NƏSİMİ ADINA DİLÇİLİK İNSTİTUTU

MƏSUD MAHMUDOV

KOMPÜTER DİLÇİLİYİ

33565

92
12.12.2013
SİMOAYI
DÖVLET İNVERSİTETİ
KİTABXANA

“Elm və təhsil”
BAKI – 2013

80

C 2 +

M15

Azərbaycan Milli Elmlər Akademiyası Nəsimi adına
Dilçilik İnstitutunun Elmi Şurasının qərarı ilə çap olunur

- Redaktorlar:** **Ağamusa Axundov**
*AMEA-nın həqiqi üzvü,
filologiya elmləri doktoru, professor*
- İlham Tahirov**
filologiya elmləri doktoru
- Rəyçilər:** **Telman Əliyev**
*AMEA-nın həqiqi üzvü,
texnika elmləri doktoru, professor*
- Məsumə Məmmədova**
Texnika elmləri doktoru, professor
- Əbülfət Fətullayev**
Texnika üzrə fəlsəfə doktoru, dosent
- Kamilə Vəliyeva**
Filologiya elmləri doktoru
- Sayalı Sadıqova**
Filologiya elmləri doktoru, professor

Məsud Mahmudov. Kompüter dilçiliyi.
Bakı, "Elm və təhsil", 2013, 356 s.

Monoqrafiya kompüter dilçiliyinin nəzəri və təcrübi məsələlərinə həsr olunmuşdur. İlk dəfə olaraq kompüter dilçiliyinin yaranma səbəbləri, mahiyyəti, istiqamətləri və sahələri haqqında məlumat verilir, Azərbaycan dilinin milli korpusu və onun tərkib hissələri olan statistik lüğətçilik, məşın tərcüməsi və formal linqvistik təhlil sistemləri hərtərəfli şərh olunur.

Monoqrafiya nəzəri və tətbiqi dilçiliyin aktual problemləri və ən yeni sahələri ilə maraqlananlar üçün nəzərdə tutulmuşdur.

ISBN 978-9952-8176-2-1

© "Elm və təhsil", 2013

“KOMPÜTER DİLÇİLİYİ” ƏSƏRİNƏ ÖN SÖZ VƏ YA DİLÇİLİK MÜASİR DƏQİQ ELMLƏR ORBITİNDƏ

Riyazi metodların və “riyazi ruh”un dilçiliyə nüfuz etməsi dilçiliyin dəqiq elmlər və obyektivlik istiqamətində inkişaf etməsinə kömək etdi. Dilçiliyin riyaziyyatla yaxınlaşmasının səbəbləri, dilçilikdə riyazi metodların tətbiq edilə bilməsinin sərhədləri, riyaziyyat və digər dəqiq elmlərlə məşğul mütəxəssislərlə dilçilərin qarşılıqlı anlaşmasını əngəlləyən faktorların təbiəti və mahiyyəti barədə elmi ədəbiyyatda kifayət qədər fikir və mülahizələr yer almaqdadır. Keçən yüzilliyin 50-ci illərində dilin strukturunun tədqiq edilməsi üçün riyazi metodların tətbiqi barədə fikirlər və bu yöndə dilçilərin riyaziyyat, kibernetika, informasiya texnologiyaları mütəxəssisləri ilə əməkdaşlığa başlama təşəbbüsləri çoxlarının təəccübünə səbəb olmuşdu: dilçiliyin də daxil olduğu humanitar elmlərin riyaziyyatla və digər dəqiq elmlərlə heç bir əlaqəsinin olmadığı və ümumiyyətlə, belə əlaqənin mümkün ola bilməyəcəyi barədə düşünürdülər. Əslində təbii dil ilə riyaziyyat arasında əlaqənin olması o dövr üçün heç də yeni bir kəşf deyildi. L.S.Vıqotski hələ 1934-cü ildə “Dil və təfəkkür” kitabında yazırdı ki, “Riyaziyyatda ilk olaraq dildən gələn təfəkkürü görənlər, lakin onun üstündən keçən, görünür, Dekart olmuşdur. Təkcə bir şeyi demək olar: bizim adi danışmaq dilimiz ona xas olan qrammatik və psixoloji tərəddüd (колебание) və uyğunsuzluqlar səbəbindən riyazi ideal və fantastik harmoniya arasında mütəhərrik tarazlıq vəziyyətində və fəzilsiz hərəkətdə olur ki, bunu biz təkamül adlandırırıq.” (Выготский Л.С. Мышление и речь. Психологические исследования. (Под ред. и со вступительной статьей В.Колбановского). с. 273).

Qədim Yunanıstanda qrammatik kateqoriyalar haqqında meydana çıxmış təlimdə dilin təşkilinin bir sıra mühüm aspektlərini mücərrəd modellərin köməyiylə təsvir etmək istiqamətində təsəvvürlər mövcud idi. Lakin həmin mücərrəd modellərin yaradılması üçün tələb olunan mücərrəd təfəkkür səviyyəsinin nə də-

rəcədə yüksək olması barədə fikir yürütmək indi imkan xaricindədir. Bununla yanaşı, məkan formalarının təsviri üçün qədim yunan riyaziyyatçılarının (Pifaqor, Fales, Evklid və s.) yaratdığı modellər də elm aləminə yaxşı məlumdur.

Əslində təəccüb ediləsi fakt ondan ibarətdir ki, dilin strukturunun öyrənilməsi üçün həqiqi riyazi vasitələrdən istifadə edilməsi təşəbbüsləri eyni vaxtda bir neçə elmin (məsələn, fəlsəfə, məntiq, riyaziyyat, dil və s.) sirlərinə vaqif olmuş şəxslərin yaşadığı zamanlarda deyil, məhz XX yüzillikdə göstərilmişdir, halbuki bu dövrdə elm sahələrində daha çox diferensiasiya getmişdir. Bu “gecikmə”nin bir çox səbəbləri göstərilə və əsaslandırıla bilər, lakin onların ikisi daha vacibdir. Bu səbəblərdən biri dilçiliyin, digəri isə riyaziyyat da daxil olmaqla digər dəqiq elmlərin inkişaf istiqamətləri bağlıdır. Onları ümumiləşdirərək aşağıdakı şəkildə göstərmək olar:

1. Dilçilik tarixindən də məlum olduğu kimi, dilçilik elmi ən qədim dövrlərdə nail olduğu əhəmiyyətli uğurlardan sonra (e.ə. IV-III əsrlər Panini qrammatikası, Qədim Yunan dilçilik məktəbinin poetika və üslub məsələləri, “İsgəndəriyyə qrammatikası” məktəbi, ərəb dilçiliyi və s.), yenidən həqiqi mənada inkişaf etməyə XIX yüzillikdə başlamışdı. Lakin XIX yüzillik boyu dilçilərin əsas diqqəti dilin tarixi inkişaf məsələlərinə yönəlmişdi. Bütün humanitar elmlər üçün strukturalizm əsri kimi tarixə düşən XX yüzillikdə qədim dövrlərdəkindən sonra ilk dəfə olaraq dilçilik elmində dil strukturlarının, artıq yeni səviyyədə, yeni baxış bucağından öyrənilməsinə başlandı. XX yüzillikdə aparılan tədqiqatlarda dilçilər dərk etməyə başladılar ki, dilə, dilin strukturuna “Ferdinand de Sössürün “Ümumi dilçilik kursu” əsərində irəli sürülən linqvistik ideyalarla, xüsusən dilin sistem səciyyəsinə olması ideyası” (Axundov A. Ümumi dilçilik. Dilçiliyin tarixi, nəzəriyyəsi və metodları. Bakı: Maarif, 1979, s. 48) əsasında yanaşmaq lazımdır. Bu ideyaya görə, dil elə işarələr sistemidir ki, burada işarələrin fiziki təbiəti deyil, yalnız onların arasındakı münasibətlər mühümdür. Buradan dilin “xalis münasibətlər sistemi” olması qənaətinə gəlmək olar, hərçənd ki, bəzi struk-

tur dilçilik məktəbləri bunu tamamilən qəbul etmədilər (məsələn, Praqa dilçilik dərnəyində hesab edirdilər ki, dildə vahidlərin öz xüsusiyyətləri, xüsusilə fonem üçün – onun səs xarakteristikası mühümdür) [Вax: Структурализм в лингвистике // <http://files.school-collection.edu.ru/dlrstore/f8d1f64a-39cb-e5af-5d0a-0276936ef13b/1008588A.htm>]. Dilçilik tədqiqatlarında bu mülahizə dərk edildikdən sonra, “xalis münasibətlər sistemi” olan riyazi birləşmələr və dil arasında paralellər aparmaq ideyası aşkar surətdə özünə yer tapdı.

2. Yeni dövrün başlanğıcında riyaziyyatda kəmiyyət metodlarının istifadəsi ön planda idi. Yalnız XIX əsrdən başlayaraq yenidən mücərrəd qeyri-kəmiyyət modelləri qurulmağa başlandı ki, bunlar antik dövrlərdəkindən daha yüksək mücərrədlik səviyyəsi ilə fərqlənirdi. Həmin qeyri-kəmiyyət modelləri arasında sonradan dilçilikdə tətbiq olunmağa başlayan modellər də var idi. Bu modellərin tətbiq edilməsi isə XX əsrin birinci yarısına təsadüf edir.

XX əsrin ortalarında dilçiliklə riyaziyyat, kibernetika, hesablama texnologiyası, statistika və s. dəqiq elmlər kompleksinin görüşməsi, birgə fəaliyyət göstərməsi riyazi dilçilik, kibernetik dilçilik, struktur-tətbiqi dilçilik, statistik dilçilik (lingvostatistika), kompüter dilçiliyi, mühəndis dilçiliyi, korpus dilçiliyi və s. müxtəlif terminlərlə adlandırılan elm sahəsinin meydana çıxmasına səbəb oldu. Bu elm sahəsinin predmetini dilçilik tədqiqatları üçün riyazi aparatın işlənilib hazırlanması təşkil edir. Burada əsas yeri formal qrammatikalar nəzəriyyəsi, xüsusilə alqoritmlər nəzəriyyəsi tutur. Hansı formada adlandırılmasından asılı olmayaraq, dilçiliyə çoxluq, funksiya, izomorfizm kimi fundamental riyazi ideya və anlayışlar daxil olur. Yalnız riyazi anlayış və metodlardan istifadə edilməsi mümkün olan dilçilik sahələrində deyil, ümumiyyətlə, dilçiliyə “riyazi ruh”un nüfuz etməsi baş verir. Bütün bunlar onu deməyə əsas verir ki, dilçilik özlüyündə humanitar elm hüdudlarında qalmaqda davam etməklə, eyni zamanda daha çox dəqiq və daha obyektiv elmə çevrilməyə meyillidir.

Azərbaycan dilçiliyi istər yaxın, istərsə də uzaq xarici ölkələrdə özünün zəngin tarixi və ənənələri ilə tanınır. Azərbaycan dilçiləri Azərbaycan dilinin qaynaqları və inkişaf tarixi, Azərbaycan dilinin qohum və qohum olmayan dillərlə müqaisəli-tipoloji tədqiqi, çeşidli lüğətlərin hazırlanması, elmi-texniki terminologiyanın linqvistik məsələləri, nəzəri, tətbiqi dilçilik və onomastika məsələlərinə dair kifayət qədər sanballı tədqiqatlara sahibdirlər. Dilin mühəndis dilçiliyi metodları ilə təhlili, avtomatik lüğətlər, maşın tərcüməsi, statistik leksikoqrafiya Azərbaycan dilçiliyinin xüsusi bir istiqamətini təşkil edir. Dilçiliyimizi bu istiqamətdə inkişaf etdirən alim və tədqiqatçılar arasında filologiya elmləri doktoru, professor Məsud Mahmudovun özünəməxsus yeri vardır. Professor Məsud Mahmudov dilin avtomatik təhlili, riyazi dilçilik, avtomatik lüğətlər, maşın tərcüməsi, statistik leksikoqrafiya sahəsində ölkədə ilk tədqiqatçılardan biridir.

Məsud Mahmudov o tədqiqatçılardandır ki, Leninqrad (indiki Sankt-Peterburq) Dövlət Pedaqoji İnstitutunda prof. R.Q.Piotrovskinin rəhbərlik etdiyi «Инженерная лингвистика» laboratoriyasında elmi təcrübə keçdiyi, eyni zamanda Leninqrad Dövlət Universitetində «Математическая лингвистика» kafedrasının aspirantı olduğu zamanlarda (1979-1982-ci illərdə) dilçiliyə “riyazi ruh”un gəlməsini canlı şəkildə müşahidə etmişdir. Bu elə bir vaxt idi ki, hesablama texnologiyaları inkişaf dövrünə qədəm qoymuşdu, dilə tətbiq olunan proqram və linqvistik təminat getdikcə təkmilləşirdi, texnoloji proqramların sürəti çoxalır, yaddaşı artır və beləliklə, onların linqvistik tədqiqatlara tətbiqi üçün geniş imkanlar açılırdı. Bu həm də elə bir dövr idi ki, proqram və linqvistik təminat texnologiyaları keçmiş SSRİ-də rus dilinə, xaricdə isə əsasən ingilis dilinə tətbiq edilirdi. Məhz belə bir dövrdə Məsud Mahmudov Azərbaycan dilinin materialına əsaslanaraq mətnlərin formal təhlili problemini sistemli şəkildə araşdırmış, müxtəlif dil səviyyələrində formal əlamətlərə görə avtomatik işlənmə alqoritmləri qurmuş və onları hesablama maşınlarında reallaşdırmışdır. Məsud Mahmudovun türk dilləri

üçün söz-formaların formal morfoloji təhlili sisteminin hazırlanmasında göstərdiyi xidmət onun elmi rəhbəri, Rusiya Federasiyasının Əməkdar elm xadimi, Humanitar Elmlər Akademiyasının akademiki, bir sıra beynəlxalq akademiya və elmi qurumların həqiqi üzvü, professor R.Q.Piotrovski tərəfindən yüksək qiymətləndirilmişdir. Müdafiə Şurasının iclasındakı çıxışında o, aspirantının elmi cəsarətini onun bütün elmi tədqiqat siklini tamamlaması ilə əlaqələndirir. Sikl isə elmi fərziyyənin formulirovkası (1), həmin fərziyyə üzrə modelin qurulması (2), fərziyyənin doğruluğunun yoxlanması (3) və praktik tətbiqi (4) mərhələlərini ehtiva edir. Mərhələləri bütünlüklə və uğurla tamamlamaqla M.Mahmudov qurduğu bərpəedici mühəndis dilçiliyi modelləri ilə türk dillərinin bir sıra qanunauyğunluqlarını yeni baxımdan tədqiq edə bilmişdir. SSRİ EA-nın Dilçilik İnstitutunun Tətbiqi dilçilik şöbəsinin müdiri, texnika elmləri dokoru, professor R.Q.Kotov və Ümumittifaq Tərcümə mərkəzinin direktoru, filologiya elmləri doktoru, professor Y.N.Marçuk da M.Mahmudovun tədqiqatlarını alqışlamış və bu tədqiqatın xarakterik cəhətini onun obyektinin təkcə maşın tərcüməsinin deyil, eyni zamanda informasiyanın işlənməsinin bütün prosesləri üçün mərkəzi və başlanğıc mərhələ olmasında görmüşlər. Türk dillərində söz, morfem strukturu və kombinatorikası etibarilə kifayət qədər mürəkkəb olduğu üçün, bu dillərin materialı əsasında fərziyyə, riyazi model və alqoritmlərin qurulması və tətbiq edilməsi yüksək qiymətləndirilmişdir. Alqoritmın dəqiqliyi və nisbi sadəliyi diqqəti cəlb edir. Məsələnin texniki həllinin uğurluluğu tədqiqatların mühüm məziyyəti kimi qeyd edilmişdir. Qurulmuş alqoritm ümumi səciyyə daşdığından onu türk dillərinin bütün qruplarına tətbiq etməyə imkan verməsi xüsusilə vurğulanmışdır.

Dilçilik və texnika sahəsində görkəmli mütəxəssislərin Məsud Mahmudovun tədqiqatları haqqında fikirləri də diqqətəlayiqdir. Bu sətirlərin müəlliflərindən biri (akademik Ağamusa Axundov) olaraq keçmiş tələbəm haqqında belə düşünürəm: "... M.Ə.Mahmudov 10.02.21 – struktur, tətbiqi və riyazi dilçilik indeksi üzrə dissertasiya yazıb müdafiəyə təqdim etmiş ilk Azər-

baycan filoloqudur. Onun sələfləri texnika və riyaziyyat sahəsində mütəxəssislər olmuşlar.” Texnika elmləri doktoru, professor G.P. Melnikov: “... M.Mahmudovun struktur və tətbiqi dilçiliyin nəzəriyyə və praktikasına verdiyi, türkologiyanın tələblərinə yönəlik töhfəsi də tamamilə aşkar görünür. Bununla əlaqəli xatırladaq ki, bu səpkidən olan məlum türkoloji tədqiqatlarda indiyə qədər türk dillərindəki söz-formaların sintezi prinsiplərinin, prosesin maşında realizasiyası ilə təkmilləşdirilməsi aspekti üstünlük təşkil edir, bu işdə isə ilk dəfə olaraq türk mətnlərində söz-formaların leksik-qrammatik səciyyəsinin söz-formalar arasında onların strukturunun sintaktik və semantik əlaqələrinin üzə çıxarılmasının növbəti mərhələlərinə keçidi nəzərdə tutan linqvistik təhlil modeli qurulmuşdur. Bu modelin mühüm cəhəti ondan ibarətdir ki, onunla affiksəl morfemlərin omonimliyi aşkarlanır, minimuma endirilərək aradan qaldırılır.”

Göründüyü kimi, M.Mahmudovun tədqiqatları haqqında fikir söyləyənlər belə bir məqamı vurğulayırlar ki, o, təkcə fərziyyə irəli sürməklə kifayətlənməmiş, həmin fərziyyənin doğruluğunu sübut etmək üçün linqvistik model qurmuş, alqoritmlər tərtib etmiş və nəzəri müddəaları praktik olaraq reallaşdırmağa nail olmuşdur. Adətən belə tədqiqatlarda yalnız fərziyyə irəli sürülür və onun formal təsəvvürü verilir. Amma Məsud Mahmudov “nəzəriyyə-təcrübə-sınaq-tətbiq” mərhələlərini özündə cəmləşdirən bütöv bir elmi konsepsiyaya sadıq qalaraq tədqiqatı tam olaraq icra edir.

Məsud Mahmudov türk dillərində söz-formaların formal morfoloji təhlili sistemini yaratdıqdan sonra tədqiqatlarını daha da genişləndirdi. Formal təhlil sistemlərinin qurulmasının nəzəri və praktik problemləri, maşın tərcüməsi, avtomatik maşın lüğətləri, müxtəlif tipli statistik (tezlik, əks əlifba və s.) lüğətlərin tərtibi sahəsində yorulmadan çalışan alim respublikada və onun hüdudlarından kənarında tanınmış mütəxəssisdur. Onun tərtibçisi olduğu “Azərbaycan dilinin tezlik lüğəti” üçcildliyinin I cildi (söz kökləri) çapdan çıxmış və qalan iki cildin çapı gözlənilir.

“Azərbaycan dilinin tezlik lüğəti” Azərbaycan dilçilik elminə sanballı töhfədir.

1994-cü ildə “Azərbaycan mətnlərinin avtomatik işlənməsi sistemi” mövzusunda müdafiə etdiyi doktorluq dissertasiyası əsasında hazırlanmış və 2002-ci ildə çapdan çıxmış «Mətnin formal təhlili sistemi» adlı monoqrafiyası informatika, dil nəzəriyyəsi və süni intellekt problemləri ilə məşğul olan mütəxəssislər tərəfindən yüksək qiymətləndirilmiş və məşin tərcüməsi sistemlərində uğurla tətbiq olunmaqdadır. Məsud Mahmudov internetdə geniş istifadə olunan “Poliqlot” elektron lüğətlər və «Dilmanc» məşin tərcüməsi sistemlərinin yaradılmasında yaxından iştirak etmişdir.

Oxuculara təqdim olunan “Kompüter dilçiliyi” adlı monoqrafiya alimin məşin tərcüməsi və formal linqvistik təhlil sistemləri haqqında araşdırmalarının davamıdır. İlk öncə, monoqrafiyanın adı diqqəti cəlb edir: kompüter dilçiliyi. Alim bu məsələyə aydınlıq gətirir. O, aydınlaşdırır ki, XX əsrin 50-ci illərində riyaziyyat, kibernetika, informatika, statistika, hesablama texnologiyası və s. dəqiq elmlərlə dilçiliyin qovuşma nöqtəsində yeni bir elm sahəsi formalaşdı. İlk əvvəl bu yeni yaranan elm sahəsi “riyazi dilçilik”, “struktur dilçilik”, “hesablama dilçiliyi”, “statistik dilçilik”, “kibernetik dilçilik”, “struktur-tətbiqi dilçilik”, “mühəndis dilçiliyi”, “korpus dilçiliyi” və s. kimi adlandırılırdı. Əslində müxtəlif elmlərin qovuşuğunda yaranan bu yeni elm sahəsinin tədqiqat obyektı və predmeti məlum idi. Çox geniş və əhatəli araşdırma sahəsinə malik olduğu üçün ayrı-ayrı tədqiqatçılar ixtisaslarına müvafiq olaraq dilçiliklə həmhüdüdü elm sahəsi yaratmağa cəhd göstərir və bu cür termin müxtəlifliyinə yol verirdilər. “Kompüter dilçiliyi” termini həmin adların hamısını ehtiva edir və dilçiliyin riyaziyyat, statistika, kibernetika elmləri kompleksi ilə qovuşma nöqtəsində yaranan yeni sahəni, alimin fikrincə, “kompüter dilçiliyi” adlandırmaq daha düzgün olardı.

Monoqrafiya 9 fəsildən ibarətdir. I-III fəsillərdə kompüter dilçiliyinin nəzəri və təcrübi məsələləri, “kompüter dilçiliyi” an-

layışı, kompüter dilçiliyinin yaranma səbəbləri, mahiyyəti, istiqamətləri və sahələri, Azərbaycan dilinin milli korpusunun nəzəri və təcrübi bazası, Azərbaycan dilinin milli korpusunun strukturu kimi məsələlər araşdırılır. IV-IX fəsillərdə Azərbaycan dilinin milli korpusu və onun tərkib hissələri olan statistik lüğətçilik, maşın tərcüməsi və formal linqvistik təhlil sisteminin leksik, morfoloji, morfonoloji, sintaktik və semantik blokları hərtərəfli izah olunur.

Diqqəti cəlb edən digər bir cəhət ondan ibarətdir ki, müəllif “Mətnin formal təhlili sistemi” monoqrafiyasında əsaslandığı prinsipləri burada da əsas götürmüşdür. Bunun səbəbini qurulmuş formal təhlil alqoritmlərinin universallığında axtarmaq lazımdır. Azərbaycan dili üçün qurulmuş formal təhlil alqoritmlərinin digər türk dilləri üçün mətnlərin formal işlənməsi sistemlərinə tətbiq edilə biləcəyi öncədən xəbər verilir. Türkcə-rusca maşın tərcüməsi sistemində Azərbaycan dilinin avtomatik lüğəti və morfoloji təhlil alqoritminin prototip kimi istifadə olunduğu artıq məlumdur və monoqrafiyada bu vurğulanır. Müəllifin əsas götürdüyü mühüm prinsipləri aşağıdakı kimi xülasə etmək olar:

1. Dili bütün səviyyələrdə tamam formallaşdırmaq olar və həmin formallaşdırma əsasında mətnlərin müxtəlif məqsədlərlə (təhlil, sintez, başqa dilə tərcümə, referatlaşdırma və s.) emalı edilməsi mümkündür.

2. Dilin formal ifadəsinə əsaslanaraq dil hadisələrinin və qanunauyğunluqlarının mahiyyətini aşkarlamağa imkan verən modellər yaradıla bilər. Bu linqvistik modellər real dil hadisələrini və qanunauyğunluqlarını təqlid edib doğurmaq keyfiyyətinə malikdir.

3. Dilin bütün səviyyələrində ciddi bir tamlıq, sistem və dəqiqlik hökm sürür. Bu prinsipə əsasən tədqiqat obyektini dil olan dilçiliyi dəqiq elmlər sırasına daxil etmək olar.

4. Dilin müəyyən bir səviyyəsi daxilində qapanıb qalmaqla onun tam təhlilini vermək olmaz, çünki sistemi onun öz elementləri əsasında tam dərk etmək mümkün deyildir. İyerarxik cəhət-

dən öndə olanın tam dərki üçün sonra gələnin imkanlarından istifadə etmək lazım gəlir.

Təməl kimi götürülən bu prinsiplər alimə dil vahidlərinin təhlili və sintezi prosesinə daha dərindən nüfuz etmək, dilin formal təhlil sisteminin riyazi-məntiqi-fəlsəfi köklərini araşdırmaq imkanı vermişdir.

Prof. M.Ə.Mahmudovun çoxillik gərgin elmi axtarırlarının və zəhmətinin məhsulu olan “Kompüter dilçiliyi” əsəri Azərbaycan dilçilik elminə layiqli töhfədir. İnanırıq ki, bu əsər nəzəri və tətbiqi dilçiliyin aktual problemləri ilə məşğul olan alimlərin, müvafiq sahələr və ixtisaslar üzrə təhsil alan tələbələrin və geniş oxucu kütləsinin rəğbətini qazanacaq.

Ona da inanırıq ki, “dil və nitqin dialektik vəhdətinin ünsürləri kimi mətn vahidlərinin sistem-struktur semioloji mərhələ-səviyyəli və çoxsaxəli-səviyyələrarası modelləşdirilməsi sahəsində işləyib hazırladığı bütöv linqvistik konsepsiya”nın (prof. Aydın Ələkbərov) müəllifi olaraq prof. Məsud Mahmudov və onun davamçıları Azərbaycan Respublikası Prezidentinin 2013-cü il 9 aprel tarixli Sərəncamı ilə təsdiq edilmiş “Azərbaycan dilinin qloballaşma şəraitində zamanın tələblərinə uyğun istifadəsinə və ölkədə dilçiliyin inkişafına dair Dövlət Proqramı”nda fəvqəladə əhəmiyyətli aktual sosial sifariş kimi nəzərdə tutulmuş, Azərbaycan dilinin daha geniş istifadəsinin təmin edilməsi üçün internet texnologiyalarının, maşın tərcüməsi sistemlərinin və digər müasir tətbiqi-linqvistik texnologiyaların yaradılması və inkişaf etdirilməsi işinə də öz töhfələrini verəcəklər.

Ağamusa Axundov

*Azərbaycan Milli Elmlər Akademiyasının həqiqi üzvü,
filologiya elmləri doktoru, professor*

İlham Tahirov

filologiya elmləri doktoru

17.06.2013.

MÜƏLLİFDƏN

Dilçilik elminin inkişafının müasir mərhələsi daha çox riyazi metodların, kibernetika elmləri kompleksinin, hesablama texnikasının dilçiliyin müxtəlif sahələrinə nüfuzu ilə səciyyələnir. Dil hadisələrinin mahiyyətinin araşdırılması ilə bağlı tədqiqatlara ehtimal nəzəriyyəsinin və riyazi statistikanın tətbiqi, dildəki qanunauyğunluqların yeni baxış bucağı altında nəzərdən keçirilməsi və izahı dil nəzəriyyəsinin yeni-yeni metod və nəzəri müddəalarla zənginləşməsi ilə nəticələnir (Пиотровский, 1975; 1979).

Hesablama texnikası və riyazi statistikanın leksikoqrafiyaya tətbiqi lüğətçilikdə yeni bir sahənin – statistik leksikoqrafiyanın yaranmasına səbəb olmuşdur (Алексеев, 1975).

Müasir hesablama texnikasının coşqun inkişafı, proqram və linqvistik təminatın getdikcə təkmilləşməsi, sürətinin çoxalması, yaddaşının artması onun linqvistik tədqiqatlara tətbiqi üçün geniş və hərtərəfli imkanlar açır. Hazırda təkcə böyük EHM-lərdə deyil, fərdi kompüterlərdə də iri həcmli dil materiallarının müxtəlif məqsədlərlə işlənməsi və müxtəlif nəticələrin əldə olunması sahəsində xeyli təcrübə toplanmışdır ki, bu da insanın əqli fəaliyyətinin müxtəlif sahələrinin avtomatlaşdırılması üçün geniş imkanlar yaratmışdır. Hesablama texnikası təbii dili təmsil edən mətn informasiyasını «hiss etmək», yaddaşa köçürmək, müəyyən sorğuya görə təhlil və sintez etmək, təsnifləşdirmək, seqmentlərə ayırmaq, başqa dilə çevirmək «qabiliyyətinə» malikdir. Məhz indiyə qədər qeyri-real hesab olunan bu xüsusiyyətlər yeni bir ünsiyyət kanalının – «insan-maşın-insan» sisteminin yaranmasının başlanğıcını qoydu. Nəticədə dilçiliyin inkişafında keyfiyyətcə yeni bir mərhələ – riyaziyyatın, kibernetika elmləri kompleksinin, hesablama texnikasının nüfuzu ilə səciyyələnən kompüter dilçiliyi mərhələsi başlandı. Hazırda elmlərin inteqrasiyası və diferensiasiyası kimi dialektik bir prosesin qanunauyğun nəticəsi kimi dilçilikdə maşın tərcüməsi, statistik leksikoqrafiya, məlumat axtarışı sistemləri, «insan-maşın-insan» dialoq sorğu sist-

emləri, avtomatlaşdırılmış məlumat xidmətləri, proqramlaşdırılmış dil tədrisi vasitələri, mətnlərin avtomatik işlənməsi sistemləri kimi yeni sahələr yaranmışdır. Ayrı-ayrı dillərin məşin fondlarının (milli korpuslarının) tərkib hissəsi olan mətnlərin formal təhlili sistemləri bir çox mürəkkəb problemlər kompleksini (məşin tərcüməsi, avtomatik lüğətlər, formal linqvistik təhlil, sintez və s.) əhatə edir.

Qeyd etmək lazımdır ki, mətnlərin formal təhlili sistemlərinin müxtəlif məsələləri ilə bağlı dünya dilçiliyində, türko-logiyada, eləcə də Azərbaycan dilçiliyində bir çox tədqiqat əsərləri yazılmış, nəzəri və təcrübi səviyyədə araşdırmalar aparılmışdır (bax: Автоматический перевод, 1971; Бабанаров, 1981; Бектаев, 1978; Велиева, 1971; Vəliyeva, 1996; Джубанов, 1987; Исхакова, 1972; Котов, 1976; Мамедова, 1984; Марчук, 1988; Махмудов, 1982, 1991; Mahmudov, 1994; Пинес, 1970; Пиотровский, 1975, 1979; Рахманов, 1988; Садыков, 1987 və b.).

Monoqrafiyada ilk dəfə olaraq mətnlərin formal təhlili problemi sistem şəklində araşdırılmış, müxtəlif dil səviyyələrində formal əlamətlərə görə avtomatik işlənmə alqoritmləri tərtib olunmuş və onların hesablama məşinlərində reallaşdırılması yolları göstərilmişdir. Tədqiqat prosesində qurulmuş bərpəedici mühəndis dilçiliyi modelləri dilin bir çox qanunauyğunluqlarının və hadisələrinin yeni baxımdan tədqiqi və yozumu imkanı yaradır.

Tədqiqat obyektini kimi Azərbaycan yazılı mətnləri, dilin lüğət tərkibi, Azərbaycan dilinin qrammatik sistemi, sözlərin qrammatik münasibətləri və ifadə vasitələri, əlamətin formal təsviri üsulları götürülmüşdür. Monoqrafiyada ilk dəfə olaraq mətnlərin formal təhlil sistemləri bütün mərhələləri nəzərə alınmaqla tam şəklində götürülmüş, sistemin təkcə nəzəri deyil, həm də praktik prinsipləri araşdırılmışdır. Tərtib olunmuş təhlil alqoritmləri təcrübi olaraq sınaqdan keçirilmiş, mətnlər üzərində yoxlanmışdır.

Tədqiqatın nəticələrindən Azərbaycan dilinin məşin fondunun (milli korpusun) təşkilində, mətnlərin müxtəlif məqsədlərlə avtomatik işlənməsi sistemlərinin yaradılması və təkmilləşdi-

rilməsində, avtomatik «insan-maşın-insan» dialoq sistemlərində, öyrədici linqvistik avtomatların hazırlanmasında, süni intellekt problemləri ilə əlaqədar, Azərbaycan dilinin normativ qrammatikasının yazılışı prosesində, avtomatlaşdırılmış məlumat axtarışı sistemlərinin işində istifadə oluna bilər.

Formal təhlil alqoritmləri unversaldır və digər türk dillərində də mətnlərin formal işlənməsi sistemlərinə də tətbiq oluna bilər. Türkçə-rusca maşın tərcüməsi sistemində Azərbaycan dilinin avtomatik lüğəti və morfoloji təhlil alqoritmı «prototip» kimi istifadə olunmuşdur (bax: Бабаһаров, 1981, 5).

Tədqiqatda dörd mühüm prinsip əsas götürülmüş, formal təhlil sistemi həmin prinsiplər əsasında yaradılmışdır.

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. Bu linqvistik fərziyyənin doğruluğu tərtib olunmuş alqoritmlərin sınaqdan keçirilməsi prosesində sübut olunmuşdur.

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. Lakin tədqiqatda istinad edilən modellər indiyə qədər mövcud ədəbiyyatlarda təsvir edilən dil modellərindən fərqlidir. Mühəndis dilçiliyi prinsiplərinə əsaslanan bu linqvistik modellər real dil hadisələrini və qanunauyğunluqlarını təqlid edib doğurmaq keyfiyyətinə malikdir. Bu isə, dilin mahiyyətinin, onun işlənmə mexanizminin dərkisi prosesində həlledici amildir (bax: Пиотровский, 1975, 1979).

Məlum olduğu kimi, linqvistik model bu və ya digər problemin daha dərinəndən araşdırılmasına, onun adı tədqiqatçı tərəfindən müşahidəsi çətin olan xüsusiyyətlərinin daha qabarıq şəkildə əks olunmasına və öyrənilməsinə xidmət edir. Bu qəbildən olan linqvistik modelin qurulması prosesində əvvəlcə hər hansı bir

lingvistik obyekt haqqında müəyyən fərziyyə irəli sürülür və bu fərziyyəyə müvafiq olaraq həmin obyektin analoqu yaradılır. Professor R.Q.Piotrovskinin irəli sürdüyü model tipi üz quruluşuna və funksiyalarına görə fərqlidir (Пiotровский, 1979). Bu modellərdə lingvistik obyektin analoquna əsasən onun özü yənidən bərpa olunur. Modelin elmi-nəzəri əhəmiyyəti lingvistik obyektlə onun analoqu əsasında bərpa olunmuş surəti arasında yaxınlıq dərəcəsi ilə ölçülür. Əgər model əsasında bərpa olunmuş obyektlə real lingvistik obyekt üst-üstə düşürsə, onlar arasında böyük yaxınlıq varsa, onda həmin modelin bu və ya digər dil hadisəsini aydınlaşdırmaq xüsusiyyətləri yüksək qiymətləndirilir. Bərpa olunmuş lingvistik obyekt reallıqdan uzaqdırsa, lingvistik modelin əsasını təşkil edən fərziyyə də həqiqətə uyğun hesab olunmur. Məhz bu obyektiv izaedici, aydınlaşdırıcı və həqiqəti real təsvir etmək kimi mühüm potensial imkanları nəzərə alınaraq tədqiqatda professor R.Q.Piotrovskinin əsaslandığı və sınaqdan keçirdiyi bərpaedici mühəndis lingvistik model tipindən istifadə olunmuşdur.

Üçüncü prinsip dilçiliyin dəqiq elmlər kateqoriyasına daxil edilməsi ilə əlaqədardır. Dilin bütün səviyyələrində ciddi bir qanunauyğunluq, tamlıq, sistem və dəqiqlik hükmü sürür. Dil araşdırmalarında müşahidə olunan qeyri-dəqiqliklər, müxtəlif yozumlar dil sisteminin deyil, dilə müxtəlif baxımdan yanaşan dilçilərin əməyinin «məhsuludur».

Nəhayət, dördüncü mühüm prinsip riyazi məntiq elminin inkişafında yeni bir mərhələ açmış Gödelin natamamlıq teoremi ilə əlaqədardır (Gödel, 1931; 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 (Mahmudov, 1994).

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 vaidi kimi formalaşması morfemlərdə başa çatırsa, morfemlərin tam izahı sözlərdə

açılırsa, analogi olaraq sözlərin tam və bitkin mənası (istər leksik, istər qrammatik) cümlədə məlum olur (bax: Солнцев, 1977). Təsadüfi deyildir ki, izahlı lüğətlərdə sözlərin mənasını tam açmaq, bütün məna çalarlarını vermək üçün onlar cümlələrdə-sintaktik əlaqələrdə verilir. Bu baxımdan cümlə də ayrılıqda tam təhlil olunub qavranılmır. Cümlənin tam «başla düşülməsi» üçün onu məndə götürmək lazım gəlir. Məndə əvvəl və sonra gələn cümlələrin informasiya tutumu və kommunikativ funksiyası cümləyə öz təsirini göstərmiş olur.

Mətnin “qavranılması” cümlənin qavranılmasından da mürəkkəbdir, daha doğrusu, formal meyarlar burada morfem səviyyəsində olduğu kimi böyük “səmərə” vermir.

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 öz 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, morfemləri “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.

Formal təhlil sistemində bu prinsip nəzərə alınmış, sistemi təşkil edən müstəqil bloklar (leksik, morfoloji, morfonoloji, semantik) arasında sıx qarşılıqlı əlaqə yaradılması təmin edilmişdir. Formal təhlil alqoritmlərinin tərtibi prosesində morfoloji səviyyədə (və ya başqa səviyyədə) təhlili qeyri-mümkün olan dil vahidlərinin araşdırılması üçün digər səviyyələrə və istisna hallar blokuna müraciət olunması mühüm şərtlərdən hesab olunur.

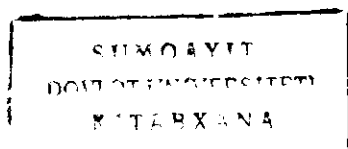
Azərbaycan dilinin materialı əsasında formal təhlil sisteminin yaradılması prosesində göstərilən dörd prinsip əsas götürülmüş və onlar sistemin iş prosesi boyu nəzərə alınmışdır.

Monoqrafiyada kompüter dilçiliyinin tədqiqat obyektı hesab olunan kompüter lüğətçiliyi, maşın tərcüməsi, formal təhlil sistemləri, milli korpus, süni intellekt, Azərbaycan dilinin internet məkanında işlənməsi və təmsil olunması ilə bağlı problemlər hərtərəfli şərh olunur, mövcud problemlərin həlli yolları təhlil edilir, konkret tövsiyələr verilir.

Müəllif monoqrafiyanın çapı ilə bağlı böyük zəhmət çəkmiş elmi redaktorlara – Azərbaycan Milli Elmlər Akademiyasının həqiqi üzvü, filologiya elmləri doktoru, professor Ağamusa Axundova və filologiya elmləri doktoru İlham Tahirova, rəyçilərə – Azərbaycan Milli Elmlər Akademiyasının həqiqi üzvü, texnika elmləri doktoru, professor Telman Əliyevə, texnika elmləri doktoru, professor Məsumə Məmmədovaya, texnika üzrə fəlsəfə doktoru, dosent Əbülfət Fətullayevə, filologiya elmləri doktoru Kamilə Vəliyevaya və filologiya elmləri doktoru, professor Sayalı Sadıqovaya, eləcə də Nəsimi adına Dilçilik İnstitutunun əməkdaşlarına öz dərin minnətdarlığını bildirir

Monoqrafiya ilə bağlı rəy, təklif, arzu və iradlarını göndərəcək oxuculara öncədən təşəkkür edirəm.

33565



I FƏSİL. KOMPÜTER DİLÇİLİYİ DİLÇİLİYİN YENİ BİR SAHƏSİ KİMİ

1.1. Ənənəvi dilçiliyə qeyri-ənənəvi baxış

Ənənəvi dilçiliyin tarixi çox qədimlərə gedib çıxır. Qədim Hind dilçiləri hələ e. ə. VII-II əsrlərdə indiki qrammatikalardan çox da fərqlənməyən qrammatika yaratmışdılar. İlk normativ qrammatika hesab olunan Panini qrammatikasında fonetika, kombinator səs dəyişmələri, vurğu və intonasiya haqqında da məlumat verilməsinə baxmayaraq, morfolojiya və sintaksis əsas yer tuturdu. Panini dili sistem kimi başa düşürdü. O, sözləri nitq hissələrinə görə təsnif etmiş, fel köklərinin dərəcələrini və adların hallanmasında şəkilçi tiplərini, paradigma əsasında kök və şəkilçi anlayışını, əsas, kök və şəkilçinin fərqlərini, daxili fleksiyanı ilk dəfə izah etmişdi. Dünya dilçiliyi tarixində ilk dəfə müqayisəli metoddan da Panini istifadə etmişdi (Rəcəbov, 1988, s.14-15).

Qədim hind dilçiləri məxrəc yeri, fəal danışq üzvü, partlayan, sürtünən (novlu), sait, yarım sait, samit, heca və s. fonetik anlayışlardan istifadə edir, samit əvəzlənmələrini tədqiq edirdilər. Onlar sözləri nitq hissələrinə – adlara və fellərə bölürdülər. Sanskritdə ismin yeddi halı göstərilirdi: 1.Adlıq, 2.Təsirlilik, 3.Birgəlik, 4.Yönlük, 5.Çıxışlıq, 6.Yiyəlik, 7.Yerlik (Axundov, 1979, s. 8-9).

Ənənəvi qrammatikanın formalaşmasında yunan dilçiliyinin də böyük rolu olmuşdur. İlk dəfə olaraq Aristotel sözlərin düzgün izahını vermiş, sözləri siniflərə ayırmışdı. Aristotel sözlərin üç sinfini ayırırdı: adlar, fellər və bağlayıcılar. (bağlayıcılara ədatlar, nidalar, önlüklər və əvəzliliklər aid edilirdi). O, “hallanma” qrammatik anlayışına hallanma və təsriflənməni, xüsusən zamana görə təsriflənməni daxil edirdi. Aristotel sözləri cinslərə görə fərqləndirirdi: kişi cinsi, qadın cinsi və onların arasında mövcud olan cins.

İsgəndəriyyə qrammatikləri 8 nitq hissəsini qeyd edirdilər: adlar, fel, feli sifət, artıkl, əvəzlik, önlük, zərf, bağlayıcı. Frakiyalı Dionisi isim, yaxud əşya bildirən ümumi, yaxud xüsusi göstərən, hallanan nitq hissəsini ad, zaman, şəxs və kəmiyyət qəbul edən, hərəkət yaxud vəziyyət ifadə edən və hallanmayan nitq hissələrini fel adlandırır. İsmi beş halı göstərirdi: əsas, yiyəlik, yönlük, təsirlik və çağırış (xıtab). Dionisi felin beş şəklini (qeyri-müəyyən, xəbər, əmr, arzu və təbəedici), üç mənə növünü (təsirli, məchul və orta növ), üç kəmiyyətini (tək, cüt və cəm kəmiyyətlərini), üç şəxsini (I-danışan, II-müraciət edilən, III-haqqında danışılan), üç zamanını (indiki, keçmiş, gələcək) şərh edirdi (Axundov, 1979, s.11-12).

Bu xülasəni eramızdan əvvəl və sonra fəaliyyət göstərmiş müxtəlif dilçilik cərəyanlarını və məktəblərini təmsil edən görkəmli zəka sahiblərinin fikirləri, əsərləri zəminində davam etdirmək, dilçiliyin müxtəlif məsələlərinə onların münasibətini geniş şərh etmək olardı. Fikrimizcə, buna ehtiyac yoxdur. Bu barədə çap olunmuş saysız-hesabsız araşdırma, xülasə, dərslik və dərs vəsaitlərində kifayət qədər məlumat verilmiş, bəzən eyni fikir ayırı-ayrı müəlliflər tərəfindən cüzi dəyişikliklərlə təkrarlanmışdır.

Azərbaycan dilinin müxtəlif səviyyələrini və sahələrini əhatə edən əsərlərlə (fonetika, morfologiya, sintaksis, leksika, frazeologiya, semasiologiya və s.) eramızın VII əsrdən başlamış indiyə qədər olan araşdırmalar arasında köklü fərqlər hiss olunmur. Xüsusilə, nitq hissələri, cümlə üzvləri, danışq səsləri, söz birləşmələri və s. məsələlər demək olar ki, dəyişilməz qalmış və ya cüzi dəyişikliklərə məruz qalmışdır. Halbuki digər elmlərdə, riyaziyyat, fizika, kimya, biologiya, mexanika, informatika sahələrində böyük sıçrayışlar və inkişaf göz qabağındadır. Hər halda nə vaxtsa fəlsəfinin bölmələri olmuş bu elm sahələrinin müasir inkişaf tempini və perspektivlərini danmaq qeyri-mümkündür.

Dilçiliyimizdə isə hələ də Aristarx dövründən məlum olan nitq hissələri (adlar, fel, feli sifət, artıkl, əvəzlik, önlük, zərf, bağlayıcı), cümlə üzvləri təkrar-təkrar tədqiq olunmaqdadır. Dilçilikdə bu ehkamçılıq və ifrat konservatizmin kökündə nə durur?

Müasir informatika texnologiyalarının, yeni kompüterlərin, mobil rabitə vasitələrinin coşqun inkişafı dövründə hələ də Mahmud Kaşğari üsulu ilə lüğət tərtib etməyin əhəmiyyəti varmı? Belə lüğətlər beynəlxalq kompüter şəbəkələrində geniş yayılmış və özünü doğrultmuş «Linqvo» tipli lüğətlərlə rəqabətə girə bilirmi?

Ənənəvi dilçiliyə qeyri-ənənəvi baxış dilçi şıltağı, professional modabazlıq deyil. Bu, hər şeydən əvvəl dilçilik elminin təbii, dialektik inkişafının obyektiv nəticəsidir.

Eramızdan əvvəl VII əsrdə yaşamış Panini ilə XXI əsrin dil tədqiqatçısının baxışları arasında fərq olmalıdır. Əks təqdirdə, bizim tədqiqatlar heç bir elmi və əməli səmərəsi olmayan boş məşğələyə çevrilər.

Dilin tədqiqi və tədrisi haradan başlanmalıdır – fonemlərdən, yoxsa mətnlərdən?

Kimya elmində əvvəlcə maddə bütövlükdə səciyyələndirilir, sonra onun atom və molekul quruluşu öyrənilir. Əvvəlcə atom və molekulaları tədqiq edərək, sonra onların əmələ gətirə biləcəyi maddələri səciyyələndirmək qeyri-mümkündür. Dilin tədqiqi və tədrisi isə fonetikadan və fonemlərdən başlanır. Bəlkə əksinə, mətndən fonemə, “yuxarıdan aşağıya” dil vahidləri ayrılmalı və tədqiq olunmalıdır. Bu məsələ K.Abdullayevi də maraqlandırmış və o, münasibətini belə ifadə etmişdir: “Hansı quruluş hansından yaranıb, sadə quruluş mürəkkəb quruluşa keçib, yoxsa mürəkkəb quruluş sadələşməyə başlayıb?” Bu suallara cavab vermək çox çətindir. Hər iki təkamül prosesinin xeyrinə məntiqi dəlillər tapmaq olar. Amma sistemin bütövlüyü naminə, qeyd etmək lazımdır ki, bütün atomar, sintaktik vahidlər, «bütün yuxarı qalxan» sintaksislər, «aşağıdan yuxarı gedən vahidlər» tədqiqatı, yəqin ki, bir vaxt gələcək öz yerini «yuxarıdan aşağı gələn vahidlərə, onları öyrənən üsullara verəcək» (K.Abdullayev, 1998, s.7).

Azərbaycan dilçiliyində dilə qeyri-ənənəvi baxış artıq formalaşmaqdadır. Mətn dilçiliyinin inkişafı, Azərbaycanda bu sahədə tədqiqat əsərlərinin yaranması ənənəvi dilçiliyə baxış bucağının dəyişməsi ilə nəticələndi. (K.Abdullayev, 1998, Ə.Abdullayev, 1998; 2002; Novruzova, 2002 və b.). Başqa sözlə desək,

mətn dilçiliyini ənənəvi dilçiliyə qeyri-ənənəvi baxışın formalaşmasının birinci mərhələsi hesab etmək olar. Ənənəvi dilçilik, normativ qrammatika yaratmaq sahəsində böyük xidmətləri olan dilçilər özləri də dilçilikdə ehkəmçiliyə qarşı çıxır, yeni baxışın formalaşmasının vacibliyindən danışırlar.

Bu baxımdan Azərbaycan dilinin sintaksisi sahəsində dəyərli tədqiqat əsərlərinin və dərsləklərin müəllifi kimi tanınan prof. Q.Kazımovun fikirləri böyük maraq doğurur. O, aktual üzlənmə və mətn məsələsindən bəhs edərək yazır ki, əgər dilin kommunikativlik imkanları, dil və nitq arasında mövcud olan sintaqmatik və paradiqmatik münasibətlər işlənərsə, müasir Azərbaycan ədəbi dilinin sintaksisi tamamilə yenidən işlənmiş olar. Professional sintaksis mütəxəssisinin bu fikrində böyük həqiqət var. Q.Kazımovun gəldiyi son nəticə belədir: «Daim belə bir paradoks meydana çıxır: cümləni bu qədər mübtədə və xəbərə ayırmağın, heç bir praktik əhəmiyyəti olmayan, insanın maddi-mənəvi inkişafında heç bir əməli rolu olmayan baş və ikinci dərəcəli üzvlər bölgüsünün yaradılması və ikinci sinifdən başlayaraq şagirdlərə öyrədilməsi hansı məntiqi zəminlə bağlıdır? İnsan bunun faydasını harada və nə zaman görməlidir?»

Lakin nitqin kommunikativ mahiyyəti nəzərə alındıqda dilin bütün elementləri hərəkətə gəlir və öz əhəmiyyətini göstərir. Bizim fikrimizcə, Azərbaycan dilin sintaksisinə aid tədqiq və araşdırmaları, xüsusən sintaksis dərsləklərini bu istiqamətə yönəltməyin vaxtı çatmışdır» (Kazımov, 1998, s.5-21).

1991-ci ildə «Elm» nəşriyyatında çap olunmuş monoqrafiyamda kompüterdə mətnlərin işlənməsi problemi ilə bağlı ənənəvi qrammatikalardan fərqli xüsusiyyətlərə və mahiyyətə malik «məşin qrammatika»larının yaradılmasının vacibliyi və zəruriliyi göstərilmişdir (Махмудов, 1991, s.10-50).

Sonralar bu ideya «Mətnin formal təhlili sistemi» əsərində başqa istiqamətdə əsaslandırılmışdır. 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 kompüter sırf formal əlamətlər və prinsiplər əsasında sistem üçün zəruri məlumatı əldə etmək imkanına malik olsun» (Məhmudov, 2002, s.81).

Ənənəvi dilçilikdə ayrı-ayrı dil səviyyələrində «nata-mamlıq», «izaholunmazlıq» müşahidə olunmaqdadır. Bu xüsusiyyət bir çox tədqiqatçılar tərəfindən ara-sıra qeyd olunmuş, onun səbəblərinin araşdırılmasına cəhdlər göstərilmişdir.

Ayrılıqda götürülmüş bir sözün mənasını tam dəqiq ifadə etmək qeyri-mümkündür. Sözün dəqiq mənası cümlədə (kontekstdə) açılır. Məhz ona görə də izahlı lüğətlərdə bəzən bir sözün onlarla mənası göstərilir.

Məsələn, Azərbaycan dilinin izahlı lüğətində «almaq» sözünün 26, «vermək» sözünün 15 müxtəlif mənası göstərilmişdir (Azərbaycan dilinin izahlı lüğəti, I cild, 1966, s.95-96; s.355-356). Ayrılıqda götürülmüş cümlə də nisbi mənada bitmiş fikir ifadə edə bilər. Əslində, cümlənin də tam mənası mətndə açılır. Bununla əlaqədar A.Məmmədov yazır: «Ayrı-ayrılıqda mətn ilə digər sintaktik vahidlər arasında çox maraqlı bir qarşılıqlı münasibətlər sistemi formalaşır. Belə ki, cümlə öz-özlüyündə heç bir mənaya malik deyil və əgər digər elementlərdən izolə edilmişsə, başa düşülə bilməz. Başqa sözlə, cümlə yalnız mətn çərçivəsində konkret hadisə ilə bağlı öz mənasını əldə edir və faktiki virtual olaraq burada aktuallaşma qabiliyyətinə malik ola bilər» (Məmmədov, 2001, s.24).

Mətdə cümlənin sərhədi və mətn-cümlə münasibətlərində mövcud mübahisəli məqamlardan bəhs edən K.Abdullayev yazır: «Bizim fikrimizcə, sintaktik nəzəriyyədə mövcud olan məhz bu cür çatışmazlıqlar mətn dilçiliyi kimi sahənin yaradılmasına xüsusi təkan verdi. Belə çatışmazlıqların sırasında ilk növbədə cümlənin sintaktik sistemdə son iyerarxik vahid kimi götürülməsini qeyd etmək olar. Bu halın ortadan qaldırılması cümlənin son deyil, orta sintaktik vahid kimi götürülüb, öyrənilib müdafiə olunması, cümlənin izomorf və homomorf, formal-semantik

parametrlərlə başqa cümlələrlə əlaqəyə girə bilməsi, qeyd etmək lazımdır ki, müasir dilçiliyin belə vacib sahəsi olan mətn dilçiliyi sahəsinin sütun daşlarına çevrilir və bu sahənin güclü inkişafını şərtləndirir» (K.Abdullayev, 1998, s.178-179).

Sintaksis və morfolojiya ilə yanaşı dilçiliyin başqa sahələri də müasir elmi araşdırmalar səviyyəsinə uyğunlaşdırılmalı, dilçilik müasir dünya elmləri arasında öz layiqli yerini tutmalıdır.

Azərbaycan dilçiliyinin ən mühüm sahələrindən olan lüğətçilik də Mahmud Kaşğari səviyyəsində qala bilməz. İnformasiya və kompüter texnologiyalarının indiki coşğun inkişafı dövründə yeni göstəricilərə və keyfiyyətlərə malik, istifadə üçün olduqca əlverişli və əhatəli elektron lüğətlərin olduğu informasiya mühitində lüğətçiliyə münasibət dəyişməli, dövrün tələblərinə uyğun lüğətçilik sahəsi formalaşdırılmalıdır.

Nəticə olaraq göstərmək lazımdır ki, kompüter dilçiliyi və onun bir qolu olan kompüter lüğətçiliyi, milli dil korpuslarının, formal qrammatikaların və sistemlərin yaradılması, maşın tərcüməsi və s. bu qəbildən olan məsələlər müasir dövrün, yaşadığımız əsrin tələbidir və bu tələb ödənilməlidir.

Azərbaycan dilçiliyində dilə konservativ baxış, ənənəyə «sadiqlik» və yeni dilçilik çərəyanlarına biganəlik ona aparıb çıxartdı ki, maşın tərcüməsi, mətnlərin formal təhlili, çeşidli lüğətlərin tərtibi, siyasi nitqin sintezi və tanınması və s. məsələlərlə daha çox qeyri-dilçilər, dəqiq elmlərin nümayəndələri, riyaziyyatçılar, kibernetiklər, informatika mütəxəssisləri məşğul olmağa başladılar. Dil materialı üzərində araşdırmalar aparən və texnika sahəsində elmi dərəcə alan mütəxəssislərin sayı da indi nəzərəcarpacaq dərəcədədir (Ə.Fətullayev 2006; Əmirov, 2006; R.Fətullayev, 2010; Mamedova, 1984 və b.).

Ənənəvi dilçiliyə yeni, qeyri-ənənəvi baxışın formalaşması vaxtı gəlib çatmışdır. Hazırda praktik məqsədlərə cavab verən yeni dilçiliyinin, maşın qrammatikalarının, fonetikasının, yeni fundamental lüğətlərin yaranmasına böyük ehtiyac var. Maşın tərcüməsi, şifahi nitqin tanınması və sintezi, müxtəlif universal, terminoloji və tərcümə lüğətlərinin tərtibi ilə bağlı da dilçilik

elminin üzərinə böyük vəzifələr düşür. Azərbaycan dilçiliyində daha çox dilləri əhatə edən, sözün tələffüzü və transkripsiyasını (hətta sözün ifadə etdiyi əşyanın şəklini) verən müasir lüğətlərin tərtibi günün vacib məsələlərindəndir (Mahmudov, 2006, s.111).

Dilə və dilçiliyə baxış bucağının təkcə dilçilik deyil, həm də psixoloji baxımdan dəyişdirilməsi məqamı çatmışdır.

1.2. Kompüter dilçiliyi anlayışı

Kompüter dilçiliyi anlayışı çox genişdir. Əslində, kompüter dilçiliyinə bir neçə elm sahəsinin qovuşuq nöqtəsində yaranan yeni bir elmi istiqamət kimi baxmaq olar. Kompüterlərin dilçilikdə istifadə olunması ilə bağlı bütün məsələlər (istər təcrübi olsun, istərsə də nəzəri) kompüter dilçiliyinin tədqiqat obyektinə hesab olunur. Kompüter dilçiliyini bir çox tədqiqatçılar tətbiqi dilçiliyin bir qolu hesab edirlər. Dilin müxtəlif məqsədlərlə modelləşdirilməsi ilə bağlı proqramlar, kompüter texnologiyaları və vasitələri tətbiqi dilçiliyin bu sahəsində geniş istifadə olunur. İnformatika və proqramlaşdırma nəzəriyyəsinin dil tədqiqatlarına istiqamətlənməsi tətbiqi dilçiliyin predmeti kimi səciyyələndirilə bilər.

XX əsrin 50-ci illərində elektron hesablama maşınlarının yaranması elmi təfəkkürdə inqilab yaratmış oldu. Qısa zaman kəsiyində elə elm sahələri yarandı ki, əvvəllər onlar heç təsəvvür belə olunmurdu. Kibernetika ilə biologiyanın qovuşuğunda «bionika», psixologiya və dilçiliyin sərhədində «psixolinqvistika», «neyrolinqvistika» yarandı. Sonralar, ümumiyyətlə, belə hesab olunmağa başladı ki, elmi axtarışlar daha çox iki və ya daha artıq elm sahəsinin qovuşma nöqtələrində səmərəli nəticələr almaq imkanı verir. Tədqiqatçılar müxtəlif elm sahələrinin qovuşma nöqtələrində yaranan elmi istiqamətlərə diqqəti artırmağa başladılar.

R.Q.Piotrovski «Mühəndis dilçiliyi və dil nəzəriyyəsi» əsərində dilçilikdə əqiq elmlərə aid metodlardan istifadə olunmasını təsadüfi hesab etmirdi. Son zamanlarda dilçilikdə yaranan və «riyazi dilçilik», «struktur-tətbiqi dilçilik» «mühəndis dilçiliyi», «kibernetik dilçilik», «hesablama dilçiliyi» və s. adlandırılan istiqamətlər müxtəlif elmi ideyaların qarşılıqlı təsiri və sosial tələbatın məhsuludur.

R.Q.Piotrovski termin müxtəlifliyini aradan qaldırmaq üçün «mühəndis dilçiliyi» terminini təklif edir və monoqrafiyasında bu termini elmi cəhətdən əsaslandırır (Пiotровский, 1979).

Dilçilik öz inkişafının elə biz mərhələsinə çatmışdı ki, artıq əvvəlki metod və üsullar az səmərə verirdi. Dil faktlarının daha dərindən və hərtərəfli araşdırılması üçün yeni metod və üsullara böyük ehtiyac hiss olunurdu. Bu metod və üsulların köməyi ilə dilin modelləşdirilməsi, statistik-riyazi baxımdan tədqiqi imkanı yarandı. Dilin indiyə qədər gizli qalan və ilk nəzərdən əhəmiyyətsiz görünən bir çox cəhətləri üzə çıxır və maraqlı elmi nəticələr alınır. Dilçi-alimlər riyazi-kibernetik, ehtimali-statistik metod və üsullardan istifadəyə maraqlı göstərməklə dili yeni baxış bucağından öyrənmək istəyirdilər.

Digər tərəfdən riyaziyyatçılar özləri də dil materiallarına bir tədqiqat obyektini kimi baxmağa başladılar. Nitqin informasiya quruluşunun riyazi tədqiqi, Markov zəncirlərinin «linqvistik» işlənməsi məhz dəqiq elm nümayəndələrinin dil materialına marağının uğurlu nəticəsi idi (Пiotровский, 1979, s.5.).

Daha mühüm bir amil insanın nitq fəaliyyətini təqlid edə biləcək süni intellektin kibernetik modelinin yaradılması sahəsində tədqiqatların intensivləşməsi idi. Bu məsələdə kibernetiklərlə dilçilərin əməkdaşlığı olmadan heç bir uğur qazanmaq olmazdı.

Məlum olduğu kimi, son zamanlar elmi-texniki mətnlərin, müxtəlif rəsmi sənədlərin operativ işlənməsi, maşın tərcüməsi, annotasiyaların və referatların avtomatik alınması məsələləri aktuallaşmışdır. Rus, ingilis, alman, fransız, yapon, çin, ərəb və bir çox başqa dillər üçün tərcümə sistemlərinin, avtomatik lüğətlərin hazırlanması da dilçiliyin yeni bir sahəsinin yaranması zərurətini doğururdu.

Böyük əhəmiyyət kəsb edən məsələlərdən biri də ana dilinin və xarici dillərin tədrisi prosesində riyazi, mühəndis-statistik tədqiqatlardan istifadə olunmasıdır.

Bütün bunlar XX əsrin ortalarında dilin daha dərindən öyrənilməsi, ilk baxışdan, müşahidəsi mümkün olmayan linqvistik obyekt və əlaqələrin müəyyənləşdirilməsi və tədqiqi məsələsini ön plana çəkdi. Dilin müşahidəsi çətin olan obyekt və əlaqələri-

nin müəyyənləşdirilib öyrənilməsi yalnız modelləşdirmə yolu ilə reallaşdırıla bilər.

R.Q.Piotrovski öz tədqiqatında yeni mühəndis dilçiliyi modeli təklif edir və onun köməyi ilə dil situasiyalarının tədqiqı yollarını göstərirdi (Піотровскій, 1979, с.8-9).

Beləliklə, XX əsrin ortalarında riyaziyyat, kibernetika, hesablama texnologiyası, statistika və s. dəqiq elmlərlə dilçiliyin qovuşma nöqtəsində yeni bir elm sahəsi formalaşmağa başlayırdı. İlk əvvəllər bu yeni yaranan elm sahəsi «riyazi dilçilik», «struktur dilçilik», «hesablama dilçiliyi», «statistik dilçilik», «kibernetik dilçilik», «struktur-tətbiqi dilçilik», «mühəndis dilçiliyi», «korpor dilçiliyi» və s. adlandırılırdı. Əslində, bu yeni sahənin tədqiqat obyektı və predmeti məlum idi. Çox geniş araşdırma sahəsinə malik olduğu üçün ayrı-ayrı tədqiqatçılar ixtisaslarına müvafiq olaraq dilçiliklə həmhüdüdü elm sahəsi yaratmağa cəhd göstərir, nəticədə bu cür müxtəlifliklərə yol verilmiş olurdu.

Fikrimizcə, «kompüter dilçiliyi» termini bu adların hamısını ehtiva edir və dilçiliyin riyaziyyat, statistika, kibernetika elmləri kompleksi ilə qovuşma nöqtəsində yaranan yeni sahəsinı «kompüter dilçiliyi» adlandırmaq düzgün olardı. Dünya elmində və rus dilçiliyində də «kompüter dilçiliyi» dilçiliyin yeni sahəsi kimi tanınır və bu termin artıq dünya elmində vətəndaşlıq hüququ qazanmışdır (Марчук, 2000; 2002; <http://elanina.narod.ru/lanina/index.files/intell//lingvistik.htm>; <http://files.school-collection.edu.ru/dlrstore/>).

Kompüter dilçiliyi yeni bir elm sahəsi kimi XX əsrin 60-cı illərindən formalaşmağa başlamışdır. Rus dilində işlənən «компьютерная лингвистика» ingilis dilindəki «computational linguistics» termininin kalka olunmuş variantıdır. Əslində, «computational» termini rus dilinə «вычислительный» kimi tərcümə olunur. Ona görə də, bəzən «hesablama dilçiliyi» termini də işlənir. «Hesablama dilçiliyi» anlayışı daha çox «kvantativ dilçilik» termininə sinonim kimi işlənir. ABŞ-da «Kompüter dilçiliyi» rüblük jurnalı çap olunur. «Kompüter dilçiliyi» assosiasiyasının bir çox Avropa ölkələrində regional qurumları fəaliyyət göstərir. Hər iki ildən bir kompüter dilçiliyi məsələlərinə həsr olunmuş beynəlxalq konfrans

(COLING) keçirilir. Süni intellekt məsələləri ilə bağlı keçirilən konfranslarda da kompüter dilçiliyi geniş təmsil olunur.

Kompüter dilçiliyini tətbiqi dilçiliyin başqa sahələrindən fərqləndirən əsas cəhət dil materialının kompüterlərdə işlənməsidir. Müstəqil bir elm sahəsi kimi formalaşan kompüter dilçiliyində bir çox yeni anlayış və terminlərdən istifadə olunur. Həmin termin və anlayışların əksəriyyəti süni intellekt sistemlərində və koqnitiv dilçilikdə işlənir. Belə terminlərə «freym» (tipik mövzuda vahid situasiyaların təqdimi üçün konseptual quruluş anlayışı), «ssenari» (stereotip situasiya və davranışlar haqqında biliklərin təqdimi üçün konseptual quruluşlar), «plan» (müəyyən məqsədə çatmaq üçün mümkün hərəkətlərin təqdimini qeyd alan bilik) və s. misal göstərmək olar. Bu anlayış və terminlərin bir çoxu yenidir, çətin izah olunur və mənimsənilir. Bəzən eyni terminin izahı müxtəlif tədqiqatçıların yozumunda fərqli olur. Bu da əlavə dolaşlıqlar yaradır.

Kompüter dilçiliyi koqnitiv dilçiliklə sıx əlaqəlidir. Koqnitiv dilçilikdə dünya modeli qurulur və onun kompüter analoqu yaradılır. Dünya modeli süni intellekt sistemlərində xüsusi blok təşkil edir və dünya haqqında ümumi biliklər sistemini formalaşdırır. Belə kompüter modellərinin linqvistik aspekti kompüter dilçiliyininin də obyektinə hesab olunur.

1.3. Kompüter dilçiliyinin sahələri

Kompüter dilçiliyi olduqca geniş bir anlayışdır və bir-birindən fərqli sahələri ehtiva edir. Bura süni intellekt problemləri, ünsiyyətin və mənimsəmənin modelləşdirilməsi, süjet quruluşunun modelləşdirilməsi, mətnlərin hipermətn texnologiyası ilə təqdimi, maşın tərcüməsi, kompüter lüğətçiliyi, ayrı-ayrı dillər üçün milli maşın fondlarının və milli korpusların yaradılması və s. daxildir. Adları sadalanan sahələr özləri də çox geniş anlayışlardır. Süni intellekt problemləri, ünsiyyətin və mənimsəmənin modelləşdirilməsi, biliyin quruluşu və dərki kimi məsələlər koqnitiv dilçilikdə, psixolinqvistika və neyrolinqvistika da öyrənilir.

Maşın tərcüməsi problemləri, avtomatik tərcümə lüğətləri, təbii mətnin təhlili və sintezi alqoritmlərinin tərtibi, formal maşın qrammatikalarının yaradılması kimi məsələlər kompleks yeni bir elmi istiqamət statusu qazanmaqdadır.

Kompüter lüğətçiliyi sürətlə inkişaf etməkdə olan olduqca perspektivli sahələrdən biridir. Hazırda müxtəlif məqsədlərlə tərtib olunan və İnternet məkanında geniş yayılmış elektron lüğətlərin əhəmiyyətini izah etməyə ehtiyac yoxdur. Bu barədə «Kompüter dilçiliyi» fəslində ətraflı məlumat veriləcəkdir.

Milli dillərin korpuslarının yaradılması da günün vacib və zəruri məsələlərindəndir. Ayrı-ayrı dillər üçün milli korpusların hazırlanması böyük zəhmət tələb edən ağır işdir. Burada dilçilərlə, informasiya texnologiyaları mütəxəssisləri ilə yanaşı başqa elm sahələrində çalışanların da iştirakı faydalı olardı. Milli dil korpusu dövlət əhəmiyyətli strateji məsələdir. Mətnlərin hipermətn texnologiyaları ilə təqdimi sahəsində bir çox işlər görülmüş və uğurlu nəticələr əldə olunmuşdur. İnternet məkanında məlumatların yerləşdirilməsi, milli korpusların təşkili, müxtəlif sorğular əsasında lazımi məlumatın əldə edilməsi, elektron lüğətlər arasında əlaqə və eyni zamanda çoxlu sayda məlumat mənbəyinə müraciət hipermətn texnologiyalarının köməyi ilə əldə edilir.

Kompüter dilçiliyinin sahələri hər biri dərin elmi araşdırmalar tələb edən yeni istiqamətlərdir. Bu sahələrin hər biri haqqında ayrılıqda bir neçə monoqrafiya yazmaq olar. Son zamanlar kompüter dilçiliyinin müxtəlif problemlərinə həsr olunmuş onlarla elmi-tədqiqat əsəri, dissertasiya yazılmış və yazılmaqdadır.

Kompüter dilçiliyi sürətlə sürətlə inkişaf edən elm sahəsidir. Kompüter texnologiyasının coşqun inkişafı, getdikcə təkmilləşməsi, insan cəmiyyətinin ən mühüm, ayrılmaz atributuna çevrilməsi kompüter dilçiliyi qarşısında da geniş üfüqlər acır.

1.4. Mətn kompüter dilçiliyinin obyektini kimi

Kompüter dilçiliyinin yarandığı ilk vaxtlarda təbii dillərin mətnlərinin avtomatik işlənməsinə böyük maraq var idi. İstifadəçilər süni dillərdən daha çox təbii dillərdə olan mətnlərlə

işləmək, ondan lazımı məlumatı əldə etmək istəyirdilər. Lakin təbii dilin mətnlərini kompüterlərdə yerləşdirmək və ondan müxtəlif məqsədlərdə istifadə etmək üçün, ilk növbədə, mətnin mahiyyətini acmaq, onun mənə cəhətini və strukturunu dərk etmək lazım idi. İnsanlar ünsiyyət prosesində mətni necə qavrayırlar? Mətnin anlama modellərini qurmaq mümkündürmü? Mətnin insan tərəfindən qavranması prosesini təqlid etmək nə dərəcədə reallaşdırıla bilər? Bütün bu proseslər formal sistemlərdə effekt verə bilərmi? Ənənəvi dilçilikdə bu məsələlərlə, demək olar ki, ciddi məşğul olmamışdılar. Bəlkə də, kompüterlərin yaranması və sürətlə inkişaf etməsi, təbii dilin avtomatik işlənməsinə maraqlı bu sualları doğurmuşdu. Kompüter dilçiliyi sahəsində çalışan mütəxəssisləri bu məsələ çox maraqlandırır. Əlbəttə, dilçilər əvvəllər də mətn və onun təşkil edən cümlələri tədqiq edirdilər. Lakin kompüterlərin yaranması ilə bu tədqiqatlar yəni-
dən aktuallaşdı, ayrıca istiqamətə çevrildi. Tərcümə prosesinin avtomatlaşdırılması sahəsində aparılan işlər mətnin formal təhlilinə elmi marağı son həddə çatdırdı. Dünyanın bir çox elm mərkəzlərində mətnin avtomatik təhlili ilə bağlı tədqiqatlar aparılmağa başladı. İlk əvvəl belə bir məsələni aydınlaşdırmaq lazım idi: cümlə və mətn strukturunu formal qaydalar əsasında qurmaq olarmı? Mətnin strukturu, onu təşkil edən cümlələr formal prinsiplər əsasında izah oluna bilərmi?

Ənənəvi dilçilikdə cümlə strukturu az-cox öyrənilsə də, mətn haqqında bu qəbildən elan tədqiqatlar yox dərəcəsində idi. Tədqiqatlar nəticəsində məlum oldu ki, hər bir mətn (eləcə də, kiçik mətn funksiyasında olan cümlə) bir neçə formal strukturdan ibarətdir. Həmin formal strukturları üç səviyyəyə ayırmaq olar: Birinci səviyyə – səthi sintaktik-struktur səviyyəsidir. Bu səviyyədə mətni təşkil edən hər bir cümlə ayrılıqda götürülür və cümlə üzvlərinə görə təhlil edilir. Cümlənin baş üzvləri – mübtəda və xəbər tapılır, sonra ikinci dərəcəli üzvlər – xəbər, təyin, tamamlıq və zərfliyin növləri müəyyənləşdirilir. Lakin yalnız səthi sintaktik strukturda tam təhlil aparmaq kafi nəticələr vermir. Sonrakı, ikinci səviyyə – dərin sintaktik-struktur səviyyə hesab olunur. Bu ideya

ona əsaslanır ki, təbii dillər xarici sintaktik xüsusiyyətlərinə görə fərqlidir. Obyekt, hadisə, onların əlamətləri və onların iştirakı ilə gedən proseslər fərqli formalarda təzahür edir. Əslində, müxtəlif təsirlərə, təzahür formalarına baxmayaraq mahiyyətinə görə dünya və dünyada baş verən hadisə və proseslər eynidir. Ayrı-ayrı dillərdə səthi strukturda onların izahı fərqlidir. Məhz ona görə, dərin struktur anlayışının işlənilib hazırlanması zərurəti yaranmışdır. Bu ideyaya görə, hər bir dilin fərqli təzahür formalarına baxmayaraq, dərin struktur bizi əhatə edən dünya haqqında adekvat təsvir verir. Bu strukturla bağlı dərin hallardan istifadə ideyası Amerika tədqiqatçısı Ç.Filmorun adı ilə bağlıdır. Sonralar bu ideya N.Xomski və onun ardıcılıarı tərəfindən inkişaf etdirilmişdir (Хомский, 1972; 1962; 1966).

Konkret nümunələr əsasında dərin sinataktik strukturu nəzərdən keçirək: «Əsgər məktub yazdı» və «Əsgər tərəfindən yazılmış məktub». Mahiyyətinə görə eyni bir hadisənin təsviri olsa da, xarici təzahür forması bizi fərqli şəkildə düşünməyə vadar edir. Sintaktik səviyyədə hər iki cümləni təhlil etsək, dərin strukturun səthi strukturdan fərqi barədə bizdə təsəvvür yaranmış olar. Birinci cümlədə «yazdı» hadisəsinin subyekti «əsgər»dir. «Əsgər» sözü adlıq halda olan isimdir. İkinci cümlədə, mübtədə funksiyasını «məktub» sözü yerinə yetirir. «Əsgər tərəfindən yazılmış» (və ya «əsgərin yazdığı») isə təyin olur. Əslində, bu cümlədə də subyekt funksiyasını «əsgər» sözü yerinə yetirir. Hər iki cümlədə situasiyanı başa düşmək üçün mətnədə hadisə, onun subyekti və obyektini müəyyənləşdiririk. Anlama prosesində subyekt və obyektin mövqeyinin, statusunun təyini dərin struktur anlayışının mahiyyətini təşkil edir. Subyekt və obyekt halının müəyyənləşdirilməsi yeganə üsul deyil. Tədqiqatçıların bəziləri müxtəlif sayda halların əlavə edilməsini təklif edirlər (alət, zaman, sahə və s.). Dərin hallır əsasında qurulmuş sintaktik struktur cümlənin səthi sintaktik strukturundan onun dərin strukturuna keçməyə imkan verir. Bu səviyyədə dilin leksik vahidlərində (sözlərdə) müşahidə olunan əlavə məlumatlar da təhlilə cəlb edilir. Sözlün mənası dil daşıyıcısına məlum olur və onun yad-

daşında saxlanır. Yaddaşa müraciət etməklə cümlə elementlərinə müvafiq mənalar əlavə edilir və bu da mətnin semantik səviyyədə qavranması imkanı yaradır.

Formal strukturun üçüncü mərhələsi semantik-struktur hesab olunur. Üçüncü mərhələdə də sintaktik struktur səviyyəsinə müvafiq olaraq səthi və dərin struktur səviyyələri mövcuddur. Səthi semantik-struktur dərin sintaktik strukturla sıx əlaqəlidir. Dərin semantik struktur daha çox eynitipli situasiyalar sinfi ilə bağlı fəaliyyət göstərir. Məsələn, «Tələbə imtahandan narazı qaldı» cümləsində səthi semantik-struktur yalnız tələbənin imtahandan sonrakı durumunu, hal-vəziyyətini əks etdirir. Dərin semantik strukturda isə tələbənin narazı qalmasının səbəbləri barədə müxtəlif ehtimallar söyləyə bilərik: – tələbə özü öz cavabından razı qalmayıb, onun cavabı düzgün qiymətləndirilməyib və s. Cümlənin təhlili prosesində yaranan dərin struktur praqmatik struktur adlandırıla bilər.

Dərin struktur cümlənin insanı nəyə təhrək etdiyini, nə tələqin etdiyini anlamaq imkanı yaradır. Praqmatik struktur mətnin cümlələri arasında əlaqə yaradır, mətni vahid bütöv şəkildə formalaşdırır. Avtomatik təhlil prosesində zəruri strukturları müəyyənləşdirmək üçün bir neçə ardıcıl mərhələdən keçmək lazımdır. Həmin mərhələlər aşağıdakılardır:

- ilkin mətn (təhlil olunan mətn);
- ilkin təhlil (qabaqcadan aparılmış təhlil);
- morfoloji təhlil;
- səthi sintaktik təhlil;
- dərin sintaktik təhlil;
- səthi semantik təhlil;
- dərin semantik təhlil;
- praqmatik təhlil;
- mətn strukturlarının müəyyənləşdirilməsi. (<http://elanina.narod.ru/lanina//idex.files//intell/linqvistik.htm>)

Tərkibində iki və ya daha artıq cümlə olan mətnlərin təhlili prosesində yeni strukturlar yaranır. Həmin strukturların köməyi ilə bu və ya digər təsvir olunan situasiya və ya situasiyalar

ardıcılığı çərçivəsində mətni təşkil edən cümlələrin bağlılığı təmin olunur. Mətnin bütövlükdə qavranmasına imkan verən frazalararası əlaqələr yaranmış olur. Bir cümlənin əsasında duran strukturlara nisbətən mətn strukturları lazımi səviyyədə öyrənilməmişdir. Mətn strukturunun tədqiqi də çoxmərhələlidir. Birinci mərhələ hazırlıq mərhələsidir. Həmin mərhələdə mətn kompüterə daxil edilmək məqsədilə hazırlanır. Həmin mərhələdə olan səhvlərə düzəliş verilir, qısaltmalar tam şəkllə gətirilir, xüsusi işarələr izah olunur və ya mətndən çıxarılır və s.

Morfoloji təhlil mərhələsində ayrı-ayrı sözlər morfoloji baxımdan araşdırılır – sözün əsası, fleksiyalar, sözönülər, şəkilçilər (prefiks, infiks, suffiks) müəyyənləşdirilir. Sözün əsası əsaslar lüğətində yerləşdirilmiş mənənin tapılması baxımından əhəmiyyətlidir. Digər hissəsiklər isə cümlə daxilində sözlər arasında qrammatik əlaqələrin aydınlaşdırılması prosesində istifadə olunur. Bu əlaqələr tam şəkildə səthi sintaktik təhlil mərhələsində əldə edilir. Beləliklə, hər bir cümlə üçün səthi struktur formalarının mümkün variantları tapılmış olur. Həmin struktur bütöv mətnin formalaşması üçün mühüm amil rolunu oynayır. Dərin sintaktik strukturun təhlili mərhələsində cümlənin dərin sintaktik strukturu müəyyənləşir. Həmin struktur sonrakı mərhələdə səthi və dərin semantik strukturunun alınması üçün ilkin baza olur. Zərurət olduqca, sonrakı mərhələlərdə praqmatik və bütün təhlil mərhələlərini ehtiva edən mətn strukturları qurmaq mümkündür. Göstərilən mərhələlər təbii dillərdə mətnlərin təhlili məsələlərini tam ehtiva edir. Qarşıya qoyulan məqsədə müvafiq olaraq konkret mətnlərin təhlili prosesində bu və ya digər mərhələdən istifadə oluna bilər.

Təhlillə yanaşı, təbii dillərdə mətnlərin sintezi də çoxmərhələlidir. Sintez təhlilə əks istiqamətdə baş verən prosesdir. Əgər mövzu və mətnin məqsəd strategiyası məlumdursa, hesab etmək olar ki, mətnin praqmatik strukturu artıq müəyyənləşdirilmişdir. Həmin praqmatik struktur mətni təşkil edən hər bir cümlədə tətbiq etməklə həmin cümlələr üçün təhlil mərhələlərini əks istiqamətdə reallaşdırmaq tələb olunur. Bu sadalanan məsələlərlə bağlı

həll olunmamış problemlər çoxdur. Məsələn, mətnin yaranma məqsədlərindən praqmatik strukturu necə seçib ayırmaq olar və həmin praqmatik strukturun dərin semantik struktura keçidin texnologiyası necə olmalıdır? Bunun üçün, hər şeydən əvvəl, mətnin yaranma mexanizmi hərtərəfli araşdırılmalıdır. Bu istiqamətdə hərəkət aktantlarından istifadə olunması da məsələnin həlli yollarından biri hesab oluna bilər. Hər bir hərəkətlə bağlı müəyyən obyekt və səciyyətlər göstərmək mümkündür. Həmin obyekt və səciyyətlər dərin struktur anlayışları ilə üst-üstə düşür. Məsələn: Əgər biz «getmək» felini bu baxımdan araşdırsaq, onunla bağlı bir çox strukturlarla mümkün əlaqələri araşdırmalıyıq:

- getmək (hərəkəti icra edən subyekt kimdir);
- getmək (haradan getmək, hərəkətin başlanğıc nöqtəsi haradan başlanır);
- getmək (hara getmək, hərəkətin son nöqtəsi haradır, hərəkət harada tamamlanır);
- getmək (nə məqsədlə getmək, gedişin məqsədi nədir);
- getmək (nə səbəbə (səbəbdən) getmək, gedişin səbəbi nədir);
- getmək (kimlə (kimlərlə) getmək);
- getmək (nə ilə (nədə) getmək, nəqliyyat vasitəsi (araba, maşın, təyyarə və s.) getmək);
- getmək (necə getmək - tələsik, tələsmədən, əsəbi, sakit və s. getmək);
- getmək (nə vaxt, nə zaman (bu il, gələn il, bu ay (gün, həftə), gecə, gündüz, səhər, axşam) getmək);
- getmək (hansı emosional durumda (razı, narazı, kədərlə, sevinclə, gülə-gülə, ağlaya-ağlaya və s.) getmək);
- getmək (hayatdan, ailədən, işdən, dostlardan ayrılıb getmək, həmişəlik, əbədi, ömürlük (müvəqqəti) bir dəqiqəlik (saatlıq) getmək).

Bu sadalananlar «getmək» felinin aktantları hesab olunur. Aktantların sayı məhdud deyil. Situasiya ilə bağlı fəllərə yeni aktantlar əlavə edilə bilər. Qeyd etmək lazımdır ki, fel aktantlarının heç də hamısı aktiv olmur. Felin (hərəkətin) bütün müm-

kün aktantları göstərilə bilər. Təhlil və sintez prosesində isə onların bir qismi istifadə oluna bilər. Başqa sözlə, hərəkət üçün onun bütün potensial (mümkün) aktant siyahısı verilə bilər. Real təhlil və sintezdə isə onların az bir qismi aktivləşə bilər (istifadə oluna bilər). Bu baxımdan hərəkət aktantlarını aktiv və passiv kateqoriyalara ayırmaq olar. Bu, təhlil və sintez prosesində vaxta və kompüterin yaddaşına qənaət baxımından da önəmlidir.

Mətnin qavranması təkcə linqvistik problem hesab oluna bilməz. Burada psixologiya, fəlsəfə, semiotika, məntiq və s. elmlərin metod və üsulları da önəmli rol oynayır. Mətni anlamaq üçün onu əvvəlcə təhlil etmək, hissələrinə ayırmaq, hər bir hissənin qrammatik və semantik səciyyələrini müəyyənləşdirmək, sonra bu ayrı-ayrı hissələri yenidən birləşdirib (sintez) bütöv mətnin izahını vermək lazımdır. Deməli, mətnin başa düşülməsi, qavranması üçün təhlil və sintez mühüm mərhələlərdir. Monoqrafiyada bu məsələlər Azərbaycan dilinin materialları əsasında şərh olunmuşdur.

Kompüter dilçiliyində təbii mətnin təhlili və sintezi çərçivəsində onun canlandırılması da yeni istiqamətlərdən biri hesab olunur. Fərdi kompüterlərin inkişafı nəticəsində nəinki mətnlərin mübadiləsi, həm də istifadəçilərin displeydə görüntülərinin verilməsi mümkün olmuşdur. İnsan təfəkkürü çoxaspektlidir (psixoloqlar bunu «müxtəlif modallıq» kimi şərh edirlər). Ənənəvi dilçilikdə modallıq danışanın ifadə olunan fikrə münasibətini, ifadə olunan fikrin obyektiv varlığa münasibətini əks etdirən qrammatik-semantik kateqoriya kimi şərh olunur. Obyektiv modallıq isə ifadə edilən fikrin obyektiv varlığa münasibətini əks etdirir. Burada reallıq/qeyri-reallıq, mümkünlük/qeyri-mümkünlük, zərurilik/ehtimallıq və s. nəzərdə tutulur (Adilov, Verdiyeva, Ağayeva, 1989, s.164).

Kompüter dilçiliyində modallıq istənilən formada təqdim olunan informasiyanın (mətn, səs, simvol, şəkil və s.) təfəkkürdə təhlili və qavranması prosesində iştirak edən kateqoriya kimi başa düşülür. Bu o deməkdir ki, bizim dünya və özümüz haqqında təsəvvürlərimizin təbiəti eyni deyil, müxtəlifdir (müxtəlif modal-

lıq çərçivəsindədir). Biz həm sözlərlə, həm də şəkil görüntüləri ilə (yuxu prosesində olduğu kimi) düşünə bilərik. Hətta qoxu və ya dad təsəvvürləri ilə düşüncələrə, xatirələrə qapılmaq olar. Başqa sözlə, insanın bütün hiss orqanları təfəkkür prosesinə özünün modallıq xüsusiyyətləri ilə təsir göstərə bilər. Amma insan təfəkkürü üçün, anlama, dərk etmə prosesi üçün iki modallıq növü – simvol (mətn) və görüntü (şəkil) daha vacib və önəmlidir. Bu iki modallıq arasında sıx əlaqə vardır. Adətən hər hansı bir ad və ya hər hansı bir situasiyanın təsvirini ifadə edən mətn eyni zamanda həmin obyekt və ya situasiyanın görüntüsünü də yaratmış olur. Biz həmin məfhum və ya hadisəni sanki göz-lərimizlə görür, videogörüntüsünü təsəvvürümüzdə canlandırırıq. Mətn, ssenari, görüntü isə həmin ssenari əsasında çəkilmiş filmi xatırladır. Bu prosesin əksi də mümkündür. Gördüyümüz hər hansı bir hadisənin, obyektin görüntüsünü yazılı mətnə çevirmək olar (görüntüdən mətnə, mətndən görüntüyə keçid). Bizim təfəkkürümüzdə görüntü və mətn ayrı-ayrılıqda deyil, bütöv şəkildə təmsil olunur. Mətn təfəkkürdə obrazlı təəcəssümünü tapır. Kompüter dilçiliyi və süni intellekt sistemləri sahəsində fəaliyyət göstərən mütəxəssislər bu iki bir-birini tamamlayan, təfəkkürdə vahid bir tama çevirən komponentlərin təbiətinin öyrənilməsi ilə daha ciddi məşğul olmalıdırlar. Bu məsələlərin hərtərəfli araşdırılması anlama prosesinin təbiətini açmaq, onun elmi izahını vermək imkanı yaradacaqdır. Mətnlə görüntünün inteqrasiyasının qanunauyğunluqlarının müəyyənləşdirilməsi istiqamətində araşdırmalar uğurla davam etdirilir. Görüntü ilə mətn arasında əlaqənin eksperimental modeli yaradılmış və süni intellekt sistemlərində mətndən görüntüyə və görüntüdən mətnə keçidin texnologiyası hazırlanmışdır. Kompüter mətnə görə mümkün görüntünü və görüntüyə görə təxmini mətni təqdim etməklə informasiya texnologiyaları sahəsində yeni bir eranın başlanğıcını qoymuş olur. Səsli və görüntülü lüğət tiplərini də bu problemlə əlaqəli şəkildə təkmilləşdirmək olar.

Müasir dövrdə insan nitqini və təbii dillərdə mətnləri qəbul edən, təhlil blokundan keçirməklə qavraya bilən süni sistemlərin

yaradılması «insan-kompüter-insan» ünsiyyət kanalının yaranmasının əsasını qoydu. Bu da kompüter dilçiliyində yeni bir ünsiyyət kanalının – dialoq sisteminin formalaşmasına marağı artırdı. Dialoqların təşkili konkret məqsədlərlə bağlı olur və ünsiyyət həmin məqsədlərə istiqamətləndirilir. Dialoqun ssenarisi elə qurula bilər ki, aktiv tərəfin məqsədləri təmin olunsun. Eləcə də, dialoqda iştirak edən ikinci tərəfin maraqları ssenaridə dəstəklənə bilər və ya onun mümkün əks hərəkətləri qabaqcadan müxtəlif vasitələrlə neytrallaşdırıla bilər. Bunun üçün qabaqcadan müəyyən hazırlıq tələb olunur. Ssenari elə qurula bilər ki, danışanın niyyəti gizli qalsın, açıq-aşkar hiss olunmasın. Bütün sadalananlar kommunikasiya modellərinin ümumi məsələləri hesab oluna bilər. Konkret məqsəd və situasiyalardan asılı olsa da, «insan-kompüter-insan» ünsiyyət kanalının mahiyyəti və quruluşunda müvafiq dəyişikliklər aparıla bilər.

1.5. Kompüterlərdə mətnlərin formal işlənmə problemi

Kompüter dilçiliyinin mühüm tədqiqat sahələrindən olan təbii dilin (mətnlərin) işlənməsinə XX əsrin 60-cı illərindən başlanılmışdır. İngilis dilində *natural language processing* termininin tərcüməsi olan bu anlayış «süni intellekt» problemləri çərçivəsinə aid edilir. Təbii dilin (mətnin) formal işlənməsi geniş anlayışdır, kompüterdə bu prosesin reallaşdırılmasının bütün aspektlərini əhatə edir. Lakin bu anlayış bəzən dar mənada da izah olunur. Konkret bir dilin (mətnin) kompüterlərdə işlənməsinin metod və texnologiyalarının hazırlanması sistemi hesablama texnikasının insanla ünsiyyət kanalının yaranmasına xidmət edir.

XX yüzilliyin 70-ci illərində mətnlərin formal işlənməsi sistemləri sürətlə inkişaf etməyə başladı. Monoqrafiyanın sonrakı fəsilərində türk dillərində, o cümlədən Azərbaycan dilində mətnlərin kompüterlərdə işlənməsi və müxtəlif məqsədlərlə istifadəsi barədə ətraflı məlumat verilmiş, bütün dil səviyələrində formal təhlil alqoritmlərinin tərtibi və reallaşdırılması yolları göstərilmişdir.

