UOT-811.11-112
Səhifə: 200-206
https://doi.org/10.59849/2663-8967.2026.1.200

Natəvan Rəhmanova
AMEA Nəsimi adına Dilçilik İnstitutu
natavanrahmanova@anas.az

AZƏRBAYCAN DİLİNDƏ QOŞMALI FEİLİ BİRLƏŞMƏLƏRİN NLP SİSTEMİ ÜÇÜN MODELLƏŞDİRİLMƏSİ

XÜLASƏ

Məqalədə Azərbaycan dilində qoşmalı feili birləşmələrin struktur xüsusiyyətləri və onların təbii dilin avtomatik emalı (NLP) sistemləri üçün modelləşdirilməsi məsələsi araşdırılmışdır. Tədqiqatın əsas məqsədi Azərbaycan dilində geniş yayılmış sintaktik konstruksiyalardan biri olan qoşmalı feili birləşmələrin formal strukturunu müəyyənləşdirmək və onların kompüter dilçiliyi sistemlərində avtomatik identifikasiyası üçün modellər təklif etməkdir. Məqalədə göstərilir ki, qoşmalı feili birləşmələrin avtomatik müəyyənləşdirilməsi təbii dilin emalı prosesində əsasən sintaktik qruplaşdırma mərhələsində həyata keçirilir. Bu mərhələdə morfoloji analiz nəticəsində əldə olunan nitq hissələri və qrammatik kateqoriyalar haqqında məlumatlardan istifadə olunur. Tədqiqat göstərir ki, təklif olunan modellər Azərbaycan dili üçün müxtəlif NLP tətbiqlərində istifadə oluna bilər. Bu tətbiqlərə morfoloji analiz sistemləri, sintaktik qruplaşdırma alqoritmləri, avtomatik sintaktik təhlil sistemləri, maşın tərcüməsi modelləri və milli korpusların avtomatik morfoloji və sintaktik annotasiyası daxildir. Qoşmalı feili birləşmələrin formal modelləşdirilməsi Azərbaycan dilində mürəkkəb sintaktik konstruksiyaların avtomatik analizinə imkan yaradır və kompüter dilçiliyi baxımından Azərbaycan dili üçün NLP texnologiyalarının inkişafına mühüm töhfə verir.

Açar sözlər: *söz birləşməsi, qoşmalı feili birləşmələr, morfoloji modelləşdirmə, sonlu avtomat morfologiyası, qaydalara əsaslanan NP, NLP.*

Giriş: Azərbaycan dilində müxtəlif növ söz birləşmələri vardır ki, onların sistemli şəkildə öyrənilməsi əsasında müəyyən modellərin yaradılıb NLP-də, həmçinin maşın tərcüməsi sistemində, milli korpusların qurulmasında, sözlərin avtomatik sıralanmasının bir çox məsələlərinin həllində tətbiqi xüsusi maraq doğurur. Bu məqalənin məqsədi Azərbaycan dilində qoşmalı feili birləşmələrin struktur modellərini müəyyənləşdirmək və onların təbii dilin avtomatik emalı sistemlərində istifadə oluna biləcək formal təsvirini təqdim etməkdir.

Azərbaycan dilindəki qoşmalı feili birləşmələrin modelləşdirməsi təyini söz birləşmələri ilə müqayisədə ümumi strukturu qurub formallaşdırmaq baxımından nisbətən çətindir.

Azərbaycan dilində əsas tərəfin nitq hissəsi ifadəsinə görə söz birləşmələrini üç növə ayırırlar. Bu növlərə ismi, feili və zərf birləşmələri daxil edilir [5, s. 42].

Dilçiliyimizdə söz birləşmələrinin tədqiqi, onların müəyyən formal əlamətlərə görə bölgüsü məsələsi ilk olaraq ismi birləşmələr əsasında aparılmışdır. İsmi söz birləşmələri "təyini söz birləşmələri" adı altında birləşdirilmiş, üç növ təyini söz birləşməsi fərqləndirilmişdir [7; 8].

