

İNFORMATİKA

AZƏRBAYCAN DİLİNDƏ SÖZ-FORMALARININ VƏ SÖZ
BİRLƏŞMƏLƏRİNİN SİNTEZİ MƏSƏLƏSİNDƏ MÜMKÜNLÜK
FUNKSİYALARINDAN İSTİFADƏ

Z.M. ƏMİROV

MTN-nin Heydər Əliyev adına Akademiyası

zamirov@inbox.ru

Məqalədə Azərbaycan dili üçün mətnlərin avtomatik emalı sistemlərinin yaradılması zamanı söz-forma və ya söz birləşmələrinin sintezi məsələsinə "Mümkünlük funksiyaları"nın tətbiqi şərh edilmişdir.

Məlumdur ki, hazırkı dövrdə insanlar tərəfindən istifadə olunan informasiyanın əsas hissəsi təbii dillərdə tərtib olunmuş elektron sənədlərdə saxlanılır. Bu səbəbdən də süni intellekt sistemlərinin bir qolu kimi təbii dili müxtəlif səviyyələrdə analiz edən, onu "başla düşən" və təbii dilin daxili qanunlarına əsasən, onun müxtəlif elementlərini sintez edə bilən proqramların yaradılması aktualıq kəsb edir. Belə proqramlar təbii dilin emalı sistemləri (*natural language processing systems*) və ya **lingvistik prosessorlar** adlanır. Təbii dillərdə hazırlanmış mətnlərin avtomatlaşdırılmış emalı (MAE) sistemləri çox böyük praktiki əhəmiyyətə malikdir. Maşın tərcüməsi sistemlərinin, mətnlərin düzgünlüyünün yoxlanılması sistemlərinin, avtomatik öyrədici sistemlərin, nitqin tanınması və emalı sistemlərinin, İnternet axtarış sistemlərinin yaradılması, həmçinin mətnlərin avtomatlaşdırılmış emalına aid digər sahələrdə tədqiqatlar getdikcə daha da genişlənir.

Linqvistik prosessorların effektivliyini təmin edən başlıca şərt təbii dillərin formal modellərinin hazırlanmasıdır. Formal modellərin hazırlanması isə dilin leksik, morfoloji, sintaktik, semantik səviyyələrdə analizi və sintezi problemlərinin öyrənilməsindən keçir.

Son 20-25 ildə linqvistik prosessorların layihələndirilməsi və onların tətbiqi aşağıdakı yeni və maraqlı sahələr hesabına genişləndirilmişdir [2,5]:

- tətbiqi intellektual sistemlərin təbii-dil interfeysi (TD-interfeysi);
- ekspert sistemlərinin (tibbi diaqnostika, texniki diaqnostika və s.) istifadəçiləri üçün təbii dildə, tövsiyə xarakterli mətnlərin sintezi;
- çoxagentli kompyuter sistemlərində intellektual kompyuter agentlərinin yaradılması;
- elektron ticarət sahəsində tətbiqi nəzərdə tutulan danışıqlar protokolu dilinin yaradılması;
- yeni nəsil semantik veb-texnologiyalar üçün semantik-struktur dillərin yaradılması;

- intellektual informasiya-axtarış sistemlərinin layihələndirilməsi;
- elektron kitabxanaların yaradılması. Elektron kitabxanaların effektivliyini yüksəldən əsas elmi məsələlərdən biri informasiya mənbələrindən informasiyanın mənaya görə axtarışının təmin olunması məqsədilə mətnlərin semantik analizinin avtomatlaşdırılmasıdır;
- mətnlərdən biliklərin avtomatik şəkildə çıxarılması istiqamətində layihələrin işlənməsi;
- İnternet şəbəkəsinin insan fəaliyyətinin müxtəlif sahələrini əhatə edən çox geniş informasiya ehtiyatlarından səmərəli şəkildə istifadə etməyə can atan hər bir kompyuter istifadəçisi üçün TD-interfeyslərinin mövcudluğu zəruridir. İnternetin əsas aparıcı təşkilatlarından biri hesab olunan *W3C (the World Wide Web Consortium)* konsorsiumu 2001-ci ilin fevralında mövcud veb-şəbəkə xidmətinin semantik şəbəkə xidməti (*Semantic Web*) ilə əvəz olunması istiqamətində çoxşaxəli tədqiqatlara başlanıldığını rəsmən bəyan etmişdir. Bu iri miqyaslı layihənin ən vacib cəhətlərindən biri İntellektual Kompyuter Agentlərindən (İKA) istifadənin gerçəkləşdirilməsi ideyasıdır. İKA bir-biri ilə qarşılıqlı əlaqəli surətdə veb-saytlarda təqdim olunmuş informasiyanı analiz edə bilər. Ən maraqlısı isə İKA-nın elektron sənədlərdə, veb-saytlarda yerləşdirilmiş təbii dildə olan mətnlərin məna analizini yerinə yetirə bilməsidir. Bu ideyanın gerçəkləşməsi, İnternet istifadəçilərinə informasiyanın axtarışını təkcə açar sözlərə görə deyil, həm də İKA-nın köməyi ilə mənaya görə axtarmağa da imkan yaratmış olar.