Kompüterlərin tədrisdə istifadə olunması ilə bağlı proqramlaşdırma dillərinin təbii dillərə yaxınlaşdırılması məsələsi dillərin dərinlən araşdırılması zərurətini doğurdu. Məhz həmin vaxtlarda «Visual Basic», «Yazı stolu», «Kitabxana» əməliyyat sistemləri meydana gəldi. Proqramlaşdırma dillərinin təbii dillərə yaxınlaşdırılması perspektiv planda nəzərdə tutulurdu. Bunun üçün istifadəçinin nitq məlumatının təhlili, məlumatın interpretasiyası, cavabın mənasının formalaşdırılması, səthi strukturun sintezi məsələsi dərin araşdırmalar tələb edirdi. «İnsan-maşın-insan» dialoq sisteminin reallaşdırılması istiqamətində aparılan işlərdə ünsiyyət prosesində meydana çıxan uğursuzluqlar ayrıca təhlil edilir və onların aradan qaldırılması mexanizmi işlənilib hazırlanırdı. Bunun üçün insanın nitq fəaliyyətinin öyrənilməsi, modelləşdirilməsi və həmin modellər əsasında sistemin təkmilləşdirilməsi aparılırdı. Bu aspektdə psixolinqvistik, neyrolinqvistik araşdırmaların intensivləşdirilməsi məsələsi aktuallaşırdı. Bütün bunlar kompüter dilçiliyinin tədqiqat sahələrinin genişlənməsi və dərinləşməsi fonunda baş verirdi.

Kompüterlərdə təbii dilin araşdırılmasında sual-cavab sistemləri xüsusi blok təşkil edir. Sual-cavab, dialoq sistemləri məlumat-axtarış sistemlərində informasiyanın tapılması və zərurət olduqda, yenidən kodlaşdırılması məqsədi ilə aparılır. Dialoq sistemlərində verilən sorğulara cavabın təbii dildə formalaşdırılması məqsəd kimi qarşıya qoyulurdu. Belə sistemlərdə sorğu təbii dildə (ingilis, rus, alman və s.) verilirdi və cavab da konkret təbii dildə alınırı. Təhlil bütün dil səviyyələrində (morfoloji, sintaktik, semantik) aparılırdı. Sintez prosesi də çoxmərhələli idi.

Əlaqəli mətnlərin işlənmə sistemi müxtəlif quruluşlarda təzahür edir. Burada biliyin təqdimi texnologiyası ilkin mühüm mərhələdir. Sistem mətni «qavramalı» və «qavradığı» səviyyədə cavab verməyi bacarmalıdır. «Qavrama» müəyyən kommunikativ niyyətlə mətndən məlumatın əldə olunması deməkdir. Mətndən potensial istifadəçinin öyrənmək istədiyi hissə götürülür. Başqa sözlə, mətndə olan informasiyadan məqsədyönlü şəkildə istifadə olunur. Vahid proqram kompleksinə daxil edilən

«RESEARCHER» və «TAILOR» sistemlərini bunlara misal göstərmək olar (<http://files.school-collection.edu.ru/dlrstore>).

Təbii dillərin kompüterlərdə işlənməsinin mühüm bir istiqaməti də məlumat-axtarış sistemlərinin işlənilib hazırlanmasıdır. Elmi-texniki mətnlərin həcminin durmadan artması bu istiqamətin yaranmasına səbəb oldu. Məlumat-axtarış sistemləri saxlanan və işlənən informasiyaların tipinə görə iki qrupa ayrılır: sənədlər və faktoqrafik məlumatlar. Sənəd bölümündə sənədlərin mətni və onların təsviri (referatlar, biblioqrafik kartoçkalar və s.), faktoqrafik bölümündə isə konkret faktların təsviri yerləşdirilir. Bu, mətn şəklində olmaya da bilər (cədvəllər, düsturlar, sxemlər, qrafik və s.). Məlumat-axtarış sistemlərinin qarışıq tipində eyni zamanda həm sənəd mətnləri, həm də faktoqrafik təsvirlər yerləşdirilə bilər. Sənədlərə müraciətlər məlumat-axtarış tezaurlarına əsaslanan məlumat-axtarış dilləri vasitəsi ilə həyata keçirilir. Məlumat-axtarış dili sənədlərinin ayrı-ayrı aspektlərinin təsviri üçün nəzərdə tutulmuş formal dildir. Məlumat-axtarış dilində sənədlərin təsviri proseduru indeksləşmə adlanır. İndeksləşmə nəticəsində hər bir sənədin məlumat-axtarış dilində formal təsviri alınır. Həmin formal təsvir sənədin axtarış obrazı hesab olunur. Analoji olaraq sorğu da indeksləşdirilir və onun axtarış obrazı hazırlanır. Məlumat axtarışı alqoritmi sənədin axtarış obrazının müqayisəsinə əsaslanır. Müqayisə nəticəsində sənədin axtarış obrazı ilə sorğunun axtarış obrazı üst-üstə düşməyə bilər və ya qismən uyğun gələ bilər. Xüsusi hallarda istifadəçi sənədin əldə edilməsi meyarlarını özü təklif edir. Bu, məlumata olan tələb əsasında qurulur. Avtomatlaşdırılmış məlumat-axtarış sistemlərində çox vaxt deskriptor məlumat-axtarış dillərindən istifadə olunur. Sənədin mövzusu deskriptorların məcmuyu ilə ifadə olunur. Deskriptor funksiyasında adi sözlər, terminlər, açar sözlər, həmin sənədə dair elementar kateqoriya və anlayışlar çıxış edirlər. Sənədin aid edilə biləcəyi bütün sahələri əhatə edən sayda deskriptor məlumat-axtarış sisteminə daxil edilir. Deskriptorların sayına məhdudiyət qoyulmur. Bu da hər hansı sənədi çoxölçülü əlamətlər matrisində axtarmaq imkanı yaradır. Des-

kriptorların mümkün birləşmələrinə bəzən məhdudiyət qoyula bilər. Deskriptor birləşmələri məlumat axtarış dilinin sintaksisinin varlığını göstərir.

Deskriptor dili əsasında qurulmuş ilk məlumat-axtarış sistemi M.Taube tərəfindən yaradılmış UNITERM sistemidir. Bu sistemdə deskriptor funksiyasını sənədin açar sözləri- unitermlər yerinə yetirir. Bu sistemdə məlumat dilinin lüğəti qabaqcadan verilmir. Lüğət sənədin və sorğunun indeksləşdirilməsi prosesində yaranır. Müasir məlumat-axtarış sistemləri tezaurussuz qurulur. Sistem məhdud təbii dildən istifadə olunmaqla fəaliyyət göstərir. Axtarış sənədlərin özləri, referat mənbələri əsasında aparılır. Tezaurussuz məlumat-axtarış sistemlərində təbii dilin söz və söz birləşmələrindən istifadə olunur.

Təhlil dillərinin kompüterlərdə işlənməsinin ən geniş yayılmış və ən çox istifadə olunan növü hipermətn sistemidir. Hipermətn sistemi kompüter dilçiliyinin obyektinə kimi XX əsrin 60-cı illərindən məlumdur. Hazırda dünya İnternet şəbəkəsini onuz təsəvvür etmək mümkün deyil.

Hipermətn mətnlərin xüsusi üsulla təşkilidir. İstifadəçi mətni və onun ayrı-ayrı fraqmentlərini müxtəlif əlaqə tiplərinə görə bağlaya bilər. Burada assosiativ münasibətlər də mühüm rol oynayır. Müxtəlif ardıcılıqla internetdə yerləşdirilmiş mətnə istənilən vaxtda müraciət etmək, oradan istənilən mövzu ilə bağlı məlumat əldə etmək mümkündür. Müxtəlif münasibətlərə görə bir-biri ilə əlaqələndirilmiş və optimal keçid sistemində malik mətn hipermətnlər adlanır. Hipermətn sistemi informasiya texnologiyaları sahəsində inqilab kimi səciyyələndirilə bilər. Məsələn, biz «Google» axtarış sistemində istənilən hadisə, elmi məlumat, bu və ya digər şəxs haqqında informasiya əldə edə bilərik. İngilis, rus, Azərbaycan və s. dillərdə yazılmış hər hansı söz, termin və ya ad barədə dərhal hərtərəfli məlumat verilir. Bu, hipermətn sistemi vasitəsi ilə mümkün olur. İnternet sistemində hipermətn qraf şəklində verilmişdir. Qrafın qovşaqlarında (kəsişmə nöqtələrində) mətnlər, onların fraqmentləri, təsvirləri, cədvəllər, videosüjetlər və s. məlumatlar yerləşdirilmişdir. Qovşaqlar müxtəlif

münasibətlər əsasında əlaqələndirilmişdir. Bu əlaqə tipləri hipermətnlərin proqram təminatı ilə məşğul olan mütəxəssislər və ya istifadəçilərin özləri tərəfindən müəyyənləşdirilir. Həmin əlaqə tiplərinə görə hipermətndə hərəkət etmək, müəyyən mövzu istiqamətinə yönəlmək mümkündür. Münasibət əlaqəsi biristiqamətli və ya çoxistiqamətli ola bilər. İkiistiqamətli ox işarəsi istifadəçiyə hər iki tərəfə, biristiqamətli isə bir tərəfə getmək imkanı yaradır. Mətnin komponentlərini nəzərdən keçirən istifadəçinin keçdiyi düyün nöqtələri zənciri *yol* və ya *marşrut* adlanır.

Hipermətnlərin kompüter realizasiyası iyerarxik və ya şəbəkə şəklində olur. Hipermətnin ağacabənzər iyerarxik quruluşu onun komponentləri arasında keçidi məhdudlaşdırır. Belə hipermətnlərdə komponentlər arasında nisbət, cins-növ əlaqələri əsasında qurulan tezaurusları xatırladır. Şəbəkə hipermətni cins-növ münasibətləri ilə məhdudlaşmadan komponentlər arasında müxtəlif tip münasibətlərdən istifadə imkanı yaradır. İnternet məkanında hipermətnlərin statik və dinamik növləri yayılmışdır. Statik hipermətn istismar müddətində dəyişməz qalır. İstifadəçi həmin mətnlərə öz şərhlərini əlavə edə bilər. Lakin bu qeydlərin hipermətnin işinə təsiri olmur. Dinamik hipermətnlərdə dəyişmələr normal mövcudluq forması hesab olunur. Dinamik hipermətnlər müxtəlif tipli məlumat xidmətlərində istifadə olunur. Bəzi məlumat xidmətlərində məlumat axınına fasiləsiz olaraq təhlil etmək lazım gəldiyindən dinamik hipermətnlərdən istifadə zəruridir.

Hipermətn elementləri arasında əlaqə onun yaradıcıları tərəfindən qeydə alınır. İstifadəçinin belə mətnlərə müraciəti prosesində də bu əlaqə yarana bilər. Birinci halda söhbət sərt, ikinci halda isə yumşaq strukturlu hipermətnlərdən gedir. Sərt strukturlu hipermətnlərin yaradılma texnologiyası məlumdur. Yumşaq strukturlu hipermətnlər isə sənədlərin və ya məlumat mənbələrinin semantik yaxınlığının təhlili əsasında qurulur. Hazırda internet məkanında açar sözlərə əsaslanan yumşaq strukturların texnologiyası geniş istifadə olunur. Hipermətn şəbəkəsində bir düyün nöqtəsindən digər düyün nöqtəsinə keçid açar sözlərin

axtarışı ilə həyata keçirilir. Açar sözlər hər dəfə müxtəlif olduğu üçün hipermətnlərin strukturu da hər dəfə dəyişmiş olur.

Hipermətn sisteminin təşkili texnologiyası mətn formasında olan və olmayan məlumatları fərqləndirmir. Vizual və səsli məlumatların bu sistemə daxil edilməsi (vidiosüjet, şəkil, foto, audioyazı və s.) interfeysin dəyişdirilməsini, daha güclü proqram və kompüter şəbəkəsinin olmasını tələb edir. Belə sistemlər hipermedia və ya multimedia adlandırılır. Tədris prosesində, ensiklopediyaların kompüter variantlarının hazırlanmasında multimedia sistemlərindən geniş istifadə olunur.

Kompüter dilçiliyinin mühüm sahələrindən biri olan kompüter lüğətçiliyində də təbii dil mühüm vasitədir. Bu tipli lüğətlərin tərtibində və istifadəsində təbii dillərdən optimal yararlanmanın texnologiyası işlənib hazırlanmışdır. Xüsusi proqramlar, məlumat bazaları, kompüter kartotekaları, mətnlərin formal işlənməsi, saxlanması, lüğət məqalələrinin avtomatik rejimdə formalaşdırılması imkanı yaratmışdır. Kompüter lüğətçiliyi proqramları iki böyük qrupa ayırır:

- lüğətçilik sahəsindəki işlərə yardımçı proqramlar;
- leksikoqrafik bazalardan təşkil olunmuş müxtəlif tipli avtomatik lüğətlər.

Avtomatik lüğətlər kompüterlərdə istifadəçi və mətnlərin işlənməsi proqramlarında istifadə üçün nəzərdə tutulmuş xüsusi məşin formatında olan lüğətlərdir. Başqa sözlə desək, avtomatik lüğətlərin bəziləri insan-istifadəçi, digərləri isə mətnlərin avtomatik işlənməsi proqramları üçün nəzərdə tutulmuşdur. İstifadəçi üçün nəzərdə tutulmuş avtomatik lüğətlər məşin tərcüməsi, avtomatik referatlaşdırma, məlumat axtarışı sistemləri üçün nəzərdə tutulmuş avtomatik lüğətlərdən interfeysə və lüğət məqalələrinin quruluşuna görə çox fərqlənir. İstifadəçilərə çox vaxt hamıya yaxşı məlum olan adi lüğətlərin kompüter versiyaları təqdim olunur. Proqram təminatı bazarında ingilis və rus dillərinin izahlı lüğətlərinin kompüter analoqları geniş istifadə olunur. Bu bərdə «Kompüter lüğətçiliyi» fəslində ətraflı məlumat veriləcəkdir. Mətnlərin işlənməsi proqramları üçün nəzərdə tutulmuş lüğətləri

avtomatik lüğətlər adlandırmaq daha düzgün olardı. Geniş istifadəçi kütləsi üçün hazırlanmış lüğətlər adi lüğətlərin elektron variantı olduğu üçün onları şərti olaraq avtomatik lüğət adlandırmaq olar.

Kompüter dilçiliyinin daha bir perspektivli sahəsi süjetlərin quruluşunun kompüter modelləşdirilməsidir. Süjet quruluşunun modelləşdirilməsi struktur ədəbiyyatşünaslıq, semiotika və kulturologiyaya aid edilir.

Süjetin modelləşdirilməsinin kompüter proqramları süjetin formal təqdiminin üç aspektinə əsaslanır:

- süjetin morfoloji təqdimi;
- süjetin sintaktik təqdimi;
- koqnitiv yanaşma.

Morfoloji təqdimatda empirik materialın təhlili əsasında sıralanmış tipikləşdirilmiş situasiyalar çoxluğu alınır. Mətnə müşahidə olunan analogi situasiyalar ardıcıl olaraq təkrarlanır və personajların görüşlərinin ssenariləri alınır. Başqa sözlə, süjetin morfoloji təqdimatında təkrarlanan eynitipli hadisələr modelləşdirilir və onun ümumiləşdirilmiş ssenarisi alınır.

Sintaktik yanaşma ilə mətnin süjetinin təqdiminin nəzəri əsasını «süjet qrammatika»ları təşkil edir. Bu qrammatika ötən əsrin 70-ci illərində N.Xomski tərəfindən təşkil olunmuş generativ qrammatika prinsipləri əsasında hazırlanmışdır. Əgər generativ qrammatikada sintaktik struktur fel və ad qruplarına ayrılırsa, süjet qrammatikalarının əsas anlayışları ekspozisiya, hadisə və epizoddur.

Kompüter dilçiliyinin ən çox müzakirə olunan və heç vaxt gündəmdən düşməyən istiqamətlərindən biri məşın tərcüməsidir. Bu barədə növbəti fəsillərdə məlumat veriləcəkdir.

II FƏSİL. AZƏRBAYCAN DİLİNİN MİLLİ KORPUSU

2.1. Ümumi məsələlər

Azərbaycan dilinin milli korpusu, onun yaradılmasının nəzəri və təcrübi məsələləri, quruluşu, tərkibi və s. məsələlərin izahına keçməzdən əvvəl bəzi anlayışları konkretləşdirilməsinə ehtiyac vardır.

Hər şeydən əvvəl, *korpus* nədir və bu termin hansı mənalarda işlənir?

“Azərbaycan dilinin izahlı lüğəti”ndə bu sözün latın mənşəli olduğu göstərilir və 5 mənası qeyd olunur:

1. Ümumi sahədə yerləşən binalardan hər biri, yaxud böyük binanın ayrıca hissəsi;
2. Bir neçə diviziya və ya briqadadan ibarət iri qoşun birləşməsi;
3. Bir hökumət yanındakı xarici dövlətlərin nümayəndələri;
4. İnqilabdan əvvəlki Rusiyada orta hərbi məktəb adlarının tərkibində işlənir;
5. On punktluq mətbəə şrifti (hürufatı) (ADİL, II cild, s. 738-739).

“Rusca-azərbaycanca lüğət”də “Azərbaycan dilinin izahlı lüğəti”ndəki bütün mənalar eyni ilə verilir. Amma birinci mənə kimi 1) “bədən”, gövdə (insanda, heyvanda); 2) korpus, gövdə, çanaq, təknə; корпус судна – gəminin korpusu (gövdəsi) göstərilir (Rusca-azərbaycanca lüğət, I cild, s. 573).

“İngiliscə - azərbaycanca lüğət”də latın mənşəli bu söz 1) bədən, gövdə (insan və heyvan haqqında); 2) toplu, külliyyat, məcmuə (sənəd və s. haqqında) kimi izah olunur (İngiliscə - azərbaycanca lüğət, s.344).

Müxtəlif lüğətlərə müraciət etməkdə məqsəd bu çoxmənəli sözün semantik tutumunu müəyyənləşdirmək, ifadə etdiyi anlayışların mənə bağlılığını müəyyənləşdirməkdir.

Kompüter dilçiliyində *korpus* bir qədər fərqli anlamda işlənir. Kompüter dilçiliyində *korpus* dedikdə, hər hansı bir konkret dildə mətnlərin elektron formada toplanmasına istiqamətlənmiş məlumat axtarışı sistemi başa düşülür. Konkret dilin milli korpusunda müəyyən mərhələdə (tarixi dövrdə) dil tam şəkildə – bütün üslubları, janrları, ədəbi dili, dialektləri ilə təmsil olunur.

Hazırda kompüter dilçiliyi ilə yanaşı, dilçiliyin daha yeni bir sahəsi – korpus dilçiliyi yaranmaqdadır. Fikrimizcə, korpus dilçiliyini müstəqil bir dilçilik sahəsi kimi səciyyələndirmək düzgün olmazdı. Korpus dilçiliyi kompüter dilçiliyinə aid edilə bilər. Kompüter dilçiliyi korpus dilçiliyinin obyekt və predmetini ehtiva edir. Korpus dilçiliyi kompüter dilçiliyi daxilində daha aydın başa düşülür və izah olunur. Monoqrafiyada korpus dilçiliyinə kompüter dilçiliyinin bir komponenti, bölməsi kimi yanaşılır və bu kontekstdə izah olunur.

2.2. Milli dil korpusları

Konkret dillər üçün milli korpusların yaradılması ideyası yeni deyil. Əvvəllər bu, maşın fondu adlandırılırdı. Göründüyü kimi, burada termin müxtəlifliyi vardır. Vaxtilə keçmiş SSRİ məkanında SSRİ xalqlarının dillərinin milli korpuslarının yaradılması ilə bağlı intensiv işlər aparılırdı. Rus dili ilə yanaşı, hər bir müttəfiq respublikanın dili üçün milli korpus yaradılması planlaşdırılırdı.

Qeyd etmək lazımdır ki, indinin özündə də *maşın fondu* və *milli korpus* paralel terminlər kimi işlənməkdədir. Hətta eyni müəllifin eyni konfransa təqdim etdiyi məruzələrdə həm maşın fondu, həm də milli korpus barədə məlumat verilir. Görünür, bu terminlərin sabitləşməsi üçün hələ müəyyən qədər vaxt lazımdır.

“Başqırd dilinin milli korpusu” ilə bağlı Z.A.Sirazitdinov, A.Ş.İşmuxametova və A.D.İbrahimovanın təqdim etdikləri məruzədə bu məsələ ilə bağlı maraqlı ideyalar irəli sürülür. Müəlliflər göstərirlər ki, informasiya texnologiyalarına dair terminologiyanın bugünkü durumundan çıxış etdikdə, “milli korpus” termini yerinə “kompüter fondu” işlətmək daha düzgün olardı.

Ötən əsrin 80-ci illərində SSRİ Elmlər Akademiyası rus dilinin maşın fondunun yaradılması barədə proqram qəbul etmişdi. Bu proqram çox geniş müzakirə olunduqdan sonra qəbul olundu. İndinin özündə də internet şəbəkələrində “Rus dilinin maşın fondu” adı qalmaqdadır. Lakin son zamanlar “Rus dilinin milli korpusu” termini əvvəlki adı – “Rus dilinin maşın fondu” ifadəsini üstələməkdədir.

Keçmiş SSRİ ərazisində yaşayan xalqların bəziləri milli korpusla bağlı müəyyən işlər görmüş, ayrı-ayrı fraqmentlər üzrə tədqiqatlar aparmışlar (qazax, litva, eston, buryant, osetin və s.). Başqırd dilinin milli korpusunun təmsalında məsələnin ümumi cəhətləri, prinsipləri, strukturu barədə müəyyən təsəvvür əldə etmək olar. Əlbəttə, keçmiş SSRİ məkanında milli dil korpusları arasında həcmi və mükəmməlliyi ilə seçilən rus dilinin milli korpusudur.

O.S.Rublyova rus dilinin milli korpusu sahəsində aparılan tədqiqatları kompüter dilçiliyinin predmetinə aid edir. Kompüterlərdə lüğət materialları maşın kartotekalarında saxlanılır və istifadə olunurdu. Kompüter texnologiyası inkişaf edib təkmilləşdikcə daha mükəmməl leksikoqrafik məlumatlar bazası yaradılmağa başladı. O.S.Rublyovanın fikrincə, “Rus dilinin milli korpusu”nu belə leksikoqrafik bazalardan biri hesab etmək olar (Рублёва, 2010 s. 34). Həmin vaxtdan başlayaraq çap olunmuş lüğətlərin elektron variantlarının yaradılması sahəsində işlər sürətləndi. Qlobal şəbəkə əlaqələri və müxtəlif axtarış sistemləri imkanı olan yeni lüğət növləri və “korpus” sistemlərinin yaranmasına maraq da getdikcə artmağa başladı. Məsələnin nəzəri və praktik cəhətləri indi də leksikoqrafların və kompüter dilçiliyi mütəxəssislərinin diqqət mərkəzindədir. Əslində, burada söhbət çoxfunksiyalı mətn məlumatları bazalarının və ya, başqa sözlə, korpusların yaradılmasından gedir.

“Korpus” sözü burada özünün lüğəvi mənasında deyil, yeni anlamda işlənir. Dünya dilçiliyində, o cümlədən rus dilçiliyində korpus elektron daşıyıcıda saxlanan, müxtəlif dil hadisələri və aspektləri ilə bağlı lingvistik tədqiqatlara material verən, müəy-

yən nizamla düzülmüş təbii dil mətnləri çoxluğudur. Korpusu təşkil edən mətnlər və onlara müraciət müəyyən qaydalar əsasında aparılır.

İlk çoxfunksiyalı məlumat bazaları və lüğətlər 1956-cı ildə ABŞ-da yaradılmışdır. Elektron məlumatlar bazası əsasında bir çox lüğətlər, məsələn, *Webster's English Dictionary* hazırlanmışdır (1996).

XX əsrin 60-70-ci illərində kompüter texnologiyasının inkişafı ilə əlaqədar ilk korpuslar meydana çıxdı. Orada sözlərin miqdarı məhdud idi (Kennedy, 1998).

Həmin dövrdə məlum olan məşhur korpuslar *The Brown Corpus*, *Lancaster-Oslo/Bergen Corpus*, *London-Lund Corpus* hesab olunur (Рублёва, 2010 s.35).

Söylənilənləri yekunlaşdıraraq, belə qənaətə gəlmək olar ki, milli korpuslar dilçilər tərəfindən elmi araşdırmalar və dilin tədrisi prosesini optimallaşdırmaq məqsədi ilə yaradılır. Korpus dili bütün janr və üslub müxtəliflikləri, dialektləri və peşə leksikası ilə müəyyən mərhələdə və ya onun mövcudluğunun bütün mərhələlərində təmsil edir.

Hazırda dünyanın bir çox ölkələrində dünyanın aparıcı dilləri üçün milli korpus tərtib olunmuşdur. 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 dilləri ailəsində rus və çex dillərinin milli korpusları xüsusilə seçilir.

2.2.1. Türk dillərinin milli korpusu

Ayrı-ayrı türk dilləri üçün maşın fondunun yaradılması ideyasının irəli sürülməsi və inkişafında R.Q.Piotrovski, A.M.Şerbak, V.Q.Quzevin xidmətləri böyükdür (Гузев, Пиотровский, Щербак, 1988, c.92-101).

1988-ci ilin mayında Moskvada Sovet Türkoloqları Komitəsinin XIV plenumunda keçmiş SSRİ miqyasında türk dillərinin maşın fondunun yaradılması ideyası irəli sürülmüş və türk dillərinin maşın fondunun yaradılmasının əsas istiqamətləri müəyyən-

ləşdirilmişdi. Qərara alınmışdı ki, türk dillərinə aid leksikoqrafik, qrammatik, statistik-üslubi, tarixi-etimoloji məlumatlar toplanılmalı və müqayisə-qarşılaşdırma planında sistemləşdirilməli, daha sonra türk dilləri haqqında istənilən məlumatı dolğun və düzgün şəkildə doğura və təqdim edə biləcək qaydalar işlənilib hazırlanmalıdır. Türk dillərinin maşın fondunun yaradılması üçün mərkəz olaraq Qazaxıstan SSR Elmlər Akademiyasının Dilçilik İnstitutu tövsiyə olunmuşdu. Başlanğıc mərhələdə türk dillərində sözlərin qrammatik formalarının yaranma proseslərinin öyrənilməsi üçün birhecalı sözlərin struktur-fonetik müxtəlifliklərinin tədqiqi ön plana çəkilmişdi. Gələcəkdə həm ümumtürk dil sistemini, həm də hər bir konkret dili modelləşdirməyə qabil olan çoxfunksiyalı böyük Türk Dillərinin Maşın Fondunun (TDMF) reallaşdırılması nəzərdə tutulurdu. TDMF müqayisə-qarşılaşdırma planında türk dillərinin leksikoqrafik, qrammatik, tarixi-etimoloji xüsusiyyətlərinə aid məlumatların toplanması və sistemləşdirilməsi işlərini tənzimləməli idi. Eyni zamanda digər türkdilli regionlarda da bu problemlə bağlı intensiv işlər aparılmağa başlanmışdı (Джубанов, 2012, s.154).

Türk dillərinin maşın fondunun (TDMF) yaradılmasının ilkin mərhələsində türk ulu dilinin rekonstruksiyası məsələsinin həlli nəzərdə tutulurdu. Bu mürəkkəb məsələnin araşdırılması üçün aşağıdakı məlumatların toplanılması zəruri idi:

- türk dillərinin birhecalı söz köklərinin müxtəlif növləri barədə struktur-fonetik informasiya;
- morfem siyahıları;
- sintaktik əlaqələri əks etdirən sxemlər;
- affikslərin qrammatik tezaurusu;
- konkret dillərin fonetik, qrammatik quruluşu barədə analitik göstəricilərin toplusu.

Beləliklə, ilkin mərhələdə ümumtürk və dillərarası fonomorfoleksik-semantik təhlil problemlərinin öyrənilməsinə imkan verən türk dillərinin linqvistik bankının yaradılması nəzərdə tutulurdu.

İkinci mərhələdə çoxfunksiyalı böyük TDMF-nin yaradılması planlaşdırılırdı. Həmin fond həm ümumtürk, həm də hər bir konkret dil sistemini modelləşdirməklə praktik istifadə üçün yararlı olmalı idi. TDMF-in əsas blokları leksik, qrammatik, fonoloji və morfonoloji bloklar hesab olunurdu.

Maşın fondunun layihələşdirilməsi və proqramlaşdırılması “yuxarıdan aşağıya” prinsipi ilə aparılmalı idi. Bu, ayrı-ayrı modulların uzlaşdırılması baxımından məqsədəuyğun idi. Hesab olunurdu ki, əvvəlcə kompüterdə hər bir konkret türk dili hər-tərəfli işlənməli, toplanmış təhlilin nəticələri sonradan ümumtürk kontekstinə çıxarılmalıdır. Qeyd etmək lazımdır ki, ideya müəllifləri ümumittifaq miqyasında türkdilli respublikaları bu işə səfərbər etmək üçün gərgin fəaliyyət göstərirdilər. Beynəlxalq elmi-nəzəri jurnal olan “Sovetskaya tyurkologiya”da türk dillərinin maşın fondunun yaradılması barədə geniş məqalə bu məqsədlə çap olunmuşdu (Гузев, Пиотровский, Щербак, 1988, s.92-101).

2.2.2. Başqırd dilinin milli korpusu

Başqırd dilinin maşın fondunun yaradılması ilə bağlı işlərə 2003-cü ildən başlanılmışdır (Сиразитдинов, Ишмухаметова, Ибрагимова, 2012, s. 35-37).

Başqırd dilinin maşın fondu xüsusi olaraq dilçilər, müəllimlər, tələbələr və yuxarı sinif şagirdləri üçün nəzərdə tutulmuşdur. Fonda xüsusi məlumatlar bazası yerləşdirilmişdir. Məlumatlar bazasında və proqram təminatında sorğuların işlənməsi və yerinə yetirilməsi məqsədi ilə interfeyslər yaradılmışdır. İnterfeys “birləşmə, təmas, əlaqə yeri, üsulu” mənasında işlənir. Əgər fərdi kompüterin, proqramın funksiyasının interfeysi dəyişməz qalırsa, həmin obyektin digər obyektlərlə qarşılıqlı təsir prinsiplərini dəyişdirmədən onun özünü modifikasiya etmək olar. Məsələn, Windows proqramlarında interfeys eynidir. Beləliklə, interfeys dedikdə – istifadəçinin müxtəlif qurğularla ünsiyyətdə istifadə etdiyi vasitələr başa düşülür (Əliquliyev, Şükürlü, Kazımova, 2009, s.87).

Başqırd dilinin maşın fondu 7 məlumat bazasını özündə birləşdirən altfondlardan ibarətdir:

- əsas kartotekalar altfondu;
- leksikoqrafik altfond;
- qrammatik altfond;
- əlyazmaların kataloqu altfondu;
- qədim çap kitablarının kataloqu altfondu;
- təcrübi-fonetik altfond;
- dialektoloji altfond.

Əsas kartoteka dilin leksik sistemi haqqında zəruri informasiyanı ehtiva edən 100000 kök söz və düzəltmələrdən ibarətdir. Həmin kartotekanın dilin bütün qatlarını əhatə etdiyi iddia olunur. Hər bir leksik vahid üçün 50-dən çox əlamət göstərilir. Əsas kartotekada sözün hansı nitq hissəsinə aidliyi, mənşəyi, üslubu, dialektə və ya ədəbi dilə mənsubluğu, tarixizm və ya arxaizm, neologizm olması, şəxslı (şəxssız, ümumi) xüsusi və s. barədə məlumatlar yerləşdirilmişdir.

Əsas kartoteka fondun digər altfond bazaları ilə əlaqəlidir. Bu, lazım gəldikdə, əlavə informasiya əldə etmək imkanı yaradır.

Leksikoqrafiya altfondunda söz haqqında müasir başqırd dilinin lüğətləri əsasında hazırlanmış məlumatlar verilmişdir. Altfondda 500000 lüğət vahidi barədə lüğət məqaləsi verilmişdir. Leksikoqrafik altfondda akademik və tədris profilli lüğətlər: birdilli, ikidilli, çoxdilli, tezlik, terminoloji, frazeoloji, sinonim, sorğu lüğətləri, yer adlarını (küçə, şəhər, yaşayış məntəqələri və s.) bildirən onomastik lüğətlər və s. təmsil olunmuşdur.

Təcrübi-fonetik altfondda başqırd dili sait və samitlərinin artikulyasiya səciyyələri yerləşdirilmişdir. Burada 8000 vahiddən ibarət fonetik lüğət verilmişdir. Həmin materialdan başqırd dilini müstəqil öyrənənlər də istifadə edə bilərlər.

Başqırd yazılı ədəbi dili barədə məlumat verilməsi məqsədi ilə fondda əlyazmaları və qədim çap kitablarını əks etdirən 2000-dən çox vahidi birləşdirən daha iki kataloq təmsil olunmuşdur. Kataloqlar mövcud əlyazma və qədim çap kitablarının təsvirini və aşağıda qeyd olunan məlumatları verir:

Başlıq (rus dilinə tərcümə ilə), başlığın transliterasiyası, müəllif (əsərin müəllifi), müəllifin adı (transliterasiyada), əlyazmanın üzünü köçürən şəxs barədə məlumat, il (nə vaxt köçürüldü), həcmi, formatı (səhifədə sətirlərin sayı), səciyyəsi, annotasiyası, kim tərəfindən tapılıb və təhvil verilib, dili (ərəb, qədim türk, osmanlı və s.), paleoqrafiyası, harada saxlanılır, şifri və s.

Dialektoloji altfond 3 müstəqil bazadan – leksik, kartoqrafik və tekstoloji bazalardan təşkil olunmuşdur.

Qrammatik altfondda akademik qrammatikalar, başqırd dilinin sözdəyişdirmə sisteminin alqoritmik təsviri, eləcə də morfemlərin statistik bazası barədə məlumatlar toplanmışdır. (Сиразитдинов, Ишмухаметова, Ибрагимова, 2012, s.35-43).

A.K.Djubanovun 75 illiyinə həsr olunmuş elmi-nəzəri konfransda daha bir məruzə başqırd dilinin maşın fondunun yaradılması problemlərinə həsr olunmuşdu (Бускунбаева, Сиразитдинов, 2012, s.54-55).

Müəlliflərin məlumatına görə başqırd dilinin maşın fondu Rusiya Elmlər Akademiyasının Ufa elm mərkəzində Tarix, Dil və Ədəbiyyat İnstitutunun Linqvistika və İnformasiya Texnologiyaları laboratoriyasında 2011-ci ildən başlayaraq yaradılır. Fonda başqırd dilinin 4 əsas – bədii, publisist, elmi-tədris, rəsmi-işgüzar – üslubunu əhatə edən, XIX əsrin 20-ci illərindən bu günə qədər çap olunan mətnlər daxil edilir.

Başqırd dilinin maşın fondunun yaradılması sahəsində qazanılmış təcrübə və materiallar korpusun hazırlanmasında istifadə olunur. Bura avtomatik təhlil və sintez alqoritmləri, leksikoqrafik baza və s. daxildir. Korpusda söz-formaları lüğət vahidi formasına gətirməyə imkan verən morfoloji təhlil alqoritmləri əsas komponentlərdən hesab olunur.

2012-ci ildə başqırd dilinin milli korpusunun pilot layihəsi hazırlanmışdır.

Hazırda korpusda XX əsrin əvvəllərindən indiyə qədərki dövrü əhatə edən 63 müəllifin 579 əsərinin elektron variantı hazırlanmış, ümumi həcmi 9277754 söz-forma (sözişlətmə) redaktə

olunmuşdur. Həmin mətnlər 1981-ci ildə qəbul olunmuş başqırd dilinin yeni orfoqrafiyasına uyğunlaşdırılmışdır.

Korpusda söz-formaların axtarışı aşağıdakı parametrlərə uyğun aparılır:

1) Ekstralinqvistik marker istifadəçiyə imkan verir ki, konkret parametrləri göstərməklə axtarış sahəsini məhdudlaşdırsın. Bu da istifadəçiyə lazım olan informasiyanın qısa müddətdə əldə etmək üçün şərait yaradır. Ekstralinqvistik məlumatda müəllif (soyadı, adı, atasının adı, təvəllüdü, cinsi və s.), əsər (adı, yaranma ili, janrı, mövzusu, təqdim növü, kitab, jurnal, elektron mətn və s.), mənbənin adı və nəşr ili barədə qeydlər əks olunmuşdur. Belə ekstralinqvistik nişanlar istifadəçiyə istədiyi əlamətlər üzrə axtarış aparmaq imkanı yaradır (burada nişan “marker” mənasında işlənir).

Morfoloji marker istifadəçinin konkret morfoloji səciyyələrinə görə bu və ya digər söz-formanı seçməsinə reallaşdırır. Hər bir söz-formanın leksem mənsubiyyəti və morfoloji əlamətləri qabaqcadan göstərilmişdir: Sözün ilkin forması, nitq hissəsi, əlaməti, qrammatik kateqoriya əlamətləri.

Semantik markerdə leksemlərin geniş tematik sinifləri və sözəmələgətirmə səciyyələri əks olunmuşdur.

Hazırda başqırd dilinin milli korpusunun periodika (qəzet, jurnal), folklor mətnləri, rəsmi-işgüzar və elmi mətnləri əhatə edən altkorpuslarının yaradılması işlərinə başlanılmışdır (Быс-кунбаева, Сиразитдинов, 2012, s.54-58).

2.2.3. Rus dilinin milli korpusu

Rus dilinin milli korpusunun yaradılmasına 2003-cü ildən Rusiya Elmlər Akademiyasının “Filologiya və informatika” proqramı çərçivəsində başlanılmışdır.

Hazırda “Rus dilinin milli korpusu” təxminən 140 milyon sözü əhatə edir və gələcəkdə onun həcminin 200 milyona çatdırılması nəzərdə tutulur (Рублёва, 2010 s.36).

Rus dilinin milli korpusunun ən mühüm xüsusiyyətlərindən biri balanslaşdırılmış mətn korpusuna malik olması və dili lazımi

dərəcədə təmsil edə bilməsindədir. Bu, o deməkdir ki, korpus dildə mövcud olan yazılı və şifahi mətnlərin bütün tiplərini (müxtəlif janrlı bədii, publisist, tədris, elmi, işgüzar, danışq, dialekt və s.) imkan daxilində ehtiva edir. Həmin mətnlər dildə malik olduqları paya müvafiq olaraq korpusda təmsil olunmuşdur. Qeyd etmək lazımdır ki, dili təmsil edə bilmək üçün korpusun həcmi kifayət qədər böyük olmalıdır (10 və 100 milyon söz-forma (sözişlətmə)).

Rus dilinin milli korpusu (RDMK) tərkibinə daxil edilmiş mətnlərin xüsusiyyətləri barədə əlavə məlumatlar verilir (xüsusi markerləmə və ya annotasiyalarının verilməsi). Mətnlərin markerlənməsi korpusun əsas səciyyələrindəndir. Bu, korpus mətnlərini sadə mətn kolleksiyalarından və ya “kitabxanalar”dan fərqləndirir.

RDMK ilə yanaşı Rusiyada “Fundamental elektron kitabxana” da yaradılır. Burada rus klassik ədəbiyyatı nüfuzlu çap nəşrlərində olduğu kimi, heç bir dəyişikliyə uğramadan elektron formada verilir. Elektron istifadə üçün işlənməmiş belə kitabxanalar elmi tədqiqat işləri üçün yararlı deyil. Nəzərə almaq lazımdır ki, kitabxanalar onun üçün yaradılmır ki, oradakı mətnlər üzərində dilçilər tədqiqat aparsın, mətnin dil xüsusiyyətlərini araşdırsın. Həmin klassik ədəbiyyat onu istifadəçilərin oxuması üçün qoyulur. Ona görə də, internetdə qoyulmuş kitabxanalar qarşıya qoyduqları məqsədə müvafiq qaydada kiber məkanda təmsil olunurlar. Milli korpus elektron kitabxanadan fərqli olaraq “maraqlı” və ya “faydalı” mətnləri toplamaqla məşğul olmur. Onu elə mətnlər maraqlandırır ki, həmin mətnlər dilin öyrənilməsi və tədqiqi baxımından maraqlı və faydalı olsun. Beləliklə, milli korpus üçün yazıçının şöhrəti, kitabın maraqlı və ya faydalı olması, yüksək bədii və ya elmi əhəmiyyətə malik olması mühümdür, lakin birinci dərəcəli deyil. Dilin öyrənilməsi baxımından isə ikinci dərəcəli bir yazıçının romanı, adi telefon danışığı, icarə müqaviləsi və s. də, klassik bədii ədəbiyyat nümunələri ilə yanaşı əhəmiyyətli ola bilər (<http://www.ruscorpora.ru/corpora-intro.html>).

Mətnlərin elektron variantları nə qədər çox işlənərsə, mətnlərə nə qədər çox nişan qoyularsa, milli korpusun elmi və tədris baxımından əhəmiyyəti bir o qədər yüksəlmiş olar. RDMK-də hazırda daha çox beş işarələnmə növündən istifadə olunur: meta-mətn, morfoloji (sözdəyişmə), sintaktik, aksent və semantik.

Qeyd etmək lazımdır ki, korpusda markerləmə sistemi getdikcə təkmilləşdirilir, yeni formalar və optimal üsullar təklif olunur.

Milli korpus ilk növbədə dilin leksika və qrammatikası sahəsində tədqiqatlara istiqamətlənmişdir. Eyni zamanda müəyyən zaman kəsiklərində (100-200 il ərzində) dildə baş verən fasiləsiz dəyişikliklərin izlənilməsi və təhlili də korpus materialının köməyi ilə hərtərəfli araşdırıla bilər. Korpusun daha bir vəzifəsi dilin leksika, qrammatika, aksentologiya tarixi və s. barədə mümkün arayışlar verməkdir. Müasir kompüter texnologiyası böyük mətn massivlərinin linqvistik işlənmə prosedurunun dəfələrlə sadələşdirir və sürətləndirir. Əvvəllər tədqiqatçı mətnə baxır və oradan lazımi informasiyaya aid nümunələri əllə köçürüb kağıza yazırdı. Mətnlərin bu cür ilkin işlənməsi ağır fəaliyyət növü olub, çox vaxt aparırdı. Həm də bu üsulla böyük mətn massivlərini işləmək qeyri-mümkün idi. İndi təhlil olunan mətnin həcminə heç bir məhdudiyət yoxdur. İnformasiyanın axtarış sürəti də kifayət qədər böyükdür. Tədqiqatçı istənilən mətn tipini istənilən həcmdə qısa müddət ərzində təhlil etmək imkanına malikdir. Bütün bunlar bizim dil haqqında biliklərimizin getdikcə dərinləşməsi, çoxalması üçün geniş üfüqlər açır. Mətnlərin kütləvi şəkildə (o cümlədən statistik cəhətdən) işlənməsi dilin quruluşu və inkişafı haqqında əvvəllər heç düşünmədiyimiz, yaxud ehtimal etdiyimiz bir çox qanunauyğunluqların aşkarlanması və əsaslandırılması üçün təkzibolunmaz faktlar verir. Dillərin qrammatik quruluşunun elmi təsviri, eləcə də nüfuzlu akademik lüğətlər yalnız milli dil korpuslarına əsaslanmaqla hazırlanmalıdır. Başqa xüsusi tədqiqat sahələrində də korpus məlumatlarının nəzərə alınması faydalı olardı.

Milli dil korpuslarının əsas istifadəçiləri, əlbəttə ki, dilçi tədqiqatçılardır. Lakin istifadəçiləri təkcə dil tədqiqatçıları ilə məşğul olanlarla məhdudlaşdırmaq olmaz. Müəyyən tarixi dövrün və ya konkret müəllifin dili ilə bağlı etibarlı statistik məlumatlar ədəbiyyatçıları, tarixçiləri və humanitar elm sahəsinin başqa nümayəndələrini də maraqlandıra bilər. Ana dilinin və ya xarici dillərin tədrisi prosesi üçün də milli korpuslar olduqca əhəmiyyətlidir. Son zamanlar dərslük və dərs vəsaitləri daha çox korpus materiallarına əsaslanmaqla hazırlanır. Tanış olmayan sözün və ya qrammatik formanın nüfuzlu müəlliflər tərəfindən istifadə olunma xüsusiyyətlərini xarici dil daşıyıcıları, məktəblilər, müəllimlər, jurnalistlər, redaktorlar, yazıçılar, bir sözlə, bütün maraqlananlar çox qısa vaxt ərzində effektiv olaraq milli korpusun köməyi ilə yoxlayıb təhlil edə bilərlər.

RDMK XVIII əsrin ortalarından müasir dövrə qədərki vaxtı əhatə edir. Həmin dövr ərzində mövcud olmuş bütün dil xüsusiyyətləri milli korpusda öz əksini tapmışdır. Korpus dil və mədəniyyət baxımından böyük maraq doğuran orijinal (tərcümə ədəbiyyatı bura daxil edilmir) bədii ədəbiyyat nümunələrini əhatə edir. Lakin milli korpus təkcə bədii ədəbiyyat korpusu deyil. Bədii mətnlərlə yanaşı, bura yazılı və şifahi dil nümunələri – memuarlar, esselər, publisistika, elmi-populyar və elmi ədəbiyyat, çıxışlar, şəxsi yazışmalar, gündəliklər, sənədlər və s. daxil edilir.

RDMK-da məlumatlar müəyyən sistemlə yerləşdirilmiş və ayrı-ayrı mətnlər sıx əlaqələndirilmişdir.

2.2.3.1. Əsas mətn korpusu

Rus ədəbi dilini təmsil edən mətnləri iki əsas massive ayırmaq olar:

- müasir yazılı mətnlər (XX əsrin ortalarından – XXI əsrin əvvəllərində);
- daha əvvəlki dövrlərə aid mətnlər (XVIII əsrin ortalarından – XX əsrin ortalarındak);

Hər iki massivə eyni zamanda müraciət etmək olar. Əsas korpusa daxil olunan bütün mətnlərdə metamarkerləmə və morfoloji markerləmə aparılır.

Morfoloji markerləmə xüsusi avtomatik morfoloji təhlil proqramlarının köməyi ilə aparılır. Əsas korpusun kiçik hissəsində (6 milyon sözləşmə; gələcəkdə bu işin artırılması nəzərdə tutulur) omonimliyin əllə aradan qaldırılması və avtomatik morfoloji təhlilin əlavə olaraq korreksiyası aparılmışdır. Bu hissə omonimliyin aradan qaldırılması korpusu adlanır və yüksək axtarış dəqiqliyi tələb edən müasir rus morfoloqiyasının tədqiqi, mətnlərin avtomatik işlənməsi və formal morfoloji təhlili, müxtəlif axtarış proqramlarının testləşdirilməsi üçün əlverişli poliqon rolunu oynayır. Bu altkorpusun nümunələri müvafiq markerlə təchiz olunur (omonimlik aradan qaldırılmışdır). Omonimliyi aradan qaldırılmış mətnlər avtomatik olaraq vurğulanmışlar.

2.2.3.2. Müasir yazılı mətnlər

Morfoloji markerlə təchiz olunmuş müasir mətnlər korpusu altkorpusların ən irihəcmli və əsaslarından sayılır. Bu korpus müasir rus ədəbi (yazılı) dilinin müxtəlif tiplərini əhatə edir:

- müxtəlif istiqamətli və janrlı müasir bədii nəsr nümunələri;
- müasir dramaturgiya;
- memuar-bioqrafik ədəbiyyat;
- jurnal publisistikası və ədəbi tənqid;
- qəzet publisistikası və xəbərlər;
- elmi, elmi-populyar və tədris mətnləri;
- dini və dini-fəlsəfi mətnlər;
- istehsalatla bağlı texniki mətnlər;
- rəsmi-işgüzar və hüquqi mətnlər;
- məişət mətnləri (o cümlədən çap üçün nəzərdə tutulmayan mətnlər: şəxsi yazışmalar, gündəliklər və s.).

Bədii mətnlər (dramaturgiya və memuarlar daxil olmaqla) korpusun 40%-ni təşkil edir.

Korpusa daxil edilən mənbələr bu mətnlərin nəşirləri tərəfindən təsdiq olunmuş kitab, jurnal və qəzet məhsulları hesab olunur.

Nəşirlərin təsdiq etmədiyi və icazə vermədiyi mətnlər korpusa daxil edilmir.

XVIII əsr və XX əsrin ortalarına aid olan mətnlər onların elmi nəşrlərində saxlanılan orfoqrafiyada – orijinal orfoqrafiyada verilir.

2.2.3.3. Dərin annotasiyalı (sintaktik) korpus

RDMK-nın bu bölümü morfo-sintaktik markerlə təchiz olunmuş mətnlərdən ibarətdir. Bu o deməkdir ki, morfoloji informasiya ilə təchiz olunmuş hər bir sözlə yanaşı, hər bir cümlə üçün də onun sintaktik strukturu verilmişdir. Dərin annotasiyalı korpusda cümlənin sintaktik quruluşu asılılıqlar ağacı kimi təsvir olunur. Onun qovşaqlarında cümləni təşkil edən sözlər, budaqlarında işə sintaktik əlaqələr göstərilir. Cümlənin sintaktik strukturunun bu cür təsviri “Məna↔Mətn” linqvistik modelinə uyğun gəlir.

Morfoloji markerlə təchiz olunmuş fraqmentlərdən fərqli olaraq burada morfolji və sintaktik omonimlik aradan götürülmüşdür.

2.2.3.4. Qəzet korpusu (müasir KİV korpusu)

RDMK-da qəzet korpusu (müasir kütləvi informasiya vasitələri korpusu) 2010-cu ildə açılmışdır və həmin bölmədə 2000-ci illərdə kütləvi informasiya vasitələrində dərc olunmuş məqalələr əhatə olunur. Real dövrdə dildə baş verən dəyişiklikləri əks etdirən və böyük maraq doğuran bütün məqalələr korpusa daxil edilə bilməz.

Qəzet mətnləri korpusuna kütləvi informasiya vasitələrindən – “Известия”, “Советский спорт”, “Труд”, “Комсомольская правда” qəzetlərindən, eləcə də elektron agentliklərdən (РИА “Новости”, РБК, “Новый регион”) götürülmüş

mətnlər daxil edilmişdir. Korpus daim təzələnir, hər il 10 milyonlarla sözişlətmədən (söz-formadan) ibarət bərabər həcmdə mətnlər əlavə olunur.

2.2.3.5. Paralel mətnlər korpusu

Bu korpusun özəlliyi ondadır ki, rus dilində olan mətn onun başqa dilə tərcüməsi ilə qarşılaşdırılır. Bəzən əksinə də olur, başqa dillərdə olan mətn onun rus dilinə tərcüməsi ilə paralel verilir. Orijinal və tərcümə mətni (adətən cümlələr) arasında xüsusi prosedurlar vasitəsilə müvafiqlik müəyyənləşdirilir. Bu prosedur müvafiqliklə "cilalanma" adlanır, mətnlər isə dəqiq-ləşdirilmiş "cilalanmış" hesab olunur.

Paralel korpuslar tərcümə nəzəriyyəsi və praktikasını baxımından əhəmiyyətli və belə paralel mətnlərdən rus və xarici dillərin tədrisi prosesində istifadə oluna bilər.

Hazırda RDMK-da aşağıda qeyd olunan dillər üçün paralel korpuslar yerləşdirilmişdir: ingilis-rus, rus-ingilis, alman-rus, rus-alman, fransız-rus, rus-fransız, ispan-rus, rus-İspan, İtalyan-rus, rus-İtalyan, polyak-rus, rus-polyak, ukrayna-rus, rus-ukrayna, belarus-rus, rus-belarus, eləcə də çoxdilli mətnlər.

2.2.3.6. Dialekt mətnləri korpusu

Dialekt mətnləri korpusu 2005-ci ildən açıqdır. Standart orfoqrafiyaya qismən uyğunlaşdırılmış, Rusiyanın müxtəlif regionlarından toplanmış dialekt nitqi bu korpusda təmsil olunmuşdur. Mətnlərin morfoloji, sintaktik və leksik xüsusiyyətləri olduğu kimi saxlanmışdır. Altkorpusda dialekt morfoloqiyasının ədəbi dildən fərqli xüsusiyyətlərini əks etdirən markerlərdən istifadə olunmuşdur. Xalis dialekt leksikası xüsusi izahla təchiz edilmişdir.

2.2.3.7. Poetik mətnlər korpusu

Poetik mətnlər korpusu nəzmlə yazılmış əsərləri ehtiva edir və 2006-cı ildə açılmışdır. Nəzmlə yazılan dramatik əsərlər hələ

korpusa daxil edilməmişdir. Adi semantik və morfoloji markerlərlə yanaşı, korpusda xüsusi şeir markerləri qoyulmuşdur. Müxtəlif şeir formalarını həmin markerlə axtarıb tapmaq olar.

2.2.3.8. Rus dilinin tədris korpusu

Rus dilinin tədris korpusu 2007-ci ildə açılmışdır. Məktəblərdə rus dilinin tədrisi ilə əlaqədar çox da böyük olmayan bu korpusda omonimliyin ləğvi markerləri qoyulmuşdur. Mətnlər məktəb proqramlarından seçilib götürülmüşdür. Məktəbdə tədris olunan çərçivədə funksional üslublar, sözdəyişdirmə tipləri və əlavə morfoloji əlamətlər markerlənmişdir.

2.2.3.9. Şifahi nitq korpusu

Şifahi nitq korpusu RDMK-da 2007-ci ildən mövcuddur. Bu altkorpusda maqnitofon yazıları, çıxışlar, şəxsi söhbətlər, kinofilmlərin transkriptləri yerləşdirilmişdir. Rus standart orfoqrafiyasının qaydaları gözlənilmiş, ən çox işlənən və hamı tərəfindən qəbul olunmuş şifahi danışq normalarına əməl olunması təmin olunmuşdur. Sosioloji parametrlər üzrə leksik, morfoloji və semantik axtarış aparmaq imkanı yaradılmışdır. Müxtəlif janrlı mətnlər korpusu cəlb edilmişdir. Rusiyanın bütün regionlarının şifahi danışq xüsusiyyətlərinin nəzərə alınması nəzərdə tutulmuşdur (Moskva, Sankt-Peterburq, Saratov, Ulyanovsk, Taqanroq, Yekaterinburq, Norilsk, Voronej, Novosibirsk və b.). Korpus 1900-2000-ci illəri əhatə edir.

2.2.3.10. Aksentoloji korpus

Aksentoloji korpus (rus vurğusunun tarixi korpusu) 2008-ci ildən fəaliyyətdədir. Rus vurğusunun tarixi haqqında məlumatları əhatə edir. Bu mətnlərin əksəriyyəti poetik korpusdan götürülmüşdür. Həmin mətnlərdə şeir vəznələrinə uyğun olaraq sözlərdə vurğunun yeri barədə məlumat verilmişdir. Şifahi nitq parçalarında, eləcə də kinofilmlərdə, real səslənən nitq axınında

vurğunun yeri hiss olunur. Bu mətnlərdə vurğunun yeri və sözün prosodik quruluşuna görə axtarış aparmaq mümkündür.

2.2.3.11. Multimediya korpusu

Multimediya korpusu 2010-cu ildə istifadəyə verilmişdir. Buraya 1930-2000-ci illərdə istehsal olunmuş filmlər daxildir. Multimediya korpusunda səslənən nitq parçalarının şifrəsi açılmış, kadrda müşahidə olunan görüntü, səslənmə və jestlər paralel olaraq yerləşdirilmişdir. Bu korpusda axtarış təkcə tələffüz olunan mətnə görə deyil, həm də jestlərə (baş işarəsi, çiyinə əl vurma və s.) və nitq hadisəsi tipinə (razılıq, şübhə, lağa qoyma, kinayə, ələsalma və s.) görə aparıla bilər. Axtarışdakı sorğuya əsasən, videogörüntülərə və səslənməyə fraqmentlər şəklində baxış keçirmək mümkündür.

RDMK daim təzələnen, tamamlanan açıq sistemdir. Ona vaxtaşırı əlavələr edilir, yeni materiallar, məlumatlar, axtarış sistemləri daxil olunur.

İnformasiya texnologiyaları, kompüterlər, internet sistemləri təkmilləşdikcə, korpus da yenilənir, təkmilləşir, daha çevik və istifadə üçün yararlı olur.

2.3. Azərbaycan dilinin milli korpusunun yaradılmasının nəzəri və təcrübi məsələləri, strukturu

2.3.1. Məsələnin qoyuluşu

Milli dillərin korpuslarının yaradılması sahəsində toplanmış təcrübə Azərbaycan dilinin milli korpusunun hazırlanması işlərinin intensivləşdirilməsi zərurətini yaratmışdır.

“Azərbaycan dilinin qloballaşma şəraitində zamanın tələblərinə uyğun istifadəsinə və ölkədə dilçiliyin inkişafına dair Dövlət Proqramı”nda da Azərbaycan dilinin dövlət dili kimi ümumi informasiya bazasının yaradılması, elektron tezlik, terminoloji lüğətlərin hazırlanması, Azərbaycan dilində internet resurslarının, elektron və interaktiv dərsliklərin yaradılması istiqamətində

fəaliyyətin gücləndirilməsi, əlyazmaların elektron portalının yaradılması, Azərbaycan dilinin daha geniş istifadəsinin təmin edilməsi üçün internet texnologiyasının, maşın tərcüməsi sistemlərinin və digər müasir tətbiqi-linqvistik texnologiyaların yaradılması və inkişaf etdirilməsi kimi məsələlər həlli vacib prioritet istiqamətlər kimi xüsusi göstərilmişdir. Bütün bunlar Azərbaycan dilinin milli korpusunun yaradılmasının hüquqi bazasını təşkil edir və Azərbaycan dilinə dövlət səviyyəsində qayğının parlaq nümunəsidir.

Dövlət Proqramının icrasına dair tədbirlər planında həmin sadalanan işlərin 2013-2020-ci illər ərzində yerinə yetirilməsi nəzərdə tutulmuşdur. Bütün bunlar dilçiləri bir daha milli korpusun hazırlanması işlərinə səfərbər edir.

Qeyd etmək lazımdır ki, milli korpusun yaradılması ideyası ötən əsrin 80-ci illərində yaranmışdır. SSRİ xalqlarının dillərinin milli korpuslarının yaradılması işlərinə də elə həmin vaxtlardan hazırlıq görülürdü. Azərbaycan dilinin milli korpusu sahəsində hazırlıq işləri türk dillərinin milli korpusu çərçivəsində aparılırdı. Bi işləri Azərbaycan dilinin milli korpusunun yaradılmasının başlanğıc mərhələsi hesab etmək olar.

2.3.2. Azərbaycan dilinin milli korpusunun nəzəri və təcrübi bazası

Ötən əsrin 80-ci illərindən SSRİ xalqlarının dillərinin milli korpuslarının yaradılması ilə bağlı vaxtaşırı olaraq müşavirələr keçirilirdi (Тбилиси, 1987; Таллин, 1988 və s.). Həmin müşavirələrdə korpusun quruluşu, iş prinsipi, vəziyyəti barədə operativ müzakirələr aparılırdı. Sonralar türk dillərinin milli korpusu ilə bağlı da müşavirələr keçirilməyə başladı. “Советская тюркология” jurnalında türk dilləri üçün milli korpusun yaradılmasının prinsipləri açıqlandı (Гузев, Пиотровский, Щербак, 1988, s.92-101).

Bütün bunlar Azərbaycan dilinin milli korpusunun yaradılması üçün də perspektivlər açırdı.

Milli korpusların yaradılmasının nəzəri bazası hesab olunan formal təhlil sistemləri, formal morfoloji, sintaktik, semantik

təhlil və sintez proqramlarının işlənilib hazırlanmasıdır. Formal təhlil və sintez sistemləri milli dil korpuslarının ən mühüm komponentləri hesab olunur. Ona görə də, korpusun hazırlanması işlərinə onlardan başlamaq lazım idi. Yalnız mükəmməl və hərtərəfli formal təhlil bazası əsasında sistemin digər komponentlərini qurmaq olardı. Xüsusilə, morfoloji, sintaktik və semantik səviyələrdə omonimliyin aradan qaldırılması problemi milli dil korpuslarının yaradılmasında əsas maneələrdən hesab olunur.

Qeyd etmək lazımdır ki, Azərbaycan dilinin bütün səviyələrində (morfoloji, sintaktik, semantik) formal təhlil alqoritmləri hazırlanmış, morfoloji və sintaktik səviyələrdə omonimliyin aradan qaldırılması yolları və üsulları işlənilib hazırlanmışdır. Monoqrafiyanın sonrakı fəsilələrində sadalananlar hərtərəfli və geniş şəkildə izah olunduğu üçün bu məsələlər üzərində dayanmayacağıq. Həmin baza əsasında milli korpusun qalan komponentlərini hazırlamaq və əlaqələndirmək mümkündür.

2.3.3. Azərbaycan dilinin milli korpusunun strukturu

Korpusa daxil edilən mətnlər Azərbaycan dilinin bütün üslublarını eyni səviyyədə və həcmdə ehtiva etməlidir. Korpusun həcmi nə qədər böyük olursa, korpus bir o qədər etibarlı və əhatəli hesab olunur. Müasir informasiya texnologiyaları milli dil korpuslarının həcmi istənilən qədər artırmağa imkan verir. Əlbəttə, burada optimal yerləşdirmə üsullarından istifadə etmək daha faydalı olardı. Həcm həddindən artıq böyük ola bilər, amma dil lazımi səviyyədə təmsil olunmaz. Məsələn, bütün korpusu qəzet, jurnal materialı ilə doldurmaq olar, dilin başqa üslubları və sahələri isə korpusdan kənar qala bilər. Ona görə də, korpusun strukturunu düzgün planlaşdırmaqdan çox şey asılıdır. Azərbaycan dilinin milli korpusuna aşağıdakı mətnlərin daxil edilməsi məqsədəuyğun olardı:

- folklor mətnləri (Azərbaycan şifahi xalq ədəbiyyatı nümunələri – dastanlar, nağıllar, bayatılar, tapmacalar, atalar sözləri, lətifələr, xalq mahnıları və s.);
- klassik Azərbaycan ədəbiyyatı;

- müasir Azərbaycan ədəbiyyatı (nəzm, nəsr, dramaturgiya);
- publisistika (qəzet, jurnal, televiziya, radio, informasiya agentlikləri, internet resursları və s.);
- elmi-texniki ədəbiyyat (elmi əsərlər, dissertasiyalar, avtoreferatlar, elmi məcmuələr, elmi jurnallar, konfrans materialları, dərsliklər, dərs vəsaitləri və s.);
- dini-fəlsəfi mətnlər;
- rəsmi-işgüzar üsluba aid mətnlər;
- məişət mətnləri;
- danışiq dili və dialekt mətnləri.

Qeyd etmək lazımdır ki, burada göstərilənlərdən əlavə də korpusa yeni mətnlər, müxtəlif sahələrə aid materiallar daxil edilə bilər. Başqa sözlə, Azərbaycan dilində yazılmış, Azərbaycan dilinə aid istənilən mətn dili təmsil edirsə, korpusa əlavə oluna bilər. Azərbaycan dilinin milli korpusu aşağıda göstərilən altkorpuslardan təşkil oluna bilər:

- mətnlər altkorpusu;
- linqvistik təhlil altkorpusu;
- kütləvi informasiya vasitələri altkorpusu;
- paralel mətnlər altkorpusu;
- dialekt mətnləri altkorpusu;
- poetik mətnlər altkorpusu;
- tədris altkorpusu;
- şifahi nitq altkorpusu;
- aksentologiya altkorpusu;
- multimediyaya altkorpusu.

Tədqiqat prosesində korpusa yeni altkorpuslar əlavə edilə bilər və ya bəzi altkorpuslar ləğv oluna bilər. Milli korpus açıq sistemdir və onun strukturunda və funksiyalarında istənilən vaxtda zərurət yarandıqda dəyişiklik aparmaq olar.

Milli korpusun altkorpuslarının hamısı eyni vaxtda fəaliyyət göstərmir. Altkorpuslar formalaşdıqca, onların təminat bazaları yarandıqca fəaliyyətləri də başlanır. İlk vaxtlarda korpusda

bir neçə altkorpus funksional olur. Onların qarşılıqlı əlaqələri, tələbatı, yeni altkorpusların yaranması üçün stimula ola bilər.

Məlum olduğu kimi, hazırda respublikamızda elektron kitabxana xidməti fəaliyyət göstərməkdədir. Milli kitabxanada, Azərbaycan Milli Elmlər Akademiyasının kitabxanasında, bir çox sahə kitabxanalarında elmi-tədqiqat müəssisələrində, ali məktəblərdə nəşr olunan ədəbiyyatın elektron variantı yaradılır və internetdə yerləşdirilir. Bütün bunlar milli korpusda mətn massivi kimi istifadə oluna bilər. Lakin kitabların elektron variantları internetdə ilkin işlənməsiz, çap olunduğu kimi qoyulur. Bu da onların müxtəlif məqsədlərlə istifadəsini çətinləşdirir. Milli korpus mətnləri isə birbaşa deyil, müəyyən işlənmə prosesindən (lingvistik işlənmə, markerlənmə və s.) sonra sistemə daxil edilə bilər. Bu barədə əvvəlki səhifələrdə məlumat verildiyi üçün təkrar şərhə ehtiyac yoxdur. İnternetdə yerləşdirilmiş bu ədəbiyyatın ilkin işlənmə prosesinin üsul və formaları barədə düşünməyə dəyər.

III FƏSİL. KOMPÜTER LÜĞƏTÇİLİYİ

3.1. Lüğətçiliyin inkişafında yeni mərhələ

Dilçiliyin ən qədim və ən çox istifadə olunan, qeyri-dilçilərin də gündəlik tələbatına çevrilmiş müxtəlif çeşidli lüğətlərin tərtibi məsələləri ilə məşğul olan sahəsi – lüğətçilik (leksikoqrafiya) yeni inkişaf mərhələsinə qədəm qoymuşdur. Bu mərhələnin aparıcı xüsusiyyəti informasiya texnologiyalarının yaradılması və inkişafı, kompüterlərin lüğətçilikdə istifadəsi ilə səciyyələnir.

İnsan həyatında, cəmiyyətdə, dillərarası, mədəniyyətlərarası dialoqda lüğətlərin rolu çox böyükdür. Bəlkə də bunun nəticəsidir ki, lüğətə və lüğətçiliyə maraq və tələb azalmır, əksinə getdikcə artır, çoxalır. Məsələ orasındadır ki, ayrı-ayrı xalqlar, ölkələr arasında əlaqələr də getdikcə artır, intensivləşir. Bu əlaqələrdə lüğətlər körpü, ünsiyyət rolu oynayır. Bütün dünya ölkələrində lüğətçiliyə, çeşidli lüğətlərin tərtibinə belə önəm verilməsinin başlıca səbəblərindən biri, bəlkə də ən mühümünü bu deyimlərlə əlaqələndirmək olar.

Lüğətçilik böyük səbir, zəhmət, səriştə, bilik tələb edən bir sahədir. Burada işin elmi cəhəti ilə yanaşı “texniki” cəhətləri də var. Kartoçkalara yığılmış lüğət vahidləri (sözlər, söz-formalar, söz birləşmələri və s.) sistemləşdirilməli, qruplaşdırılmalı və əlifbaya düzülməlidir. Əlifbaya düzülmə təkcə lüğət vahidinin birinci hərfinə görə deyil, sözü təşkil edən bütün hərflərə görə aparılır. Bu isə olduqca ağır, yorucu bir fəaliyyət növüdür. Bəzən işin sırf texniki cəhəti, yəni əlifbaya düzülmə sözlüyün toplanması və izahından da çox vaxt aparır. Nəticədə həcmcə çox da böyük olmayan bir lüğətin əlifbaya düzülməsi, aylarla, hətta illərlə vaxt aparır. Lüğət çap olunmamışdan köhnəlir, aktuallığını itirir, onun yenidən tərtibinə, yeniləşməsinə, tamamlanmasına ehtiyac yaranır.

Kompüter texnologiyası, müasir kompüterlər bu mərhələdə lüğətçilərə böyük kömək göstərə bilər. Sözlüyün qruplaşdırılması, sistemləşdirilməsi və əlifba ardıcılığı ilə düzülməsi işləri kompüterlərdə qısa zaman kəsiyində yerinə yetirilə bilər. Lüğətin

həcmnin böyüklüyü, sözlərin çoxluğu əvvəlki vaxtlarda olduğu kimi tərtibçini çox narahat etməməlidir. Tərtibçi bir lüğətçi kimi öz işini keyfiyyətlə yerinə yetirməlidir (lüğətin sözlüyünün düzgün müəyyənləşdirilməsi, sözlərin tərcüməsinin və izahının düzgün verilməsi və s.).

Məsələnin başqa bir tərəfi həmin lüğətlərin internet məkanında yerləşdirilməsidir. Deməli, kompüterlər iki məsələdə lüğətçilərə yardım etmiş olur: lüğətlərin tərtibinin sürətləndirilməsi (avtomatlaşdırılması) və onların qlobal şəbəkələrdə yerləşdirilməsi. Burada bir məsələni dəqiq izah etmək lazımdır. Elektron məkana qoyulan hər hansı lüğət elektron lüğət deyil. Lüğət çap olunduğu şəkildə, yəni kitab formatında internetdə yerləşdirilə bilər. Həmin kitaba müraciət adı lüğətin kitab kimi ekranda oxunmasıdır. Heç bir ilkin işlənmə, sorğuya görə markerlənmə və s. aparılmadığı üçün müxtəlif məqsədlərlə ona müraciətlər də yox dərəcəsində olur.

Elektron lüğətlərin quruluşu, yerləşdirilməsi, onların sorğulara görə məlumat vermək imkanları ilkin işlənmə tələb edir. Bu barədə sonra ətraflı məlumat veriləcəkdir.

Adi lüğətlərdə (kağız variantda), internetdə yerləşdirilmiş lüğətin (heç bir işlənmə olmadan internetdə yerləşdirilmiş variant) və elektron lüğətlərin oxşar və fərqli cəhətlərinin öyrənilməsi gələcəkdə bu sahədə görülmək işlərin istiqamətinin müəyyənləşdirilməsində əhəmiyyəti ola bilər. Bunun üçün kağız çap variantında və elektron variantda olan lüğətlərin quruluşu ilə tanış olmaq, onların səciyyəvi cəhətlərini aydınlaşdırmaq lazımdır. Digər tərəfdən, müasir dövrdə hansı lüğət növünə tələbatın daha çox və ya az olduğunu öyrənmək və lüğətçilik sahəsində işlərin təşkilini ona uyğun qurmaq lazımdır. Lüğətçiliyin inkişafının yeni mərhələsi həm də lüğət və onun tərtibi məsələlərinə yeni baxışın yaranması və formalaşması ilə səciyyələnir.

3.2. Kompüter lüğətçiliyi anlayışı

Lüğətçilikdə kompüterlərdən istifadə olunması müxtəlif cür izah olunur. Bəziləri kompüterin burada yardımçı bir vasitə olduğunu, alət funksiyası daşdığını iddia edir, onun lüğətçilik

üçün önəmli olmadığını göstərir. Kompüterlərin lüğətlərin tərtibindəki rolunu şişirdənlər, onların inqilab yaradacağı fikrinin də tərəfdarları az deyil. Məncə, məsələyə obyektiv yanaşmaq, lüğətçiliyin inkişafındakı yeni mərhələni elmi cəhətdən düzgün qiymətləndirmək lazımdır.

Qeyd etmək lazımdır ki, lüğətlərin tərtibində kompüterlərdən istifadə olunması məsələləri dilçiliyin müxtəlif sahələrində öyrənilir: kompüter lüğətçiliyi, hesablama lüğətçiliyi, maşın lüğətçiliyi, avtomatik (avtomatlaşdırılmış) leksikoqrafiya (Рублёва, 2010, s.32), [http://www./linevoda.ru](http://www.linevoda.ru); www.tech-tr.ru).

“Kompüter dilçiliyinin əsasları” kitabının müəllifi Y.N.Marçuk bu terminlərin hamısının kompüter dilçiliyinin predmeti olduğunu göstərir və bu termin altında onların birləşdirilməsini məqsəduyğun hesab edir (Marçuk, 2002, s.4). Fikrimizcə, *hesablama lüğətçiliyi, maşın lüğətçiliyi, elektron lüğətçilik, avtomatik (avtomatlaşdırılmış) lüğətçilik* terminlərini konkret bir terminlə əvəz etməklə termin müxtəlifliyindən və dolaşılığından qurtulmaq olar.

Kompüter lüğətçiliyi kompüter dilçiliyinin bir sahəsi kimi həmin kontekstdə öyrənilməlidir. Bu termin vaxtilə dəbdə olan və çox işlənən *statistik leksikoqrafiya* terminini də ehtiva edir.

Monoqrafiyada kompüter lüğətçiliyinin ayrıca bir fəsil kimi verilməsi və bu tipli məsələlərin həmin fəsildə araşdırılması məqsəduyğun hesab olunmuşdur. Əslində, yuxarıda sadalanan bu terminlər (hesablama lüğətçiliyi, maşın lüğətçiliyi), avtomatik lüğətçilik, statistik lüğətçilik) sinonim terminləridir və onların tədqiqat obyektini kompüterlər vasitəsi ilə tərtib olunan, elektron məkanda yerləşdirilən, müxtəlif sorğu və məlumat elementləri ilə təchiz olunmuş lüğətlərdir. Kompüter lüğətçiliyi kompüter dilçiliyinin bir sahəsi, kompüter lüğətləri isə bu sahənin obyektidir.

Kompüter lüğətləri adi lüğətlərdən aşağıdakı səciyyələrinə görə fərqlənir:

- lüğətə müraciət və ondan müəyyən sorğular əsasında məlumatların alınması proseduru daha sadə və sürətlidir;
- eyni zamanda bir neçə lüğətə müraciət edib, sözün mənasının qısa zamanda dəqiqləşdirilməsi və ümumi nəticənin

- çıxarılması mümkündür. Adi lüğətlərdə bu işin yerinə yetirilməsi üçün xeyli vaxt və zəhmət tələb olunardı;
- bu lüğətlərin hər birində hər hansı bir dili ön plana çəkmək və başqa lüğətlərdə müqayisə aparmaq imkanı var. Məsələn, 8-dilli tərcümə lüğətində hər hansı bir dil əsas götürülə bilər. Yəni çoxdilli lüğətlərdə təmsil olunmuş bütün dillər eyni hüquqa və axtarış imkanlarına malikdir;
 - adi lüğətlər çap olunduqdan sonra təkrar nəşrə qədər dəyişməz qalır, sözlərin sayı və izahı dəyişdirilə bilməz. Kompüter lüğətləri açıq, hərəkətli sistemlərdir. Belə lüğətlərə yeni sözləri anı olaraq əlavə etmək, düzəliş aparmaq və köhnəlmiş sözləri çıxarmaq (və ya həmin sözlər ayrıca qrupda yerləşdirmək) mümkündür.

Kompüter lüğətlərinin əksəriyyəti Geniş Ümumdünya Şəbəkəsində (*World Wide WEB*, qısaltma variantı – *WWW*) yerləşdirilmişdir. Bu, ümumdünya kompüter şəbəkəsi – internetin populyar elektron resurslarından hesab olunur. Vəbdə bu qəbildən olan əksər sənədlər kimi lüğətlər HTML (HyperText Markup Language – HiperMətnin Markerləmə Dili) formatında yaradılmışdır. Belə format lüğətlərlə işi sadələşdirir və onların effektivliyini artırır (Рублёва, 2010, s.42).

Belə formatda yerləşdirilmiş lüğətlərdə hər bir lüğət məqaləsinin başlığı (sərlövhəsi) eyni zamanda istinad funksiyasını yerinə yetirir və həmin söz başqa yerdə təkrar olunursa, istifadəçi dərhal bu barədə onu maraqlandıran məlumatı əldə edə bilər. Beləliklə, hipermətn formatında yaradılmış lüğətlərdə bütün lüğət məqalələrinin başlıqları birbaşa öz aralarında əlaqələndirilir. Yeni tərtib olunan hipermətn formatlı lüğətlərdə sinonim, antonim, omonimlərə hipermətn istinadlar edilməsi və paralel olaraq bir neçə lüğətə eyni zamanda baxış keçirmək imkanı reallaşdırılmışdır.

3.3. Kompüter lüğətlərinin kağız (kitab) lüğətlərdən üstünlükləri

Yuxarıda qeyd etmişdik ki, kompüter lüğətlərinin hipermətn quruluşu və buna müvafiq hipermətn istinadlar belə lüğət-

lərin kağız lüğətlərdən üstün cəhətlərini və optimal istifadə olunma imkanlarını səciyyələndirən mühüm xüsusiyyətlərdir. Bunlardan əlavə, kompüter lüğətlərinin üstün cəhətlərinə aşağıdakıları aid etmək olar:

- həcmi qeyri-məhdudluğu;
- həcmi çox olmasına baxmayaraq zəruri informasiyanın axtarış sürətinin çox böyük olması;
- istinadlar vasitəsilə digər lüğətlərdən əlavə məlumatların əldə edilə bilməsi imkanı;
- lüğətdəki məlumatların yeniliyi və müasir həyatla bağlılığı, yeni söz və ifadələrin dəqiq izahının olması;
- lüğət məqalələri daim təzələndiyi və tamamlandığı üçün köhnəlməməsi və uzunömürlüü;
- irihəcmli olmasına baxmayaraq, eyni zamanda yığcamlığı.

Kompüter lüğətlərində kağız lüğətlərdə olmayan əlavə informasiyalar da yerləşdirilə bilər. Məsələn, sözün ifadə etdiyi əşya və ya məfhumun şəkli, sözün səslənmə variantları və s. (Селегей, [http://www./lineveda.ru/-www.tech-tr.ru](http://www.lineveda.ru/-www.tech-tr.ru); Вадяев, 2003, c.83-92; Карпова, 2001, c.54-64).

Gələcəkdə bu tipli lüğətlərin daha da təkmilləşdirilməsi və genişləndirilməsi imkanları da böyükdür. Belə ki, şəkil və səslənmədən əlavə belə lüğətlərdə sözün bütün birləşmə imkanlarının verilməsi, xüsusi adların transliterasiya və tələffüzünün göstərilməsi, lüğətlərin daim təzələnməsinə nəzarət xidmətinin yaradılması nəzərdə tutulur. Kompüter lüğətlərinin daha bir üstün cəhəti onların canlı orqanizm kimi daim inkişafda və ətraf mühitlə, cəmiyyətlə təmasda olmasıdır. Bu tipli lüğətlərin gələcəkdə elə yeni cəhətləri, imkanları ola bilər ki, elmin indiki durumunda, cəmiyyətin indiki inkişaf səviyyəsində biz hələ onları heç təsəvvürümüza belə gətirmirik.

3.4. Kompüter lüğətləri elektron resurs kimi

Kompüter lüğətləri müxtəlif elmi mərkəzlər tərəfindən müxtəlif məqsədlərlə tərtib olunur. Məhz ona görə də, bu tipli lüğətlər informasiya texnologiyaları sistemində müxtəlif para-

metrlərə görə təsnif olunur. Onların bəziləri yalnız internet sistemlərində geniş istifadə üçün nəzərdə tutulmuşdur və onların kağız (kitab) variantları yoxdur. Bəzi kompüter lüğətlərinin isə kağız variantları mövcuddur. Daha doğrusu, bəzən elmi ictimaiyyət tərəfindən bəyənilmiş və gündəlik həyatda istifadə olunan lüğətlərin elektron versiyaları yaradılır. Bu versiyalar həm elektron, həm də kağız lüğətlərinə xas bəzi xüsusiyyətləri özündə əks etdirir. Yalnız kağız variantda olan lüğətin elektron variantı yaradılır və bu həmin lüğətlərə daha geniş istifadəçi kütləsinin mürciət etməsinə imkan yaradır. Bu lüğətlərin özlərinin də qismən elektronlaşdırılmış və tam elektronlaşmış variantları mövcuddur. Deyilənlərin yekunu olaraq mövcud kompüter lüğətlərini ümumi şəkildə aşağıdakı kimi təsnif etmək olar:

- maşın tərcüməsi və süni intellekt sistemlərində istifadə olunanlar;
- terminoloji bankların bazasını təşkil edən lüğətlər və tərcüməçilərə kömək məqsədi ilə yaradılmış "Poliqlot" tipli lüğətlər;
- müxtəlif bibliografik və ensiklopedik məlumatları əhatə edən elektron kartotekalar;
- kağız (kitab) variantında çoxdan mövcud olan və geniş istifadə dairəsinə malik lüğətlərin elektron variantları.

Maşın tərcüməsi üçün nəzərdə tutulmuş lüğətlər ənənəvi lüğətlərdən fərqli struktura və funksiyalara malikdir. Belə lüğətlərdə lüğət vahidləri əsaslardan, söz-formalardan, əsas və söz-formalardan ibarət ola bilər. Məsələn, Azərbaycan dili üçün tərtib olunmuş avtomatik lüğət əsas və işlənmə tezliyi böyük olan söz-formalardan təşkil olunmuşdur. Bu seçim aqlütinativ quruluşlu dillər üçün özünü doğruldur və daha məqsədəuyğun hesab olunur. Bu barədə monoqrafiyanın sonrakı fəsilərində ətraflı məlumat veriləcəkdir.

İngilis dilində ən çox yayılmış və geniş istifadə olunan kompüter lüğətlərinə "Dictionary.com", "Dictionaries" və "One-look Dictionaries" elektron resurslarını nümunə göstərmək olar. Bu lüğətlərin hər biri 200-dən çox lüğətdən ibarətdir. Onları

təşkil edən ümumi (ənənəvi lüğətlər tipində) və xüsusi (müəyyən elm sahəsinə aid terminoloji) lüğətlər bir axtarış sistemində birləşdirilir. Adları sadalanan lüğətlər müəyyən tematik qruplara ayrılmış birdilli (ingilis, alman, fransız, ispan, rus və b.) və ikidilli (ingilis-ispan, çin-ingilis, rus-ingilis və b.) lüğətləri əhatə edirlər.

İngilis dilində olan “Your Dictionary” 240 dildə (ingilis, fransız, alman və b.) olan lüğətlərdən və 30-dan artıq çoxdilli sorğu kitablarından ibarətdir. Kataloq daxilində lüğətlər dil qrupları üzrə (german dilləri, roman dilləri, slavyan dilləri və s.) və başqa əlamətlərə görə (qədim dillər, müasir dillər və s.) yerləşdirilmişlər. Sorğular əlifbaya görə sıralanmışlar.

“Your Dictionary” ingilisdilli lüğətlərin xüsusi “On-line lüğət” və “On-line thesaurus” kataloqları istənilən sözü eyni zamanda “Hypertext Dictionary”, “WordNet” və “Intralinguas English Language Dictionary” tipli elektron resurslarda da axtarmağa, onların fonetik xüsusiyyətlərini öyrənməyə və mənalərini dəqiqləşdirməyə imkan verir.

“Word Net” lüğətində axtarılan sözün bu və ya digər nitq hissəsinə mənsubluğu, eləcə də həmin sözün sinonim, antonim və omonimləri haqqında məlumat əldə etmək olar. Kataloqda “American Heritage Dictionary”, Thesaurus and Quotes”, “Cambridge Dictionaries Online”, “Dictionary of Difficult Words”, “The Jargon Dictionary”, “Online Dictionary”, “Database Query”, “Ono-look Dictionary”, “Grandiloquent Dictionary” və s. lüğətlər yerləşdirilmişdir. Bu lüğətlərin hər birinin çoxdilli və xüsusi sahə lüğətlərinə çıxışı vardır. Xüsusi sahə lüğətləri 86 fəaliyyət dairəsini əhatə edir. Bura biznes, maliyyə, tibb, kompüter texnologiyası, idman və s. daxildir.

Vəbdə çoxlu sayda qlossarilər, lüğətlər, tezauruslar var ki, onların hər biri haqqında məlumat vermək imkan xaricindədir və buna heç ehtiyac da yoxdur. “Google”da və ya başqa axtarış serverlərində sorğu verməklə onların hər biri haqqında məlumat almaq, xidmətlərindən istifadə etmək olar. “Google Azərbaycan”

axtarış blokunda bu lüğətlərlə Azərbaycan dilində də tanış olmaq imkanı var.

“Yahoo!” kataloqunda müxtəlif lüğətlər barədə 50-dən artıq kateqoriya üzrə 1500-dən artıq sayt vardır. Onlar müəyyən iyerarxiya üzrə sıralanmışlar: əvvəlcə tam və ümumi, sonra isə xüsusi sahə sorğu lüğətləri (kimya, tibb, kompüter texnologiyası, iqtisadiyyat və s.). Qeyd etmək lazımdır ki, bu lüğətlər Avropa və Şərq dillərini əhatə edir.

İnternet şəbəkəsində lüğətlərlə yanaşı ensiklopediyalar da geniş təmsil olunmuşlar. Belə ensiklopediyalara “Encarta”, “Britannica”, “Amerikana”, “Encyclopedia Smithsonian” “AUTOPEdia” və başqalarını misal gətirmək olar.

Rusiyada “Yandex” ensiklopediya resurs sistemi geniş tanınır. “Yandex” vasitəsi ilə “Böyük Sovet Ensiklopediyası”, “Kiçik Sovet Ensiklopediyası”, 11 cildlik “Ədəbiyyat ensiklopediyası”, “Rus sinonimlər lüğəti”, “Vətən tarixi”, V.Dahın “İzahlı lüğəti” və s. populyar nəşrlərə müraciət etmək olar. Bu lüğətlərə birbaşa çıxış var və istənilən söz barədə məlumat əldə etmək olar. “Dic.academic.ru” saytı vasitəsi ilə axtarılan söz barədə bütün mövcud ensiklopediyalardan eyni zamanda ətraflı izah almaq olar.

«Россия Он-лайн» – «Крыгосвет» elektron resursunda dilçilik, psixologiya, sosiologiya, iqtisadiyyat, tarix və b. sahələr üzrə məlumat yerləşdirilmişdir.

“Кирилл и Мефодий” saytında “Megabook” elektron resurslarda 10 onlayn lüğət bazasında 100000 məqalə, 30000 illüstrasiya yerləşdirilmişdir. Bundan başqa, orada xüsusi sahələri əhatə edən kiçik həcmli ensiklopediyalar bu elektron resurs formatında verilmişdir: “Musiqi ensiklopediyası”, “İncəsənət terminləri lüğəti”, “Orta əsrlər fəlsəfə kitabxanası”, “Latınca-rusca fəlsəfə terminləri lüğəti”, “Qədim Roma tarixi” və s.

Ayrı-ayrı ali məktəblərin saytlarında da böyük elmi maraq doğuran ensiklopediyalar özünə yer tapmışdır. Buna misal olaraq “M.V.Lomonosov adına Moskva Dövlət Universiteti tarix fakültəsinin elektron resurslar kitabxanası”nı göstərmək olar.

Rusiyanın məşhur “Mail.ru” elektron poçt serverində də lüğət və ensiklopediyaların elektron resurslarına rast gəlmək mümkündür: “Российский энциклопедический словарь», «Всемирный биографический энциклопедический словарь», «Всемирная История», «Мифы народов мира», «Народы и религии мира» və s.

«Словарь и энциклопедии on-line» məlumat-sorğu kataloqunda aşağıdakı izahlı və sahə terminoloji lüğətləri yerləşdirilmişdir: “Толковый словарь Ушакова», «Толковый словарь В.И.Даля», «Финансовый словарь», «Экономический словарь», «Англо-русский словарь финансовых терминов», «Словарь терминов антикризисного управления», «Исторический словарь», «Этимологический словарь русского языка М.Фасмера» və s.

3.4.1. Tərcümə kompüter lüğətləri

Kompüter lüğətlərinin bu növü də internet məkanında geniş yer tutur. Bu barədə internet axtarış sistemlərində geniş məlumat verildiyi üçün qısa icmal məlumatı ilə kifayətlənəcəyik. Kompüter tərcümə lüğətlərinin ən çox istifadə olunanları Abbyy Lingvo 12, Multilex Delux 6.4, Context 7.0, Multitran 3.44, VERDiet 2.0 Qigant, Polyglossum 3.20 və s. hesab olunur. Onların hər biri lüğət dəstlərindən ibarət olub lüğət sistemi təşkil edirlər. Məsələn, Abbyy Lingvo kompüter lüğətləri sistemi 47 lüğətdən ibarətdir və sistem mütəmadi olaraq yenilənir. Lüğətin yeni versiyasına ingilis dilinin 120 min sözdən ibarət izahlı lüğəti (Collins English Dictionary), «Новый большой русско-английский словарь» (tərtibçilər D.İ.Yermoloviç və T.M.Krasavin, 110 min sözdən ibarət), “Русско-английский словарь религиозной лексики» (tərtibçi A.A.Azarov, 14 min söz), “Большой англо-русский фразеологический словарь» (tərtibçi A.V.Kunin, 20 min söz) kimi lüğətlər əlavə olunmuşdur. Bütövlükdə sistemin lüğət bazası 2,5 milyon lüğət vahidinə çatdırılmışdır. Orada hər bir söz haqqında ətraflı qrammatik məlumat verilir, onların real mətnlərdə işlənməsi nümunələr əsasında göstərilir, sözlərin bir qismi səsləndirilir.

Abby Lingvo proqramında sözlərin morfoloji formaları barədə sorğulara cavab verilir və onun vasitəsi ilə www.lingvo.ru saytında başqa lüğətlərə pulsuz çıxış mümkündür.

Multilüks elektron lüğətlər versiyasına 645 ümumi, terminoloji, izahlı və ölkəşünaslığa aid lüğət 6 dildə daxil edilmişdir. İngilis dilinin izahlı lüğəti (Webster New World Dictionary – Thesaurus), eləcə də 23 sahə terminoloji lüğəti sistemdən işlənən məlumatı almağa imkan verir. Terminoloji lüğətlər iqtisadiyyat, bank işi, hüquqşünaslıq, texnika, texnologiya, biologiya, tibb, ölkəşünaslıq və s. sahələri ehtiva edir.

3.4.2. Çoxdilli Lingvo 8.0 elektron lüğəti

8.0 versiyasına qədər bu lüğət ingilis-rus-ingilis variantında mövcud idi. İndiki versiyada həmin lüğətə alman, fransız, italyan dilləri də əlavə olunmuşdur. Bu lüğətlər bir-biri ilə əlaqəlidir və istənilən söz və ya söz birləşməsini axtarmaq və sadalanan dillərə tərcümə etmək mümkündür. Amma bu dillərin bəzilərində lüğət bazası zəifdir və həmin dillərə tərcümədə boşluqlar yaranır. Məsələn, italyan dilinin ümumişlək sözlər lüğəti sistemə daxil edilməmişdir və italyan dili orada yalnız politexnik sözlər lüğəti ilə təmsil olunmuşdur. Proqramda sözün müxtəlif qrammatik formalarda tanınması mümkündür. İstənilən söz-forma sistemə daxil edildikdə, onun başlanğıc forması lüğətdə avtomatik əks olunur. Lingvoda bütün sözlərin paradiqmaları işlənib hazırlanmışdır. Bəzi dar ixtisas sahələrinə aid terminlər lüğətə düşməyə bilər. Əgər yeni daxil edilən sözün paradiqma tipi məlumdursa, lüğətdə avtomatik olaraq həmin söz paradiqmaları ilə alınır, yeni daxil edilən sözün hansı nitq hissəsinə mənsubluğu, bəzi əlamətləri (xüsusi isim, qısaldılmış söz və s.) və ilkin forması məlumdursa, paradiqma tipi düzgün seçilə bilər.