Azərbaycan dilində sözün strukturunun ümumi modelini aşağıdakı kimi təsəvvür etmək olar: Kök + sözdüzəldici affiks +...+ sözdüzəldici affiks+ sözdəyişdirici affiks+ ...+ sözdəyişdirici affiks

Bu modeli kanonik şəkildə belə göstərmək mümkündür:

K+Sd + ... + Sd + Sdf + ... + Sdf

Burada K=kök, Sd=sözdüzəldici affiks, Sdf=sözdəyişdirici affiksdir.

Kök sözdüzəldici şəkilçi qəbul edərək yeni mənalı sözə çevrildikdə bu struktura malik söz əsas adlanır (Ə = K + Sd). Təhlil göstərir ki, əsas sözdüzəldici şəkilçi qəbul edərək yeni mənalı

sözə çevrilmək imkanına malikdir. Bu hal yeni düzəltmə sözdə kökdən sonra iki fərqli sözdüzəldici şəkilçinin olmasına uyğun gəlir: $Ə = Ə_1 + Sd = K + Sd_1 + Sd_2$; $Sd_1 \neq Sd_2$.

Sözdüzəldici şəkilçi üçün qəbul etdiyimiz işarəni indeksli dəyişənlə (Sd_n) əvəz etmək olar. Buradakı "n" indeksi dəyişir. Onun dəyişmə intervalı fərqli ola bilər. Təhlil və araşdırmalar göstərir ki, bu indeksin 3-dən böyük qiymət almasına rast gəlinmir. Sözdəyişdirici şəkilçi üçün "m" indeksini seçmək olar. Məsələn, dəmir, dəmirçi, dəmirçilik sözləri aşağıdakı modelə uyğun gələcəkdir: K = dəmir, $K+ Sd_1$ = dəmir + çi = dəmirçi, $K+ Sd_1 + Sd_2$ = dəmir + çi + lik = dəmirçilik.

Bu cür halları nəzərə alsaq, ümumi model aşağıdakı kimi olacaqdır:

$$\mathbf{K+Sd_1 + ... + Sd_n + Sdf_1 + ... + Sdf_m}$$

Modelin ümumi formulundakı "n" və "m" əlavə olunan affiksin sayından asılı olaraq dəyişəcəkdir. Beləliklə, Azərbaycan dilinin söz-forması üçün ümumi struktur model SF = $\mathbf{K+Sd_1 + ... + Sd_n + Sdf_1 + ... + Sdf_m}$ olacaqdır. Ümumiləşdirmə aparsaq, Azərbaycan dilində söz-formaların altı fərqli modelini ayıra bilərik:

I. $\mathbf{S = K}$
II. $\mathbf{S = K + Sd_n}$
III. $\mathbf{S = K+Sd_1 + ... + Sd_n}$
IV. $\mathbf{SF = K + Sf_n}$
V. $\mathbf{SF = K + Sd_1 +... + Sd_n}$
VI. $SF = K+Sd_1 + ... + Sd_n + Sdf_1 + ... + Sdf_m$

Azərbaycan dilində söz-formanın strukturunun **K + Sd$_1$ … + Sdf**□ modeli ilə ifadə olunması morfoloji analiz üçün son dərəcə əlverişli formal baza yaradır. Bu cür struktur təsviri müasir kompüter dilçiliyindəki sonlu avtomat morfologiyası (FSM) ilə tam uyğunluq təşkil edir [9, s.7]. FSM modelində sözün quruluşu ardıcıl keçidlər sistemi kimi təsvir olunur: kök halından başlanaraq sözdüzəldici və daha sonra sözdəyişdirici şəkilçilərin artırılması vəziyyətlərarası keçid kimi modelləşdirilir [2]. Aqlütinativ dillərdə şəkilçilərin ardıcıllığının sabitliyi bu modelin tətbiqini xüsusilə effektiv edir.

Təqdim edilən $SF = K + Sd_1 + … + Sdn + Sdf_1 + … + Sdfm$ modeli Azərbaycan dili üçün FSM əsaslı morfoloji analizatorun nəzəri təsviri funksiyasını yerinə yetirir və söz-formaların həm avtomatik analizi, həm də avtomatik sintezində tətbiqini tapa bilər.

Azərbaycan dilçiliyində söz birləşmələrini Y.Seyidov kompleks şəkildə tədqiq etmişdir. O, dilin söz birləşmələrini 3 qrupa – ismi, zərf və feili söz birləşmələrinə bölmüş, feili birləşmələrdə qoşmalı feili birləşmələri ayrıca qrup şəklində təqdim etmişdir [7, s. 322-331].