MAE sistemlərinin müasir variantlarında təbii dillərin semantik analiz və sintez məsələləri olduqca aktualdır və layihələndirilən hər bir sistemin uğurla sənaye miqyasında tətbiq oluna bilməsi məhz bu məsələlərin həllindən asılıdır. Təbii dilin semantik-sintaktik analiz alqoritmlərinin qurulması isə morfoloji və sintaktik analiz mərhələlərinin nəticəsi olan biliklər toplusundan, semantik səviyyədə leksik vahidlərin qarşılıqlı əlaqəsi haqqında məlumatlardan və digər biliklərdən asılıdır. Aqlütinativ, yəni şəkilçilərin sayı çox olan dillərdə bu asılılıq daha qabarıq şəkildə özünü büruzə verir.

İnkişaf etmiş ölkələrin əksəriyyətində mətnlərin avtomatik emalı sahəsində ciddi nailiyyətlər əldə olunmuşdur. Müxtəlif mətn redaktorlarında daxil edilən mətnlərin avtomatlaşdırılmış şəkildə düzgünlüyünün yoxlanılması, skanerlərin və optik tanıma sistemlərinin köməyi ilə mətnlərin kompyuterlərin yaddaşına daxil edilməsi, müxtəlif elektron lüğətlərdən istifadə, mətnlərin bir dildən digərinə tərcüməsinin avtomatlaşdırılması müasir dövrün reallıqlarıdır. İngilis, alman, fransız, italyan, yapon, çin, rus və s. dillərində mövcud olan belə sistemlərə çoxsaylı misallar göstərmək olar. Lakin Azərbaycan dili üçün son 10-15 ilə qədər belə sistemlər işlənilməmişdir. Həmçinin adı çəkilən bu dillərlə Azərbaycan dilinin strukturundakı köklü fərqlərə görə Azərbaycan dilində mətnlərin avtomatlaşdırılmış emalı məqsədilə bu dillər üçün işlənmiş yanaşmalardan birbaşa istifadə də mümkün deyildir. Belə sistemlərə bir çox sahələrdə kəskin tələbat olduğundan Azərbaycan dili üçün MAE sistemlərinin yaradılması hazırda informasiya texnologiyalarının və linqvistikanın kəsişməsində yerləşən ən aktual problemlərdən birinə çevrilmişdir.

Yuxarıda göstərilən MAE sahələrinin hər birində Azərbaycan dili üçün linqvistik prosessorların tətbiq edilməsi, geniş spektrli linqvistik tədqiqatların aparılması tələb olunur. Dünya dövlətləri və xalqları ilə birbaşa əlaqələrin qurulmasında, müxtəlif sferalarda baş verən inteqrasiya proseslərində daha fəal iştirak etməkdə maraqlı

olan respublikamız üçün yuxarıda qeyd olunmuş məsələlərin əhəmiyyəti olduqca yüksəkdir. Bir tərəfdən yaradılmış mövcud sistemlərin mütəmadi olaraq təkmilləşdirilməsi, digər tərəfdən isə yeni layihələrin gerçəkləşdirilməsi aktual məsələlər olaraq öz həllini gözləyir.