Lüğət məqalələri başlıq (giriş), tərcümə zonası, nümunələr zonası və qeydlər zonasından ibarətdir. Lüğət siyahısında ardıcılıq tərcümə mətnindən asılı olaraq dəyişdirilə bilər. Əgər tərcümə olunacaq mətn kompüter texnologiyasına aiddirsə, onda kompüter terminləri birinci sırada duracaq.

Nümunələr zonasında tipik sözişlənmə formaları, frazeologizmlər, sabit söz birləşmələri əks olunmuşdur. Linqvoda təkcə ayrı-ayrı sözləri deyil, söz birləşmələri və hətta cümlələri də tərcümə etmək mümkündür.

Lingvo 8.0-da 41 lüğət yerləşdirilmişdir. Onların ikisini xüsusi qeyd etmək lazımdır: Lingvo-Thesaurus (RU)-rus işgüzar leksikası tezaurusu və Lingvo Grammatical (En-Ru) – ingiliscə-rusca qrammatik lüğət.

Mahiyyətinə görə qrammatik lüğət ingilis dilinin qrammatikası üzrə hipermetn sorğusu mənbəyidir. Əsas qrammatik anlayışlar, qaydalar və onların tətbiqini əks etdirən nümunələr orada yerləşdirilmişdir.

Kartoçkalarda təkcə hər bir ayrıca götürülmüş qrammatik informasiya deyil, həm də yaxın konstruksiyaların fərqləri verilmişdir. Bundan başqa, nitq etiketi normaları, ingilis dilinin Britaniya və Amerika variantları arasında fərqlər, xüsusilə sözlərin tələffüz variantları barədə də kifayət qədər məlumat toplamış və lüğətdə yerləşdirilmişdir. Lüğət həm dərs vəsaiti, həm də ingilis dilinin qrammatikası üzrə məlumat mənbəyi kimi istifadə oluna bilər. Xarici dili öyrənmək istəyənlər də professional tərcüməçilər də linqvodan etibarlı bir mənbə kimi geniş istifadə edirlər. Linqvonun yeganə çatışmayan cəhəti leksik bazanın keyfiyyətinin aşağı olmasıdır. Lakin ABBYY kompaniyası bu işə diqqəti artırmışdır: mövcud lüğət bazası yeni sözlər və terminlərlə daim zənginləşdirilir, interfeys təkmilləşdirilir.

3.4.3. Context 4.0 elektron lüğəti (“İnformatik” kompaniyası)

“Context” lüğətində 46 xüsusişdirilmiş lüğət yerləşdirilmişdir ki, onların da 44-ü hər iki istiqamətə işləyir. Bu lüğətlərə onlayn rejimində müraciət mümkündür. “Context”i təşkil edən lüğətləri aşağıdakı qruplara ayırmaq olar:

- ingiliscə-rusca-ingiliscə lüğətlər (Böyük ümumişlək sözlər lüğəti, kompüter terminləri, kommertiya ter-

minləri, tibb terminləri, hüquq terminləri və “İngilis dilinin çətinlikləri” lüğətləri).

- rus dili lüğətləri (“Rus sinonimləri”, “Rus dilinin izahlı lüğəti”, “Frazeoloji lüğət, “Xarici sözlər lüğəti”).
- başqa dillər üzrə lüğətlər (ingilis-italyan, ingilis-ıspan, ingilis-alman, alman-portuqal, ingilis-serb, ingilis-fransız, ingilis-xorvat, fransız-italyan, fransız-alman).

“Context” lüğətlərinin bir hissəsi kağız lüğətlərin elektron versiyalarıdır (S.U.Ojeqovun “İzahlı lüğət”i; “Xarici sözlər lüğəti”, “Frazeoloji lüğətlər”).

Lüğət məqalələri giriş (tərcümə olunan söz), tərcümə zonası, transkripsiya (ingilis dilindən tərcümədə), qrammatik izahlar zonası (nitq hissəsi, felin növləri, zamanları və s.) nümunələr zonası və izahlar zonasından ibarətdir. Üslubi xüsusiyyətlər və sahə mənsubluğu qeyd olunmur. Vurğu və transkripsiya yalnız ingilis dilindən rus dilinə tərcümədə nəzərə alınır. Context 4.0 lüğətinin üstün cəhəti ingilis və rus dillərinin morfolojiyasına bələdlidir. Bu imkan verir ki, istənilən qrammatik formada olan söz və ya söz-forma düzgün tərcümə olunsun.

Axtarış aparılan mətnlər “Sorgular tarixi” bölməsində siyahıya alınır. Axtarılan sözü yenidən axtarmaq lazım gəlmir, söz avtomatik olaraq ekrana gəlir. “Çətin sözlər” xüsusi blokda yerləşir və onlar haqqında geniş qrammatik izahat verilir. Lüğətin çatışmayan cəhəti sinonim və antonimlər barədə məlumatın olmamasıdır. Sözlərin səsləndirilməsi də nəzərə alınmamışdır.

3.4.4. Polyglossum 3.2 elektron lüğəti

Polyglossum elektron lüğəti haqqında ümumi şəkildə verilmiş məlumata əlavə olaraq bəzi məsələləri nəzərdən keçirək.

Polyglossum sistemi 1995-ci ildən fəaliyyətə başlasa da, onun haqqında məlumat geniş yayılmamışdır. Bu sistem yarandıqda kompüter lüğətləri və kompüter vasitəsilə tərcümə məsələləri nadir hadisələrdən sayılırdı. O vaxtlar qrammatik təhlil və sözlərin informasiya ilə təchizi məsələləri gündəmdə deyildi. Polyglossum kağız lüğətlərə yaxın variantda hazırlanırdı.

Windows sisteminə keçid və Multileks, Lingvo kimi tərcümə lüğətləri yaranandan sonra, polyglossum bir müddət sanki unuduldu. Lakin, bir çox professional tərcüməçilər indi də polyglossuma üstünlük verirlər.

Bu sistemə daxil olan elektron lüğətlər professional lüğətçilər tərəfindən hazırlanır. Elektron variant tərtibçi ilə birlikdə hazırlanır və onun kağız variantı sonralar elektron versiya əsasında hazırlanır.

Polyglossum 3.2 lüğət dəsti 1.8 versiyasının təkrarıdır. Bu dəstə aşağıdakılar daxildir:

Böyük ümumleksik lüğət, Böyük politexnik lüğət, Hesablama texnikası və proqramlaşdırmaya aid lüğət, Riyaziyyat və patent lüğəti, Qısaltmalar və rəsmi adlar lüğəti, Rabitə üzrə qısaltmalar lüğəti, Mühəndisin lüğəti, Mikroelektronika üzrə lüğət, Telekommunikasiya və rabitə üzrə lüğət (ingiliscə-rusca və rusca-ingiliscə).

Yeni versiyada eyni zamanda bir sözün bütün lüğətlərdə axtarışı mümkündür. Əgər hər hansı bir lüğət lazım deyilsə, onu siyahının axırına atmaq mümkündür.

Lüğət məqalələrində informasiya minimuma endirilib, başlıq və tərcümə zonasından ibarətdir. Polyglossum dar ixtisas sahələrini əhatə etdiyi üçün terminlərin tərcüməsi söz-birləşmələri ilə bir yerdə verilir. Linqvistik məlumat minimaldır: qaydasız fellər üçün formalar, isimlər üçün cins göstərilir. Sözlərin transkripsiyası və üslubi əlamətlər qeyd olunmur.

Bu proqram professional tərcüməçiyə öz lüğətini tərtib etmək imkanı verir. Polyglossumda bu məqsədlə istifadəçi tərəfindən tamamlanan lüğətlər var: English-French-German-Finnish-Russian.

Yuxarıda göstərilənlər “Lingvo”, “Polyglossum”, “Multilex” və “Context” kimi kompüter lüğətlərinin üstün və çatışmayan cəhətlərini təhlil etməyə imkan verir. Təhlil aşağıdakı kriteriyalar əsasında aparılır: morfoloji prinsip, lüğət məqalələrinin tamlığı və əlverişliliyi, axtarış rejimi, ikitərəfli tərcümənin mümkünlüyü, söz birləşmələri və cümlələrin tərcüməsi, prioritet lü-

ğətlərin seçimi və onlara qoşulma, əlavə məlumatların mövcudluğu (qrammatik, fonetik və s.). Təhlilə cəlb olunmuş lüğətlərin hamısı bu və ya digər dərəcədə bu meyarlara cavab verir.

Linqvo lüğətlərinin əsas üstünlükləri aşağıdakılardır:

Unikal qrammatik məlumatlar, fonetik proqramlar, bütün sözlər üçün paradıqmaların olması. Bu, təkcə söz-birləşmələrini deyil, cümlələri də tərcümə etməyə imkan verir).

Polyglossum daha çox xüsusi lüğətlərdən ibarətdir, həcmi kiçikdir və bütövlükdə Windows əməliyyat sisteminə yerləşdirilə bilər. İnformasiyanın təqdimatına görə də Polyglossum əlverişlidir.

Multilex nüfuzlu kağız lüğətlərin dəqiq sürətidir. Bu onun keyfiyyətinə zəmanətdir. İngilis və rus sözlərində vurğunun qoyulması, transkripsiya, üslubi işarələr bu lüğətin üstünlüklərindən hesab oluna bilər. Məhz bu xüsusiyyətlərinə görə Multilex bahalı proqram hesab olunur.

Context-in əsas üstünlüyü tərcümə üçün çoxlu tematik lüğətlərin mövcudluğu, istənilən formada olan söz birləşmələrinin çıxarılmasının mümkünlüyü, bütün birləşmələrin digər sözlərlə tərcüməsinin siyahı halında müşahidəsinin mümkünlüyüdür. Sinonim və antroponimlərin lüğətlərdə əks olunmaması və səslənmənin yoxluğu kontekstin çatışmazlıqlarındandır.

Təhlil göstərir ki, təhlilə və əhatə dairəsinə görə Polyglossum birinci yerdədir. Lüğət məqalələrinin ümumi həcmi 3442000 termindən ibarətdir. Lüğətlərdə dar ixtisas sahələri hesab olunan kitabxanaçılıq işi, aviasiya, kosmos, maşın detalları, geodeziya, kartoqrafiya, yeyinti və toxuculuq sənayesi və s. əhatə olunmuşdur.

Multilex və Linqvo lüğətləri həcmcə təxminən eynidirlər: 180222700 və 1217500 lüğət məqaləsi. Hər iki lüğət tibb, texnika, poliqrafiyaya aid mövzuları geniş əks etdirir və onlarda ümumişlək leksikaya önəm vermişdir. Bu da bədii ədəbiyyatın tərcüməsi üçün əhəmiyyətli amildir. İtalyan və ispan dillərinin ümumişlək lüğətlərinin mövcudluğu Multilex-in roman dillərinin öyrənilməsi prosesində istifadəsini ön plana çəkir. Linqvo-da isə fransız dilinə aid lüğətlər daha çox təmsil olunmuşdur.

Texniki mətnlərin professional tərcüməsində mühəndislər, alimlər daha çox Polyglossuma üstünlük verirlər. Kontekst rus dilini öyrənmək istəyən xaricilər üçün daha əlverişlidir.

Azərbaycan dilinin elektron lüğətləri təhlilə cəlb olunmamışdır. Həmin lüğətlər haqqında ayrıca məlumat veriləcəkdir.

3.4.5. Elektron ensiklopediyalar

Kompüter lüğətləri şəbəkəsində elektron ensiklopediyalar da mühüm komponentlərdən hesab olunur. Axtarış sistemlərinin bir çoxunda “sorgu nəşrləri” kateqoriyasında ensiklopediyalara və elektron ensiklopediya lüğətlərinə istinadlara rast gəlmək olar. İnternetdə ümumi (məsələn, Encyclopedia Americana Online, Encarta), eləcə də xüsusi ensiklopediyalar (məsələn, The Encyclopedia Mythica, The Encyclopedia of Canadian, Thistle on the WWW, standart Encyclopedia of Polyglossum, Class Encyclopedia) geniş təmsil olunmuşlar.

Nüfuzlu Britaniya ensiklopediyası, Americana, Encarta, Encyclopedia, FREE Internet Encyclopedia ilə yanaşı minlərlə az tanınan, lakin faydalı və əhəmiyyətli ensiklopediyalar da internetdə yerləşdirilmişdir. Elm və texnikaya aid belə ensiklopediyalara misal olaraq Creation-Evolution Encyclopedia, Encyclopedia Smithsonian, AUTOPEdia, Natural Remedies Encyclopedia, Encyclopedia Metallica və s. göstərmək olar. Bu ensiklopediyaların bir çoxunun əsasında kağız ensiklopediyalar durur. Məsələn, «Кирилл и Мефодий» multimediyaya elektron ensiklopediyasının əsasında “Böyük ensiklopediya lüğəti” durur. Bu ensiklopediya sistemində 89500-dən artıq ensiklopediya və sorgu lüğəti, 40000-dən çox multimediyaya illüstrasiyası, 860 audiofraqment, 570 videofraqment, 1600 ilkin mənbə sitatı, dünyanın interaktiv xəritəsindən 520 xəritə, 140000 anlayışın izahı, 470 bədii ədəbiyyat və humanitar elmlərə aid əsərlər, 1000 biblioqrafik təsvir, 12 interaktiv fotonoram, 900-dən artıq illüstrasiyanı əhatə edən 150 fotoalbom yerləşdirilmişdir.

“Encarta” elektron multimediyaya ensiklopediyası 68 mindən artıq məqalədən ibarətdir. Burada müxtəlif statistik məlumatlar,

çoxlu sayda təsvirlər, videofraqlentlər, tarixi xəritələr, dünyanın interaktiv xəritəsi əks olunmuşdur. Encarta ensiklopediyası internetdə tam təmsil olunmamışdır. Onun tam variantı disk şəklində buraxılmışdır.

«Кругосвет» universal rus elektron ensiklopediyası 2000-ci ildən fəaliyyət göstərir. 1200-dən artıq məqalə, 600 bioqrafiya, 11500-ə yaxın illüstrasiya və yüzlərlə tarixi sənədi əhatə edir. Yalnız elektron variantda “krugosvet.ru” saytında pulsuz istifadə oluna bilər.

Ən populyar elektron ensiklopediyalardan biri də “Vikipediya” hesab olunur. “Viki” konsepsiyası ilk dəfə 25 mart 1995-ci ildə ABŞ-ın Oreqona ştatında yaşayan Vord Kenningem tərəfindən işlənmiş, veb sahifələrin idarəsi, birbaşa veb-brauzerdən redaktə olunması üçün nəzərdə tutulmuşdur. “Viki” havay dilində “daha tez” deməkdir. “Viki” texnologiyası dedikdə “sürətli informasiya texnologiyası” başa düşülür. Bəzən Viki texnologiyasını “canlı sənədlər” texnologiyası da adlandırırlar (Əliquliyev, Ələkbərova, 2010, s.13-14). Çoxdilli, açıq sistem olan Vikipediya da könüllü həvəskar ensiklopediyaçıları tərəfindən hazırlanmışdır. 2001-ci il yanvarın 15-dən fəaliyyət göstərən bu layihənin koordinatoru “Wikimeyia Foundation Inc” amerikan fondudur. Yaradıcıları Ohayo Universitetinin fəlsəfə doktoru Larri Senqer və amerikalı sahibkar Cimmi Uelsdir (Əliquliyev, Ələkbərova, 2010, s.19).

İstifadəçilər orada öz məqalələrini yerləşdirə bilər və ya başqa istifadəçilərin məqalələrinə düzəlişlər edə bilərlər. İstifadəçilər öz məqalələrini ABŞ-ın Florida ştatında yerləşən serverlə müəllif hüququ qanununa əməl etməklə razılaşıdırırlar.

“Vikipediya”nın üstünlüyü ondadır ki, məqalə müəllifi eyni zamanda redaktor və korrektor funksiyasını yerinə yetirir, o, öz məqaləsini təqdim edir, yaxud başqalarının məqalələrində düzəliş edir və bunları bütün istifadəçilər müşahidə etmək imkanına malik olurlar. Məqalələr neytral olmalı, tədqiqatın orijinal versiyası olmamalıdır. Vikipediya elmi məcmuə deyil, ikinci məlumat mənbəyidir. Məqalələrin ədəbi dilin elmi üslubunda yazılması tövsiyə olunur, şəxsi imza və qeydlərə yol verilmir.

Vikipediya neytral mövqedən çıxış edir. Məqalə müəllifi məqalə ilə bağlı bütün mübahisəli məsələləri əks etdirməli heç bir fikrə öz şəxsi münasibətini bildirməməlidir.

Vikipediya materialları istifadə üçün açıqdır, hər kəs istənilən vaxt öz düzəlişini, əlavəsini etmək hüququna malikdir.

Dünyanın ingilisdilli ən böyük “Vikipediya”sında 2 milyondan çox məqalə və 800 mindən artıq illüstrasiya vardır. Hər gün Vikipediyanın ingilis versiyasına (en.wikipedia.org) 300 min adam müraciət edir. Vikipediyanın alman versiyasında 188 min məqalə var. Üçüncü yerdə gedən yapon Vikipediyasında 94 min məqalə barədə məlumat verirlər. Vikipediyanın müxtəlif dilləri əhatə edən versiyalarında 250 mindən çox məqalə yerləşdirilmişdir.

Vikipediya günbəgün təkmilləşən, inkişaf edən bir sistemdir. Hazırda onun tərkibində kitabxana, lüğət, dərslik, yeniliklər və s. bölmələr yaradılmışdır.

Vikipediya yalnız elektron variantda mövcuddur. Orada çoxlu hipermətn istinadları şəbəkəsi yaradılmışdır ki, bu da istifadəçilərin başqa mənbələrə istinad etmək imkanını artırır.

Rus Vikipediyasında hazırda 791 “yaxşı” məqalə yerləşdirilmişdir. Bu kateqoriyaya daxil edilən məqalələr istifadəçilərin əksəriyyəti tərəfindən bəyənilən məqalələrdir. Həmin məqalələr xüsusi olaraq seçilir və işarələnir.

“Yaxşı məqalə”də hər şey dəqiq, tam və ya neytral olmalı, üslubi baxımdan istifadəçiləri qane etməlidir.

Vikipediyada azərbaycanlı istifadəçilərin sayı da getdikcə artmaqdadır. “Vikipediya (virtual ensiklopediya)” kitabında azərbaycanlı istifadəçilərin fəallıq dinamikası getdikcə yüksəlir: “Belə ki, əgər 2007-ci ildə Vikipediyanın Azərbaycan dili bölməsində məqalələrin sayı cəmi 3 min idisə, 2010-cu ildə məqalələrin sayı artıq 35 minə çatmışdır” (Əliquliyev, Ələkbərova, 2010, s.9).

3.4.6. Assosiativ lüğətlər

Assosiativ tərcümələr göstərir ki, sözün semantik strukturunu ancaq sssosiativ münasibətlər fonunda tam şəkildə açmaq mümkündür. Belə təcrübələrlə lüğətlərdə indiyə qədər təsbit

olunmayan leksik-semantik variantların tapılması, sözlərin milli-mədəni xüsusiyyətlərinin müəyyənləşdirilməsi mümkündür. Sonralar məlum oldu ki, sözlərin mənasında baş verən dəyişikliklər cəmiyyətin həyatında olan köklü dəyişikliklərlə sıx bağlıdır. Uşaq nitqinin böyüklərlə müqayisədə assosiativ təcrübələrlə tədqiqi də sözün mənacə dəyişməsi prosesinin dinamikasını izləməyə imkan verir (Залевская, 1990; 2007).

Hazırda assosiativ lüğətlər daha çox kağız variantda geniş yayılmışdır, lakin bu lüğətlərin elektron variantlarına da rast gəlmək mümkündür.

Hər şeydən əvvəl, kompüterlərdə assosiativ təcrübələrin nəticələri əsasında bu və ya digər linqvomədəni cəmiyyətin nümayəndələrinin dili təhlil edilə və sözün mənasının dərin və hər-tərəfli səciyyələri müəyyənləşdirilə bilər. Bu mənada assosiativ normalar lüğəti ilə assosiativ tezaurusları fərqləndirmək lazımdır.

İlk assosiativ tezaurus Edinburq Universitetinin tələbələri tərəfindən hazırlanmış “İngilis dilinin assosiativ tezaurusu” hesab olunur. Bu lüğətdə stimül sözlərdən assosiativ münasibətlərin birbaşa və tərsinə siyahılarının alınmasına qədər bütün proses avtomatlaşdırılmışdır. Nəticədə 9 min təcrübə iştirakçısının 8400 stimül sözə münasibəti əsasında 55 min lüğət məqaləsi alınmışdır. Bütün təhlil prosesi kompüterlərdə aparılmış və nəticələr ümumiləşdirilmişdir.

“İngilis dilinin assosiativ tezaurusu”nun prinsipləri əsasında “Rus dilinin assosiativ lüğəti” tərtib olunmuşdur. Kompüterdə hazırlanmış bu lüğət dil bacarığının təhlilində yeni linqvistik baza və alət kimi səciyyələndirilir (Черкасова, 1996). Bütün bunlar bir daha sübut edir ki, assosiativ lüğətlər gələcəkdə elektron resurslarda mühüm bir komponent kimi təmsil olunmalıdır. Belə lüğətlər olmadan sözlərin tam semantik izahı qeyri-mümkündür.

3.5. Azərbaycanda kompüter lüğətçiliyi

Azərbaycanda kompüter lüğətçiliyinin tarixi ötən əsrin 70-ci illərindən başlanır. Həmin illərdə ilk dəfə olaraq Azərbaycan-

da qəzet dilinin tezlik lüğəti “BESM – 2” elektron hesablama maşınlarının köməyi ilə tərtib olunmuşdu.

Azərbaycanda kompüter lüğətçiliyi məsələsinə iki yanaşma mövcuddur. Birinci yanaşmada kompüter vasitəsi ilə tərtib olunan müxtəlif statistik lüğətlər nəzərdə tutulur. İkinci yanaşma internet məkanında təmsil olunan lüğətləri əhatə edir. Kompüterdə təmsil olunan lüğətləri də iki qrupa ayırmaq olar: heç bir ilkin işlənmə olmadan, kağız (kitab), variantında olduğu kimi internetdə yerləşdirilmiş lüğətlər və xüsusi olaraq internet məkanında istifadə üçün işlənmiş elektron lüğətlər.

3.5.1. Kompüterin köməyi ilə tərtib olunmuş lüğətlər

Kompüter dilçiliyinin predmeti kimi əvvəlcə kompüterlərin köməyi ilə tərtib olunmuş lüğətləri nəzərdən keçirək:

1. Yuxarıda qeyd olunduğu kimi, ilk belə lüğət “Azərbaycan qəzet dilinin tezlik lüğəti”dir. Lüğət 70-ci illərdə tərtib olunsada 2004-cü ildə işıq üzü görmüşdür (Azərbaycan qəzet dilinin tezlik lüğəti. Bakı, “Nurlan” 2004. – 264 s.). (Tərtibçilər – K.Vəliyeva, M.Mahmudov, V.Pines, redaktorlar – akad. M.Şirəliyev, prof. A.Ələkbərov, rəyçilər – AMEA-nın müxbir üzvü A.Axundov, filologiya elmləri namizədi B.Abdullayev, proqramlaşdırıcı – O.E.Kulqavina, ön sözün müəllifi – M.Mahmudov).

2. “Kitabi-Dədə Qorqud” dastanlarının statistik təhlili (tərtibçilər Vəliyeva K.A., Mahmudov M.Ə., Pines V.Y., Rəhmanov C.Ə., Sultanov V.S., redaktor – prof. Axundov A.A., rəyçilər – prof. Vəliyev K.N., Əhmədov F.Z., mühəndis-proqramlaşdırıcı – Məmmədov O.S., ön sözün müəllifi – Mahmudov M.Ə.). Bakı, “Siyasət”, 1997. – 78 s.

3. “Kitabi-Dədə Qorqud” dilinin statistik təhlili (ilkin nəticələr), tərtibçilər – Vəliyeva K.A., Mahmudov M.Ə., Pines V.Y., Rəhmanov C.Ə., Sultanov V.S., redaktorlar – filologiya elmləri doktoru, professor Axundov A.A., filologiya elmləri doktoru Mahmudov M.Ə., rəyçilər – filologiya elmləri doktoru, professor Vəliyev K.A., filologiya elmləri namizədi Əhmədov F.Z., mü-

həndis-proqramlaşdırıcı – Məmmədov O.S., ön sözün müəllifi – Mahmudov M.Ə.). Bakı, “Elm”, 1999. – 248 s.

4. Məhəmməd Füzulinin nəzm əsərlərinin əlifba-tezlik sözlüyü (tərtibçilər: K.A.Vəliyeva, M.Ə.Mahmudov, C.Ə.Rəhmanov, V.S.Sultanov; redaktorlar: AMEA-nın müxbir üzvü A.A.Axundov, filologiya elmləri namizədi A.B.Mirzəyev rəyçilər: filologiya elmləri doktoru, prof. R.C.Məhərrəmov, filologiya elmləri doktoru, prof. A.Q.Ələkbərov, mühəndis-proqramlaşdırıcı – O.S.Məmmədov, ön sözün müəllifi – filologiya elmləri doktoru M.Ə.Mahmudov). Bakı, “Elm”, 2004. – 548 s.

5. “Azərbaycan dilinin əks-əlifba lüğəti” (tərtibçilər: Məsud Mahmudov, Əbülfət Fətullayev, redaktorlar: AMEA-nın müxbir üzvü, filologiya elmləri doktoru, professor Ağamusa Axundov, filologiya elmləri doktoru Kamilə Vəliyeva, rəyçilər: filologiya elmləri doktoru İsmayıl Məmmədov, filologiya elmləri doktoru Sayalı Sadıqova). Bakı, “Nurlan”, 2004. – 524 s.

6. Azərbaycan qəzet dilinin tezlik lüğəti (söz kökləri), I cild (tərtibçilər: M.Mahmudov, Ə.Fətullayev, S.Məmmədova, R.Fətullayev, R.Hüseynli, B.Talıbov, G.Əzimova, N.Abdullayev, elmi məsləhətçi: akad. Ə.Abbasov, redaktor: A.Axundov, rəyçilər: filologiya elmləri doktoru K.Vəliyeva, filologiya elmləri doktoru, prof. S.Sadıqova). Bakı, “Elm”, 2010. – 464 s.

Bunlardan əlavə, məşin tərcüməsi ilə bağlı tezlik lüğətləri bazasında təcrübi azərbaycanca-rusca avtomatik lüğət hazırlanmış və onun əsasında avtomatik linqvistik təhlil aparılmaqla məşin tərcüməsi nümunələri alınmışdır (Mahmudov, 1991, Mahmudov 1992).

Kompüterlərin köməyi ilə tərtib olunmuş bu lüğətlərin linqvistik xüsusiyyətləri və özəllikləri barədə monoqrafiyanın sonrakı fəsilərində ətraflı məlumat veriləcəkdir.

Azərbaycan Milli Kitabxanasının, ali tədris müəssisələrinin saytlarında ayrı-ayrı şəxslərin hazırladığı bir çox lüğətlərin kağız (kitab) variantı internetə qoyulmuşdur. Onların çoxu ancaq oxunmaq üçün nəzərdə tutulmuşdur, yəni elektron lüğət kimi öncədən işlənməmişdir. Müəyyən sorguya əsasən onlardan məlumat al-

maq, tədqiqat üçün müəyyən parametrlərə görə məlumat toplamaq qeyri-mümkündür. Kağız (kitab) variantında da olsa, lüğətlərin internet məkanına çıxarılması müsbət haldır və bunu gələcəkdə elektron varianta keçidin ilk pilləsi hesab etmək olar.

3.5.2. İnternetdə yerləşdirilmiş elektron lüğətlər

Bu qrupa daxil edilən lüğətlər internetə qoyulmazdan əvvəl işlənmiş və istifadəçilərin sorğusuna görə məlumatvermə imkanına malikdirlər. Belə lüğətlər həm müstəqil, həm də tərcümə sistemlərində komponentlərdən biri kimi elektron məkana daxil olunmuşlar:

1. “Dilmanc” Azərbaycanda ilk maşın tərcüməsi sistemidir. Azərbaycanca-İngiliscə, İngiliscə-azərbaycanca və türkcə-azərbaycanca “Dilmanc” tərcümə sahəsində artıq tanınır. Bu sistemdə böyük mətnlər də tərcümə olunur. “Dilmanc” maşın tərcüməsi sistemində alınmış tərcümənin və ayrı-ayrı sözlərin düzgün tələffüzü səsləndirilə bilər. “Dilmanc”ın yaranması ilə Azərbaycan dilinin formal qrammatikası yaradılmış və onun əsasında ilkin dildə formal morfoloji, sintaktik və semantik təhlil, tərcümə olunmuş dildə cümlənin sintezi alqoritmləri işlənilib hazırlanmışdır. Leksik baza 250000-dən artıq ümumişlək söz-forma və söz birləşməsindən ibarətdir.

“Dilmanc”da maşın tərcüməsi sahəsində işlərə 2003-cü ilin əvvəllərindən başlanılmışdır. Bu müddət ərzində Azərbaycan dilindən digər dillərə və əksinə tərcümə prosesini avtomatlaşdırmağa imkan verən xüsusi modelləşdirmə metodu (rəqəmsal modelləşdirmə metodu) yaradılmış və bu metod əsasında Azərbaycan-İngilis maşın tərcüməsi sisteminin yaradılmasına başlanılmışdır. Bu sahədə aparılan işlər Azərbaycan Respublikası Rabitə və İnformasiya Texnologiya Nazirliyi ilə BMT-nin İnkişaf Proqramının Azərbaycandakı Nümayəndəliyinin birgə layihəsi ilə aparılır. Layihə üzrə işlərə 2005-ci ilin apreldən başlanılmışdır. Layihənin həyata keçirilməyə başladığı vaxtdan etibarən Azərbaycan-İngilis, İngilis-Azərbaycan və Türk-Azərbaycan istiqamətləri üzrə tərcümənin ilkin versiyası hazırlanmış və istifa-

dəçilərə təqdim edilmişdir. Layihənin elmi məsləhətçisi – akademik Əli Abbasov, direktoru – texnika üzrə fəlsəfə doktoru, dos. Əbülfət Fətullayev, sahə üzrə məsləhətçilər – filologiya elmləri doktoru, professor Məsud Mahmudov (Dilçilik), texnika elmləri doktoru, professor Bayram İbrahimov (Sıqnaların rəqəmli emalı). (Bax: <http://az.wikipedia.org/w/index.php/title=Dilmanc&oldid=1609121>; <http://az.wikipedia.org/wiki/Dilmanc>; <http://www.dilmanc.az/az>

2. Poliqlot Azərbaycan lüğətləri

Poliqlot lüğətlər sistemi istifadəçilərin dərin rəğbətini qazanmış populyar 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. Lüğətlərin elektron variantının hazırlanması prosesində aşağıdakı lüğətlərdən istifadə olunmuşdur:

- İngiliscə-azərbaycanca lüğət (Q.Rzayevin redaktəsi ilə). 1994, Bakı, Azərbaycan Tərcümə Mərkəzi (20 minə yaxın söz və ifadə);
- Məktəblinin ingiliscə-azərbaycanca lüğəti. 2000, Bakı, Tələbə Qəbulu üzrə Dövlət Komissiyası (3500-ə yaxın söz);
- Azərbaycanca-ingiliscə lüğət (O.Musayevin redaktəsi ilə). 1998, Bakı, Azərbaycan Dillər Universiteti (45 min söz və ifadə, 10 min cümlə);
- Məktəblinin almanca-azərbaycanca lüğəti. 2001, Bakı, Tələbə Qəbulu üzrə Dövlət Komissiyası (3000-ə yaxın söz);
- Məktəblinin fransızca-azərbaycanca lüğəti. 2003, Bakı, Tələbə Qəbulu üzrə Dövlət Komissiyası (5000-ə yaxın söz);

- Məktəblinin azərbaycanca-fransızca lüğəti. 2003, Bakı, Tələbə Qəbulu üzrə Dövlət Komissiyası (5000-ə yaxın söz);
- Rusca-azərbaycanca lüğət. (Ə.Orucovun redaktəsi ilə). 1990, Bakı, AMEA Dilçilik İnstitutu (90000-ə yaxın söz 8000 ifadə);
- Azərbaycanca-rusca lüğət (X.Əzizbəyovun redaktəsi ilə). 1984, Bakı, (2000-ə yaxın söz);
- İngiliscə-rusca lüğət (V.K.Müllərin redaktəsi ilə), 7-ci nəşr. 1961, Moskva, Xarici və Milli lüğətlər üzrə Dövlət Nəşriyyatı” (5000-ə yaxın söz) (www.Poliqnot.az).

Qeyd etmək lazımdır ki, “Kitab.az” saytının qızıl onluğunda “Poliqnot” elektron lüğətlər sistemi 5-ci yerdədir (12.02.2013-cü ilin məlumatı əsasında).

3. Azərbaycan dilinin orfoqrafiya lüğəti (elektron variant). Elektron variant “Azərbaycan dilinin orfoqrafiya lüğəti” (Bakı, “Lider nəşriyyatı”, 2004. – 728 s.) kitabı əsasında hazırlanmışdır. Redaktor və ön sözün müəllifi AMEA-nın müxbir üzvü, filologiya elmləri doktoru, professor Ağamusa Axundov, tərtibçilər filologiya elmləri doktorları İsmayıl Məmmədov və Aydın Ələkbərov, filologiya elmləri namizədləri Bəhrüz Abdullayev və Nəriman Seyidəliyev, elmi işçi Nigar Xəlifəzadədir. Orfoqrafiya lüğətindəki texniki işlərin yerinə yetirilməsi, yeni sözlərin toplanması, əlifba sırasının yoxlanması və sözlərin sayılmasında kömənin əməkdaşı, baş laborant Əminə Həbibzadə tərtibçilərə kömək etmişdir (Azərbaycan dilinin orfoqrafiya lüğəti, 2004, s. 6). Qeyd olunan kitab Azərbaycan Respublikası Nazirlər Kabinetinin 05 avqust 2004-cü il tarixli qərarı ilə təsdiqlənmiş “Azərbaycan dilinin orfoqrafiya qaydaları” əsasında latın qrafikalı əlifba ilə hazırlanmış ilk nəşrdir. Lüğətə əvvəlki nəşrdə olmayan 18 minə yaxın söz əlavə olunmuşdur və 80000-ə yaxın sözü əhatə edir. Elektron variant filologiya elmləri doktoru, professor Məsud Mahmudov və texnika elmləri üzrə fəlsəfə doktoru, dosent Əbülfət Fətullayev tərəfindən hazırlanmışdır və internetdə yerləşdirilmişdir. (<http://www/dilmanc.az/az/mode/164>).

4. Azərbaycan dilinin əks-əlifba lüğətinin elektron variantı 2004-cü ildə “Nurlan” nəşriyyatı tərəfindən çap olunmuş “Azərbaycan dilinin əks-əlifba lüğəti” əsasında hazırlanmışdır (Azərbaycan dilinin əks-əlifba lüğəti, 2004).

Azərbaycan dilinin əks-əlifba lüğəti orfoqrafiya lüğətinin yeni sözlüyü əsasında tərtib olunmuşdur və təxminən 80000-ə yaxın sözü əhatə edir. Elektron variantın tərtibçiləri filologiya elmləri doktoru Məsud Mahmudov və texnika elmləri üzrə fəlsəfə doktoru Əbülfət Fətullayevdir.

İnternetdə Azərbaycandilli lüğətlərin bibliografyası ilə bağlı materialda elektron lüğətlər barədə də məlumat verilmişdir (müəllif U.Nəsibbəyli). Həmin materialda yuxarıda sadalanan lüğətlərdən əlavə daha bir neçə elektron lüğət barədə məlumat verilmişdir:

1. AzerDict – Azərbaycanın ən böyük pulsuz onlayn Azərbaycanca-İngiliscə, İngiliscə-Azərbaycanca lüğəti (<http://azerdict.com>);
2. Intelsoft – rus dilindən Azərbaycan dilinə tərcümə sistemi (<http://translate.intelsoft.az>);
3. Linqvosoft – İngiliscə-azərbaycanca, azərbaycanca-İngiliscə, rusca-azərbaycanca, azərbaycanca-rusca lüğətlər sistemi (<http://www.linqvozone.com/> Linqvosoft-Online-English-Azeri-Dictionary).
4. Google Translate - İngiliscə-azərbaycanca, azərbaycanca-İngiliscə məşin tərcüməsi (<http://translate.com/?hl=en#/az/>);
5. Azərbaycanca-türkcə sözlük (<http://www/azerice.da.ru>);
6. AMEA Terminologiya Komissiyası. 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-İngiliscə-azərbaycanca baytarlıq terminləri lüğəti; izahlı tarix terminləri lüğəti; Soyuq silahın kriminalistik ekspertizası üzrə xüsusi terminlərin izahlı sözlüyü; Rusca-azərbaycanca arıçılıq terminləri lüğəti (<http://www.azterm.az>).

Onlayn və endirilə bilən elektron lüğətlər:

1. Maliyyə bazaları terminlərinin izahlı lüğəti ([www.scs.gov.az /upload/Dictionary.pdf](http://www.scs.gov.az/upload/Dictionary.pdf));
2. Torpaqların səhrələşməsi üzrə izahlı lüğəti (www.un-az.org/doc/az/sahralasma.pdf);
3. Təhsil lüğəti (www.cil.az/library/kitab.TL.pdf);
4. Azərbaycanca-İngiliscə-rusca frazeoloji lüğət (<http://yazikilitra.narod.ru/nik.htm>);
5. Günəş enerjetikası terminləri lüğəti (<http://library.adau.edu.az/view.php?Lang=az.&menu=book&id=139>);
6. İngiliscə-rusca-azərbaycanca aqrar istehsal və maşınqayırmağa aid terminlər lüğəti (<http://library.adau.edu.az/view.php?lang=az&menu=book&id=71>); (<http://www.dil-manc.net/3|post|2011|03>).

IV FƏSİL. MƏTNİN FORMAL TƏHLİLİ SİSTEMİ AZƏRBAYCAN DİLİNİN MİLLİ KORPUSUNUN TƏRKİB HİSSƏSİ KİMİ

4.1. Mətnin formal təhlili sisteminin Azərbaycan dilinin milli korpusunda yeri

Artıq sübuta etiyacı olmayan bir həqiqətdir ki, dilçilik elminin gələcək inkişaf perspektivləri müasir kompüterlərin geniş tətbiqi olmadan mümkün deyil. Əks təqdirdə dilçilik, elmlərin ümumi inkişaf axınından kənar qalmış olardı.

Məhz müasir kompüter texnikasının geniş imkanları başqa elm sahələri ilə bərabər dilçilikdə də bir çox yeni istiqamətlərin, perspektivli sahələrin yaranması və inkişafına şərait yaratmışdır. Son zamanlar məlumat axtarışı sistemləri, maşın tərcüməsi, statistik leksikoqrafiya, tezaurus hazırlanması, mətnlərin avtomatik işlənməsi sistemləri sahəsində tədqiqatlar yeni məzmun kəsb etmiş, onların qarşısında yeni tələblər qoyulmuş, tətbiq sahələri genişlənmişdir. Dilçilik ədəbiyyatında mətnlərin avtomatik işlənməsi sistemləri ilə yanaşı ayrı-ayrı dillərin milli korpuslarının yaradılması ilə əlaqədar fikirlər də getdikcə daha geniş yayılmaqdadır (Бах: Гузев və başq., 1988; Касевич, Насилов, 1989 və s.).

Hər şeydən əvvəl mətnlərin formal təhlili sistemləri ilə milli korpusların əlaqəsi məsələsinə diqqət yetirək.

Türkologiyada mətnlərin formal təhlili sistemləri və milli korpuslar üçün ümumi sayıla biləcək xeyli tədqiqat işləri aparılmış və beləliklə, daha dərin, sistemli elmi axtarışlar üçün zəmin yaradılmışdır. Azərbaycan, türk, qazax, özbək, qaraqalpaq, qırğız, tatar dilləri üçün hesablayıcı texnikanın köməyi ilə tezlik, əks lüğətlər, tezauruslar hazırlanmış, bu dillərin bəziləri üçün analiz və sintez alqoritmləri, təcrübi səviyyədə azərbaycanca-rusca, türkcə-rusca, azərbaycanca-ingiliscə maşın tərcüməsi sistemləri hazırlanmışdır (Бах: Бектаев, 1978; 1986; Велиева, 1971; Пинес, 1970; Исхакова, 1972; Садыков, 1987; Мах-

мудов, 1991; Мамедова, 1984, Рахманов, 1988; Якубова, 1979 və s.). Bu tədqiqatlar istər mətnlərin formal təhlili sistemləri, istərsə də milli korpusların yaradılması üçün bünövrə rolunu oynaya bilər. Lakin tədqiqatçıların bir qrupu (В.Г.Гузев, Р.Г.Пиотровский, А.М.Шербак) belə “sistemsiz” şəkildə aparılmış tədqiqat işlərinin sonradan birləşdirilərək bütövlükdə məşin fondu (milli korpus) təşkil etmək yolunu məqbul hesab etmirlər. Onların fikrinə görə, belə konkret fraqmentlərin sonradan sistem yaratması prosesində bir çox qüsurlar müşahidə olunur ki, bu da “yuxarıdan aşağıya” prinsipi ilə işləmək zərurətini yaradır (Quzev və başqaları, 1988). Bütün bunlardan belə nəticə çıxartmaq olar ki, milli korpusa bu tədqiqat işləri heç də “birbaşa, qeyd-şərtsiz” daxil edilə bilməz. Mətnlərin formal təhlili sistemlərində də vəziyyət analoji olaraq qalır. İndiyə qədər məşin tərcüməsi, avtomatik sintez, təhlil və s. ilə bağlı ayrı-ayrı fraqmentlər kifayət qədər işlənib hazırlansa da, bütövlükdə problem həll olunmamış qalmışdır. Bu onunla şərtlənir ki, ayrı-ayrı fraqmentlər bütöv sistemdə heç də həmişə özünü doğrultmur və onların bir çox “naqis cəhətləri” bütöv sistemin işi zamanı aşkar olur.

Qeyd etmək lazımdır ki, milli korpus fondu mətnlərin avtomatik işənməsi sistemindən daha geniş anlayışdır və onu tam ehtiva edir.

1987-ci ildə Tbilisidə SSRİ xalqlarının milli dillərinin korpuslarının layihələrinin hazırlanması məsələlərinə həsr olunmuş müşavirədə bu məsələ ilə bağlı 27 məruzə və məlumat dinlənilmişdi. Müşavirədə linqvistik tədqiqatların avtomatlaşdırılmasının çoxaspektli sistemi kimi milli dillərin məşin fondunun (korpuslarının) yaradılmasının zəruriliyi qeyd edilmişdi.

Türk dillərinin materialları əsasında aparılmış tədqiqat işlərinin hazırkı səviyyəsi məsələyə bütöv sistem halında yanaşılması, fraqment təcrübəçiliyindən əl çəkmək və mühüm nəticələr əldə etmək imkanı yaradır. Bu baxımdan mətnlərin formal təhlili sisteminin Azərbaycan dilinin milli korpusu (ADMK) problemləri çərçivəsində araşdırılması məqsədəuyğun olardı.

Şübhəsiz, ADMK-yə daxil olan bütün məsələlərin həlli qarşıya məqsəd qoyulmamışdır. ADMF böyük bir kollektiv tərəfindən uzun illər ərzində yaradıla bilər. Belə sistem açıq olmalı, daim zənginləşməli, təkmilləşdirilməli, yeni-yeni lüğətlər, ədəbi-bədii nümunələr hesabına genişləndirilməlidir. Ona görə də, ADMK daim dəyişməkdə, inkişafda olan bir sistem kimi kompüterlərin texniki imkanları artdıqca yeniləşməli, dövrün tələbinə uyğun, ictimai sifarişə müvafiq bloklar əlavə olunmalıdır. Mövcud bəzi bloklar müvəqqəti “ehtiyatda” saxlana, dondurula bilər. Bu müddəanı əsas tutaraq mətnin formal təhlil sistemi ADMK-nin tərkib hissəsi kimi tədqiqata cəlb olunmuş, problem sistem halında araşdırılmışdır.

4.2. Azərbaycan dilinin milli korpusunun linqvistik bölmü

ADMK-nin linqvistik bölmü aşağıdakı bloklardan təşkil oluna bilər:

- fonetik (fonoloji) blok;
- morfoloji blok;
- morfonoloji blok;
- leksik blok;
- sintaktik blok;
- semantik blok;
- müstəsna hallar bloku.

Sistem açıq olduğu üçün zərurət olduqda blokların sayı artırıla və ya azaldıla bilər.

Fonetik (fonoloji) blokda dilin fonem tərkibi, ahəng qanunu, assimlyasiya, dissimlyasiya, vurğu, fonemlərin distribusiyası və statistikasısı və s. məlumatlar toplanır.

Morfonoloji blokda morfem birləşmələrinin sərhədində baş verən fonetik dəyişikliklər, morfemlərin distribusiyası və statistikasısı verilir.

Morfoloji blokda morfemlər sözlüyü, sözdüzəldici şəkilçilər, sözdəyişdirici şəkilçilər, birhecalı sözlər, morfoloji sintez və təhlil alqoritmləri, sözdüzəltmə və sözdəyişdirmə modelləri,

morfem ardıcılığı, nitq hissələri, onların təsnifi və s. məsələlər əhatə olunur.

Leksik blok leksikoqrafik arayışları, müxtəlif məqsədli lüğətləri (orfoqrafiya, orfoepiya, izahlı, terminoloji, tezlik, əks, avtomatik, dialektoloji, tarixi, etimoloji, onomastik, şəxs adları, ikidilli və çoxdilli (azərbaycanca-rusca, ingiliscə-almanca və s), ədəbi-bədii nümunələri (illüstrativ mətnlər), yazılı abidələri (Kitabi-Dədə Qorqud) və s. ehtiva edir.

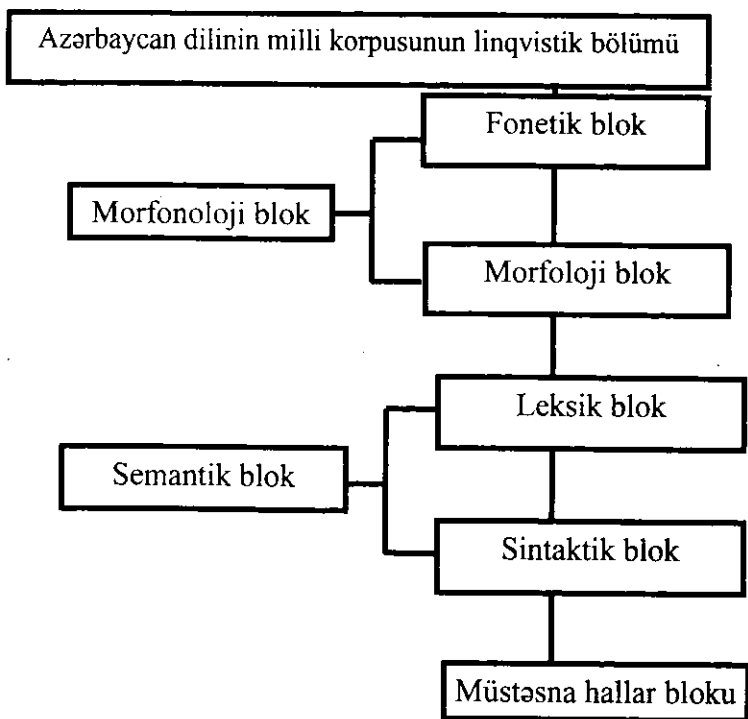
Sintaktik blokda sadə cümlə tipləri, mürəkkəb cümlə tipləri, cümlədə söz sırası, cümlə üzvlərinin ənənəvi yeri, söz birləşmələrinin tipləri, sintaktik təhlil və sintez alqoritmləri, sintaktik nümunələr və s. əhatə edilir.

Semantik blokda morfoloji və sintaktik səviyyədə sintezi və təhlili mümkün olmayan söz və ifadələrin müxtəlif vasitələrlə araşdırılması, semantik əlamətlər, leksik və sintaktik-semantik münasibətlər, semantik təhlil və sintez alqortmləri və s. əhatə olunur.

Müstəsna hallar blokuna istisna təşkil edən fonetik (fonoloji), morfoloji, morfonoloji, sintaktik qaydalar, istisna təşkil edən sözlərin siyahısı, müstəsna hallar blokuna müraciət alqoritmləri daxil olur.

ADMK-nin linqvistik bölümünü şərti olaraq I sxemdəki kimi göstərmək olar.

Əlbəttə, tədqiqatda ADMK-nin tam şəkildə işlənilib hazırlanması prinsiplərinin araşdırılması qarşıya məqsəd qoyulmamışdır. Tədqiqat obyektini olaraq ADMK-nin tərkib hissəsi - mətnin formal təhlili sistemi götürülmüşdür. Müasir dil tədqiqatlarının mühüm istiqamətlərindən olan mətnlərin avtomatik işlənməsi sistemləri Azərbaycan dilçiliyində müxtəlif baxımlardan araşdırılmış, onun ayrı-ayrı komponentləri ilə bağlı tədqiqatlar aparılmışdır. Ona görə də problemin sistemli şəkildə öyrənilməsi və ümumiləşdirilmələr aparılması, eləcə də dilçiliyimiz üçün əhəmiyyət kəsb edən bir çox cəhətlərin araşdırılması ön plana çəkilmişdir.



Sxem 1. Azərbaycan dilinin milli korpusunun linqvistik bölmü

Qrafem və fonem səviyyəsində Azərbaycan mətnlərinin statistik-distributiv təhlili ilə bağlı tədqiqat işlərinin olduğunu və mətnlərin avtomatik işlənməsi sistemlərində fonem səviyyəsində araşdırmalardan daha çox yuxarı dil səviyyələrinin tədqiqinin ön plana çəkildiyini nəzərə alıb, leksik, morfoloji, morfonoloji, sintaktik, semantik səviyyələrdə avtomatik işlənmə məsələləri üzərində dayanacaq (Ахундов, 1963; Рахманов, 1988).

4.3. Mətnin formal təhlili sisteminin ümumi məsələləri

4.3.1. Maşın lüğətləri

Mətnin formal təhlili sisteminin əsas komponentlərindən biri leksik blokdir. Sistemin tamlığı, etibarlılığı və dəqiqliyi daha çox onu təşkil edən lüğətlərdən asılıdır. Leksik blok təkcə leksikoqrafik məlumatlar toplusu deyil. Bu blok elə qurulmalıdır ki, maşında mətnlərin tam və hərtərəfli işənməsi üçün zəruri məlumat vermiş olsun. Dilin müxtəlif səviyyələrdə araşdırılması (fonetik, morfoloji, morfonoloji, leksik, sintaktik, semantik) maşına daxil edilmiş lüğətlərin optimal quruluşu, kamilliyi ilə şərtlənir. Maşın lüğətləri adi lüğətlərdən fərqli olaraq həm də dilin qrammatikasını əvəz edir.

Mətnin formal təhlili sistemində maşın lüğətləri aşağıdakı məqsədlərə xidmət edir:

- sorğuya görə saxlanılan informasiyanı (bu informasiya dilin müxtəlif səviyyələrinə aid ola bilər) və ya onun yerləşdiyi ünvanı, kodunu verir;
- bir dildən başqa dilə (Azərbaycan dilindən başqa dillərə və əksinə) leksik-morfoloji səviyyədə tərcümə nümunələri almağa imkan yaradır;
- məlumat-axtarış sistemləri üçün zəruri məlumatların əldə edilməsinə ilkin şərait yaradır;
- müxtəlif təyinatlı linqvistik proseslər (avtomatik təhlil və sintez və s. bu qəbildən olan əməliyyatlar) üçün təminat hazırlayır.

Avtomatik işlənmə sistemlərində eyni vaxtda bir neçə müxtəlif lüğət verilə bilər. Avtomatik işlənmə prosesinin müxtəlif mərhələlərində zəngin leksikoqrafik məlumatlar müxtəlif məqsədlərlə ayrı-ayrı dil səviyyələrində istifadə oluna bilər.

Mütəxəssislərin fikrincə, maşın lüğətlərindən üç əsas məqsədlə istifadə oluna bilər:

- informasiyanın (leksik, qrammatik, semantik məlumatın) məşində istifadəsi üçün yararlı və hazır vəziyyətdə qorunub saxlanması;
- yığcam informasiyadan müxtəlif məqsədlər üçün istifadə;
- translyasiya (Марчук, 1983).

Maşının yaddaşında yerləşdirilmiş eyni lüğət müxtəlif funksiyalar yerinə yetirə bilər.

Mətnin formal təhlili sistemində istifadə olunan maşın lüğətləri adi lüğətlərdən fərqli xüsusiyyətlərə malikdirlər. Belə ki, maşın lüğətləri istifadə prosesində dəqiqləşdirilə və tamamlana bilər. Ayrı-ayrı sözlərlə birlikdə söz-formalar, ən çox işlənən ifadələr də maşın lüğətlərində vahid kimi götürülə bilər. Qeyd etmək lazımdır ki, bu və ya digər elm sahəsinə aid mətnlərin işlənməsi prosesində sahə terminalogiyasından istifadə olunmaqla yanaşı, ən çox işlənən ümumişlək leksika da öz əksini tapmalıdır. Formal sistemlər üçün maşın lüğətləri tərtib olunan konkret hesablama texnikasının yaddaşının həcmi, lüğətlərdə sözlərin axtarılma tezliyi (sürəti), tərtib olunacaq proqramın mürəkkəbliyi də nəzərə alınmalıdır. Göstərilən amillər heç də nəzərə alınmayacaq dərəcədə əhəmiyyətsiz hesab olunmamalıdır. Bütün bunlar lüğətin quruluşuna, sözlərin orada yerləşdirilməsinə təsir göstərir.

Formal təhlil sistemləri üçün maşın lüğətlərinin tərtibi zamanı üç mühüm cəhətə xüsusi diqqət yetirilmişdir:

- sözlərin (sözlüyün) seçilməsi;
- sözlərin (lüğət vahidlərinin) yerləşdirilməsi üsulu (informasiyanın sürətlə və yığcam əldə olunması üçün);
- alqoritm və proqramın daha sadə və optimal tərtibi məsələlərinin bir çox xüsusiyyətlərinin nəzərə alınması.

Yeni, mükəmməl avtomatik lüğətin tərtibi dilin tam, yeni nəzəri müddəalara əsaslanan modelinin qurulması deməkdir. Bu, dilə münasibətin və ondan istifadənin yeni baxış bucağı altında tədqiqi ilə bağlı bir çox maraqlı linqvistik məsələlərin həllini ön plana çəkir.

Şübhəsiz, müəyyən bir tədqiqatçı və ya tədqiqatçılar qrupu tərəfindən hazırlanmış maşın lüğətləri fərdi xüsusiyyət daşıyır. Buna səbəbdən belə lüğətlərin konkret xüsusiyyətləri qarşıya qoyulmuş məqsədlə səciyyələnir.

Mətnin formal təhlili sistemini təşkil edən lüğətlərin təkcə quruluş xüsusiyyətləri deyil, say tərkibi və həcmi də linqvistik tədqiqata müvafiq dəyişilə bilər.

Maşın lüğətlərinin əsasını tezlik lüğətləri təşkil edir. İstər müxtəlif sahələr üzrə, istərsə də ümumişlək leksika da daxil olmaqla dilin tam tezlik lüğətinin tərtibi bir çox linqvistik problemlərin həllini asanlaşdırır (Алексеев, 1976). Təhlil sistemində ikidilli (və ya çoxdilli) tərcümə lüğətləri, əks lüğətlər, terminoloji sahə lüğətləri, sintaktik birləşmələri əhatə edən lüğətlər (köməkçi fellər nəzərə alınmaqla), semantik lüğətlər (sintaktik əlaqələrə əsaslanmaqla) ehtiva olunur.

Maşın lüğətlərinin riyazi modelləri ideal lüğət strukturuna əsaslanır. Ən mühüm və vacib cəhətlər qabarıq nəzərə çatdırılır. İkinci dərəcəli xüsusiyyətlər ya nəzərə alınmır, ya da qismən əhatə olunur. Semantik blokun dolğun fəaliyyəti üçün müxtəlif tezaurların tərtibi də mühüm əhəmiyyət kəsb edir. Tədqiqat işində formal təhlil sistemlərinin mühüm tərkib hissəsi olan müxtəlif səpkili maşın lüğətləri barədə sonrakı fəsildə geniş məlumat veriləcək. Məlum olduğu kimi, maşın lüğətləri kororanə tərtib olunmur, müəyyən linqvistik məqsədə xidmət edir. Ona görə də, leksik blokda yerləşdirilmiş maşın lüğətlərindən danışmadan əvvəl formal təhlil sistemlərinin bəzi ümumi istiqamətlərinə diqqət yetirək.

4.3.2. Avtomatik linqvistik təhlil və sintez

Formal təhlil sistemlərində mühüm komponentlərdən biri avtomatik linqvistik təhlil (ALT) və sintezlə bağlı problemlərdir. Müxtəlif dil səviyyələrində ALT-nin sonrakı fəsillərdə şərh olunacağını nəzərə alıb bu problemlə bağlı ümumi cəhətləri nəzərdən keçirək.

ALT üç mühüm blokdan ibarətdir: morfoloji, sintaktik, semantik.

Qeyd etmək lazımdır ki, maşın tərcüməsi (MT) və avtomatlaşdırılmış məlumat axtarışı sistemlərinin konkret tələblərinə uyğun olaraq qurulmuş bu sistemdə fonetik blok tamamilə nəzərə alınmamış, morfonoloji və müstəsna hallar bloku isə qismən verilsə də, ayrılıqda təmsil olunmamışdır. Linqvistik baza rolunu oynayan leksik blok təhlildən tamamilə təcrid olunduğundan nəzərdən keçirilir.

ALT-nin morfoloji bloku söz-formaların və sözdəyişdirici şəkilçilərin təhlili mərhələləri ilə fəaliyyətə başlayır (Mahmudov, 1981). Sonra morfoloji səviyyədə əsas və söz formaların omonimliyinin aradan qaldırılması həyata keçirilir. Əgər omonimlik morfoloji səviyyədə aradan qaldırıla bilmirsə, sintaktik və semantik bloklara müraciət olunur.

Sintaktik blokda təhlil prosesi aşağıdakı kimi gedir: əvvəlcə mübtədə və xəbər əlamətlərinə görə axtarılıb tapılır. Sonra köməkçi sözlərin və durğu işarələrinin təhlili aparılır, sözlər arasında sintaktik əlaqələr müəyyənləşdirilir.

Semantik blokda iş daha mürəkkəbdir. Digər bloklarda omonimlik və çoxmənalığın müəyyənləşdirilməsi nəticəsində qarşıya çıxan problemlər burada öz həllini tapır. Təhlil prosesində alınmış naməlum sözlərin təsnifi də bu blokun üzərinə düşür. Bundan başqa, sintaktik blokda mübahisə doğuran bir çox problemlər (məsələn, söz-formalar arasında sintaktik əlaqələrin dəqiqləşdirilməsi) semantik blokun fəaliyyəti olmadan mümkün deyil. Tam linqvistik təhlil həyata keçirildikdən sonra sintez alqoritmləri reallaşdırılır. Qeyd etmək lazımdır ki, sintezin reallaşdırılması fonetik və morfonoloji səviyyədə kifayət qədər məlumatın maşının yaddaşında olmasını tələb edir.

4.3.3. Maşın tərcüməsi

Formal təhlil sistemlərindən danışarkən MT məsələlərinə toxunmamaq mümkün deyil. Problemin həlli ilə bağlı müxtəlif dövrlərdə 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 probleminə aid külli miqdarda ədəbiyyat (Вах: АП, 1971; Герасимов, Марчук, 1975; Котов, 1976; Чхайдзе, 1976; Пиотровский, 1975; Марчук, 1983; Махмудов, Пинес, 1989 və s.) olduğu üçün məsələnin 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 mənaəmləyəgətirmə 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.

Avtomatik tərcümənin mümkünlüyü ideyası ilk dəfə 1946-cı ildə Amerika alimləri A.But və U.Uiver tərəfindən söylənmışdir (Кулагина, 1979, 13). 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 olunandan 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 bilmədiyi 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 və hazırlanması ilə bağlı araşdırmalar, insanın beyin fəaliyyəti ilə maşın 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 do-

ğurur. Müxtəlif idomatik 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) məşinlərin 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 məşinin 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 məşin 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. Məşinə 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 lüğətdə 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 mənə yazılışına və universal sintaktik münasibətlərə malikdir.

MT probleminin həlli ilə əlaqədar tək-cə 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 görülmə işlərinin 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 müddəatı əsaslandırmağa çalışırlar ki, hazırda hətta elmi-texniki ədəbiyyatın belə tamamilə avtomatlaşmış, 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 ekstralingvistik 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» (Инҟве, 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 olma-

- lıdır. 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 olması fikrini irəli sürürlər. Onlar belə hesab edirlər ki, MT üçün zəruri olan linqvistik məlumatlar nəzəri olmaqdan da çox təcrübədirlər və konkret məsələlərin həlli zamanı alına bilərlər. Belə ki, keçmiş ümumittifaq «Nitq statistikası» qrupunun rəbəri 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) «Nitq statistikasını» qrupunun MT-nin reallaşdırılması

proqramı hazırlanmışdır. Həmin proqramı aşağıdakı mərələlərə 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).

4.3.4. Yarımavtomatik tərcümə

Qeyd etmək lazımdır ki, yarımavtomatik tərcümə də böyük maraq və əhəmiyyət kəsb edir. Mətnlərin avtomatik işlənməsi sistemləri ilə əlaqədar yarımavtomatik tərcümə məsələlərinin araşdırılması çox vacibdir. Yarımavtomatik tərcümədə kompüterlər köməkçi vasitə rolu oynayır, tərcüməçini ağır, az məşhuldar və yaradıcı olmayan işlərdən azad etmək üçün maşınların köməyi ilə tərtib olunmuş xüsusi lüğətlərdən istifadə olunur. Proses aşağıdakı kimi həyata keçir: tərcüməçi məndəki 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 dərinləşməsi ilə yaradılmış sistem sözün və onun daxil olduğu kontekstin maşın lüğətində axtarışına imkan verir. Məsələ orasıdır ki, sözün tam, dolğun və dəqiq mənası kontekstdə məlum olur və istər adi, istər yarımavtomatik tərcümədə bu amili nəzərə almamaq qeyri-dəqiq nəticələr alınmasına aparıb çıxarır. Tərcüməçi məndə 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 lazımi 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.

4.3.5. 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 hərtərəfli semantik təhlili tələbi ilə üzvləşir. 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 mümkünliyünü inkar edirlər.

Parodoksal haldır ki, bəzən süni intellektin qarşısına ağlasığmaz tələblər qoyulur. İnsan intellektinin həll edə bilmədiyini və ya həll etməyə cəhd göstərdiyi məsələlərin ortimal 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 üçün də 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, freymlər, 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 və tərcümə etməyə cəhd göstərir. Tərcümə zəif alınanda, başa düşülməyəndə yenə insan-tərcüməçiyə müraciət edir, tərcümənin uğursuz alınmasının səbəblərini araşdırırıq. Bir vaxtlar, maşın tərcüməsinin təşəkkül dövründə sözbəsöz tərcümədən istifadə olunurdu. Məsələn: «tanış etmək», «tanış» – «знакомый», «etmək» – «делать», «сделать» və s. Sonra məlum oldu ki, insan bu tipli ifadələri bütövlükdə, söz birləşməsi kimi tərcümə edir: «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 aprobeasiyası mühüm əhəmiyyət kəsb edir və tərcümənin sərhədləri bir qədər də genişlənməmiş olur. Sözü mənası cümlədə tam aydınlaşdığı kimi, cümlə də mətn səviyyəsində daha dəqiq, daha düzgün anlaşılır.

4.4. 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, 1970; Барабанов, 1981; Байрамова, 1966; Бектаев, 1978; 1986; Велиева, 1971; Джубанов, 1987; Исхакова, 1979; Кенесбаев, Махмудов, 1991; Пиотровский, Бектаев, 1970; Халитова, 1964; Садыков, 1987 və s.)

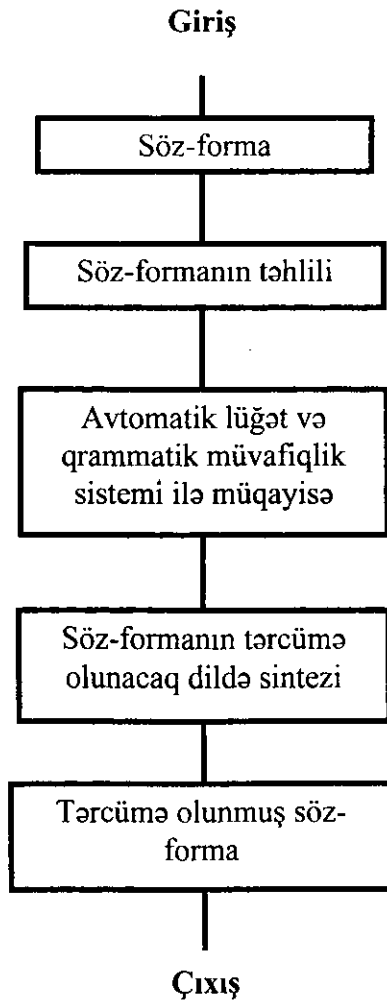
Yuxarıda qeyd edildi ki, istər avtomatik lüğətlər, istər kompüterlərdə istifadə olunmuş digər lüğətlər müxtəlif məqsədlərə

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 məşində «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 (sxem 2.)

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 şəkilçilərə ayrılır.

Əsas dedikdə kök və ya kök birləşməsi ilə sözdüzəldici şəkilçinin birləşməsi başa düşülür. Mövcud tədqiqatlarda MT və təhlil sistemlərinin tələblərinə uyğun olaraq sözün əsası daha kiçik mənalı vahidlərə ayrılmaz, çünki əsas ayrılıqda leksik vahid kimi bütün lüğətlərdə təmsil olunur. Lakin bu o demək deyildir ki, əsası daha kiçik vahidlərə ayrılması qeyri-mümkündür və ya elmi maraq doğurmur. Sözdüzəldici şəkilçilərin 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 2. 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.

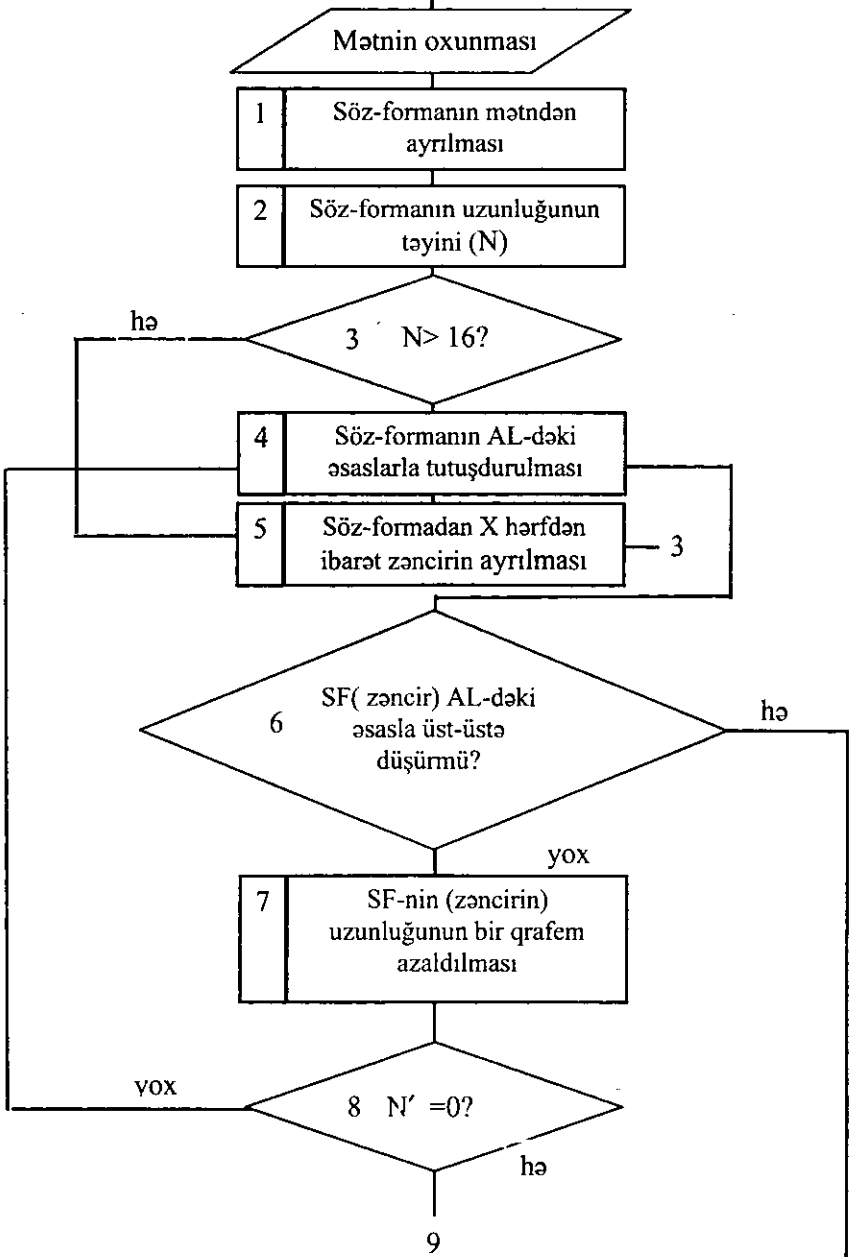
Təhlil alqoritmlərinin köməyi ilə həmin söz-forma «məktəbli», «lər» və «in» hissələrinə ayrılır. «Məktəbli» sözü lüğətdə (avtomatik lüğətdə) təmsil olunduğu üçün onun mənası tapılır və üzərinə şəkilçilər cədvəlində təsbit olunmuş «lər» və «in» şəkilçilərinin qrammatik əlamətləri əlavə olunur. MT çərçivəsində «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 məşinin «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 çətinlik və maneələrin hamısını əhatə etmək mümkün deyil. Ayrı-ayrı blokların işi ilə əlaqədar bu problemlərə bir daha toxunulacaq. Məqsəd leksik, morfoloji, morfoloji, sintaktik-semantik səviyyələrdə təhlil sistemləri işlənilib hazırlanarkən hansı detalların nəzərə alınıb-alınmaması məsələləri barədə ümumi təsəvvür yaratmaqdan ibarət idi.

MT problemi indi daha aktual məzmun kəsb etmişdir. 1983, 1985, 1989-cu illərdə maşın tərcüməsinin müxtəlif məsələlərinə həsr olunmuş beynəlxalq seminar və konfranslar bunu 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 (Барабанов, Махмудов, Садыков, 1979; Велиева, Махмудов, Пинес, 1985; Велиева, Мамедова, 1985; Велиева, Мамедова, Махмудов, Пинес, Рахманов, 1985; Махмудов, 1980, 1981, 1982, 1985; Махмудов, Пинес, 1983, 1989; Пинес, Махмудов, 1983).

İstər ADMF, istərsə də formal təhlil sistemlərindən asılı olaraq maşının leksik bloku qurulur. Maşına daxil edilən lüğətlərin riyazi modelləri ideal lüğət strukturlarına əsaslanır. Tədqiqatı maraqlandıran mühüm və vacib cəhətlər qabarıq nəzərə çarpdırılır. İkinci dərəcəli xüsusiyyətlər ya nəzərə alınmır, ya da qismən nəzərə alınır. Göstərilən məsələlərin geniş şərhli formal təhlil sisteminin leksik bloku fəslində verilir.

Sxem 3-də leksik-morfoloji səviyyədə azərbaycanca-rusca maşın tərcüməsinin ümumi blok sxemi göstərilmişdir. Sxemdə N - əsasların, n - şəkilçilərin, X - söz-formaların, L - qalığın uzunluğunu göstərir.

Giriş



9	SF-nin AL-də olmayan əsaslar siyahısına yazılışı	30
---	--	----

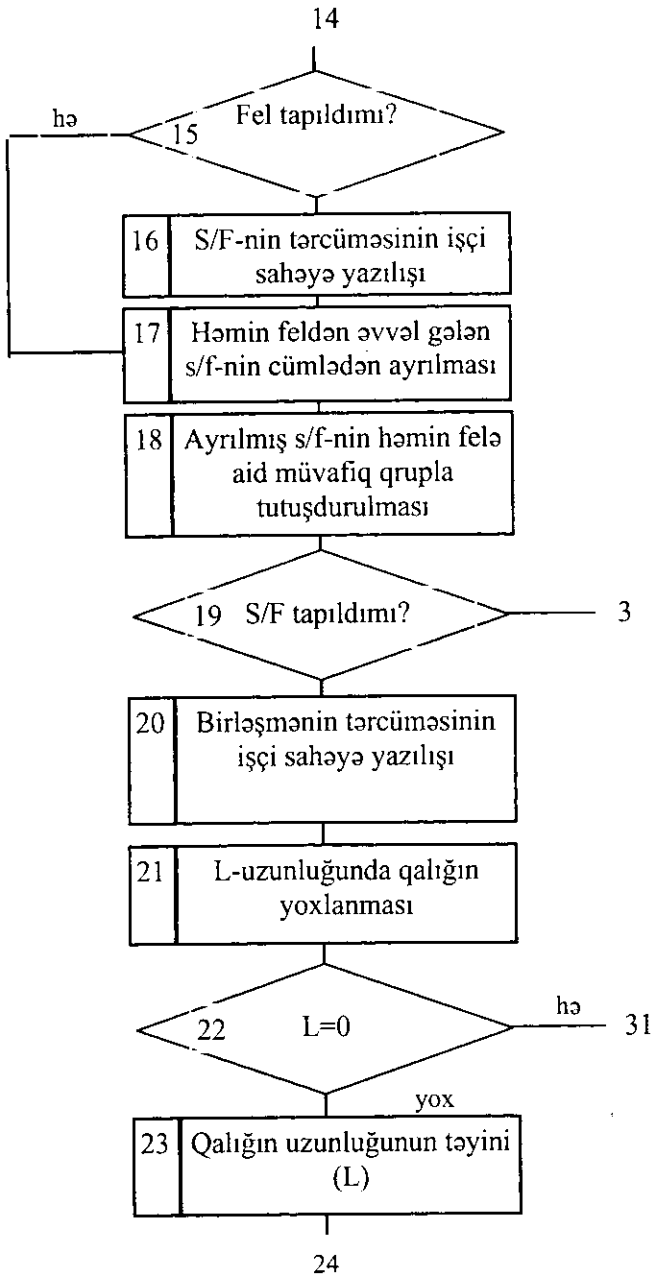
10	Yoxlamalı əsas omonimdirmi?	yox
----	-----------------------------	-----

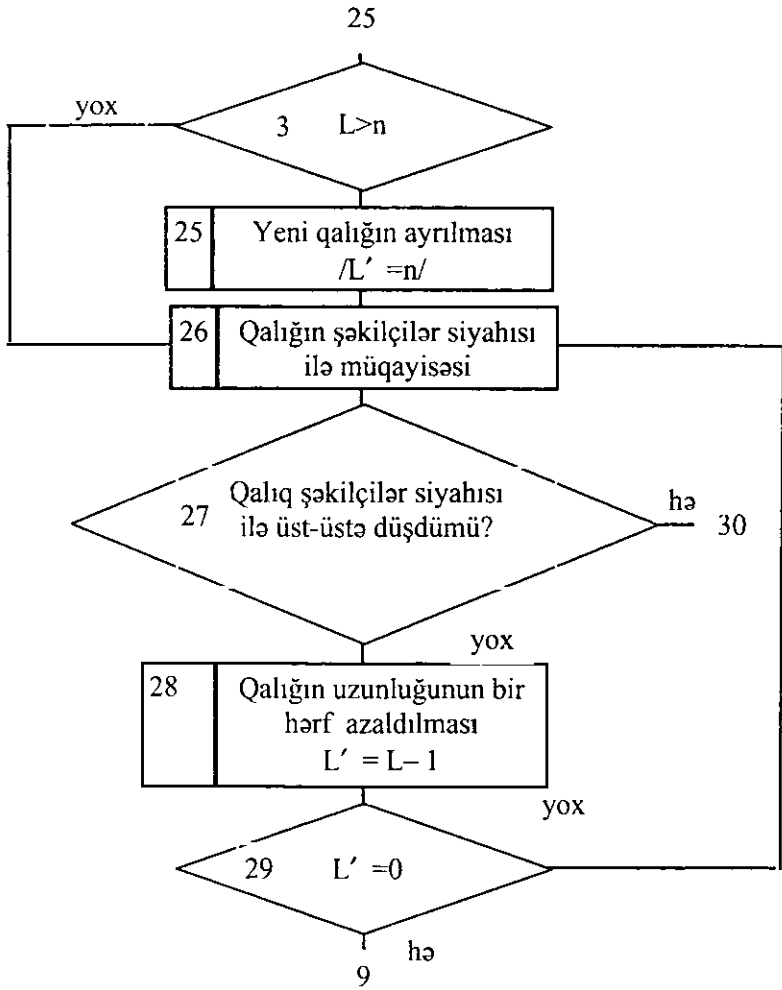
11	Əsasların omonimliyinin aradan qaldırılması blokuna müraciət
----	--

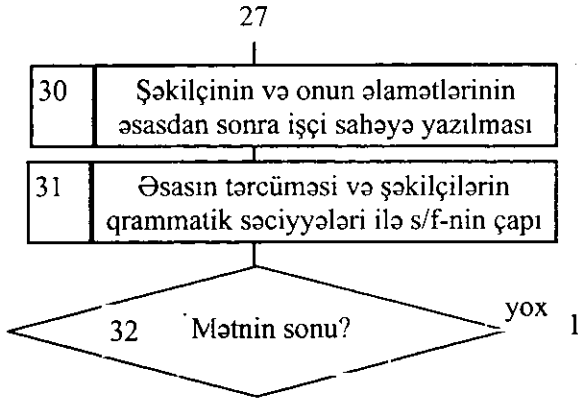
12	Şəkilçi omonimdir?	
----	--------------------	--

13	Şəkilçilərin omonimliyinin aradan qaldırılması blokuna müraciət
----	---

14	S/F-nin fellər siyahısı ilə tutuşdurulması
----	--







Sxem 3. Leksik-morfoloji səviyyədə azərbaycanca-rusca məşın tərcüməsinin ümumi blok sxemi

V FƏSİL. MƏTNİN FORMAL TƏHLİLİ SİSTEMİNİN LEKSİK BLOKU

5.1. Leksik blok mətnlərin formal təhlili sisteminin aparıcı komponenti kimi

Formal təhlil sistemlərinin hansı səviyyədə, necə işləməsi leksik təminatdan, leksik bloku təşkil edən lüğətlərin strukturundan, həcmindən, leksik-qrammatik-semantik informasiyanın verilməsi üsulu və əhatəliliyindən çox asılıdır. Təsadüfi deyil ki, leksik blok sistemin aparıcı, əsas bölmələrindən sayılır. Leksik blok sistemin gələcək «fəaliyyətini», müxtəlif dil səviyyələrində təhlil və sintezin aparılması və tənzimlənməsi işlərini müəyyənləşdirir. Sistemin etibarlılığı, dəqiqliyi onun leksik blokunun məzmunu ilə səciyyələnir. Məsələn, əgər sistem kifayət qədər sözlüyə malik deyilsə, lazımi səviyyədə leksik-qrammatik-sintaktik informasiya ilə təmin olunmamışsa, onda, şübhəsiz ki, avtomatik təhlil və sintez, tərcümə və s. aşağı səviyyədə olacaq, bütövlükdə sistemin fəaliyyəti sifarişçini təmin etməyəcək.

Leksik blokun təşkilində bir çox mühüm amillər nəzərə alınmalıdır:

- maşına daxil edilmiş lüğətlərlə avtomatik təhlil, sintez və tərcümə alqoritmlərinin nisbəti; belə ki, göstərilən alqoritmlərin tərtibi, fəaliyyəti, optimallığı lüğətlə müəyyənləşir;
- müxtəlif növlü maşın lüğətlərinin tipologiyası;
- çoxsəviyyəli iyerarxiya kimi dilin kommunikativlik xüsusiyyəti;
- tərtib olunmuş alqoritmlərin və lüğətlərin müstəqil və bir-birindən asılı olmayaraq mövcudluğu. Bu, çox vacib şərt-dir. Sistemin fəaliyyəti zamanı lüğət təkmilləşdirilir, həcmi artır və ya azalır, bir çox yeni əlamətlər (leksik-qrammatik-sintaktik) əlavə edilir və bütün dəyişikliklər alqoritmə təsir göstərmir. Eləcə də təhlil, sintez, tərcümə və s. alqoritmləri iş prosesində tədqiqatın məqsədindən asılı

olaraq dəyişdirilir, bu da öz növbəsində lüğətin tamamilə yenidən tərtibi ilə nəticələnmişdir.

Optimal sistemlərdə bütöv bir sistemin tərkib hissələri kimi lüğətlər qabaqcadan elə qurulur ki, alqoritmlərlə həmahəng olsun. Qeyd etmək lazımdır ki, sistemin açıqlığı, istənilən mərhələdə təkmilləşdirilə bilməsi onun qiymətli xüsusiyyətidir (Зорев, 1973; Марчук, 1985; Пиотровский, 1977).

Lüğət vahidi ayrılıqda mücərrəd və natamamdır. Onun tam mənası sintaktik əhatədə açılır, məlum olur. Bu baxımdan adi lüğətlərdən fərqli olaraq leksik bloku təşkil edən sözlüklərdə sintaktik əhatənin və əlaqələrin nəzərə alınması və bunların əksi çox müüm xüsusiyyətdir.

Leksik blokun geniş izahına keçməzdən əvvəl bu sahədəki bəzi mühüm perspektivləri nəzərdən keçirək. Maşın lüğətləri təkcə MT problemi və ya avtomatik təhlil və sintezə xidmət etmir. Dil barədə bir çox ensiklopedik səciyyəli məlumatlar bu lüğətlərdə ehtiva olunur. Gələcəkdə maşın lüğətləri adi ikidilli və ya çoxdilli, terminoloji, etimoloji, dialektoloji, izahlı və s. Lüğətlərin tərtibinin avtomatlaşdırılmasında da mühüm rol oynaya bilər. Tezaurusların təhlil sistemlərində istifadə olunması imkanları və perspektivləri də çox böyükdür (Никитина, 1978). Azərbaycan dilçiliyində tezaurusların tərtibi və tətbiqi məsələlərinin praktik və nəzəri həlli ilə bağlı işlərə son vaxtlarda başlanılmışdır (Мамедова, 1984).

Mətnlərin formal təhlili sistemi ilə əlaqədar Azərbaycan dilinin kontekstoloji lüğətinin tərtibi də böyük əhəmiyyət kəsb edir. Belə lüğətlərdə sözlərin (lüğət vahidlərinin) işlənmə tezliyi, onların işlənmə dairəsi, mümkün işlənmə formaları, sintaktik əlamətləri və əlaqələri öz əksini tapır. Nəticədə adi ikidilli lüğətlərdən daha düzgün və dəqiq tərcümə nümunəsi alınır. Sözlərin mümkün birləşmə imkanları sadalandığı üçün onlar haqqında tam, sistemli təsəvvür yaranmış olur. Lüğət vahidlərinin qarşısındakı statistik məlumat onun hansı mənasının daha mühüm olduğunu və çox müşahidə olunduğunu göstərir (Марчук, 1983, 127).

5.2. Statistik lüğətçilik

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 (Пинес, Махмудов, 1987).

Dil bir çox mürəkkəb qarşılıqlı əlaqəli müxtəlif obyektlər sistemini özündə 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 metoddların məcmusu (lingvostatistika) nitqdə və mətnədə hansı vahidlərin tez-tez və ya çox az rast gələ bilməsi hallarını müəyyən etməyə imkan verir. Dil vahidlərinin işləkliyinin təhlili nitq ifadələrinin əmələgəlmə prosesləri üzrə müəyyən qanunauyğunluqların aşkara çıxarılmasını təmin edir.

5.3. 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 (Frumkina, 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 müəyyən edilən bir sıra xüsusiyyətləri dəqiq ölçülərlə göstərir. Tezlik lüğətləri təhlil sistemlərində və MT-də olan avtomatik lüğətlərin əsasını təşkil etdiyi üçün onun üzərində geniş dayanaq.

5.3.1. Tezlik lüğətlərinin növləri

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 mətndə rolu və mətnmələgətirmə imkanları müəyyən edilir.

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 üz 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 seçmələri 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 bir-birindən 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, maşınların köməyi ilə və ya qarışıq üsulla tərtib oluna bilər.

Tezlik lüğətlərinin tərtibi prosesi ümumi şəkildə aşağıdakı mərhələlərdən ibarətdir:

- mətndən 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).

Əgər əl üsulu ilə tərtibatda mətnin təhlili ilə bağlı bütün əməliyyatı (tezliklərin hesablanması və s.) tezlik lüğətini tərtib edənlərin özləri aparırsa, maşınların köməyi ilə lüğət tərtib edilərkən yalnız mətnlərin kompüterlərə daxil edilməsinə qədər görülən işlər bilavasitə insan əməyinin məhsulu olur. Doğrudur, mətnləri avtomatik olaraq birbaşa maşına daxil etməyə imkan verən qurğu mövcuddur. Ancaq bu qurğu o qədər də yüksək sürətə malik deyil. Ona görə də mətnlər çap qurğusu vasitəsilə kompüterə daxil edilir. Kompüterlərin tezlik lüğətlərinin tərtib

olunması işlərinə tətbiqi dilçiləri mexaniki, yorucu işlərdən azad etmiş olur.

Məsələn, adi qayda ilə 200 min söz-formadan ibarət seçmə mətn əsasında tezlik lüğəti tərtib etmək üçün tədqiqatçı ən azı iki il gərgin iş aparmalıdır. Halbuki orta gücə malik maşının köməyi ilə bu işin öhdəsindən iki ay yarımına gəlmək mümkündür. Həm də bu zaman vaxtın çox hissəsi mətnin kodlaşdırılmasına sərf olunur. Kompüterə verilmiş mətnləri maşın özü təhlil edir və bütün statistik göstəriciləri ilə sözlərin tezlik və s. əlifba siyahısını verir. Göründüyü kimi, leksikoqrafik məqsədlər üçün qənaətbəxş və iqtisadi cəhətdən daha əlverişli alqoritmlərin alınması metodlarının işlənilib hazırlanmasına xidmət edən tədqiqat işlərinin böyük əhəmiyyəti vardır.

Məlumdur ki, kompüterlərdə ilk tezlik lüğətini 1953-cü ildə C.Yosselson tərtib etmişdir. Ondan sonra istər xaricdə, istərsə də keçmiş SSRİ-də bir çox tezlik lüğətləri hazırlanmışdır. Sovetlər dövründə ümumittifaq «Nitq statistikası» qrupunun əməkdaşları prof. R.Q.Piotrovskinin rəhbərliyi altında maşınlardan istifadə etməklə çoxlu iriəcimli tezlik lüğətləri nümunələri hazırlamışlar. Latın və rus dillərinin tezlik lüğətləri tərtib edilərkən hesablayıcı texnikadan istifadə olunmuşdur. Müasir ingilis yazılı nitqinin hərtərəfli statistik təhlili isə ABŞ-da Q.Kuçera və U.Frensisin sanballı lüğətində öz əksini tapmışdır (Pines, Mahmudov, 1977). Bu tədqiqat zamanı həcmi bir milyon söz-formadan çox olan mətnlər üzərində iş aparılmışdır. Hər bir mətn perfokartlara vurulmuş və «İBM» tipli elektron hesablayıcı maşında işlənmək üçün maqnit lentlərinə köçürülmüşdür. Qeyd etmək lazımdır ki, yazılışın maqnit lentində uzun müddət saxlanması nəticəsində mətnlərə istənilən qədər müraciət etmək imkanı yaranır.

Təkcə leksikoqrafiyada deyil, ümumiyyətlə, lingvistik tədqiqatların avtomatlaşdırılması prosesində meydana çıxan bir məsələ də insanla maşının qarşılıqlı əlaqəsidir.

Məlum olduğu kimi, əvvəllər mətnlərin maşında işlənməsini təmin etmək üçün qabaqcadan müəyyən qaydalar əsasında

kodlaşdırılmış mətnləri perfokartlara, perfolentlərə və ya maqnit lentlərinə yazmaq lazım gəlirdi. Hazırda informasiya kompüterə birbaşa puldan daxil edilir və yardımçı vasitələrə ehtiyac qalmır.

Məlumatların maşına daxil edilməsi işini teleqraf aparatı tipli qurğuda işləyən operator yerinə yetirir. Beləliklə, elə həmin mərhələdə biz «insan-maşın-insan» sistemində iş görmüş oluruq.

Mətn maşına daxil edildikdən sonra kompüter alqoritmdə, yəni qabaqcadan maşına daxil olunmuş təlimatlarda nəzərdə tutulmuş istənilən əməliyyatı yerinə yetirə bilər.

5.3.2. Tezlik lüğətlərinin əhəmiyyəti və tətbiqi

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.

Lüğət vahidlərinin azalan tezlik siyahısına görə düzülüşü müxtəlif həcmli mətnlərin müqayisəli tədqiqi üçün istifadə oluna bilər. Bununla əlaqədar sözün tezliyi ilə onun rəqəmi (tezlik siyahısındakı nömrəsi) arasındakı asılılığı müəyyənləşdirməyə imkan verən qayda xüsusi olaraq diqqəti cəlb edir. Bir neçə lüğətin təhlili əsasında C.Sipf müəyyənləşdirmişdir ki, sözün tezliyi ilə onun rəqəminin hasili sabit kəmiyyətdir: $F \cdot i = \text{const}$. $15 < i < 1500$ intervalında olmasına baxmayaraq, Sipf qanunu bir növ universaldır.

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 məsələlərindən 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 müəyyənləşdirməyə yönəldilmiş qlottoxronologiya da belə statistik metodlardan biri hesab olunur.

Dilin ayrı-ayrı üslublarını – elmi, publisistik, məişət, rəsmi və s. üslubların kəmiyyətə öyrənilməsi məsələsi də böyük əhəmiyyət kəsb edir. Üslubi statistika fərdi üslubların kəmiyyət parametrlərini dəqiqləşdirməklə eyni zamanda ədəbiyyatşünaslıq elminə xeyir vermiş olur.

Müəllif üslubunun kəmiyyət qiyməti anonim və mübahisəli mətnlərin əsl müəllifinin təyin edilməsində mühüm meyar sayılır. Burada statistika ənənəvi tekstologiya və kriminalistikanın köməyinə gəlir (Ермоленко, 1988).

Sözün işlənmə dərəcəsi üzərində aparılan dəqiq müşahidələr fərdi nitqin normadan kənar çıxma hallarını qiymətləndirməyə 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 – dərslik tərtibçilərinə tezlik lüğətlərinin böyük köməyi ola bilər. Əhəmiyyətini nəzərə alaraq, tezlik lüğətlərinin tədrisdə istifadə imkanlarını nəzərdən keçirək.

Dillərin tədrisi ilə məşğul olan müəllimlər dil materiallarının seçilməsində statistik metodlardan geniş istifadə edə bilərlər. Dilin tədrisi ilə əlaqədar lüğət-minimumlar üçün sözlüyün 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).

Leksikanın tez bir zamanda və daha asan üsullarla mənimsənilməsi ilkin materialın keyfiyyətindən və lüğət-minimumun düzgün müəyyənləşdirilməsindən çox asılıdır. Lüğət-minimumun tərkibi ən işlək sözlərdən ibarət olmalı, qarşıya qoyulan bu və ya digər məqsədə müvafiq başqa mənbələrdən gətirilən materiallar hesabına tamamlanmalıdır. Bundan başqa, şifahi nitqin və yazılı dilin müxtəlif kod sistemlərinin təkmilləşdirilməsi söz birləşmələrinin işlənmə tezliyinin qeydə alınmasına əsaslanaraq, ayrı-ayrı abidələrin və müəlliflərin dilinin üslubi cəhətdən öyrənilməsində də tezlik lüğətlərindən geniş istifadə olunur.