Y.Seyidov qoşmalı feili birləşmələri formaca "müstəqil mənalı söz + qoşma + feil" şəklində təqdim etmişdir. Müəllif qoşmaların feili birləşmələrdəki rolunu səciyyələndirərkən dörd fərqli xüsusiyyəti diqqət mərkəzinə çəkmişdir.

Birinci xüsusiyyət olaraq *can, cən, dək* qoşmalarının, eləcə də ilə qoşmasının ixtisar formasının sözlə bitişik yazılması qeyd olunmuşdur. Məsələn, evədək qaçmaq, evəcən getmək, qələmlə yazmaq [7, s. 322-323]. Təhlil göstərir ki, ilə qoşmasının bitişik deyil ayrı yazılması halı da mümkündür. Məsələn, *araba ilə getmək, metro ilə getmək* və s.

Y.Seyidov ikinci xüsusiyyəti şərh edərkən yazmışdır: " Qoşmalar öz-özlüyündə feillə əlaqə saxlaya bilməyən, feilin tələblərinə tabe olmayan yiyəlik halın feillə əlaqəsini təmin edir; məs.: *onun kimi oxumaq, bizim kimi yaşamaq*" [7, s. 323].

Çıxışlıq hallı feili birləşmələrdə qoşma hal birinci komponentdən sonra birləşməyə əlavə oluna bilir. Məsələn, *günortadan bəri yatmamaq, evdən bəri gəlmək, evdən sarı darıxmaq* və s.

Qeyd edilmiş üç xüsusiyyətin qoşmalı feili birləşmələrin modelləşdirilməsində nəzərə alınması vacibdir. Əgər qoşmalı feili birləşmələrin ümumi modelini **MMS + Q+ F** götürsək (MMS - müstəqil mənalı söz, Q – qoşma, F – feil), onda MMS + hal şəkilçisi + qoşma + feil, MMSQ + feil modellərinin mövcudluğunu da qeyd etməliyik. MMSQ + feil modelinə *evəcən getmək, avtobusla getmək, qələmlə yazmaq*, MMS + hal şəkilçisi + qoşma + feil modelinə aid *günortadan bəri yatmamaq, evdən sarı darıxmaq* nümunələrini göstərmək olar.

Q.Kazımov feili birləşmələrdən bəhs edərkən qoşmalı feili birləşmələri də təhlilə cəlb etmişdir. O, qoşmalı feili birləşmələrdə adlıq, yiyəlik, yönlük və çıxışlıq hallarda olan sözlərin işlənməsini qeyd edir [5, s.59]. Adlıq hal ümumi modelə (MMS + Q+ F) uyğun gəlir. Digər halların işlənməsi əlavə üç modelin mövcud olduğunu təsdiq edir:

1. Ad + yiyəlik hal şəkilçisi + qoşma + feil; A + YHŞ + Q + F modeli
2. Ad + yönlük hal şəkilçisi + qoşma + feil; A + YNHŞ + Q + F modeli
3. Ad + çıxışlıq hal şəkilçisi + qoşma + feil; A + ÇHŞ + Q + F modeli

Verilmiş modellərdə A – ad, YNŞ – yiyəlik hal şəkilçisi, YNHŞ – yönlük hal şəkilçisi, ÇHŞ – çıxışlıq hal şəkilçisi kimi başa düşülməlidir.

Hər üç modelə aid nümunələr vermək olar: *onun kimi oxumaq, bizim kimi yaşamaq, sizin kimi işləmək, mənim kimi düşünmək; evə tərəf getmək, məktəbə sarı getmək, küçəyə tərəf dönmək, evədək getmək, qapıyacan ötürmək və s.; çörəkdən sarı getmək, yeməkdən ötrü almaq, günortadan bəri görünməmək, səhərdən bəri yeməmək* və s.