Azərbaycan dilinin leksik və morfoloji analiz modelləri, ayrı-ayrı alqoritm və metodlar, həmçinin bu dilin fonetik, sintaktik, semantik bölmələrinə uyğun formal modellər müəyyən dərəcədə tərcümə sistemləri, elektron lüğətlər, mətnlərin orfoqrafik düzgünlüyünün yoxlanılması, mətnlərin optik tanınması sistemləri (mətnin skaner vasitəsilə sürətlə daxil edilməsini yerinə yetirən *Fine Reader* tipli proqramlar nəzərdə tutulur), dil öyrədən sistemlər kimi MAE sistemlərinin yaradılmasında tətbiq olunmuşdur.

“Mümkünlük qaydaları” və “Mümkünlük funksiyaları”. Azərbaycan dilinin avtomatlaşdırılmış morfoloji analizinin M-Avtomat modelinin yerinə yetirilməsi prosesində istifadə olunan linqvistik bilgiler içərisində “Mümkünlük qaydaları” və “Mümkünlük funksiyaları” adlandırılan anlayışlar xüsusi rol oynayırlar. Belə ki, söz-formaların analizi prosesində əsasla şəkilçilərin, həmçinin şəkilçilər ardıcılığına daxil olan ayrı-ayrı şəkilçilərin bir-birlərinə düzgün birləşib-birləşməməsi məsələsinin araşdırılmasında bilavasitə bu anlayışlara aid qrammatik qaydalardan istifadə edilmişdir [1, 3, 4].

“Mümkünlük qaydaları” cüt-cüt qrammatik kateqoriyalar – şəkilçi qrupları arasında əlaqələrin mümkünlüyünü göstərir və aşağıdakı kimi təsvir olunurlar:

(I Kateqoriyaya daxil olan şəkilçilər qrupu – II Kateqoriyaya daxil olan şəkilçilər qrupu) → Mümkündür.

Bu yazılış onu göstərir ki, I Kateqoriyaya daxil olan şəkilçilər qrupundan heç olmasa biri II Kateqoriyaya daxil olan şəkilçilər qrupundan heç olmasa biri ilə birləşdirildikdə düzgün birləşmə alınır [5]. Bu qaydaların əsl mahiyyətini göstərmək üçün bəzi qrammatik kateqoriyalar arasında münasibətləri əks etdirən aşağıdakı misallara baxaq:

Misal 1. Əsası isim olan söz-formaların analizində rast gəlinən kateqoriyaların müəyyən qrup şəkilçiləri üçün “Mümkünlük qaydaları”na nümunələr:

✓ “Hal” və “Xəbərlilik” kateqoriyalarının müəyyən qrup şəkilçiləri üçün:

(sında, sində, sunda, sündə; da, də; nda, ndə; sından, sindən, sundan, sündən; dan, dən; ndan, ndən – am, əm, yam, yəm; ıq, ik, yıq, yik; san, sən; dır, dir; sınız, siniz; dırlar, dirlər) → **Mümkündür.** Məsələn, *məhəlləsindən dirlər, küçə dəyəm* və s.;

✓ “Xəbərlilik” və “Şərt” kateqoriyalarının müəyyən qrup şəkilçiləri üçün:

(am, əm; ıq, ik, uq, ük; san, sən; dır, dir, dur, dür; sınız, siniz, sunuz, sünüz; dırlar, dirlər, durlar, dürlər – sa, sə) → **Mümkündür.** Məsələn, *tələbə sənə, gedirlərsə* və s.

Misal 2. Əsası fel olan söz-formalara aid müxtəlif kateqoriyalar arasındakı əlaqələri əks etdirən nümunələr:

“Zaman” və “Kəmiyyət” kateqoriyalarının müəyyən qrup şəkilçiləri üçün:

✓ (dı, di, du, dü; mış, miş, muş, müş; ıb, ib, ub, üb; yıb, yib, yub, yüb; ır, ir, ur, ür; yır, yir, yur, yür; mır, mir, mur, mür; acaq, əcək; yacaq, yəcək; ar, ər; yar, yər; maz, məz – lar, lər) → **Mümkündür.** Məsələn, *görmüşlər, almazlar, görüşmüşlər* və s.