Təbiidir ki, dil öyrənərkən, lüğət tərkibini təşkil edən bütün sözləri yadda saxlamaq mümkün deyildir. Buna görə də, ən mühüm sözləri yadda saxlamaq daha əlverişlidir. Ən mühüm sözlər işləmə, nitq prosesində və yazılı mətnlərdə daha çox işlənən sözlərdir. Beləliklə, mətnlərdə daha çox və çox az işlənən sözlərin müəyyənləşdirilməsi tədrisdə mühüm əhəmiyyət kəsb edir. Dilin tədrisi prosesində işlənmə tezliyi çox olan sözlər fəal söz ehtiyatını təşkil etməlidir. Tədris prosesi üçün zəruri komponent olan lüğətlərin sözlüyünün müəyyənləşdirilməsində və bir çox linqvo-statistik məlumatların əldə edilməsində tezlik lüğətləri əvəzsiz mənbədir.

Elmi-texniki tərəqqi dövründə tədris prosesinin uğurlu təşkili bir sıra çətinliklərlə bağlı olur. Bu çətinliklərdən biri ali məktəblərdə tədris proqramlarının həddindən artıq yüklənməsidir. İxtisası dil olmayan ali məktəblərdə xarici dili tədris edən müəllimlər xüsusilə böyük çətinlik qarşısında qalırlar, çünki xarici dillərin tədrisinə olduqca az vaxt ayrılır. Proqramların bu cür tərtibinin psixoloji nəticələri də var: tələbədə belə bir təsəvvür yaranır ki, belə az miqdarda tədris saati müqabilində xarici dili öyrənmək olmaz və ya onun öyrənilməsi bir o qədər də vacib deyil.

Tədris prosesinin səmərəliliyinin artırılması vacib məsələlərdəndir. Mütəxəssislərin xarici dili öyrənmələrinə tələbat get-

dikcə artır, lakin xarici dilin tədrisi şəraitinin yaxşılaşdırılması sahəsində isə (dilin tədrisinə ayrılmış dərs saatlarının miqdarının artırılması, ayrı-ayrı qruplarda xarici dil ilə məşğul olanların sayını azaldılması) heç bir dəyişiklik gözlənilmir. Bu məsələlərin bir çoxu tədrisin planlaşdırılması və təşkili ilə bağlıdır.

Xarici dillərin tədrisi prosesinin optimallaşdırılması rasionel 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 rasionel metodikada üç mühüm məsələ – tələbənin (şagirdin) öyrəndiyi dili tətbiq edəcəyi şərait, bunun üçün zəruri olan dil vasitələri (nitq vərdisləri) və minimum vasitə və vərdislərin 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 sözlətmələrin (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 orfoqrafiyanın təkmilləşdirilməsi ilə də sıx əlaqəlidir. 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, korrəktura oxunun öyrədilməsi və stenoqrafiya 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.

Qazaxıstan EA Dilçilik İnstitutunda «Statistik dilçilik sahəsində tədqiqatlar və avtomatlaşdırma» adlı bir qrup aşağıdakı istiqamətlərdə fəaliyyət göstərir:

- Qazax və başqa türk dillərinin bütün dil səviyyələrində (fonetika, fonologiya, leksikologiya, morfolojiya, sintaksis, üslubiyyət və s.) statistik təsviri;
- dilin informasiya (kod) səciyyəsinin müəyyən edilməsi;
- ayrı-ayrı yazıçıların dili (üslubu) üzərində tədqiqat aparılması;
- mətnin avtomatik işlənməsi;
- müasir elmi tələblər səviyyəsində dilin və nitqin öyrənilməsinin ümumi metodoloji prinsiplərinin işlənilib hazırlanması.

Qrup mətnlərin indeksləşdirilməsi işlərini yerinə yetirməklə kompüterlərdə müxtəlif növ tezlik lüğətlərinin alınması sahəsində də xeyli iş görmüşdür. Belə tezlik lüğətlərinə qrafem və onların birləşmələrini əhatə edən tezlik siyahılarını, söz-forma, söz birləşməsi, termin, heca və sairəni təsbit edən tezlik lüğətlərini göstərmək olar.

Qazax publisistikası və qazax qəzetlərinin dilini əhatə edən ilk tezlik lüğətlərinin tərtibi artıq başa çatdırılmışdır. Qrup tərəfindən anonim əsərlərin müəllifini riyazi-statistik metodla təyin etmək sahəsində də iş aparılır.

5.3.3. Azərbaycan dilinin tezlik lüğəti

Təcrübi xarakter daşıyan ilk tezlik lüğəti müasir Azərbaycan dilində olan qəzetlərin fəal leksikası əsasında azırlanmışdır.

Lüğətdən Azərbaycan dili üçün dərs vəsaitlərinin, lüğət-minimum, müxtəlif danışq kitabçalarının və s. tərtib edilməsi üçün istifadə oluna bilər. Bu lüğət bədii dilin normallaşdırılması, eləcə də nitq mədəniyyəti məsələləri ilə əlaqədar tədqiqatlarda da böyük əhəmiyyət kəsb edir.

Lüğət əlifba və tezliklərin azalma sırası siyahılarından ibarətdir. Əlifba siyahısında da hər bir sözün qarşısında onun tezliyi qeyd olunur.

Mətnlərin seçilməsi, seçmənin həcmi tezlik lüğətinin mühüm səciyyəsi hesab olunur. Çünki lüğətin statistik etibarlılığı və buna müvafiq olaraq onun müxtəlif tətbiqləri üçün yararlılığını təhlil edilmiş materialın həcmi müəyyən edir. Seçmənin həcmi qeydə alınmış vahidlərin bütün işlənmə hallarının məcmuyundan ibarətdir ki, bu da tezlik lüğətində bütün vahidlərin tezliyinin cəminə bərabərdir. Mövcud tezlik lüğətləri seçmənin həcminə görə bir-birindən çox fərqlənirlər: indiyə kimi 20 mindən 1 milyona qədər və daha çox söz-formanı əhatə edən tezlik lüğətləri çap olunmuşdur. Məlumdur ki, seçmənin həcmi nə qədər böyük olarsa, tezlik lüğətindən alınan məlumat da bir o qədər etibarlı olar. Lüğətin Azərbaycan dilçiliyində ilk təcrübə olması, texniki imkanların məhdudluğu və qısa müddətdə bu sahədə müəyyən elmi nəticə olmaq zərurəti ilə əlaqədar olaraq seçmənin həcmi 100 000 söz-forma qəbul edilmişdir.

«Azərbaycan qəzet dilinin tezlik lüğəti»nin tərtib olunması üçün materialı, tezlik lüğətlərinin bir çoxunda olduğu kimi, qəzet mətnlərindən götürülmüş seçmələr təşkil edir.

Qəzet kütləvi məlumat vasitəsi kimi nitq fəaliyyətində mühüm yerlərdən birini tutur və digər mətn növlərindən bəzi xüsusiyyətlərinə görə fərqlənir. Ümumi həcmdə çap məhsullarının böyük bir hissəsi qəzetlərin payına düşür. Yazılı nitqin mühüm bir hissəsi də qəzet mətnlərində əks olunur. Qəzet təkcə kütləvi oxucuya deyil, həm də kütləvi müəllifə malikdir.

Bunun nəticəsində qəzet dil daşıyıcısı olan kollektivin dilinə güclü normalaşdırıcı təsir göstərir, dildə müəyyən normanın sabitləşməsinə şərait yaradır. Bundan başqa, müəlliflər öz dillərini müəyyən dərəcədə oxucuların gündəlik danışmaq dilinə yaxınlaşdırmağa çalışaraq, tez-tez normadan uzaqlaşır, dil vahidlərinin nitqdə qeyri-normativ hesab olunan yeni formalarından istifadə edirlər. Qəzet bədii mətnlərdə və normativ sorğu kitablarında qeyd olunmağa imkan tapmayan dil faktlarını və dildə baş verən bütün dəyişiklikləri, demək olar ki, ani qeydə almağa imkan verir. Beləliklə, qəzet mətnləri ən yeni dil məlumatlarının mənbəyi rolunu oynayır.

Son zamanlar bir sıra dillər üçün, o cümlədən türk dillərindən biri olan qazax dili üçün qəzet materialları əsasında tezlik lüğəti tərtib olunmuşdur.

Azərbaycan dilinin tezlik lüğəti üçün seçmələrin əsasını «Kommunist», «Bakı», «Azərbaycan gəncləri» kimi respublika qəzetlərinin 1976-cı il nömrələrindən götürülmüş mətnlər təşkil edir. Seçmə vahidi olaraq qəzetin hər səhifəsindən götürülmüş 100 söz-forma qəbul edilmişdir. Qəzet mətnlərini «janrlara» görə bölmək xüsusi məqsəd kimi qarşıya qoyulmamışdır. Lakin qəzetin müxtəlif səhifələrində mətnlərin məzmununa görə bölgüsü onları aşağıdakı qruplara ayırmağa imkan verir: baş məqalə, daxili həyat (əsas etibarilə sənaye, kənd təsərrüfatı, elm və texnika yenilikləri), xarici xəbərlər, incəsənət, idman və s. məlumatlar. Göstərilən bölgü ayrı-ayrı mətnlərin seçilməsi üçün müəyyən istiqamət rolunu oynayır.

Əgər adi lüğətlərdə lüğət vahidi kimi söz götürülürsə, tezlik lüğətlərində başqa dil vahidləri (söz-forma, morfem və s.) də qeydə alın bilər. Yəni dil tədqiqatının məqsədindən asılı olaraq göstərilən vahidlərdən biri seçilib götürülür. Əgər tezlik lüğəti dilin sözdəyişdirmə və qrafik sisteminin öyrənilməsi məqsədi ilə tərtib olunmuşsa, lüğət vahidi kimi söz-formalar götürülə bilər. Lakin tədqiqatçı bir lüğət vahidi kimi söz (məsələn, dili öyrənməklə əlaqədar) və onun mənası maraqlandıırırsa, onda hesablaşma vahidi söz-leksəm ön plana çəkilir. Leksik vahidlərin geniş siyahısı tədqiq olunan dilin bütün zənginliyini açmağa və statistik hesablamaların etibarlılığına təminat verə bilər.

Bəzi tədqiqatçıların haqlı olaraq qeyd etdiyi kimi, indiyə qədər dilçilikdə «lüğət vahidi» anlayışı dəqiq müəyyənləşdirilməmişdir. Bu da öz növbəsində tezlik lüğətləri üçün vahidlərin seçilməsini çətinləşdirir. Mətnlərin obyektiv təhlili üçün mətndəki material olduğu kimi əks olunmalıdır. Söz-formanın leksemlərlə hər cür əvəzlənməsi isə obyektiv metoddan yayınmaq deməkdir, çünki bu zaman müəyyən subyektiv amil tədqiqata nüfuz etmiş olur. Şübhəsiz, bu müddəa türk dilləri, eləcə də Azərbaycan dili üçün tam əsaslandırıla bilər. Çünki bu dillərdə

sözlərin nitq hissələrinə görə təsnifi və söz-formaların hər hansı bir leksemə aid edilməsi bir çox hallarda tamamilə mübahisəli hesab olunur (nitq hissələrinin təsnifi məsələsinə sonrakı fəsillərdə toxunulmuşdur).

Məhz buna görə, mətn işlənmərkən 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 dilçiliyində müvafiq məsələlərin kifayət qədər öyrənilməməsi, mətnlərin maşına verilməzdən öncə işlənməsi zamanı bir sıra sadələşdirmələr aparmaq zərurətini doğurur. Məsələn, sözlərin nitq hissələrinə görə indeksləşdirilməsindən imtina etmək lazım gəlirdi. Mürəkkəb felləri quruluşlarına görə onlara oxşar söz birləşmələrindən ayırmaq üçün hamı tərəfindən qəbul olunmuş ümumi, obyektiv bir qaydanın olmaması ilə əlaqədar göstərilən konstruksiyalar komponentlər üzrə müstəqil yazılmış oldu. Etibarlı standart lüğət olmadan təhlili mümkün olmayan frazeoloji vahidlər üçün də analogi qayda tətbiq olundu.

5.3.4. Tezlik lüğətlərinin statistikasi

Lüğətdə vahidlərin mütləq tezliyi hesablanır. Bu göstərici tədqiq olunan vahidin mətndə işlənməsinin əsas, vacib və minimal zəruri səciyyəsi hesab olunur. Mütləq tezlik əsasında əlavə hesabatlar vasitəsilə başqa törəmə səciyyələr də əldə etmək olar.

Lüğətdə həmçinin sözlərin yayılma göstəricisi də qeyd alınır. Bu göstərici tədqiq olunan mətnə sözlərin bərabər ölçüdə müşahidə olunması dərəcəsi haqqında məlumat verir. Sözlün yayılma dərəcəsi onun heç olmasa bir dəfə qeydə alındığı mətnlərin miqdarı ilə müəyyənləşdirilir. Ən çox işlənən sözlər, bir qayda olaraq, bütün tədqiq olunan mətnlərdə, az işlənən sözlər isə mətnlərin bəzilərində müşahidə olunur.

Tezlik siyahısında sözlün sıra nömrəsini göstərən və tezlikdən doğan səciyyələrdən biri də rənqdır. Eyni tezliyə malik sözlərə müəyyən interval verilir.

Kompüterlərin lüğətlərin tərtibi ilə əlaqədar olaraq tətbiqi nəticəsində bir çox kəmiyyət səciyyələrinin alınması mümkün olmuşdur. Aşağıda «Nitq statistikas» qrupunun tədqiqatlarında qəbul edilmiş statistik göstəriciləri nəzərdən keçirək:

i - ranq, F - tezlik, m - eyni tezliyə malik sözlərin miqdarı, F* - toplanmış mütləq tezlik, f - nisbi tezlik, f* - toplanmış nisbi tezlik.

Sözün yığım tezliyi həmin sözün tezliyinin, tezlik siyahısında ondan əvvəl gələn sözlərin tezlikləri ilə cəmlənməsindən alınır. Məsələn, söz sırasına görə onuncu olan sözün yığım tezliyi siyahının əvvəlindən başlayaraq bütün on sözün tezlikləri cəminə bərabərdir. I cədvəldə Nəsimi adına Dilçilik İnstitutunun Riyazi dilçilik qrupunun əməkdaşları tərəfindən tərtib olunmuş "Azərbaycan qəzet dilinin tezlik lüğəti"ndə tezliklərin paylanması qanunauyğunluqları göstərilmişdir.

Nisbi tezlik alınmış tezliyin mətnin ümumi uzunluğuna olan nisbəti ilə ölçülür. Başqa sözlə desək, bu sözün mətndə paylanma miqdarını göstərir. Bu kəmiyyət göstəricisi onluq kəsrə və ya faizlə ifadə oluna bilər.

Nisbi yığım tezliyi də bu qayda ilə – yığım tezliyinin mətnin ümumi uzunluğuna nisbəti ilə hesablanır. Bu halda bir sözün deyil, bu sözün və tezliklərin azalma sırası ilə düzülüş siyahıda ondan əvvəl gələn bütün sözlərin mətnə ümumi paylanma miqdarı müəyyənləşdirilmiş olur.

5.3.5. Tezlik lüğətlərinin tərtibi qaydaları

Yuxarıda göstərdik ki, türkologiyada nitq hissələri, söz birləşmələrinin təsnifində bəzi mübahisə doğuran səbəblər əsas götürülərək tezlik lüğətinin tərtibində müəyyən sadələşdirmələr aparılmış, nitq hissələrinə görə indeksləşdirilmə aparılmamış, mürəkkəb fel konstruksiyaları komponentlər üzrə müstəqil götürülmüşdür. Azərbaycan qəzet dilinin tezlik sözlüyünün tərtibində aşağıda sadalanan prinsiplər əsas götürülmüşdür:

- isimlər adlıq halında, mənsubiyyət və cəm şəkilçisi nəzərə alınmadan yazılmışdır: əldən - *əl*, atım - *at*, kitablar - *kitab*,
- iki və daha artıq sözün birləşməsindən əmələ gələn mürəkkəb isimlərdə komponentlərin hər biri ayrılıqda götürülmüşdür: Sovet kəndi - *sovet, kənd*, Azərbaycan Dövlət Filarmoniyası - *Azərbaycan, Dövlət, Filarmoniya*,
- qısaldılmış mürəkkəb adlar bir söz kimi yazılmışdır: *AHIŞ, BDU*,
- antroponimlər (ad, familiya və s.), toponimlər (yer adları) nəzərə alınmamışdır;
- defislə yazılan qoşa isimlər hər iki komponentin və ya komponentlərdən yalnız ikincisinin hal şəkilçisi qəbul edib-etməmələrindən asılı olmayaraq bir söz kimi qəbul edilmişdir: *ata-baba, mal-qara, uşaq-muşaq*,
- isimdə və müxtəlif nitq issələrinə müşahidə olunan omoformlar fərqləndirilməmişdir;
- sifətin azaltma dərəcəsində -*ımtıl*, -*raq* və s. şəkilçilərlə düzələn sifətlər müstəqil sözlər kimi götürülmüşdür: *bozumtul, yaxşıraq, qırmızımtraq*,
- analitik yolla düzələn sifətin üstünlük dərəcəsində hər iki söz ayrı-ayrılıqda yazılmışdır: ən yaxşı - *ən, yaxşı*, lap gözəl - *lap, gözəl*, daha maraqlı - *daha, maraqlı*,
- qoşa sifətlər bir söz kimi qəbul edilmişdir;
- miqdar sayları adlıq halda, tək halda götürülmüşdür: beşdən - *beş*, altısında - *altı*, ikiləri - *iki*,
- sıra sayları müstəqil söz kimi yazılmışdır: birincisi - *birinci*, doqquzuncunu - *doqquzuncu*,
- mürəkkəb saylar komponentlər üzrə verilmişdir: altmış iki - *altmış, iki*, on beşinci - *on, beşinci*,
- rəqəmlə yazılış hərfi yazılışla əvəz olunmuşdur;

Cədvəl 1. Azərbaycan dilinin tezlik lüğətinin statistik quruluşu

RANQ N-si I	Mütləq tezlik, F	Eyni tezliyə malik sözlərin miqdarı, m	Toplanmış mütləq tezlik, F*	Nisbi tezlik, f	Toplanmış nisbi tezlik, f*
1	2	3	4	5	6
1	2535	1	2535	0,02535	0,02535
2	1256	1	3791	0,01256	0,03791
3	1040	1	4831	0,01040	0,04831
4	770	1	5601	0,00770	0,05601
5	695	1	6296	0,00695	0,06296
6	719	1	7015	0,00719	0,07015
7	668	1	7683	0,00668	0,07683
8	611	1	8294	0,00611	0,08294
9	602	1	8896	0,00602	0,08896
10	589	1	9485	0,00589	0,09485
11	543	1	10028	0,00543	0,10028
12	542	1	10570	0,00542	0,10570
13	513	1	11083	0,00513	0,11083
14	511	1	11594	0,00511	0,11594
15	498	1	12092	0,00498	0,12092
16	487	1	12579	0,00487	0,12579
17	386	2	73351	0,00386	0,13351
18	385	1	13736	0,00385	0,13736
19	383	1	14119	0,00383	0,14119
20	378	1	14497	0,00378	0,14497
21	376	2	15249	0,00376	0,15249
22	365	1	15614	0,00365	0,15614
23	354	1	15968	0,00354	0,15968
24	342	1	16310	0,00342	0,16310
25	339	1	16649	0,00339	0,16649
26	328	1	16977	0,00328	0,16977
27	322	1	17299	0,00322	0,17299
28	321	1	17620	0,00321	0,17620

29	320	1	17940	0,00320	0,17940
30	312	2	18564	0,00312	0,18564
31	311	2	19186	0,00311	0,19186
32	305	1	19491	0,00305	0,19491
33	303	1	19794	0,00303	0,19794
34	302	1	20096	0,00302	0,20096
35	292	1	20393	0,00297	0,10393
36	293	1	20686	0,00293	0,20686
37	292	1	20978	0,00292	0,20978
38	281	1	21259	0,00281	0,21259
39	280	1	21539	0,00280	0,21539
40	276	1	21815	0,00276	0,21815
41	275	1	22090	0,00275	0,22090
42	272	1	22362	0,00272	0,22362
43	267	1	22629	0,00267	0,22629
44	264	1	22893	0,00264	0,22893
45	256	1	23149	0,00256	0,23149
46	251	1	23400	0,00251	0,23400
47	243	1	23643	0,00243	0,23643
48	239	1	23882	0,00239	0,23882
49	235	1	24117	0,00235	0,24117
50	231	1	24348	0,00231	0,24348
51	229	2	24866	0,00229	0,24866
52	227	1	25032	0,00227	0,25032
53	226	1	25250	0,00226	0,25250
54	218	1	25477	0,00218	0,25477
55	201	1	25678	0,00201	0,25678
56	200	1	25878	0,00200	0,25878
57	197	1	26075	0,00197	0,26075
58	193	2	26461	0,00193	0,26461
59	190	1	26651	0,00190	0,26651
60	188	2	27027	0,00188	0,27027
61	185	2	27212	0,00185	0,27212
62	183	1	27395	0,00183	0,27395
63	180	2	27755	0,00180	0,27755

64	179	1	27934	0,00179	0,27934
65	178	1	28112	0,00178	0,28112
66	177	1	28289	0,00177	0,28289
67	175	1	28464	0,00175	0,28464
68	170	1	28634	0,00170	0,28634
69	169	1	28803	0,00169	0,28803
70	168	1	28972	0,00168	0,28972
71	163	2	29297	0,00163	0,29297
72	162	1	29459	0,00162	0,29459
73	159	1	29618	0,00159	0,29618
74	158	1	29776	0,00158	0,29776
75	157	3	30247	0,00157	0,30247
76	153	2	30553	0,00153	0,30553
77	152	2	30857	0,00152	0,30857
78	151	1	31008	0,00151	0,31008
79	149	5	31753	0,00149	0,31753
80	148	1	31901	0,00148	0,31901
81	146	1	32047	0,00146	0,32047
82	144	2	32335	0,00144	0,32335
83	142	1	32477	0,00142	0,32477
84	140	1	32617	0,00140	0,32617
85	139	1	32756	0,00139	0,32756
86	137	1	32893	0,00137	0,32893
87	135	1	33028	0,00135	0,33028
88	133	1	33161	0,00133	0,33161
89	127	1	33288	0,00127	0,33288
90	126	1	33414	0,00126	0,33414
91	125	2	33664	0,00125	0,33664
92	124	1	33788	0,00124	0,33788
93	123	4	34280	0,00123	0,34280
94	122	1	34402	0,00122	0,34402
95	121	1	34523	0,00121	0,34523
96	120	1	34643	0,00120	0,34643
97	119	2	34881	0,00119	0,34881
98	118	3	35235	0,00118	0,35235

99	117	1	35352	0,00117	0,35352
100	116	1	35468	0,00116	0,35468
101	115	3	36813	0,00115	0,36813
102	114	2	36041	0,00114	0,36041
103	112	1	36153	0,00112	0,36153
104	111	2	36375	0,00111	0,36375
105	110	1	36595	0,00110	0,36595
106	109	1	36704	0,00109	0,36704
107	107	3	37025	0,00107	0,37025
108	106	2	37237	0,00106	0,37237
109	105	1	37342	0,00105	0,37342
110	104	4	37758	0,00104	0,37758
111	103	3	38067	0,00103	0,38067
112	100	3	38367	0,00100	0,38367
113	98	3	39151	0,00098	0,39151
114	96	2	39343	0,00096	0,39343
115	95	2	39533	0,00095	0,39533
116	94	5	40003	0,00094	0,40003
117	93	2	40189	0,00093	0,40189
118	91	2	40371	0,00091	0,40371
119	90	2	40651	0,00090	0,40551
120	89	2	40729	0,00089	0,40729
121	88	2	40905	0,00088	0,40905
122	87	2	41075	0,00087	0,41075
123	84	4	41415	0,00084	0,41415
124	83	4	41747	0,00083	0,41747
125	82	3	41993	0,00082	0,41993
126	81	2	42155	0,00081	0,42155
127	80	4	42475	0,00080	0,42475
128	79	2	42633	0,00079	0,42633
129	78	5	43023	0,00078	0,43023
130	77	2	43177	0,00078	0,43177
131	76	4	43481	0,00076	0,43481
132	74	4	43777	0,00074	0,43777
133	73	4	44069	0,00073	0,44069

134	72	7	44573	0,00072	0,44573
135	71	7	45070	0,00071	0,45070
136	70	2	45210	0,00070	0,45210
137	69	5	45555	0,00069	0,45555
138	68	5	45895	0,00068	0,45895
139	67	5	46230	0,00067	0,46230
140	66	2	46362	0,00066	0,46362
141	65	3	46557	0,00065	0,46557
142	63	5	46872	0,00063	0,46872
143	62	2	46996	0,00062	0,46996
144	61	2	47118	0,00061	0,47118
145	60	6	47478	0,00060	0,47478
146	59	6	47851	0,00059	0,47851
147	58	2	48007	0,00058	0,48007
148	57	7	48406	0,00057	0,48406
149	56	7	48798	0,00056	0,48798
150	55	6	49128	0,00055	0,49128
151	54	5	49398	0,00054	0,49398
152	53	7	49769	0,00053	0,49769
153	52	2	49873	0,00052	0,49873
154	51	4	50077	0,00051	0,50057
155	50	6	50377	0,00050	0,50377
156	49	2	50475	0,00049	0,50475
157	48	5	50715	0,00048	0,50715
158	47	10	51185	0,00047	0,51185
159	46	9	51599	0,00046	0,51599
160	45	12	52139	0,00045	0,52139
161	44	13	52711	0,00044	0,52711
162	43	5	52926	0,00043	0,52926
163	42	10	53346	0,00042	0,53346
164	41	12	53838	0,00041	0,53838
165	40	13	54358	0,00040	0,54358
166	39	7	54631	0,00039	0,54631
167	38	4	54783	0,00038	0,54783
168	37	15	55338	0,00037	0,55338

169	36	16	55914	0,00036	0,55914
170	35	13	56369	0,00035	0,56369
171	34	13	56811	0,00034	0,56811
172	33	15	57306	0,00033	0,57306
173	32	12	57690	0,00032	0,57690
174	31	17	58217	0,00031	0,58217
175	30	14	58637	0,00030	0,58637
176	29	19	59188	0,00029	0,59188
177	28	17	59664	0,00028	0,59664
178	27	21	60231	0,00027	0,60231
179	26	18	60699	0,00026	0,60699
180	25	18	61149	0,00025	0,61149
181	24	19	61605	0,00024	0,61605
182	23	17	61996	0,00023	0,61996
183	22	21	62458	0,00022	0,62458
184	21	19	62857	0,00021	0,62857
185	20	23	63317	0,00020	0,63317
186	19	27	63830	0,00019	0,63830
187	18	38	64514	0,00018	0,64514
188	17	34	65092	0,00017	0,65092
189	16	43	65780	0,00016	0,65780
190	15	44	66440	0,00015	0,66440
191	14	58	62252	0,00014	0,62252
192	13	67	68123	0,00013	0,68123
193	12	57	68807	0,00012	0,68807
194	11	51	69368	0,00012	0,69368
195	10	73	70098	0,00010	0,70098
196	9	93	70935	0,00009	0,70935
197	8	97	71711	0,00008	0,71711
198	7	135	72656	0,00007	0,72656
199	6	131	73442	0,00006	0,74442
200	5	203	74457	0,00005	0,74452
201	4	299	75653	0,00004	0,75653
202	3	466	77051	0,00003	0,77051
203	2	852	78755	0,00002	0,78755
204	1	3023	81778	0,00001	0,81778

- larca, -larla şəkilçiləri ilə əmələ gəlmiş qeyri-müəyyən miqdar sayları ayrıca söz kimi verilmişdir: *minlərcə, on-larca, yüzlərlə*;
- defislə yazılan miqdar sayları bir söz kimi yazılmışdır: *bir-iki, beş-altı, beş-beş, on-on*;
- kəsr sayları iki söz kimi qəbul edilmişdir: *beşdə iki - beş, iki, ikidə bir - iki, bir*;
- şəxs əvəzliləri, “onlar” istisna olmaqla, adlıq halda verilmişdir: *mənim - mən, bizdə - biz, sizlərdən - siz*;
- “onlar” şəxs əvəzliyi və başqa növlərdən olan əvəzlilər əsas halda, təkdə yazılmışdır: *onlar - o, bunlarda - bu, özlərinə - öz*;
- əvəzlilə başqa sözlərin birləşməsindən əmələ gəlmiş mürəkkəb əvəzlilər komponentlər üzrə, ayrıca sözlər kimi verilmişdir: *heç nə - heç, nə, kim isə - kim, isə, nə üçün - nə, üçün, hər bir kəs - hər, bir, kəs*;
- defislə yazılan əvəzlilər bir söz kimi verilmişdir: *mənli-mənsiz, sizli-bizli, elə-belə, belə-belə*;
- fel məsdər formasında, təsdiqdə, feli sifət və feli bağlama ayrıca söz kimi yazılmışdır;
- mürəkkəb fellərin komponentləri və növ xüsusiyyətlərini özündə əks etdirən analitik fellər ayrıca sözlər kimi verilmişdir: *yadına salır - yad, salmaq, əl tutmamışam - əl, tutmaq, ad qoydular - ad, qoymaq, aşıb-daşdı - aşmaq, daşmaq, saralıb-solmaq - saralmaq, solmaq, yazıb qurtarmışdı - yazmaq, qurtarmaq*;
- felin bacarıq forması əsas felin məsdərinə verilmişdir: *ala bilmədim - almaq, oxuya bilərəm - oxumaq, otura bilməmişdir - oturmaq*. Belə yanaşma çərçivəsində *-a, -ə bilmək* birləşməsi bacarıq formasının əsas göstəricisi, sözdəyişmə kateqoriyalarından biri kimi qəbul edilir;
- felin məchul, qayıdış, qarşılıq, icbar kimi sözdüzəltməyə meyl edən növləri məsdər şəklində yazılmışdır: *açılır - açılmaq, başlamamışdan - başlamaq, tutuşdurdular - tutuşdurulmaq*;
- keçmiş zamanın mürəkkəb formalarının əmələ gəlməsində iştirak edən *idi, imiş* köməkçi felləri və ekvivalent variant hesab olunan *-di, -miş* ayrılıqda qeyd olunmamış-

- dır: *almış idi* - almaq; *durur imiş* - durmaq; *oturlmalı idim* - oturmaq; *itirsə idi* - itirmək;
- *olmaq* feli ilə işlənən perifrastik formalar məsdər formasına gətirilmişdir: *oxumalı oldum* - oxumaq, *yazası oldu* - yazmaq, *almış olar* - almaq;
 - defislə yazılan qoşa zərflər bir söz kimi yazılmışdır: *astasta*, *uzun-uzadı*, *az-çox*;
 - mürəkkəb zərflər komponentlərinə ayrılmışdır: hər yerdə *hər, yer*, bir az - *bir, az*, birə-iki - *bir, iki*;
 - mürəkkəb bağlayıcılar komponentlər üzrə verilmişdir: ona görə - *o, görmək*; bunun üçündür ki - *bu, üçün, ki*, amma buna baxmayaraq - *amma, bu, baxmaq*;
 - *görəsən, olmaya, qoy, qoysana, gəlsənə* ədatları ilə müvafiq fel formaları fərqləndirilmişdir;
 - defislə yazılan nidalar bir söz kimi qeyd olunmuşdur: *a-a, bə-bə-bə*;
 - mürəkkəb modal sözlər (birləşmələr) komponentlərinə ayrılmışdır: ola bilsin ki - *olmaq, ki*, bir söz - *bir, söz*; nə təhər deyərlər - *nə, təhər, demək*;
 - *demək, deməli* modal sözləri müvafiq fel formalarından fərqləndirilmişdir.

Burada göstərilən prinsiplərin bir çoxu şərti olaraq qəbul edilmişdir və tezlik lüğətlərinin tərtibində tədqiqatçıların qarşılıqlarına qoyduqları məqsədə müvafiq olaraq qaydalar müxtəlif ola bilər.

Azərbaycan qəzet dilinin tezlik lüğəti iki hissədən - əlifba-tezlik sözlüyündən və tezlik sözlüyündən ibarətdir.

Əlifba-tezlik sözlüyü mətndə müşahidə olunan bütün leksəmləri əhatə edir. Ümumiyyətlə, sözlükdə 10 000 söz-formadan ibarət seçmələrdən alınmış 6 532 müxtəlif söz qeydə alınmışdır.

Lüğətdəki hər bir leksik vahidin iki kəmiyyət səciyyəsi-tezliyi və həmin sözün qeydə alındığı mətnlərin sayı göstərilir.

Sözün tezliyinə görə onun tezlik lüğətindəki yerini və lüğətin statistik quruluşundakı sıra nömrəsini müəyyənləşdirmək olar. Sözün işləkliyinin müəyyənləşdirilməsi üçün təkcə onun tezliyini deyil, həm də onun işləndiyi mətnlərin sayını nəzərə almaq lazımdır.

Cədvəl 2. Tezlik sözlüyündən bir fraqment

Sıra- sı	Söz	Tezlik	Yayıma
1	və	2535	100
2	etmək	1256	100
3	Bu	1040	100
4	Bir	770	100
5	Yüz	695	90
6	O	719	100
7	Olmaq	668	91
8	İl	611	70
9	Partiya	602	65
10	ilə	589	68
11	Sovet	543	65
12	Min	542	91
13	Ki	513	67
14	Azərbaycan	511	63
15	İş	498	68
16	üçün	487	66
17	Doqquz	386	95
18	Respublika	386	66
19	İyirmi	385	95
20	Xalq	383	84
21	təsərrüfat	378	80
22	Yer	376	93
23	çox	376	95
24	On	365	81
25	də	354	64
26	Böyük	351	78

Cədvəl 3. Əlifba-tezlik sözlüyündən bir fragment

Sırası	Söz	Tezlik	Yayıllma
1	Abad	3	3
2	Abadlaşdırılma	2	2
3	Abadlaşdırılmış	1	1
4	Abadlaşdırma	1	1
5	Abadlıq	2	1
6	Abidə	13	9
7	Ab-hava	1	1
8	ABŞ	26	13
9	Abşeron	9	3
10	Avadanlıq	32	20
11	Avanqard	5	5
12	Avantüra	1	1
13	Avqust	9	7
14	Aviasiya	12	5
15	Avropa	28	10
16	Avtobaza	1	1
17	Avtobus	10	7
18	Avtomat	4	3
19	Avtomatik	1	1
20	Avtomatika	1	1
21	Avtomatlaşdırılmaq	1	1
22	Avtomatlaşdırma	1	1

Əgər kifayət qədər böyük tezliyə malik sözün bir neçə mətndə rast gəlinməsi halı müşahidə olunarsa, belə sözləri çox işlək söz hesab etmək olmaz. Məsələn, *birgə* və *maddə* - sözləri eyni tezliyə malikdirlər. Hər ikisinin tezliyi 13-ə bərabərdir. Lakin birinci söz 13 mətndə, ikinci söz isə cəmi beş mətndə müşahidə olunmuşdur.

Tezlik sözlüyü tezliyi 5 və ondan çox olan 1892 vahidi əhatə edir. Onlar tezliklərin azalması sırasına görə düzülüşlər. Tezlik sözlüyündə tezliyi ən böyük olan sözlər bütün sözlüyün 28,9 faizini təşkil edir. Buna baxmayaraq həmin sözlər bütün

mətnlərin 91 faizini, qalan 4640 söz isə mətnin 9 faizini əhatə edir.

Azərbaycan dilini tam əhatə edən fundamental tezlik lüğətinin hazırlanması olduqca vacib və aktual problemlərin həllinə böyük kömək olardı. İstər formal təhlil sistemləri üçün, istər ənənəvi lüğətçilik, istərsə də bütövlükdə Azərbaycan dilçiliyi üçün belə lüğətin olduqca böyük nəzəri və praktik əhəmiyyəti var.

Haqqında söhbət gedən “Azərbaycan qəzet dilinin tezlik lüğəti” AMEA Nəsimi adına Dilçilik İnstitutunun tətbiqi dilçilik qrupu əməkdaşları tərəfindən ötən əsrin 70-ci illərində hazırlanmışdı. Həmin illərdə belə bir lüğətin tərtibi olduqca çətin və mürəkkəb bir iş idi. Müasir kompüterlərin mövcud olmadığı bir zamanda, sürəti az və yaddaşı məhdud olan elektron hesablama maşınlarında linqvisistik tədqiqatlar aparılmırdı. Kibernetika və dilçilik psixoloji cəhətdən belə araşdırmalara hazır deyildi. Hesab olunurdu ki, elektron hesablama maşınları yalnız dəqiq elmlər sahəsində “ciddi” məqsədlərlə istifadə oluna bilər. Dilçilərin kompüterlərdən istifadə etmələri təəccüblə (bəlkə də, istehza ilə) qarşılanırdı. Kibernetika İnstitutunda, yaxud Bakı Dövlət Universitetinin Hesablama Mərkəzində o vaxtkı elektron hesablama maşınlarının xidmətindən dilçilərə yalnız işdən kənar vaxtlarda istifadə etməyə icazə verilirdi.

70-ci illərdə informasiyanın (dil materialının) maşına daxil edilməsi də çətin bir proses idi. Köhnə elektron maşınlarda informasiya perforasiya maşınlarında qabaqcadan perfolent və perforatlara yazılardan sonra yaddaşa daxil edilə bilərdi. Həmin prosesin necə ağır və çətin olduğunu o nəslin nümayəndələri yaxşı bilirlər.

Dilin riyazi-statistik metodların köməyi ilə tədqiqinə dilçilərin də münasibəti birmənalı deyildi. Hətta o dövrün görkəmli dilçiləri də belə tədqiqatların perspektivinə inanmır, bu sahədə aparılacaq işlərə dəstək vermir, süni maneələr yaradırdılar. “Azərbaycan qəzet dilinin tezlik lüğəti” və “Azərbaycan dilinin

əks əlifba lüğəti” kimi tədqiqatların həmin vaxtlarda çap olunmasının bir səbəbi, bəlkə də, həmin münasibətin nəticəsi idi.

1979-cu ildə A.Axundovun “Riyazi dilçilik” kitabının çapından sonra dilçiliyin bu yeni sahəsinə olan münasibət tədricən dəyişməyə başladı. Sonralar “Kitabi-Dədə Qorqud” dastanlarının və Məhəmməd Füzulinin nəzm əsərlərinin statistikasını sahəsində araşdırmalar aparılmağa başlandı. Həmin vaxtlarda məşin tərcüməsi, dilin mühəndis dilçiliyi metodları ilə tədqiqi istiqamətində bir çox uğurlar qazanıldı. Həmin işlərlə bağlı müxtəlif nüfuzlu məcmuələrdə, xarici ölkələrdə, Moskva və Leninqradda (indiki Sankt-Peterburqda) məqalələr çap olundu.

Azərbaycan qəzet dilinin tezlik lüğəti 70-ci illərdə çap olunsaydı, həmin dövr üçün sensasiya olardı. Təəssüf ki, bu baş vermədi. Kompüter texnologiyalarının inkişafı səviyyəsində hazırda daha mükəmməl, iri həcmli, dili tam əhatə edə biləcək tezlik lüğətləri hazırlamaq imkanı yaranmışdır. Buna baxmayaraq ilk təcrübə kimi göstərilən lüğətin əhəmiyyəti çox böyükdür.

Mətnlərin toplanması və məşindən qabaq işlənməsi prosesində institutun əməkdaşları Gülşən Axundova və Mübariz Yusifov da iştirak etmişlər (Azərbaycan qəzet dilinin tezlik lüğəti, 2004).

2010-cu ildə üçcildlik “Azərbaycan dilinin tezlik lüğəti”nin I cildi çapdan çıxmışdır. I cild söz köklərini əhatə edir. Tərtibçilər M.Mahmudov, Ə.Fətullayev, S.Məmmədova, R.Fətullayev, R.Hüseynli, B.Talıbov, G.Əzimova, N.Abdullayevdir. Kitab Azərbaycan Respublikası Rabitə və İnformasiya Texnologiyaları Nazirliyinin dəstəyi ilə çap olunmuşdur.

Azərbaycan dilinin tezlik lüğətinin tərtibi üçün bütün üsullar (elmi, bədii, publisist, rəsmi, məişət və s.) təmsil olunmaqla təxminən 50 milyona yaxın söz-forma linqvostatistik baxımdan kompüterlərdə işlənməmiş, rəqəmlər, müxtəlif simvollar, təsadüfi sözlər, xüsusi adlar və s. “saf-çürük” edildikdən sonra təxminən 300 minə yaxın söz-forma qalmışdır. Həmin baza əsasında müxtəlif tipli tezlik lüğətlərinin (əsasların, söz-formaların, söz köklərinin tezlik, əlifba-tezlik və əks-əlifba tezlik siyahıları,

qrammatik və leksik şəkilçilərin, qrafem və qrafem birləşmələrinin, durğu işarələrinin tezlik səciyyələrini əks etdirən siyahılar və s.) hazırlanması nəzərdə tutulur.

Azərbaycan dilinin tezlik lüğəti gələcəkdə çapı nəzərdə tutulan bu tipli lüğətlərin birincisi hesab oluna bilər. Birinci mərhələdə lüğətin üç cildən ibarət olması nəzərdə tutulur. I cild söz köklərinin, tezlik səciyyələrini əks etdirən siyahılardan ibarətdir.

Sözün kökü dedikdə, sözün şəkilçiləri atıldıqdan sonra qalan və müəyyən əşyavi məna bildirən hissə başa düşülür. Söz kökləri tezlik lüğətində müasir inkişaf vəziyyəti baxımından araşdırılır, tarixi, etimoloji aspekt nəzərə alınır. Ünsiyyət prosesində müəyyən morfoloji dəyişmələrə məruz qalmış söz kökləri ilkin qrammatik formaya gətirilmir. Bu, mətnlərin formal işlənməsi baxımından zəruridir. Məsələn, “gedir” söz-formasında şəkilçi atıldıqdan sonra qalan hissə “ged” kökü hesab olunur. Əslində, sözün kökü ilkin formada “get/mək” hesab olunur (Azərbaycan dilinin tezlik lüğəti, I cild, 2010).

II cild əsasların tezliyini əks etdirən statistik göstəriciləri əhatə edir.

Məlum olduğu kimi, əsas dedikdə, söz-formadan (ünsiyyət prosesində müəyyən qrammatik formada çıxış edən söz) sonluqları və forma düzəldici şəkilçilər atıldıqdan sonra yerdə qalan və sözün leksik mənasını bildirən hissə başa düşülür. Tərkibindəki sözdüzəldici şəkilçilərin sayından asılı olaraq əsaslar aşağıdakı kimi qruplaşdırılır: birinci əsas (kökə yalnız bir sözdüzəldici şəkilçi artırılmaqla düzələn əsas), ikinci əsas (kökə iki sözdüzəldici şəkilçi artırılmaqla düzələn əsas), üçüncü əsas (kökə üç sözdüzəldici şəkilçi artırılmaqla düzələn əsas) və s. Məsələn, “yazıçılıq” leksik vahidində “yaz” – kök, “yazı” – birinci əsas, “yazıçı” – ikinci əsas, “yazıçılıq” – üçüncü əsasdır (Adilov, Verdiyeva, Ağayeva, 1989, s. 107).

Azərbaycan dilinin tezlik lüğətinin III cildi söz-formaları əhatə edir.

Söz-forma dedikdə, müəyyən qrammatik formada çıxış edən sözlər qrupu nəzərdə tutulur. Məsələn, “Mən dənizi se-

virəm”, “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 gəlmişdir.

Təbiidir ki, söz-formalar tezlik lüğəti həcmcə söz-kökləri və əsaslardan böyük olmalıdır. Ona görə də, bu lüğətdə tezliyi yüksək olan söz-formaların əhatə olunması nəzərdə tutulur.

Söz kökləri dilin lüğət tərkibinin əsasını təşkil edir. Azərbaycan dilinin orfoqrafiya lüğətinin son çapında (2004) 80000-ə yaxın sözün düzgün yazılışı verilmişdir. Söz köklərinin sayı isə əslində xeyli azdır. Azərbaycan dilinin sözdüzəltmə imkanları hesabına lüğətdəki sözlərin sayı artırılmış və 80000-ə çatdırılmışdır. Bu baxımdan orfoqrafiya lüğəti Azərbaycan dilinin lüğət tərkibi barədə obyektiv məlumat vermir. Sözdüzəldici şəkilçilərin köməyi ilə yaranmış yeni sözlər (əsaslar) lüğətin həcmnin artmasına səbəb olmuşdur. Bu, həm ad nitq hissəsinə, həm də fellərə aiddir. Təbii nitq prosesində sözdəyişdirici (qrammatik) şəkilçilərin də fəallığı artır. Nəticədə söz köklərinə artırılan sözdüzəldici (leksik) və sözdəyişdirici şəkilçilərin köməyi ilə bir söz kökündən minlərlə (on minlərlə) yeni sözlər (söz-formalar) yaranmış olur. Fel kökləri sözdüzəltmə və sözdəyişdirmə imkanlarına görə daha fəaldır.

Məsələn: *bax* fel kökündən leksik şəkilçilərlə düzələn lüğət vahidlərini nəzərdən keçirək: – *baxar, baxarlı, baxdırma, baxdırmaq, baxəbər, baxıcı, baxıcılıq, baxılış, baxılma, baxılmaq, baxım, baxımlı, baxımlılıq, baxımsız, baxımsızlıq, baxınma, baxınmaq, baxış, baxışlı, baxışma, baxışmaq, baxma, baxa-baxa, baxmadan, baxmaq, baxmalı, baxmayaraq* və s. (Azərbaycan dilinin orfoqrafiya lüğəti, 2004, s. 68). Təbii nitq prosesində işlənmə imkanları da nəzərə alınsa, eyni fel kökündən minlərlə söz və söz-formanın yaranmasının şahidi olarıq: *baxıram, baxırsan, baxır, baxırıq, baxırsınız, baxırlar, baxdım, baxdın, baxdı, baxdıq, baxdınız, baxırlar, baxacağam* və s.

Deyilənlər ad nitq hissəsinə də aiddir. Məsələn: *beton, betonçu, betonçuluq, betonlama, betonlamaq, betonlanma, betonlanmaq,*

betonlaşma, betonlaşmaq, betonlatdırma, betonlatdırmaq, betonlatma, betonlatmaq, betonlu və s.

Bura birinci komponenti *beton* olan mürəkkəb sözləri də əlavə etmək olar: – *betondaşıyan, betondələn, betondöşəyən, betonqarışdıran, betonqarışdırıcı, betonqarışdırma, betonsındıran, betontökən* və s.

Nəzərə alsaq ki, *beton* sözü müxtəlif qrammatik şəkilçilərlə də işlənə bilər, onda eyni söz kökündən əmələ gəlmiş söz və söz formalarının sayını xeyli artırmaq olar.

Bütün bunlar belə bir nəticəyə gəlməyə imkan verir ki, türk dillərində, o cümlədən Azərbaycan dilində söz köklərinin müəyyənləşdirilməsi və təhlili dilin lüğət tərkibi, onun nüvəsi barədə dəqiq məlumat baxımından olduqca əhəmiyyətlidir.

Məhz bu baxımdan “Azərbaycan dilinin tezlik lüğəti (söz kökləri)” olduqca əhəmiyyətli və faydalıdır.

Azərbaycan dilinin tezlik lüğətinin (söz kökləri) tərtibi üçün bütün üslublar (elmi, bədii, publisist, rəsmi, məişət və s.) təmsil olunmaqla 50 milyondan artıq söz-forma linqvostatistik təhlilə cəlb olunmuşdur. Müqayisə üçün qeyd edək, 1977-ci ildə çap olunmuş rus dilinin tezlik lüğətində cəmi 1 milyon söz-forma (sözləşmədən) istifadə olunmuşdu. Nəticədə təxminən 40 min sözdən ibarət lüğət tərtib olunmuşdu. Həmin lüğətin tərtibində 100-ə yaxın adam iştirak etmişdi (Bax.: Частотный словарь русского языка, 1977, s.3). Azərbaycan dilinin tezlik lüğətində 50 dəfədən çox (50 milyon) söz-forma işlənmişdir. Lüğətin tərtibində isə cəmi 8 nəfər iştirak etmişdir. Burada informasiya texnologiyalarının müasir inkişaf səviyyəsi, yeni, daha sürətli və böyük yaddaşa malik kompüterlərin mövcudluğu və informasiyanın daha müasir üsullarla və vasitələrlə kompüterə daxil edilməsi amili mühüm rol oynayır.

Elmi, bədii, publisist, rəsmi, məişət və s. üslublara aid materialların elektron variantların mövcudluğu tədqiqatın daha geniş əhatə dairəsinə malik olması imkanı yaratmışdır. Statistik seçmə baxımından da etibarlılıq dərəcəsi yüksəkdir.

Azərbaycan dilinin tezlik lüğətində ilkin material kimi götürülmüş söz-formalar çoxluğu “saf-çürük” edilmiş, rəqəmlər, müxtəlif simvollar, təsadüfi sözlər, xüsusi adlar, lüzumsuz olaraq işlənmiş əcnəbi söz və ifadələr və s. sözlükdən çıxarılmışdır. İlkin işlənmədən sonra sözlükdə təxminən 300 min söz-forma qalmışdır. Həmin baza əsasında müxtəlifçeşidli tezlik lüğətlərinin (əsasların, söz-formaların tezlik, əlifba-tezlik və əks-əlifba lüğətləri, eləcə də qrammatik və leksik şəkilçilərin, qrafem və qrafem birləşmələrinin, durğu işarələrinin tezlik səciyyələrini əks etdirən siyahılar və s.) hazırlanması nəzərdə tutulur.

5.4. 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.ə. 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 əvvəlində doğru (sağdan sola) davam etdirilir. Məsələn, əks lüğətdə *sınaq* sözü *alim* sözündən əvvəl yerləşdirilir. Çünki *q* hərfi əlifba sırasına görə *m* hərfindən daha irəlidedir. *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:

<i>adi lüğət</i>	<i>əks lüğət</i>
ağır	şura
çiçəkli	çiçəkli
oturmaq	təsəlli
şura	udmaq
təsəlli	oturmaq
udmaq	ağır

Beləliklə, adi lüğətlərdə əvvəli eyni hərfərdən ibarət olan sözlər bir yerdə yerləşirsə, əks lüğətdə sonu eyni hərfərdən ibarət sözlər bir sırada düzülüb olur. Belə ki, Azərbaycan dilinin əks lüğətində bütün fellər, daha doğrusu *-maq* və *-mək* şəkilçili fel 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 fellərin) 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 (ОЧСЛЯ, 1976).

Əks lüğət morfologiya 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 sözformaların 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.

Əks lüğətlər adətən lüğət-indeks səciyyəsi daşıyır. Bu və ya digər adi lüğətə daxil olan bütün sözlər burada əks əlifba sırası ilə düzülür. Bununla belə, bir neçə adi lüğət əsasında tərtib edilmiş lüğət-indeks də ola bilər. Azərbaycan lüğətçiliyinin ilk təcrübəsi olan əks lüğət “Azərbaycan dilinin orfoqrafiya lüğəti” (Bakı, 1975) əsasında tərtib olunmuşdur. Bundan başqa bir neçə yüz söz “Azərbaycanca-rusca lüğət”dən (Bakı, 1965) əlavə edilmişdir. Bu vaxta qədər türkologiyada yalnız özbək, qazax, türk dilləri üçün əks lüğətlər tərtib olunmuşdur (Куньиров, Тихонов, 1968; Бекраев, 1971; Vietze, Zenker, Warnke, 1975). Təxminən 60 min sözü əhatə edən Azərbaycan dilinin əks lüğəti türk dilləri

üzrə tərtib olunmuş lüğətlər içərisində həcminə görə ən böyüyüdür.

Azərbaycan dilinin əks lüğəti Azərbaycan EA Nəsimi adına Dilçilik İnstitutunda tətbiqi dilçilik qrupunun əməkdaşları tərəfindən hazırlanmışdır.

Yuxarıda göstəriləni kimi, Azərbaycan dilinin leksikasını daha tam şəkildə əhatə etmək məqsədilə sözlük kimi orfoqrafiya lüğəti əsas götürülmüşdür. Ona görə də, əks lüğət xeyli terminoloji, məhəlli sözləri ehtiva etmiş olur. Bundan başqa, əks lüğətdəki bir sıra sözlərin mənasını mövcud lüğətlərdə tapmaq qeyri-mümkündür. Lakin lüğətlər üçün mühüm amil sayılan sözlük tamlığı cəhətdən bu hala müəyyən dərəcədə haqq qazandıran olar.

Haqqında danışılan lüğət ötən əsrin 80-ci illərində “Elm” nəşriyyatına təqdim olunmuşdu. Çox təəssüf ki, “Azərbaycan qəzet dilinin tezlik lüğəti” kimi o da müxtəlif səbəblərə görə həmin illər çap olunmadı. İş elə gətirdi ki, çap növbəsində köhnəlmiş, “elmi əhəmiyyətini” itirmiş həmin lüğətin yeni nəşrinin hazırlanması zərurəti yenidən yarandı. Əslində, informatikanın, internet sisteminin, elektron lüğətlərin inkişaf etdiyi, geniş yayıldığı indiki dövrdə əks əlifba lüğətlərinə böyük ehtiyac vardır.

80-ci illərdə çapa təqdim olunmuş Azərbaycan dilinin əks əlifba lüğəti Azərbaycan dilinin orfoqrafiya lüğətinin artırılmış və təkmilləşdirilmiş üçüncü nəşri əsasında hazırlanmışdı (Azərbaycan dilinin orfoqrafiya lüğəti. Bakı, “Elm”, 1975). Məlum olduğu kimi, həmin lüğətin sözlüyü qismən köhnəlmiş, son dövrlər, xüsusən müstəqillik illərində dilimizə xeyli yeni söz və ifadələr keçmiş, dilimizin lüğət tərkibi zənginləşmişdir. Təbii ki, həmin sözlük yeni nəşrdə əsas götürülə bilməzdi. 2004-cü ildə “Nurlan” nəşriyyatında çap olunmuş “Azərbaycan dilinin əks əlifba lüğəti” Azərbaycan dilinin orfoqrafiya lüğətinin yeni, son variantı əsasında tərtib olunmuşdur. Latın qrafikalı yeni Azərbaycan əlifbasına keçidlə əlaqədar lüğətin əlifba düzülüşü də dəyişdirilmiş, yeni əlifbaya uyğunlaşdırılmışdır. Əgər əvvəlki variantda lüğətdə təxminən 60 min söz verilmişdisə, çap olunmuş

kitab təxminən 80 minə yaxın sözü əhatə edir (Azərbaycan dilinin əks əlifba lüğəti, 2004).

5.5. 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 səciyələndirilməsi 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 işlənə 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 sözdəyişdirmə 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. (Bax: Vəliyeva, Mahmudov, Pines, Rəhmanov, Sultanov, 1997; Şıxıyeva, 1995.)

1997-ci ildə “Kitabi-Dədə Qorqud” dastanlarının 1300 illik yubileyi ərəfəsində “Kitabi-Dədə Qorqud” dastanlarının statistik təhlili kitabı çap olundu (Bakı, “Siyasət”, 1997).

“Kitabi-Dədə Qorqud” dastanlarının statistik təhlili Azərbaycan dilinin tarixinin öyrənilməsi baxımından olduqca əhəmiyyətliyətlidir. İlk əvvəllər abidənin tənqidi mətni olmadığı üçün akad. Həmid Araslı tərəfindən nəşr edilmiş variant üzərində iş aparılırdı (Kitabi-Dədə Qorqud. Bakı, “Gənclik”, 1978). Məlum olduğu kimi, həmin sadələşdirilmiş variantda dastan mətni müasir Azərbaycan dilinə uyğunlaşdırılmış, bir çox fonetik, orfoqrafik, qrammatik təhriflərə yol verilmişdi. Təbii ki, həmin variant əsasında aparılmış təhlilin nəticələri dil tarixi üçün qiymətli faktlar

verə bilməzdi. Məhz bu səbəbdən 1988-ci ildə “Kitabi-Dədə Qorqud”un mükəmməl tənqidi mətni çap olunduqdan sonra (tərtibçilər: F.Zeynalov, S.Əlizadə) dastanın linqvostatistik təhlilinə başlandı. Maliyyə çətinliyinə görə əsəri tam şəkildə nəşr etmək imkanı olmadığından, ilk növbədə, əlifba-tezlik siyahısının çapı məqsəduyğun hesab olunmuşdur. Kitabça tərtibçilərin öz vəsaitləri hesabına çap olunmuşdu və Azərbaycan xalqının yazılı abidəsi “Kitabi-Dədə Qorqud”un 1300 illik yubileyinə müəlliflərin kiçik hədiyyəsi idi (“Kitabi-Dədə Qorqud” dastanlarının statistik təhlili, 1997).

Həmin kitabçanın çapı tərtibçilər üçün düşərli oldu. “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 (Bakı, “Elm”, 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. Kitabın tərtibçiləri K.A.Vəliyeva, M.Ə.Mahmudov, V.Y.Pines, C.Ə.Rəhmanov, V.S.Sultanov, redaktorları filologiya elmləri doktoru, professor A.A.Axundov və filologiya elmləri doktoru M.Ə.Mahmudov, mühəndis-programlaşdırıcı O.S.Məmmədov, rəyçiləri filologiya elmləri doktoru, professor K.N.Vəliyev, filologiya elmləri namizədi F.Z.Əhmədov, ön sözün müəllifi filologiya elmləri doktoru M.Ə.Mahmudovdur. (“Kitabi-Dədə Qorqud” dilinin statistik təhlili (ilkin nəticələr), Bakı, “Elm”, 1999, -248 s.).

“Kitabi-Dədə Qorqud” dastanlarının statistik təhlili sahəsində uğurlu tədqiqatlar dil tarixinə kompüterlərin tətbiqinə stimul yaratmış oldu. Hazırda görkəmli Azərbaycan klassiki Məhəmməd Füzulinin dilinin kompüterlərlə tədqiqi ilə əlaqədar iş

baş a çatdırılmışdır. Füzulinin Azərbaycan divanlarının statistik baxımdan hərtərəfli tədqiqi böyük şairin söz xəzinəsi barədə bir çox qiymətli faktlar almaq imkanı yaratmışdır.

2004-cü ildə “Məhəmməd Füzulinin nəzm əsərinin əlifba-tezlik sözlüyü çapdan çıxdı (Bakı, “Elm”, 2004. -548 s.).

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, sözdəyişdirmə 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 mə-

lumat 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.

Tərtibçilər – K.A.Vəliyeva, M.Ə.Mahmudov, C.Ə.Rəhmanov, V.S.Sultanov, redaktorlar – AMEA-nın müxbir üzvü A.A.Axundov, filologiya elmləri namizədi A.B.Mirzəyev, rəyçilər – filologiya elmləri doktoru, professor R.C.Məhərrəмова, filologiya elmləri doktoru, professor A.Q.Ələkbərov, mühəndis-proqramlaşdırıcı O.S.Məmmədov, ön sözün müəllifi filologiya elmləri doktoru M.Ə.Mahmudovdur.

Tərtib olunmuş tezlik lüğətləri Azərbaycan dilinin quruluşu, qrammatik xüsusiyyətləri, leksik özəllikləri, sözdəyişdirmə və sözəmələgətmə tipləri, inkişaf meylləri 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: *administ-rasiya* (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), *de-partament* (6276), *şou* (1184), *şou-biznes* (709), *şou-proqram* (74), *kommersiya* (5106), *presedent* (1074), *informasiya* (32242), *informasiya-kommunikasiya* (764), *informatizasiya* (36), *disk* (1492), *display* (91), *modern* (1964), *modifikasiya* (178), *modul* (1061), *kompyuter* (6873), *kompyuter-terminal* (14) və s.

Bəzi söz və ifadələr isə paralel olaraq işlənməkdədirlər. 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); Sözköklərinin şəkilsiz variantı ilə saitlə başlanan şəkilsiz variantlarının işlənmə tezlikləri də müxtəlifdir (söhbət sonu -q, -k şəkilsizləri ilə bitən sözköklərindən gedir. Saitlə başlanan şəkilsiz qəbul etdikdə -q(-ğ), -k(-y) əvəzlənməsi baş verir. Eləcə də sonu -t ilə bitən sözkökləri 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) — qayıt (10954)
 öyrət (1377) — öyrəd (1745)
 kömək (22799) — köməy (6860)
 ördək (129) — ördəy (28)
 örnək (1544) — örnəy (178)
 ömək (54) — öməy (51)
 bayraq (3222) — bayraq (1796)
 bıçaq (315) — bıçağ (3734)
 budaq (1032) — budağ (264)
 çıraq (1256) — çırağ (305)
 çiçək (3282) — çiçəy (431)
 çörək (3324) — çörəy (1603)
 xörək (674) — xörəy (180)

Mənsubiyyət və hal şəkilsizləri qəbul etdikdə sözün ikinci hecasında qapalı saitin 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 fellər (50 fel) 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 vardır. Azərbaycan fellərinin tarixi inkişafı və bütövlükdə leksik qatın inkişaf tarixi və meyilləri 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 meyilləri, yeni söz və ifadələrin dilimizə keçməsi məsələləri linqvostatistik fonda araşdırma obyektinə ola bilər.

Cədvəl 4. Azərbaycan dilinin tezlik lüğətində yüksək tezlikli fellər (50 fə)

1. ol(maq)	26. çatmaq
2. ed(et(mək)	27. qal(maq)
3. gör(mək)	28. danış(maq)
4. ver(mək)	29. qoy(maq)
5. keç(mək)	30. üz(mək)
6. de(mək)	31. ged/get(mək)
7. bil(mək)	32. çək(mək)
8. gəl(mək)	33. vur(maq)
9. çıx(maq)	34. dəyiş(mək)
10. göstər(mək)	35. gətir(mək)
11. art(maq)	36. yet(mək)
12. al(maq)	37. saxla(maq)
13. böyü(mək)	38. tik(mək)
14. bildir(mək)	39. qaz(maq)
15. apar(maq)	40. yaran(maq)
16. istə(mək)	41. öl(mək)
17. seç(mək)	42. çalış(maq)
18. tut(maq)	43. ötmək
19. yarad/yarat(maq)	44. at(maq)
20. qur(maq)	45. din(mək)
21. bax(maq)	46. kəs(mək)
22. yaz(maq)	47. götür(mək)

- | | |
|---------------|----------------|
| 23. elə(mək) | 48. topla(maq) |
| 24. düş(mək) | 49. qazan(maq) |
| 25. yaşa(maq) | 50. sat(maq) |

Cədvəl 5. Azərbaycan qəzet dilinin tezlik lüğətində yüksək tezlikli fellər (50 fel)

- | | |
|------------------|----------------------|
| 1. et(mək) | 26. başla(maq) |
| 2. ol(maq) | 27. keç(mək) |
| 3. ver(mək) | 28. get(mək) |
| 4. göstər(mək) | 29. hazırla(maq) |
| 5. yetir(mək) | 30. çat(maq) |
| 6. edil(mək) | 31. qoy(maq) |
| 7. artır(maq) | 32. danış(maq) |
| 8. olun(maq) | 33. bildir(mək) |
| 9. de(mək) | 34. çıx(maq) |
| 10. keçir(mək) | 35. çatdır(maq) |
| 11. al(maq) | 36. öyrən(mək) |
| 12. gör(mık) | 37. götür(mək) |
| 13. art(maq) | 38. istə(mək) |
| 14. istə(mək) | 39. sal(maq) |
| 15. gəl(mək) | 40. aç(maq) |
| 16. bil(mək) | 41. qal(maq) |
| 17. işlə(mək) | 42. genişləndir(mək) |
| 18. keçiril(mək) | 43. ödə(mək) |
| 19. yüksəlt(mək) | 44. tutul(maq) |
| 20. yarat(maq) | 45. yaz(maq) |
| 21. veril(mək) | 46. düş(mək) |
| 22. etdir(mək) | 47. qarşıla(maq) |
| 23. apar(maq) | 48. burax(maq) |
| 24. görül(mək) | 49. açıl(maq) |
| 25. çalış(maq) | 50. salın(maq) |

Cədvəl 6. "Kitabi-Dədə Qorqud" da yüksək tezlikli fellər

(50 fel)

- | | |
|----------------|-----------------|
| 1. de(mək) | 26. tut(maq) |
| 2. ay(dır) | 27. dön(mək) |
| 3. gəl(mək) | 28. çap(maq) |
| 4. ol(maq) | 29. tur(maq) |
| 5. soylə(maq) | 30. qal(maq) |
| 6. gör(mək) | 31. dilə(mək) |
| 7. al(maq) | 32. sor(maq) |
| 8. var(maq) | 33. et(mək) |
| 9. ver(mək) | 34. ögin(mək) |
| 10. söylə(mək) | 35. qaq(maq) |
| 11. gir(mək) | 36. çal(maq) |
| 12. yat(maq) | 37. as(maq) |
| 13. qoy(maq) | 38. qaç(maq) |
| 14. ur(maq) | 39. qıy(maq) |
| 15. baq(maq) | 40. yet(mək) |
| 16. qalq(maq) | 41. at(maq) |
| 17. keç(mək) | 42. dər(mək) |
| 18. eylə(mək) | 43. kəs(mək) |
| 19. get(mək) | 44. bil(mək) |
| 20. ayıt(maq) | 45. çıxar(maq) |
| 21. en(mək) | 46. sıçra(maq) |
| 22. gətir(mək) | 47. öldür(mək) |
| 23. ağla(maq) | 48. sal(maq) |
| 24. bin(mək) | 49. bas(maq) |
| 25. düş(mək) | 50. qurtar(maq) |

Cədvəl 7. Məhəmməd Füzulinin azərbaycandilli nəzm əsərlərində yüksək tezlikli fellər (50 fel)

- | | |
|--------------|----------------|
| 1. ol(maq) | 26. gəz(mək) |
| 2. eylə(mək) | 27. qaç(maq) |
| 3. et(mək) | 28. qətir(mək) |
| 4. qıl(maq) | 29. keç(mək) |

- | | |
|----------------|-----------------|
| 5. de(mək) | 30. ayrıl(maq) |
| 6. ver(mək) | 31. dol(maq) |
| 7. yet(mək) | 32. götür(mək) |
| 8. gör(mək) | 33. başla(maq) |
| 9. düş(mək) | 34. bax(maq) |
| 10. gəl(mək) | 35. görün(mək) |
| 11. bil(mək) | 36. neylə(mək) |
| 12. qal(maq) | 37. unut(maq) |
| 13. çək(mək) | 38. əy(mək) |
| 14. tut(maq) | 39. iç(mək) |
| 15. çıx(maq) | 40. tap(maq) |
| 16. bul(maq) | 41. yandır(maq) |
| 17. dilə(mək) | 42. dön(mək) |
| 18. get(mək) | 43. dur(maq) |
| 19. ur(maq) | 44. oyna(maq) |
| 20. tök(mək) | 45. yaz(maq) |
| 21. görün(mək) | 46. ağla(maq) |
| 22. açıl(maq) | 47. bənzət(mək) |
| 23. art(maq) | 48. cizgin(mək) |
| 24. burax(maq) | 49. döndər(mək) |
| 25. eşit(mək) | 50. yor(maq) |

Cədvəl 8. Azərbaycan dilinin tezlik lüğətində yüksək tezlikli sözlər (100 söz)

- | | |
|---------------|---------------|
| 1. və | 51. daha |
| 2. ol(maq) | 52. bağ |
| 3. bu | 53. yol |
| 4. ed,et(mək) | 54. apar(maq) |
| 5. bir | 55. iki |
| 6. ki | 56. bütün |
| 7. gör(mək) | 57. hazır |
| 8. ilə | 58. həm |
| 9. il | 59. var |
| 10. ver(mək) | 60. məsafə |
| 11. o | 61. həyat |

- | | |
|-----------------|-----------------|
| 12. keç(mək) | 62. söz |
| 13. baş | 63. inkişaf |
| 14. de(mək) | 64. nə |
| 15. iş | 65. istə(mək) |
| 16. üçün | 66. barə |
| 17. bil(mək) | 67. üzrə |
| 18. öz | 68. dünya |
| 19. çox | 69. xalq |
| 20. dövlət | 70. aç(maq) |
| 21. biz | 71. zaman |
| 22. pılkə | 72. rayon |
| 23. mən | 73. əl |
| 24. kim | 74. nəticə |
| 25. gəl(mək) | 75. haqq |
| 26. gün | 76. seç(mək) |
| 27. isə | 77. ad |
| 28. yer | 78. insan |
| 29. onlar | 79. sahə |
| 30. məlumat | 80. tut(maq) |
| 31. sonra | 81. sir |
| 32. çıx(maq) | 82. fəaliyyət |
| 33. on | 83. əmək |
| 34. nazir | 84. şəhər |
| 35. ara | 85. qədər |
| 36. prezident | 86. milli |
| 37. göstər(mək) | 87. iştirak |
| 38. əsas | 88. son |
| 39. tərəf | 89. yarad/yarat |
| 40. art(maq) | 90. hesab |
| 41. al(maq) | 91. əlaqə |
| 42. yeni | 92. vaxt |
| 43. böyü(mək) | 93. qur(maq) |
| 44. respublika | 94. şəxs |
| 45. belə | 95. bax(maq) |
| 46. bildir(mək) | 96. yan |

47. hər	97. xidmət
48. qarşı	98. daxil
49. təşkil	99. heç
50. qeyd	100. həmin

Cədvəl 9. Azərbaycan qəzet dilinin tezlik lüğətində yüksək tezlikli sözlər (100 söz)

1. və	51. göstər(mək)
2. et(mək)	52. yetir(mək)
3. bu	53. istehsal
4. bir	54. komitə
5. yüz	55. edil(mək)
6. o	56. ssri
7. ol(maq)	57. yetmiş
8. il	58. ad
9. partiya	59. həyat
10. ilə	60. artır(maq)
11. sovet	61. yüksək
12. min	62. bir
13. ki	63. inkişaf
14. azərbaycan	64. üç
15. iş	65. yoldaş
16. üçün	66. beşinci
17. doqquz	67. yoldaş
18. respublika	68. kolxoz
19. iyirmi	69. vəzifə
20. xalq	70. təşkilat
21. təsərrüfat	71. yol
22. yer	72. qərar
23. çox	73. olan
24. on	74. olun(maq)
25. də	75. de(mək)
26. böyük	76. həmin
27. iki	77. ittifaq
28. da	78. keyfiyyət

29.	sov.ikp	79.	siyasi
30.	rayon	80.	tədbir
31.	daha	81.	müəssisə
32.	məhsul	82.	idarə
33.	qurultay	83.	katib
34.	bütün	84.	kollektiv
35.	beşillik	85.	dörd
36.	sahə	86.	şəhər
37.	ver(mək)	87.	ton
38.	əmək	88.	sənaye
39.	kənd	89.	mübarizə
40.	hər	90.	kimi
41.	ölkə	91.	nəzər
42.	plan	92.	səkkiz
43.	kommunist	93.	keçir(mək)
44.	beş	94.	yeddi
45.	mk	95.	uğrunda
46.	baş	96.	pambıq
47.	dövlət	97.	proqram
48.	gün	98.	vaxt
49.	öz	99.	al(maq)
50.	yeni	100.	yarış

Cədvəl 10. "Kitabi-Dədə Qorqud" da yüksək tezlikli sözlər (100 söz)

1.	de(mək)	51.	iki
2.	ay(dır)	52.	ana
3.	bir	53.	məgər
4.	qara	54.	baş
5.	qazan	55.	və
6.	nə	56.	qaba
7.	gəlmək	57.	kafir
8.	ağ	58.	yağı
9.	xan(m)	59.	dünya
10.	bu	60.	gir(mək)
11.	ol	61.	baba

- | | | | |
|-----|------------|-----|------------|
| 12. | oğul | 62. | əl |
| 13. | mərə | 63. | ya |
| 14. | dəxi | 64. | yat(maq) |
| 15. | oğuz | 65. | saqqal(lu) |
| 16. | yigit | 66. | xatun |
| 17. | kim | 67. | gəl(mək) |
| 18. | ol(maq) | 68. | ox |
| 19. | soyla(maq) | 69. | xoş |
| 20. | mən | 70. | altmış |
| 21. | man(a) | 71. | bin |
| 22. | gör(mək) | 72. | göz |
| 23. | at | 73. | içun |
| 24. | ala | 74. | ur(maq) |
| 25. | al(maq) | 75. | baq(maq) |
| 26. | oğlan | 76. | qalq(maq) |
| 27. | sən | 77. | qatı |
| 28. | kafər | 78. | tənri |
| 29. | bəg | 79. | alp |
| 30. | üzəri(nə) | 80. | keç(mək) |
| 31. | var(maq) | 81. | beş |
| 32. | var | 82. | qadir |
| 33. | qız | 83. | eylə(mək) |
| 34. | san(a) | 84. | necə |
| 35. | qırq | 85. | sağ |
| 36. | ver(mək) | 86. | hey |
| 37. | gün | 87. | get(mək) |
| 38. | qoca | 88. | allah |
| 39. | üç | 89. | qurban |
| 40. | bura | 90. | çıq(maq) |
| 41. | dəli | 91. | biz |
| 42. | ər | 92. | qan |
| 43. | söylə(mək) | 93. | dədə |
| 44. | xəbər | 94. | ev |
| 45. | görk | 95. | en(mək) |
| 46. | yer | 96. | gətür(mək) |

- | | |
|-----------|---------------|
| 47. qalın | 97. ağla(maq) |
| 48. yoq | 98. düş(mək) |
| 49. adam | 99. tut(maq) |
| 50. ilə | 100. gərək |

Cədvəl 11. Məhəmməd Füzulinin azərbaycandilli nəzm əsərlərində yüksək tezlikli sözlər (100 söz)

- | | |
|---------------|--------------|
| 1. kim | 51. nola |
| 2. ki | 52. mey |
| 3. bir | 53. dəxi |
| 4. ol | 54. yet(mək) |
| 5. bu | 55. fələk |
| 6. ilə | 56. aşıq |
| 7. ey | 57. dil |
| 8. hər | 58. izhar |
| 9. nə | 59. tut(maq) |
| 10. ol(maq) | 60. böylə |
| 11. eylə(mək) | 61. yox |
| 12. mən | 62. müdam |
| 13. var | 63. əcəb |
| 14. həm | 64. əlbəttə |
| 15. qər | 65. imiş |
| 16. yenə | 66. dərd |
| 17. et(mək) | 67. guya |
| 18. o | 68. məgər |
| 19. yox | 69. aləm |
| 20. tək | 70. daim |
| 21. gül | 71. əmma |
| 22. qıl(maq) | 72. həmişə |
| 23. idi | 73. neçə |
| 24. can | 74. rəhm |
| 25. çün(üçün) | 75. zar |
| 26. çox | 76. ərz |
| 27. eşq | 77. key |
| 28. gün | 78. öz |

29.	ta	79.	əhl
30.	min	80.	iki
31.	içrə	81.	qələm
32.	q(ğ)əm	82.	ləhzə
33.	dəm	83.	gəl(mək)
34.	yar	84.	gərçi
35.	və	85.	şad
36.	gör(mək)	86.	tərk
37.	qöz	87.	lütf
38.	üzrə	88.	zahir
39.	könül	89.	xəlq
40.	değil	90.	billah
41.	hiç	91.	gərək
42.	xoş	92.	yad
43.	necə	93.	ancaq
44.	özgə	94.	gərdun
45.	yüz	95.	nəzər
46.	qan	96.	çək(mək)
47.	nə	97.	çərt
48.	ah	98.	mütləq
49.	ver(mək)	99.	zaman
50.	ikən	100.	gecə

5.6. 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 (Плютовский, 1977). Əgər Hind-Avropa dilləri üçün AL söz-formalar

əsasında tərtib oluna bilsə, türk dilləri üçün bu xüsusiyyət məqbul sayıla bilməz. Aqlütinativ quruluşa malik olan türk dillərində, o cümlədən Azərbaycan dilində eyni fel 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 aşağıdakılardan ibarətdir:

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

5.6.1. 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 saxlanması 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. Hazırda 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 (bax: Кесенбаев, Пиотровский, Бектаев, 1970, 16).

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 (Бектаев, 1978, 55).

Yuxarıda göstərilənləri nəzərə alaraq 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 əsasları 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 (Cədvəl 4.). 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. Fəllər 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 fəllin məsdərdə verilməsi təhlil prosesində bir çox lüzumsuz anlaşılmazlıqların yaranmasına gətirib çıxardı. 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 alqoritminin işindən sonra söz-forma morfemlərə ayrılır: *yaz-ır-am*. Maşın *-ır* və *-am* şəkilçilərinin, şəkilçilər cədvəli ilə müqayisəsindən sonra müvafiq qrammatik informasiyanı əldə edəcək:

- *ir*-indiki zaman (fel)

- *am* – xəbərlik kateqoriyası, I şəxs, tək və s.

Şəkilçilərə görə məlum olur ki, *yaz* omonimi ad mənasında (“*yaz*”- “*вечна*”) deyil, fel kimi tərcümə olunmalıdır. Əsaslar AL-də *yaz* felini tapır və onun mənasına şəkilçilərin mənalərini əlavə etməklə söz-formanın bütövlükdə qrammatik sintezini və dəqiq tərcüməsini alır. Əgər göstərilən fel AL-də məsdərdə təmsil olunsaydı, onda *yaz* əsasını orada birbaşa tapmaq mümkün olmazdı, təhlil alqoritmlərindən sonra yardımçı sintez alqoritmlərinə müraciət etmək zərurəti yaranardı. Sözü əsasının (bizim misalda *yaz*) fel olduğunu müəyyənləşdirdikdən sonra sintez alqoritm *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. Fellərin lüğətdə əmr şəklində verilməsi prosesdəki bu artıq əməliyyatları ixtisar etməyə imkan verir.

Cədvəl 12 Azərbaycanca-rusca əsaslar avtomatik lüğətindən bir fragment

Əsas	Nitq hissəsinə mənsubiyyəti	Tərcüməsi
Bayraq	isim	флаг, знамя
bayraq		
yoldaş	isim	товарищ
parlaq	sifət	яркий, блестящий
paltar	isim	одежда, одеяние
səkkiz	say	восемь
səksən	say	восемьдесят
səhifə	isim	страница

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 fel kök-

ləri saitlə başlanan şəkilçi qəbul etdikdə *t-d* cingiltiləşməsi 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. (Cədvəl 13.)

***Cədvəl 13. Azərbaycanca-rusca əsaslar AL-də sonu "q"
samiti bitən fəllərin verilməsi***

Əsas (fel)	Tərcüməsi
unut unud	забыть, забывать
eşit eşid	слышать, услышать
qayıt qayıd	вернутся, возвращаться
get ged	идти, уходить
et ed	делать, сделать

Morfoloji proses sonu *q*, *k* samitləri ilə bitən çoxecali adlarda da geniş müşahidə olunur:

Torpaq - torpağa, torpağı
Yarpaq - yarpağa, yarpağı
çiçək - çiçəyə, çiçəyi
çörək - çörəyə, çörəyi və s.

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

torpaq / torpağ
yarpaq / yarpağ
çiçək / çiçəy

çörək / çörəy

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

Morfoloji səviyyədə omonimliyi aradan qaldırıla bilməyən əsaslar AL-də nitq hissələrinə potensial mənsubiyyətləri göstərməklə verilmişdir (Cədvəl 14.).

Cədvəl 14. Omonim əsasların AL-də verilməsi

Omonim əsas	Nitq hissəsinin mənsubiyyəti	Tərcüməsi
Qaz	существительное/ глагол	гусь/газ копать
Yaz	глагол существительное/ глагол	весна писать

Sonrakı mərələlərdə əsasların omonimliyi ona birləşmiş şəkilçilərə 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 hesablandıq-

dan 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 (Барабанов, 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 işlənmə 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) “Statistika reçi” 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 “ölənlərin” (umerevşix) söz forması da verilmiş, “politika” sözündən əlavə “politikamızın” (нашей политики) söz-forması ayrılıqda təsbit olunmuşlar və s.

5.6.2. 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 mətnindən 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” küməkçi felləri ilə və bir çox başqa fellərlə (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. Aşağıdakı cədvəldə (Cədvəl 15) komponentləri ayrılıqda təhlil və tərcümə olunan birləşmələr timsalında bütün bu qəbildən olan feli birləşmələr haqqında müəyyən təsəvvür əldə etmək olar.

Cədvəldəki misalların sayını bir qədər də çoxaltmaq olardı. Lakin elə həmin misallardan da Azərbaycan dilində ikinci komponenti feldən 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. Cədvəldən görüldüyü kimi, *baş* sözü ikinci komponentdəki feldən asılı olaraq *случиться, возникнуть*, “наведываться, навестить”, “вмешиваться, обращать внимание” kimi tərcümə olunur. Analoji xüsusiyyət ikinci komponenti yuxarıda sadaladığımız fellər ola biləcək bütün birləşmələrə xasdır. Deməli, birləşmələr AL-də sintaktik bloklara əlaqədar bir çox amillər mütləq nəzərə alınmalı, ikinci komponenti fel olan bütün birləşmələr ayrılıqda deyil, sintaktik vahid kimi tədqiq olunmalı, təhlil və tərcümə edilməlidir. Bunun üçün hər şeydən əvvəl ayrı-ayrı fel 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 fel olan sintaktik vahidlər iyirmi qrupa ayrılmış, hər qrupda konkret felə (məsələn, “et”) birləşmək potensialına malik sözlər verilmişdir.

Cədvəl 15. Feli birləşmələrin komponentlər üzrə tərcüməsi

Söz birləşməsi	Ayrı-ayrı komponentlər üzrə tərcümə	Söz birləşməsinin düzgün tərcüməsi
baş ver (mək)	qolova, головной+давать, отдавать	случится, возникать вмешиваться,
baş qoş (maq)	головной+запрягать, присоединять	обращать внимание
baş çək (mək)	голова, головной+тянуть, таскать сердце,	наведываться, навестить
ürək ver (mək)	душа+давать, отдавать	ободреть, подбодрять
dərd çək (mək)	горе, скорбь+тянуть, таскать рука,	горевать, подписывать,
qol çək (mək)	ветвь+тянуть, таскать	расписывать

Formal təhlil və MT sistemləri üçün çox vacib olan bir xüsusiyyəti də qeyd etmək lazımdır: birləşmələr AL açıq sistemdir. İstər fel 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 hərflərin sayı nə-

zərə alınmaqla, azalan siyahı ilə verilir. Eyni uzunluğa malik sözlər öz aralarında əlifba siyahısı ilə düzülürlər. 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. Birləşmələrin ikinci komponenti olan fellər imperativdə verilir. Aşağıdakı cədvəldə birləşmələr AL-də vahidlərin yerləşdirilməsi göstərilmişdir (Cədvəl 16).

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* sözündən 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. (Cədvəl 17.)

Cədvəl 16. Et/ed qrupunda sözlərin yerləşdirilməsi

Birinci komponent	İkinci komponenti et/ed olan birləşmələrin tərcüməsi
Azad	освобождать,
Arzu	желать, мечтать,
Qəsd	покушаться, посягать,
Zənn	пологать, предполагать

Cədvəldəki sözlərin sintaktik əhatədən asılı olaraq müxtəlif mənalar kəsb etməsi xüsusiyyəti nəzərə alınmaqla təhlil və MT alqoritmlərinin məzmununa müəyyən əlavələr etmək, sintaktik elementlərdən istifadə olunmaqla düzgün təhlil və tərcümə nümunələri almaq olar. Fərz edək ki, cümlənin xəbəri “et(mək)” köməkçi feli ilə ifadə olunmuşdur. Azərbaycan dilində xəbər cümlənin sonunda gəldiyi üçün “et(mək)” köməkçi felini 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, суетливый” kimi başa düşüləcək. Cədvəldən göründüyü kimi, “ol(maq)” qrupunda bu söz tamamilə başqa məna kəsb edəcək. Bütün yuxarıda göstərilənləri alqoritmləşdirməklə prosesin dəqiq ifadəsini vermək olar. Birləşmələr AL-də sintaktik vahidlərin müəyyənləşdirilməsi alqoritmləri barədə mətnin formal təhlili sisteminin sintaktik blokunda geniş izahat veriləcəkdir (Şəkil 1.).

Cədvəl 17. Birləşmələr AL-də sözlərin ayrılıqda və müxtəlif qruplarda mənalari

Sözlər I kompo- nent	Əsaslar AL-də mənası	Birləşmələr AL-də mənası və ya tərcüməsi	
		et/ed qrupu	ol qrupu
taniş	знакомый, знакомец, счет, расчет	ознакомить, познакомить считать, подсчитать	познакомить ся считаться, приниматься
hesab	внутренняя	считать, подсчитать	в расчет вспутать, вступать,
daxil	часть, внутренности,	вводить, включать	входить беспокоится
narahat	беспокойный, суетливый	беспокоить, тревожить	тревожиться

VI FƏSİL. MƏTNİN FORMAL TƏHLİLİ SİSTEMİNİN MORFOLOJİ BLOKU

6.1. Mətnin formal təhlili sisteminin morfoloji blokun yeri

Morfoloji blok mətnin formal təhlili sisteminin ən mühüm və zəruri mərhələlərindəndir. Mətni təşkil edən ayrı-ayrı vahidlərin (söz-formaların) formal təhlili - formal təhlil sisteminin sonrakı mərhələləri üçün ilkin zəruri qrammatik səciyyələrin alınmasını təmin edir.* Digər tərəfdən formal morfoloji təhlil məlumat axtarışı sistemlərində informasiyanın toplanması, yığılması, axtarışı, saxlanması və ötürülməsi proseslərini avtomatlaşdırmaq imkanı yaradır (Белогов, Котов, 1968; Белогов, Новоселский, 1979).

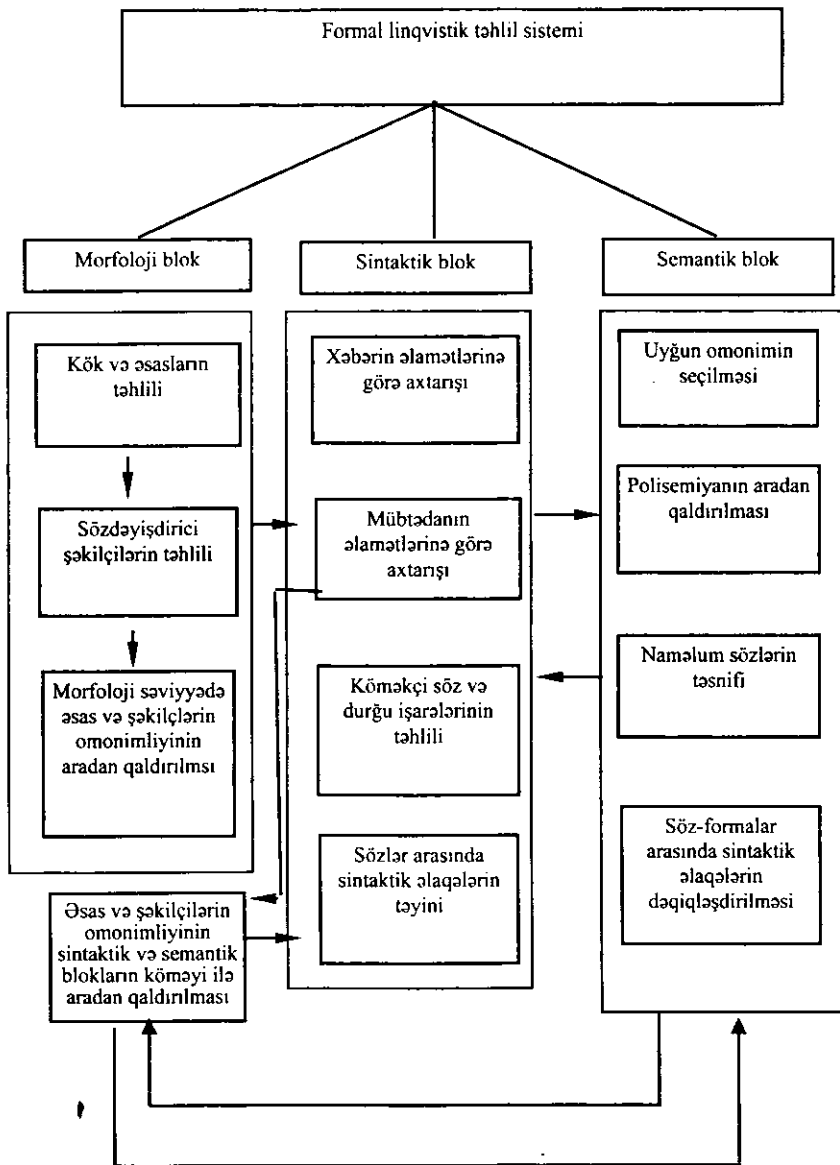
Mahiyyətinə görə formal morfoloji təhlil - formadan (qrafik) mənaya keçid deməkdir. Lakin təkcə morfoloji blokun köməyi ilə qrafik formadan mənaya keçid qeyri-mümkündür. Omonimlik (omograflıq), söz sırası və s. spesifik linqvistik xüsusiyyətlər sintaktik-semantik blokların köməyi ilə «tənzimlənilir». Sxem 4-də formal linqvistik təhlil sisteminə morfoloji blokun yeri göstərilmişdir. Sintaktik və semantik bloklar morfoloji blokdən sonra fəaliyyət göstərirlər. Sintaktik təhlili çətinlik törədən söz-formalar (omonim sözlər, çoxmənalı sözlər, AL-də təmsil olunmamış leksik vahidlər və s.) semantik blokda yoxlanılır.

* Türkologiyada tatar dilindən rus dilinə məşin tərcüməsi zamanı morfoloji təhlil məsələlərinə həsr edilmiş ilk işlərdən biri N.A.Xalitova və b. tərəfindən aparılmışdır (Халитова, Закирова, Гимадуудинова, 1962, 98-105). Tədqiqatçılar morfoloji təhlilin aşağıdakı komponentlərini göstərirlər: ümumi qaydalar (A), Məlumat siyahısı (B), Şəkilçilərin məlumat siyahısı (V), Qeyri-standart yazılış (Q), Omonim hallar (D). Söz-formalar üçün məlumatların seçilməsi və onların təhlili proqramları kiçik riyazi mətnlər (120s/f) üçün tərtib olunmuşdu. Əsasın lüğətdə axtarılması mətn vahidlərinin əsaslar lüğəti ilə hərfbəhərf tutuşdurulması yolu ilə aparılır. Hər bir əsas və şəkilçi üçün ona müvafiq qrammatik məlumat yazılır. Müəlliflər şəkilçilərin xüsusi cədvəllərdə axtarışı proqramlarını da tərtib etmişlər.

Bu bloklar bir-biri ilə qarşılıqlı əlaqədə fəaliyyət göstərirlər və bütövlükdə vahid bir sistem təşkil edirlər. Semantik blok öz növbəsində morfoloji blokla əlaqəlidir. Morfoloji təhlil blokunda çətin təhlil olunan söz-formalar semantik «sözgəcdən» keçirilir və sonra yenidən sintaktik bloka qaytarılır. Ayrı-ayrı konkret dillərin xüsusiyyətlərindən, formal təhlil alqoritmlərinin genişlənməsi və təkmilləşməsindən asılı olaraq morfoloji, sintaktik və semantik bloklar müəyyən dəyişikliklərə məruz qalaraq tamamlana bilər və onun strukturuna əlavələr oluna bilər (Бельская, 1969; Пиотровский, Бектаев, 1977; Пиотровская, Пиотровский, 1974).