Qoşmalar köməkçi nitq hissələridir və onlar şəkilçiləşməyə doğru inkişaf edirlər. M.Hüseynzadə dörd qrupa ayırmışdır. Onun bölgüsünə görə, birinci qrup həm əsas nitq hissəsi kimi, həm də qoşma kimi işlənənlər, ikinci qrupa həm qoşma, həm də bağlayıcı kimi işlənənlər, üçüncü qrupa yalnız qoşma olanlar, nəhayət, dördüncü qrupa həm qoşma, həm də şəkilçi kimi işlənənlər daxildir. Birinci qrupda *əvvəl, sonra, qabaq, bəri, tərəf, başqa, özgə, ayrı, əlavə, qeyri, qarşı* və s. qoşmaları, ikinci qrupda *ilə, üçün, görə, ötrü,* üçüncüdə *kimi, tək, sarı, içrə, aid, dair, məxsus*, dördüncü qrupda isə *ilə (-la/-lə), -can/-cən* qoşmaları yer almışdır [3, s. 293-294].

Qoşmaların bu qruplar əsasında ayrılması modelləşdirməyə də eyni aspektdən yanaşmağa zəmin yaradır. Modelləşdirməni bu şəkildə aparmaq üçün əvvəlcə qoşmalı feili birləşmələrin əsas ümumi modellərini qurmaq lazım gəlir.

QFB = MMS + Q + F ümumi modeldir. Qeyd olunduğu kimi MMS-dən asılı olaraq ümumi model dörd alt modelə bölünür:

I alt model: **QFB = Ad + Q + F** (Ad + qoşma + feil);

II alt model: QFB = İ + YHŞ + Q + F (İsim + yiyəlik hal şəkilçisi + qoşma + feil);

III alt model: **QFB = İ + YNHŞ + Q + F** (İsim + yönlük hal şəkilçisi + qoşma + feil);

IV alt model: QFB = İ + ÇHŞ + Q + F (İsim + çıxışlıq hal şəkilçisi + qoşma + feil).

Qeyd edildiyi kimi şəkilçiləşmiş qoşmanın bilavasitə müstəqil mənalı sözə bitişməsi halı da vardır ki, bu beşinci alt model olacaqdır.

V alt model: QFB = MSMQ + F (müstəqil mənalı söz şəkilçiləşmiş qoşma + feil).

Beləliklə, modelləşdirmənin birinci mərhələsində beş alt model müəyyənləşir. Aşağıdakı cədvəldə alt model və ona aid nümunələr verilmişdir (Cədvəl 1).

Cədvəl 1.

	Alt model	**Nümunə**
1.	QFB = Ad + Q + F	ev üçün almaq; maşın kimi sürmək
2.	QFB = İ + YHŞ + Q + F	onun üçün almaq, sənin kimi demək
3.	QFB = İ + YNHŞ + Q + F	dağa sarı baxmaq, qapıya doğru getmək
4.	QFB = İ + ÇHŞ = Q + F	Sizdən ötrü çalışmaq, maşından sonra gəlmək
5.	QFB = MSMQ + F	Evədək getmək, qapıyacan ötürmək

Qoşmalı feili birləşmələrin qeyd edilən modellərində feilin bilavasitə qoşmadan sonra işlənməsi hallarını əhatə edir. Dildə qoşmalı feili birləşmələrdə qoşma ilə feil arasına sözlərin girməsi mümkündür. Məsələn, ev üçün çörək almaq, sizin üçün kitab almaq, sizdən ötrü gecə-gündüz çalışmaq, evəcən piyada getmək və s. Göründüyü kimi qoşma ilə feil arasına bir və daha artıq sözün daxil olması mümkündür. Bu cəhət qoşmalı feili birləşmələrin komponent tərkibinə görə modellərin çoxluğunu təsdiq edir.

Qoşmalı feili birləşmələrin birinci tərəfinin təyini söz birləşməsi ola bilir. Məsələn, daş hasar üçün almaq, daş hasar üçün sement almaq, maşın təkəri üçün getmək, maşın təkəri üçün mağazaya getmək, maşının təkəri üçün usta çağırmaq və s. Bu tipli modellər iki model əsasında

formalaşan mürəkkəb modellərə uyğun gəlir və TSB + QFB şəklində ümumi mürəkkəb model kimi qeydə alına bilər.

NLP sistemində sadə və mürəkkəb modellərin avtomatik təyini məsələləri bir-birindən fərqlənir və təbii ki, sadə modellərin aşkara çıxarılması daha sadə prosesdir.