“Zaman” və “Şəxs və Xəbərlilik” kateqoriyalarının bir qrup şəkilçiləri üçün:

✓ (dı, di, du, dü – m, n; nız, niz, nuz, nüz) → **Mümkündür.** Məsələn, *gördüm, soruşdunuz, ağladınız, demədiniz* və s.

✓ (miş, miş, muş, müş – am, əm; ıq, ik, uq, ök; san, sən; sınız, siniz, sunuz, sünüz; dır, dir, dur, dür) → **Mümkündür**. Məsələn, *görmüşük, gəlmişsiniz* və s.

✓ (ıb, ib, ub, üb; yıb, yib, yub, yüb – san, sən; sınız, siniz, sunuz, sünüz; dır, dir, dur, dür) → **Mümkündür**. Məsələn, *görübsən, sürüyübsünüz* və s.

“Mümkünlük funksiyaları” isə hər hansı bir şəkilçinin bilavasitə özündən sonra ona birləşən digər şəkilçi ilə düzgün birləşmə əmələ gətirdiyini göstərir və aşağıdakı kimi təsvir olunur:

(Şəkilçi – Şəkilç₁, Şəkilç₂, ... , Şəkilç_N) → Mümkündür.

Bu yazılış onu göstərir ki, şəkilçilər zəncirində “Şəkilçi”dən sonra “Şəkilç₁”, “Şəkilç₂” və nəhayət, “Şəkilç_N” şəkilçilərindən hər hansı biri gələrsə, bu iki şəkilçidən alınan birləşmə düzgündür [5]. Bu anlayışın mahiyyətini dəqiqləşdirmək üçün misallara baxaq:

Misal 3: İsmi “Hal”, “Kəmiyyət”, “Xəbərlik” kateqoriyalarından olan “da”, “lə”, “dır”, “dir”, “dur”, “dür” şəkilçiləri üçün “Mümkünlük funksiya”na nümunələr:

✓ (da – yam, yıq, san, dır, sınız, dırlar, kı) → **Mümkündür**. Məsələn, *otaqdayıq, maşındasınız, qaladadırlar* və s.;

✓ (lə – sə, mi, ik, dir, dirlər, siniz, i, in, im, iniz, imiz, ə, də, dən, lə, miş) → **Mümkündür**. Məsələn, *müəllimlərimiz, mələklərdir, böyüklərdən* və s.;

✓ (dır – sa, mı) → **Mümkündür**. Məsələn, *yarpızdırsa, pambıqdırmı* və s.;

✓ (dir – sə, mi) → **Mümkündür**. Məsələn, *yeməkdirsə, şəkildirmi* və s.;

✓ (dur – sa, mu) → **Mümkündür**. Məsələn, *odundursa, çovğundurmu* və s.;

✓ (dür – sə, mü) → **Mümkündür**. Məsələn, *köpükdürsə, körükdürmü* və s.

Misal 4: Əsası fel olan söz-formaların araşdırılması zamanı rast gəlinən “Mümkünlük funksiyaları”na nümunələr:

✓ (ur – am, uq, san, sunuz, sa, lar, mu) → **Mümkündür**. Məsələn, *dururlar, uçurmu, doluşursunuz, soruşurlar* və s.;

✓ (ür – əm, ök, sən, sünüz, sə, lər, mü) → **Mümkündür**. Məsələn, *sökürük, döyüşürsünüz, sönürmü* və s.;

✓ (acaq – san, sınız, dır, sa, lar, dı, mı) → **Mümkündür**. Məsələn, *alacaqsınız, yazacaqlarmı, batacaqsa* və s.;

✓ (əcək – sən, siniz, dir, sə, lər, di, mi) → **Mümkündür**. Məsələn, *deyəcəksiniz, biləcəksə, gələcəkmi* və s.;

✓ (malı – ya, yıq, san, sınız) → **Mümkündür**, (məli – yəm, yik, sən, siniz) → **Mümkündür**. Məsələn, *satılmalıya, almalısan, baxmalısınız* və s.

✓ (dən – sə, əm, ik, sən, siniz, dir, mi) → **Mümkündür**. Məsələn, *gülməkdənsə, rənglənməkdəndir* və s.