Bütövlükdə, formal linqvistik təhlil sistemi qapalı deyil və proses zamanı onun quruluşunda zəruri düzəlişlər aparmaq mümkündür. Morfoloji blokun strukturu sonrakı mərhələlər üçün mühüm əhəmiyyət kəsb etmir. Əsas məqsəd ondan ibarətdir ki, formal morfoloji təhlil nəticəsində sintaktik və semantik səviyyələr üçün zəruri informasiya əldə olunmuş olsun.

Bütün təhlil sistemi dilin qrammatik quruluşuna formal cəhətdən yanaşmanı tələb edir. Bu da öz növbəsində mətnlərin avtomatik işlənməsi sistemləri üçün məşin qrammatikalarının yaradılması problemini ön plana çəkir.



Sxem 4. Formal linqvistik təhlilin ümumi sxemi

6.2. Maşın qrammatikası anlayışı

Məlum olduğu kimi, hər hansı bir mətnin tam təhlili, onun ayrı-ayrı cümlələrə ayrılması, söz-formaların təhlili, izahı formal təhlil sistemlərində ənənəvi dilçilikdə insan-dilçi tərəfindən aparılan əməliyyatlardan fərqlidir. Buna görə də, kompüterin əsaslandığı, istifadə etdiyi qrammatika və onun kateqoriyaları da ənənəvi qrammatikadan fərqli olmalıdır. Daha doğrusu, elə bir qrammatika yaradılmalıdır ki, onun köməyi ilə kompüter sırf formal əlamətlər və prinsiplər əsasında sistem üçün zəruri məlumatı əldə etmək imkanına malik olsun. Əgər insan-dilçi təhlil və ya sintezdə sırf formal əlamətlərdən başqa özünün dilçilik sahəsindəki biliyinə, lazım gəldikdə, ekstralinqvistik məlumatlara əsaslanırsa, kompüter bu imkanlardan tamamilə məhrumdur. Kompüter bu və ya digər məlumatı «anlamır», «dərk etmir», sadəcə olaraq tanıyır, çünki maşın «obrazı» deyil, kodu «görür». Ona görə də, maşında bu və ya digər məlumatı «anlama, dərk etmə» prosesi kodun tanınmasına aparıb çıxaran müəyyən əməliyyatların yerinə yetirilməsi ilə əldə edilir (Зубов, 1971). Deməli, elə bir qrammatika işlənilməlidir ki, maşın onu «başla düşə bilsin» və ondan lazımı səviyyədə məlumat əldə etməklə sistemin normal fəaliyyətini təmin etsin. Lakin burada bir mühüm cəhət də nəzərə almaq olduqca vacibdir: maşın qrammatikaları sırf nəzəriyyə, yeni istiqamət xatirinə yaradılmır. Hər şeydən əvvəl belə qrammatikalar praktik məqsədlərə xidmət edir və praktik əhəmiyyət kəsb edir (Пиотровский, Аникина, Аполлонская, Билан, Боркун, Лесохин, Сингарева, 1980; СТРААТ, 1980).

Digər tərəfdən yeni baxış bucağı altında qrammatikaya nəzər yetirilməsi, ənənəvi dilçilik üçün qərribə görünənə biləcək kriteriyalara əsaslanma bir çox maraqlı faktların aşkarlanmasına səbəb olur. Tədqiqatın bu istiqamətdə dərinləşdirilməsi əsas məqsəddən yayınma ilə nəticələnə bilər, praktik məqsəddən uzaq, sırf nəzəri formal maşın qrammatikalarının işlənilməsi kimi cəzbedici bir işə meyil yaranmış olardı. Maşın qrammatikalarında tədqiqatçı marağını müəyyən məqsəd ətra-

fında cəmləşdirməklə, konkret bir məqsədə xidmət edə biləcək qrammatik kateqoriyaların formal sistemə uyğunlaşdırılmış forma və metodları ön plana çəkilmişdir. Lakin, lazım gəldikdə, ənənəvi qrammatikanın bu və ya digər məsələsinə münasibət nəzəri dilçilik üçün heç də faydasız olmazdı.

Mövcud leksik səviyyədə və sözbəsöz tərcümə nümunələri almağa imkan verən digər avtomatik işlənmə sistemlərinin təcrübəsi göstərir ki, bu qəbildən olan proseslərdə maşın qrammatikaları mühüm və zəruri amildir. Sistemin sonrakı inkişafı və təkmilləşməsi maşın qrammatikaları olmadan real hesab oluna bilməz. Hətta xüsusi morfoloji və sintaktik quruluşa malik çin dilində və analitik-flektiv dillərdə belə (məsələn, ingilis dilində) maşın qrammatikalarının nəzərə alınmaması son nəticədə özünü doğrultmur (bax: Бектаев, Беляева, Королева, Марчук, Пиотровский, Садчикова, Цингарева, 1981, Действующие системы МП и автоматические словари, 1979, Бабанаров, 1981, Жеребин, 1963, Лингвистическое обеспечение в системе автоматического перевода третьего поколения, 1978, Шевенко, 1964, Арзикулов, Пиотровский, Попеску, Хажинская, 1978, Махмудов, 1982; 1991; 1994). Aqlütinativ quruluşlu dillərdə (xüsusilə, türk dillərində) maşın qrammatikası olmadan aparılan təhlil tamamilə perspektivsiz və səmərəsizdir. Nəzərə almaq lazımdır ki, türk dillərində bir fel və ya ad əsasında yüzlərlə və hətta nəzəri olaraq minlərlə söz-forma yaratmaq olar. Ona görə də, belə dillərdə təhlil prosesi hökmən mükəmməl maşın qrammatikalarına əsaslanmalıdır (Пиотровский, 1975, 272-190).

Təcrübələr göstərir ki, bir fel əsasından şəkildilər vasitəsi ilə tatar dilində 17947, türk dilində 11390, özbək dilində 13592, söz-forma düzəltmək olar (Исхакова, 1972, 17). Hətta beş formant ilə məhdudlaşdıqda belə Azərbaycan dili üçün bu göstərici 8000-dən yuxarı olur. Azərbaycan dilində fel əsaslı söz-formaların tərkibində şəkildilərin sayı ondan yuxarı ola bilər (Пинес, 1974, 104). Məsələn, Azərbaycan dilində *məktəblərimizdəkilərdəndirbirmi?* söz-forması 10 morfemdən (onlardan 9-u şəkildçi morfem) ibarətdir.

Qazax dilində isimlərdən forma əmələ gətirən şəkilçilərin sayı 500-ə çatır, feldən isə 1000-ə qədər müxtəlif forma əmələ gətirmək olar (Бектаев, 1978, 39). Burada daha mühüm bir amil diqqəti cəlb edir: əgər elmi-texniki mətnlərdə 13 mindən az müxtəlif söz-formalardan istifadə olunursa, flektiv dillərdə bu söz-formaların sayı 20-25 minə çatır, aqlütinativ dillərdə (məsələn, qazax dilində) isə 60 minə yaxın müxtəlif söz-formadan istifadə olunur (Кенесбаев, Пиотровский, Бектаев, 1970, 13). М.Аuezovun «Абай jолу» romanının tezlik lüğəti 466 min sözişlətmədən ibarətdir ki, bunların da 61 mini müxtəlif söz-formalardır. Başqa türk dillərində də vəziyyət analoji qaydadır (Бектаев, Мырзабеков, 1973).

Hind-Avropa dilləri leksikası və morfologiyasının modelləşdirilməsi və formal təhlili təcrübəsi göstərir ki, mətnlərin formal təhlili sistemləri və mühəndis dilçiliyi modellərinin tərtibi ilə əlaqədar maşın qrammatikalarının yaradılması olduqca zəruridir (bax: Статистика речи, 1968, Статистика речи и автоматический анализ текста, 1971, 1973, 1974, 1980). Belə qrammatikalar qeyri-ənənəvi «maşın fleksiyalarına» və «maşın əsasları (kvazi əsas) lüğətlərinin» yaradılmasına əsaslanır (Зорев, 1973, 245-259). Maşın qrammatikası dedikdə elə bir qrammatika başa düşülür ki, onun köməyi ilə kompüterlərdə müxtəlif məqsədlərlə (formal təhlil, mətnlərin avtomatik işlənməsi, maşın tərcüməsi, avtomatlaşdırılmış məlumat-axtarış sistemləri, avtomatik referatlaşdırma və s.) sırf formal əlamətlər əsasında tam qrammatik təhlil aparılması və linqvistik təhlilin sonrakı mərhələləri üçün zəruri qrammatik məlumatların alınması mümkün olsun. Maşın qrammatikasında türk dillərində (eləcə də başqa dillərdə) sözdüzəltmə və sözdəyişdirmənin elə cəhətləri müəyyənləşdirilərək nəzərə alınmalıdır ki, ənənəvi qrammatikalarda adətən onlar açıq şəkildə göstərilmişdir. Məsələn orasıdır ki, təbii dilin bir çox özünəməxsusluqları formal təhlil sistemlərində aşkar olunur (Исхакова, 1956; Конопов, 1976; Совпель, 1980; Статистико-информационное изучение тюркских языков, 1969; Махмудов 1991).

Türk dillərinin maşın qrammatikalarının yaradılması həmçinin söz-formaların yaranmasının məntiqi xüsusiyyətlərinin tədqiqi, eləcə də onların struktur-funksional modellərinin qurulması, linqvistik avtomatda yoxlanması baxımından da zəruridir. Maşın qrammatikaları, sözdüzəltmə və sözdəyişdirmənin struktur-funksional modelləri olmadan türk dillərində formal təhlil sistemləri mövcud ola bilməz. Bu həm də türk dillərinin aqlu-tinativ təbiətindən doğur (Пиотровская, 1977; Совпель, 1980).

Qeyd etmək lazımdır ki, mətnlərin kompüterlərdə işlənməsi prosesində maşın əsasları və şəkilçilərindən istifadə etmək daha məqsədəuyğundur (Зорев, 1972; 1973; Пиотровская, 1973).

Maşın qrammatikalarında mühüm məsələlərdən biri nitq hissələrinin formal əlamətlərə görə təsnifidir.

6.3. Türk dillərində nitq hissələrinin təsnifi

Türk dillərində nitq hissələrinin təsnifi məsələsi mübahisə doğuran problemlərdən biri hesab olunur. Bu problemin formal təhlil üçün böyük əhəmiyyət kəsb etdiyini nəzərə alıb məsələni bir daha nəzərdən keçirək. Türk dillərində nitq hissələrinin bölgüsü (təsnifi) məsələsində türkoloqlar arasında yekdil fikir mövcuddur (АДГ, 1960; Баскаков, 1957, 1963, 1968; Закиев, 1973; Исхаков, 1956; Самоилович, 1926; Hüseynzadə, 1973; Шербак, 1957, 1977). Türkologiyada mövcud fikri qəbul etsək, morfemləri aşağıdakı üç qrupa ayırmaq olar:

- əşyanın adını və ya statik əlamətini bildirən kök-morfemlər;
- hərəkət - dinamik əlamət - fel bildirən kök-morfemlər;
- söz birləşmələri və cümlələrdə qrammatik münasibətləri bildirən köməkçi kök morfemlər.

Birinci qrupa adlar - isim, sifət, say, zərf, nida, mimemlər, əvəzlilik, ikinci qrupa - fellər, üçüncü qrupa isə - ədat, qoşma, bağlayıcı, modal sözlər daxildir. Birinci iki qrup müstəqil və ya əsas, üçüncü qrup isə qeyri-müstəqil və ya köməkçi nitq hissələrinə aid olunur. Beləliklə, bütün kök morfemləri iki böyük sinfə ayırmaq olar:

- müstəqil və ya əsas,
- qeyri müstəqil və ya köməkçi.

Bundan əlavə, türk dillərində morfoloji səviyyədə, kontekstdənkənar heç bir nitq hissəsinə aid oluna bilməyən sözlər (omonim tipli) qrupunu da ayırmaq olar.

Əsas nitq hissələrinə aid edilən söz qrupları bir-birindən fərqli, xüsusi sözmələgətirmə və sözdəyişdirmə şəkilçi sistemlərinə malik olurlar (Cəfərov, 1968; Qiyasbəyli, 1987).

Şəkilçilərin özünü də bəzən formal olaraq fərqləndirmək mümkün deyildir (omograf şəkilçilər). Omonim sözlər həm birinci, həm də ikinci tipə birləşən şəkilçilər qəbul edə bilərlər. Köməkçi nitq hissələri heç bir şəkilçi elementi qəbul etmirlər.

Sözün məna məzmununa, bu və ya digər sözmələgətirmə və ya sözdəyişdirmə xüsusiyyətlərinə, eləcə də sintaktik funksiyalarına görə professor M.Hüseynzadə Azərbaycan dilində nitq hissələrini dörd qrupa ayırır (Hüseynzadə, 1973, 18):

- müstəqil (əsas),
- səs-obraz-təqlidi (kalka),
- nida,
- köməkçi.

Müstəqil (əsas) nitq hissələri cümlə üzvü ola bilir və ya cümlə əmələ gətirirlər. Onlar keyfiyyət, əlamət, miqdar, hərəkət-vəziyyət, hərəkət - vəziyyətin əlamətini, eləcə də əlamətin əlamətini, əşyanın keyfiyyətini bildirirlər. Əsas nitq hissələrinə isim, sifət, say, fel və əvəzlik daxildir. Qeyd etmək lazımdır ki, digər əsas nitq hissələrindən fərqli olaraq əvəzlik əşyanın miqdarını və keyfiyyətini deyil, bu miqdar və keyfiyyətin adına işarə funksiyasını yerinə yetirir.

Səs-obraz-təqlidi sözlər (kalkalar) əsas nitq hissələri kimi sözmələgətirmə prosesində fəal iştirak edirlər. Lakin, əsas nitq hissələrindən fərqli olaraq, onların konkret leksik mənası olmur. Məhz buna görə də, professor M.Hüseynzadənin təsnifində səs-obraz-təqlidi sözlər (yamsılamalar) əsas və köməkçi nitq hissələri arasında aralıq mövqe tutur.

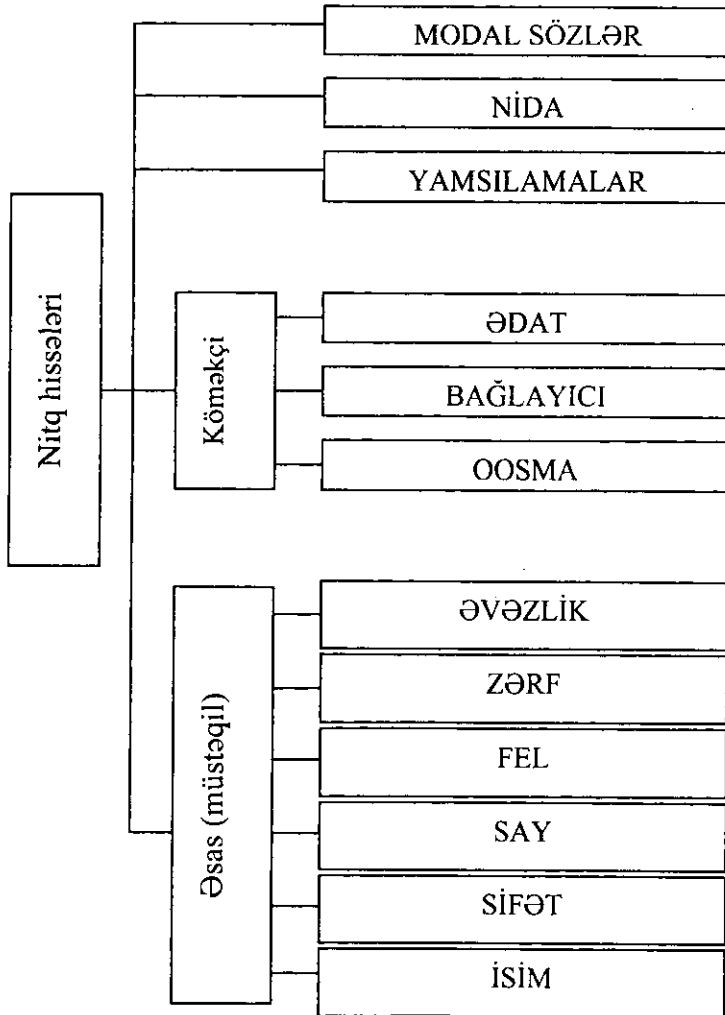
Nida-subyektin müəyyən hiss-əyəcanı ifadə etmə vasitəsidir (həyəcan, hiss, kədər, sevinc). Nidalar konkret anlayışı bildirmir və cümlə üzvü olmurlar. Buna görə də, nida həm əsas, həm də köməkçi nitq hissələrindən fərqləndirilir.

Müasir Azərbaycan dilində bir çox sözlər həm nida, həm də müstəqil nitq hissələri kimi işləyə bilər. Bu sözlər adlara aid olan sözdəyişdirici şəkilçiləri qəbul edə bilərlər (bax: Hüseynzadə, 1973, 287). Məsələn, *aman, a, haray, afərin* və s.

Köməkçi nitq hissələri cümlə üzvü ola bilmir və cümlə əmələ gətirmir. Onların leksik mənası olmur, cümlədə müxtəlif sintaktik funksiyalar daşıyırlar. Köməkçi nitq hissələrinə qoşma, bağlayıcı və ədat daxildir (Sxem 5.)

Qeyd etmək lazımdır ki, səs-obraz-təqlidi sözlər (yamsılamalar) sözmələgətirmə və sözdəyişdirmə xüsusiyyətlərinə görə özlərini isim kimi aparır. Təsnifatda nitq hissələri söz və forma-mələgətirmə funksiyalarına görə qruplaşdırıldığı üçün formal təhlilin morfoloji blokunda belə sözlərin isimlərin tərkibində verilməsi daha məqsədəuyğundur. Əlbəttə, türk dillərində, eləcə də Azərbaycan dilində nitq hissələrinin yuxarıdakı şəkildə təsnifi, qruplaşdırılması ənənəvi dilçilikdə heç bir mübahisə doğurmur. Təbiidir ki, formal təhlil sistemlərində sözlərin qrammatik sinif, kateqoriya və nitq hissələrinə ayrılması ciddi surətdə formallaşdırılmış prinsiplərə əsaslanır və buna görə də, bu məqsədlə yaradılmış qrammatika ənənəvi qrammatika ilə eyni olmaya bilər. Formal təhlil sistemləri üçün nəzərdə tutulmuş, yenidən yaradılmış və ya qismən uyğunlaşdırılmış qrammatika konkret elmi məqsədə xidmət edir. Tədqiqatın səciyyəsinə əsasən olaraq bu qrammatika genişləndirilə və ya mədudlaşdırıla bilər. Ayrı-ayrı kateqoriyalar məşin qrammatikalarında nəzərə alınmaya da bilər. Lakin bu problemlə əlaqədar bir məsələ ciddi mübahisə doğurur. Nitq hissələrinin təsnifi ənənəvi dilçiliyin özünü təmin edirmi? Nitq hissələrinin bölgüsündəki məntiqi uyğunsuzluq, bölgü prinsipinin pozulması türk dillərində bu problemin həllinə sistemli şəkildə bir daha baxılmasını tələb edir. Fikrimizi aydınlaşdırmaq məqsədi ilə bir daha türk dillərində nitq hissələrinin

bölgüsü məsələsinə qayıdaq. Hər hansı bir elmi sahədə toplanmış bilik və ya nəzəri fərziyyələrin sistemli şəkildə öyrənilməsində təsnifatın mühüm rolu var. Təsnifatın məqsəd və vəzifəsindən asılı olaraq cisim və hadisələrin müəyyən zonalara, siniflərə, qruplara, yarımqruplara və s. bölünməsi müxtəlif prinsiplərə əsasən aparıla bilər.



Sxem 5. Azərbaycan dilində nitq hissələrinin təsnifi

Bütün hallarda təsnifat elmi həqiqətlərin aşkar edilməsi və sistemli şəkildə öyrənilməsinə xidmət etməlidir. Təsnifatın düzgün aparılmaması elmi dolaşılığa, fikir dağınıqlığına və sistem-sizliyə aparıb çıxarır.

Məntiq elminə görə, təsnifat bir ümumi əlamət əsasında aparılmalıdır və təsnifatın sonuna kimi seçilmiş əlamət dəyişməz qalmalıdır. Əgər əlamət dəyişilərsə, bölgü prinsipi pozular və ziddiyətli təsnifat alınar. Nəticədə bölünən məfhumla bölgü üzvlərinin həcmi arasında bərabərsizlik yaranmış olar. Məntiqin bölgü prinsiplərinə ardıcıl olaraq əməl olunmadığından ayrı-ayrı türk dillərində nitq hissələrinin təsnifində uyğunsuzluqlar müşahidə olunur.

Türk dillərində nitq hissələri müstəqil mənalarına, morfoloji əlamətlərinə və sintaktik roluna görə (bax: Кононов, 1956; Зейналов, 1974; Махмудов, Гулиев, Кербалаева, 1982; Гусейнзаде, 1973, Пиотровская, 1964), morfonoloji və fonetik əlamətlər toplusu ilə fərqlənən struktur-semantik söz sinifləri kimi (Грамматика киргизского литературного языка, 1987), bəzən isə sözəmələgətirmə və formaəmələgətirmə tiplərinə, söz birləşməsi və cümlə tərkibində sintaktik funksiyalarına görə birbirindən fərqlənən leksik-qrammatik söz qrupları kimi təsnif edilir (Махмудов, Гулиев, Кербалаева, 1982).

Göründüyü kimi, bölgünün əsasında bir ümumi əlamət deyil, müxtəlif əlamətlər toplusu durur. Belə olduqda, bölgü üzvlərinin həcmi bölünən məfhumun həcminə bərabər ola bilməz. Bölgü üzvləri də biri digərini inkar etmədiyi üçün ardıcılıq pozulur və qarışdırılmış söz qrupları yaranmış olur (bax: Ожерочик, 1965; Уемов, 1958).

Bölgünün əsasında əlamətlər çoxluğunun qoyulması onunla nəticələnmişdir ki, bu və ya digər sözün hər hansı nitq hissəsinə mənsubluğu mübahisə doğurur. Məsələn, *dəmir* sözü ayrılıqda isimdir. Cümlədə mübtədə rolunda çıxış edir. Təyin kimi işləndikdə isə sifətə çevrilir. Məsələn: *Dəmir qapı. Yaxşı* sözü əşyaya və hərəkətə aid olub-olmamasından asılı olaraq müxtəlif nitq hissələri kimi izah olunur.

Məsələn, *cəsurluq, bərklik, mərdlik, sərtlik, şəffaflıq* və s. kimi sözlər keyfiyyət bildirdikləri halda isim nitq hissəsi sayılır.

Nitq hissələrinin, ayrı-ayrı qrammatik kateqoriyaların bir-birinə qarışdırılması, ayrı-ayrı türk dillərində söz qruplarının müxtəlif nitq hissələri kimi izahına dair çoxlu misallar göstərmək olar.

Köməkçi nitq hissələri məfhumu özü də olduqca yaygın, qeyri-müəyyəndir və türk dillərində bu məsələ ilə əlaqədar fikir yekdilliyi yoxdur. Sözlər və cümlələr arasında müxtəlif əlaqə və münasibət yaratma xüsusiyyətləri, müxtəlif emosional-ekspressiv və hissi münasibətlərin ifadəsi köməkçi nitq hissələrinin səciyyəvi əlamətləri hesab olunur. Məntiq elmi baxımından bu əlamətlər dəqiq müəyyənləşdirilməmişdir və bölgünün əsasında qoyula bilməz.

Yuxarıda sadalanan bütün mümkün əlamətləri bölgünün əsasına qoymaqla nitq hissələrinin vahid təsnifatını vermək məntiqi baxımdan qeyri-məqbuldur. Tədqiqatın məqsədindən asılı olaraq nitq hissələrinə baxış bucağı dəyişilməli, başqa sözlə tədqiqat istiqamətinə müvafiq model yaradılmalıdır. Məsələn: sintaksislə məşğul olan mütəxəssis nitq hissələrinin sintaktik əlamətlərinə əsaslanmalı və həmin model üzrə lazımi elmi nəticələr əldə etməyə cəhd göstərməlidir. Bu o deməkdir ki, lazım gəldikdə təsnifatın əsasına sintaktik, semantik, morfoloji və s. əlamətlərdən biri qoyula bilər. Bu əlamətlərin hamısının eyni zamanda bölgünün əsasına qoyulması isə məntiqə ziddir və qeyri-məqbul təsnifata aparıb çıxarır.

Morfoloji (formal), sintaktik və semantik əlamətlər əsasında aparılmış təsnifatlar bir-birindən köklü surətdə fərqli olacaqdır.

Hər bir bölgünün nisbi səciyyə daşdığını nəzərdə tutaraq qeyd etmək lazımdır ki, təklif olunan təsnifat hər dəfə bir mühüm, aparıcı əlamət əsasında aparılır. Digər əlamətlər isə göstərilən təsnifat zamanı nəzərə alınmır. Məsələyə bu cür yanaşıldıqda məntiqi cəhətdən daha dəqiq və düzgün təsnifat almağa imkan yaranır.

6.4. Nitq hissələrinin formal əlamətlərə görə təsnifi

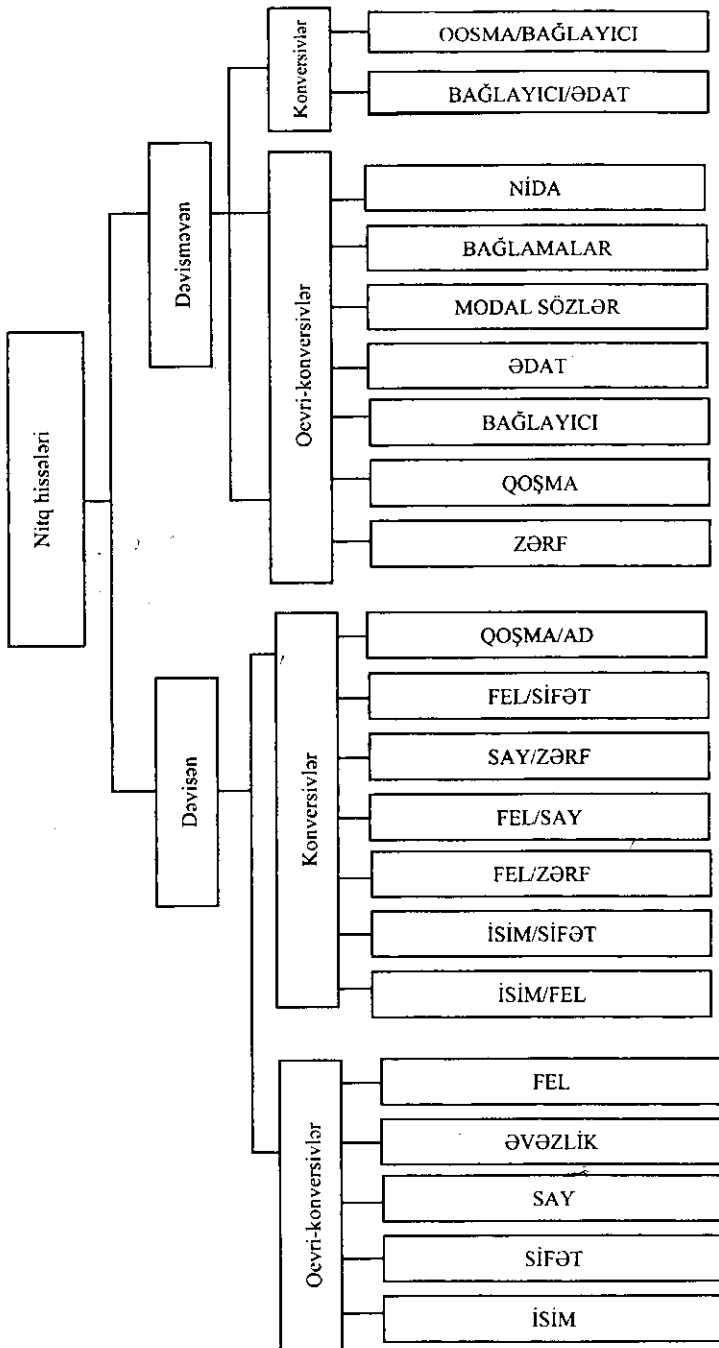
Türk dillərində mətnlərin avtomatik işlənməsi ilə əlaqədar nitq hissələrinin formal əlamətlərə görə təsnifi sahəsində artıq

müəyyən işlər görülmüşdür (Babanarov, 1981; Mahmudov, 1991). Nitq hissələrinin sırf formal əlamətlərə görə təsnifi o demək deyildir ki, konkret dil və dil qruplarının təsvirindən və ümumi qrammatika nəzəriyyələrinin nailiyyətlərindən tam imtina edilməlidir. Formal təsnifatın izahında ənənəvi dilçilikdə formallaşmış nitq hissələri anlayışından istifadə olunur. Bu ondan irəli gəlir ki, mətnin formal təhlil sistemi mücərrəd, qeyri-müəyyən məqsədlərə xidmət edən xəyali sistem deyil, Azərbaycan mətnlərinin praktik məqsədlərlə işlənməsinə yardım edir və onun gəldiyi nəticələrdən hamı istifadə edə bilmək imkanına malik olmalıdır.

Sözdəyişdirici şəkilçilərə münasibətinə görə mətnin formal təhlili sisteminin morfoloji blokunda nitq hissələri iki mühüm qrupa ayrılır: (Sxem 6)

- dəyişən (bura sözdəyişdirici şəkilçilər qəbul edə bilən söz-formaları ehtiva edən nitq hissələri daxildir);
- dəyişməyən (bura sözdəyişdirici şəkilçilər qəbul etməyən söz-formaları ehtiva edən nitq hissələri daxildir).

Dəyişən nitq hissələrinə isim, səs-obraz-təqlidi sözlər, sifət, say, əvəzlik, fellər aid edilir. Bu qrup nitq hissələri aşağıdakı konversivləri də tərkibinə alır («konversiv» - sözü altında fonetik və morfoloji quruluşuna görə eyni olub, bir nitq hissəsindən digərinə keçən söz qrupları başa düşülür): isim/fel (qaz, at), isim/sifət (qoca, göy), fel/zərf (düz), fel/sifət (burma), fel/say (*min, qırx*), say/zərf (*çox, xeyli*), qoşma/ad (*qabaq*). Dəyişməyən nitq hissələri zərf, qoşma, bağlayıcı, ədat, modal sözlər, bağlama və nidadan ibarətdir. Qeyd etmək lazımdır ki, Azərbaycan dilində bəzən zərflər hallanaraq ismin hal şəkilçilərini qəbul edə bilirlər (məsələn; *içəri, içəriyə, içəridə, içəridən, geri, geriyyə, geridə, geridən, yaxın, yaxına, yaxında, yaxından* və s.). Lakin bütövlükdə sözdəyişdirici şəkilçilərlə işlənmə zərflər üçün səciyyəvi xüsusiyyət deyildir.



Sxem 6. Maşın morfologiyasında nıttq hissələrinin təsnifi

Dəyişməyən nitq hissələrinə aşağıdakı konversiz siniflər də aid olur: bağlayıcı/ədat (*ancaq, ki*), qoşma/bağlayıcı (*ibə, üçün, görə, ötrü*).

Hazırda türk dilinin materialları əsasında formal əlamətlərə görə nitq hissələrinin təsnifi mövcuddur (Барабанов, 1981, 10) Həmin təsnifatda bir sıra mübahisə doğuran cəhətlər vardır:

- təsnifatda saylar (miqdar, sıra və kəsr sayları), eləcə də sifət dəyişən deyil, dəyişməyən nitq hissələrinə aid edilir. Sayların və sifətin dəyişməyən nitq hissələrindən hesab olunması düzgün sayıla oluna bilməz, çünki türk dillərində istər say, istərsə də sifət sözdəyişdirici şəkilçilər qəbul edib dəyişə bilər.

Göstərilən təsnifatdan fərqli olaraq tədqiqatda nida, qoşma və bağlayıcı ayrıca nitq hissəsi kimi təsbit olunur. Səs-obraz təq-lidi sözlər isimlərin tərkibində verilir. Fel/sifət, fel/say, say/zərf, bağlayıcı/ədat, qoşma/bağlayıcı kimi yeni konversivlər təsnifata daxil edilir. Formal əlamətlərinə görə Azərbaycan dilində nitq hissələrinə isim/say, isim/zərf, sifət/əvəzlik, qoşma/ədat konversivlərin daxil olması lüzumsuz hesab olunmuşdur.

Təqdim etdiyimiz təsnifatda dəyişən nitq hissələri qrupuna daxil edilən nitq hissələri və kateqoriyaları (məlum olduğu kimi, bu nitq hissələri və kateqoriyalara sözdəyişdirici şəkilçilər qəbul edən söz-formaları daxildir) aşağıdakı alt qruplara bölmək olar:

- ad nitq hissələrinin sözdəyişdirici kateqoriya və formaları (bura ismin hal, mənsubiyyət və xəbərlilik kateqoriyaları daxildir);
- fel nitq hissəsinin sözdəyişdirici kateqoriya və formaları (bura felin növləri, şəkilləri, zamanları, eləcə də inkarlıq, məsdər, feli sifət və feli bağlama daxildir);
- ümumi kateqoriyalara (həm adlara, həm də fellərə aid olan kateqoriyalar) - təklik (cəmlik), şəxs və sual kimi kateqoriyalar daxildir.

Formal təhlil sistemində sözdəyişdirici şəkilçilərlə işlənən söz-formaları əhatə edən nitq hissələri (dəyişən nitq hissələri) tədqiqatın mərkəzi obyektinə hesab olunur. Dəyişməyən nitq hissələri və onlara aid edilən kateqoriyalar sistemin leksik blokunda

ehtiva olunduğu üçün, o cümlədən bütün sözdüzəldici kateqoriya və formalar barədə kifayət qədər informasiya toplandığı üçün mətnlərin işlənməsi üçün zəruri olan və AL-də təsbit olunması qeyri-mümkün olan digər qrammatik məlumatlar formal təhlil sistemində həyata keçirilən müxtəlif linqvistik əməliyyatlar nəticəsində əldə edilir.

6.5. Morfoloji blokun funksiyası

Morfoloji blokunun əsas funksiyası mətnlərdə müşahidə olunan söz-formaların tanınması, təhlili və onun müxtəlif məqsədlər üçün zəruri olan müvafiq mənalarla (qrammatik-leksik və s.) əvəz olunmasıdır. Bu məqsəd həyata keçirilərkən belə bir müddəə əsas götürülür ki, mətnin formal cəhətdən fərqləndirilə bilən ən kiçik vahidi sayılan söz-forma ilə onun mənaları arasında müəyyən qanunauyğunluqlar əsasında müəyyənləşdirilə bilən müvafiqlik mövcuddur. Bu müvafiqliyin araşdırılması aşağıdakı mərhələlərlə həyata keçirilir:

- massivdən mətnlərin bir-birindən ayrılması və yaxud tədqiqata cəlb olunacaq mətnin kompüterin operativ yaddaşına verilməsi;
- həmin mətnin cümlələrə ayrılması və yaxud mətnin cümlə-cümlə hissələrə bölünməsi;
- ayrılmış cümlələrin formal cəhətdən fərqləndirilə bilən ən kiçik vahidlərə - söz-formalara parçalanması;
- söz-formaların təhlil alqoritmlərinin köməyi ilə əsas və şəkilçilərə ayrılması;
- əsasın bütün mənalarının müəyyənləşdirilməsi və yazılması, ən mühüm əlamətlərinin təyini;
- şəkilçilər zəncirində şəkilçilərin bir-birindən ayrılması, şəkilçinin qrammatik funksiyalarının təyini, bu əlamətlərin əsasa və ya onun digər dilə tərcümə olunmuş ekvivalentinə əlavə olunması.

Göstərilən vəzifələrin yerinə yetirilməsi morfoloji blokun əsas funksiyası sayılır və bu funksiyanın nə səviyyədə, hansı

keyfiyyətlə reallaşması mətnin formal təhlili sisteminin özünün səviyyə və keyfiyyətini müəyyənləşdirir.

6.6. Formal təhlil anlayışı

Türk dillərində, eləcə də Azərbaycan dilində formal təhlil prosesini həyata keçirən linqvistik modellər qurarkən iki əsas müddəə əsas götürülür. Hər şeydən əvvəl qəbul edilir ki, dilin və nitqin hər bir leksik və ya qrammatik vahidinin məna (əks olan) və forma (əks olunan) kimi mürəkkəb işarə sistemində malik bir quruluşu var. Qeyd etmək lazımdır ki, təkcə leksik, qrammatik səviyyədə deyil, fonetik səviyyədə də dilin mənalı vahidləri mövcuddur. Fonemlərin semantikasi problemi ilə bağlı tədqiqatlar maraqlı nəticələr əldə edilməsi ilə nəticələnmişdir. (Жыпаров, 1974). Lakin mətnlərin avtomatik işlənmə sistemləri ilə əlaqədar fonetik semantika məsələlərinin araşdırılmasına ehtiyac hiss olunur. İkinci mühüm cəhət odur ki, istənilən linqvistik model nəzəri olmaqdan daha çox təcrübədir. Yəni hətta yüksək səviyyədə qurulduğu ehtimal olunan modelin özü belə mütləq praktik sınaqdan keçirilməlidir. Yalnız təcrübədə özünü doğrultmuş təhlil modeli məqbul sayıla bilər. Linqvistik modeldə məntiqi ziddiyyət olmamalı, o, linqvistik cəhətdən anlaşılqı və kifayət qədər aydın olmalıdır.

Formal təhlil sisteminin müxtəlif səviyyələri üçün nəzərdə tutulmuş linqvistik modellərin yoxlanmasının ən yaxşı vasitəsi kifayət qədər dil materialı (mətnlər) əsasında onların kompüterdə sınaqdan keçirilməsidir.

Formal təhlil sistemində çoxluqlar nəzəriyyəsinin anlayışlarından da geniş istifadə olunur. Formal linqvistik təhlil sistemlərində çoxluq termini altında linqvistik obyektlərin (hərflər, şəkilçi, söz-forma, cümlə və s.) məcmusu başa düşülür. Hər hansı bir elementin A çoxluğuna daxil olması aşağıdakı kimi göstərilir:

$$X \in A$$

Əgər C çoxluğunun hər bir X_i elementi A çoxluğuna aiddirsə, onda C çoxluğunu A çoxluğunun altçoxluğu hesab edirlər.

Analoji olaraq, A' çoxluğunun hər bir X_i' elementi C' çoxluğuna aiddirsə, onda A' çoxluğu C' çoxluğunun altçoxluğu adlanır.

X_i - linqvistik simvollarının ixtiyari sonlu ardıcılığını zəncir adlandıracağıq. Zəncirin uzunluğu ona daxil olan a_i elementlərinin uzunluğu ilə ölçülür.

Formal təhlil sisteminin mərkəzi və əsas anlayışı söz-formadır. 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.

Cümlə ixtiyari sonlu hərflər ardıcılığı olan zəncirlərdən (zəncirdən) ibarət olub, iki nöqtə, nöqtə ilə nöqtəli vergül, nöqtə ilə sual və ya nida işarələri arasında yerləşir.

Araşdırılan formal təhlil sistemində mətn dedikdə kontekstdən kənarında götürülmüş (kontekstə daxil olmayan) söz-formalar çoxluğu başa düşülür.

İndiyə qədər araşdırdığımız anlayışlar mətnin formal olaraq müəyyənləşdirilə bilən, formal sərhədləri olan elementləridir. Bu vahidlərin formal təyini üçün əlavə məlumatların olması lazım gəlmir. Əsas və şəkilçi kimi anlayışların təyini üçün isə dilin öz sistemi ilə bağlı olan zəruri linqvistik məlumatın olması vacib şərtidir. Məndən söz-forma ayrıldıqdan sonra formal təhlil sisteminin «fəaliyyəti» başlanır. Hər şeydən əvvəl, söz-formada olan əsasla şəkilçilər zənciri, eləcə də ayrı-ayrı şəkilçilər arasında «sərhəd» zonaları müəyyənləşdirilməlidir, yəni elə bir formal təhlil metodu (alqoritmi) hazırlanmalıdır ki, sırf formal prinsiplər əsasında söz-formadan sözün əsasını, şəkilçilər zəncirini və ayrı-ayrı şəkilçiləri ayırmaq mümkün olsun.

Təqdim olunan formal təhlil sisteminin morfoloji blokunun əsas funksiyası bu göstərilən prinsiplərin reallaşdırılmasıdır.

Formal linqvistik təhlil sistemində mərkəzi anlayış olan söz-formaya əsas və sözdəyişdirici şəkilçilərin konkatenasiyası kimi baxmaq olar. Əsas termini altında sözün kökü və ya

kök+sözdüzəldici şəkilçi başa düşülür. Bunların hər ikisi müstəqil lüğət vahidi kimi əsaslar avtomatik lüğətinə daxil edilir. Mürəkkəb əsaslar (bir və ya bir neçə əsasın birləşməsindən ibarət əsaslar) da müstəqil lüğət vahidi kimi əsaslar avtomatik lüğətinə təmsil olunurlar (bax: Грамматика азербайджанского языка, 1971, 38; Махмудов, 1991, 52-60).

Şəkilçilər (sözdəyişdirici şəkilçilər) avtomatik lüğətə daxil edilmirlər. Mətnlərin formal təhlili sistemlərində əsas məsələlərdən biri də sözdəyişdirici şəkilçilərin təhlili və onların qrammatik səciyyələrinin müəyyənləşdirilməsidir. Bu proses belə bir linqvistik fərziyyəyə əsaslanır ki, mətni təşkil edən söz-formalar çoxluğu ilə mətnin ifadə etdiyi mənalar çoxluğu arasında müəyyən əlaqə, müvafıqlıq mövcuddur. Formal təhlil sisteminin əsas vəzifəsi bu əlaqələri, müvafıqlıyı əsas götürməklə mətndən (söz-formalar çoxluğundan) mənaya keçidi təmin etməkdir. Tədqiqatda məhz bu linqvistik konsepsiyadan doğan vəzifələrin həyata keçirilməsi yolları araşdırılmışdır.

Formal baxımdan linqvistik təhlil (o cümlədən morfoloji təhlil) prosesi mətndən (nitqin qrafik formasından) mənaya (leksik, qrammatik, semantik) mətnin hər bir vahidinə (hərflər zəncirinə) müvafiq mənənin tapılmasına imkan verən qaydalara keçid deməkdir. Mətndən mənaya (formadan mənaya) keçidin modelini aşağıdakı kimi ifadə etmək olar:

$$A \xrightarrow{F} A'$$

Burada A- qrafik forma, söz-formalar çoxluğu, mətndir.

$$A = \{a_i\}, i=1, \dots, n.$$

$$A' = \{a'_i\}, i=1, \dots, n.$$

F-A -da müəyyən olunan və $a \in A$ -nın hər bir elementi üçün yalnız bir $a' \in A'$ verən qaydalardır.

Verilmiş modelə A çoxluğunun A' çoxluğunda əksi kimi baxmaq olar. A və A' çoxluqları arasında elə bir F müvafıqlıyı vardır ki, həmin müvafıqlıya görə hər hansı bir $a \in A$ üçün $a' \in A'$ doğrudur. Təhlil sistemində A çoxluğu siyahı şəklində (morfem siyahıları, sözlüklər və s.) verilir. Leksik bloku təşkil

edən bütün linqvistik materiallar (AL, şəkilçilər siyahısı) və leksik-qrammatik informasiya A çoxluğunun elementləri hesab oluna bilər.

Morfem siyahılarının vahidləri (əsaslar, söz-formalar, sözdəyişdirici şəkilçilər) onları təşkil edən hərflərin sayına görə sıralanır. Söz-formaların, əsasların və şəkilçilərin uzunluğu əsas götürülməklə onlar müəyyən qruplarda yerləşdirilir. Birinci qrupda k sayda, ikinci $k-1$ sayda hərflərdən ibarət morfemlər yerləşir. Bu qaydada yerləşdirilmə təhlil alqoritmlərinin tələbinə uyğun olaraq qurulmuşdur. Leksik blokun təsviri verilərkən əsas, söz-forma və birləşmələrin AL-də yerləşdirilməsi prinsipləri ilə tanış olmuşduq. İndi morfoloji blokun mühüm elementi sayılan sözdəyişdirici şəkilçilərin yerləşdirilməsi qaydasını nəzərdən keçirək.

6.7. Sözdəyişdirici şəkilçilərinin morfoloji blokda yerləşdirilməsi

Sözdüzəldici şəkilçilərin digər konstruksiyalarla birlikdə leksik blokda ehtiva olunduğunu nəzərə alaraq morfoloji blokda sözdəyişdirici şəkilçilərin yerləşdirilməsi prinsiplərini nəzərdən keçirək. Yuxarıda göstərmişdik ki, leksik blokda prinsiplərə uyğun olaraq şəkilçilər də onları təşkil hərflərin sayına görə düzülür. Qeyd etmək lazımdır ki, şəkilçi siyahısı artan və azalan istiqamətdə ola bilər. Bunun mahiyyət etibarı ilə təhlil alqoritmlərinin işinə heç bir təsiri yoxdur. Yəni şəkilçilər $K-\mathcal{K}+1$ və ya $K-\mathcal{K}-1$ istiqamətində düzülə bilər. Təhlil prosesində istənilən qrupa müraciət etmək imkanı vardır. Şəkilçilərin qarşısında onların bütün mümkün qrammatik əlamətləri və kodları göstərilir.

Hazırda formal təhlil sistemləri və MT ilə bağlı ayrı-ayrı nitq hissələrinin, qrammatik kateqoriyaların kodlaşdırılmasında vahid prinsipin olması çox vacibdir. Bu istənilən tədqiqatdan səmərəli istifadə üçün imkan yaradır. Kodlaşdırma ilə bağlı materialların ayrılıqda veriləcəyini nəzərə alaraq şəkilçi qruplarında kodların izahının verilməməsi məqsədəuyğun hesab olunur. Şəkilçilər üçün dörd mühüm əlamət qeyd olunur. Əgər şəkilçinin

yalnız bir qrammatik əlaməti varsa, qalan üç mövqedə sıfırlar yazılır. Təhlilin tələbinə uyğun olaraq bəzi şəkilçilər birləşmə şəklində (ən çox müşahidə olunan formada) verilmişdir. Omonim şəkilçilər üçün bütün mümkün qrammatik variantlar verilmişdir. Şəkilçiliyin omonimliyin aradan qaldırılması çoxmənalılığın və omonimliyin ləğvi alqoritmlərinin köməyi ilə reallaşdırılır. Azərbaycan dilinin sözdəyişdirici şəkilçiləri barədə daha geniş izahat və şərtlərin verilməsinə ehtiyac hiss olunmamışdır (bu haqda bax: Махмудов, 1991).

Məlum olduğu kimi, türk dillərində şəkilçilər mətnlərin avtomatik işlənməsi sistemləri üçün və başqa nəzəri məqsədlərlə bir çox tədqiqatçılar tərəfindən araşdırılmışdır (Баскаков, 1957; Велиева, 1968, 1971; Глисон, 1959, 163-170; Ревзин, Юлдашева, 1969; Пинес, 1970; Пинес, Ревзин, 1969; Şerbak, 1963; Якубова, 1969, 1979). İ.İ.Revzin və Q.D.Yuldaşevanın tədqiqatında müasir özbək ədəbi dilində feli şəkilçilərin təsnifi və yerləşdirilməsi təsvir olunmuşdur (Ревзин, Юлдашева, 1969). Müəlliflər feli şəkilçiləri 7 tərtib üzrə cədvəl şəklində yerləşdirmişlər. Bu cədvəl əsasında fonetik şərtlərdən asılı olaraq alomorfun seçilməsini tənzimləyən fonoloji və müvafiq tərtib seçilməsindən asılı olaraq morfoloji qaydaların köməyi ilə özbək dilində feli şəkilçilərin generativ qrammatikası yaradılmışdır.

V.Y.Pinesin namizədlik dissertasiyası da bu qəbildən olan tədqiqatlardandır (Пинес, 1970). Burada Azərbaycan fel formalarının morfoloji strukturu araşdırılmış, qurulmuş model azərbaycanca-ingiliscə avtomatik lüğətdə sınaqdan keçirilmişdir. Müəllifin nəzəri müddəaları İ.İ.Revzin və Q.D.Yuldaşevanın fikirlərinə yaxındır.

K.A.Vəliyeva sözün sintezinin formal təsvirinə həsr olunmuş namizədlik dissertasiyasında şəkilçiləri də araşdırmış, sintezlə əlaqədar onların formal səciyyələrini işləyib hazırlamışdır (Велиева, 1971).

Konkret məqsəddən asılı olaraq təqdim olunan tədqiqatda Azərbaycan dilinin şəkilçiləri başqa prinsip üzrə təsnif edilmişdir. Avtomatik əsaslar lüğətində olduğu kimi, şəkilçilər də ibarət

olduqları hərflərin sayından asılı olaraq qruplaşdırılmışdır. Birinci qrupa doqquz, ikinci qrupa səkkiz və nəhayət, doqquzuncu qrupa bir hərfdən ibarət şəkilçilər daxil edilmişdir. Əgər şəkilçinin yalnız bir morfoloji əlaməti varsa, onda bu əlamət birinci pozisiyada qeyd olunur, qalan pozisiyalara isə sıfır yazılır. Məsələn, *-di, -di, -du, -dü* şəkilçisi üçün birinci pozisiyada onun əsas morfoloji səviyyəsi (şühudi keçmiş zaman) göstərilir. Bu şəkilçinin başqa qrammatik əlamətləri olmadığına görə qalan pozisiyalara sıfır yazılır (bax: cədvəl 18).

Şəkilçilər siyahısında çoxmənalı şəkilçilərin bütün mümkün qrammatik əlamətləri qeyd olunur. Məsələ burasındadır ki, morfoloji səviyyədə çoxmənalı şəkilçilərin dəqiq identifikasiyası qeyri-mümkündür. Məsələn *-acaq, -əcək (-acaq, -yəcək)* şəkilçiləri həm felin qəti gələcək zaman formasının, həm də feli sifətlərin əmələ gəlməsində iştirak edir. Bu şəkilçi ancaq sintaktik səviyyədə fərqləndirilə bilər: *kitab oxunacaq* (qəti gələcək zaman), *oxunacaq kitab* (feli sifət). Məz bu xüsusiyyətlərinə görə həmin şəkilçinin bütün mümkün qrammatik əlamətləri siyahıda qeyd olunur. Bəzi omonim şəkilçilər üçün isə dörd qrammatik əlamət göstərilir. Məsələn, *-ın, -in, -un, -ün* şəkilçisi ad nitq hissələrinə əlavə olunduqda yiyəlik halı və ya mənsubiyyəti (II şəxs, tək), fellərə birləşdikdə isə əmr şəklini (II şəxs, cəm) göstərir.

Avtomatik əsaslar lüğətində olduğu kimi, *k>y, q>g* əvəzlənmələrini nəzərə alaraq sonu *-k, -q* ilə bitən şəkilçilər iki variantda verilir: *-acaq -acağ, -əcək -əcəy*.

Formal təhlil sisteminə xas bəzi xüsusiyyətləri və qanunauyğunluqları nəzərə alaraq tez-tez birləşmə şəklində, yanaşı işlənən şəkilçi birləşmələri müstəqil şəkilçi kimi qeyd olunur. Təcrübə göstərir ki, bu hal mətnlərin avtomatik işlənməsi sistemlərində özünü doğruldur. Məsələn, *-sının, -sinin, -sunun, -sünün, -sında, -sində, -sunda, -sündə* və s.

Azərbaycan dilində söz-formaların tərkibində bəzi şəkilçilər sait və ya samit düşümü nəticəsində 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+sən - gəl+mir+sən

Alınmış yeni şəkilçi forması (-*mir*, -*mir*, -*mur*, -*mür*) inkarlıq+indiki zaman qrammatik əlamətlərinə malikdir. Göstərilən şəkilçi müşahidə olunan variantda şəkilçilər siyahısında təmsil olunmuşdur.

Nəqli keçmiş zaman şəkilçisi (-*miş*, -*miş*, -*muş*, -*müş*) sonda gələn *ş* samiti olmadan da işlənə bilər: *görmü(ş)sən* – *görmüsən*. Qeyd etmək lazımdır ki, bu xüsusiyyət artıq ədəbi dildə də möhkəmlənməkdədir. Digər tərəfdən, *ş* samiti olmadan bu şəkilçi -*mi*, -*mi*, -*mu*, -*mü* sual şəkilçiləri ilə qarışdırıla bilər. Bu isə öz növbəsində formal təhlil prosesini çətinləşdirə bilər. Məhz bu cəhət əsas tutularaq həmin şəkilçi də birləşmə formasında verilmişdir: -*misan*, *misən*, -*musan*, -*müsən* və s.

Köməkçi *idi*, *imiş* felləri avtomatik əsaslar lüğətində verildiyinə baxmayaraq, şəkilçilər siyahısında da təmsil olunmuşdur. Məsələ orasındadır ki, onlar müvafiq fonetik dəyişikliklərlə şəkilçi formasında da müşahidə olunurlar: *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*.

Şəkilçilər siyahısına qeyd olunan küməkçi fel və qoşmaların şəkilçi variantları da daxil edilmişdir: -*ydi*, -*ydi*, -*yla*, -*ylə*, -*çin*, -*çin*, -*çun*, -*çün* (-*yçin*, -*yçin*, -*yçun*, -*yçün*), -*ymış*, -*ymış*, -*ymuş*, -*ymüş* və s.

Ənənəvi qrammatikalarda ayrıca şəkilçi kimi göstərilməyən, yalnız mətnlərin avtomatik işlənməsi sistemlərində qeyd olunan bu şəkilçiləri maşın şəkilçiləri də adlandırırlar. Analogi olaraq yalnız formal təhlil sistemlərində müşahidə olunan *maşın əsasları* ifadəsinə də elmi ədəbiyyatda rast gəlinir (bax: Зорев, 1979; Пиотровская, 1973; Махмудов, 1991).

Cədvəl 18. Azərbaycan dili sözdəyişdirici şəkilçilərinin morfoloji blokda yerləşdirilməsi

Qruplar	Sözdəyişdirici şəkilçilər	Qrammatik əlamətləri	Kodu
1	2	3	4
1.0	-yacağımız, -yəcəyimiz oxuyacağımız, deyəcəyimiz	Feli sifət, I şəxs, cəm, qəti gələcək zaman	EAJW
1.1	-yacağınız, -yəcəyiniz oxuyacağınız, deyəcəyiniz	Feli sifət, II şəxs, cəm, qəti gələcək zaman	E+JW
1.2	-yacaqları, -yəcəkləri oxuyacaqları, deyəcəkləri	Feli sifət, III şəxs, cəm, qəti gələcək zaman	ELJW
2.0	-acağımız, -əcəyimiz yazacağımız, gürəcəyimiz	Feli sifət, I şəxs, cəm, qəti gələcək zaman	EAJW
2.1	-acağınız, -əcəyiniz yazacağınız, gürəcəyiniz	Feli sifət, II şəxs, cəm, qəti gələcək zaman	E+JW
2.2	-acaqları, -əcəkləri yazacaqları, gürəcəkləri	Feli sifət, III şəxs, cəm, qəti gələcək zaman	ELJW
2.3	-yacaqmış, -yəcəkmiş oxuyacaqmış, deyəcəkmiş	Gələcək zamanın rəvayəti) Xφφ
3.0	-yacağım, -yəcəyim oxuyacağım, deyəcəyim	Feli sifət, I şəxs, tək, qəti gələcək zaman	EALW
3.1	-yacağın, -yəcəyin oxuyacağın, deyəcəyin	Feli sifət, II şəxs, tək, qəti gələcək zaman	E+LW
3.2	-dıqları, duqları,	Feli sifət, III şəxs,	ELJĞ

	-dikləri, -dükləri yazdıqları, oxuduqları, gəldikləri, gördükləri	cəm, şühudi keçmiş zaman	
3.3	-yacaqdi, -yəcəkdi oxuyacaqdi, deyəcəkdı	Gələcək zamanın hekayəsi) φφφ
3.4	-acaqmış, -əcəkmiş oxuyacaqmış, deyəcəkmış	Gələcək zamanın hekayəsi) Xφφ
4.0	-yarkən, -yərəkən, - yürkən, -yurkən, -yirkən, - yirkən oxuyarkən, deyarkən, oxuyurkən, deyirkən, başlayirkən, sürüyürkən	Feli bağlama, qeyri-qəti gələcək, indiki zaman	= [φφ
4.1	-mişkən, -muşkən, -mişkən, -müşkən almışkən, oxumuşkən, gəlmışkən, görmüşkən	Feli bağlama, nəqli keçmiş zaman	= Qφφ
4.2	-yandan, -yəndən oxuyandan, deyəndən	Feli bağlama	= φφφ
4.3	-dırlar, -durlar, -dirilər, -dürlər adamdırlar, buzdurlar, tələbədirlər, güldürlər	III şəxs, cəm	Lφφφ
4.4	-acağım, -əcəyim yazacağım, görəcəyim	Feli sifət, I şəxs, tək, qəti gələcək zaman	EALW
4.5	-acağın, -əcəyin yazacağın, görəcəyin	Feli sifət, tək, qəti gələcək zaman	E+LW
4.6	-sından, -sundan, -sindən, -sündən atasından,	Mənsubiyyət, III şəxs, tək, çıxışlıq hal	RLLX

	qonşusundan, bibisindən, günüsündən		
4.7	-mişmiş, -muşmuş, -mişmiş, -müşmüş yazmışmış, oxumuşmuş, gəlmişmiş, görmüşmüş	Nəqli keçmişin rəvayəti	I*φφφ
4.8	-yirmiş, -yurmuş, -yirmiş, -yürmüş başlayırmış, oxuyurmuş, deyirmiş, sürüyürmüş	İndiki zamanın rəvayəti	Eφφφ
4.9	-acaqdi, -əcəkdi yazacaqdi, görəcəkdidi	Qəti gələcək zamanın hekayəsi	*φφφφ
4.10	-yarmış, -yərmiş oxuyarmış, deyərmiş	Qeyri-qəti gələcək zamanın rəvayəti]Xφφφ
4.11	-yacağı, -yəcəyi oxuyacağı, deyəcəyi	Fəli sifət, III şəxs, tək, qəti gələcək zaman	ELLW
4.12	-yıbmış, -yubmuş, -yibmiş, -yübmüş başlayıbmiş, oxuyubmuş, deyibmiş, sürüyübmiş	Keçmiş zamanın rəvayəti	*Xφφφ
5.0	-maqda, -məkdə oxumaqda, gəlməkdə	İndiki davamedici zaman	/φφφφ
5.1	-yaraq, -yərək oxuyaraq, deyərək	Fəli bağlama	=φφφφ
5.2	-yanda, -yəndə oxuyanda, deyəndə	Fəli bağlama	=φφφφ
5.3	-arkən, -ərkən, -irkən, -urkən, -irkən, -ürkən yazarkən, gələrkən,	Fəli bağlama	=φφφφ

	baxırkən, dururkən, gəlirkən, görürkən		
5.4	-dıqda, -dikdə, -duqda, -dükdə yazdıqda, gəldikdə, oxuduqda, gördükdə	Feli bağlama	=φφφ
5.5	-dıqca, -duqca, -dikcə, -dükcə yazdıqca, oxuduqca, gəldikcə, gördükcə	Feli bağlama	=φφφ
5.6	-yınca, -yunca, -yincə, -yüncə başlayınca, oxuyunca, deyincə, sürüyüncə	Feli bağlama	=φφφ
5.7	-madan, -mədən baxmadan, demədən	Feli bağlama	=φφφ
5.8	-andan, -əndən baxandan, gələndən	Feli bağlama	=φφφ
5.9	-dığın, -duğun, -diyın, -düyün aldığın, oxuduğun, gəldiyin, gördüyün	Feli sifət, II şəxs, tək, keçmiş zaman	E+LF
5.10	-acağı, -əcəyi yazacağı, görəcəyi	Feli sifət, III şəxs, tək, qəti gələcək zaman	ELLW
5.11	-yacaq, -yəcək, - yacağ, -yəcəy oxuyacaq/oxuyacağ, deyəcək/deyəcəy	Qəti gələcək zaman, feli sifət	Öφφφ
5.12	-sınız, -sınız, -sunuz, -sünüz yazırsınız, deyirsiniz, oxuyursunuz, görürsünüz	Xəbərlik II şəxs, cəm	Y+Jφ
5.13	-sının, -sunun,	Mənsubiyyət, III	RLLS

	-sinin, -sünün atasının, qonşusunun, bibisinin, günüsünün	şəxs, tək, yiyəlik hal	
5.14	-sında, -sunda, -sində, -sündə atasında, qonşusunda, bibisində, günüsündə	Mənsubiyyət, III şəxs, tək, yerlik hal	RLLZ
5.15	-ından, -undan, -ından, -ündən başından, qolundan, əlindən, gözündən	Mənsubiyyət, II şəxs, çıxışlıq hal	R+LX
5.16	-misan, -musan, -misən, -müsən yazmisan, oxumusan, demisən, görmüsən	Nəqli keçmiş zaman, II şəxs, tək	I+Lφ
5.17	-mişdi, -muşdu, -mişdi, -müşdü yazmışdı, oxumuşdu, gəlmişdi, görmüşdü	Keçmiş zamanın hekayəsi	*Xφφ
5.18	-irmiş, -urmuş, -irmiş, -ürmüş yazırmış, dururmuş, gəlirmiş, görürmüş	İndiki zamanın rəvayəti	[Xφφ
5.19	-armış, -ərmiş yazarmış, görərmiş	Qeyri-qəti gələcək zamanın rəvayəti] Xφφ
5.20	-yardı, -yərdi oxuyardı, deyərdi	Qeyri-qəti gələcək zamanın hekayəsi] φφφ
5.21	-yırdı, -yurdu, -yirdi, -yürdü başlayırdı, oxuyurdu, işləyirdi, sürüyürdü	İndiki zamanın hekayəsi	[φφφ
5.22	-ibmiş, -ubmuş, -ibmiş, -übmüş yazıbmiş, durubmuş, gəlibmiş, görübmiş	Keçmiş zamanın rəvayəti	*Xφφ

6.0	-yası, -yəsi oxuyası, deyəsi	Felin lazım şekli/feli sifət	Dφφφ
6.1	-malı, -məli oxumalı, görməli	Felin vacib şekli/feli sifət	Tφφφ
6.2	-araq, -ərək baxaraq, görərək	Feli bağlama	=φφφ
6.3	-anda, -əndə baxanda, görəndə	Feli bağlama	=φφφ
6.4	-ınca, -incə, -unca, - üncə alınca, gəlinçə, durunca, görünçə	Feli bağlama	=φφφ
6.5	-yalı, -yəli oxuyalı, deyəli	Feli bağlama	=φφφ
6.6	-acaq, -əcək, -acağ, - əcəy baxacaq/baxacağ, görəcək/görəcəy	Feli sifət	Dφφφ
6.7	-ımız, -imiz, -umuz, -ümüz kitabımız, dəftərimiz, odunumuz, gülümüz	Mənsubiyyət, I şəxs cəm	RAJφ
6.8	-ınız, -iniz, -unuz, - ünüz kitabınız, dəftəriniz, odununuz, gülünüz	Mənsubiyyət, II şəxs, cəm	R+Jφ
6.9	-ının, -inin, -unun, - ünün kitabının, dəftərinin, odununun, gülünün	Mənsubiyyət, II/III şəxs, tək, yiyəlik al	R+LŞ
6.10	-ında, -ində, -unda, - ündə kitabında, dəftərində, odununda, gülündə	Mənsubiyyət, II/III şəxs, tək, yerlik al	R+LZ
6.11	-sına, -sinə, -suna, -	Mənsubiyyət, II/III	RLLŞ

	sünə atasına, əmisinə, qonşusuna, günüsünə	şəxs, tək, yönlük hal	
6.12	-sını, -sini, -sunu, - sünü atasını, əmisini, qonşusunu, günüsünü	Mənsubiyyət, II/III şəxs, tək, tə'sirlik hal	RLLZ
6.13	-yçın, -yçin, -yçun, - yçün atasıyçın, əmisiyçin, qonşusuyçun, günüsüyçün	Qoşma	Rφφφ
6.14	-ymış, -ymiş, atasıymış, əmisiymiş,	Keçmiş zamanın rəvayəti	Iφφφ
6.15	-ycən, -ycən atasıycən, əmisiycən	Qoşma	Rφφφ
6.16	-ması, -məsi yazması, görməsi	Feli isim, mənsubiyyət, III şəxs, tək	-RLL
6.17	-dırt, -dırt, -durt, -dürt yazdırt, gəzdirt, vurdurt, sürdürt	İcbar növ	Fφφφ
6.18	-ıncı, -inci, -uncu, - üncü qırxıncı, beşinci, otuzuncu, üçüncü	Sıra sayı	I φφφ
6.19	-ardı, -ərdi yazardı, gələrdi	Qeyri-qəti gələcəyin hekayəsi] φφφ
6.20	-ırdı, -irdi, -urdu, -ürdü yazırdı, gəlirdi, dururdu, sürdürdü	İndiki zamanın hekayəsi	[φφφ
7.0	-ası, -əsi yazası, gələsi	Lazım şəkli/feli sifət	Dφφφ
7.1	-maq, -mək, -mağ, -	İnfinitiv	İ φφφ

	məy yazmaq/yazmağ, gəlmək/gəlməy		
7.2	-yıb, -yib, -yub, -yüb başlayıb, deyib, quruyub, sürüyüb	Feli bağlama/keçmiş zaman	: φφφ
7.3	-alı, -əli yazalı, görəli	Feli bağlama	=φφφ
7.4	-yar, -yər oxuyar, deyər	Qeyri-qəti gələcək zaman	DWφφ
7.5	-maz, -məz oxumaz, dinməz	Qeyri-qəti gələcək, inkarlıq/feli sifət	D-φφ
7.6	-caq, -cək alcaq, görcək	Feli bağlama	=φφφ
7.7	-yan, -yən oxuyan, deyən	Feli sifət/feli bağlama	Dφφφ
7.8	-miş, -miş, -muş, -müş yazmış, gəlmiş, oxumuş, görmüş	Nəqli keçmiş zaman/feli bağlama	Qφφφ
7.9	-yır, -yir, -yur, -yür başlayır, deyir, oxuyur, sürüyür	İndiki zaman	Bφφφ
7.10	-nın, -nin, -nun, -nün qapının, əlinin, qonşunun, sürünün	Mənsubiyyət, II şəxs, tək, yiyəlik hal	R+LŞ
7.11	-dan, -dən qapıdan, dəftərdən	Çıxışlıq hal	Xφφφ
7.12	-lar, -lər kitablar, dəftərlər	Cəm hal	Jφφφ
7.13	-yam, -yəm atayam, tələbəyəm	Xəbərlik, I şəxs, tək	YALφ
7.14	-yım, -yim, -yum, - yüm başlayım, deyim, oxuyum, sürüyüm	Əmr şəkli, I şəxs, tək	ALÜφ

7.15	-yaq, -yək oxuyaq, deyək	Felin əmr şəkli/arzu şəkli, I şəxs, cəm	AJÜΦ
7.16	-mız, -miz, -muz, - müz atamız, əmimiz, qonşumuz, sürümüz	Mənsubiyyət, I şəxs, cəm	RAJΦ
7.17	-san, -sən oğlansan, tələbəsən	Xəbərlik, II şəxs, tək	+LYΦ
7.18	-nız, -niz, -nuz, -nüz atanız, əminiz, qonşunuz, sürünüz	Mənsubiyyət, II şəxs, cəm/xəbərlik	R+JΦ
7.19	-yın, -yin, -yun, -yün başlayın, deyin, oxuyun, sürüyün	II şəxs, cəm, əmr şəkli	Ü+JΦ
7.20	-sın, -sin, -sun, -sün yazsın, desin, dursun, sürsün	III şəxs, tək, əmr şəkli	ÜLLΦ
7.21	-dır, -dir, -dur, -dür kitabdır, dəftərdir, odundur, sürüdür, yazdır, gəzdir, qurdur, sürdür	III şəxs, tək, xəbərlik/icbar növ	ILLΦ
7.22	-yığ, -yik, -yuq, -yük atayıq, tələbəyik, quruyuq, sürüyük	I şəxs, cəm/xəbərlik	YAJΦ
7.23	-ını, -ini, -unu, -ünü kitabını, dəftərini, odununu, gülünü	Mənsubiyyət, II/III şəxs, tək, təsirlik hal	RLLZ
7.24	-ına, -inə, -una, -ünə kitabına, dəftərinə, odununa, gülünə	Mənsubiyyət, II/III şəxs, tək, yönlük hal	RLLH
7.25	-mır, -mir, -mur, -mür almır, gəlmir, oxumur, gülmür	İndiki zamanın inkar forması	-VΦΦ

7.26	-yla, -ylə atasıyla, əmisiylə, başıyla, əliylə	Qoşma/tərzi ərəkət zərfi	Rφφφ
7.27	-ydı, -ydi balacaydı, tələbəydi	Keçmiş zamanın hekayəsi	İĞφφ
7.28	-çın, -çin, -çun, -çün atançın, səninçin, onunçun, üzünçün	Qoşma	Rφφφ
7.29	-can, cən oracan, evəcən	Qoşma	Rφφφ
7.30	-yıl, -yil, -yul, yül yuyulmaq, deyilmək	Məcəul növ	<φφφ
7.31	-ncı, nci altıncı, ikinci	Sıra sayı	Iφφφ
7.32	-dığ, -diy, -duğ, -düy Bu şəkilçi ayrılıqda işlənmiş. Şəxs şəkilçiləri ilə birləşmiş şəkildə geniş müsaidə olunur. başladığımız, başladığım, başladığınız, başladığın, başladıkları, başladığı	Feli sifət, şühudi keçmiş zaman	EFφφ
8.0	-ya, -yə oxuya, deyə, qapıya, gəmiyə	Arzu şəkli/yönlük hal	2φφφ
8.1	-sa, -sə alsa, gəlsə	Şərt şəkli	Gφφφ
8.2	-ıb, -ib, -ub, -üb alıb, gəlib, durub, görüb	Feli bağlama/keçmiş zaman	: φφφ
8.3	-ar, -ər alar, gələr	Qeyri-qəti gələcək zaman/feli sifət	DWφφ

8.4	-an, -ən alan, gələn	Feli sifət/feli bağlama	Dφφφ
8.5	-dı, -di, -du, -dü aldı, gəldi, durdu, gördü	Şühudi keçmiş zaman	Ğφφφ
8.6	-ır, -ir, -ur, -ür alır, gəlir, durur, görür	İndiki zaman	Vφφφ
8.7	-ma, -mə yazma, gəlmə, burma (saç), döymə (metal)	İnkarlıq/feli sifət	:φφφ
8.8	-ın, -in, -un, -ün kitabın, dəftərin, odunun, gülün, atın, gəlin, oturun, gülün	Yiyəlik al/mənsubiyyət, II şəxs, tək/ əmr şəkli, II şəxs, cəm	9φφφ
8.9	-ni, -ni, -nu, -nü qapını, gəmini, qutunu, sürünü	Mənsubiyyət II/III şəxs, tək/təsirlik hal	3φφφ
8.10	-da, -də kitabda, dəftərdə	Yerlik hal	Zφφφ
8.11	-am, -əm oğlanam, müəlliməm baxıram, görürəm	Xəbərlik, I şəxs, tək	ALYφ
8.12	-ım, -im, -um, -üm kitabım, dəftərim, odunum, gülüm, alım, gəlim, oturum, gülüm	Mənsubiyyət, I şəxs, tək/əmr şəkli/ I şəxs, tək	Bφφφ
8.13	-aq, -ək alaq, gələk	Əmr şəkli, I şəxs, cəm/arzu şəkli, I şəxs, cəm	AJÜφ
8.14	-sı, -si, -su, -sü atası, əmisi, qonşusu, sürüsü	Mənsubiyyət, III şəxs, tək	RLLφ
8.15	-mı, -mi, -mu, -mü alırımı, gəlirimi,	Sual	Gφφφ

	dururmu, görürmü		
8.16	-la, -lə atamla, əmimlə	Qoşma	Rφφφ
8.17	-kı, -ki, -ku, -kü Bakıdakı, bizdəki, onunku, üzünkü	Mənsubiyyət	Rφφφ
8.18	-na, -nə qapına, evinə	Yönlük hal, mənsubiyyət, II/III şəxs	Hφφφ
8.19	-ıl, -il, -ul, -ül baxılmaq, dərilmək, qurulmaq, görülmək	Məchul növ	<φφφ
8.20	-ış, -iş, -uş, -üş yazışmaq, girişmək, vuruşmaq, döyüşmək	Qarşılıq növ	>φφφ
8.21	-ıq, -ik, -uq, -ük adamıq, əsirik, qohumuq, gülük, baxırıq, gəzirik, dururuq, görürük	Xəbərlilik, I şəxs, cəm	YAJφ
9.0	-a, -ə yaza, gələ	Arzu şəkli/yönlük hal	2φφφ
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	Əφφφ
9.3	-q, -k aldıq, gəldik	I şəxs, cəm	AJφφ
9.4	-n qapın	Mənsubiyyət, II şəxs, tək	+Lφφ
9.5	-ş mələşmək	Qarşılıq növ	>φφφ

6.8. Morfoloji blokda nitq hissələrinin, ayrı-ayrı qrammatik kateqoriyaların kodlaşdırılması

Formal təhlil sistemlərdə və MT-də vahid kodlaşdırmanın çox mühüm əhəmiyyəti vardır. Respublikamızda göstərilən problemlər çərçivəsində fəaliyyət göstərən kollektivlər eyni kod sistemindən istifadə edirlər. Bu və ya digər problemlə əlaqədar konkret alqoritmlərdən istifadə zamanı vahid kodlaşdırma bir çox ziddiyyətlərin və uyğunsuzluqların aradan qaldırılmasına xidmət edir. Şübhəsiz, istənilən kod sistemi şərtidir və tədqiqatın müəyyən pillələrində təkmilləşdirilə və yenisi ilə əvəz oluna bilər (Аршинов, Садовский 1983; Борисевич, Гончаренко, Гречишкина, Добрускина, Кочеткова, Крисевич, Нехай, Пиотровский, Штирбу, Ястребова, 1970). Lakin kod sisteminin qismən təzələnməsi və əvəzlənməsi zərurət olduqda həyata keçirilməlidir və bu barədə həmin kodlardan istifadə edən mütəxəssislərə məlumat verilməlidir. Bu, kodların universallığı və vahidliyi prinsiplərinin tələbidir. Hazırda mövcud təhlil sistemlərində və MT-də istifadə olunmuş kodlar keçmiş ümumittifaq “Статистика речи” qrupunda işlənilib hazırlanmış sxem əsasında aparılmışdır. Yığcamlıq və konkretlik üçün nitq hissələrinin, ayrı-ayrı kateqoriyaların, leksik və leksik-qrammatik informasiyaların kodları qruplar üzrə cədvəllərdə verilmişdir:

Cədvəl 19. Şəkilçisiz işlənmə söz- formaları ehtiva edən nitq hissələrinin kodları

Nitq hissələri (söz qrupları)	Kodu	Misallar
Qoşma	P	əvvəl, ilə, üçün
Bağlayıcı	B	və, həm
Ədat	U	bəs,kaş
Modal sözlər	B	demək, şübhəsiz
Bağlamalar	B ₁	idi, imiş
Nida	!	ay, of

Cədvəl 20. Tərkibində şəkilçi qəbul edə bilən söz-formalar olan nitq hissələrinin kodları

Nitq hissələri	Kodu	Misallar
İsim	I	ölkə, çörək
Sifət	S	yaxşı, soyuq
Say	S _q	beş, beşinci
Əvəzlik	Ə _v	mən, bu
Fel	F	almaq, vermək
Zərf	Z	yavaş, aşağı

Cədvəl 21. Ad nitq hissəsi sözdəyişdirici kateqoriyaları və formalarının kodları

Adların sözdəyişdirici kateqoriya və formaları	Kodu	Misallar
Hal kateqoriyası		
Yiyəlik	Ş	adamin, qapinin
Yönlük	H	adama, qapıya
Təsirlik	Z	adami, qapini
Yerlik	Z	adamda, qapıda
Çıxışlıq	C	adamdan, qapıdan
Mənsubiyyət kateqoriyası	R	kitabın, dəftərin
Xəbərlik kateqoriyası	Y	kitabdır, dəftərdir

6.9. Morfoloji təhlilin əsas istiqamətləri və mərhələləri

Kompüterlərdə hər hansı bir dil səviyyəsində təhlil və sintez keçirilməsi üçün müəyyən linqvistik bazanın mövcudluğu mühüm şərtidir. Linqvistik bazanın mühüm komponenti hesab olunan leksik təminat leksik blok vasitəsi ilə həyata keçirilir. Yuxarıda mətnin formal təhlili sisteminin leksik bloku barədə geniş məlumat verilmiş, onu təşkil edən sözlüklər, lüğətlərin quruluşu, ən mühüm linqvistik xüsusiyyətləri və işlənmə perspektivləri izah olunmuşdu. Linqvistik təhlil prosesi həm də sistemin güclü qrammatik informasiya ilə təchizi ilə şərtlənir. Qrammatik informasiya sistemə müxtəlif üsullarla və hissə-hissə

daxil edilir. Əgər informasiyanın müəyyən hissəsi AL-də ehtiva olunursa, digər qismi şəkilçilər cədvəlində öz əksini tapır. Qrammatik qaydalar isə alqoritmlərlə ifadə olunur. Formal təhlil sisteminə alqoritm mətnlərdən (formadan) mənaya keçidə imkan verən, müəyyən ciddi ardıcılıqla düzülmüş qaydalar çoxluğuur.

Cədvəl 22. Felin sözdəyişdirici kategoriya və formalarının kodları

Felin sözdəyişdirici kategoriya və formaları	Kodu	Misallar
1	2	3
Növ kategoriyası		
Məchul	F _(m)	yazılmaq, atılmaq
Qarşılıq	F _(q)	yazışmaq, atışmaq
İcbar	F _(i)	yazdırmaq, atdırmaq
Felin şəkilləri		
Əmr	F _(ə)	yazım, atım
Xəbər	F _(x)	oxuyuram, atıram
Arzu	F _(a)	oxuyam, atam
Vacib	F _(v)	oxumalıyam, atmalıyam
Lazım	F _(L)	oxuyasıyam, atasıyam
Şert	F _(e)	oxusam, atsam
Davam	F _(d)	oxumaqdayam, yazmaqdayam
Felin zamanları		
Şühudi keçmiş	F	oxudum, yazdım
Nəqli keçmiş	I	oxumuşam, yazmışam
İndiki	B	oxuyuram, yazıram
Qəti gələcək	W	oxuyacağam, yazacağam
Qeyri-qəti gələcək	O	oxuyaram, yazaram
Nəqli keçmişin rəvayəti	I *	oxumuşdu, yazmışdı
Qəti gələcək zamanın rəvayəti	W *	yazacaqmış, oxuyacaqmış
Qeyri-qəti gələcək	O *	oxuyarmış, yazarmış

zamanın rəvayəti		
İndiki zamanın hekayəsi	{	oxuyurdu, yazırdı
İnkarlıq	-	oxumaram, yazmaram
Məsdər	F _(m)	oxumaq, yazmaq
Feli sifət	F _(s)	oxuyan, yazan
Feli bağlama	F _(Bi)	oxuyub, yazıb

Cədvəl 23. Ad və fel nitq hissələri üçün ümumi olan kateqoriyaların kodları

Kateqoriya və formalar	Kodu	Misallar
<u>Kəmiyyət kateqoriyası</u>		
tək	L	kitab, dəftər
cəm	J	kitablar, dəftərlər
<u>Şəxs kateqoriyası</u>		
I şəxs	A	oxuyuram, oxuyuruq
II şəxs	+	oxuyursan, oxuyursunuz
III şəxs	L	oxuyur, oxuyurlar
Sual	G	oxuyurmu? gəlirmi?

Cədvəl 24. Omonim əsasların kodları

Omonim əsaslar (mümkün qrammatik səciyyələri)	Kodu	Misallar
isim/fel	I/F	qaz, yaz
isim/sifət	I/S	qoca, gözəl
fel/zərf	F/Z	düz, dik
fel/sifət	F	sarı, burma, qorxaq
say/fel	S _a /F	min, qırx
bağlayıcı/ədat	B/Ə _d	ancaq, ki
İnkarlıq/feli isim	-/F _L	oxuma, yazma

Cədvəl 25. Omonim şəkilçilərin kodları

Omonim şəkilçilər /mümkün qrammatik səciyyələri/	Kodu	Misallar
Lazım şəkli /feli sifət	0	-ası/-əsi,-yası, -yəsi/
Feli bağlama/feli sifət	D	-an/-ən, -yan, -yən/
İcbar növ/xəbərlik	I	-dır/-dir, -dur, -dür/
Arzu şəkli/yönlük hal	2	-a/-ə, -ya, -yə/
Mənsubiyyət/təsirli k hal	Ə	-ı/-u, -ü, -i/
Vacib şəkli/feli sifət	T	-malı/-məli/
Mənsubiyyət, I şəxs, tək/əmr şəkli	B	-ım/-im, -um, -üm/
Qəti gələcək zaman/feli sifət	ö	-acaq/-əcək,-yacaq,-yəcək/
Feli sifət/nəqli keçmiş zaman	Q	-mış/-miş,-muş,-müş/
Yiyəlik hal/mənsubiyyət II-III şəxs, tək/əmr şəkli	9	-ın/-in, -un, -ün/
Feli sifət/nəqli keçmiş zaman	:	-yıb/-yib,-yub,-yüb/-mış/-miş, -muş, -müş/

Göstərilən alqoritmləri formal prinsiplər əsasında qurulmuş və formal kriteriyalarla izah oluna biləcək qrammatika kimi səciyyələndirmək olar. Formal qrammatikalardan fərqli olaraq alqoritmlərdə qaydaların tətbiqində ixtiyarilik və «sərbəstlik» yoxdur (formal qrammatikalar barədə bax: Гладкий, 1973).

Alqoritmlərin quruluşundakı mükəmməl «intizam», dəqiqlik, qayda-qanuna ciddi əməl olunma – bütün bunlar onların tərtibinə məsuliyyətlə yanaşılmasını tələb edir. Göstərmək lazımdır ki, linqvistik təhlil alqoritmlərinin tərtibində cüzi diqqətsizlik elə ilk sınaq zamanı aşkar çıxır və bütün sistemin işi dayanmış olur. Ona görə də, «başdansovma», «yarımçıq», «müvəqqəti» istifadə üçün nəzərdə tutulmuş alqoritmlər sistemə daxil edilə bilməz. Bu cəhətdən alqoritmin tərtibi prosesində tam məqsəd aydınlığı olmalı, təhlilin əsas istiqamətləri və mərhələləri tam müəyyənləşdirilməlidir. Başqa sözlə, tərtib zamanı məlum olmalıdır ki, alqoritm nəyə xidmət edir, təhlil hansı linqvistik bazaya əsaslanır, haradan başlanır, nə vaxt qurtarır, təhlilə hansı linqvistik material cəlb edilir və nəticədə hansı informasiyaların alınması vacibdir. Əgər belə dəqiq məqsəd aydınlığı olmazsa, alqoritmin, eləcə də bütövlükdə təhlil prosesinin səmərəsindən, aktuallığından danışmaq olmaz.

Məhz yuxarıda göstərilənləri əsas tutaraq formal sistemlərdə morfoloji təhlilin əsas istiqamətlərini nəzərdən keçirək. Təhlil prosesində tədqiqat obyektini Azərbaycan mətnləridir. Məqsəd həmin mətnin (və ya mətnlərin) təhlili, ən kiçik dil vahidlərinə ayrılması (dil vahidlərinə ayrılmanın səviyyəsi tədqiqatçının konkret məqsədlərinə görə müxtəlif ola bilər, məsələn, mətnlərin işlənməsi prosesində fonem və morfem səviyyəsindən daha yuxarı səviyyələrdə təhlil qarşıya məqsəd qoyula bilər. Başqa bir tədqiqatda isə yalnız fonem səviyyəsində araşdırmalar ön plana çəkilə bilər və s.), hər bir ayrılmış vahidin qrammatik səciyyəsinin müəyyənləşdirilməsi (onun səciyyəvi qrammatik əlamətlərinin tapılması və yazılması), tədqiq olunmuş mətnin qrammatik cəhətdən formal əlamətlərə görə tam mənimsənilmiş, qavranmış şəkildə yenidən tam sintezidir. Göstərmək lazımdır ki, tədqiq olunmuş dil vahidləri haqqında informasiya bir çox elementar məlumatların məcmuyundan ibarət ola bilər. Elə bir halda müşahidə edilə bilər ki, informasiya bir elementar məlumatdan təşkil olunmuş olsun. Məsələn, -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 məlumat beş elementar məlumatdan təşkil olunmuşdur. Semantik blokun «müdaxiləsindən» sonra bu məlumatların hansının təhlil mətni (kontekst) üçün səciyyəvi və məqbul olması aydınlaşdırılır.

Formal təhlil sisteminin morfoloji blokunun əsasını təşkil edən alqoritmin ən ümumi cəhətlərini nəzərdən keçirək.

Yuxarıda qeyd etmişdik ki, $\langle A, F, A' \rangle$ modelində F-funksiyalar kompozisiyası, rekursiv qaydalar çoxluğu (Роджерс, 1972; Мальцев, 1965). F funksiyası formadan (A) mənaya (A') keçidi təmin edir. Formal təhlil sistemində F-funksiyası alqoritm şəklində verilmişdir. Bu alqoritmi formal qrammatika, dilin müəyyən vahidləri haqqında məlumat almağa imkan verən (qrammatik və ya leksik) rekursiv qaydalar məcmuunu kimi səciyyələndirmək olar (Гросс, 1971).

F-funksiyasına bir-biri ilə sıx əlaqədə olan K, P, R, C, P funksiyalarının kompozisiyası kimi baxılır (Фиталов, 1961, 2; Кузнецов, 1980, 50). Göstərilən kompozisiya aşağıdakı ifadə oluna bilər:

$$F(x)=P(C(R(R(K(x')))))$$

Burada K, R, R, S, P - funksiyalarının aşağıdakı mənaları vardır:

- K-funksiyası mətni mətndən, cümləni mətndən, söz-formanı cümlədən ayırır;
- R-funksiyası hər bir söz-formanı morfem siyahıları (əsaslar avtomatik lüğəti və şəkilçilər siyahısı) ilə tutuşdurur və beləliklə, söz-formanın əsas və şəkilçilərə ayrılmasını təmin edir;
- R-funksiyası əsas və şəkilçilərin omonimliyini aradan qaldırır;
- C-funksiyası hər bir X-morfemi üçün müəyyən informasiyanı (qrammatik və ya leksik) göstərir;

- P-funksiyası giriş söz-formasına müvafiq leksik-qrammatik informasiyanın çıxışda əldə olunmasını təmin edir.

Söz-forma haqqında informasiya - elementar informasiyalar çoxluğudur (elementar informasiyalar disfunksiyasıdır). İnformasiya birmənalı olaraq bir elementar informasiyadan ibarət ola bilər. Elementar informasiyalar bir-biri ilə konyuktiv əlaqələrlə bağlanmış bir neçə əlamətdən ibarət ola bilər. R-funksiyasının (omonimliyın aradan qaldırılması funksiyası) işindən sonra bu elementar informasiyanın müvafiq gələnini seçilir.

Təhlil vahidlərinin ayrılması funksiyası K aşağıdakı kimi fəaliyyət göstərir: əvvəlcə mətn (məyyyə yazılı mətn parçası) mətnlər çoxluğundan (massivdən) ayrılır. Bunun üçün məyyyə edilmiş qayda üzrə təhlil vahidinin həcmi (uzunluğu) qabaqcadan məyyyənləşdirilir və onun əvvəlində, eləcə də sonunda xüsusi şərti işarə qoyulur. Sonra həmin təhlilin vahidi kimi ayrılmış mətndən cümlə seçilir. Cümlənin seçilib ayrılması onun sonunu göstərən durğu işarələrinin axtarışı ilə reallaşdırılır. Sonrakı mərhələdə söz-forma cümlədən ayrılır. Bu vaxt nəzərdə tutulur ki, söz-forma bir boşluqdan (probeldən) digər boşluğa (probelə) qəddər olan məsafədə yerləşən hərflər ardıcılığı zənciridir.

Söz-formaların hissələrinə ayrılması funksiyası R-nin işləməsi üçün morfem siyahılarının olması əsas şərtidir. Belə siyahılar olmadan söz-formaların hissələrinə bölünməsi və qrammatik cəhətdən səciyyələndirilməsi qeyri-mümkündür. Morfem elementləri çoxluğundan ibarət bu siyahılar əsas və şəkilçiləri ehtiva edir. Söz-formaların əsas və şəkilçilərə ayrılması, eləcə də qrammatik səciyyələri göstərilməklə şəkilçilərin «düzgün» olaraq bir-birindən fərqləndirilməsi mürəkkəb prosesdir. P-funksiyası söz-formanın tərkibində heç bir dəyişiklik aparmadan onu morfem hissələrinə ayırır və lazımi qrammatik informasiyanın alınmasını təmin edir.

S-funksiyasının işi zamanı söz-forma ilə onun ifadə etdiyi məna arasında uyğunluq söz-formaların morfem siyahıları ilə tutuşdurulması ilə təmin edilir.

Əsas və şəkilçilərin omonimliyinin aradan qaldırılması isə R-funksiyasının köməyi ilə reallaşdırılır. R-funksiyasının işi formal təhlil sisteminin semantik blokunda ətraflı şərh olunmuşdur.

Morfoloji təhlildə ilk mərhələ təhlil vahidlərinin ayrılması mərhələsidir. Fərz edək ki, n sayda mətnlər çoxluğu verilmişdir. Tələb olunur ki, həmin çoxluq kompüterdə təhlil olunsun və təhlil sistemi üçün zəruri informasiya əldə edilmiş olsun. Hər şeydən əvvəl təhlilin həcmi müəyyənləşdirilməlidir. Kompüterin operativ yaddaşından və tədqiqatın qarşısında qoyulmuş vəzifələrdən asılı olaraq təhlil mətninin uyğunluğu müxtəlif ola bilər. Təhlil mətni kimi səhifə, abzas, cümlə, işarə 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 mətnlərə müəyyən şərti işarələr qoyulmamışsa, sadə sayğac üsulu ilə təhlil mətnini ayırmaq olar. Məsələn, elə qəbul etmək olar ki, hər 100 söz-forma bir təhlil mətnidir. Kompüter sayğacı mətnlər çoxluğundan 100 söz-forma ayıran kimi proses dayanır və təhlil başlanır. Təhlil başa çatdıqdan sonra sayğac alqoritmin müəyyən blokuna müraciət edir və yeni mətn tələb olunub-olunmadığını «öyrənir» və lazım gələrsə, yeni mətn ayırır. Əgər yeni mətn tələb olunmursa, sayğac mətnin ayrıldığı sərhəd nöqtəsini yadda saxlayır və iş bununla bitmiş hesab olunur. Qeyd etmək lazımdır ki, mətnlər çoxluğundan təhlil mətninin ayrılması elə bir prinsipial çətinlik doğurmur və sırf texniki məsələ hesab olunur.

Növbəti mərhələ - mətni təşkil edən digər vahidlərin ayrılmasıdır. Mətn nəzəri olaraq bir cümlədən və ya abzasdan, eləcə də çoxlu sayda cümlə və abzaslardan ibarət ola bilər. Elə qəbul edək ki, bizim tədqiqat prosesində mətn dedikdə, N sayda abzaslardan və n sayda cümlələrdən təşkil olunmuş kompozisiya başa düşülür:

$$n < N < T$$

Burada T -mətni (teksti) göstərir: $T = \{N_i\}$, $i=1, \dots, k$. N də öz növbəsində özündən kiçik vahidlərdən təşkil olunmuşdur: $N = \{n_i\}$, $i=1, \dots, k$.