✓ (da – yam, yıq, san, sınız, dır) → **Mümkündür**. Məsələn, *almaqdayıq, yatmaqdasınız* və s.

Düzgünlük məsələsinin tam yoxlanılmasını təmin etmək məqsədilə nitq hissələrinin hər biri hərtərəfli tədqiq edilmiş və əsas qrammatik kateqoriyalar üçün “Mümkünlük qaydaları”, “Mümkünlük funksiyaları” və hər bir nitq hissəsinin bilavasitə özünə düzgün birləşə bilən şəkilçilər çoxluğu müəyyənləşdirilmişdir [5].

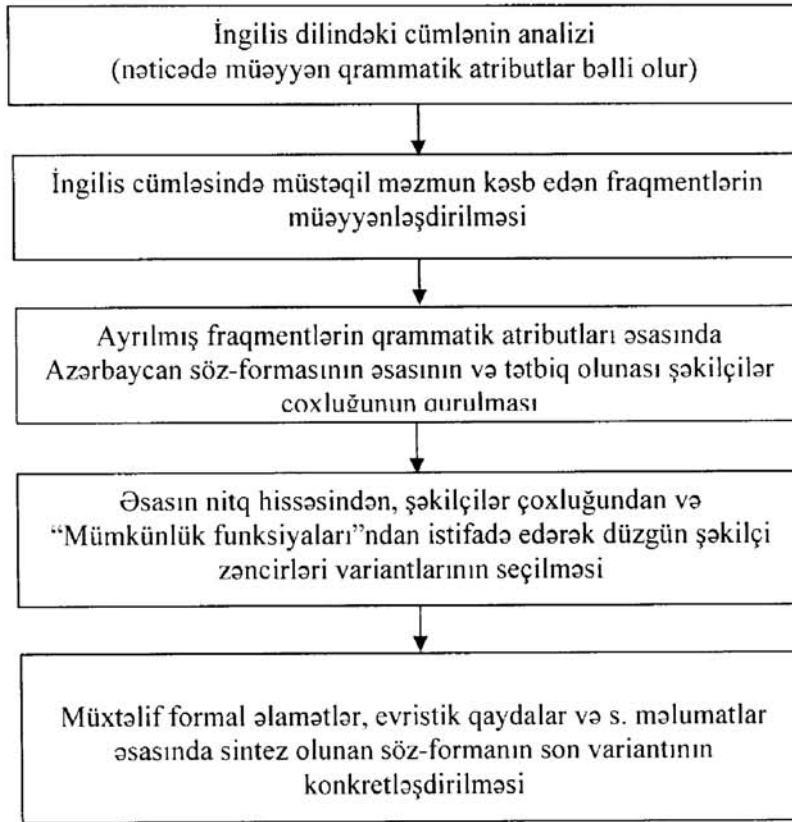
Yuxarıda daxil edilmiş aparat və terminlərdən istifadə edərək aşağıdakı hökmləri söyləmək olar:

Hökm 1. Əgər verilmiş şəkilçi zəncirinin birincidən başlayaraq bütün şəkilçi cütləri hər hansı bir mümkünlük funksiyasının elementidirsə (yəni düzgün şəkilçi cütüyüdürsə), onda şəkilçi zənciri bütövlükdə qrammatik düzgün şəkilçi zənciridir.

Eynilə, bu cür şəkilçi zəncirinin müvafiq əsasa qoşulması ilə alınan söz-forma da düzgün söz-forma olacaqdır.

Höküm 2. Əgər verilmiş şəkilçi zənciri düzgün zəncirdirsə, onda birincidən başlayaraq zənciri təşkil edən hər bir ardıcıl şəkilçi cütü də düzgün şəkilçi cütlüyüdür.

Təkliflər. Yuxarıda təqdim olunmuş hökmlərə əsaslanaraq Azərbaycan dili üçün müxtəlif xarakterli MAE sistemlərinin işlənməsində “Mümkünlük funksiyaları”nı uğurla tətbiq etmək olar. Birinci tətbiq istiqaməti kimi başqa dillərdən Azərbaycan dilinə tərcümə sistemlərində Azərbaycan dilində düzgün söz-formanın, söz birləşməsinin sintez edilməsi prosesini göstərmək olar. Məsələn, ingilis dilindən Azərbaycan dilinə avtomatik tərcümə sistemi üçün söz-formanın sintez prosesinin qlobal alqoritmini sxematik olaraq belə təsvir etmək olar (şəkil 1):



Şəkil 1. Azərbaycan dilində düzgün söz-formanın sintezi.