Cümlənin özü də söz-formalardan təşkil olunmuşdur: $n = \{c_i\}$, c-söz-formanı göstərir və $i = 1, \dots, k$. Bu prosesi ardıcıl olaraq izləsək, daha kiçik vahidlərə gedib çıxırıq. Söz-forma əsas və sözdəyişdirici şəkilçilərdən (əsasların özləri də kök və sözdüzəldici şəkilçilərdən təşkil olunur) ibarət bir vahiddən ibarətdir. Alınmış morfemlərin özlərinin də daha kiçik vahidlərə - fonemlərə ayrıldığını qəbul etsək, aşağıdakı kimi bir iyerarxiya alınır (bu barədə bax: Vəliyev, 1989, 36, 39; Cəlilov, 1985, 13):

Mətn—abzas—cümlə—söz-forma—morfem—fonem.

Mətnin formal təhlili sistemində fonem səviyyəsində təhlil tədqiqatın məqsəd və vəzifələrinə müvafiq olaraq aparılmışdır.

Azərbaycan dilçiliyində fonem səviyyəsində müxtəlif səpili bir çox araşdırmalar mövcuddur və konkret məqsədlərlə həmin tədqiqatlardan istifadə olunması zərurəti istisna edilmir (bax: Ахундов, 1984; Мамедов, 1985; Рахманов, 1988; Cəlilov, 1988).

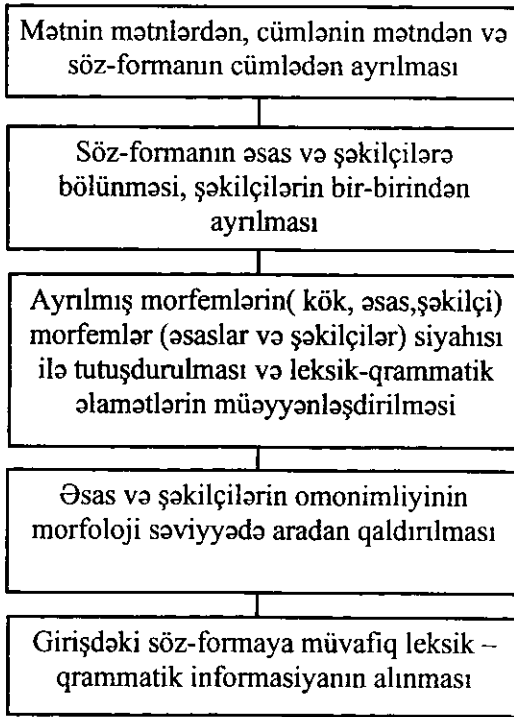
Qeyd etmək lazımdır ki, təqdim olunan sistemdə sözdüzəldici şəkilçilərin və söz köklərinin ayrılıqda geniş təhlili də nəzərə alınmamış, belə informasiyalar imkan daxilində leksik blokda lüğətlərə əlavə olunmuşdur. Bu həm də formal təhlil sistemlərinin praktik tətbiqi ilə əlaqədardır. Yuxarıdakıları yekunlaşdıraraq belə bir nəticəyə gəlmək olar ki, fonem səviyyəsi, sözdüzəldici şəkilçilər və onlarla düzələn vahidlər (lüğət vahidi kimi təsbit olunmuş konstruksiyalar) təhlil bloklarında ayrıca araşdırılır:

Mətn—abzas—cümlə—söz-forma/əsas—sözdəyişdirici şəkilçilər (Sxem 7).

Mətndən sonrakı vahidin – abzasın ayrılması da şərti səciyyəyə daşıyır. Abzaslar bir-birindən və mətndən probel (boşluq) işarəsi və ya qabaqcadan şərtləşdirilmiş hər hansı işarə ilə ayrıla bilər. Cümlə abzasdan kiçik, söz-formadan böyük vahid sayılır. Göstərmək lazımdır ki, abzas və cümlənin ayrıca vahidlər kimi verilməsi semantik blokun fəaliyyəti üçün mühüm əhəmiyyət kəsb edir. Məlum olduğu kimi, bir çox sözlərin həqiqi mənaları yalnız cümlə və abzaslarda, müəyyən kontekstdə, sintaktik əhatədə məlum olur.

Cümlə abzasdan onun əvvəlində və sonunda yerləşmiş durğu işarələri əsasında 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.

Sonra cümlədən söz-formaların ayrılması mərhələsi başlanır. Bu vaxt nəzərdə tutulur ki, söz-forma bir probel (boşluq) işarəsindən digər probel işarəsinə qədər olan məsafədə yerləşən hərflər ardıcılığı zənciridir.



Sxem 7. Mətnin formal morfoloji təhlil algoritminin mərhələləri

Söz-formanın hissələrə ayrılması ən mühüm və mürəkkəb mərhələlərdəndir. Bu prosesin həyata keçməsi üçün qabaqcadan tərtib olunmuş morfem siyahılarının mövcudluğu çox zəruridir. Leksik blokda yerləşdirilmiş lüğət və məlumatlar, eləcə də

morfoloji blokda yerləşdirilmiş şəkilçilər siyahısı, şəkilçilərin qrammatik səciyyələri əsasında söz-formaların «düzgün» olaraq hissələrə ayrılması (əsasə və ayrı-ayrı şəkilçilərə bölünməsi) baş verir. Qeyd etmək lazımdır ki, söz-formaların əsasə və şəkilçilərə, eləcə də şəkilçilərin bir-birindən «düzgün» ayrılması türk dillərində, o cümlədən Azərbaycan dilində çox mürəkkəb səciyyə daşıyır. Məsələn, ilk nəzərdən çox sadə və adi görünən «oxuya-caq» söz-forması kompüterdə aşağıdakı şəkildə hissələrə (kök və şəkilçilərə) ayrıla bilər:

ox - u - ya - caq

oxu - ya - ca - q

oxu - y - ac - aq

oxu - yac - aq

oxu - yacaq

oxu - ya - caq və s.

Praktik olaraq 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 azırlanmalıdır ki, söz-formadakı morfemlər tamamilə düzgün olaraq ayrıla bilsin və bu zaman söz-formanın təhlildən əvvəlki mənası tərif olunmasın. Morfoloji blok söz-formadan morfemləri ayıraraq, onları leksik və morfoloji bloklardakı siyahılarla tutuşdurmaqla sistemi tam təmin edən nəticələr əldə etməyə imkan verir. Morfoloji blokda morfem variantları (allomorflar) müstəqil morfemlər kimi təmsil olunur. Məsələn, indiki zaman şəkilçilərinin bütün allomorfları -ır, -ir, -ur, -ür, eləcə də -yır, -yir, -yur, -yür müstəqil şəkilçi kimi avtonom qrammatik informasiya ilə təciz olunmuşdur. Əlbəttə, bir allomorfdan bu allomorfları təmsil edən əsas morfemə keçməyə imkan verən xüsusi qaydalar işləyib hazırlamaq olardı. Lakin bu qaydalar təhlil alqoritmlərinin olduqca mürəkkəbləşməsinə və onun kompüterlərdə reallaşdırılmasının çətinləşməsinə səbəb olardı. Ona görə də, ilk nəzərdən primitiv təsir bağışlasa da kodlaşdırma zamanı hər bir allomorf üçün ayrılıqda qrammatik məlumatın təkrarlanması məqsədəuyğun hesab olunmuşdur. Sonralar kompüter sınaqları bu metodun düzgünlüyünü bir daha təsdiq etmişdir. Sintez prosesi üçün bir əsas morfemdən allomorflara

keçidin qanunauyğunluğunu işləyib hazırlamaq çətinlik törətmir və təcrübədə özünü doğruldur.

Məsələn, hər hansı bir söz-formanın sintezi zamanı biz məlum qaydaları (şəkilçilərin sıralanması ardıcılığı, ahəng qanunu və s.) qabaqcadan «saf-çürük» edərək yeganə düzgün variantı tapmaqla sintez apara bilirik. Məsələn, fərz edək ki, bizə aşağıdakı morfemlər verilmiş, onların hamısını düzgün işlətməklə Azərbaycan ədəbi dilinə zidd olmayan ifadələr almaq tələb olunur:

«baba», «çiçək», -lar (-lər), -ımız(-imiz, -umuz, -ümüz), -dan(-dən).

Təbiidir ki, «baba» sözündən sonra -ümüz, -dən, sonra isə -lar şəkilçilərini işlətmək dilimizin qrammatikasına zidd olardı. Əvvələn, burada sözdüzəldici və sözdəyişdirici şəkilçilərin sıralanması qaydası, sonra isə ahəng qanunu pozulmuşdur. Sintez alqoritmləri «babaümüzdənlar» söz-formasının düzgün olmadığını «hiss edirlər». Çünki alqoritmlərdə şəkilçilərin sıralanması və ahəng qanunu özünün formal ifadəsini tapmışdır və sintez demək olar ki, heç bir xəyata yol verilmədən aparılır. Şəkilçilərin sıralanma ardıcılığına görə tapırıq ki, bu göstərilən şəkilçilər aşağıdakı ardıcılıqla düzülə bilər: əvvəlcə cəm şəkilçisi, sonra mənsubiyyət və hal şəkilçisi. Konkret misaldan 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 alqoritmlərinin işi barədə, bir qədər səthi və bəsit görünsə belə, verilən izahatdan da görünür ki, burada bütün qrammatik qanunauyğunluqlar və istisnalar qabaqcadan nəzərə alınabilir (Белиева, 1971; Пинес, 1970; Садыков, 1987).

Təhlildə isə proses mürəkkəb və ziddiyyətli gedir. Təhlil prosesinə cəlb olunmuş söz, onu təşkil edən, ona birləşmiş şəkilçilər, səsdüşümü, səstartımı, kök-şəkilçi, şəkilçi-şəkilçi sərhədlərində baş verən morfonoloji dəyişikliklər qabaqcadan məlum

olmur və bu səbəblərə görə omonim və omoqraf olmuş, əslində isə müxtəlif kateqoriyaları təmsil edən morfemlər sistemdə bir qarmaqarışıqlıq, nizamsızlıq yaradır. Allomorfların bir morfemlə təmsil olunması bu qarmaqarışıqlıq və nizamsızlığı bir qədər də artırmış olardı.

Fərz edək ki, alqoritmdəki hər hansı P-operatorunun köməyi ilə biz söz-formadakı şəkilçilərin düzgün ardıcılığını əldə etdik. C-operatorunun köməyi ilə söz-formadan alınmış hər bir morfem üçün (X) müəyyən informasiya alınır: C(X'). Başqa sözlə, bu operatorun fəaliyyətindən sonra hər bir X morfemi üçün $\varphi(x')$ məlumatı alınır.

F(X) -elementar məlumatların dizyünksiyasıdır. Əgər a morfemi M_1, \dots, M_n -ə daxildirsə, $M_i = P(X_i)$ -dirsə, onda həmin morfemi $\varphi_a \cup_{i=1}^n C(X_i)$ məlumatı ilə müqayisə etmək olar (Фитиалов, 1961).

Burada M_i - söz-formanın P-operatorunun işindən sonra hissələrə ayrılmasından sonra alınmış morfemdir. Qeyd olunduğu kimi, söz-forma əsasdan və şəkilçilər zəncirindən ibarətdir:

$$X_i = \sum_{i=1}^n M_i$$

C-operatorunun işindən sonra alırıq:

$$C(X_i) = \cup_{i=1}^n C(M_i)$$

C-operatorunun işini konkret misallar əsasında nəzərdən keçirək. Fərz edək ki, «kitablarımızdan» söz-formasını təhlil etməliyik. P-operatorunun işindən sonra həmin söz-formanın düzgün olaraq morfemlərə bölünmüş şəklini alırıq:

$$P(\text{kitablarımızdan}) = (\text{kitab}) + (\text{lar}) + (\text{ımız}) + (\text{dan})$$

C-operatorunun tətbiqindən sonra alırıq:

$$C(X_i) = C(M_1) \cup C(M_2) \cup C(M_3) \cup C(M_4)$$

$$C(M_1) = a_1$$

$$C(M_2) = L_1$$

$$C(M_3) = \{n_1, n_2, n_3\}$$

$$C(M_4) = Z_1$$

Burada $L_1, n_1, n_2, n_3, Z_1, a_1$ - məlumatları aşağıdakıları bildirirlər:

$$a_1 - \text{isimr}$$

L_1 - cəm
 n_1 - mənsubiyyət
 n_2 - I şəxs
 n_3 - cəm
 Z_1 - çıxışlıq hal
 Nəticədə alırıq:

$$C(P(\text{kitablarımızdan})) = C(\text{kitab}) \cup C(\text{lar}) \cup C(\text{ımız}) \cup C(\text{dan}) = \{a_1\} \cup \{L_1\} \cup \{n_2\} \cup \{n_3\} \cup \{Z_1\}$$

Burada a-nın qrammatik mənadan başqa, həm də leksik mənası vardır.

Başqa bir misalı nəzərdən keçirək. Tutaq ki, bizə *dağın, satın, kitabın* söz-formaları verilmişdir. R-operatorundan sonra alırıq:

$P(\text{dağ-ın})$

$P(\text{sat-ın})$

$P(\text{kitab-ın})$

Dağ, sat, kitab əsaslarını müvafiq olaraq a_1, a_2, a_3 ilə, *-ın* şəkilçisini isə l ilə işarə etsək alırıq:

$$X_1 = P(\text{dağ-ın}) = \{a_1, L\}$$

$$X_2 = P(\text{sat-ın}) = \{a_2, L\}$$

$$X_3 = P(\text{kitab-ın}) = \{a_3, L\}$$

C-operatorundan sonra alırıq:

$$S(X_1) = S\{a_1\} \cup C\{Lx_1\}$$

$$C(X_2) = C\{a_2\} \cup C\{Lx_2\}$$

$$C(X_3) = C\{a_3\} \cup C\{Lx_3\}$$

Burada

$$C\{Lx_1\} = \{y_1\}$$

$$C\{Lx_2\} = \{k_1, k_2, k_3\}$$

$$C\{Lx_3\} = \{z_1, z_2, z_3\}$$

Göstərilən ifadələrdə $y_1, k_1, k_2, k_3, z_1, z_2, z_3$ aşağıdakı məzmununa malikdirlər:

U_1 - yiyəlik al, K_1 - əmr şəkli, K_2 - II şəxs, K_3 - tək, Z_2 - II şəxs, Z_3 - tək, Z_1 - mənsubiyyət

$$S(L) = C\{Lx_1\} \cup \{Lx_2\} \cup \{Lx_3\} = \{y_1\} \cup \{k_1, k_2, k_3\} \cup \{z_1, z_2, z_3\}$$

Buradan belə bir nəticə alınır ki, omonim şəkilçi olan -ın üç qrammatik məna ifadə edə bilər:

I. Yiyəlik al

II. Mənsubiyyət, II şəxs, tək

III. Əmr şəkli, II şəxs, tək

Misaldan göründüyü kimi, burada $K_2=Z_2$ (K_2, Z_2 - ikinci şəxsi göstərir), $K_3=Z_3$ (K_3, Z_3 -tək halı göstərir).

Əgər K_2, Z_2 -ni t_1 və K_3, Z_3 — t_2 ilə işarə etsək, onda yuxarıdakı ifadə belə şəkil alır:

$$S(L) = \{u_1\} \cup \{k_1, t_1, t_2\} \cup \{Z_1, t_1, t_2\} = \{y_1\} \cup \{(k_1 \cup Z_1)(t_1, t_2)\}$$

Göstərilən düstur göstərir ki, -ın morfemi omonim morfem kimi ya yiyəlik halı, ya da felin əmr şəklini (mənsubiyyəti - II şəxs, tək) göstərir.

6.10. Azərbaycan mətnlərinin formal morfoloji təhlil alqoritmi

Yuxarıda göstərilən leksik, morfoloji bazaya əsaslanaraq ümumi şəkildə morfoloji təhlil alqoritmlərini qurmaq olar. Bu alqoritmlərdə omonimliyin aradan qaldırılması ilə bağlı hissə verilməmişdir. Belə hesab olunur ki, istər leksik-morfoloji, istərsə də sintaktik səviyyələrdə omonimliyin aradan qaldırılması, polisemiyanın ləğvi məsələləri bütövlükdə sistemin semantik blokunun öhdəsinə düşür (Белогогов, Калинин, Поздняк, Хорошилов, Яфаева, 1983).

Morfoloji təhlil alqoritmlərinin konkret izahına keçməzdən əvvəl aşağıdakı şərtləri qəbul edək:

N_i - AL-də müəyyən qruplar üzrə yerləşdirilmiş əsasların uzunluğu; burada $i=1, \dots, k$.

N_i - I qrupdakı əsasların uzunluğu, N_2 - II qrupdakı əsasların uzunluğu və nəhayət, N_k -k qrupdakı 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 söz-formadakı əsasın, Z isə şəkilçilərin (və ya şəkilçinin) uzunluğuna işarədir.

$x_i=y+z$ mətndəki söz-formanın uzunluğu onun tərkibindəki əsasla sözdəyişdirici şəkilçilərin uzunluqları cəminə bərabərdir.

Bütün bu göstərilən şərtləri nəzərə alsaq, aşağıdakı kimi ümumi bir alqoritm qurmaq olar:

1. Söz-formanın təhlil üçün nəzərdə tutulmuş mətndən ayrılması.

2. Söz-formanın uzunluğunun təyini: əgər $x_i > N_i$, onda 5-ə, əks halda 3-ə keçid.

3. Söz-formanın AL-də əsaslar siyahısının müvafiq qrupu ilə müqayisəsi: əgər söz-forma tapılıbsa 4-ə, əks halda 6-ya keçid.

4. Tapılmış əsasın əlamətlərinin bu məqsədlə ayrılmış işçi sahəyə yazılması və 7-yə keçid.

5. $X_i = N_i$ -yə bərabər uzunluqda söz-formanın ayrılması və 3-ə keçid.

6. Söz-formanın uzunluğunun bir hərf azaldılması ($x_i - 1$): əgər $x=0$ -dirsə 13-ə, əks halda 3-ə keçid.

7. Qalığın varlığının yoxlanılması (Z): əgər $Z=0$ -dirsə 14-ə, əks halda 8-ə keçid.

8. Qalığın uzunluğunun müəyyənləşdirilməsi: əgər $Z > n_i$ -dirsə 12-yə, əks halda 9-a keçid.

9. Qalığın müvafiq şəkilçilər qrupu ilə müqayisəsi: əgər şəkilçi tapılıbsa 10-a, əks halda 11-ə keçid.

10. Şkilçinin əlamətlərinin əsasın əlamətlərindən sonra yazılması və 7-yə keçid.

11. Qalığın uzunluğunun Z bir hərf azaldılması ($Z - 1$): əgər $Z=0$ -dirsə 9-a, əks halda 13-ə keçid.

12. n_i - uzunluğunda qalığın ayrılması və 9-a keçid.

13. Söz-formanın naməlum sözlər siyahısına yazılması və 15-ə keçid.

14. Söz-formanın onun qrammatik əlamətləri ilə çapı və 15-ə keçid.

15. Mətnin sonu işarəsinin varlığının yoxlanması: əgər son işarəsi varsa, 16-ya, əks halda 1-ə keçid.

16. Təhlili qurtarmalı (Sxem 8).

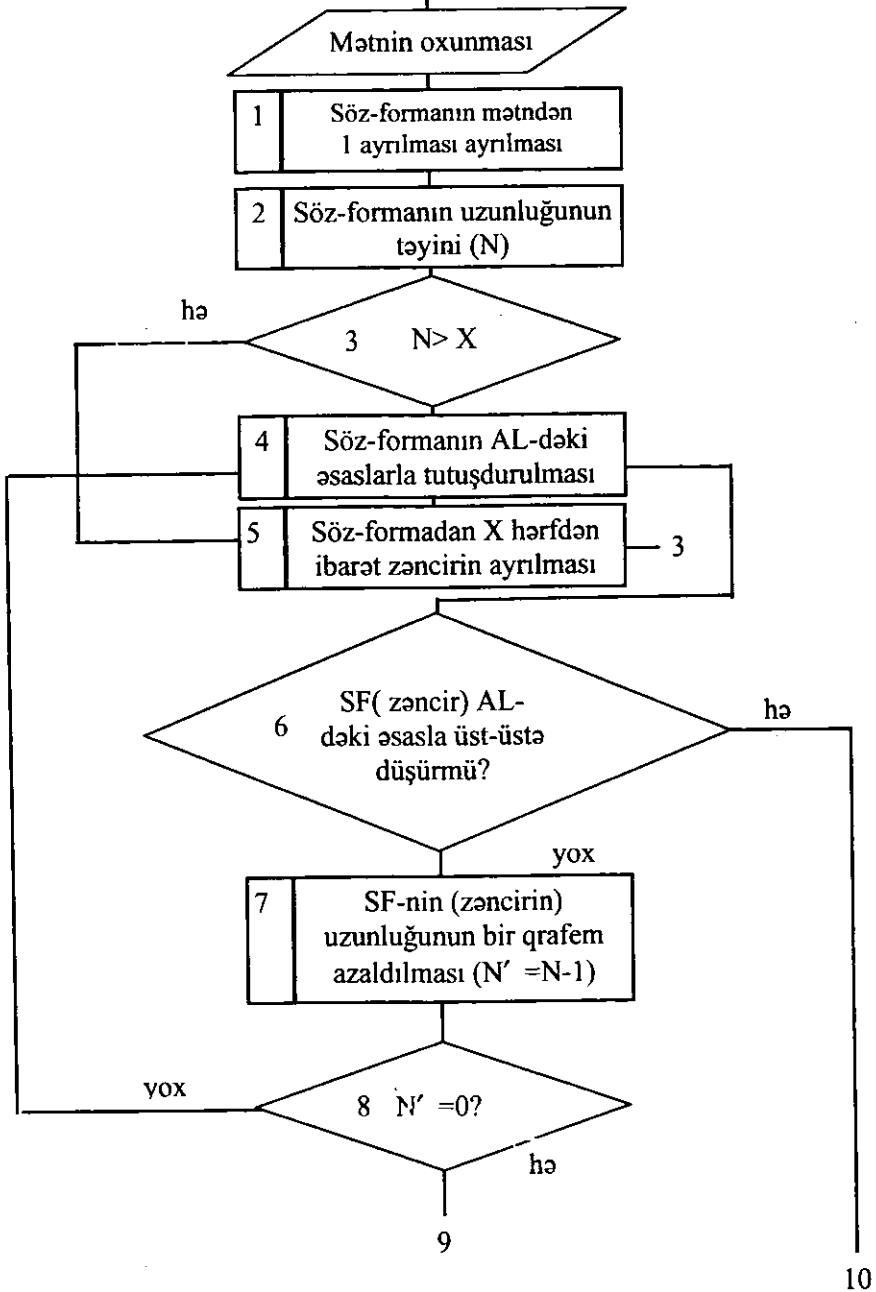
Alqoritm haqqında tam əyani təsəvvür əldə etmək üçün «məktəblərimizdəndir» söz-formasının kompüterdə təhlil prosesini izləyək (alqoritm A.A.Piotrovskaya tərəfindən proqramlaşdırılmışdır).

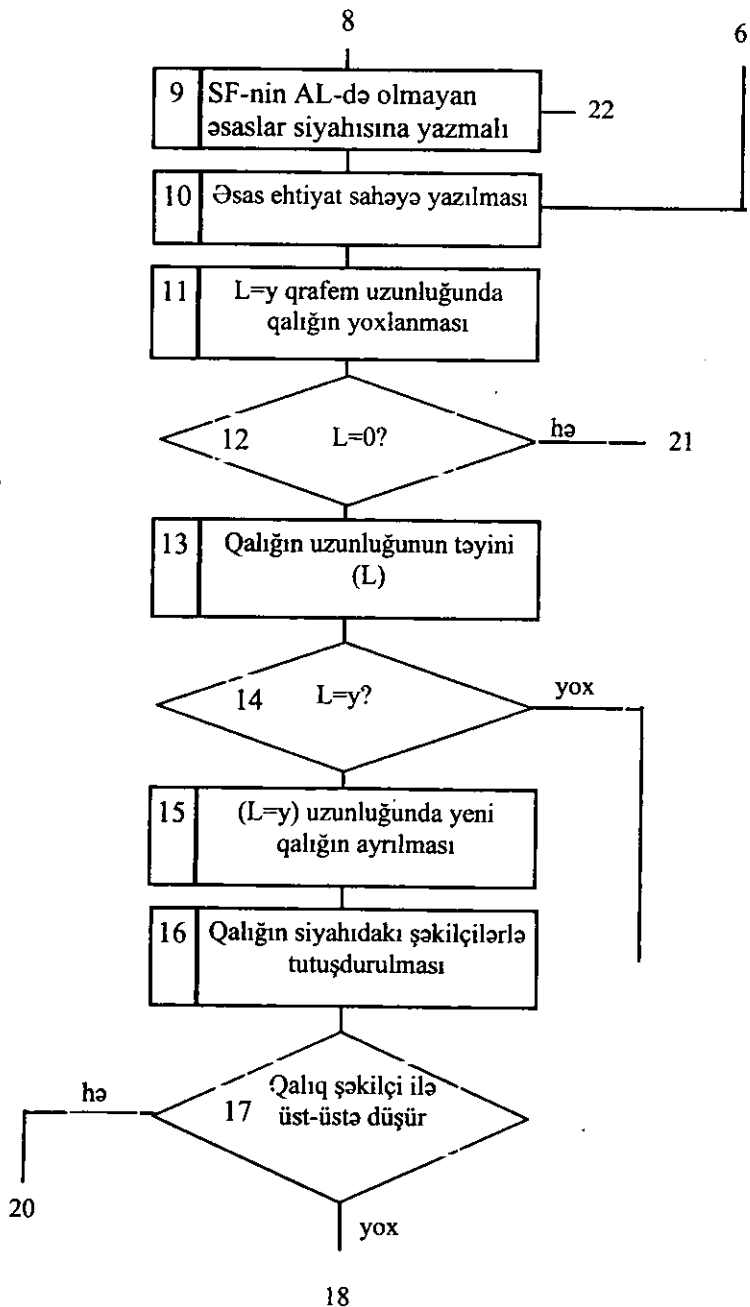
Mətndən ayrılmış bu söz-formanın uzunluğunu tapırıq: $x_i=19$. Əsaslar AL-də isə əsasların maksimal uzunluğu 16 hərfdir. $N_i=16$. Deməli, $x_i>N_i$ (yəni $19>16$). Alqoritmə görə, bu şərt daxilində 5-ci addıma keçməliyik. 5-ə görə $x_i=N_i$ -yə bərabər sayda hərf saxlamaqla söz-formanın uzunluğunu azaldırıq: $N_i=16$ olduğu üçün $x_i=16$ hərf saxlayırıq, qalığı isə yadda saxlayırıq: «məktəblərimizdən». Bu söz-formanı 16 hərfdən ibarət əsaslar yerləşən qrupla müqayisə edirik. Şərtə görə, əgər söz-forma orada tapılsaydı, 4-ə, əks halda 6-ya keçməliyik. Təbiidir ki, bu söz-forma bir lüğət vahidi kimi əsaslar AL-ində təsbit oluna bilməz, çünki onun tərkibində əsas + sözdəyişdirici şəkilçi var. Ona görə də şərtə əsasən 6-ya keçirik.

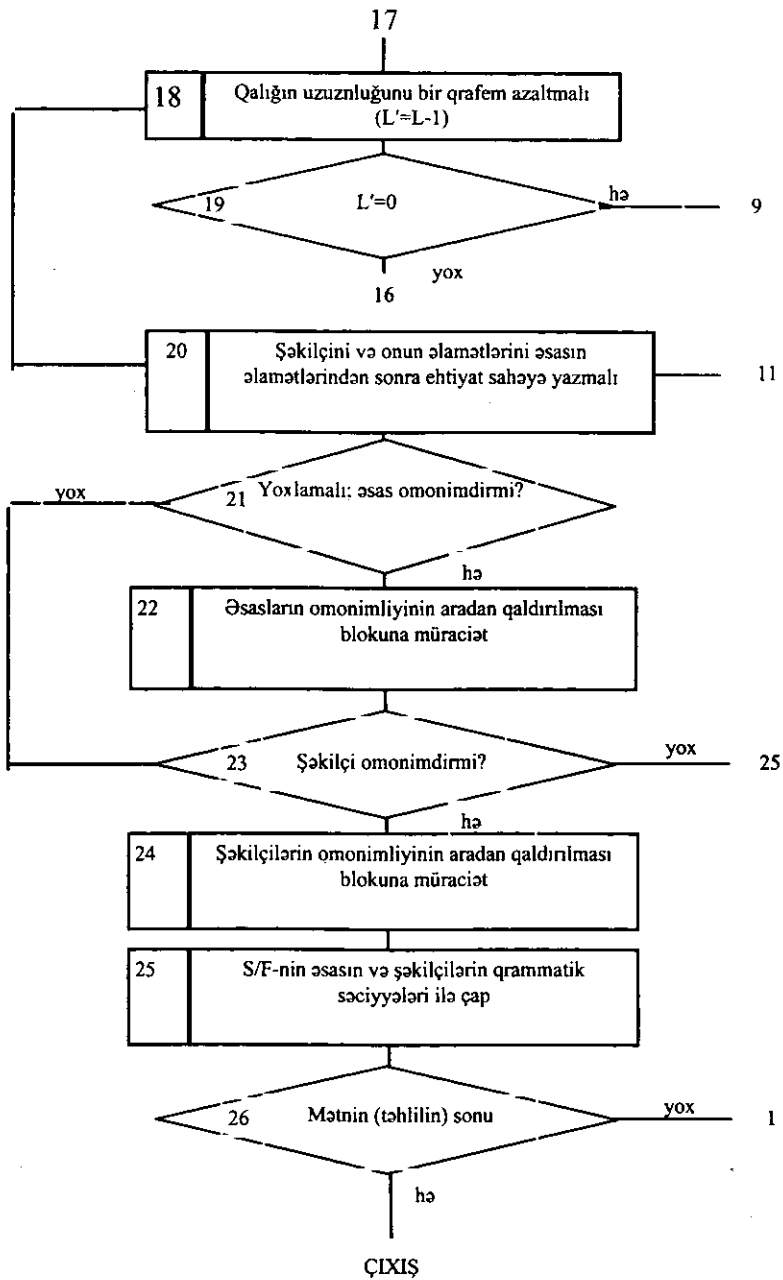
Altıncı addımın şərtinə görə, söz-formanın uzunluğunu bir hərf azaldırıq. Əgər $X_i=0$ olsa, 13-ə, əks halda 3-ə keçməliyik. X_i sıfırdan fərqli olduğu üçün 3-ə keçirik. “Məktəblərimizdə” ifadəsini 15 hərfdən ibarət əsaslar yerləşdirilmiş qrupla müqayisə edirik. Təbiidir ki, bu ifadə də orada təsbit oluna bilməz. Dövri təkrarlar nəticəsində proses (hər dəfə söz-formanın sonundan bir hərf ayırmaqla, onun əsasları AL-nin müvafiq qrupu ilə müqayisəsi) o vaxta qədər davam edir ki, «məktəb» əsası beş hərfdən ibarət əsasların yerləşdiyi qrupda tapılır. Tapılmış «məktəb» sözünün tərcüməsi, qrammatik əlamətləri qabaqcadan ayrılmış işçi sahəyə yazılır. Sonra qalığın (şəkilçilərin) olub-olmaması müəyyənləşdirilir. Qalığın varlığı aşkar olunduqdan sonra 8-ci addımın şərti yerinə yetirilir. Əgər qalığın uzunluğu $Z>n_i$ -dirsə (yəni

şəkilçilər siyahısında yerləşən şəkilçilərin maksimal uzunluğundan böyükdürsə), 12-yə, əks halda 9-a keçirik. Bizim misalda qalığın (*lərimizdəndir*) uzunluğu şəkilçilərin maksimal uzunluğundan (doqquz hərf) böyükdür. Ona görə də, 12-ci addımın şərtlərini yerinə yetiririk: şəkilçilərin maksimal uzunluğuna bərabər uzunluqda qalıq ayırırıq, qalan hissəni yaddaşa yerləşdiririk: $Z=n_i=\underline{\text{lərimizdə}}$. Alınmış hərflər zənciri müstəqil şəkilçi deyil, şəkilçilər birləşməsi olduğu üçün doqquz hərfdən ibarət olan şəkilçilərin yerləşdiyi qrupda təsbit oluna bilməz. Ona görə də, doqquzuncu addımın şərtinə əsasən 11-ə keçirik. On birinci addıma görə, həmin hərflər zəncirinin sonundan bir hərf ayrılır və alınmış ifadə «lərimizdə» ona müvafiq qrupla- səkkiz hərfdən ibarət şəkilçilər qrupu ilə müqayisə olunur. Bu qalıq orada da təmsil oluna bilməz. Proses o vaxta qədər davam edir ki, -lər şəkilçisi üç hərfdən ibarət şəkilçilər qrupunda tapılır və onun qrammatik əlamətləri «məktəb» sözünün qrammatik əlamətlərinə əlavə olunur. Sonra yenə yeddinci addıma qayıdırıq və qalığın qalıb-qalmamasını yoxlayırıq. Məlum olur ki, qalıq var: *imizdəndir*. Dövri təkrarların köməyi ilə hər dəfə qalığın sonundan bir hərf ayıraraq onu şəkilçilər siyahısının müvafiq qrupu ilə müqayisə etməklə söz-formada olan bütün şəkilçilərin qrammatik səciyyələri işçi sahəyə yazılır. Beləliklə, təhlil prosesi nəticəsində *məktəblərimizdəndir* sözü düzgün olaraq kök və şəkilçilərə bölünür, ayrı-ayrı şəkilçilər bir-birindən səhsiz ayrılır və onların qrammatik səciyyələri əldə olunur. Əgər təhlil etdiyimiz söz-formada omonim söz və şəkilçi müşahidə olunsaydı, onda biz semantik bloka müraciət etməli olardıq. Semantik blok istər morfoloji səviyyədə, istərsə də daha yuxarı dil səviyyələrində omonimliyin aradan qaldırılması və polisemiyanın optimal şəkildə həlli problemləri ilə bağlı alqoritmlər sistemi ilə təchiz olunmuşdur. Əlbəttə, omonimlik və polisemiya qismən morfoloji səviyyədə də aradan qaldırıla bilər. Bunun üçün morfoloji blokda əlavə alqoritmlərin olması zəruridir. Morfoloji səviyyədə həlli qeyri-mümkün olan problemlər semantik blokda yeni səviyyədə araşdırılır. Təcrübə göstərir ki, omonimlik və polisemiya semantik problemlər çərçivəsində daha tam, ertərəfli və sistemli tədqiq oluna bilər (Коптилов, 1962).

Giriş







Sxem 8. Mətnlərin formal morfoloji təhlilinin blok sxemi

Söz-forma;
“çağırduğumuzlardansınız mı”

N=25

Əsas;

çağır AL-də tapıldı

Çağır Fel

Şəkilçilər;

dığ	siyahıda tapıldı
dığ	feli sifət, şühudi keçmiş zaman
ımız	siyahıda tapıldı
ımız	mənsubiyyət, I şəxs, cəm
lar	siyahıda tapıldı
lar	cəm
dan	siyahıda tapıldı
dan	çıxışlıq hal
sınız	siyahıda tapıldı
sınız	II şəxs, cəm, xəbərlik
mı	siyahıda tapıldı
mı	sual
mı	sonuncu şəkilçi

Şəkil 2. Söz-formanın avtomatik (formal) morfoloji təhlili

VII FƏSİL MƏTNİN FORMAL TƏHLİLİ SİSTEMİNİN MORFONOLOJİ BLOKU

7.1. Morfonoloji blokun funksiyası

Təbii dil materialının formal təhlil obyektı olması bir çox morfoloji, fonetik, morfonoloji hadisələrin qanunauyğunluqlarının, əsl təbiətinin hərtərəfli araşdırılması zərurətini doğurmuşdur. Əgər morfemlər, onların sıralanma xüsusiyyətləri, sözdüzəltmə və sözdəyişmə paradıqması morfolojiyada öyrənilsə, dilin iki səviyyəsi – fonetika və morfolojiya arasında aralıq mövqedə duran morfolojiyada fonetik vahidlərlə morfoloji vahidlərin qarşılıqlı əlaqəsi öyrənilir. (Bax. Cəlilov, 1985; 1988). Dilin morfoloji sisteminin ən mühüm elementləri olan morfemlərin fonetik tərkibi və onların müxtəlif fonetik dəyişmələrə məruz qalması (səsartımı, səsdüşümü, karlaşma, cingiltiəşmə, ixtisar və s.) morfonoloji səciyyə daşıyır. Formal təhlil və sintez sistemlərində bu xüsusiyyətlərin nəzərə alınmaması bütövlükdə sistemin mümkünlüyünü şübhə altına alır.

Ümumiyyətlə, mətnlərin kompüterlərdə işlənməsi prosesində morfonoloji xüsusiyyətlər də qabarıq müşahidə olunur və bu baxımdan müəyyən qayda-qanun, nizamlı sistem yaradılması daha böyük əhəmiyyət kəsb edir. Azərbaycan mətnlərinin formal işlənməsi sistemlərində müşahidə olunan morfonoloji dəyişiklikləri aşağıdakı kimi qruplaşdırmaq olar:

- söz kökündə baş verən morfonoloji dəyişmələr (səsartımı, səsdüşümü);
- söz kökü ilə sözdüzəldici şəkilçi sərhədində morfonoloji dəyişmələr;
- söz kökü ilə sözdəyişdirici şəkilçi sərhədində morfonoloji dəyişmələr;
- sözdüzəldici şəkilçi ilə sözdəyişdirici sərhədində morfonoloji dəyişmələr;
- sözdəyişdirici şəkilçilər sərhədində morfonoloji dəyişmələr.

Göstərilən qruplara aid istisna sözlər siyahı şəklində göstərilir ki, lazım gəldikdə alqoritmdə onlara müraciət etmək imkanı nəzərə alınsın.

7.2. Söz kökündə morfonoloji dəyişmələr

Azərbaycan dilində iki və ya çoxecalı sözlər mənsubiyyət və hal şəkilçilərini qəbul edərkən ikinci hecadakı (qapalı heca) qapalı saitlərin (-ı,-i,-u,-ü) -düşməsi hadisəsi baş verir. *Şəkil, sinif* sözlərinin mənsubiyyət və hal şəkilçisi qəbul edərkən uğradığı morfonoloji dəyişikliyi nəzərdən keçirək:

Cədvəl 26. Mənsubiyyət şəkilçili sözlərdə səs düşümü

Şəxs	Tək	Cəm
I	şəklim, sinfim	şəklimiz, sinfimiz
II	şəklin, sinfin	şəkliniz, sinfiniz
III	şəkli, sinfi	şəkli, sinfi

Cədvəl 27. İsmın hal şəkilçisini qəbul etmiş sözlərdə səs düşümü

İsmın alları	Hal şəkilçisi qəbul etmiş sözlər
Adlıq	şəkil, sinif
Yiyəlik	şəklin, sinfin
Yünlük	şəklə, sinfə
Təsirlilik	şəkli, sinfi
Yerlik	şəkildə, sinifdə
Çıxışlıq	şəklindən, sinfindən

Göstərmək lazımdır ki, səsdüşümü daha çox alınma sözlərdə özünü göstərir. Qərribə burasıdır ki, bu qəbildən olan sözlərin bir çoxu əslində alındığı dillərdə, məsələn, «isim», «ömür», «şəkil» formasında deyil, «is/i/m», «öm/ü/r», «şək/i/l» və s. kimi işlənilirlər. Azərbaycan dilinə keçdikdən sonra onlar dilimizin fo-

notaktikasına uyğunlaşdırılmış, sait yığımının qarşısı sait artımı vasitəsilə aradan qaldırılmışdır (bu barədə bax: Cəlilov, 1988, 55).

Qeyd etmək lazımdır ki, analogi morfonoloji hadisə digər türk dillərində də geniş müşahidə olunur (Коронов, 1958; Поцелуевский, 1975).

Assimlyasiya, səsdüşümü və digər morfonoloji dəyişmələr türk dillərinin sırf ciddi aqlütinativ dil olması ilə bərabər, onlarda aqlütinativ-fuzional, eləcə də flektiv dillərə məxsus fleksiyanın müşahidə olunduğunu göstərir. Ümumiyyətlə, dünya dillərini bu və ya digər xüsusiyyətinə görə bir-birindən tam təcrid edib ayırmaq olmaz. Flektiv dillərdə də aqlütinativ dil qruplarına xas bir çox xüsusiyyətlər mövcuddur. Göstərilən morfonoloji xüsusiyyət alqoritmlərdə müəyyən qanunauyğunluq şəklində təsbit oluna bilərdi. Lakin məsələ orasındadır ki, bu qəbildən olan bütün sözlərdə analogi səsdüşümü hadisəsi baş vermir. Məsələn, əgər «sətin» sözü bu dəyişikliyə məruz qalırsa, eyni quruluşa və sonluğa malik «əsir» və «dəmir» sözlərində ikinci hecadakı qapalı sait düşür:

dəmirim, dəmirin, dəmiri əsirim, əsirin əsiri

dəmirimiz, dəmiriniz, dəmiri əsirimiz, əsiriniz, əsiri

Söz kökündə belə morfonoloji dəyişmələrin Azərbaycan dili üçün mühüm səciyyəvi xüsusiyyət sayılmadığını və belə sözlərin kəmiyyətə azlığını nəzərə alıb mənsubiyyət və ismin hal şəkilçilərini (yiyəlik, yönlük, təsirlik) qəbul edərkən sözün ikinci hecasındakı qapalı saitin düşməsi hadisəsinə məruz qalan sözlər ayrıca siyahı şəklində verilmişdir. AL-lərdə bu sözlərin hər iki variantı təsbit olunmuşdur: ağıl/ağl, ağız/ağz və s. Göstərilmiş variantlar eyni qrammatik səciyyə daşıyırlar. Verilmiş siyahıda Azərbaycan dilində bütün bu qəbildən olan sözlərin tam şəkildə verildiyi iddia olunmadığı üçün sonradan ora müşahidə edilmiş bu qəbildən sözlər əlavə oluna bilər. Bunun formal təhlil sisteminin işinə mənfi təsiri yoxdur.

*Cədvəl 28. Mənsubiyyət və hal şəkilçilərini (yivəlik, yön-
lük, təsirlilik) qəbul edərəkən sözün ikinci hecasındakı qapalı sait
düşməsi hadisəsinə məruz qalan sözlərin siyahısı*

1. abır	20. könül
2. ağıl	21. misir
3. ağız	22. nəsil
4. alın (isim)	23. oğul
5. bağır (isim)	24. ömür
6. boyun	25. səbir
7. burun (isim)	26. sətir
8. qayın	27. sinif
9. qarın	28. sıfır
10. qəbir	29. üfüq
11. qədir	30. fəsil
12. qisim	31. fikir
13. qoyun (ağuş- isim)	32. xeyir
14. əyin (isim)	33. xətir
15. əsil	34. çətir
16. ətir	35. çiyin
17. zəhin	36. cisim
18. isim	37. şəkil
19. gəlin	

*7.3. Söz kökü ilə sözdüzəldici şəkilçi sərhədində morfo-
noloji dəyişmələr*

Azərbaycan dilində söz kökü ilə sözdüzəldici şəkilçi sərhədində morfonoloji dəyişikliklər təhlil prosesində bir qisim «anlaşılmazlıqların» aradan qaldırılmasına imkan verir.

Göstərilən morfonoloji dəyişmələr söz kökü ilə sözdəyişdirici şəkilçi sərhədində baş verən dəyişmələrə nisbətən az müşahidə olunsa da, formal təhlil sistemlərində onların nəzərə alınması vacibdir. Azərbaycan dilində söz kökü ilə sözdüzəldici şəkilçi sərhədində baş verən və söz kökünün morfonoloji dəyişikliyə uğraması ilə nəticələnən prosesi aşağıdakı kimi qruplaşdırmaq olar:

a) çağır/maq/, qovur/maq/, devir/mək/, yoğur/maq/, sovur/maq/, sivr/mək/ və s. bu qəbildən olan fel kökləri felin məchul növ şəkilçilərini qəbul etdikdə ikinci hecadakı qapalı saitlərin düşməsi hadisəsi baş verir: çağr-ıl-/maq/, qovr-ul-/maq/, devr-il-/mək/, çevr-il-/mək/ və s. Təhlil prosesində nəzərə alınmalıdır ki, çağır/çağr, qovur/qovr, devir/devr, yoğur/yoğr, sovur/sovr, sivr/sivr, çevir/çevr, eyni fel kökləridir.

b) Azərbaycan dilində bir çox söz kökləri -i, -vi sözdüzəldici şəkilçisini qəbul etdikdə söz kökündə incələşmə baş verir və həmin sözlər Azərbaycan dilinə uyğunlaşmış olur. Göstərmək lazımdır ki, burada ciddi bir qanunauyğunluq və sistem yoxdur. Bu sözlərin əksəriyyətində söz kökləri alındığı dillərdən fərqli şəkildə dilimizə keçmiş, sonra yenidən morfonoloji dəyişikliyə uğrayaraq mənşəyinə yaxınlaşmışdır:

dünya- vi- dünyəvi, kimya- vi- kimyəvi,
xalq- i- xəlqi (bundan da xəlqilik) və s.

Misallardan görüldüyü kimi, bu sözlərdə söz kökü (dünya, kimya, xalq) -i, -vi şəkilçilərinin ahənginə uyğun olaraq morfonoloji dəyişmələrə məruz qalmışdır. Azərbaycan dilinin əks lüğətinin materialları əsasında belə misalların sayını bir qədər də artırmaq olardı. Lakin tədqiqatdan məqsəd bütün bu qəbildən olan sözləri sadalamaq deyil, onların təbiətini, dəyişmə mexanizmini izləməkdir. Bəzən alınma söz köklərindəki -ya, -yyat, -a, -ət və s. hissəcikləri atmaqla -i, -vi şəkilçilərinin yaratdığı məzmun alınır: coğrafiya - coğrafi, riyaziyyat - riyazi, fizika - fiziki, iqtisadiyyat - iqtisadi, siyasət - siyasi.

Göstərilən xüsusiyyətlərin hərtərəfli tədqiqi və təhlil alqoritmlərində nəzərə alınması qarşıya çıxacaq çətinliklərin, yanlış yozumların aradan qaldırılması imkanı yaratmış olardı.

Söz kökləri ilə sözdüzəldici şəkilçi sərhədində baş verən morfonoloji dəyişmələr ilə bağlı -ıı (-icı, -ucu, -ücü) şəkilçilərinin bəzi fel köklərinə təsirinin izlənilməsi də maraqlıdır. Aşağıdakılar göstərir ki, sonu «t» samiti ilə bitən fel kökləri saitlə başlanan şəkilçilər qəbul etdikdə t-d cingilləşməsi baş verir: böyüt – böyüd-ücü, dağıt – dağıd-ıı, qurut – qurud-ucu, ərit – ərid-

ici və s. Belə morfonoloji dəyişmələr söz kökü ilə sözdəyişdirici şəkilçi sərhədində daha çox baş verdiyi üçün bu məsələ tədqiqatda geniş araşdırılacaq.

V.Radlov köklə şəkilçinin bir-birinə qarışaraq birləşməsi və bölünmə xüsusiyyətini itirməsini hiperaqlütinasiya adlandırır. Şəkilçilərin dinamik təbiəti türk dillərində təhlil zamanı kök və şəkilçi sərhədinin müəyyənləşdirilməsini çətinləşdirir. Bir çox qeyri-məhsuldar və az məhsuldar şəkilçilərin, eləcə də bölünməz mürəkkəb əsasların formalaşması və inkişafı ilə əlaqədar mübahisəli məsələlər də məhz bunun nəticəsidir (Коронов, 1971).

7.4. Söz kökü ilə sözdəyişdirici şəkilçi sərhədində morfonoloji dəyişmələr

Söz kökü ilə sözdüzəldici şəkilçi sərhədində sonu «t» samiti ilə bitən fellərdən sonra saitlə başlanan şəkilçi işləndikdə «t»-nin intervokal vəziyyətdə özünün «d» korrelyatı ilə əvəzlənməsini göstərmişdik. Həmin qəbildən olan fellər saitlə başlanan sözdəyişdirici şəkilçilərlə işləndikdə də analoji morfonoloji dəyişmələr baş verir:

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

Fellərin sözdəyişdirici şəkilçilərlə işlənmə tezliklərinin çox böyük olduğunu nəzərə alsaq, bu qəbildən olan proseslərə xüsusi diqqət yetirilməsinin zəruriliyi bir daa aşkar olar.

Söz kökü ilə sözdəyişdirici şəkilçi sərhədində morfonoloji dəyişmələr - daha çox sonu -q və -k samiti ilə bitən çoxecalı sözlərdə müşahidə olunur. Belə sözlərə saitlə başlanan şəkilçi əlavə olunduqda q-ğ, k-y əvəzlənməsi baş verir;

çiçək - çiçəyim, çiçəyə; budaq - budağım, budağa; ayaq - ayağım, ayağa, torpaq - torpağım, torpağa.

Sonu -oloq və -ik şəkilçiləri ilə bitən Avropa mənşəli və -q, iq və -ik ilə bitən ərəb mənşəli sözlər istisna təşkil edir: *bioloq, texnoloq, fizik, elektrik* (Avropa mənşəli sözlər); *müstəqil, şərik, məntiq* (ərəb mənşəli sözlər) və s.

Saitlə bitən şəkilçi qəbul etdikdə *k*-y və *q-ğ* əvəzlənməsinə məruz qalmayan istisna sözlər ayrıca siyahı şəklində verilir.

Belə sözlərin təhlili zamanı morfonoloji rekonstruksiya prinsipindən istifadə olunur və söz kökünün leksik vahid kimi ənənəvi lüğətlərdə təsbit olunmuş variantı bərpa olunur.

Fərz edək ki, bizdən «torpağımızın» söz-formasının təhlili tələb olunur. Morfoloji blokdakı alqoritmlərin köməyi ilə söz-forma hissələrinə ayrılır və hər bir ayrılmış vahidin qrammatik səciyyələri göstərilir: **Torpağ-ımız-ın**

AL-də «torpağ» sözünün qrammatik səciyyələri «torpaq» sözü ilə eynidir. Q - Ğ əvəzlənməsinin şərtlərini əsas tutaraq söz kökünün ilkin variantını - «torpağ» variantını əldə edib, qrammatik səciyyələri ilə birlikdə qabaqcadan nəzərdə tutulmuş işçi sahəyə yazırıq. Analoji olaraq səsdüşümü, səsartımı və digər morfonoloji deformasiyalara məruz qalmış söz köklərinin də rekonstruksiya mümkündür. Söz kökləri ilə yanaşı şəkilçi formalarının da ilkin formaları alına bilər.

Cədvəl 29. Saitlə bitən şəkilçi qəbul etdikdə k-y əvəzlənməsinə məruz qalmayan istisna sözlərin siyahısı

1. abrek	28. konyak	55. skeptik
2. aqnostik	29. kubok	56. statistik
3. dinamik	30. mayak	57. taktik
4. akademik	31. malik	58. tacik
5. akveduk	32. materik	59. tematik
6. aptek	33. mənşevik	60. texnik
7. barak	34. mexanik	61. təbrik
8. botanik	35. məslək	62. tədarük
9. bərk	36. məhək	63. təhrik
10. naçalnik	37. narkotik	64. tiryək
11. qrafik	38. opriçnik	65. tragik
12. daşnak	39. optik	66. uryadnik
13. dialektik	40. özbək	67. utopik
14. eklektik	41. parik	68. fabrik
15. elastik	42. plastik	69. faktik

16. elektrik	43. platonik	70. fanatik
17. empirik	44. podpolkovnik	71. fantastik
18. julik	45. poetik	72. fizik
19. zirək	46. politexnik	73. fonetik
20. əmlak	47. polkovnik	74. herik
21. zontik	48. poruçik	75. həlak
22. iqrek	49. praktik	76. çaynik
23. iştirak	50. praporşik	77. çevik
24. kazak	51. ritorik	78. şəkkak
25. kazbek	52. rolik	79. şarik
26. kauçuk	53. romantik	80. şirmik
27. kinomexanik	54. sinoptik	81. şlak
		82. ştok

Cədvəl 30. Saitlə başlanan şəkilçi qəbul etdikdə q-ğ cingil-tib şəmsinə məruz qalmayan istisna sözlərin siyahısı

1. aysberq	28. mioloq	55. rövneq
2. akvalanq	29. mikrobioloq	56. sabiq
3. arxeoloq	30. monoloq	57. semasioloq
4. aşiq	31. müvafiq	58. semitoloq
5. bioloq	32. müstəntiq	59. seroloq
6. vərəq	33. müttəfiq	60. sosioloq
7. dialektoloq	34. müşfiq	61. speleoloq
8. dialoq	35. natiq	62. stamotoloq
9. entomoloq	36. nevroloq	63. teoloq
10. əxlaq	37. nevropotoloq	64. texnoloq
11. zooloq	38. nifaq	65. təsdiq
12. ideoloq	39. odontoloq	66. tətbiq
13. ilhaq	40. okeanoloq	67. tədqiq
14. istintaq	41. onkoloq	68. təşviq
15. ixtioloq	42. ontoloq	69. tifaq
16. kardioloq	43. ornitoloq	70. tokoloq
17. geoloq	44. oftalmoloq	71. toksikoloq
18. ginekoloq	45. paleontoloq	72. türkoloq
19. layiq	46. parazitoloq	73. uroloq

20.larinqoloq	47.pataloq	74.üfûq
21.leksikoloq	48.pedaqoq	75.farmakoloq
22.meteoroloq	49.pedoloq	76.fizioloq
23.metroloq	50.piroq	77.xaliq
24.metodoloq	51.proloq	78.xələyiç
25.məntiq	52.psixoloq	79.hersoq
26.məşriq	53.radioloq	80.histoloq
27.məşuq	54.rentgenoloq	81.hüquq
		82.şəfəq

7.5. Sözdüzəldici şəkilçi ilə sözdəyişdirici şəkilçi sərhədində morfonoloji dəyişmələr

Təkcə söz kökləri deyil, müəyyən sözdüzəldici şəkilçi qəbul etmiş əsaslarla saitle başlanan sözdəyişdirici şəkilçi sərhədində analoji morfonoloji dəyişmələr baş verə bilər. Azərbaycan dilinin əks lüğətində (sözlüyündə) belə əsasların tam siyahısını əldə etmək mümkündür. Daha yığcam olmaq üçün əsasları deyil, sonu q-ğ ilə bitən sözdüzəldici şəkilçiləri nəzərdən keçirək:

1. -ıq (-ik, -uk, -ük, -q, -k)

qırıq, kəsik, buruq, sükkük.

Bu sözlər mənsubiyyət və hal şəkilçiləri qəbul etdikdə q - ğ, k - y əvəzlənməsinə məruz qalır: taxtanın qırığ-ı, dəmirin kəsiy-i, saçın buruğ-u, paltarın söküy-ü. Substantivləşdikdə h al şəkilçisi qəbul edirlər:

qırıq - qırığın, qırığa qırığı və s.

2. -laq (-lək)

qışlaq, işlək

qışlağım, qışlağın, qışlağa və s.

işləyə, işləyin, işləyim və s.

3. -lıq (-lik, -luq, -lük)

adamlıq, qohumluq, gözəllik, böyüklük

adamlığı, adamlığı, adamlığı və s.

gözəlliyim, gözəlliyin, gözəlliyi və s.

qohumluğum, qohumluğun, qohumluğu və s.

-maq (-mək)

qazmaq, yemək

qazmağım, qazmağın, qazmağı və s.

yeməyim, yeməyin, yeməyi və s.

4. -çılıq (-çilik, -çuluq, -çülük)

mehribançılıq, cütçülük, qoyunçuluq, əkinçilik

mehribançılığım, mehribançılığın, mehribançılığa və s.

cütçülüyüm, cütçülüyün, cütçülüyə və s.

qoyunçuluğum, qoyunçuluğun, qoyunçuluğa və s.

əkinçiliyim, əkinçiliyin, əkinçiliyə və s.

5. -caq (-cək)

yelləncək

yelləncəyim, yelləncəyin, yelləncəyə, yelləncəyi və s.

6. -cıq (-cik, -cuq, -cük)

almacıq, təpəcik, qutucuq, gözcük

almacığım, almacığa almacığı və s.

təpəciyim, təpəciyə təpəciyi və s.

qutucuğum, qutucuğa, qutucuğu və s.

gözcüyü, gözcüyə, güzcüyü və s.

7. -aq (-ək, -q, -k)

sorağım, sorağın, sorağı və çökəyim, çökəyi, çökəyə və s.

dayağım, dayağın, dayağa

ələyim, ələyin, ələyə və s.

8. -anaq (-ənək)

boğanaq, döyənək

boğanağa, boğanağı, boğanağın və s.

döyənəyim, döyənəyə, döyənəyi və s.

9. -acaq (-əcək)

dayanacaq söykənəcək

dayanacağım, dayanacağa, dayanacağı və s.

söykənənəcəyim, söykənəcəyə, söykənəcəyi və s.

10. -dıq (-dik, -duq, -dük) feli sifət şəkilçisi ayrılıqda deyil, şəkilçilərlə işənir və sait əhatəsində -q, -k samitləri dəyişməyə məruz qalır:

aldığım, aldığı, aldığı və s.

gəzdiyim, gəzdiyin, gəzdiyi və s.

oxuduğum, oxuduğun, oxuduğu və s.
gördüyüm, gördüyün, gördüyü və s.

7.6. Sözdəyişdirici şəkilçilər sərhədində morfonoloji dəyişmələr

Məlum olduğu kimi, sözdəyişdirici şəkilçilər təkcə söz kökündən və əsasdan sonra deyil, həm də sözdəyişdirici şəkilçidən sonra işlənə bilər. Sonu -q və -k samitləri ilə bitən samitlərdən sonra saitlə başlanan sözdəyişdirici şəkilçi gəldikdə məlum morfonoloji dəyişmələr baş verir.

Sonu -q və -k samitləri ilə bitən bütün şəkilçilər deyil, yalnız -q və -k samitləri ilə bitən və sonra saitlə başlanan sözdəyişdirici şəkilçi qəbul etdikdə morfonoloji dəyişmələrə məruz qalan şəkilçilər tədqiqata cəlb olunur. Məhz göstərilənlərə görə də, öz-lərindən sonra şəkilçi qəbul etmədikləri üçün -araq (-ərək, -yaraq, -yəmək), -caq (-cək), -aq (-ək, -yaq, -yək), -ıq (-ik, -uq, -ük), -yıq (-yik, -yuq, -yük) bu halda araşdırılmır. Aşağıdakı sözdəyişdirici şəkilçilərdə morfonoloji dəyişmələr özünü göstərir:

-acaq (-əcək, -yacaq, -yəmək). Qəti gələcək zamanı şəkilçisi şəxslərə görə təsrifləndikdə I şəxsin tək və cəmində iki sait arasında düşmüş -q və -k samitləri uyğun olaraq -ğ və -y samitinə çevrilir:

yazacağam, yazacağım
gələcəyəm, gələcəyik
oxuyacağam, oxuyacağım
böyüyəcəyəm, böyüyəcəyik

-maq, -mək. Felin məsdər şəkilçisi dilimizdə ən çox işlənən şəkilçilərdəndir. Məlum olduğu kimi, lüğətlərdə fel çox vaxt məsdər formasında təsbit olunur. Tədqiqatda formal təhlil sistemi alqoritmlərinin işinə müvafiq olaraq fel məsdər şəkilçisiz, II şəxsin təkində (at, yaz, bax, və s.) verilmişdir.

Felin məsdər formasından sonra hal və şəxs şəkilçiləri işlənə bildiyi üçün analoji xüsusiyyətləri burada da görmək olar:

baxmağa, baxmağı, baxmağım, baxmağın və s.
getməyə, getməyi, getməyim, getməyin və s.

İki sözdəyişdirici şəkilçi sərhədində baş verən morfonoloji dəyişmələrin araşdırılması və alqoritmlərdə bu xüsusiyyətlərin nəzərə alınması təhlil prosesi üçün mühüm əhəmiyyət kəsb edir. Şübhəsiz, morfonoloji dəyişmələr bizim sadaladığımız məsələlərlə bitmir və daha geniş fono-morfoloji hadisələr kompleksini əhatə edir. Formal təhlil sisteminin morfonoloji blokunda bu qəbildən olan sistemlərdə gedən proseslərlə bağlı, təhlil və sintez alqoritmlərinin işində nəzərə alınması zəruri hesab olunan morfonoloji dəyişmələr tədqiq edilmişdir.

VIII FƏSİL. MƏTNLƏRİN FORMAL TƏHLİLİ SİSTEMİNİN SİNTAKTİK BLOKU

8.1. Sintaktik blokun funksiyası

Mətnin formal təhlili sistemində və eləcə də mövcud digər sistemlərdə sintaktik təhlil və sintez həlledici xüsusiyyətə malikdir. Formal sistemi təşkil edən digər bloklar (leksik, morfoloji, morfonoloji semantik) bütövlükdə cümlənin (daha geniş götürdükdə, mətnin) tam təhlili üçün zəruri olan informasiyanın toplanması və sintaktik bloka verilməsinə xidmət edirlər. Sintaktik blok həm də mürəkkəbliyinə görə digər bloklardan fərqlənir. Mətnlərin avtomatik işlənməsi və MT sistemlərində sintaktik təhlil və sintez problemi müxtəlif cəhətlərdən araşdırılmış, bir çox nəzəri və təcrübi sxemlər qurulmuşdur (Bax: Kuno, Ettinger, 1971 a; 1971 b; Лайонз, 1978; Налимов, 1974; Фитиалов, 1961; Хоул, 1968; Хомский, 1962; 1966; 1972).

Ayrı-ayrı tədqiqat işlərinin səciyyəvi xüsusiyyətləri üzərində dayanmazdan əvvəl sintaktik blokla bu və ya digər dərəcədə əlaqəli olan, onun mahiyyətinə təsir edə biləcək bəzi məsələləri aydınlaşdırmaq.

Hər şeydən əvvəl, müəyyənləşdirilməlidir ki, sintaktik blok nəyə xidmət edir. Məlum olduğu kimi, mətnlərin avtomatik işlənməsi və MT sistemlərinin özündə də sintaktik blokun funksiyası, fəaliyyət dairəsi dəqiq məlum olmalıdır. Əgər bizi sözbəsöz tərcümə, leksik-morfoloji səviyyədə avtomatik işlənmə maraqlandırarsa, onda belə sistemlərdə sintaktik təhlil səthi səciyyə daşıyır. Bu halda sintaktik təhlil deyil, onun ayrı-ayrı elementləri sistemə daxil edilir.

Əgər sintaktik bloku kontekstə daxil olmayan, ayrılıqda götürülmüş cümlənin təhlili və sintezi maraqlandırarsa, onun meyar və üsulları da buna müvafiq olacaqdır.

Tamamlanmış, bitmiş bir mənaya malik mətnlərin (cümle çoxluqlarının) təhlili isə avtonom cümlədən fərqli olacaqdır. Belə mətnlərdə ayrı-ayrı cümlələr arasında sıx əlaqə olur.

Mətnlərdəki cümlə ardıcılığında əvvəl gələn cümlə, sonrakı cümlənin qarşısında duran kommunikativ əlaqəni izah edir (Гальперин, 1981). Ayrılıqda hər bir cümlə mətn daxilində həm müstəqil bir tamlıqdır, həm də eyni zamanda müəyyən bir asılılığı ifadə edir. Bu cümlə özündən əvvəlki cümlədəki fikri davam etdirir, eyni zamanda özündən sonrakı cümlə üçün zəmin hazırlayır. Məz 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 morfemlərdə, morfemlərin tam izahı sözlərdə baş verirsə, sözlərin də tam və bitkin mənası (istər leksik, istərsə də qrammatik) cümlədə məlum olur (Солнцев, 1977). Təsadüfi deyil ki, izahlı lüğətlərdə sözlərin bütün məna çalarlıqlarını vermək üçün onlar müxtəlif cümlələrdə - sintaktik əhatələrdə verilir. Müşahidə göstərir ki, sözün həqiqi mənası yalnız kontekstdə məlum olur. Bu baxımdan cümlə də ayrılıqda tam təhlil olunub qavranılmır. Cümlənin tam «başə düşülməsi» üçün onu mətndə götürmək lazımdır, ondan əvvəl və sonra gələn cümlələrin informasiya tutumu və kommunikativ funksiyası cümləyə üz təsirini göstərməmiş olmur. Lakin mətnin «qavranması» cümlənin «qavranmasından» daha mürəkkəbdir, daha doğrusu, formal meyarlar burada morfoloji blokda 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 mənimsəyərək insan təkcə həmin mətn çərçivəsində qapanıb qalmır, digər ekstralinqvistik bilik və təcrübələrdən faydalanır. Məhz buna görə, avtomatik işlənmə sistemlərində sintaktik bloka yanaşma müxtəlif olmuşdur. Sintaktik təhlil təkcə cümlə konstruksiyalarının araşdırılması və sintezi üçün deyil, həm də morfoloji səviyyədə sözlərin hələ tam açılmamış mənalarının tapılması, omonimliyin və polisemiyanın aradan qaldırılması baxımından da çox vacibdir. Sintaktik blokla əlaqədar MT-nin ilk illərində geniş yayılmış aralıq-keçid dili ideyasına da münasibətin bildirilməsi çox vacibdir.

8.2. Aralıq-keçid dili (vasitəçi dil)

MT-nin təşəkkül tapdığı və geniş yayıldığı vaxtlarda aralıq-keçid dili ilə bağlı qızgın diskussiyalar keçirilir, geniş layihələr işlənib hazırlanırdı (Вах: Мельчук, Равич, 1967).

Aralıq-keçid dili (vasitəçi dil) ideyasının meydana çıxmasının bir çox obyektiv səbəbləri var idi. Bu, hər şeydən əvvəl, dillərin leksik, qrammatik, semantik cəhətdən tam müvafiqliyinin əldə oluna bilməməsindən irəli gəlirdi. Başqa sözlə, tamamilə izomorf iki dil göstərmək mümkün deyil. İki və daha artıq dildə mövcud olan iki sözün bütün mənalарının qismən uyğunluğu nadir hallarda müşahidə olunur. Belə olduqda, leksik və hətta sintaktik müvafiqlikdən danışmaq nə dərəcədə doğru olardı? Məz buna görə də, elə bir süni dil-aralıq-keçid dili yaradılmalı idi ki, bütün digər dillər birbaşa bir-birinə deyil, əvvəlcə bu dilə çevrilsin, sonra bu dil vasitəsilə tərcümə olunacaq dildə ifadə olunsun. Aralıq-keçid dili bütün dünya dilləri üçün qəbul edilə bilən fonetik, morfoloji, sintaktik və semantik quruluşa, qaydalara müvafiq qurulmalı idi. Belə dil mətnlərin avtomatik işlənməsi sistemlərində təhlil və sintez alqoritmlərinin xeyli sadələşdirilməsinə, yığcamlığına şərait yaratmalı idi (Андреев, 1957; Андреев, Кремнева, Волков, 1961; Генкин, 1962; Лейкина, 1961).

Aralıq-keçid dilinin qrammatikası ilə təbii dillərin qrammatikaları arasında üç tip əlaqə göstərilir və bu əlaqə tipindən asılı olaraq linqvistik təhlil və sintez alqoritmlərinin quruluşu və məzmunu müxtəlif olur:

- aralıq-keçid dilinin qrammatikası təbii dillərin qrammatikalarından tamamilə fərqli, müstəqil bir qrammatikadır;

- aralıq-keçid dilinin qrammatikası təbii dillərdən qismən asılıdır;

- aralıq-keçid dilinin qrammatikası təbii dillərin qrammatikalarından tam asılıdır (Андреев, Кремнева, Волков, 1961). Burada təbii dillərin qrammatikaları dedikdə, daha çox avtomatik işlənmə və MT sistemlərində istifadə edilən dillər nəzərdə tutulur. Sonralar aralıq-keçid dilinə dünyanın müxtəlif geniş yayılmış dilləri (İngilis, rus, alman, fransız və s.) arasında leksik, morfoloji,

sintaktik, semantik uyğunluqlar tapılması ilə bağlı müxtəlif layihə və alqoritmlər aid edilməyə başladı. Sintaktik təhlil alqoritmlərinin qismən sadələşdirilməsi ilə əlaqədar aralıq-keçid dillərinə bir çox elementlər daxil edildi. Hər şeydən əvvəl sintaksis ikiyelementli mötərizə kimi qəbul edildi. Bu elementlərdən biri digərindən əvvəl gəlir və özündən sonrakı mövqedə duranı izah edirdi. Cümlənin ümumi səciyyəsi mötərizənin sonunda qoyulmuş xüsusi işarə ilə göstərilir: is- nəqli cümlə, ?-sual cümləsi, !-nida cümləsi.

Beləliklə, bütün göstərilənləri nəzərə alaraq aralıq-keçid dilinin xüsusiyyətlərini ümumi şəkildə belə göstərmək olar:

- aralıq-keçid dilində sinonimlər, çox mənalı sözlər və omonimlər olmur. Sinonim və omonim sözlərdən birinin mənası əsas götürülür. İstənilən mətni aralıq-keçid dilinə keçirməklə, biz bir növ mətni semantik süzgecdən keçirmiş oluruq. Aralıq-keçid dilindən də istənilən dilə tərcümə bu baxımdan asanlaşır.

Mütəxəssislər belə hesab edirdilər ki, bütün dünya linqvistlərinin iştirakı ilə aralıq-keçid dili üçün xüsusi qrammatika hazırlamaq məqsədəuyğun olardı. Müzakirəyə üç layihə təqdim olunmuşdu (Мельчук, Равич, 1967):

1. Aralıq-keçid dili kimi təbii dillərdən biri – standartlaşdırılmış və proqramlaşdırma qaydalarına uyğunlaşdırılmış «requlyar» ingilis dili təklif olunurdu.

2. Aralıq-keçid dili kimi süni beynəlxalq dil tiplərindən biri götürülə bilər. Belə esperanto və ya interlingvo dillərinin bir üstünlüyü də ondadır ki, bir neçə Avropa dilləri üçün ümumi olan qrammatik kateqoriyaları özündə birləşdirir.

3. Daha çox ehtimal olunan, əlverişli yollardan biri də tərcümə maşınları üçün xüsusi olaraq hazırlanmış, ümumi sintaktik qaydaları özündə cəmləşdirən bir süni dil yaradılmasıdır.

Aralıq-keçid dili ilə bağlı bir çox linqvistik problemlərin semantik yaxınlıq və oxşarlıq, semantik lüğətlərin - tezausların tərtibi məsələlərinin də həlli tələb olunurdu.

Göstərmək lazımdır ki, aralıq-keçid dili ideyası (eləcə də beynəlxalq dil – esperanto, beynəlxalq əlifba ideyaları) həyata keçməsə də, onun mətnlərin sintaktik təhlili və sintezi alqoritm-

lərinin tərtibi, formalaşması və inkişafında böyük rolu olmuşdur. Mətnlərin avtomatik işlənməsi alqoritmlərinin sadələşdirilməsi və optimallığı üçün aralıq-keçid dili ideyası ilə bağlı meydana çıxmış bir çox elementlərdən geniş istifadə olunur.

8.3. Sintaktik təsadüflər və ümumiləşdirmə

Sintaktik təhlil və sintez alqoritmlərini iki istiqamətə əsaslanmaqla qurmaq olar. Bu istiqamətləri şərti olaraq sadədən mürəkkəbə və mürəkkəbdən sadəyə doğru hesab etmək olar. Əvvəlcə, xüsusi olaraq seçilmiş kiçik mətnlər götürülür. Həmin mətnlər məhdud elm sahələrinə (daha çox texniki elmlər nəzərdə tutulur) aid olur və onun üçün terminlər daha səciyyəvi olur. Həmin mətnlər üzərində avtomatik sintaktik təhlil və sintez alqoritmləri sınaqdan keçirilir. Bütün müşahidə olunan istisna və təsadüfi hallar qeyd edilir və alqoritmə nəzərə alınır. Sonra bu nisbətən sınaqdan keçirilmiş və təkmilləşdirilmiş qaydalar həcm-cə daha böyük, geniş mətnə tətbiq olunur, yoxlanılır, təkmilləşdirilir. Getdikcə mətnlər daha da genişləndirilir və alqoritmə müşahidə olunan istisna və təsadüfi hallara müvafiq korreksiyalar aparılmaqla avtomatik işlənmənin sintaktik bloku üçün səciyyəvi olan qaydalar, qanunauyğunluqlar sistemi işlənilir və hazırlanır. Tədqiqat işində bu metod aparıcı olmuş, morfoloji və morfonoloji bloklarda, eləcə də leksik blokda aşağıdan yuxarıya, sadədən mürəkkəbə, kiçik massivlərdə sınaqdan keçirilməklə daha böyük həcmli massivlərə keçid prinsipi əsas götürülmüşdür.

Sintaktik blokun ayrı-ayrı komponentlərinin konkret izahına keçməzdən əvvəl yuxarıda qeyd etdiyimiz sintaktik nəzəriyyələri ümumi şəkildə nəzərdən keçirək. Sintaktik blokun qurulması üçün bunların araşdırılmasının əhəmiyyəti böyükdür.

8.4. Sintaktik modellər

Məlum olduğu kimi, dilə müəyyən işarələrin (hərflər və ya söz) ardıcıl düzülmüş siyahısı kimi baxmaq olar. Lakin istənilən hərflər zənciri söz və ya cümlə məzmunu vermir. Məsələn, Azər-

baycan dilində «çörək» hərflər ardıcılığı həm leksik, həm də qrammatik məzmunu malik bir məfhumu bildirir. Həmin hərflərin iştirak etdiyi, başqa ardıcılıqla düzülmüş hərflər zəncirinin («çrköə») isə heç bir leksik və qrammatik mənası yoxdur. Belə düzülüş qaydası Azərbaycan dilinin söz kökləri və əsaslarında sait və samitlərin düzülüşü qanunauyğunluğuna da ziddir (Cəlilov, 1988, 47).

Dilin formal qrammatikası əlifba işarələrinin müxtəlif ardıcılıqla düzülmüş formalarından dilə uyğun gələni seçib ayırmaq «qabiliyyətinə» malik olmalıdır. Yəni, «çrökə», «çrəkö», «çrköə», «ökrəç», «əkörç», «rçökə», «çörək» və s. ardıcılıqla düzülmüş hərflər zəncirindən Azərbaycan dilinin leksik quruluşunda təsbit olunmuş «çörək» sözünü «taniya» bilməlidir. Əlbəttə, belə qrammatikada dildəki bütün sözləri verməklə vəziyyətdən çıxış yolu tapmaq olardı. Digər tərəfdən, məlumdur ki, dildəki bütün sözləri (bura sözdüzəltmə və sözdəyişdirmə kateqoriyaları vasitəsilə düzələn sözləri də əlavə etsək) siyahı şəklində vermək qeyri-mümkündür. Əgər siyahı şəklində verilsə belə, həmin qrammatika praktik cəhətdən istifadə oluna bilməz.

Deməli, müəyyən qayda və qanunauyğunluqlar əsasında qurulmuş xüsusi qrammatikalar yaratmaq zərurəti qarşıda durur. Belə qrammatikalar qurularkən təbii dillər üçün yaradılmış mövcud qrammatikaların bir çox xüsusiyyətləri araşdırılmalı, nəzəri ümumiləşdirmələr əsasında dünya dilləri üçün yararlı qayda və qanunauyğunluqlar müəyyənləşməlidir.

Formal əlamətlərə əsaslanan belə qrammatikaları şərti olaraq söz-formaları araşdıran morfologiyaya (əlifba işarələrinin müəyyən ardıcılıqla düzülmüş zəncirini söz-forma kimi götürmək olar), həmin baza özülündə qurulmuş bütün cümlə tiplərini izah edən sintaksisə ayırmaq olar.

Yuxarıda göstərilən üsullardan istifadə edərək söz qrupları qrammatikası yaratmaq olar. Şərti olaraq belə adlandırılan qrammatikalarda cümlələr bir-birilə kəşisməyən əlaqəli söz qruplarına parçalanır. Bu söz qrupları konfigurasiya adlanır. Hər bir konfigurasiya aşağı tərtibdə daha kiçik konfigurasiyalara ayrılır. Ən

kiçik konfiqurasiyalarda onun tərki b hissəsi sayılan söz-formalar qeyd olunur. Bu üsulla cümlə onu təşkil edən konfiqurasiyalara, konfiqurasiyalar isə öz növbəsində daa kiçik konfiqurasiyalara ayrılır. Konfiqurasiyalar cümlə tərki bində vasitəsiz iştirakçıların (təşkiledicilərin) ayrılması kimi başa düşülür.

Belə vasitəsiz iştirakçılara ayrılma və söz qruplarının cümlədə müəyyən ardıcılıqla düzümü mətnlərin avtomatik işlənməsi və MT alqoritmlərinin tərki bində istifadə olunur. Cümlələrin konfiqurasiyalara ayrılması bütün dünya dillərinə tətbiq oluna bilər və bu üsulla istənilən dilin qrammatik quruluşu təsvir oluna bilər. (Лайонз, 1978).

Formal qrammatikaların başqa bir növü söz-formalar arasında qrammatik asılılıq əsasında meydana çıxan qanunauyğunluqların araşdırılması ilə səciyyələnilir.

Fərz edək ki, söz-formaların müəyyən sonlu ardıcılığı zənciri verilmişdir: $S=a, v, s, \dots, n$.

Bu söz-formaları zəncirdə düzülmə ardıcılığı ilə nömrələyəək:

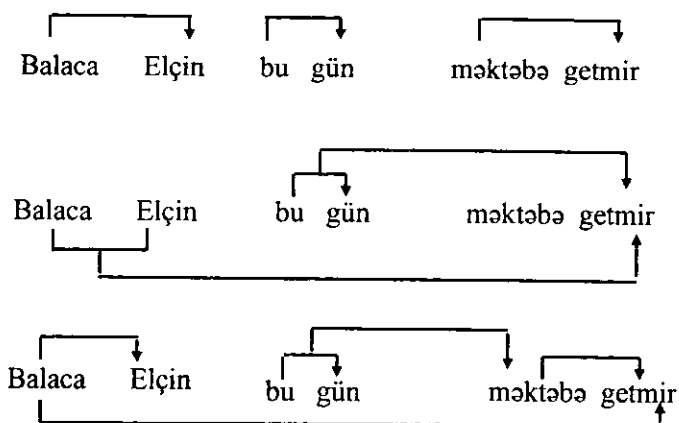
$a_1, v_2, s_3, \dots, n_i$

Sonrakı mərhələlərdə bu söz-formalar ardıcılığı üçün bütün mümkün konfiqurasiyalar axtarılıb tapılır və ayrı-ayrı konfiqurasiyalar arasında qrammatik asılılıq müəyyənləşdirilir (Хомский, 1962, 1966, 1972). Məsələn: «Balaca Elçin bu gün məktəbə getmir» - cümləsində aşağıdakı söz-formaları ayırmaq olar:

Balaca, Elçin, bu, gün, məktəbə, getmir.

Bunlar da konfiqurasiyalar şəklində belə göstərilə bilər:

Balaca Elçin, bu gün getmir, məktəbə getmir, Elçin getmir, balaca Elçin getmir və s. Bu göstərilən konfiqurasiyalar arasında qrammatik əlaqələr müxtəlifdir:



Bu əlaqə formalarının qeydə alınması və onların təhlili istər sintez, istərsə də təhlil alqoritmlərinin quruluşuna ciddi təsir göstərir. Xüsusilə tərcümə olunacaq dilin söz-formaları arasında mümkün qrammatik əlaqələrin analoji olaraq sintezi üçün əhəmiyyətlidir.

Göstərilən qrammatikalar dildə olan hər bir söz-forma üçün mümkün asılılıq kontekstini müəyyənləşdirməyə imkan verir. Bu mümkün asılılıq kontekstləri müəyyən qaydalar şəklində qeyd olunur. Başqa sözlə desək, yuxarıda təsvir olunan qrammatikalar söz-formaların valentliyini müəyyənləşdirir və onların müəyyən ardıcılıqla düzülməsi prosesindəki məhdudlaşdırmaları göstərir.

Söz-formalar arasındakı qrammatik asılılığa əsaslanan qrammatikalar tabeli mürəkkəb cümlə səviyyəsində optimal nəticələr vermir və bu qrammatikaların mürəkkəb cümlə tiplərinə tətbiqi məqsədəuyğun sayılmır. Tabesiz mürəkkəb cümlələr və həmcins üzvlü sadə geniş cümlə tipləri də bu qrammatika çərçivəsində çətin izah olunur. Bütün bunlar göstərir ki, belə qrammatikaların tətbiq dairəsinin genişləndirilməsi üçün yeni-yeni metodlar və üsullar işlənilməli, sınaqdan keçirilməlidir. N.Xomski ideyalarına əsaslanan qrammatikaların təhlil və sintez alqoritmlərinin hazırlanması problemi ilə bağlı böyük əhəmiyyəti qeyd

olunmalıdır (Фиталов, 1961). Bu tip qrammatikalar dilin universal təsviri, bütün dillər üçün yararlı qrammatikalar hazırlanması işində uğurlu axtarışlar prosesidir. Daha çox sintez alqoritmlərinin tərtibi ilə bağlı transformasiya qrammatikaları geniş tətbiq edilir. Azərbaycan mətnlərinin kompüterlərdə təhlili və sintezi sistemində transformasiya qrammatikalarının ayrı-ayrı elementlərindən yeri gəldikcə istifadə olunmuşdur.

8.5. Universal qrammatikalar nəzəriyyəsi

Mətnlərin avtomatik işlənməsi sistemi sintaktik blokunun hərtərəfli və tam fəaliyyəti məşın sintaksisinin işlənilib hazırlanması zərurətini doğurur. Əvvəlcə, Xomskinin universal qrammatika nəzəriyyəsinin əsas müddəalarını nəzərdən keçirək. Aşağıdan yuxarı, sadədən mürəkkəbə doğru istiqamətə əsaslanan qrammatikalardan fərqli olaraq bütün təbii dillər üçün yararlı qrammatikalar işlənilib hazırlanması bu qrammatikaların əsasını təşkil edir (Хомский, 1962, 1966, 1972).

Dil dedikdə, müəyyən uzunluğa malik, müəyyən sayda elementlər çoxluğundan təşkil olunmuş cümlələr çoxluğu başa düşülür. Hər bir təbii dil fonem və ya əlifbanı təşkil edən hərflərin sonlu sayından ibarətdir. Dilin cümlələri bu fonemlərin və ya hərflərin müəyyən qanunauyğunluqla düzölmüş zəncirindən ibarətdir. Bu qayda ilə dilə - cümlələr çoxluğuna təbii dil kimi baxmaq olar. Linqvistik təhlil - cümlə adlanan fonemlər (morfemlər) zəncirinin qrammatik cəhətdən düzgün sıralandığını müəyyənləşdirmək və onu qrammatik cəhətdən düzgün hesab oluna bilməyən fonemlər ardıcılığından fərqləndirməkdir. Linqvistik təhlil sübut edə bilməlidir ki, qrammatik cəhətdən düzgün olmayan cümlələr bu və ya digər təbii dilin cümlələri hesab oluna bilməz. Təhlilin düzgün hesab etdiyi cümlələr isə tədqiqat obyektinə çevrilməli, qrammatik cəhətdən düzgün cümlə strukturları daha dərindən araşdırılmalıdır. N.Xomskiyə görə, hər hansı bir təbii dilin qrammatikası qrammatik cəhətdən düzgün sayılan fonemlər (morfemlər) ardıcılığını - cümləni almağa, sintez etməyə imkan verə biləcək mexanizmdən başqa bir şey deyildir. Bu qramma-

tika o vaxt kamil sayılır ki, yalnız düzgün cümlə formaları doğursun və heç bir düzgün olmayan fonem (morfem) ardıcılığına yol verməsin. Əgər bizim formal qrammatikamız təbii dil üçün yabançı olmayan, düzgün cümlə formalarını törədə bilirsə, onu məqbul hesab etmək olar. Yuxarıda göstərilənlərdən belə bir nəticə çıxır ki, fonemlərin hər cür ardıcılıqla düzülüşü cümlə sayıla bilməz, yəni elə fonem ardıcılığı var ki, onlar cümlə deyil. Bu baxımdan sintaksislə semantika arasında əlaqənin araşdırılmasının böyük əhəmiyyəti var. Bu əlaqə yalnız heç bir təbii dilin qrammatikasıdan asılı olmayan universal qrammatika yaradıldıqdan sonra mümkün olardı.

Xomskinin sintaktik nəzəriyyəsinə görə, kontekstə daxil olmayan dil materiallarını əhatə edən formal qrammatika təbii dilin əmələ gətirə biləcəyi bütün düzgün cümlə formalarının qeydə alınması yolu ilə qurulur. Bu formal qrammatika bir çox ilkin simvoldan, qaydalar çoxluğundan, törəmə zəncirlərdən və aksiomlardan təşkil olunmuşdur. Nəzəri olaraq bütün mümkün cümlə strukturları bura daxildir (Солнцев, 1977). Qaydalar çoxluğu düzgün cümlə strukturunu doğura bilməklə yanaşı, düzgün olmayanları da seçib ayırmaq imkanına malikdir. Sintaktik təhlil bu qrammatikada vasitəsiz təşkilədicilər (iştirakçılara) ayrılma kimi başa düşülür. N.Xomskiyə görə, bu elə bir qrammatika olmalıdır ki, bütün təbii dilləri əhatə və izah etmək imkanı olsun. Sonralar vasitəsiz iştirakçılar qaydalarını daha da inkişaf etdirməklə N.Xomski transformasiya qrammatikaları nəzəriyyəsinin irəli sürdü. Transformasiya qrammatikaları üç mühüm komponentdən təşkil olunur:

- vasitəsiz iştirakçılar üçün qaydalar sistemi;
- transformasiya qaydaları sistemi;
- morfonoloji qaydalar sistemi.

Məlum olduğu kimi, N.Xomski sintaktik nəzəriyyəsinin sintaktik əsasını təşkil edən bu qaydalar daha mükəmməldir və cümlələrin nüvə quruluşu ilə səthi quruluşu arasında əlaqəni müəyyənləşdirməyə xidmət edir. Bu əlaqə transformasiya təhlil və sintez vasitəsilə həyata keçirilir. Transformasiya qaydaları hər

bir dilin özünəməxsusluğunu nəzərə alır. Cümlənin dərin quruluşu (nüvəsi) dedikdə təfəkkürlə bilavasitə bağlı olan və dilin bu və ya digər xüsusiyyətindən asılı olmayaraq fikrin cümlə şəklində semantik interpretasiyası başa düşülür. Səthi quruluş cümlənin dərin quruluşunun (nüvəsinin) müxtəlif interpretasiyaları kimi təzahür edir. Səthi interpretasiyalar bəzən dərinlik məzmununu tam ehtiva etmir, təhrif edir. Transformasiya qaydaları cümlənin dərin quruluşundan səthi quruluşa keçid zamanı invariant mənanı müəyyənləşdirir, yeganə düzgün interpretasiyanın seçilməsi yollarını göstərir. Yuxarıda qeyd etmişdik ki, cümlə çoxluğu mətni (dili) təşkil edir. Cümlənin özü də söz-formalar çoxluğudur:

C--- a**U**b**U**c**U**d**U**...**U**n

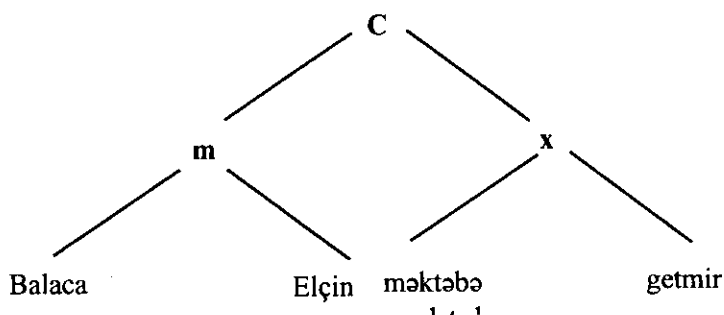
Burada, C- cümlə, a, b, s, ..., n onu təşkil edən söz-formalar, U-konkatenasiya (birləşmə) işarəsidir. Konkatenasiya göstərir ki, söz-formalar müəyyən qanunauyğunluq əsasında (məsələn, sintaktik əlaqələr nəzərə alınmaqla) yanaşı gəlir. Onların göstərilmiş ardıcılıqla düzülüşü qrammatikaya zidd deyildir. Əgər cümləni təşkil edən elementlər arasında + (plyus) işarəsi qoyulsaydı, onda belə başa düşülərdi ki, C-sintaqmı – a, b, c,, n nizamlanmamış elementlərinin birləşmələrindən ibarətdir. Transformasiya qaydalarının izahına keçməzdən əvvəl vasitəsiz təşkiledicilər (iştirakçılar) termini üzərində dayanmaq. Belə bir cümləni nəzərdən keçirək:

Balaca Elçin məktəbə getmir.

Bu cümləni aşağıdakı tərkib qruplarına bölmək olar:

Balaca Elçin və məktəbə getmir. Əlbəttə, bu tərkib qruplarını mübtəda qrupu və xəbər qrupu kimi də təsnif etmək olardı. Lakin vasitəsiz təşkiledicilər anlayışı altında mübtəda və xəbər qrupundan söhbət getmir. Bu termin hər şeydən əvvəl onu göstərir ki, cümlə, cümlədən kiçik elementlərin hər hansı xətti düzülüşü deyil. Bu xətti düzülüşün özündə də müxtəlif səviyyələr var və həmin səviyyələrin konkatenasiyası cümləni təşkil edir:

(iştirakçılara) ayrılması



Sxem 9. Cümlənin vasitəsiz təşkiledicilərə (iştirakçılara) ayrılması

Sxemdən göründüyü kimi, cümlə *balaca, Elçin, məktəbə, getmir* elementlərindən ibarətdir. Burada *Balaca* və *Elçin* sözləri bir qoldan (m) ayrılırlar və eyni konstruksiyanın vasitəsiz təşkilediciləridir. *Məktəbə, getmir* sözləri isə başqa qoldan (x) ayrılırlar və digər konstruksiyanın vasitəsiz iştirakçılarıdır. Bu iki konstruksiya m (*Balaca Elçin*) və x (*məktəbə getmir*) öz növbəsində daha yuxarı səviyyənin (C) - cümlənin vasitəsiz iştirakçısıdır. Məhz ona görə də, bu iki qol C qovuşma nöqtəsindən ayrılır. İlk dəfə Blumfild tərəfindən irəli sürülmüş bu ideya sonralar N.Xomskinin əsərlərində özünün geniş riyazi interpretasiyasını tapdı (Блумфилд, 1981).

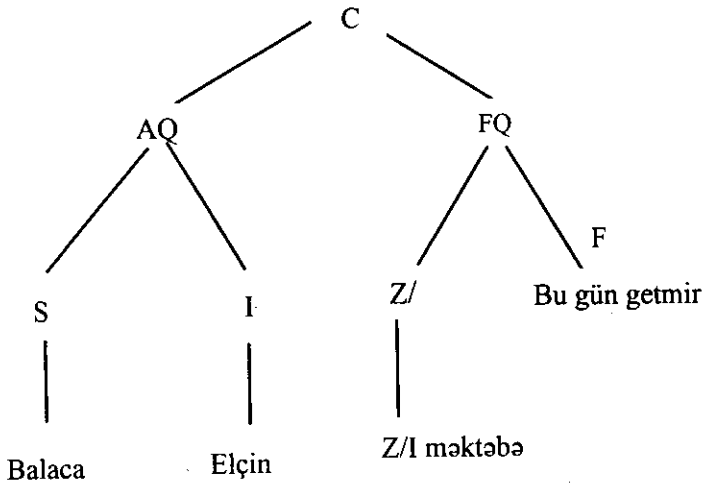
Azərbaycan dilində cümlənin vasitəsiz iştirakçılarna ayrılmasını aşağıdakı nümunə üzərində nəzərdən keçirək:

Balaca Elçin bu gün məktəbə getmir.