İkinci tətbiq istiqaməti kimi Azərbaycan dilində yazılmış cümlədə orfoqrafiya qaydalarının avtomatik yoxlanılması məsələsini göstərmək olar. Cümlədəki hər bir söz-formanın orfoqrafik cəhətdən düzgün yazıldığını yoxlamaq üçün əvvəlcə onun əsasını və şəkilçi zəncirini ayırmaq lazımdır (belə ayırma alqoritminin bir variantı [5]-də verilib). Adətən, belə ayırma prosesində əsasın düzgünlüyünün yoxlanılması məsələsi də həll olunmuş olur. Şəkilçi zəncirində hər şeyin düzgün olduğunu yoxlamaq isə daha mürəkkəb məsələdir. Məhz bu məqsədlə “Mümkünlük qaydaları” və “Mümkünlük

funksiyaları” aparatından və yuxarıda verilmiş hökmlərdən istifadə etmək olar. Şəkilçi zənciri elementlərinin cüt-cüt düzgünlüyünü yoxlamaqla zəncirin düzgün olduğunu üzə çıxarmaq olar. Lakin təbii ki, şəkilçi zəncirinin göstərilən nöqtəyi-nəzərdən düzgünlüyü yalnız söz-formada formal olaraq ziddiyyətin olmadığını göstərir, onun qrammatik və məzmun baxımından tam düzgünlüyünə isə zəmanət vermir.

ƏDƏBİYYAT

1. Səfərov S.S., Əmirov Z.M. Azərbaycan dilinin morfoloji təhlilinin avtomat modeli // Azərbaycan Respublikası EA. Xəbərlər. Riyaziyyat, fizika, texnika elmləri seriyası. 1994, №1-2, s. 24-31.
2. Məmmədova M.H., Əmirov Z.M. Maşın tərcüməsi və onun inkişaf perspektivləri // AMEA. Xəbərlər. Riyaziyyat, fizika, texnika elmləri seriyası. Bakı: 2002, №2-3, s. 99-103.
3. Əmirov Z.M. Nitq hissələrinə birləşən şəkilçilərin düzgünlüyünün yoxlanılması məsələsi // AR “Təhsil” cəmiyyəti “Bilgi” dərgisi. Sosial bilgiler seriyası. Bakı: 2003, №2, s. 15-19.
4. Əmirov Z.M. Azərbaycan dilinin morfoloji analizinin M-Avtomat modeli // AMEA. Xəbərlər. Riyaziyyat, fizika, texnika elmləri seriyası. Bakı: 2005, №3, s. 114-118.
5. Əmirov Z.M. Azərbaycan dilinin formal modellərinin yaradılması və onların əsasında linqvistik prosessorun qurulması // Texnika elmləri namizədi alimlik dərəcəsinə almaq üçün işlənmiş dissertasiyanın avtoreferatı. Azərbaycan MEA-nın Kibernetika İnstitutunun kitabxanası. Bakı: 2006, 20 s.

ИСПОЛЬЗОВАНИЕ ФУНКЦИИ ВОЗМОЖНОСТЕЙ В ЗАДАЧЕ СИНТЕЗА СЛОВОФОРМ И СЛОВСОЧЕТАНИЙ НА АЗЕРБАЙДЖАНСКОМ ЯЗЫКЕ

З.М.АМИРОВ

РЕЗЮМЕ

В этой статье изложено применение «Функции возможностей» к задаче синтеза словоформ и словосочетаний при создании систем автоматической обработки текстов для азербайджанского языка.

USE OF POSSIBILITY FUNCTIONS IN THE SYNTHESIS OF WORD-FORMS AND WORD COMBINATIONS IN THE AZERBAIJANI LANGUAGE

Z.M.AMIROV

SUMMARY

The article deals with the application of the Possibility Functions to the synthesis of word-forms, or word combinations during the creation of the systems for the automated processing of text for the Azerbaijani language.