Bu cümləni simvollarla aşağıdakı kimi yazmaq olar:

$S\{ AQ /S [Balaca] \cup I [Elçin] \cup FQ/Z [bu\ gün] \cup Zi [məktəbə] \cup F [getmir]$

Burada C- cümlə, AQ -ad qrupu, FQ- fel qrupu, S-sifət, I- isim, Z- zərf, F- fel nitq hissəsini göstərir:



Sxem 10. Cümlənin qruplar üzrə vasitəsiz təşkeildicilərə (iştirakçılara) ayrılması

Yuxarıdakı ifadədə və sxemdə Z/I təsadüfi işlənməmişdir. Burada *məktəbə* sözü hərəkətin icra olunduğu (bizim misalda olunmadığı) yeri göstərir və hara? sualına cavab olur. Bu baxımdan *məktəbə* sözü cümlədə yer zərfliyi vəzifəsi daşıyan və yer zərflərinin suallarına cavab verən zərflərdən heç nə ilə fərqlənmir. Lakin professor M.Hüseynzadə göstərir ki, yer zərfləri ilə yer bildirən isimlərin fərqi şəkli əlamətdə deyil, onların əşya anlayışına malik olub-olmamasındadır. Müəllifə görə «yer bildirən isimlər ismin hansı halında işlənirsə-işlənsin, cümlədə hansı vəzifə daşıyırsa-daşısın, müəyyən yer bildirən varlıq kimi də təsəvvür olunur, yer zərfləri isə belə bir anlayış bildirmir» (Hüseynzadə, 1988, 242). Tədqiqat işinin morfoloji blokunda türk dillərində, o cümlədən Azərbaycan dilində nitq hissələrinin təsnifinin məntiqi bölgü qanunlarına zidd şəkildə aparıldığı göstərilmişdi. Professor M.Hüseynzadənin yuxarıda qeyd olunmuş müd-

dəalarna əsaslansaq, onda zərf nitq hissəsinin və yer zərfinin tərifini bir qədər dəyişdirilmiş formada qəbul etməliyik:

- İş və hərəkətin icra tərzini, zamanını, yerini, miqdarını, eləcə də əlamətini bildirən, müəyyən əşya anlayışına malik olmayan və yer bildirən varlıq kimi təsəvvür edilməyən sözlərə zərf deyilir.
- hərəkətin və hadisənin icra yerini, istiqamətini, son nöqtəsini, çıxış nöqtəsini bildirən, müəyyən əşya anlayışına malik olmayan və yer bildirən varlıq kimi təsəvvür edilməyən sözlərə yer zərfi deyilir.

Mübahisəli səciyyə daşıyan bu qəbildən olan məsələləri kənara qoyaraq təhlil alqoritmlərində transformasiya qrammatikalarının istifadəsi perspektivlərini nəzərdən keçirək.

8.6. Təhlil sistemlərinə transformasiya qrammatikalarının tətbiqi

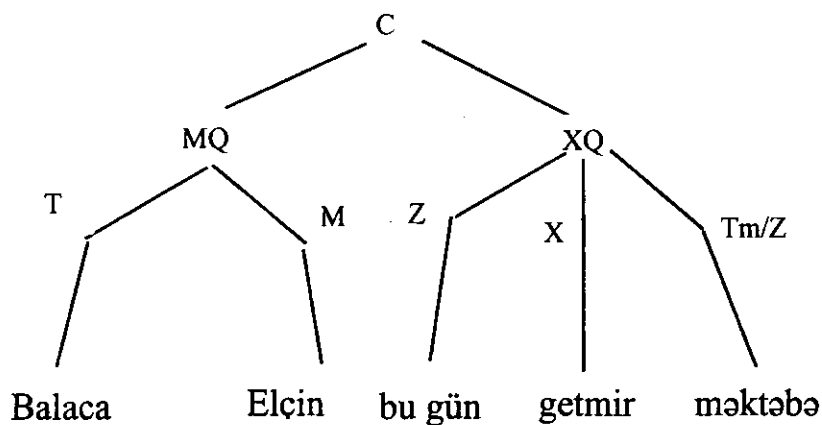
Transformasiya qrammatikaları əsasən sintez məsələləri ilə bağlı qurulur. Həmin qrammatikanın əsasında duran linqvistik modellər də törədici sintez modelləri adlanır (Лайонз, 1978). Ona görə də, həmin modellərin təhlil məsələlərinə tətbiqi daha böyük maraq doğurur. Türk dillərində sintaksisin modelləşdirilməsi sahəsində işlər başlanğıc mərhələdədir. Ayrı-ayrı fraqmentlər nəzərə alınmazsa, mətnlərin avtomatik işlənməsi sistemlərində sintaktik blok mükəmməl təhlil və sintez alqoritmləri ilə də təchiz olunmamışdır.

Yuxarıda göstərilmiş cümlənin sintaktik baxımdan vasitəsiz təşkeildicilərə ayrılmasını formal şəkildə belə ifadə etmək olar:

$$S \{ MQ (T [balaca] \cup m [Elçin] \cup XQ (Z [bu gün] \cup T/m/Z [məktəb] \cup \cup X [getmir]) \}$$

Burada C - cümlə, MQ - mübtəda qrupu, T - təyin, M - mübtəda, XQ - xəbər qrupu, Z - zərflik, T/m/Z - tamam-

lıq/zərflilik, X - xəbəri göstərir. Həmin ifadəni sxem şəklində aşağıdakı kimi göstərmək olar:



Sxem 11. Sintaktik baxımdan vasitəsiz təşkiledicilərə (iştirakçılara) bölünmə

Sxemdən görüldüyü kimi, sintaktik təhlil blokunda mühüm mərhələ mübtəda qrupu və xəbər qrupunun müəyyənləşdirilməsidir. Buna görə də, tədqiqatda əsas diqqət formal əlamətlərinə görə mübtəda və xəbər qrupunun tapılması ətrafında cəmləşdirilmişdir.

8.7. N.Xomskinin linqvistik fərziyyələrinin mahiyyəti və praktik istifadə dairəsi

N.Xomskinin linqvistik fərziyyələrindən ən başlıcası dərin (nüvə) quruluşu və səthi quruluşla bağlıdır. XX sərin 50-ci illərində dərin (nüvə) və səthi quruluş, generativ qrammatika, vasitəsiz iştirakçılar kimi dilçilik üçün yeni olan terminlər getdikcə daha geniş yayılır, linqvistikada yeni bir eranın başlandığından xəbər verirdi (Лайонз, 1978).

N.Xomskiyə görə, dərin və ya nüvə quruluşu əqli məfhumdur, məna tərkibi, fikrin ideal hissəsi, maddiləşməmiş, dərinlikdə yerləşən, cümlənin mənasını təşkil edən nüvə, bazis rolunu oynayır. Səthi quruluş isə onun maddiləşmiş, səs tərkibi (və ya qrafik işarə) ilə ifadə olunmuş surətidir. Dərin quruluştan səthi quruluşa keçid transformasiya qaydaları vasitəsilə mümkün olur. N.Xomski belə hesab edirdi ki, qrammatika dərin quruluştan səthi quruluşa keçid qaydalarını, bu iki dil səviyyəsi arasında əlaqələri öyrənməklə məşğul olmalıdır. Dərin quruluştan fərqli olaraq səthi quruluş daha mütəhərrikdir, onu hiss etmək (əşitmək və ya görmək) mümkündür. Bu konsepsiyanın sonrakı inkişafı vasitəsiz iştirakçılar təliminin və törədici (generativ) qrammatikanın yaranması ilə nəticələndi. Dilçilikdə müxtəlif cərəyanların yarandığı, inkişaf etdiyi bir dövrdə Xomskinin linqvistik fərziyyələri yeni bir məktəbin, istiqamətin başlanğıcından xəbər verirdi. Qısa müddətdə bu linqvistik ideyalar dünya dilçilərinin tədqiqatlarında özünün yeni-yeni əlavələrini, interpretasiyalarını tapdı. Bir çox tədqiqatçılar Xomski ideyalarının dilin mahiyyətinin aydınlaşdırılmasında, maşın tərcüməsində, təhlil və sintez məsələlərində bir inqilab yaradacağını, tezliklə bu sadalanan sahələrdə bütün problemlərin asanlıqla həll olunacağını, maneələrin aradan qaldırılacağını zənn edirdilər. Lakin hər hansı riyazi model, hətta bu model yüksək əqli ümumiləşdirmələrin, məntiqi qaydaların professional yekunu olsa belə, doğurduğu, əks etdirirdiyi, təmsil etdiyi obyektə tam ehtiva edə bilməz. Təcrübə göstərir ki, hər hansı bir model, o cümlədən linqvistik model dilin bütün zənginliyini, özünəməxsusluğunu, fəlsəfəsini, mahiyyətini əks etdirməkdə acizdir. Linqvistik model yalnız dili bu və ya digər cəhətdən qismən izah etməyə qadirdir və bu müddəni dilçiliyin sonrakı inkişafı, bir çox tədqiqatçıların təklif etdikləri model nümunələri bir daha sübut etdi (Piotrovskiy, 1979).

N.Xomskinin linqvistik fərziyyələri də məhz bu göstərilənlərə görə dilçilikdə möcüzə yaratmadı. Güclü məntiqi arqumentləri, riyazi dəqiqlik, fəlsəfi tamlıq, evristik alqoritmlər belə

N.Xomskinin linqvistik fərziyyələrinə gözlənilən müvəffəqiyyəti qazandıra bilmədi (Солнцев, 1976; Андреев, 1976).

Dərin zəkanın, analitik təfəkkürün nəticəsi kimi bu linqvistik fərziyyələr elmi baxışların formalaşmasında, düzgün, həqiqi yol axtarışında, şübhəsiz ki, əhəmiyyətsiz sayıla bilməz.

Dilçilikdə cümlə konstruksiyalarının təhlil və sintezi, sintaktik semantika və bir çox başqa nəzəriyyələrin inkişafında Xomski qrammatikalarının xüsusi rolunu qeyd etmək lazımdır. Yuxarıda təhlil və sintez alqoritmlərinin məzmununa, quruluşuna bu qrammatikaların təsiri qeyd olunmuşdu. Deyilənlərdən ələ nəticə çıxarılmamalıdır ki, transformasiya qrammatikaları öz tarixi vəzifələrini yerinə yetirmişlər və onları «arxivə» vermək vaxtı çatmışdır. Dilçiliyin sonrakı inkişafı ilə bağlı ələ məqamlar yarana bilər ki, tədqiqatçılar yenidən N.Xomski ideyalarına müraciət etməli olacaqlar. Real formal sistemlərin yaradılması ilə əlaqədar N.Xomski konsepsiyasının bəzi elementlərindən istifadə etməklə kifayətlənmək olar. Belə sistemlərin qurulması işində artıq özünü doğrultmuş və praktik sınaqdan uğurla çıxmış linqvistik nəzəriyyələrə, konsepsiyalara əsaslanmaq daha məqsəddəyğündür.

Mətnlərin formal təhlili sisteminin morfoloji və sintaktik blokunun strukturu və alqoritmlərin tərtibində professor R.Q.Pitrovski tərəfindən təklif olunmuş mühəndis dilçiliyi modelləri nəzəriyyəsi əsas götürülmüşdür.

8.8. Bərpaedici mühəndis dilçiliyi modelləri

Məlum olduğu kimi, riyazi-linqvistik modellər bu və ya digər problemin daha dərindən araşdırılmasına, onun adı tədqiqatçı tərəfindən müşahidəsi çətin olan xüsusiyyətlərinin daha qabarıq şəkildə əks olunmasına və öyrənilməsinə xidmət edir. Bu qəbildən olan linqvistik modellərdə hər hansı bir linqvistik obyekt haqqında müəyyən fərziyyə irəli sürülür və həmin fərziyyəyə müvafiq olaraq həmin obyektin linqvistik obyektini yaradılır. Bərpaedici mühəndis dilçiliyi modelləri isə öz quruluşuna və keyfiyyətinə görə fərqli xüsusiyyətlərə malikdir. Belə modellərdə linq-

vistik obyektin analoquna əsasən özü yenidən bərpa olunur. Modelin elmi-nəzəri əhəmiyyəti linqvistik obyektlə onun analoqu əsasında bərpa olunmuş surəti arasında yaxınlıq dərəcəsi ilə ölçülür. Əgər bərpa olunmuş linqvistik obyekt real obyektlə tam eynidirsə, onda linqvistik fərziyyənin həqiqəti aydınlaşdırmaq xüsusiyyətləri yüksək qiymətləndirilə bilər (Пиотровский, 1979). Bərpaedici linqvistik modellərin üstünlüyü orasındadır ki, onun linqvistik obyektinə nə dərəcədə «təqlid» etməsi açıq-aydın müşahidə olunur. Əgər bərpa olunmuş linqvistik obyekt reallıqdan uzaqdırsa, deməli, linqvistik modelin əsasını təşkil edən fərziyyə həqiqətə uyğun deyil.

Məhz bu obyektiv, izahedici, aydınlaşdırıcı və həqiqəti real təsvir etmək kimi mühüm potensial imkanlarını nəzərə alaraq tədqiqat işində sintaktik-semantik qanunauyğunluqların aşkarlanmasında professor R.Q.Piotrovskinin əsaslandığı və sınaqdan keçirdiyi bərpaedici mühəndis linqvistik model tipi əsas götürülmüşdür. Tərtib olunmuş alqoritmlər nəinki bu və ya digər obyektə təhlil etməyə, həm də onun real analoqunun alınmasına xidmət edir (Кенесбаев, Пиотровский, Бектаев, 1970; Пиотровский, 1978).

8.9. Sintaktik blokun əsas komponentləri

Morfoloji və morfonoloji blokların fəaliyyətindən sonra söz-formalar tapılır, əsas və şəkilçilər zəncirinin qrammatik əlamətləri qeydə alınır və avtomatik işlənmənin sonrakı mərhələləri - sintaktik-semantik bloklar üçün zəruri xüsusiyyətlər aşkar olunur. Sintaktik blokun üzərinə cümləni mətnədən ayrılması, mübtədə və xəbərin, eləcə də digər cümlə üzvlərinin tapılması, durğu işarələrinin təhlili, söz birləşmələrinin axtarışı kimi mürəkkəb kompleksə malik problemin həlli düşür. Məlum olduğu kimi, mübtədə və xəbər müxtəlif nitq hissələri ilə ifadə olunur. Təhlil prosesində mübtədə və xəbərin hansı nitq hissəsi ilə ifadə olunduğu da göstərilir. Köməkçi fellərlə, feli frazeologizmlərlə ifadə olunmuş xəbərin müəyyənləşdirilməsi daha mürəkkəb bir prosesdir. Şübhəsiz, sintaktik blokda bu məsələlərin tam və qəti həl-

linə zəmanət verilmir. Sintaktik təhlil və sintez məsələləri bir çox tədqiqatçılar tərəfindən araşdırılmış, müxtəlif metodlar, üsullar, fərziyyələr irəli sürülmüşdür (Бельская, 1969; Дрейзин, 1962; Дрейзин, Рашитов, 1962; Кулагина, 1979; Иорданская, 1967; Лейкина, 1958; Мельчук, 1964).

Eləcə də, təkcə avtomatik sintaktik təhlil və sintez, MT ilə bağlı problemlərlə deyil, həmçinin avtomatik işlənmənin digər məsələləri ilə bağlı da (avtomatlaşdırılmış məlumat axtarışı sistemləri, linqvistik məlumatların saxlanması, təsnifi və müxtəlif sorğulara görə verilməsi, mətnlərin avtomatik «tanınması» və «izahı» və s.) xeyli araşdırmalar aparılmışdır (Беляева, 1979; Белоногов, Новоселов, 1979; Домогацких, 1979; Гройс, 1979; Котов, 1977; Кулик, 1977; ММП, 1958; Нелюбин, 1983; Никитин, 1977; Пиотровский, 1981, Сахарный, 1979; Журавлев, 1979; Чемерис, 1979; Селтон, 1973; Севбо, 1979; Стрелковский, Латышев, 1980; Чубарев, Холодный, 1976; Шуляк, Макаренко, 1979; Штерн, 1979).

Göstərilən ədəbiyyatlarda müxtəlif sintaktik quruluşa, söz sırasına, söz birləşməsi tərkibinə malik dillərdə avtomatik sintaktik təhlil məsələləri şərh olunur. Tədqiqat işində Azərbaycan dilinin sintaktik quruluşunun əsas xüsusiyyətləri nəzərə alınmaqla orijinal təhlil üsulları təklif olunur. Qeyd etmək lazımdır ki, təhlil alqoritmləri Azərbaycan və digər türk dillərinin materialları üzərində yoxlanmış və kifayətləndirici nəticələr alınmışdır.

Sintaktik təhlil prosesində mübtəda və xəbər araşdırılarkən hər şeydən əvvəl onların ifadə vasitələrini müəyyənləşdirmək lazımdır. Hər hansı bir nitq işsəsi 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 ya onların araşdırılması əhəmiyyətsiz hesab olunurdu. 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.

8.10. Xəbərin formal əlamətləri

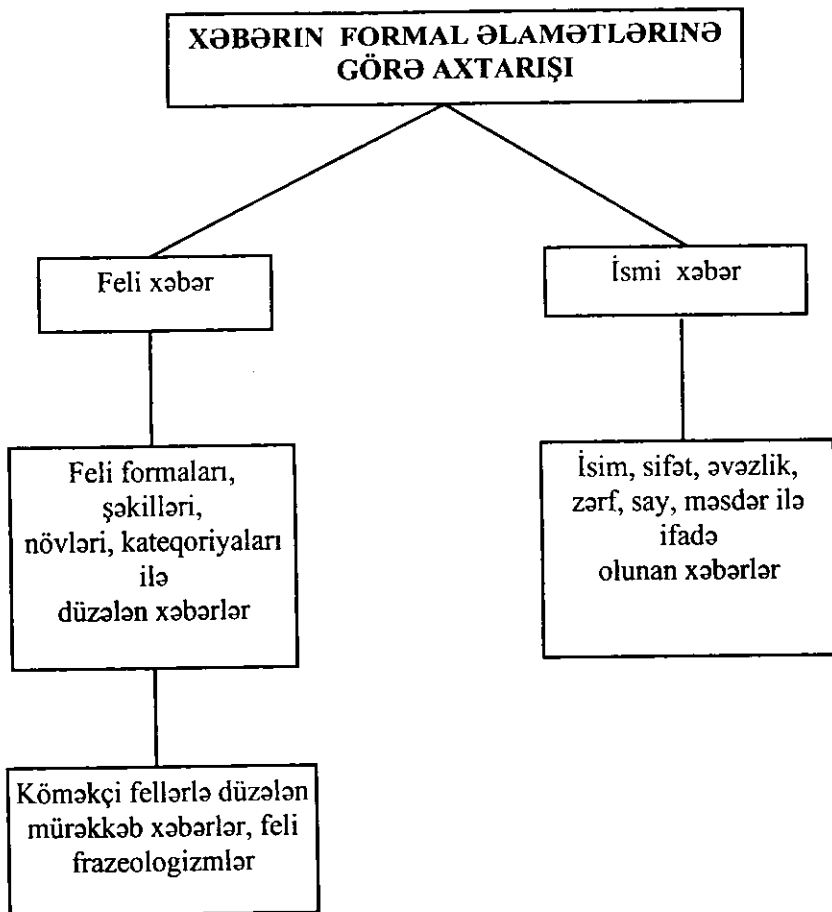
Cümlənin sintaktik təhlilində xəbərin təyin edilməsi mühüm şərtlərdəndir. Xəbər cütlükəməli cümlənin elə bir baş üzvüdür ki, qrammatik cəhətdən yalnız mübtədadan asılı olub, mübtədanın ifadə etdiyi əşyanı geniş mənada bildirir (Abdullayev, Seyidov, Həsənov, 1972).

Azərbaycan dilində xəbər ən çox felin dəyişən şəkilləri ilə və ad birləşmələri ilə ifadə olunur. Ona görə də, xəbərlər ismi və feli xəbərlər kimi iki qrupa ayrılırlar. Fellər zaman, şəxs, təsirlik, növ, forma, inkarlıq və tərz kateqoriyalarına malik olmaqla daha çox predikativlik əlaqəsi yaradır və cümlədə çox vaxt xəbər olur. Azərbaycan dilində fel formalarının zənginliyi xəbərin formal əlamətlərinin də zənginliyinə gətirib çıxarır. Felin dəyişən şəkilləri, növləri, formaları ilə ifadə olunan feli xəbərləri sistemli şəkildə nəzərdən keçirək (Cəlilov, 1983, 102).

Göstərmək lazımdır ki, xəbər köməkçi fellər *etmək, olmaq*, və feli frazeologizmlər (*gözə gəlmək, ələ salmaq, dilə tutmaq, başa salmaq* və s.) ilə də ifadə olunur ki, bu məsələyə mürəkkəb xəbərlərlə bağlı toxunulacaq. Xəbərin formal axtarışı sxem 12-də göstərilmişdir.

8.10.1. Felin əmr şəkli ilə ifadə olunmuş xəbərlər

Felin 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. İzahın yığcamlığı üçün a ə hərflərini V, k, g, ğ, q hərflərini X, ı, i, u, ü hərflərini isə \wedge ilə işarə edək:



*Sxem 12. Xəbərin formal əlamətlərinə görə axtarışının
ümumi sxemi*

Cədvəl 31. Felin əmr şəkli ilə ifadə olunmuş sadə xəbərlər

Şəxs	Şəxs şəkilçiləri		Misallar	
	tək	cəm	tək	cəm
I	-^M	-VX	al-ım gəl-im	al-aq gəl-ək
II	-	-^H	al gəl	al-ın gəl-in
III	-c^H	-C^H/LVP/	al-sın gəl-sin	al-sın(lar) gəl-sin(lər)

8.10.2. Felin xəbər şəklinin xəbərəmələgətirmə imkanları

Məlum olduğu kimi, felin xəbər şəklinin heç bir morfoloji əlaməti yoxdur. Zaman və şəxs şəkilçiləri ilə işlənərək xəbər şəkli icra olunan və ya olunması bir işi 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 fel köklərindən sonra, şəxs şəkilçilərindən əvvəl artırılır. Aşağıdakı cədvəllərdə felin xəbər şəklinin zamanlarla ifadəsi təsvir olunmuşdur.

Cədvəl 32. Şühudi keçmiş zamanla ifadə olunmuş xəbərlər

Şühudi Keçmiş zaman	Şəxs şəkilçiləri		Misallar	
	tək	cəm	tək	cəm
-D^	-M	-X	al-dı-m	al-dı-q
	-H	-N^Z	al-dı-n	al-dı-nız
	-	-/LVP/	al-dı	al-dı-(lar)

Cədvəl 33. Nəqli keçmiş və indiki zamanla ifadə olunmuş xəbərlər

Nəqli Keçmiş zaman	Şəxs şəkilçiləri		Misallar	
	tək	cəm	Tək	cəm
-M \wedge Ş	-VM	- \wedge X	al-mış-am	al-mış-ıq
	-CVH	- S \wedge H \wedge Z	al-mış-san	al-mış-sınız
	-D \wedge P	-/LVP/	al-mış(dır)	al-mış-lar
	-	-	al-ır-am	al-ır-ıq
İndiki Zaman	-	-	al-ır-san	al-ır-sınız
	-	-	al-ır	al-ır-lar
	-	-	-	-

Felin xəbər şəkli ifadə olunmuş xəbərlərin sadəsi zaman və xəbərlik şəkilçilərinin, mürəkkəbi isə idi, imiş hissəciklərinin iştirakı ilə yaranır. M.Hüseynzadə idi, imiş hissəciklərini köməkçi fel kimi səciyyələndirir (Hüseynzadə, 1973, 160). Tədqiqat işində *idi*, *imiş* bağlama kimi verilmişdir.

Cədvəl 34. Qəti və qeyri-qəti gələcək zamanla ifadə olunmuş xəbərlər

Qəti gələcək zaman	Şəxs şəkilçiləri		Misallar	
	tək	cəm	Tək	cəm
-VCVX	-VM	- \wedge X	Al-acağ-am	al-acağ-ıq
	-CVH	- S \wedge H \wedge Z	Al-acaq-san	al-acaq-sınız
	-D \wedge P	-LVP	Al-acaq-(dır)	al-acaq-lar
	-	-	-	-
Qeyri-qəti gələcək zaman	-VM	- \wedge X	Al-ar-am	al-ar-ıq
	-CVH	- C \wedge H \wedge Z	Al-ar-san	al-ar-sınız
	-	-LVP	Al-ar	al-ar-lar

Felin əmr şəkli ilə ifadə olunmuş xəbərlərin inkarı - m şəkilçisi ilə düzəlir. Xəbər şəklinin inkarı isə şühudi keçmişdə,

nəqli keçmişdə və qəti gələcək formasında, fel kökünü sonra inkarlıq şəkilçisinin, sonra isə zaman və şəxs şəkilçilərinin əlavə edilməsi ilə düzəlir. İndiki zaman və qeyri-qəti gələcək zamanlarda *-m* inkar hissəciyi əlavə edilir.

Felin arzu şəkli ilə ifadə olunan xəbərlər *-V* şəkilçisi ilə ifadə olunur:

Sabah gərək işə tez *gələsiniz*.

Felin vacib forması ilə ifadə olunan xəbərlər *-MVL* şəkilçisi ilə düzəlir:

Mən ona bu məsələ barədə *danışmalıyam*.

Felin lazım şəkli ilə ifadə olunan xəbərlər fel kökünə *-C* şəkilçisi artırılmaqla əmələ gəlir:

Deyilənə görə, o, dünən bizdə qalası imiş.

Felin davam şəkli ilə ifadə olunan xəbərlər *-MVXDV* şəkilçisi ilə düzəlir:

Hava soyuqlaşmaqdadır.

Felin lazım şəklinin inkarı digər fel şəkillərindən fərqli olaraq daha çox «deyil» sözünün köməyi ilə sintaktik yolla düzəlir.

Məchul növü əmələ gətirən *-L* şəkilçisinin artırılması ilə də xəbər düzəlir:

Bu barədə lazımı orqanlara məlumat *verilmişdir*.

Yuxarıda göstərilənlərdən əlavə felin qarşılıq, qayıdış növləri, şərt şəkli ilə də sadə xəbər formaları müşahidə olunur:

Qayıdış növlə düzələn xəbərlər *-N*, *-L* şəkilçilərini qəbul edirlər:

Əsgər tələsik *yuyundu*.

Qarşılıq növ şəkilçisi *-Ş*, *-VŞ* artırılmaqla alınan sadə xəbərlər:

Dostlar mehribanlıqla *qucaqlaşdılar*.

İcbar növlü sadə xəbərlər (*-D* *R* şəkilçisi ilə düzələnlər):

Mən evə teleqram *vurdurdum*.

Sadə feli xəbərlər tam şəkildə 24-cü cədvəldə göstərilmişdir.

8.11. İsmi xəbərlər

Feldən, daha doğrusu, təsriflənən feldən başqa digər nitq hissələri və söz birləşmələri ilə ifadə olunan xəbərlər ismi xəbərlər adlanır. İsmi xəbərlərdə xəberlik şəkilçisi funksiyasını -DΛR şəkilçisi yerinə yetirir. Bu şəkilçidən başqa, şəxs şəkilçiləri də xəbər formaları yarada bilirlər. Xəbərlərin bir qismi də «var», «yox», «deyil», «gərək» sözlərinə -DΛR şəkilçisi əlavə olunmaqla düzəlir.

İsmi xəbərlərə aşağıdakıları misal göstərmək olar:

a) İsim xəbər funksiyasında:

Samirə *şagird*dir.

b) Əvəzlik xəbər funksiyasında:

Bu oğlan *kimdir*?

Anasının yeganə oğlu *odur*.

v) Sifət bütün məna növləri və dərəcələri ilə xəbər funksiyasında:

Səma *mavidir*. Qərənfil *qıpqırmızıdır*.

q) Zərf xəbər funksiyasında:

Hələ vaxta *var*, *tezdir*.

ğ) Məsdər xəbər funksiyasında:

Mənim məqsədim *işləmək* və *oxumaqdır*.

Bundan başqa *var/vardır*, *yox/yoxdur*, *deyil/deyildir*, *gərək/gərəkdir*, *gərək deyil* sözləri də xəbər əmələ gətirir:

Vaqif öyünmə ki, kamalımız *var*,

Allaha şükür ki, kamal da *yoxdur*.

Deyil/deyildir sözləri- MV şəkilçisi ilə birlikdə xəbərin inkarını əmələ gətirirlər. *Var*, *yox* sözləri xəberlik şəkilçisi qəbul etmədən də bu funksiyanı yerinə yetirə bilirlər.

Əkəndə *yox*, biçəndə *yox*, yeyəndə orta qardaş (Atalar sözü).

Yuxarıda göstəriləyi kimi, ismi xəbərlər ad nitq hissəsinə şəxs şəkilçiləri əlavə etməklə düzəlir:

Siz təcrübəli *müəllimsiniz*.

Mənim qardaşım *həkimdir* və s.

8.12. Feli xəbərləri ifadə edən formal modellər

Sadə feli xəbərlərin 22-ci cədvəldə göstərilmiş formal əlamətlərini yığcam şəkildə belə ifadə etmək olar:

I. $\check{G}/\emptyset/\pm/MV/+Ş$

II. $0 \check{G}/X/\pm/MV/+D\wedge/+Ş$

1 $\check{G}/X/\pm/MV/+M\wedge Ş/+Ş$

2 $\check{G}/X/\pm/M/+/\wedge P/+Ş$

3 $\check{G}/X/\pm/MV/+VCVX/+Ş$

4 $\check{G}/X/\pm/MV/ /MVZ/+VP/+Ş$

III. $\check{G}/A/\pm/MV/+V/+U$

Burada “U”-”idi”, ”imiş” və “o” ifadələrini əvəz edir.

IV. $\check{G}/V/\pm/MV/+MVL\wedge/+U+Ş$

V. $\check{G}/L/\pm/VC\wedge/+U+Ş$

VI. $\check{G}/d/+MVXDV/+U+Ş$

VII. $0 \check{G}/M_I/\pm/\wedge L/\pm/MV/+DL/+Ş$

1 $\check{G}/M_I/-/\wedge L/\pm/MV/+M\wedge Ş/+Ş$

2 $\check{G}/M_I/+/\wedge L/\pm/M/+VP/+Ş$

3 $\check{G}/M_I/+/\wedge L/\pm/MV/+VCVX/+Ş$

4 $\check{G}/M_I/+/\wedge L/\pm/MV/+/\wedge P/+Ş$

VIII. $0 \check{G}/q/+/\wedge Ş/\pm/MV/+D\wedge/+Ş$

1 $\check{G}/q/+/\wedge Ş/\pm/MV/+M\wedge Ş/+Ş$

2 $\check{G}/q/+/\wedge Ş/\pm/M/+/\wedge P/+Ş$

3 $\check{G}/q/+/\wedge Ş/\pm/MV/+VCVX/+Ş$

4 $\check{G}/q/+/\wedge Ş/\pm/M/+VP/+Ş$

IX. $0 \check{G}/i/+/ D\wedge P/\pm/MV/+M\wedge Ş/+Ş$

1 $\check{G}/i/+/ D\wedge P/\pm/MV/+/\wedge P/+Ş$

2 $\check{G}/i/+/ D\wedge P/\pm/MV/+VCVX/+Ş$

3 $\check{G}/i/+/ D\wedge P/\pm/M/+VP/+Ş$

4 $\check{G}/i/+/ D\wedge P/\pm/MV/-/ D\wedge/+Ş$

X. $0 \check{G}/qy/+/\wedge H/\pm/MV/+D\wedge/+Ş$

1 $\check{G}/qy/+/\wedge H/\pm/MV/+M\wedge Ş/+Ş$

- 2 Ğ/qy/+/\^H/\pm/MV/+/\^P/+Ş
 3 Ğ/qy/+/\^H/\pm/MV/+/\^VCVX/+Ş
 4 Ğ/qy/+/\^H/\pm/M/+/\^VP/+Ş

Yuxarıda göstərilmiş əlamətləri daha yığcam ifadə etmək olar. Bunun üçün inkarlıq şəkilçisini şərti olaraq «İ» hərfi ilə, zaman şəkilçilərini isə «Z» ilə işarə edək. Onda aşağıdakı ifadələri alırıq:

- I. Ğ/\Theta/\pm/MV/+Ş
 II. Ğ/X/\pm İ+Z+Ş
 III. Ğ/A/\pm İ+V/+U+Ş
 IV. Ğ/V/\pm İ+MVL\^/+U+Ş
 V. Ğ/L/+/\^VC\^/+U+Ş
 VI. Ğ/d/+/\^MVXDV/+U+Ş
 VII. Ğ/m_l/+/\^L/\pm İ+Z+Ş
 VIII. Ğ/q/+/\^Ş/\pm İ+Z+Ş
 IX. Ğ/i/+/\^D\^P/\pm İ+Z+Ş
 X. Ğ/qy/+/\^H/\pm İ+Z+Ş

Göstərilən ifadələr daha çox ümumi və mücərrəd olduğu üçün konkret xüsusiyyətləri, bəzi variantları əhatə etmir.

Bütün yuxarıda göstərilənlər morfoloji yolla düzələn sadə feli xəbərlər idi. Bundan başqa felin ayrı-ayrı formaları, şəkilləri və növlərindən biri əsas olmaqla, morfoloji yolla düzələn mürəkkəb qrammatik mənaya malik feli xəbərlər də var.

- I. Qayıdış növün əsasında düzələn feli xəbərlər;
 Ğ/qy/+/\^H/\pm/MV/+/\^MVL\^/+U+Ş
 Ğ/qy/+/\^H/\pm/MVXDV/+U+Ş
 Ğ/qy/+/\^H/\pm/\^VC\^/+U+Ş
 II. Qarşılıq növün əsasında düzələn feli xəbərlər;
 Ğ/q/+/\^Ş/+/\^MVL\^/+U+Ş
 Ğ/q/+/\^Ş/+/\^MVXDV/+U+Ş
 Ğ/q/+/\^Ş/+/\^VC\^/+U+Ş
 III. Məchul növün əsasında düzələn feli xəbərlər;
 Ğ/m_l/+/\^L/+/\^MV/+/\^MVL\^/+U+Ş
 Ğ/m_l/+/\^L/+/\^MVXDV/+U+Ş
 Ğ/m_l/+/\^L/+/\^VC\^/+U+Ş

$\check{G}/m_l/+/^{\wedge}L/+/MVL^{\wedge}/+U+\$$
 $\check{G}/m_l/+/^{\wedge}L/+/^{\wedge}\$/+/MVL^{\wedge}/+U+\$$
 $\check{G}/m_l/+/^{\wedge}L/+/^{\wedge}\$/+/VC^{\wedge}/+U+\$$

Belə feli xəbər növlərinin bəzilərinə Azərbaycan dilində az təsadüf olunur.

İsmi xəbərlər:

I. I +/D^{\wedge}P/ /\$/ /U/ VII. I+\$
 II. S+/D^{\wedge}P/ /\$/ /U/ VIII. S+ \$
 III. Ə+/D^{\wedge}P/ /\$/ /U/ IX. Ə+ \$
 IV. Z+/D^{\wedge}P/ /\$/ /U/ X. Sa+ \$
 V. F(M₀)+ /D^{\wedge}P/ /\$/ /U/
 VI. Sa+/D^{\wedge}P/ /\$/ /U/

Feli xəbərləri ifadə edən formal modellər:

$\check{G}/\Theta/ \pm \dot{I}+\$$
 $\check{G}/X/ \pm \dot{I}+Z+\$$
 $\check{G}/A/ \pm \dot{I}+V \pm U \pm \$$
 $\check{G}/V/ \pm \dot{I}+MVL^{\wedge} \pm U+\$$
 $\check{G}/L/ + VC^{\wedge} \pm U+\$$
 $\check{G}/d/ \pm \dot{I}+MVXDV \pm U+\$$
 $\check{G}/m_l/ + ^{\wedge}L \pm \dot{I}+Z+\$$
 $\check{G}/q/ + ^{\wedge}\$ \pm \dot{I}+Z+\$$
 $\check{G}/i/ + D^{\wedge}P \pm \dot{I}+Z+\$$
 $\check{G}/qy/ + ^{\wedge}H \pm \dot{I}+Z+\$$
 $\check{G}/qy/ + ^{\wedge}H \pm MVLL +U +\$$
 $+ ^{\wedge}H \pm D^{\wedge}P+ \dot{I}+MVL^{\wedge} \pm U+\$$
 $+ ^{\wedge}H \pm \dot{I}+MVXDV \pm U+\$$
 $+ ^{\wedge}H \pm D^{\wedge}P \dot{I}+MVXDV \pm U+\$$
 $+ ^{\wedge}H \pm VC^{\wedge} \pm U \pm \$$
 $+ ^{\wedge}H \pm D^{\wedge}P+ VC^{\wedge} \pm U \pm \$$
 $\check{G}/q/ + ^{\wedge}\$ \pm \dot{I}+MVLL +U +\$$
 $+ ^{\wedge}\$ \pm \dot{I}+MVXDV \pm U+\$$
 $+ ^{\wedge}\$ \pm \dot{I}+VC^{\wedge} \pm U+\$$
 $\check{G}/m_l/ + ^{\wedge}L \pm \dot{I}+MVL^{\wedge} \pm U \pm \$$
 $+ ^{\wedge}L \pm \dot{I}-MVXDV \pm U \pm \$$
 $+ ^{\wedge}L \pm \dot{I}+VC^{\wedge} \pm U \pm \$$

+^L + LŞ ± İ + MVXDV ± U + Ş
 +^L + LŞ ± İ - MVL^ + U + Ş
 +^L + LŞ + VC^ ± U ± Ş
 Ğ/I s/ + D^P + İ + MVL^ ± U + Ş
 + D^P + İ + MVXDV ± U + Ş
 + D^P + ^L + VC^ ± U + Ş
 + D^P + VC^ ± U + Ş

Cədvəl 35. Fəli xəbər şəkilçibrinin tam yazılışı

Ş	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
Z	-ır, -ur, -ür, -ir, -yır, -yur, -yür, -yir, -dı, -di, -du, -dü, -mı, -miş, -muş, -müş -ıb, -ib, -ub, -üb, -yıb, -yib, -yub, -yüb -ar, -ər, -yar, -yər -acaq, -əcək, -yacaq, -yəcək		
V	-a, -ə, -ya, -yə		
MVL^	-malı, -məli		
VC^	-ası, -əsi, -yası, -yəsi		
MVXDV	-maqda, -məkdə		
^L	-ıl, -il, -ul, -ül		
^Ş	-ış, -iş, -uş, -üş		
^H	-i, -in, -un, -ün, -in		
İ	-m, -ma, -mə, -maz, -məz		
U	-idi, -imiş		
D^P D^PT	-dır, -dir, -dur, -dür -dırt, -dirt, -durt, -dürt		

8.13. Mübtədanın formal əlamətləri

Mübtədə cümlədə başqa üzvlərdən asılı olmur, adlıq alda isimlərlə, əvəzləklərlə, 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 (Abdullayev, Seyidov, Həsənova, 1972).

Qeyd etmək lazımdır ki, mübtədanın formal əlamətlərinə görə müəyyənləşdirilməsi alqoritmində bir çox qayda və qanunauyğunluqların nəzərə alınması lazım gəlir.

I QAYDA. Mübtədə həmişə ismin adlıq alında olur: $A/I/X/F/m_2/-I$ mübtədə. *Dəmir* faydalı qazıntıdır.

Lakin bu qayda heç də o demək deyildir ki, bütün adlıq alında olan isimlər mütləq mübtədə olmalıdırlar. Adlıq alında olan isim cümlənin başqa üzvləri yerində də işləyə bilər. Göstərmək lazımdır ki, formal cəhətdən qeyri-müəyyən təsirlik hal adlıq aldan seçilmir.

II QAYDA. Cümlədə aralarında vergül olmayan adlıq halda işlənmiş iki isim varsa və cümlənin xəbəri felin məchul növü ilə ifadə olunmuşsa, onda bu isimdən birincisi təyin, ikincisi mübtədə rolunda çıxış edir.

Bu şərti yığcam şəkildə belə ifadə etmək olar:

$A(I_1 I_2) - X / F / m_1 \} I_1$ -təyin

I_2 -mübtədə

Burada A-adlıq hal, I_1 -birinci isim, I_2 -ikinci isim, X-xəbər, $F/m_1/-$ felin məchul növünü göstərir.

III QAYDA. Əgər cümlədə aralarında vergül olan adlıq halda işlənmiş iki isim varsa, onlar çox vaxt həmcins mübtədə rolunda çıxış edirlər. Bu xüsusiyyətlər durğu işarələrin təhlili sxemləri ilə aşkar olunur.

$A(I_1, I_2) - X / F / m_1 / m_2 / I_1$ - mübtədə

I_2 -mübtədə

Samir, Elçin məktəblidirlər.

IV QAYDA. Cümlədə aralarında vergül olmayan iki isim varsa və xəbər felin məlum növü ilə ifadə olunmuşsa, onda bu iki isimdən birincisi mübtədə, ikincisi tamamlıq olur:

$A(I_1 I_2) — X /F/m_2/ I_1$ - mübtədə

I_2 -tamamlıq

Uşaqlar dəmir daşıyırdılar.

Burada m_2 - felin məlum növü ilə ifadə olunmuş xəbəri göstərir.

V QAYDA. Cümlədə adlıq halında olan bir isim, bir zərf varsa və xəbər felin məlum növü ilə ifadə olunmuşsa, onda isim mübtədə, zərf isə zərflik funksiyası daşıyır:

$A(I Z) — X /F/m_2/ I$ - mübtədə

Z -zərflik

Tələbə yuxarı qalxdı.

VI QAYDA. Cümlədə bir adlıq halında olan zərf varsa, o, mübtədə olur:

$A Z — X [F/m_2] Z$ -zərflik

Yuxarı aşağıdan sərindir.

VII QAYDA. Şəxs əvəzliləri adlıq halında cümlədə yalnız mübtədə olurlar:

$A (\Theta\text{ş}) — X [F/ m_1/ m_2/]$ $\Theta\text{ş}$ -mübtədə

O , mənə nə isə demək istəyirdi.

Azərbaycan dilində fellər şəxslə olduğu üçün çox vaxt mübtədalar şəxs əvəzliləri ilə ifadə olunduqda ixtisar oluna bilər. Belə hallarda fellə ifadə olunmuş xəbərdəki şəxs şəkilsinə əsasən mübtədanı bərpa etmək olar:

$/O/$ Mənə nə isə demək istəyirdi.

İsim, zərf və əvəzlildən başqa, məsdər, sifət və say da cümlədə mübtədə ola bilər:

VIII QAYDA. Adlıq halında məsdər mübtədə funksiyası daşıya bilər.

$A [F/m_0/]$ } m_0 - mübtədə

Yaşamaq yanmaqdır, yanasan gərək. (B.Vahabzadə)

IX QAYDA. Əgər cümlədə eyni zamanda adlıq halında işlənən isim və məsdər varsa, onda isim mübtədə, məsdər isə tamamlıq olur:

A [I F/M₀] —X [F/M₂/] } I- mübtədə

F/M₀/- tamamlıq

Sevil oxumaq istəyirdi.

X QAYDA. Cümlədə adlıq halında işlənən sifətdən sonra isim gəlmirsə, onda sifət mübtədə olur:

A (S) —X [F/ m₁/ m₂/] } S-mübtədə

Tənbəl deyər: -ver yeyim, ört yatım, gözlə canım çıxmasın (məsəl).

Əgər sifətdən sonra isim gəlsəydi, onda sifət təyin olardı. Əslində bu qəbildən olan cümlələrdə işlənən sifətlər aid olduqları isimlərlə yanaşı təsəvvür olunurlar: *Tənbəl (adam) deyər: -ver yeyim, ört yatım, gözlə canım çıxmasın.*

XI QAYDA. Cümlədə adlıq halında olan say ondan sonra isim işlənmədikdə mübtədə, əks təqdirdə təyin olur:

A (Sa) —X [F/ m₁/ m₂/] Sa- mübtədə

Yeddinc i finişə hamıdan tez çatdı.

A (Sa I) —X [F/ m₁/ m₂/] } Sa-təyin

I- mübtədə

Üç yoldaş kəndə yollandı.

Misallardan 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 və tərkibləri ilə də ifadə oluna bilər.

XII QAYDA. Adlıq halında təyini söz birləşmələri xəbərin felin məlum və ya məchul növü ilə ifadə olunub-olunmamasından asılı olmayaraq mübtədə olurlar.

A (TSB) —X [F/ m₁/ m₂/] TSB- mübtədə

Yaşlı kənd sakini ehtiramla baş əydi.

Səfəli dağ kəndi onun ürəyinə yatmışdı.

XIII QAYDA. Adlıq halında olan məsdər tərkibi mübtədə olur.

A [(F/m₀/t) —X [F/ m₁/ m₂/] F/m₀/t/-mübtədə

Ata himayəsinə can atmaq mənə yaraşmaz.

XIV QAYDA. Adlıq halında feli sifət tərkibi bütövlükdə mübtədə hesab olunur.

A [(FSt) —X [F/ m₁/ m₂/] } FSt- mübtədə

Bütün əsirliklərdə çəkdikləri onu əymişdi.

Mübtədanın araşdırılan formal əlamətlərini yığcam şəkildə göstərək (burada M-mübtədədir):

I. A(I) —X /F/T₁/ -I-M

II. A(I₁ I₂) —X /F/T₁/ I₁-T

I₂-M

III. A(I₁ I₂) —X /F/T₂/ I₁-M

I₂-T

IV. A(I₁ I₂) —X /F/T₁ T₂/ I₁-M

I₂-M

V. A(I Z) —X [F/m₂/] I - M

Z - Z_c

VI. A(Z) —X /F/m₂/ Z - M

VII. A (Əş) —X [F/ m₁/ m₂/] Əş-M

VIII. A /F/m₀/ —m₀-M

IX. A [I F/m₀/ —X [F/ m₂/] } I-M

F/m₀-T_m

X. A /S/ —X [F/ m₁/ m₂/] } S-M

XI. A /Sa/ —X [F/ m₁/ m₂/] } Sa-M

A /Sa I/ —X [F/ m₁/ m₂/] } Sa-T

I-M

XII. A (TSB) —X [F/ m₁/ m₂/] TSB-M

XIII. A /F/m₀/t /—X [F/ m₁/ m₂/] F/m₀/t- M

XIV. A [/Fst] —X F/ m₁/ m₂/ Fst- M

Mübtədə və xəbərin formal əlamətlərinin bu cür təsviri onların sintaktik təhlil alqoritmlərində müvafiq blokda yerləşdirilməsi və lazım gəldikdə onlara müraciət etmək imkanını reallaşdırır.

Cümlənin baş üzvlərinin formal təsviri uzlaşma əlaqəsi nəzərə alınmadıqda natamam təsir bağışlayır. Cümlənin təhlili və sintezi prosesində, xüsusilə türk dillərində bu əlaqənin araşdırılması və alınmış nəticələrin formal təsviri AMAİS-də işin keyfiyyətini xeyli yüksəltmiş olur.

8.14. Uzlaşma əlaqəsinin formal təsviri

Azərbaycan dilində mübtəda və xəbəri formal əlamətlərinə görə axtaran sxemlərin qurulması ilə əlaqədar uzlaşma əlaqəsinin formal təsvirinin alqoritmlərdə 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 (Abdullayev, Seyidov, Həsənov, 1972). Mübtəda xəbərə cümlədə uzlaşma əlaqəsinə görə bağlanır. Xəbər mübtədanın tələb etdiyi şəxs və kəmiyyətdə olur. Bunu belə ifadə etmə olar:

$$M/\$n/ - X/\$n/$$

Burada n-I,II,III - şəxsin tək və cəmini göstərir. Bu ümumi ifadə mübtəda 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 də hər iki tərəf bir neçə sözdən ibarət olur. Bu halda üç vəziyyət özünü büruzə verir.

1. Tabe edən bir neçə, tabe olan bir sözlə ifadə olunur:

a) Tabe edən tərəfdə başqa şəxslərə aid sözlərlə yanaşı birinci şəxsə aid əvəzliliklər iştirak edirsə, tabe söz birinci şəxsin cəmi ilə uzlaşır:

$$M(\$1 \$2, \$3) - X (\$1C).$$

Burada $\$1$ -I, $\$2$ -II, $\$3$ -III şəxsin təkini, $\$1C$ - I şəxsin cəmini göstərir.

b) Əgər tabe edən sözlər arasında birinci şəxs əvəzlilikləri iştirak etmirsə, həmin sözlərdən biri ikinci şəxsə aid əvəzlilik (sən, siz) olarsa, tabe söz ikinci şəxsin cəmi ilə uzlaşacaqdır:

$M(\$2, \$3) \rightarrow X(\$2C)$.

v) Tabe edən sözlər ancaq üçüncü şəxsə aid olarsa, tabe söz həmişə üçüncü şəxsə aid olur.

$M(\$3, \$3, \$3) \rightarrow X(\$3C)$.

2. Tabe edən bir, tabe olan bir neçə sözlə ifadə olunur.

a) Tabe olan sözlərin hamısı şəxsə və kəmiyyətə görə uzlaşır:

$M(\$1) \rightarrow (\$1, \$1)$

b) Tabe sözlərdən ancaq sonuncusu əsas sözlə şəxsə və kəmiyyətə görə uzlaşır:

$M(\$1) \rightarrow X(\$0, \$0, \$1)$.

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:

$M(\$3) \rightarrow X(\$1, \$2, \$3 / \$30)$.

$M(\$30) \rightarrow X(\$10, \$20, \$30)$.

Formal təhlil sistemlərində mübtədə və xəbəri əlamətlərinə görə axtaran sxemlərin işini konkret misallarda nəzərdən keçirək. Belə bir cümlənin təhlilinə baxaq:

Elçin dünən anasına məktub yazdı.

Morfoloji təhlil alqoritmlərinin işindən sonra aşağıdakıları müəyyənləşdiririk:

Elçin (isim, adlıq hal, canlı, tək, xüsusi, insan adı və s.)

dünən (zərf, zaman zərfi, adlıq hal, sadə)

məktub (isim, adlıq hal, cansız, tək, ümumi)

anasına (isim, yönlük hal, canlı, tək, ümumi və s.)

yazdı (fel, keçmiş zaman, III şəxs, tək, təsirli və s.).

Morfoloji blokun işindən sonra əldə etdiyimiz bu informasiyalar sintaktik blokun fəaliyyətinə şərait yaradır. Sintaktik blokun işindən sonra aşağıdakıları alırıq. Xəbəri əlamətlərinə görə axtaran sxem göstərir ki, cümlədə xəbər

$\check{G}(X) \pm (MV) + (D \wedge) + \check{S}$ modelinin $\check{G}(X) - (MV) + (D \wedge) + \check{S}$ variantına uyğun gəlir.

Mübtədanın əlamətlərinə görə axtarışı sxemlərinin işini mərhələ-mərhələ izləyək.

$A/\bar{I}-X[\bar{G}(m_1)]$ - modelinə görə cümlədə adlıq halında bir isim varsa, onda o, mübtədə olmalıdır. Lakin bizim konkret misalda adlıq halında olan iki isim vardır. Durğu 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 apardığımız əməliyyatlar nəticəsində məlum olur ki, cümlə sadədir və onun bircə xəbəri var (həmcins xəbərlili cümlə deyil). Onda sintaktik blokda aşağıdakı modelə müraciət etməli oluruq:

$A(I_1, I_2)-X/F/m_2/ \} I_1-M$

$I_2 -Tm$

Buradan belə nəticə çıxır ki, *Elçin* mübtədə, *məktub* isə tamamlıqdır. Məlum olduğu kimi, zərflər də cümlədə mübtədə yerində işləyə bilər. *Dünən* sözünün mübtədə olmadığına aşağıdakı model zəmanət verə bilər:

$A/I Z/ -X[F m_2] I-M$

Yəni cümlədə adlıq halında olan isim və zərf varsa, isim mübtədə olur.

Sintaktik təhlil blokunda köməkçi sözlərin və durğu işarələrinin təhlili sxemlərinin böyük əhəmiyyəti var. Məsələn, mübtədanın müəyyən edilməsi üçün cümlədə heç bir durğu işarəsinin (xüsusən vergülün) olmamasının özü də mühüm, həlledici əlamət sayıla bilər. Əgər bizim verdiyimiz misalda omonim, çoxmənalı və naməlum sözlər müşahidə olunsaydı, onda semantik blok da işə «qarışmalı» olardı. Əgər cümlənin kompüterdə tərcüməsi prosesini izləsək, morfoloji səviyyədə sintezdən sonra hər bir Azərbaycan sözü, özünün ekvivalenti ilə əvəz olunduqdan sonra alarıq:

Yesterday Elchin to his mother the letter wrote.

Sintaktik sintezdə xəbər və mübtədə qrupu, köməkçi və durğu işarələri bir daha araşdırılır, müəyyənləşdirilir və dəqiqləşdirilir.

Sintaktik-semantik sintezə qədər artıq tərcümə edilmiş cümlənin tam sintaktik təhlili başa çatır. İndi qarşıda cümlənin tərcümə olunmuş dildə düzgün sintaktik düzümü məsələsi durur. Bu işdə sintaktik quruluş və sıralanma alqoritmləri mühüm rol oynayır. Belə alqoritmlərdə cümlənin sintaktik quruluşu, sintaktik vahidlərin sıralanması öz əksini tapmalıdır. Sintaktik-semantik sintez alqoritmlərinin işindən sonra cümlənin tam, düzgün və ingilis dili üçün məqbul tərcüməsini əldə edə bilərik:

Yesterday Elchin wrote the letter to his mother.

İndi isə semantik təhlil alqoritmlərinin işini araşdırmaq məqsədilə aşağıdakı cümlənin təhlilini nəzərdən keçirək:

Fikrət ağacdən 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 (fel, əmr şəklinin inkarı, II şəxs, tək, təsirli);

ağacdən (isim, çıxışlıq hal, cansız, tək, ümumi);

dərdi (fe'l, 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 (isim və ya fel) qəbul etməliyik.

Xəbərin əlamətlərinə görə axtaran modellərə görə iki variant alırıq:

$$X [F / \Theta /] \pm (mv) + \S$$

$$X [F / X /] - (mv) + / D \wedge / + \S$$

Bizim cümlə üçün bu model aşağıdakı variantda işləyir:

$$X [F / \Theta /] + (mv) + O$$

$$X [F / X /] - (mv) + / D \wedge / + O$$

Semantik təhlil sxemi sintaktik təhlillə qarşılıqlı əlaqə sayəsində bu omoqraflardan birini düzgün hesab edib, bu barədə morfoloji təhlil blokuna məlumat verməlidir. Sintaktik bloka qayıdaq.

Əgər biz *alma* sözünün isim nitq hissəsi olduğunu hələ qəbul etməmişiksə, onda mübtədani əlamətlərinə görə axtaran aşağıdakı model üzrə mübtədani tapmaq imkanı yaranır:

$$A /I/ \text{---} X [F/ m_1/ m_2/ I] I - M$$

Bu modelə görə *Fikrət* sözünün mübtədə olduğu aşkar olur. Mübtədə ilə xəbərin uzlaşma əlaqəsinə görə alırıq:

$$\text{Əgər } M(\text{Ş}_3) \text{---} X(\text{Ş}_1, \text{Ş}_2, \text{Ş}_3) \text{.- sə}$$
$$\text{onda } M(E_1) \text{---} X(E_1).$$

Bizim tədqiq etdiyimiz cümlədə xəbərin III şəxsin təkində olması tələb olunur:

$$M(E_3) \text{---} X(E_3)$$

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» bilir. Birinci variant isə bu sözün isim olduğunu göstərir:

alma (isim, adlıq hal, cansız, tək, ümumi, sadə).

Bu əlamətlərə görə onu mübtədə da hesab etmək olardı (kompüter qeyri-müəyyən təsirlik halını adlıq haldan seçə bilmir:

$$A /I/ \text{---} X [F/ m_1/ m_2/ I] I - M$$

modeli cümlədə mübtədəyə 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.

$$A(I_1 I_2) \text{---} X [F/m_1 m_2/] I_1 - M$$
$$I_2 - T_m$$

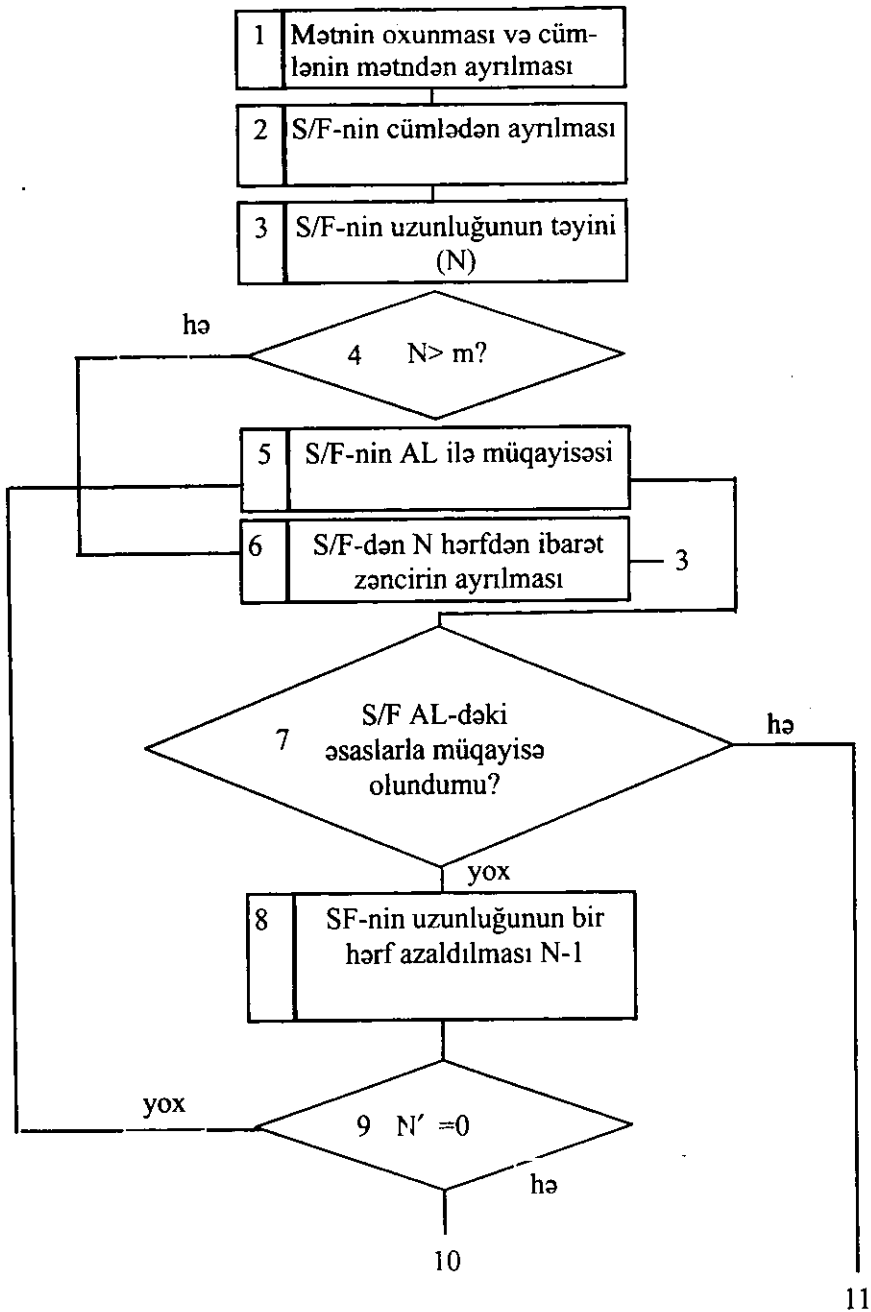
modelinə əsasən *Fikrət* sözünün mübtədə, *alma* sözünün isə tamamlıq olduğu aşkar olur. Analoji qaydada sintaktik sıralanma və uzlaşma əlaqələrinin nəzərə alınması ilə *dərdi* omoqrafının isim (təsirlik alında) deyil, fel (fel, keçmiş zaman, xəbər) olduğu müəyyənləşir.

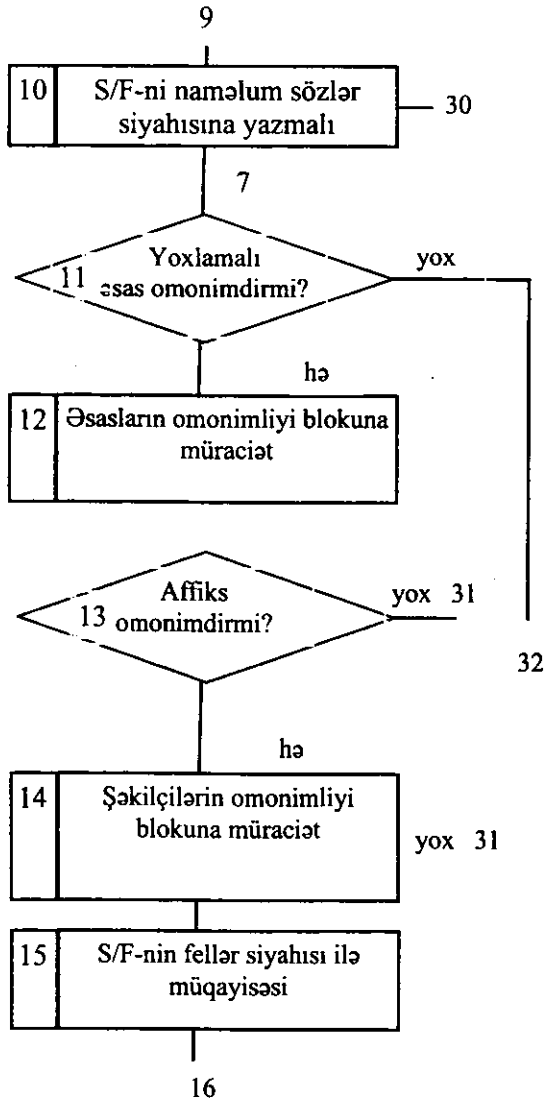
Bununla da təhlil bloklarının işi bitmiş esab olunur. Sintez proqramları təhlil bloklarına əsaslanaraq morfoloji, sintaktik, semantik səviyyədə sintezi həyata keçirir. Semantik təhlil bloku

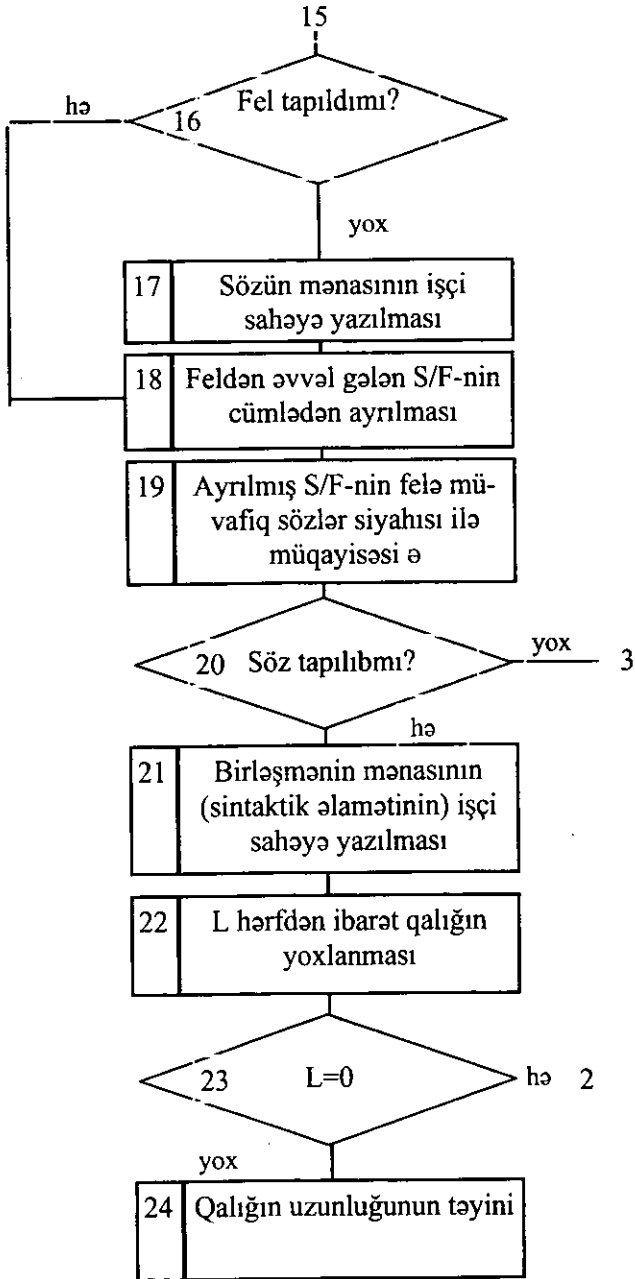
sintez sxemlərinin işini xeyli asanlaşdırır. Belə ki, MT-də tərcümə olunacaq dildə polisemiyanı müəyyənləşdirməyə, omonim variantlarının seçilməsinə ehtiyac qalmır. Sintezin keyfiyyəti təhlil alqoritmlərinin keyfiyyətindən çox aslıdır. AMAİS-in semantik blokunda bu məsələlər öz geniş izahını tapacaqdır.

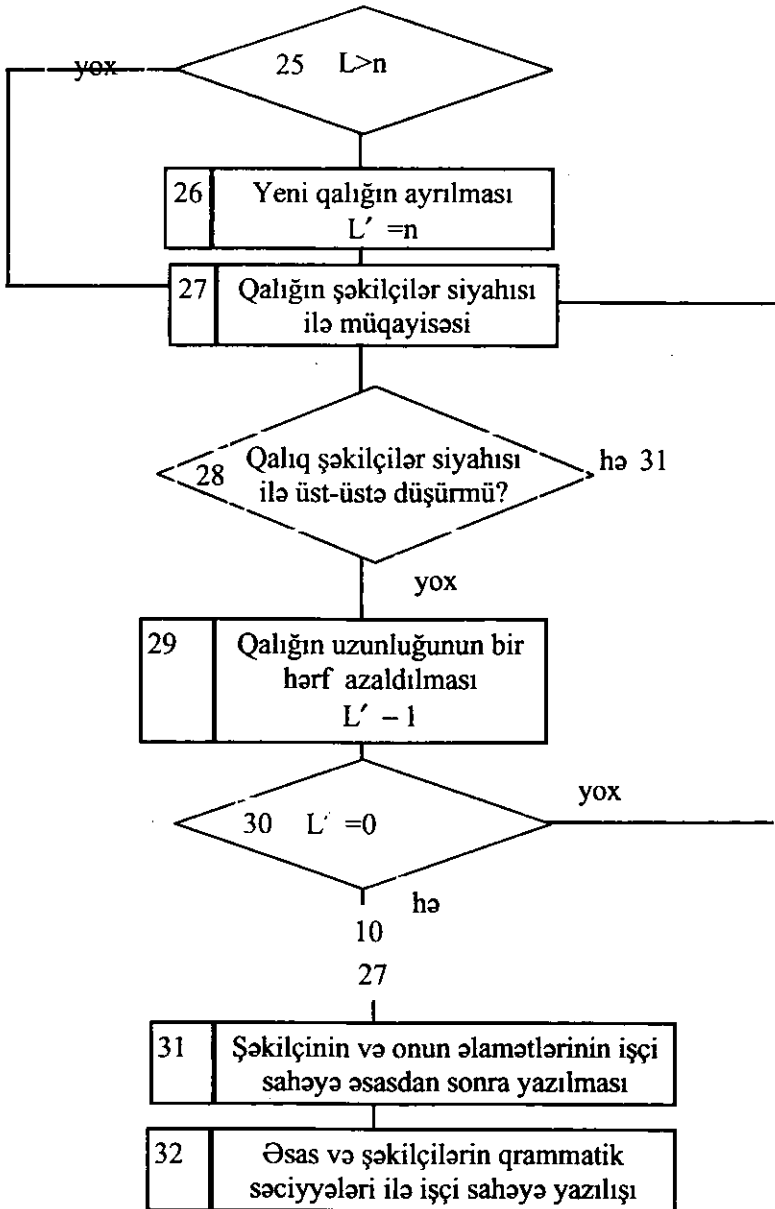
8.15. Azərbaycan dilinin avtomatik sintaktik təhlili alqoritmi (sxem 13)

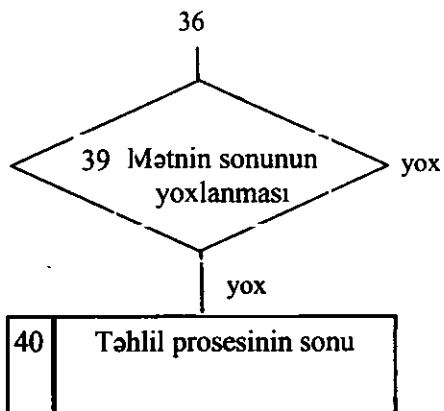
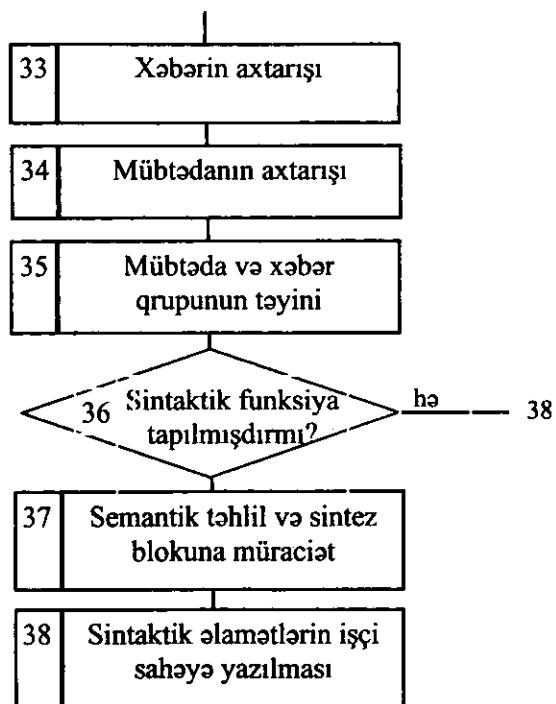
1. Mətnin kompüterə daxil edilməsi, oxunması, cümlənin mətndən ayrılması və 2-yə keçid;
2. Söz-formanın cümlədən ayrılması və 3-ə keçid;
3. Söz-formanın uzunluğunun təyini (N);
4. $N > m$ şərtinin yoxlanması: əgər $N > m$ -dirsə 6-ya, əks halda 5-ə keçid;
5. Söz-formanın avtomatik lüğətin (AL) müvafiq qrupu ilə müqayisəsi: əgər söz-forma tapılıbsa 11-ə, əks halda 8-ə keçid;
6. Söz-formanın ayrılması ($N = m$) və 7-yə keçid;
7. Söz-formanın AL-in əsaslar siyahısı ilə müqayisəsi; Əgər söz tapılıbsa 11-ə, əks halda 8-ə keçid;
8. Söz-formanın uzunluğunun bir hərf azaldılması ($N - 1$);
9. $N = 0$ şərtinin yoxlanması: Əgər $N = 0$ -dirsə. 10-a, əks təqdirdə 5-ə keçid;
10. Söz-formanın naməlum sözlər siyahısına yazılması və 30-a keçid;
11. Əsasın omonimliyi şərtinin yoxlanması: əgər omonim-dirsə 12-yə, əks təqdirdə 32-yə keçid;
12. Əsasların omonimliyi blokuna müraciət və 32-yə keçid;
13. Şəkilçinin omonimliyi şərtinin yoxlanması: əgər omonimdirsə 14-ə, əks təqdirdə 31-ə keçid;
14. Şəkilçilərin omonimliyi blokuna müraciət və 31-ə keçid;











Sxem 13. Azərbaycan dilinin sintaktik təhlil alqoritminin blok sxemi

15. Söz-formanın fellər siyahısı ilə müqayisəsi və 16-ya keçid;
16. Felin tapılma şərtinin yoxlanması:əgər fel siyahıda tapılmışsa 18-ə, əks təqdirdə 17-yə keçid;
17. Sözüün mənasının işçi sahəyə yazılması və 18-ə keçid;
18. Feldən əvvəl gələn söz-formanın cümlədən ayrılması və 19-a keçid;
19. Ayrılmış söz-formanın felə müvafiq sözlər siyahısı ilə müqayisəsi və 20-yə keçid;
20. Sözüün AL-də tapılması şərtinin yoxlanması: Əgər söz tapılmışsa 21-ə, əks halda 3-ə keçid;
21. Birləşmənin mənasının (sintaktik əlamətinin) işçi sahəyə yazılması və 22-yə keçid;
22. L-hərfdən ibarət qalığın varlığının yoxlanması və 23-ə keçid;
23. $L=0$ şərtinin yoxlanması: Əgər $L=0$ -dırsa 2-yə, əks halda 24-ə keçid;
24. Qalığın uzunluğunun (L) təyini və 25 şərtinə keçid;
25. $L>n$ şərtinin yoxlanması: əgər $L>n$ -dırsa 26-ya, əks halda 27-yə keçid;
26. Yeni qalığın ayrılması ($L' =n$) və 27-yə keçid;
27. Qalığın şəkilçilər siyahısı ilə müqayisəsi və 28-ə keçid;
28. Qalığın şəkilçilər siyahısında tapılması şərtinin yoxlanması; tapılmışsa 31-ə, əks halda 29-a keçid;
29. Qalığın uzunluğunun bir hərf azaldılması ($L-1$) və 30-şərtinə keçid;
30. $L=0$ şərtinin yoxlanması: əgər $L=0$ -dırsa, 10-a, əks halda 27-yə keçid;
31. Şəkilçinin və onun əlamətlərinin işçi sahəyə əsasdan sonra yazılışı və 32-yə keçid;
32. Əsas və şəkilçilərin qrammatik səciyyələri ilə işçi sahəyə yazılışı və 33-ə keçid;
33. Xəbərin əlamətlərinə görə axtarışı və 34-ə keçid;
34. Mübtədanın əlamətlərinə görə axtarışı və 35-ə keçid;

35. Mübtədə və xəbər qrupunun müəyyənləşdirilməsi və 36-şərtinə keçid;

36. Cümlə üzvlərinin sintaktik funksiyalarının tapılması şərtinin yoxlanması: əgər tapılmışsa, 38-ə, əks alda 37-yə keçid;

37. Semantik təhlil və sintez blokuna müraciət 38-ə keçid;

38. Cümlə üzvlərinin sintaktik əlamətlərinin işçi sahəyə yazılması və 39 şərtinin yoxlanması;

39. Mətnin sonunun yoxlanması: əgər mətn bitmişsə, 40-a, əks halda 1-ə keçid;

40. Təhlil prosesinin sonu.

Burada N söz-formaların uzunluğu, m-əsasların maksimal uzunluğu, L-şəkilçi zəncirinin uzunluğu, n-şəkilçilərin maksimal uzunluğu, s/f -söz-formanı göstərir.

Azərbaycan dilinin avtomatik sintaktik təhlil alqoritmini misal üzərində nəzərdən keçirək. Sadəlik üçün *hesab etmək* birləşməsinin sintaktik təhlil prosesini izləyək. Hər şeydən əvvəl cümlənin sonundan başlayaraq söz-formanı ayırıb 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 məqsədilə fərz edilir ki, fel infinitiv formasında təmsil olunmuşdur). Göstərilən sözün morfoloji, sintaktik əlamətlərini və tərcüməsini alırıq. Tapılmış sözün müstəqil və ya söz birləşməsinin komponenti olub-olmaması yoxlanılır. Bu məqsədlə söz fellər siyahısı ilə müqayisə olunur. Tapırıq ki, bu söz fellər siyahısında birləşmənin ikinci tərəfi kimi qeyd olunmuşdur. Sonrakı mərhələdə cümlədə həmin feldən 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 təhlil olunur.

Avtomatik morfoloji və sintaktik təhlil alqoritmlərinin işi semantik blokla sıx əlaqəlidir. Belə ki, omonimlik və çoxmənalılıq yalnız semantik blokun işindən sonra tam şəkildə həll oluna bilər. Ayn-ayrı cümlələrin təhlili nümunələri də semantik blokun iş mexanizminin izahından sonra daha anlaşılıqlı və aydın şərhini tapır.

IX FƏSİL. MƏTNLƏRİN FORMAL TƏHLİLİ SİSTEMİNİN SEMANTİK BLOKU

9.1. Məsələnin ümumi qoyuluşu. Sintaktik əhatədə çoxmənalılığın və omonimliyin aradan qaldırılması

Mətnin formal təhlili sisteminin ən mühüm komponentlərindən biri semantik blokdir. Semantik blokun məzmunu, quruluşu, məqsəd və vəzifələri, mahiyyəti, ümumiyyətlə, lingvistik vahidlərin semantik xüsusiyyətləri tədqiqatçıların həmişə diqqət mərkəzində olmuş, mübahisə və diskussiya mövzusunə çevrilmişdir (Алекперов, 1983; Городецкий, 1973,1976; Здоровов, 1976; Звегинцев, 1973; Кибрик, 1973, 1965; Кравчук, 1969; Марчук, 1979; Новицкая, 1979; Пашенко, 1979; Розенцвейг, 1979; Труб, 1979; Шаляпина, 1976; Шовкун, 1973 və s.)

Dil səviyyələri bir-birindən təcrid olunmuş şəkildə deyildir. Əksinə, dilin müxtəlif səviyyələri qarşılıqlı əlaqədə və münasibətdədirlər və aşağı səviyyənin çatışmazlıqları daha yuxarı səviyyələrdə tamamlanır. Şübhəsiz, təbii dilin səviyyələrə bölünməsi şərtidir və dilin mənimsənilməsi, onun mahiyyətinin dərkinə xidmət edir. Lakin nəzərə almaq lazımdır ki, bütövlükdə dil sistemdir və səviyyə çərçivəsində qapanıb qalır.

Məlum olduğu kimi, məsələdə lingvistik təhlil morfoloji səviyyədən başlanır. Morfemlər haqqında (kök və şəkilçi) ətraflı məlumat əldə elədikdən sonra sintaktik təhlil mərhələsi başlanır. Çünki sintaktik təhlil üçün qabaqcadan nitq hissələri və sözlərin qrammatik kateqoriyaları məlum olmalıdır. Sintaktik təhlildə cümlə ayrılır, onu təşkil edən söz-formalar arasında sintaktik əlaqələr müəyyənləşdirilir, cümlə üzvləri, uzlaşma əlaqəsi araşdırılır. İstər morfoloji, istər sintaktik səviyyələrdəki «həll olunmamış problemlər» semantik blokda araşdırılır. Sözlərin semantik təbiəti heç də həmişə morfoloji və sintaktik əlamətlərlə səciyyələnmir. Morfoloji və sintaktik səviyyələr əslində sözün tam mənasını açmaqda semantik bloka yardım edirlər. Maraqlıdır ki, bəzən dilin morfoloqiyası və sintaksisini bilmədən də kontekstdəki söz-

lərin mənasına görə sözlərin qrammatik əlaqələrini bərpa etmək mümkün olur. Kontekstdəki sözlərin mənalарına görə cümlə konstruksiyaları da qurmaq mümkündür. Buradan belə nəticə çıxarmaq olar ki, təbii ünsiyyət prosesində morfologiyadan sintaksisə, oradan da semantikaya deyil, əksinə semantikadan birbaşa sintaksisə keçid mümkündür. Burada insanın həyat təcrübəsi, intuisiya və bir çox başqa ekstralinqvistik amillər köməyə gəlir.

Kompüterlərdə canlı insana məxsus bu keyfiyyətlər olmadığı üçün semantikanı morfologiya və sintaktik səviyyələrsiz 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 «işləməlidir».

Semantik problemin həlli, hər şeydən öncə, semantik vahidlərin ayrılması kriteriyası, onların statusu və s. bu kimi məsələlərin aydınlaşdırılması zərurətini doğurur. Mətnlərin tam mənasını açmağa xidmət edən semantik vahidlərin müəyyənləşdirilməsi və onların mətnlərdəki 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 misallara baxaq.

Əgər biz hər hansı bir sintaktik əlaqə 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, boybuxunun 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 əlaqəyə 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 allarda doğrultmur. Bu və ya digər söz müxtəlif distribusiyalarda, müxtəlif sintaktik-semantik əlaqə formalarında olur. Bəzən bir distribusiya başqa bir distribusiya ilə izah olunur və bu problemi bir qədər də çətinləşdirir.

Elementar diferensial əlamətin göstərilməsi bizə nə verə bilər?

Lügət vahidi kimi sözlərin səciyyələndirilməsində qeyd olunmuşdu ki, sözlərin həqiqi mənası sintaktik əhatədə məlum olur. Bu baxımdan sözün həqiqi mənası bəzən sözün ənənəvi mənasından kiçik və böyük ola bilər. Müəyyən diferensial mənaya malik ən kiçik vahid kimi monosema götürülə bilər (Зализняк, 1963). 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 birləşmələri bəhsində «təyini söz birləşməsi» eyni məna bildirir və bu qəbildən olan sözləri avtomatik referatlaşdırma prosesində makrosema hesab etmək olar. Tədqiqat işində birləşmələr AL-də sözlərin feldən asılı olaraq qruplara ayrılması prosesini makrosemalara bölünmə kimi qəbul etmək olar.

Mətn informasiyalarının avtomatik işlənmə sistemlərində təhlilinin semantik problemlərindən bəhs edərkən B.Y.Qorodetski semantik inventarlaşdırma məsələlərini ön plana çəkir. Semantik inventarlaşdırma aşağıdakı istiqamətdə aparılır:

- leksik-semantik səviyyələrdə inventarlaşdırma;
- morfo-semantik səviyyələrdə inventarlaşdırma.

Göstərilən səviyyələrdə paradigmatik və sintaqmatik aspektlər mütləq nəzərə alınır. Semantik inventarlaşdırma prosesində söz və morfemlərin, onların müəyyən birləşmələrinin tezauruslarının tərtibi mühüm amil hesab edilir (Городецкий, 1976). Tədqiqat işində omonimliyin və çoxmənalılığın aradan qaldırılması alqoritmlərinin tərtibində sintaktik əlaqə və tezauruslar sistemindən istifadə olunmuşdur.

Göstərmək lazımdır ki, nəinki ayrıca götürülmüş söz və ya söz birləşməsi, hətta bütövlükdə cümlənin də mənası tam şəkildə kontekstdə açılır. Ona görə də mətnlərin avtomatik işlənməsi ilə bağlı mətn semantikasi məsələlərinin ertərəfli araşdırılması mühüm əhəmiyyət kəsb edir (Колшанский, 1980). Yeni nəsillə kompüterlərin getdikcə daha da inkişafı və təkmilləşdirilməsi sistemin semantik blokunun yeni baxımdan izahı, təkmilləşdirilməsi və qurulması imkanı yaradır.

9.2. Morfoloji səviyyədə əsas və şəkilçilərin omonimliyi

Morfoloji səviyyədə, kontekstdə götürülməmiş hər hansı bir omonim əsasın 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 müəyyənləşdirilə 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 kənarında verilə bilər. Yəni fonetik səviyyədə öz həllini tapmayan hər hansı bir linqvistik problem morfoloji

səviyyədə, morfoloji səviyyədə izah olunmayan linqvistik vahidlər isə sintaktik və semantik səviyyələrdə şəhr oluna bilər. Bu baxımdan morfoloji səviyyədə aradan qaldırıla bilməyən omonimlik sintaktik və semantik blokların müdaxiləsindən sonra öz izahını tapa bilər. Gödelin natamamlıq fərziyyəsi bu baxımdan linqvistika üçün əhəmiyyətlidir (Гёдел, 1981; Успенский, 1982).

Əsasların və şəkilçilərin omonimliyi AL-də və şəkilçilər siyahısında öz əksini tapır. Məsələn, əgər əsas omonimdirsə, onda onun mümkün nitq hissələrinə mənsubluğu qeyd olunur: yaz - «pişi», «vesna» - fel/isim. Yaz sözünün hansı nitq hissəsinə mənsubluğunu linqvistik təhlilin daha yüksək - sintaktik və semantik səviyyələrində tam aydınlaşdırmaq mümkündür, çünki omonimlik və çoxmənalılıq - birinci növbədə semantik problemlər çərçivəsində həll olunmalıdır (Коптилов, 1962, 76; Ингве, 1957).

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* – бросать, лошадь - şəkilçisiz həm fel, 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 alda *at* əsası zaman şəkilçiləri qəbul edir ki, bu da əsasın fel olduğunu göstərir.

İkinci alda *at* - hal və mənsubiyyət kateqoriyaları şəkilçiləri ilə işlənir. Bu şəkilçiləri isə yalnız ad nitq hissələrinə aid olan sözlər qəbul edirlər. Beləliklə, misallardan görüldüyü kimi, əsaslara birləşən şəkilçilərə görə onların omonimliyini aradan qaldırmaq mümkündür.

Azərbaycan dilində elə omonim şəkilçilərə rast gəlmək olur ki, onlar həm fellərə, həm də adlara qoşularaq müxtəlif qrammatik funksiyalar yerinə yetirirlər. Məsələn, -im (-im, -um, -üm) şəkilçiləri fel əsaslarından sonra əmr şəklini, isimlərdən sonra isə mənsubiyyəti göstərir.

- yazım - «напишу-ка я»

- kitabım - «МОЯ КНИГА»

Belə hallarda şəkilçilərin omonimliyi əsasa görə aradan qaldırılır. Yuxarıda göstərilənləri nəzərə alaraq omonimliyin aradan qaldırılması üçün formal qaydalar vermək olar.

9.3. Omonimliyin formal əlamətə görə aradan qaldırılması

Burada iki halı fərqləndirmək lazımdır:

a) şəkilçilərə görə əsasların omonimliyinin aradan qaldırılması;

b) əsaslar vasitəsilə şəkilçilərin omonimliyinin aradan qaldırılması.

Əsasların omonimliyinin aradan qaldırılması üçün aşağıdakı formal qaydaları qəbul edirik:

QAYDA 1.1. Əgər əsas omonim (AF), şəkilçi isə ad nitq hissəsi kateqoriyalarına aiddirsə (a), onda əsas da ad nitq hissəsinə aiddir (A).

Burada A- ad nitq hissəsini, F- feli, a- ad nitq hissəsinə aid şəkilçiləri, f isə fel nitq hissəsinə aid şəkilçiləri göstərir.

QAYDA 1.2. Əgər əsas omonim (AF), şəkilçi isə felə aiddirsə (f), onda əsas fel nitq hissəsinə mənsubdur (F).

QAYDA 1.3. Əgər həm əsas (AF), həm də şəkilçi (af) omonimdirsə, onda əsasın omonimliyi morfoloji səviyyədə aradan qaldırılma bilməz. Bu qayda əsasdan sonra yalnız bir şəkilçi gəldikdə doğrudur. Əgər omonim şəkilçidən sonra yenə şəkilçilər gəlsə, onda omonimliyi aradan qaldırmaq üçün omonim şəkilçidən sonra gələn şəkilçiyə müraciət etmək lazım gəlir.

Əgər omonim əsasdan sonra yalnız bir omonim şəkilçi gəlsə, onda omonimlik aradan qalxmır və əsas üçün onun bütün mümkün qrammatik səciyyələri qeyd olunur.

Analoji olaraq, şəkilçilərin omonimliyinin də aradan qaldırılması üçün formal qaydalar vermək olar:

QAYDA 2.1. Əgər affiks omonimdirsə (af), əsas ad nitq hissəsinə aiddirsə (A), onda şəkilçi də adlara (a) aiddir.

QAYDA 2.2. Əgər şəkilçi omonimdirsə (af), əsas isə felə aiddirsə (F), onda şəkilçi də fel nitq hissəsinə (f) aiddir.

QAYDA 2.3. Əgər həm şəkilçi (af), həm də əsas omonimdirsə (AF), onda şəkilçinin omonimliyi morfoloji səviyyədə aradan qaldırıla bilməz (bu o halda doğrudur ki, şəkilçi yeganədir, ondan sonra şəkilçi gəlmir. Əgər şəkilçidən sonra yenə şəkilçi gəlsə, onun hansı nitq hissəsinə mənsubluğuna görə omonimlik aradan qaldırıla bilər).

Qaydalarda göstərilənləri cədvəl şəklində vermək olar (cədvəl 36).

2-3 qaydasına görə omonim şəkilçi üçün onun mümkün qrammatik səciyyələri qeyd olunur. Məsələn, şəkilçilər siyahısında *-im (-im, -um, -üm), -in (-in, -un, -ün)* şəkilçiləri üçün onların həm ad, həm də fel nitq hissələrinə xas xüsusiyyətləri əks olunmuşdur.

Morfoloji səviyyədə əsas və şəkilçilərin omonimliyinin aradan qaldırılması təkcə yuxarıda göstərilən qaydalarla məhdudlaşmır. Morfoloji səviyyədə omonimlik probleminin həllinin digər istiqamətlərini nəzərdən keçirək.

9.4. Şəkilçi birləşmələri vasitəsi ilə omonimliyin aradan qaldırılması

Omonimlik yanaşı işlənən şəkilçi birləşmələrinin köməyi ilə də aradan qaldırıla bilər. Belə birləşmələri maşın affiksləri də hesab etmək olar.

Cədvəl 36. Əsasların və şəkilçilərin omonimliyinin aradan qaldırılması

Omonim əsas və ya şəkilçi	Əsas və ya şəkilçi	Əsasın və şəkilçinin omonimliyinin aradan qaldırılması (düzgün variantın seçilməsi)
AF	a	A
AF	f	F
AF	af	AF
af	A	a
af	F	f
af	AF	af

Məsələn, feli sifət şəkilçisi *-yacağımız*, *-yacaq* və *-ımız* şəkilçilərinin birləşməsindən ibarətdir. Birləşmənin birinci tərəfi *-yacaq* felin qəti gələcək zamanı, ikinci tərəfi *-ımız* isə adlara (isimlərə) xas olan mənsubiyyət şəkilçisidir. Birləşmə şəklində bu şəkilçilər feli sifəti göstərir.

Qeyd etmək lazımdır ki, bu şəkilçi birləşməsi Azərbaycan dilinin qrammatikasında da feli sifət şəkilçisi kimi təsvir olunmuşdur (Hüseynzadə, 1983, 217).

-diğ – mənsubiyyət şəkilçisi tipli feli sifət şəkilçilərinə gəldikdə isə, bu şəkilçinin komponentlərinə ayrılmış şəkildə verilməsi daha məqsədəuyğun hesab olunur. Qrammatika kitablarında təsbit olunmasa da, *-diğ* (*-duq*, *-diy*, *-düy*) maşın şəkilçisi kimi verilmişdir. İstisna olaraq bu qəbildən olan şəkilçilərdən *-diğın* (*-duğun*, *-diyın*, *-düyün*) birləşmə şəklində verilir.

Qrammatik səciyyələri: feli sifət, II şəxs, tək hal. Bu şəkilçinin komponentlərinə ayrılması təhlil prosesini bir qədər də mürəkkəbləşdirə bilər.

Mənsubiyyət şəkilçisi - hal şəkilçisi birləşməsi - məsələn, -sının (*-sunun*, *-sünün*, *-sinin*) omonimliyin aradan qaldırılması ilə bağlı olaraq qəbul olunmuşdur. Formal təhlillə əlaqədar bu şəkilçilərin bir hissəsi *-sın* (*-sun*, *-sin*, *-sün*) analoji formaya malik əmr şəklinin şəkilçiləri ilə qarışdırıla bilər. Analoji olaraq qeyri-qəti gələcəklə keçmiş zaman şəkilçiləri də birləşmə şəklində verilmişdir: məsələn, *-armış* (*-ərmış*, *-yarmış*, *-yərmış*), *-ardı* (*-ərdi*, *-yardı*, *-yərdi*) və s.

9.5. Maşın şəkilçiləri vasitəsi ilə omonimliyin aradan qaldırılması

Əsas və şəkilçilərin omonimliyi həmçinin ənənəvi morfoloqiyalarda öz əksini tapmamış xüsusi maşın şəkilçiləri vasitəsilə də aradan qaldırıla bilər:

a) *-mır* (*-mur*, *-mir*, *-mür*). İnkər şəkilçisi *-ma* (*-mə*) ilə indiki zaman şəkilçilərinin *-ır* (*-ir*, *-ur*, *-ür*) birləşməsi nəticəsində yaranır.

b) *-mısan* (*-musan*, *-mısən*, *-müsən*). Keçmiş zaman şəkilçisi *-miş* (*-muş*, *-miş*, *-müş*) ilə II şəxsin təkini bildirən *-san* (-

sən) şəkilçisinin birləşməsi ilə düzəlir. Burada birləşmənin birinci komponentindəki -ş samitinin düşməsi nəticəsində qalan -mı (-mu, -mi, -mü) sual şəkilçisi ilə üst-üstə düşə bilər;

v) *idi* və *imiş* bağlamalarının və *ilə* qoşmasının şəkilçi formaları da müstəqil şəkilçi kimi siyahıda təmsil olunmuşdur:

-ydi (-ydi), -ymış (-ymış), -yla (-ylə)

Bütün yuxarıda göstərilənlər şəkilçilər siyahısında əks olunmuşdur.

9.6. 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 dil səviyyələrində öz həllini tapır.

Aşağıdakı hallarda şəkilçi və əsasların omonimliyi aradan qaldırıla bilməz:

a) əgər omonim əsas şəkilçisiz işlənirsə:

at - «бросать» / «лошадь» - fel /ad;

b) 1.3. və 2.3. allarında - əsas və şəkilçi eyni vaxtda omonim olduqda:

Məsələn, *at* (fel/ad) U -im (əmr və ya mənsubiyyət) I şəxs (tək) - iki cür başa düşülə bilər:

atım «*моя лошадь*» və ya «*бросать бы мне*»

Bu mənalardan hansının doğru olduğu yalnız kontekst daxilində, təhlilin daha yüksək - sintaktik və semantik mərhələlərinin köməyi ilə müəyyənləşdirilə bilər.

Şəkilçi siyahısında həm feli, 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 feli və ismi əsaslar üçün doğru olduğu göstərilir (cədvəl 37). Məsələn, *-im* (*-um, -im, -üm*) omonim şəkilçisi BALØ kimi kodlaşdırılmışdır. Kontekst daxilində nəzərdən keçirildikdə cədvəldə qeyd olunmuş ÜALQ və RALC kodlarından birini alır. Bu kodlar göstərir ki, əgər bu şəkilçi feli əsasa birləşmişsə, onda əmr şəklinin I şəxs, tək halını (ÜALQ), ismi

əsasdan sonra gəlsə, 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 omonimliyinin aradan qaldırılması alqoritmini qurmaq olar. 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.

9.7. Əsasların omonimliyinin aradan qaldırılması alqoritmi

1. Əsasdan sonra gələn şəkilçinin mövcudluğunun yoxlanması; əgər şəkilçi varsa 2-yə, əks əldə 7-yə keçməli;
2. Əsasdan sonra gələn şəkilçini ayırmalı və 3-ə keçməli;
3. Yoxlamalı: əgər şəkilçi omonimdirsə 1-ə, əks təqdirdə 4-ə keçməli;

Cədvəl 37. Omonim şəkilçilər

Omonim şəkilçilər və onların ümumi kodları		Mümkün qrammatik səciyyələr	Kodlar
-dır (-dur -dir, -dür)	1 ALφ	Əmr şəkli, 3 şəxs, tək	FLYQ
-a (-ə, -ya, -yə)	2 φ φ φ	Xəbərlik, 3 şəxs, tək	LLYC
		Arzu şəkli	Kφ φ Q
-in (-in, -un, -ün)	9+Lφ	Yönlük hal	Hφ φ C
		Əmr şəkli, 2 şəxs, tək	Ü+LQ
-im (-um, -im, -üm)	BALφ	Yiyəlik hal Mənsubiyyət, 2 şəxs, tək	Hφ φ C R+LC
		Əmr şəkli, 1 şəxs, tək	ÜALQ
		Mənsubiyyət, 1 şəxs, tək	RALC

4. Yoxlamalı: əgər şəkilçi adlara aiddirsə 5-ə, əks halda 6-ya keçməli;

5. Əsas adlara aiddir, 8-ə keçməli;

6. Əsas fellərə aiddir, 8-ə keçməli;

7. Morfoloji səviyyədə əsasın omonimliyi aradan qaldırıla bilməz - sintaktik və semantik bloklara müraciət;

8. Əsas haqqında müvafiq qrammatik informasiyanın işçi sahəyə yazılması (sxem 12);

Misal olaraq *atımdan* «от моей лошади» söz-formasının təhlilini nəzərdən keçirək. Təhlil prosesində *atımdan* - aşağıdakı morfemlərə ayrılmış olur:

at ∪ im ∪ dan. Alınmış *at* əsasının omonimliyinin aradan qaldırılması tələb olunur. Bu aşağıdakı kimi yerinə yetirilir:

1. *At* - əsasından sonra şəkilçi olub-olmadığını yoxlayırıq. Konkret misala görə şəkilçi gəldiyi üçün 2-yə keçirik.

2. *At* - əsasından sonra gələn şəkilçini (-im) ayırıb 3-ə keçirik.

3. Yoxlayırıq: şəkilçi omonimdirmi? Bizim misalda -im şəkilçisi omonimdir. Onda 1-ə keçirik. 1-ə görə əsasdan sonra yenə şəkilçi gəlib-gəlmədiyini müəyyənləşdiririk. Əgər -im şəkilçisi omonim əsasdan sonra gələn yeganə şəkilçi olsaydı, onda alqoritm öz işini 7-yə verməli olardı. Morfoloji səviyyədə əsasın omonimliyi aradan qaldırılmadığı üçün sintaktik və semantik bloklara müraciət etmək lazım gələrdi. Lakin omonim əsas (*at*) və ondan sonra gələn omonim şəkilçidən (-im) sonra yenə şəkilçi gəldiyi üçün (-dan) 2-yə keçirik.

4. 2-yə görə omonim şəkilçidən (-im) sonra gələn şəkilçini (-dan) ayırıb yenidən 3-ə keçirik.

5. 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.

6. 4-ə görə şəkilçini adlara və ya felə aid olduğunu yoxlayırıq. Şəkilçi felə aid olsaydı, 6-ya keçməli idik. Lakin şəkilçi adlara aid olduğu üçün 5-ə keçirik.

7. 5-in şərtinə görə əsasın adlara aid olduğu müəyyənləşdirilir və 9-ə keçilir.

8. 8-ə görə əsas haqqında müvafiq qrammatik informasiya işçi sahəyə yazılır: *Atımdan* - söz-formasında *at* - əsasının qrammatik səciyyələri aşağıdakılardır:

at - isim, adlıq hal, tək 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.

9.8. Şəkilçilərin omonimliyinin aradan qaldırılması algoritmi

1. Yoxlamalı: əgər şəkilçiyə birləşmiş əsas omonimdirsə, 5-ə, əks halda 2-yə keçməli;

2. Yoxlamalı: əgər əsas ad nitq hissəsinə aiddirsə, 3-ə, əks halda 4-ə keçməli;

3. Şəkilçi adlara aiddirsə, 6-ya keçməli;

4. Şəkilçi felə aiddirsə, 6-ya keçməli;

5. Şəkilçinin omonimliyi morfoloji səviyyədə aradan qaldırılır;

6. Omonim şəkilçi haqqında qrammatik məlumatın işçi sahəyə yazılması (sxem 15).

Məsələn, *kitabım* söz-formasındakı *-ım* şəkilçisinin yuxarıda göstərilən algoritmin köməyi ilə omonimliyinin aradan qaldırılması mərhələlərini nəzərdən keçirək. Morfoloji təhlilin ilk pilləsində söz-forma əsas (kök) və şəkilçiyə ayrılır:

kitab ∪ ım

Alqoritmə nəzərdə tutulduğu kimi yoxlayırıq: şəkilçiyə birləşmiş 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-olma-

ması yoxlanır. Əsas ad nitq hissəsinə (isim) aid olduğu üçün 3-ə keçirik.

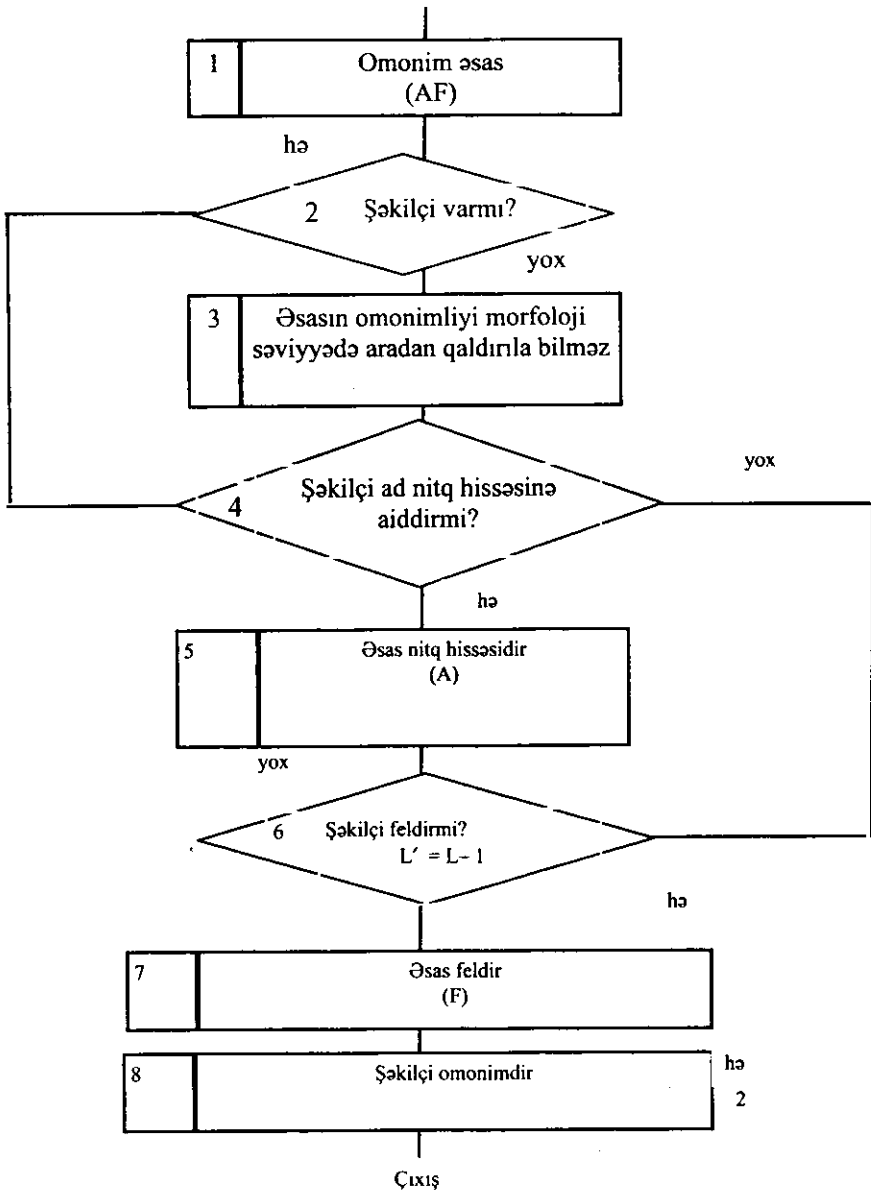
Əsas isim olduğu üçün qaydalara görə şəkilçi də isim olmalıdır. Beləliklə, -im omonim şəkilçinin ad nitq hissəsi ilə işləndikdə malik olduğu qrammatik səciyyələr müəyyənləşdirilir və işçi sahəyə yazılır:

-im - mənsubiyyət şəkilçisi, I şəxs, tək və s.

9.9. Tezauruslar vasitəsi ilə omonimliyin və çoxmənallığın aradan qaldırılması

Tezaurusların mətnlərin avtomatik işlənməsi və məlumat axtarışı sistemlərində tətbiqi bir çox maraqlı və optimal nəticələrin əldə olunması imkanı yaradır (Герд, Богданов, Буторов, Роменская, Тисенко, Андреева, 1978; Шемякин, 1974; Мамедова, 1990 və s.).

Mətnlərin avtomatik işlənməsi sistemlərində çoxmənali sözlərin və söz birləşmələrinin müəyyənləşdirilməsi, cümlədə ayrı-ayrı ifadələrin düzgün variantının seçilməsi sistemin keyfiyyətini səciyyələndirən mühüm şərtlərdəndir. Bu məsələ ilə bağlı çoxkomponentli söz birləşmələrinin çoxmənallığının aradan qaldırılmasına xüsusi diqqət yetirilməlidir. Türkcə-rusca məşin 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 çoxmənallığın həlli bir çox allarda qeyri-mümkün olur.



Sxem 14. Əsasların omonimliyinin aradan qaldırılması algoritminin blok-sxemi

Sintaktik səviyyədə də çoxmənalılıq heç də həmişə «müvəffəqiyyətə» aradan qaldırılmaz. Linqvistik vahidlər birləşmə tərkibində müxtəlif mənalarda işləyə bilərlər. Tezuruslar bu müxtəlif mənalarından düzgün variantın seçilməsinə imkan verir. Misal olaraq aşağıdakı iki söz birləşməsinə nəzərdən keçirək:

Müasir istehsal vasitələri – современные средства производства

Beynəlxalq əmək təşkilatı – международная организация труда

Əgər mətndə 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 işlənilirlər.

Sintaktik əlaqəli elementlər çoxluğundan düzgün variantın müəyyənləşdirilməsi, 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 müvafiq surətdə semantik münasibətlərin aydınlaşdırılmasına xidmət edir. Tezaurusların ikinci mühüm funksiyası- təhlil olunan elm sahəsinə dair 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) mətndəki konkret mənalarını açıqlamağa imkan verir.

Formal təhlil sistemində tezaurusların istifadə olunması perspektivlərini ictimai-siyasi leksika əsasında tərtib edilmiş Al sözlüyü üzərində əyani göstərmək olar. Bunun üçün göstərilən konkret sahənin real əlaqələrini səciyyələndirən semantik-statistik əlaqələri əsas götürülmüş sahə tezaurusunun yaradılması lazımdır. İctimai-siyasi terminologiyanın mənə əlaqələrinin for-

malaşması üçün termin fərqləndirilmələrindən istifadə olunmuşdur.

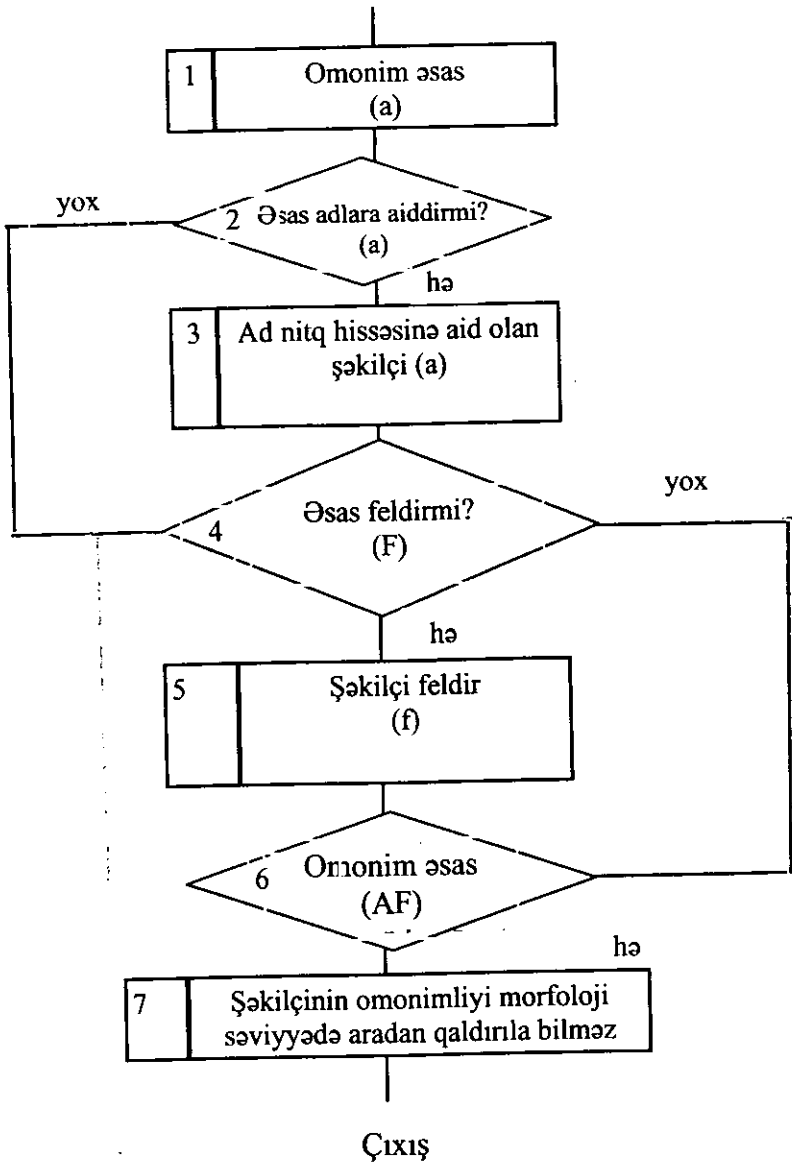
Tezaurus modeli kimi semantik şəbəkə götürülür. Həmin şəbəkənin zirvəsində sözlərin leksik mənalari, yanlarında isə sözlər arasındakı semantik-paradiqmatik münasibətlər qarşılaşdırılır. Semantik şəbəkənin tərtibi üçün aşağıdakı metodika əsas götürülür. Qrafın yanlarında semantik derivatlar, başqa sözlə 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.

Belə hesab olunur ki, başqa sözün təyin olunmasında iştirak edən söz sonuncunun vasitəsiz komponenti hesab olunur. Beləliklə, əgər bir 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 diferensiasiyasını mümkün edir. Tezaurus «cins-növ» «tam-hissə», istənilən iyerarxiya, assosiativ münasibətlər, ekvivalentlik münasibətləri (sinonimlik) və s. əsasında qurula bilər.

Hansı məqsədə xidmət edəcəyi və məzmunu nəzərə alınıraq tezauruslar müxtəlif prinsiplərə müvafiq qurulur. Formal təhlil sistemlərində çoxmənalılığın aradan qaldırılması, düzgün ekvivalentin seçilməsi ilə əlaqədar tezauruslar uğurla istifadə oluna bilər.



Sxem 15. Şəkilçilərin omonimliyinin aradan qaldırılması algoritminin blok-sxemi

SONLUQ

XX əsrin 50-ci illərində riyaziyyat, kibernetika, informatika, statistika, hesablama texnologiyası və s. dəqiq elmlərlə dilçiliyin qovuşma nöqtəsində yeni bir elm sahəsi formalaşdı. İlk əvvəl bu yeni yaranan elm sahəsi “riyazi dilçilik”, “struktur dilçilik”, “hesablama dilçiliyi”, “statistik dilçilik”, “kibernetik dilçilik”, “struktur-tətbiqi dilçilik”, “mühəndis dilçiliyi”, “korpus dilçiliyi” və s. kimi adlandırılırdı. Əslində müxtəlif elmlərin qovuşduğunda yaranan bu yeni elm sahəsinin tədqiqat obyektı və predmeti məlum idi. Çox geniş və əhatəli araşdırma sahəsinə malik olduğu üçün ayrı-ayrı tədqiqatçılar ixtisaslarına müvafiq olaraq dilçiliklə həmhüdüdü elm sahəsi yaratmağa cəhd göstərir və bu cür termin müxtəlifliyinə yol verirdilər.

“Kompüter dilçiliyi” termini bu adların hamısını ehtiva edir və dilçiliyin riyaziyyat, statistika, kibernetika elmləri kompleksli ilə qovuşma nöqtəsində yaranan bu yeni sahəsinə “kompüter dilçiliyi” adlandırmaq düzgün olardı.

Milli korpusun tərkib hissəsi olan mətnin formal təhlili sistemi Azərbaycan dilinin tədqiqində yeni bir istiqamət, keyfiyyətə yeni bir mərhələdir. Bu mərhələnin ən aparıcı xüsusiyyəti dilin formal əlamətlərinə əsaslanan linqvistik modellərinin qurulmasıdır. Mühəndis dilçiliyi prinsipləri əsasında yaradılmış linqvistik modellər real dil hadisələrini və qanunauyğunluqlarını tam təqlid etmək, doğurmaq keyfiyyətinə malikdir. Bu isə, dilin mahiyyətinin dərkı prosesində həlledici amildir.

Mətnin formal işlənməsi sistemi ayrı-ayrı dillər üçün yaradılması nəzərdə tutulmuş milli korpusun ən mühüm tərkib hissələrindən biridir. Azərbaycan dilinin milli korpusunun keyfiyyətə və kəmiyyətə zənginləşməsində, təkmilləşməsində təhlil alqoritmlərinin rolu danılmazdır. ADMK-də dilin fonetik, qrammatik, semantik və leksik sistemi ilə əlaqədar bu və ya digər məlumatın əldə olunması prosesində formal təhlili sisteminin alqoritmləri və proqramlarının xidmətindən istifadə etmək səmərəli nəticələr verə bilər.

Ənənəvi dilçilikdəki mövcud qrammatik qaydalar təhlil sistemlərində özünü konkret məqsədə xidmət edən alqoritmlər şəklində göstərir. Formal təhlil alqoritmləri mətnlərin müxtəlif səviyyələrdə işlənməsi ilə bağlı yerinə yetirilən əməliyyatlar ardıcılığıdır. Onlar əsasında tərtib olunmuş proqramlar dildə mövcud qanunauyğunluqların formal ifadəsinə əsaslanmaqla mətnlərin kompüterlərdə işlənməsinə imkan yaradır.

Mətnin formal təhlili alqoritmləri bəzən təbii dildə əsrlər boyu formalaşmış, daşlaşmış və ekama çevrilmiş bir çox dil qaydalarına və hadisələrinə yeni baxış bucağı altında baxmaq zərurəti doğurur. Bu zərurət ondan irəli gəlir ki, avtomatik işlənmə dilin yeni metod və vasitələrlə tədqiqini tələb edir.

Hesablayıcı texnikanın, fərdi kompüterlərin məişətimizə belə kütləvi nüfuzu dövründə mətnlərin internet məkanı üçün vacib olan formal sistemlərin yaradılması vacib linqvistik vəzifədir. Dilçiliyin gələcək inkişaf perspektivlərini mühəndis dilçiliyi modelləri, müxtəlif məqsədli alqoritmlər və kompüterlərsiz təsəvvür etmək çətindir. Mətnin formal təhlili sistemi ənənəvi dilçiliyin özünün inkişafına da müəyyən dərəcədə təkan vermiş olur. Bu kompüter lüğətlərinin tərtibində statistik metodlardan istifadə olunmasında daha qabarıq əks olunur. Yeni yaranmış sahə - statistik leksikoqrafiya, kompüter lüğətçiliyi lüğətçiliyin inkişafında yeni mərhələdir.

Mətnin formal təhlili sisteminin leksik bloku bütün sonrakı bloklar (morfoloji, morfonoloji, sintaktik, semantik) üçün zəmin hazırlayır. Bütün dil səviyyələrində avtomatik işlənmənin keyfiyyət göstəriciləri leksik blokun özünün keyfiyyət göstəricilərindən çox asılıdır. AL-in elmi prinsiplər üzrə optimal tərtibi, ora kifayət qədər linqvistik informasiyanın daxil edilməsi linqvistik işlənmə prosesinin səmərə və keyfiyyətini xeyli artırmış olur.

Leksik blok təkcə zəngin informasiyaya malik olması ilə deyil, həm də bu informasiyanın qorunması, yerləşdirilməsi və lazımi sorğular əsasında verilməsi üsuluna görə optimal olmalıdır. Bu baxımdan ənənəvi prinsiplərin «pozulması» (lüğəti təşkil edən linqvistik vahidlərin əlifba sırası ilə deyil, onları təşkil

edən hərflərin sayına görə düzümü, bir çox morfoloji və morfonoloji dəyişmələrin lüğətlərdə verilməsi və s.) zəruri proses kimi qiymətləndirilməlidir.

Təhlil sisteminin mühüm komponenti olan AL-in sözlüyü elmi prinsiplərlə müəyyənləşdirilməlidir. Təqdim olunan sistemdə AL-in sözlüyü Azərbaycan dilinin tezlik lüğəti əsasında seçilmişdir.

Morfoloji və morfonoloji xüsusiyyətlərin hərtərəfli araşdırılması əks lüğətlər olmadan qeyri-mümkündür. Sonu bu və ya digər samitlə bitən sözlər, məhsuldar və qeyri-məhsuldar şəkilçilərin təsnifi ilə əlaqədar Azərbaycan dilinin əks lüğəti əvəzsiz mənbədir. Həmin lüğətin köməyi ilə əldə olunmuş zəngin materiallar təhlil sistemində morfonoloji blokun qurulması üçün əsas götürülmüşdür.

Azərbaycan dilinin sintaktik quruluşu və xüsusiyyətləri nəzərə alınaraq birləşmələr AL-də lüğət vahidləri qruplar üzrə yerləşdirilir. Bu da təhlil və sintezin səmərə və keyfiyyətini artırır, düzgün tərcümə variantının alınmasını təmin edir. Göstərilən xüsusiyyət türk dillərinin sintaktik təbiətindən irəli gəlir və həmin cəhət digər türk dilləri üçün yaradılacaq analoji sistemlərdə nəzərə alınmalıdır.

N.Xomskinin transformasiya qrammatikaları sintez prosesi üçün qismən məqbul sayılsa da, təhlil, xüsusilə sintaktik təhlil Azərbaycan dili üçün tamamilə başqa prinsiplər əsasında qurulmalıdır. Mühəndis dilçiliyində sınaqdan çıxarılmış metodlarla qurulmuş morfoloji, sintaktik və semantik təhlil və sintez modelləri mətnlərin işlənməsi sistemlərinin qurulması üçün əsas götürülə bilər.

Gödelin natamamlıq teoreminin linqvistik interpretasiyasını əsas götürərək göstərmək olar ki, morfoloji səviyyədə aradan qaldırıla bilinməyən omonimlik daha yuxarı səviyyələrdə - sintaktik və semantik səviyyələrdə həll olunmalıdır. Formal təhlil sistemlərində bu tezis aparıcı olmalı, aşağı dil səviyyələrində həlli qeyri-mümkün olan omonimlik və çoxmənalılıq yuxarı səviyyələrdə araşdırılmalıdır.

Mətnin formal təhlili sistemi ilə bağlı dilin quruluşuna, dil hadisələrinə və qanunauyğunluqlarına yeni nəzərlə baxılması Azərbaycan dilinin bir çox xüsusiyyətlərinin yeni izahını tələb edir. Bu baxımdan nitq hissələrinin təsnifində bölgü prinsipinin pozulması Azərbaycan dilində (eləcə də digər türk dillərində) müxtəlif dolaşq təsniflərin yaranmasına gətirib çıxarmışdır.

Formal təhlil sisteminin linqvistik alqoritmləri dilin yaranması, inkişafı və ünsiyyət funksiyalarının araşdırılması baxımından da əhəmiyyətlidir. Göstərilən araşdırmalar dilin tədrisi ilə bağlı öyrədici sistemlərin yaradılmasında geniş istifadə oluna bilər.

Leksik blokda təsvir olunmuş Azərbaycan dilinin tezlik, əks və avtomatik lüğətləri ənənəvi lüğətçilikdə (lüğət-minimumların, danışq və sorğu kitabçalarının, tezaurusların tərtibi və s.), dilin tədrisi prosesində tətbiq oluna bilər.

Elmin sonrakı inkişaf pilləsi dəqiq statistik təhlilin tədqiqata cəlb edilməsini şərtləndirir. Bu baxımdan tarixi abidələrin kompüterlərlə tədqiqi çox vacibdir. «Kitabi-Dədə Qorqud» داستانlarının və M.Füzulinin dilinin statistik təhlili qiymətli linqvistik axtarışlar üçün əvəzsiz faktlar verə bilər.

Mətnin formal təhlili alqoritmləri universal səciyyəyə malikdir. Onlardan dil tədqiqatlarında müxtəlif məqsədlərlə istifadə etmək olar. Məlumat axtarışı sistemlərində, avtomatik sintez və təhlil sistemlərində, MT ilə bağlı göstərilən alqoritmlər öz geniş tətbiqini tapa bilər. Formal təhlil alqoritmləri digər türk dilləri üçün də məqbuldur.

Formal təhlil sistemlərində müxtəlif elm sahələrini əhatə edən tezauruslar çoxmənalılığın aradan qaldırılması, düzgün ekvivalentin seçilməsi, sintaktik əhatənin təyini kimi məsələlərlə bağlı geniş istifadə oluna bilər.

Kompüter dilçiliyi yenidən formalaşmaqda olan, lakin böyük gələcəyi olan, bəşəriyyətə fayda gətirəcək bir elm sahəsidir. Cəmiyyətin gələcək inkişafını kompüterlərsiz təsəvvür etmək mümkün olmadığı kimi, dilçilik elminin də gələcək inkişaf istiqamətlərini və perspektivlərini kompüter dilçiliyi olmadan təsəvvür etmək qeyri-mümkündür.

QISALTMALAR VƏ ŞƏRTİ İŞARƏLƏR

- АВЛ - Автоматизация в лингвистике
ADQ - Azərbaycan dilinin qrammatikası
АДД - Автореферат докторской диссертации
ADMK – Azərbaycan dilinin milli korpusu
ADMF -Azərbaycan dilinin məşın fondu
АКД - Автореферат кандидатской диссертации
AL - Avtomatik lüğət
AMAİS - Azərbaycan mətnlərinin avtomatik işlənməsi sistemi
АП - Автоматический перевод
АПТ - Автоматическая переработка текста
ВЛ - Вычислительная лингвистика
ВП - Вопросы кибернетики
VPR - Вероятностное прогнозирование в речи
ГАЯ - Грамматика азербайджанского языка
DDA - Doktorluq dissertasiyasının avtoreferatı
ИМЛ - Исследования по математический лингвистике
МЛ - Математическая лингвистика
ММП - Материалы по машинному переводу
МППЛ - Машинный перевод и прикладная лингвистика
МТ - Məşın tərcüməsi
МЯМП - Моделирования языка и машинный перевод
NDA - Namizədlik dissertasiyasının avtoreferatı
ОРОСИИ - Описание и распознавание объектов в системах искусственная интеллета
ППЛ -Проблемы прикладной лингвистики
СЕИЯСС - Семантика естественных и искусственных языков в специализированных системах
СКТ - Статистика казахского текста
СМЛ - Структурная и математическая лингвистика
СПЛ - Структурная и прикладная лингвистика
СТ - Советская тюркология
СТРААТ - Статистика речи и автоматический анализ текста

ТППЛ - Теоретические проблемы прикладной лингвистики

ЧСРЯ - Частотный словарь русского языка

ОЧСЛЯ - Обратный частотный словарь латышского языка.

F/A/ - Felin arzu şəkli

F/V/ - Felin vacib şəkli

F/Ə/ - Felin əmr şəkli

F/X/ - Felin xəbər şəkli

F/L/ - Felin lazım şəkli

F/d/ - Felin davam şəkli

F/m1/ - Felin məlum şəkli

F/m2/ - Felin məchul şəkli

F/m0/ - Məsdər

F/q/ - Felin qarşılıq növü

F/i/ - Felin icbar növü

F/qy/ - Felin qayıdış növü

F/E/ - Felin şərt şəkli

I - isim

S - sifət

Z - zərf

TSB - təyini söz birləşməsi

MT- məsdər tərkibi

FST - feli sifət tərkibi

Ş/Ş - şəxs şəkilçisi

S - cümlə

AQ - ad qrupu

FQ - fel qrupu

MQ - mübtədə qrupu

N - təyin

M - mübtədə

XQ - xəbər qrupu

ZS - zərflik

Tm/ZC - tamamlıq/ zərflik

X - xəbər
P - qoşma
B - bağlayıcı
Ə - ədat
b - modal sözlər
B₁ bağlamalar
Sa - say
Əv - əvəzlik

İSTİFADƏ OLUNMUŞ ƏDƏBİYYAT

Azərbaycan dilində

1. Abdullayev K.M. Azərbaycan dili sintaksisinin nəzəri problemləri. Bakı, "Maarif", 1998, –281 s.
2. Axundov A.A. Ümumi dilçilik. Bakı, "Maarif", 1979, – 209 s.
3. Abdullayev Ə. Aktual üzvlənmə və mətn. Bakı, Xəzər Universiteti nəşriyyatı, 1998, –190 s.
4. Abdullayev Ə. Mətni anlama modelləri. Bakı, "Səda" nəşriyyatı, 1999, –344 s.
5. Abdullayev Ə., Seyidov Y., Həsənov A. Müasir Azərbaycan dili (sintaksis), Bakı, «Maarif» nəşriyyatı, 1972, – 464s.
6. Azərbaycan dilinin izahlı lüğəti. Dörd cilddə. Bakı, «Şərq-Qərb» nəşriyyatı, 2006 (I cild – 744 s.), II cild (– 792 s., III cild –672 s., IV cild –712 s.)
7. Azərbaycan qəzet dilinin tezlik lüğəti (tərtibçilər: K.A.Vəliyeva, M.Ə.Mahmudov, V.Pines), Bakı, «Nurlan», 2004, –264 s.
8. Azərbaycan dilinin əks əlifba lüğəti (tərtibçilər: M.Ə.Mahmudov, Ə.Fətullayev), Bakı, «Nurlan», 2004, – 524 s.
9. Azərbaycan dilinin tezlik lüğəti (söz kökləri), I cild (tərtibçilər: M.Ə.Mahmudov, Ə.Fətullayev və b.), Bakı, «Elm», 2010, –464 s.
10. Azərbaycan dilinin qrammatikası. I hissə (morfologiya). Bakı, Azərbaycan SSR EA-nın nəşri, 1960, –334 s.
11. Azərbaycan dilinin orfoqrafiya lüğəti. Bakı, «Elm» nəşriyyatı, 1975, –598 s.
12. Azərbaycanca-rusca, rusca-azərbaycanca lüğət (tərtibçilər: L.Kərimov, M.Mahmudov). Bakı, Azərənşr, 1991, – 312 s.
13. Axundov A.A. Azərbaycan dilinin fonetikasi. Bakı, «Maarif» nəşriyyatı, 1984, –391 s.

14. Axundov A.A. Azərbaycan dili saiflərinin statistik təhlili təcrübəsindən. «ADU-nun elmi əsərləri (Dil və ədəbiyyat)», №5, 1963, 63-67 s.
15. Axundov A.A. Riyazi dilçilik: Xülasələr və təfriqələr. Dərs vəsaiti. Bakı, ADU, 1979, –79 s.
16. Vəliyev K. Linqvistik poetikaya giriş. ADU-nun nəşri, 1989, –101 s.
17. Qiyasbəyli M.N. Azərbaycan dilində morfoloji söz yarıdıcılığı (düzəltmə sözlərin lüğəti), «Maarif» nəşriyyatı, Bakı, 1987, –179 s.
18. Zeynalov F.R. Türk dillərinin müqayisəli qrammatikası. Bakı, ADU nəşri, 1974, -142 s.
19. Mahmudov M.Ə. Bir daha nitq hissələrinin təsnifi barədə. Tarixi və tətbiqi dilçilik. Bakı, 1993, s. 36-39
20. Mahmudov M.Ə. Gödelin natamamlıq teoreminin linqvistik interpretasiyası. Dil, etnos və etnonimiya. Bakı, «Örnək», 1994, s. 82-83
21. Mahmudov M.Ə. Tezlik lüğətləri dil tədrisində. Azərbaycan EA-nın xəbərləri, 1992, № 3, 4, s. 99-102
22. Mahmudov M.Ə. Azərbaycan mətnlərinin avtomatik işlənməsi sistemi. DDA, Bakı, 1994. –64 s.
23. Mahmudov M.Ə., Sultanov V.S. Dilimizin tarixi və kompüter dilçiliyi: problemlər, perspektivlər, tarixi və tətbiqi dilçilik. Bakı, 1993, s. 3-4
24. Mahmudov M., Cəlilov F. Tezlik lüğətlərinin tərtib olunma qaydaları haqqında. S.M.Kirov adına ADU-nun elmi əsərləri. Dil və ədəbiyyat seriyası, №3, 1977, s. 49- 53
25. Pines V.Y., Mahmudov M.Ə. Dilçiliyin başqa elmlərlə həmhüdüdü sahələri və perspektivlər. Böyük Oktyabr Sosialist inqilabının 70 illiyinə həsr olunmuş Azərbaycan dilçiliyi məsələləri elmi-nəzəri konfransının məruzələrinin tezisləri. Bakı, 1987, s. 22-24
26. Pines V.Y., Mahmudov M.Ə. Elektron hesablayıcı maşınlar dilçilikdə, Bakı, AzETETMİ, 1977, –32 s.

27. Hüseynzadə M. Müasir Azərbaycan dili. Morfologiya, III hissə. Ali məktəblər üçün dərslik. Bakı, «Maarif», 1973, –358 s., 1983, –319 s.
28. Cavadov Ə.M. Müasir Azərbaycan ədəbi dilində sintaktik vahidlərin sırası. «Elm», 1977, –112 s.
29. Cəlilov F.A. Azərbaycan dilinin morfonologiyası. Bakı, «Maarif» nəşriyyatı, 1988, - 286 s.
30. Cəlilov F. Azərbaycan dili morfonologiyasından öçerklər. Bakı, S.M.Kirov adına ADU-nun nəşri, 1985, –108 s.
31. Cəlilov F. Mürəkkəb cümlə sintaksisi. Bakı, «Maarif» nəşriyyatı, 1983, –116 s.
32. Cəfərov S. Azərbaycan dilində sözdüzəldici və sözdəyişdirici şəkilçilər. Bakı, «Maarif», 1968, - 107 s.
33. Vəliyeva K.A., Mahmudov M.Ə., Pines V.Y., Rəhmanov C.Ə., Sultanov V.S. “Kitabi-Dədə Qorqud” dilinin statistik təhlili (ilkin nəticələr). Bakı, “Elm”, 1999, –248 s.
34. Mahmudov M. Kompüterlər linqvistik araşdırmalarda: problemlər, perspektivlər. Dil dergisi. Turkoloji. Ocak, 1995, 27, Ankara, s.10-13.
35. Sultanov V. Cümlənin semantik özəyi. Bakı, “Elm”, 1997.
36. Kazımov Q. Mətn dilçiliyi problemləri. Ə.Abdullayevun “Aktual üzvlənmə və mətn” kitabına ön söz. Bakı, Xəzər Universiteti nəşriyyatı, 1998, s. 5-21.
37. Rəcəbov Ə. Dilçilik tarixi. Bakı, “Maarif”, 1988. –538 s.
38. Mahmudov M.Ə. Maşın tərcüməsi sistemində avtomatik morfoloji analizin yeri. Azərbaycan Elmlər Akademiyasının xəbərləri. ədəbiyyat, dil və incəsənət seriyası, №1, 1981, s. 68-72.
39. Novruzova N. Mətn sintaksisi. Bakı, “Təhsil” nəşriyyatı, 2002, –184 s.
40. Rusca-azərbaycanca lüğət. Üç cildə, Bakı, «Elm» nəşriyyatı, I cild, 1971, 606 s., II cild, 1976, 712 s., III cild, 1979, 556 s.
41. Əmirov Z.M. Azərbaycan dilinin formal mətnlərinin yaradılması və onların əsasında linqvistik prosessorun qu-

- rulması. Nam. disser. avtoreferatı, Bakı, AMEA akad. Ə.Hüseynov adına Kibernetika İnstitutu, 2006, –198 s.
42. 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.
 43. Məmmədov A. Mətn yaranmasında formal əlaqə vasitələrinin sistemi. Bakı, «Elm», 2001, –143 s.
 44. İsrəfilov M.M. Məntiq. Dərs vəsaiti. Bakı, «Maarif» nəşriyyatı, 1987, –334 s.
 45. Fətullayev Ə.B. Azərbaycan-İngilis maşın tərcüməsi sistemi üçün rəqəmsal modelləşdirmə metodunun işlənilib hazırlanması və tətbiqi. Namizəd. disser. avtoreferatı, Bakı, AMEA akad. Ə.Hüseynov adına Kibernetika İnstitutu, 2006, –19 s.
 46. Məmmədov A. Mətn yaranmasında formal əlaqə vasitələrinin sistemi. Dokt. disser. avtoreferatı, Bakı, 2003, –54 s.
 47. Vəliyeva K.A., Məmmədova M. Mətnlərin avtomatik redaktəsi. Bakı, «Azərbaycan Universiteti» nəşriyyatı, 2003, –809 s.
 48. Vəliyeva K.A. Mətnin avtomatik sintaktik təhlili və sintezi. Bakı, «Örnək», 1996, –158s.
 49. Vəliyeva K.A., Mahmudov M.Ə., Pines V.Y., Rəhmanov C.R., Sultanov V.S. «Kitabi-Dədə Qorqud» dastanlarının statistik təhlili. Bakı, «Siyasət», 1977, –77 s.
 50. Əliquliyev R., Ələkbərova İ. Vikipediya. Virtual ensiklopediya. Bakı, «İnformasiya Texnologiyaları», 2012, –184s.
 51. Fətullayev R.Ə. Maşın tərcüməsi sisteminə Azərbaycan dilində cümlənin formal sintezi alqoritmlərinin yaradılması və tətbiqi. Nam. disser. avtoreferatı, Bakı, AMEA akad. Ə.Hüseynov adına Kibernetika İnstitutu, 2010, –19 s.
 52. Xəlili Ə.M. Deduktiv maşının biliklər bazasının tərkib hissəsi kimi «Məhdud Azərbaycan dili»nin formal qrammatikasının işlənilib hazırlanması. Nam. disser. avtoreferatı, Bakı, AMEA nəsimi adına Dilçilik İnstitutu, 2009, –33 s.

53. Əliyev Ə.A. İngiliscə-azərbaycanca məşin tərcüməsi sistemində felin çoxmənalılığının alqoritmik həlli. Nam. disser.avtoreferatı. Bakı, Azərbaycan Dillər Universiteti, 2012, –20 s.
54. Mahmudov M.Ə. Süni intellekt məşin tərcüməsi sistemlərində. Tətbiqi dilçiliyin müasir problemləri. Beynəlxalq elmi konfrans (Azərbaycan, Bakı, 27-28 noyabr 2012-ci il) Bakı, «Mütərcim», 2013, s.61-63.
55. «Kitabi-Dədə Qorqud»un statistik təhlili (tərtibçilər: Vəliyeva K., Mahmudov M., Pines V., Rəhmanov C., Sul-tanov V), Bakı, «Elm», 1999, –248 s.
56. Mahmudov M. Ənənəvi dilçiliyə qeyri-ənənəvi baxış. AMEA-nın Xəbərləri. Humanitar Elmlər Seriyası, 2006, №2-4, s.101-113.
57. Mahmudov M. Mətnin formal təhlili sistemi. Bakı, «Elm», 2002, 244 s.
58. Əliquliyev R., Şükürlü S., Kazımova S. Elmi fəaliyyətdə istifadə olunan əsas terminlər. Bakı, «İnformasiya texnologiyaları», 2009, –201 s.
59. İngiliscə-azərbaycanca lüğət. «Qismət» nəşriyyatı, Bakı, 2003, –1696 s.

Rus dilində

60. Автоматизация в лингвистике. Сборник статей, переведенных с английского, французского, немецкого и чешского языков. Москва-Ленинград, «Наука» 1966, –158 с.
61. Автоматический перевод. Сб. Статей /под ред. и с предисл. О.С.Кулагиной и др./ Москва, «Прогресс» 1971, –368 с.
62. Автоматическая переработка текста, системный подход к решению проблем машинного перевода, Кишинев, Штиинца, 1978, –93 с.

63. Алекперов А.К. Лексическая семантика простых глаголов в современном азербайджанском языке. Баку, изд. «Элм», 1983, –157 с.
64. Алексеев П.М. Статистическая лексикография. Учебн. пособие. Ленинград, Ленинградский пед. Институт им. А.И.Герцена, 1975, –120 с.
65. Андреев Н.Д. Квазилингвистика Хомского (о причинах неудачи порождающих грамматик). Вопросы языкознания, №5, 1976, 58-73 с.
66. Андреев Н.Д. Кремнева Н.Д., Волков Н.Н. Влияние типа грамматики языка-посредника на схему алгоритмов машинного перевода. АН СССР, Инс-т научн. информации, вып. 8, Москва, 1961, –32 с.
67. Андреев Н.Д. Машинный перевод и проблема языка-посредника. ВЯ, 1957, №5, 117-121 с.
68. Арзикулов Х.А., Пиотровский Р.Г., Попеску А.Н., Хажинская М.В. Автоматизированная система тезаурусного аннотирования научно-технического документа. Научно-техническая информация, сер. 2, №12, 1978, 12-20 с.
69. Аршинов М.Н., Садовский Л.Е. Коды и математика, Москва, «Наука», 1983, –143 с.
70. Губанова Г.Г. Системно-аспектуальное функционирование компьютерной терминологии. Диссертация на соискание ученой степени доктора филологических наук. Омский, гос.пед.унив., Омск, 2009, –306 с.
71. Джубанов А.К. Казахские языкознание: прикладная лингвистика. Алматы, «КИЕ», 2012, –696 с.
72. Современное казахское языкознание: актуальные вопросы прикладной лингвистики. Материалы международной научно-теоретической конференции посвященная 75-летию юбилею Жубанова А.К., Алматы, И-т Языкознания им.А.Байтурсынова, 2012, –488 с.

73. 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. теорет. конф. Алматы, 2012, с.35-43.
74. Бускунбаева Л.А., Сиразитдинов З.А. О проблемах создания национального корпуса башкирского языка. Современное казахское языкознание: актуальные вопросы прикладной лингвистики. Материалы межд. научн. теорет.конф.Алматы, 2012, с.54-58.
75. Акулинина Т.В. Экстралингвистическая обусловленность особенностей английской терминологии компьютерной информатики. Автореф.дис.канд. филол. наук. Омск, 2003, –23 с.
76. Бабанаров А., Махмудов М.А., Садыков Т. Морфологический анализ в ОЛА. Школа-семинар по оптимизации преподавания иностранных языков с помощью технических средств (тезисы докладов и сообщении), Кишинев, 1979, –21 с.
77. Бабанаров А. Разработка принципов построения словарного обеспечения турецко-русского машинного перевода, АКД, Ленинград, 1981, –17 с.
78. Байрамова П.К. Перевод на татарский язык русских придаточных определительных предложений с относительным словом «который» и экспериментальная оценка на ЭЦВМ некоторых вопросов перевода этих предложений, АКД, Ленинград, 1966, –16 с.
79. Баскаков Н.А. Двойственная природа слова и проблема частей речи (на материале тюркских языков). -В кн.: Вопросы теории частей речи. Ленинград, «Наука», Ленинградская отд., с.1968, 237-237
80. Баскаков Н.А. Морфологическая структура слова в тюркских языках различных типов. -В сб.: Морфоло-

- гическая структура слова в языках различных типов. Москва-Ленинград, «Наука» 1963, 76-83 с.
81. Баскаков Н.А. Морфологическая структура слова и части речи в тюркских языках. Сов. Востоковедение, №1, 1957, 72-85 с.
 82. Бектаев К.Б. Русско-казахский математический словарь (на казахском языке), Алматы, «Мектеп», 1986, -295 с.
 83. Бектаев К.Б. Статистическо-информационная типология тюркского текста. Алма-Ата, Наука, 1978, -183 с.
 84. Белоногов Г.Г., Калинин Ю.П., Поздняк М.Ф., Хорошилов А.А., Яфаева Г.М. Алгоритм многоступенчатого морфологического анализа русских слов. НТИ, сер. 2, №1, 1983, с. 6-10.
 85. Белоногов Г.Г., Кузнецов Б.А. Языкованные средства автоматизированных информационных систем. Москва, «Наука», 1983, -287 с.
 86. Белоногов Г.Г., Новоселов А.П. Автоматизация процессов накопления, поиска и обобщения информации. Москва, «Наука», 1979, -256 с.
 87. Бельская И.К. Язык человека и машина. Москва, изд.Моск. Университета, 1969, -410 с.
 88. Беляева Л.Н. Структура лингвистического обеспечения для задач автоматической переработки текста. Тезисы докладов конференции «Семантика естественных и искусственных языков в специализированных системах». Ленинград, изд. Ленинградского университета, 1979, с. 22-23.
 89. Билан В.М., Боркун М.Н., Пиотровский Р.Г. Методы автоматического анализа и синтеза текста. Учебное пособие. Минск, Мин. Гос. Пед. Инс-т иностр. яз., 1977, -110 с.
 90. Блумфилд Л. Язык. Москва, «Прогресс», 1968, -607 с.

91. Борисевич А.Д., Гончаренко В.В., Гречишкина В.И., Добрускина Э.М., Кочеткова В.К., Крисевич В.С., Нехай С.А., Пиотровский Р.Г., Штирбу Т.А., Ястребова С.В. Кодирование грамматической информации в машинном словаре. В кн.: Статистика текста и автоматическая переработка текста, Минск, изд. БГУ, 1970, с. 331-614
92. Велиева К.А., Мамедова М.Г., Махмудов М.А., Пинес В.Я. Проблема автоматической переработки текстов на азербайджанском языке. Тезисы докладов семинара-совещания. «Республиканская система научно-технической информации и опыт организации информационного обеспечения народного хозяйства». Баку, АЗНИИНТИ, с.1985, 35-37.
93. Бабалова Г.Г. Лингвистические аспекты информатики (терминология и лексикография). Учебное пособие. Омск. Изд.-во ОмГПУ, 2003, –84 с.
94. Бабалова Г.Г. Математические методы в лексикографии: лексические функции, лингвистические модели и алгоритмы // Алгоритмы в лингвистике. Омск, Изд.-во ОмГПУ, 2003, –84 с.
95. Баранов А.Н. Введение в прикладную лингвистику. Изд.-во 2-е, испр.; М., Едиториал, УРСС, 2003, 358 с.
96. Велиева К.А., Мамедова М.Г., Махмудов М.А., Пинес В.Я., Рахманов Дж.А. Принципы построения систем лексико-грамматического машинного перевода с тюркских языков. Международная конференция «Теория и практика научно-технического перевода». Тезисы докладов. Москва., 1985, 113-114 с.
97. Велиева К.А. Формальное описание синтеза азербайджанского слово. АКД, Москва, с. 1971, –20.
98. Вероятностное прогнозирование в речи. Москва, «Наука», 1971, –199 с.

99. Вопросы кибернетики. Выпуск 41. Статистика речи и автоматический анализ текста. Москва-Ленинград, 1978, -107 с.
100. Вычислительная лингвистика. Москва, «Наука», 1976, -231 с.
101. Гальперин И.Р. Текст как объект лингвистического исследования. Москва, «Наука», 1981, -139 с.
102. Генкин С.Е. Числовой язык-посредник. -В кн.: Симпозиум по структурному изучению знаковых систем. Тезисы докладов. Москва, 1962, с. 90-91.
103. Герасимов В.Н., Марчук Ю.М. Современное состояние машинного перевода. Сб.: «Машинный перевод и Автоматизация информационных процессов». Москва, 1975, с. 5-18.
104. Герд А.С., Богданов В.В., Бугоров В.Д., Роменская В.Ф., Тисенко Э.В., Андреева Е.С. Информационно-поисковый тезаурус как объект лексикографии. В сб.: «Структурная и прикладная лингвистика». Ленинград, 1978, с.160-171.
105. Гжегорчик А. Популярная логика. Перевод с польского. изд. «Наука», Москва, 1965, -107 с.
106. Гладкий А.В., Мельчук И.А. Элементы математической лингвистики. Москва, «Наука», 1969, -192 с.
107. Гладкий А.В. Формальные грамматики и языки. Москва, «Наука», 1973, -368 с.
108. Глисон Г. Введение в дескриптивную лингвистику. Москва, 1959, -486 с.
109. Головин Б.Н. Язык и статистика. Москва, изд. «Просвещение», 1971, -191 с.
110. Городецкий Б.Ю. О принципах инвентаризации семантики языка и подязыка. В сб.: Теоретические и экспериментальные исследования в области структурной и прикладной лингвистики. Москва, изд. Московского университета, 1973, с. 20-59.

111. Городецкий Б.Ю. Семантические проблемы построения автоматизированных систем обработки текстовой информации. С сб.: «Вычислительная лингвистика», Москва, «Наука» 1976, с.16-33.
112. Грамматика азербайджанского языка, Баку, Элм, 1971, - 413 с.
113. Грамматика киргизского литературного языка. Часть I. Фонетика и морфология. Фрунзе, «ИЛИМ» 1987, - 399 с.
114. Гройс Б.Е. Анализ простого повествовательного предложения. Тезисы докладов конференции «Семантика естественных и искусственных языков в специализированных системах». Ленинград, изд. Ленинградского университета, 1979, с.51-52.
115. Грунина Э.А., Благова Г.Ф., Гузев В.Г., Насилов Д.М., Нигматов Х.Г., Шербак А.М. Приоритетные направления лингвистических исследований для планирования и выбора тем диссертационных работ. Советская тюркология. №5, 1989, с. 97-101.
116. Гузев В.Г., Пиотровский Р.Г., Шербак А.М. О создании машинного фонда тюркских языков. Советская тюркология. 1988, №2, с. 92-101.
117. Джубанов А.Х. Квантитативная структура казахского текста. Алма-Ата, «Наука» 1987, -147с.
118. Домогацких В.М. Естественный и информационный языки в автоматизированных информационных системах. Тезисы докладов конференции «Семантика естественных и искусственных языков в специализированных системах». Ленинград, изд. Ленинградского университета, 1979, с. 59-60.
119. Дрейзин Ф.А. Об одном способе синтаксического анализа простого предложения. Научные труды Ташкентского Ун-та, Новая серия, вып. 208, 1962, с. 76-81.

120. Дрейзин Ф.А., Рашитов Р.С. Принцип синтаксического анализа татарской фразы. - В кн.: Машинный перевод, Москва, труды АН СССР, вып. II, 1961, с.295-303.
121. Дрейфус Х. Чего не могут вычислительные машины. Москва, «Прогресс», 1978, -334 с.
122. Ермоленко Г.В. Анонимные произведения и их авторы, Минск, изд. «Университетское» 1988, -118 с.
123. Журавлев А.П. Автоматический анализ и синтез фонетической значимости слова и текста. Тезисы докладов конференции «Семантика естественных и искусственных языков с специализированных системах». Ленинград, изд. Ленинградского университета, 1979, с. 61-63
124. Закиев М.З. Классификация частей речи и аффиксов в тюркских языках. Вопросы языкознания, №6, 1973, с. 3-8.
125. Зализняк А.А. Лингвистические задачи. Сб.: «Исследования по структурной типологии». АН СССР, Москва, 1963.
126. Звегинцев В.А. История языкознания XIX-XX веков в очерках и извлечениях. М., 1965, -495с.
127. Звегинцев А.А. Смысл и значение. В сб.: «Теоретические и экспериментальные исследования в области структурной и прикладной лингвистики». Москва, изд. Московского университета, с. 91-98
128. Здоровов Ю.А. К вопросу о синтактико-семантической структуре связного текста. В сб.: «Вычислительная лингвистика», Москва, «Наука», с. 125-133.
129. Зореф М.Г. Машинная морфология и ее применение в немецко-русском автоматическом словаре. В кн.: «Статистика речи и автоматический анализ текста», Ленинград, «Наука», 1973, с. 245-259.

- 130.Зубов А.В. Переработка текста естественного языка в системе «человек-машина». В кн.: «Статистика речи и автоматический анализ текста. Ленинград, «Наука», 1971, с. 286-434.
- 131.Иванов С.Н. Курс турецкой грамматики. Учебные пособие. Ленинград, изд. Ленинградского университета, 1977, –87 с.
- 132.Ингве В. Значение исследования в области машинного перевода. НТИ, сер. 2, №7, 1965, с. 42-47.
- 133.Ингве В. Синтаксис и проблема многозначности. -В сб.: «Машинный перевод», Москва, 1957, с. 281-304.
- 134.Иорданская Л.Н. Автоматический синтаксический анализ. Том II, Новосибирск, «Наука», 1967, –231 с.
- 135.Исследования по математической лингвистике, математической логике и информационным языкам. Москва, «Наука», 1972, -297 с.
- 136.Исхакова Х.Ф. Исследования в области формальной морфологии тюркских языков (на материале татарского литературного языка в сопоставлении с турецким и узбекским). АКД, 1972, –23 с.
- 137.Исхаков Ф.Г. Лексико-грамматическая классификация слов и части речи в тюркских языках. В кн.: «Исследования по сравнительной грамматике тюркских языков». Москва, изд. АН СССР, 1956, с.72-77.
- 138.Касевич В.Б., Насилов Д.М. О проекте машинного фонда тюркских языков. Советская тюркология, №2, 1989, с. 97-100.
- 139.Кенесбаев С.К., Пиотровский Р.Г., Бектаев К.Б. Инженерная лингвистика и тюркология. Советская тюркология, №6, 1970, с. 3-16.
- 140.Кибрик А.Е. Полевая лингвистика и теория языка. В сб.: «Теоретические и экспериментальные исследования в области структурной и прикладной

- лингвистики. Москва, изд. Московского университета, 1973, с. 99-108.
141. Кибрик А.Е. Семантическая проблематика гетерологического кодирования. В сб.: «Теоретические проблемы прикладной лингвистики. Москва, изд. Московского Университета, 1965, с. 67-83.
142. Кононов А.Н. Грамматика современного турецкого литературного языка. Москва-Ленинград, изд. АН СССР, 1956, -569 с.
143. Кононов А.Н. О фузии в тюркских языках. -В кн.: «Структура и история тюркских языков», Москва, «Наука», 1971, с.108-120.
144. Колшанский. Контекстная семантика. Москва, «Наука», 1980, -148 с.
145. Котов Р.Г. Лингвистика и современное состояние машинного перевода в стране. «Вопросы языкознание», №5, 1976, с. 37-49.
146. Котов Р.Г. Лингвистические аспекты автоматизированных систем управления. Москва, «Наука» 1977, - 167 с.
147. Коптилов В.В. О формальном разграничении омонимии и полисемии. -В кн.: «Прикладная лингвистика и Машинный перевод», Киев, изд. Киевского университета, 1962, с.76-78.
148. Кравчук И.С. Разрешение омонимии конфигурации при автоматическом анализе. Тезисы конференции «Проблемы прикладной лингвистики». Москва, 1969, с.168-169.
149. Кулагина О.С. Исследования по машинному переводу. Москва, «Наука», 1979, -320 с.
150. Кулик А.Н. Информационные сети и языковая совместимость дескрипторных ИПС. Москва, «Советское радио», 1977, -247 с.

151. Куно С., Эттингер А. Многовариантный синтаксический анализатор. -В сб.: «Автоматический перевод». Москва, изд. «Прогресс», 1971, с.102-120.
152. Куно С., Эттингер А. Синтаксические структуры фраз и синтаксическая неоднозначность в английском языке. -В сб.: «Автоматический перевод», 1971, с.121-181.
153. Лайонз Дж. Введение в теоретическую лингвистику. Москва, изд. «Прогресс», 1978. -543 с.
154. Лейкина Б.М. К вопросу о структуре языка-посредника. АН СССР. Инс-т научн. Информации, вып. 3, Москва, 1951 -30 с.
155. Лейкина Б.М. Первый этап независимого анализа структуры простого предложения в английском языке. -В кн.: «Материалы по машинному переводу», вып. 1, Ленинград, изд. Ленинградского университета, 1958, с. 215-222.
156. Мамедов А.М. Тюркские согласные: анлаут и комбинаторика. Баку, изд. «Элм», 1985, -206 с.
157. Мамедова М.Г. Автоматизированный отбор лексики в информационно-поисковый тезаурус на основе анализа терминологических словарей. АКД, Москва, 1984, -25 с.
158. Мамедова М.Г. О создании банка терминологических данных азербайджанского языка. Советская тюркология, №2, 1990, с. 84-89.
159. 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.
160. Манин Ю.И. Доказуемое и недоказуемое. Москва, «Советское радио», 1979, -167 с.

161. Марчук Ю.Н. Методы моделирования перевода. Москва, «Наука», 1985, –201 с.
162. Марчук Ю., Н. Проблемы машинного перевода. Москва, «Наука», 1983, –233 с.
163. Марчук Ю.Н. Способы использования семантики в моделировании перевода. Тезисы докладов конференции «Семантика естественных и искусственных языков в специализированных системах». Ленинград, изд. Ленинградского университета, 1979, с. 115-116
164. Математическая лингвистика. Москва, «Наука», 1973, –184 с.
165. Материалы по машинному переводу. Ленинград, изд. Ленинградского университета, 1958, –228 с.
166. Махмудов А.Р., Гулиев А.Д., Кербалаева А.И. Сравнительная типология русского и азербайджанского языков. Учебное пособие. изд. «Маариф», Баку, 1982, –217 с.
167. Махмудов М.А. Инженерно-лингвистические исследования в тюркском языкознании. Всесоюзная конференция «Переработка текста методами инженерной лингвистики (тезисы докладов), Минск, 1982, с. 17-19
168. Махмудов М.А. Лексико-морфологический МП азербайджанского текста на русский язык. Материалы семинара «Статистическая оптимизация преподавания языков и инженерная лингвистика». Чимкент, 1950, с. 257- 260.
169. Махмудов М.А. Машинная морфология тюркских языков. Всероссийская третья школа-семинар «Обучающие лингвистические машины и оптимизация обучения языкам» (тезисы докладов), Махачкала, 1981, –103 с.
170. Махмудов М.А. Морфологический анализ тюркской словоформы с помощью ЭВМ. Материалы республиканской научной конференции «Функци-

- ональные и конструктивные уровни языковой системы (18-19 ноября, 1980), Баку, 1981, с.101-103.
171. Махмудов М.А. Разработка системы формального морфологического анализа тюркской словоформы. АКД, Баку 1982, –26 с.
172. Махмудов М.А., Пинес В.Я. База лексико-грамматических данных в системе азербайджанско-русского МП. Тезисы докладов научно-технической конференции «Проблемы развития и совершенствования республиканской системы научно-технической информации. Баку, 1983, с. 78-79.
173. Махмудов М.А., Пинес В.Я. Экспериментальная система турецко-русского МП. В б.: Международный семинар по машинному переводу «ЭВМ и перевод - 89». Тбилиси, 22 ноября - 2 декабря, 1989 г./ Тезисы докладов. Москва, 1989, с. 207-208.
174. Махмудов М.А. Разработка систем формального морфологического анализа тюркской словоформы. Материалы первой межреспубликанской конференции молодых лингвистов-тюркологов. Алма-ата, «Наука», 1985, с. 118-125.
175. Махмудов М. Система автоматической переработки тюркского текста на лексико-морфологическом уровне. Баку, «Элм», 1991, –143 с.
176. Марчук Ю.Н. Основы компьютерной лингвистики. М., Изд.-во МГУ, 2000, –227 с.
177. Машинный фонд русского языка: идеи и суждения. М., Наука, 1986, –240 с.
178. Марчук Ю.Н. Основы компьютерной лингвистики. М.МГОУ, 2002, –234 с.
179. Машинный перевод и прикладная лингвистика. Выпуск 6. Москва, I Московский гос. пед. ин-т иностранных языков, 1961, –100 с.
180. Мельчук И.А. Автоматический синтаксический анализ. Том I, Новосибирск АН СССР, 1964, – 359 с.

181. Мельчук И.А., Равич Р.Д. Автоматический перевод. (1949-1963). Москва, 1967, –517 с.
182. Моделирование языка и машинный перевод. Иркутск, 1979, –92 с.
183. Москович В.А. Информационные языки. Москва, «Наука», 1971, –144 с.
184. Налимов В.В. Вероятностная модели языка. Москва, «Наука», 1974, –272 с.
185. Нелюбин Л.Л. Перевод и прикладная лингвистика. Москва, «Высшая школа», 1983, –207 с.
186. Никитин П.И. Автоматизированные системы обработки и поиска документальной информации. Москва, «Статистика», 1977, –136 с.
187. Никитина С.Е. Тезаурус по теоретической и прикладной лингвистике (автоматическая обработка текста). Москва, «Наука», 1978. –374 с.
188. Новицкая И.М. Семантическое предсказание на уровне текста. Тезисы докладов конференции «Семантика естественных и искусственных языков в специализированных системах». Ленинград. изд. Ленинградского университета, 1979, с.128-129.
189. Обратный частотный словарь латышского языка. Рига, изд. «Зинатне», 1976, –500 с.
190. Описание и распознавание объектов и системах искусственного интеллекта. Москва, «Наука», 1980. –37 с.
191. Пашенко Н.А. Семантический анализ омонимичных и многозначных слов при автоматической индексировании русского научно-технического текста. Тезисы докладов конференции «Семантика естественных и искусственных языков в специализированных системах». Ленинград, изд. Ленинградского университета, 1979, с. –135.
192. Пинес В.Я., Махмудов М.А. Некоторые аспекты построения систем МП с тюркских языков. Меж-

- дународный семинар по машинному перевод. Москва, 1983, с. –177.
- 193.Пинес В.Я. Моделирование структуры азербайджанских глагольных форм в связи с проблемой автоматического словаря. АКД. Москва, 1970, –19 с.
- 194.Пиотровская А.А. Автоматическое приведение именных словоупотреблений к канонической форме. Научно-техническая информация, сер.2, №1, 1977, с. 32-36.
- 195.Пиотровский Р.Г., Аникина Н.В., Аполлонская Т.А., Билан В.Н., Боркун М.Н., Лесохин М.М., Щингарева Е.А. Формальное распознавание смысла текста. В кн.: Статистика речи и автоматический анализ текста. Ленинград, «Наука», 1980, с. 5-51.
- 196.Пиотровский Р.Г., Бектаев К.Б., Пиотровская А.А. Математическая лингвистика. Москва. «Высшая школа», 1977. –383 с.
- 197.Пиотровский Р.Г., Бектаев К.Б. Машинный перевод: теория-эксперимент-реализация. Вопросы языкознания, № 5, 1977, с. 27-33.
- 198.Пиотровский Р.Г. Инженерная лингвистика и теория языка. Ленинград, «Наука», 1979. –112 с.
- 199.Пиотровский Р.Г. Инженерная лингвистика: теория-эксперимент-реализация. изд. АН СССР, сер. лит. и языка, №1, 1978, с. 10-19.
- 200.Пиотровский Р.Г. Информационные измерения языка. Ленинград, «Наука», 1968, –116 с.
- 201.Пиотровский Р.Г. Лингвистические аспекты «искусственного разума». Вопросы языкознания, №3, 1981, 27-39 с.
- 202.Пиотровский Р.Г. Текст, машина, человек. Ленинград, «Наука», 1975, –327 с.
- 203.Покровская Л.А. Грамматика гагаузского языка. Фонетика и морфология. Москва, «Наука», 1964, –298 с.

204. Поцелуевский А.П. Избранные труды. Ашхабад, «Бллым», 1975, –337 с.
205. Проблемы прикладной лингвистики. Часть I. Тезисы межвузовской конференции. Москва, пед. ин-т иностр. языков им. Мориса Тореза, 1969, –192 с.
206. Пумпянский А.Л. Информационная роль порядка слов в научной и технической литературе, Москва, «Наука», 1974, –247 с.
207. Рахманов Дж.А. Статистико-дистрибутивный анализ азербайджанского текста (на уровне графем и фонем). АКД, Баку, 1988, –23 с.
208. Ревзин И.И. Пути преодоления кризиса в вычислительной лингвистике. НТИ, сер. 2. №2, 1968.
209. Ревзин И.И., Розеньвейг В.Ю. Основы общего и машинного перевода. Москва, Высшая школа, 1964. – 242 с.
210. Реформатский А.А. Очерки по фонологии, морфонологии и морфологии. Москва, «Наука», 1979, –103 с.
211. Розеньвейг В.Ю. Семантическая валентность и автоматический анализ текста. тезисы докладов конференции «Семантика естественных и искусственных языков в специализированных системах». Ленинград, изд. Ленинградского университета. 1979, – 147 с.
212. Садыков Т. Проблемы моделирования тюркской морфологии. Фрунзе, «Илим», 1987, –121с.
213. Самойлович А.Н. Краткая учебная грамматика современного османско-тюркского языка. Ленинград. 1926. –154 с.
214. Сахарный Л.В. Некоторые закономерности построения текста на естественном языке и их использование в информационном поиске. Тезисы докладов конференции «Семантика естественных и искусственных языков в специализированных сис-

- темах». Ленинград. изд. Ленинградского университета, 1979, с. 154-156.
215. Севба И.П. Анализы предложения в общей теории синтаксиса. В. сб.: «Структурная и математическая лингвистика», Киев, изд. Киевского университета, 1979, с. 82-98.
216. Сейгер Н. «Синтаксический анализ естественного языка». В сб.: «Автоматический перевод». Москва, «Прогресс», 1971, с. 182-213.
217. Семантика естественных и искусственных языков в специализированных системах. Тезисы. Ленинград, изд. Ленинградского университета, 1979, –192 с.
218. Солнцев В.М. Относительно концепции «Глубинной структуры». Вопросы языкознания, №5, 1976, с. 13-25.
219. Солнцев В.М. Язык как системно-структурное образование. Москва, «Наука», 1977, –341 с.
220. Статистика казахского текста. Выпуск II. Труды группы автоматизации лингвистических исследований и статистика речи. Алма-ата, «Гылым», 1990, –207 с.
221. Статистика речи и автоматический анализ текста. «Наука», 1971, –464 с.; 1973, –339 с.; 1974, –404 с.
222. Стрелковский Г.М., Латышев Л.К. Научно-технический перевод. Москва. «Просвещение», 1980, –175 с.
223. Структурная и математическая лингвистика. Киев. изд. Киевского университета, 1979, –121 с.
224. Структурная и прикладная лингвистика. Межвузовский сборник. Выпуск I. Ленинград, изд. Ленинградского университета, 1978, -208 с.
225. Сухотин Б.В. Оптимизационные методы исследования языка. Москва, «Наука», 1976, –170 с.
226. Сэлтон Б.Г. Автоматическая обработка, хранение и поиск информации. Москва, «Советское радио», 1973, –560 с.

227. Теоретические и экспериментальные исследования в области структурной и прикладной лингвистики. Москва, изд. Московского университета, 1973, –303 с.
228. Теоретические проблемы прикладной лингвистики. Москва, изд. Московского университета, 1965, –138 с.
229. Труб В.М. К проблеме семантической интерпретации некоторых атрибутивных связей в русском языке. В сб.: «Структурная и математическая лингвистика», Киев, изд. Киевского университета, 1979, с. 89-95.
230. Уемов А.И. Логические ошибки. Как они мешают правильно мыслить. изд. Политической литературы. Москва, 1958, –119 с.
231. Успенский В.А. Теорема Геделя о неполноте. Москва, «Наука», 1982, –111 с.
232. Фитиалов С.Я. О моделировании синтаксиса для машинного перевода. Доклады на конференции по обработке информации, машинному переводу и автоматическому чтению текста. Выпуск 10, Москва, 1961, –13 с.
233. Фитиалов С.Я. О построении формальной морфологии в связи с машинном переводом. Москва. АН СССР, Инс-т Научн. информации. Доклады на конференции по обработке информации машинному переводу и автоматич. чтению текста, вып. 2, 1961, – 24 с.
234. Фрумкина Р.М. Статистические методы изучение лексики. Москва, «Наука», 1964, –110 с.
235. Халитова Н.А. Начальный этап работы над машинным переводом (МП) с татарского на русский. - В кн.: «Проблемы тюркологии и истории востоковедения» (материалы конференции), Казан, изд. Казанского университета, 1962, с. 266-274.
236. Хойл В.Г. Автоматическая классификация и индексирование. Международный конгрес по научной

- информации (Москва, 16-18 сентября 1968 г.),
Москва, ВИНТИ, 1968, -20 с.
237. Хомский Н. Аспекты теории синтаксиса. Москва, изд. Московского университета, 1972, -259 с.
238. Хомский Н. Синтаксические структуры. В сб.: «Новое в лингвистике», вып.2, Москва, изд. Инст. лит., 1962, с. 412-527.
239. Хомский Н. Формальные свойства грамматик. -В кн.: «Кибернетический сборник», Новая серия, вып. 2, Москва, «Мир», 1966, с. 121-230.
240. Шаляпина З.М. К проблеме коммуникативной организации текста и её отражения в семантической записи. В сб.: «Вычислительная лингвистика», Москва, «Наука», 1976, с. 53-80.
241. Шемакин Ю.М. Тезаурус в автоматизированных системах управления и обработки информации. Москва, Воениздат, 1974, -188 с.
242. Шерба Л.В. опыт общей теории лексикографии. -В кн.: «языковая система и речевая деятельность», Ленинград, 1974.
243. Ширалиев М.Ш. Азербайджанский язык. Баку, «Элм», 1989, -48 с.
244. Шовкун В.И. Некоторые вопросы синтаксиса, семантики и их взаимосвязи в системе общего лингвистического описания. -В сб.: «Теоретические и экспериментальные исследования в области структурной и прикладной лингвистики», Москва, изд. Московского университета, 1973, с. 248-278.
245. Штерн И.Б. О лингвистической проблематике построения систем автоматического распознавания и понимания текстов. В сб.: «Структурная и математическая лингвистика» Киев, изд. Киевского университета, 1979, с. 103-108.
246. Шуляк Н.Т., Макаренко В.Г. К вопросу о статистическо описании синтаксических конструкции. В

- сб.: «Структурная и математическая лингвистика», Киев, изд. Киевского университета, 1979, с. 108-115.
247. Шербак А.М. Очерки по сравнительной морфологии тюркских языков (имя), Ленинград, «Наука», 1977, – 191 с.
248. Шербак А.М. Способы выражения грамматических значений в тюркских языках. Вопросы языкознания, №1, 1957, с. 18-26.
249. Якубова Н.К. Расстановка аффиксов в узбекских словоформах. Ташкент. ФАН, 1979, –122 с.
250. Частотный словарь русского языка. Москва, изд. «Русский язык», 1977, –935 с.
251. Чемерис Н.Н. Моделирование связности текста при переводе с коммуникативно-информационного языка ситуации на русский язык. В сб.: «Структурная и математическая лингвистика», Киев, изд. Киевского университета, 1979, с. 95-103.
252. Чижиковский В.А., Беляева Л.Н. Тезаурус в системах автоматической переработки текста, Кишинев, «Штиинца», 1983, –163 с.
253. Чубарев А.М., Холодный В.С. Невероятная вероятность (О прикладном значении теории вероятностей), Москва. изд. «Знание», 1976, –127с.
254. Чхаидзе М.П. Ещё раз о мифах и правде машинного перевода. «Вопросы языкознания», №5, 1970, с. 50-57.
255. Леонтева Н.Н. Автоматическое понимание текстов: Системы, модели, ресурсы. Москва, Академия, 2006, 304 с.
256. Селегей В. Электронные словари и компьютерная лексикография // [http://www linevoda.ru/](http://www.linevoda.ru/) – www.tech-tr.ru
257. Рублева О.С. Слово в электронном словаре. Диссертация на соиск. уч. степ. канд. наук, Киров, 2010, 160 с.
258. Вадяев С.Е. Электронная лексикография и корпусная лингвистика. Аспекты становления и функционирования

- ния западно-германских языков. Самара: Изд-во «Самарский ун-т», 2003, с. 83-92.
259. Гольдин В.Е., Сдобнова А.П. Русская ассоциативная лексикография: Учеб. пособие. Саратов: Научная книга, 2008. 77 с.
260. Городецкий Б.Ю. Компьютерная лингвистика: моделирование языкового общения // Новое в зарубежной лингвистике. М.: Прогресс, 1989. Выпуск XXXIV: Компьютерная лингвистика. с. 5-29.
261. Карпова О.М. словари нового столетия: поиски и решения // Лексика и лексикография: Сб. науч. тр. М., Ин. Языкознания РАН, 2001. Вып. 12, с. 54-64.
262. Национальный корпус русского языка. <http://www.ruscorpora.ru/> [10.06.2008].
263. Нелюбин Л.Л. Компьютерная лингвистика и машинный перевод. М., ВЦП, 1991, 154 с.
264. Черкасова Г.А. Русский ассоциативный тезаурус: компьютерная создания и издания // Этнокультурная специфика языкового сознания/ Под ред. Н.В. Уфимцевой. М., 1998, с. 129-134.
265. Шаров С.А. Корпусная лингвистика. <http://www.dialog-21.ru>
266. Большой энциклопедический словарь <http://www.edic.ru/res/>
267. Шаров С.А. Частотный словарь русского языка <http://www.artint.ru/projects/frylist.asp>.
268. Словарь Кембриджского унив. Dictionary. cambridge.org.
269. Словарь компьютерных терминов. foldoc.doc.ic.ac.uk.
270. Мамедова М.Г., Мамедова З.Ю. Машинный перевод эволюция и основные аспекты моделирования. Баку, изд. «İnformasiya texnologiyaları», 2006, –158 с.
271. Искендерова Н.А. Программно-частотный и качественно-семантический анализ адекватности переводов и его применение к переводам на различных языках.

- Автореф.канд.дисс., Баку, И-т Языкознания им. Насими, 2008, –196 с.
- 272.Ивашенко В.М. Язык и стиль научно-технического текста (на материалах английского азербайджанского и русского языков). Автореф. канд. дисс., Баку, 2009, –42 с.
- 273.Махмудов М.А. Лингвостатистика: проблемы, перспективы. Современное казахское языкознание: Актуальные вопросы прикладной лингвистика. Международная научно-теоретическая конференция посвященная 75-летию юбилею профессора Жубанова Аскара Кудайбергенулы, Алматы, 2012, с. 10-14.
- 274.Залевская А.А. Слово о лексикона человека: психолингвистическое исследование. Воронеж, изд. ВГУ, 1990, – 206 с.
- 275.Залевская А.А. Введение в психолингвистику. Москва, Российский гуманитарный университет, 2007. - 560 с.
- 276.Черкасова Г.А. Русский ассоциативный тезаурус: компьютерная технология создания и издания // Этнокультурная специфика языкового сознания / Под ред. Н.В.Уфимцевой, Москва, 1996, с. 181-190.
- 277.[www/mfbl.ru](http://www.mfbl.ru)
- 278.<http://www.cfri.ru>
- 279.Машинные фонды языков народов СССР: Материалы рабочего совещание (Тбилиси 15-22 сент. 1987 г.) Тбилиси, 1988.
- 280.Машинные фонды языков народов СССР: Материалы рабочего совещание (Таллин, 19-22 декабря, 1988 г.) Таллин: Инст.Яз. и литературы, АН Эст ССР, 1988, 21 с.
- 281.<http://www.ruscorpora.ru/corpora-info>
- 282.<http://www.ruscorpora.ru/corpora-structurl.html>
- 283.<http://www.ruscorpora.ru/>
- 284.info@corpora.ru

285. <http://www.dilmanc.az/az>
286. <http://az.wikipedia.org/wiki/dilmanc>
287. www.poliglot.az
288. <http://www.multilax.ru>
289. <http://www.krugosvet.ru>
290. Linqvo Online linqvo.ru
291. Britannica Online <http://www.eb.com>
292. Online Dictionaries. <http://www.onelook.com>
293. Britannica On-Line <http://www.eb.com>.
294. English-Russian Online Dictionary. http://www.terem.ru/cgi-bin/www_dic
295. English-English Dictionary. <http://www.dictionary.com/>
296. On-Line Dictionaries <http://www.onelook.com/>
297. On-Line Dictionaries // <http://www.factstaff.bucknell.edu//rbeard>
298. Oxford Dictionaries <http://www.oup.com/>
299. Bank of English // http://titania.cobuild.collins.co.uk/bol_info.html/
300. Bektaev K.B. Kəzəkı tilinin kerı alfavitti sezdiqi. Almatı, 1971, -173 s.
301. Banguoğlu T. Türkçenin qrameri. 3-cü basılış. Türk Tarih Kurumu Basım evi, Ankara, 1990.-628 s.
302. Kunğırov R., Tixonov A. Uzbek tilining çappa luğati, Samarkand, 1968.
303. Gödel K. Üeber formal unentscheşdbare satze der pricipia matematika und verwandter systeme. Monatshette für mathematiks and phusiks, xxxviii, 1931.
304. Vşetze H.P., Zenker Z., Warnke I. Rückaufiges Wörterbuch der türkischen sprache. Leipzig, 1975.
305. Veliyeva K.A., Mahmudov M.A., Pines V.Y. Automatic dictionary in the system of Azerbaijan-Russian mashine translation. Symposium on automatic compilation of dictionary, Tallin, november, 25-27, 1985, p. 81.
306. Kennedy G. An introduction tu Corpus, Linguistics / Addisson Wesley: Longman Ltd, 1998, 315 p.

MÜNDƏRİCAT

“KOMPÜTER DİLÇİLİYİ” ƏSƏRİNƏ ÖN SÖZ VƏ YA
DİLÇİLİK MÜASİR DƏQİQ ELMLƏR ORBİTİNDƏ3

MÜƏLLİFDƏN.....12

I FƏSİL. KOMPÜTER DİLÇİLİYİ DİLÇİLİYİN YENİ BİR
SAHƏSİ KİMİ.....18

- 1.1. Ənənəvi dilçiliyə qeyri-ənənəvi baxış18
- 1.2. Kompüter dilçiliyi anlayışı24
- 1.3. Kompüter dilçiliyinin sahələri27
- 1.4. Mətn kompüter dilçiliyinin obyektı kimi.....28
- 1.5. Kompüterlərdə mətnlərin formal işlənmə problemi36

II FƏSİL. AZƏRBAYCAN DİLİNİN MİLLİ KORPUSU43

- 2.1. Ümumi məsələlər.....43
- 2.2. Milli dil korpusları.....44
 - 2.2.1. Türk dillərinin milli korpusu46
 - 2.2.2. Başqırd dilinin milli korpusu.....48
 - 2.2.3. Rus dilinin milli korpusu51
 - 2.2.3.1. Əsas mətn korpusu.....54
 - 2.2.3.2. Müasir yazılı mətnlər.....55
 - 2.2.3.3. Dərin annotasiyalı (sintaktik) korpus.....56
 - 2.2.3.4. Qəzet korpusu (müasir KİV korpusu)56
 - 2.2.3.5. Paralel mətnlər korpusu57
 - 2.2.3.6. Dialekt mətnləri korpusu57
 - 2.2.3.7. Poetik mətnlər korpusu57
 - 2.2.3.8. Rus dilinin tədris korpusu.....58
 - 2.2.3.9. Şifahi nitq korpusu.....58
 - 2.2.3.10. Aksentoloji korpus.....58
 - 2.2.3.11. Multimediya korpusu.....59
- 2.3. Azərbaycan dilinin milli korpusunun yaradılmasının
nəzəri və təcrübi məsələləri, strukturu.....59

2.3.1. Məsələnin qoyuluşu.....	59
2.3.2. Azərbaycan dilinin milli korpusunun nəzəri və təcrübi bazası	60
2.3.3. Azərbaycan dilinin milli korpusunun strukturu.....	61
III FƏSİL. KOMPÜTER LÜĞƏTÇİLİYİ.....	64
3.1. Lüğətçiliyin inkişafında yeni mərhələ	64
3.2. Kompüter lüğətçiliyi anlayışı	65
3.3. Kompüter lüğətlərinin kağız (kitab) lüğətlərdən üstünlükləri	67
3.4. Kompüter lüğətləri elektron resurs kimi.....	68
3.4.1. Tərcümə kompüter lüğətləri	72
3.4.2. Çoxdilli Lingvo 8.0 elektron lüğəti	73
3.4.3. Context 4.0 elektron lüğəti (“İnformatik” kompaniyası)	74
3.4.4. Polyglossum 3.2 elektron lüğəti	75
3.4.5. Elektron ensiklopediyalar	78
3.4.6. Assosiativ lüğətlər	80
3.5. Azərbaycanda kompüter lüğətçiliyi.....	81
3.5.1. Kompüterin köməyi ilə tərtib olunmuş lüğətlər	82
3.5.2. İnternetdə yerləşdirilmiş elektron lüğətlər.....	84
IV FƏSİL. MƏTNİN FORMAL TƏHLİLİ SİSTEMİ AZƏRBAYCAN DİLİNİN MİLLİ KORPUSUNUN TƏRKİB HİSSƏSİ KİMİ.....	89
4.1. Mətnin formal təhlili sisteminin Azərbaycan dilinin milli korpusunda yeri	89
4.2. Azərbaycan dilinin milli korpusunun linqvistik bölümü...91	
4.3. Mətnin formal təhlili sisteminin ümumi məsələləri	94
4.3.1. Maşın lüğətləri.....	94
4.3.2. Avtomatik linqvistik təhlil və sintez.....	96
4.3.3. Maşın tərcüməsi.....	97
4.3.4. Yarımavtomatik tərcümə	104
4.3.5. Süni intellekt maşın tərcüməsi sistemlərində	105
4.4. Türk dillərinin maşın tərcüməsi sistemi	107

V FƏSİL. MƏTNİN FORMAL TƏHLİLİ SİSTEMİNİN	
LEKSİK BLOKU	117
5.1. Leksik blok mətnlərin formal təhlili sisteminin aparıcı komponenti kimi.....	117
5.2. Statistik lüğətçilik	119
5.3. Tezlik lüğətləri.....	119
5.3.1. Tezlik lüğətlərinin növləri	120
5.3.2. Tezlik lüğətlərinin əhəmiyyəti və tətbiqi.....	123
5.3.3. Azərbaycan dilinin tezlik lüğəti.....	128
5.3.4. Tezlik lüğətlərinin statistikasını.....	131
5.3.5. Tezlik lüğətlərinin tərtibi qaydaları	132
5.4. Azərbaycan dilinin əks lüğəti	149
5.5. Abidələrin statistik təhlili və sözlüklərin tərtibi	153
5.6. Azərbaycan dilinin avtomatik lüğəti.....	168
5.6.1. Əsaslar avtomatik lüğəti	169
5.6.2. Birləşmələr avtomatik lüğəti	174
5.7. Tezauruslar mətnlərin işlənməsi sistemlərində.....	179

VI FƏSİL. MƏTNİN FORMAL TƏHLİLİ SİSTEMİNİN	
MORFOLOJİ BLOKU	181
6.1. Mətnin formal təhlili sistemində morfoloji blokun yeri	181
6.2. Maşın qrammatikası anlayışı	184
6.3. Türk dillərində nitq hissələrinin təsnifi	187
6.4. Nitq hissələrinin formal əlamətlərə görə təsnifi	192
6.5. Morfoloji blokun funksiyası	196
6.6. Formal təhlil anlayışı	197
6.7. Sözdəyişdirici şəkilçilərinin morfoloji blokda yerləşdirilməsi	200
6.8. Morfoloji blokda nitq hissələrinin, ayrı-ayrı qrammatik kateqoriyaların kodlaşdırılması	216
6.9. Morfoloji təhlilin əsas istiqamətləri və mərhələləri.....	217
6.10. Azərbaycan mətnlərinin formal morfoloji təhlil alqoritmi.....	231

VII FƏSİL MƏTNİN FORMAL TƏHLİLİ SİSTEMİNİN	
MORFONOLOJİ BLOKU	239
7.1. Morfonoloji blokun funksiyası	239
7.2. Söz kökündə morfonoloji dəyişmələr	240
7.3. Söz kökü ilə süzdüzəldici şəkilçi sərhədində morfonoloji dəyişmələr	242
7.4. Söz kökü ilə sözdəyişdirici şəkilçi sərhədində morfonoloji dəyişmələr	244
7.5. Süzdüzəldici şəkilçi ilə sözdəyişdirici şəkilçi sərhədində morfonoloji dəyişmələr	247
7.6. Sözdəyişdirici şəkilçilər sərhədində morfonoloji dəyişmələr	249
VIII FƏSİL. MƏTNLƏRİN FORMAL TƏHLİLİ SİSTEMİNİN	
SİNTAKTİK BLOKU	251
8.1. Sintaktik blokun funksiyası	251
8.2. Aralıq-keçid dili (vasitəçi dil)	253
8.3. Sintaktik təsadüflər və ümumiləşdirmə	255
8.4. Sintaktik modellər	255
8.5. Universal qrammatikalar nəzəriyyəsi	259
8.6. Təhlil sistemlərinə transformasiya qrammatikalarının tətbiqi	264
8.7. N.Xomskinin linqvistik fərziyyələrinin mahiyyəti və praktik istifadə dairəsi	265
8.8. Bərpəedici mühəndis dilçiliyi modelləri	267
8.9. Sintaktik blokun əsas komponentləri	268
8.10 . Xəbərin formal əlamətləri	270
8.10.1. Felin əmr şəkli ilə ifadə olunmuş xəbərlər	270
8.10.2. Felin xəbər şəklinin xəbərəmələgətirmə imkanları	272
8.11. İsmi xəbərlər	275
8.12. Feli xəbərləri ifadə edən formal modellər	276
8.13. Mübtədanın formal əlamətləri	280
8.14. Uzlaşma əlaqəsinin formal təsviri	284
8.15. Azərbaycan dilinin avtomatik sintaktik təhlili alqoritmi	289

IX FƏSİL. MƏTNLƏRİN FORMAL TƏHLİLİ SİSTEMİNİN SEMANTİK BLOKU.....	297
9.1. Məsələnin ümumi qoyuluşu. Sintaktik əhatədə çoxmənalılığın və omonimliyin aradan qaldırılması.....	297
9.2. Morfoloji səviyyədə əsas və şəkilçilərin omonimliyi....	300
9.3. Omonimliyin formal əlamətə görə aradan qaldırılması.	302
9.4. Şəkilçi birləşmələri vasitəsi ilə omonimliyin aradan qaldırılması.....	303
9.5. Maşın şəkilçiləri vasitəsi ilə omonimliyin aradan qaldırılması.....	304
9.6. Morfoloji səviyyədə aradan qaldırılmalı bilinməyən omonimlik.....	305
9.7. Əsasların omonimliyinin aradan qaldırılması algoritmi	306
9.8. Şəkilçilərin omonimliyinin aradan qaldırılması algoritmi.....	308
9.9. Tezauruslar vasitəsi ilə omonimliyin və çoxmənalılığın aradan qaldırılması.....	309
SONLUQ.....	314
QISALTMALAR VƏ ŞƏRTİ İŞARƏLƏR.....	318
İSTİFADƏ OLUNMUŞ ƏDƏBİYYAT.....	321

MƏSUD MAHMUDOV

KOMPÜTER DİLÇİLİYİ

«Elm və təhsil» nəşriyyatının direktoru:

professor Nadir MƏMMƏDLİ

Dizayn:

Zahid Məmmədov

Texniki redaktor:

Rövşənə Nizamiqızı