Aparılmış tədqiqat göstərir ki, təbii dilin avtomatik emalı sistemlərində mürəkkəb sintaktik vahidlərin avtomatik müəyyənləşdirilməsi xüsusi əhəmiyyət kəsb edir. Azərbaycan dilində qoşmalı feili birləşmələr sintaktik strukturların mühüm növlərindən biri olduğuna görə onların formal modellərinin müəyyənləşdirilməsi avtomatik analiz sistemlərinin qurulması üçün vacibdir.

NLP sistemlərində mətnin emalı adətən bir neçə mərhələdən ibarət olur. Bu mərhələlər aşağıdakı ardıcıllıqla həyata keçirilir: mətn → tokenizasiya → morfoloji təhlil → morfoloji omonimlik → sintaktik qruplaşdırma → sintaktik təhlil.

Qoşmalı feili birləşmələrin avtomatik müəyyənləşdirilməsi əsasən sintaktik qruplaşdırma mərhələsində həyata keçirilir. Bu mərhələdə morfoloji analiz nəticəsində əldə edilmiş nitq hissəsi və qrammatik kateqoriyalar haqqında məlumatdan istifadə olunur [6, s. 123].

Azərbaycan dili aqlütinativ dil olduğuna görə sözlərin morfoloji quruluşu əvvəlki bölmədə göstərildiyi kimi formal model əsasında təsvir edilə bilər:

$SF = K + Sd_1 + \dots + Sdn + Sdf_1 + \dots + Sdfm$

Bu model əsasında qurulan morfoloji analizator sözün kökünü və şəkilçilərini avtomatik müəyyənləşdirir. Morfoloji analiz nəticəsində hər bir söz üçün aşağıdakı kimi etiketlənmiş struktur əldə edilir:

ev/isim üçün/qoşma almaq/feil
maşın/isim kimi/qoşma sürmək/feil
onun/əvəzlik kimi/qoşma danışmaq/feil

Bu məlumat əsasında sintaktik strukturların müəyyənləşdirilməsi mümkün olur. Qaydalara əsaslanan sistemlərdə qoşmalı feili birləşmələrin müəyyənləşdirilməsi üçün xüsusi qaydalar tətbiq olunur. Bu qaydalar sintaktik model əsasında qurulur. Qoşmalı feili birləşmələrin əsas modeli aşağıdakı kimi formal şəkildə göstərilə bilər:

QFB → MMS + Q + F; Bu model NLP sistemində daha ümumi şəkildə aşağıdakı sintaktik qayda ilə ifadə edilə bilər:

Qoşmalı feili birləşmələrin modelləri maşın tərcüməsi sistemləri üçün də mühüm əhəmiyyət daşıyır. Azərbaycan dilində qoşma ilə ifadə olunan münasibətlər bir çox dillərdə ön qoşma və ya sintaktik konstruksiya vasitəsilə ifadə olunur. Məsələn:

ev üçün almaq → buy for the house
sənin kimi danışmaq → speak like you
qapıya doğru getmək → go towards the door

Bu səbəbdən qoşmalı feili birləşmələrin struktur modelləri avtomatik tərcümə sistemlərində sintaktik transformasiya qaydalarının qurulmasına imkan yaradır.

Beləliklə, qoşmalı feili birləşmələrin formal modelləşdirilməsi Azərbaycan dili üçün NLP sistemlərinin aşağıdakı komponentlərində istifadə oluna bilər:

• morfoloji analiz sistemlərində
• sintaktik qruplaşdırma alqoritmlərində
• avtomatik sintaktik təhlil sistemlərində
• maşın tərcüməsi modellərində
• milli korpusların sintaktik annotasiyasında.

Bu modellər Azərbaycan dilində mürəkkəb sintaktik konstruksiyaların avtomatik identifikasiyası üçün formal baza yaradır və kompüter dilçiliyi baxımından praktiki tətbiq imkanlarını genişləndirir.

Müasir kompüter dilçiliyində aqlütinativ dillərin morfoloji quruluşunun modelləşdirilməsində ən geniş istifadə olunan formal metodlardan biri sonlu avtomatlar nəzəriyyəsinə əsaslanır və sözlərin morfoloji quruluşunu vəziyyətlər və keçidlər sistemi kimi təsvir edir [1, p.85].

Sonlu avtomat formal olaraq aşağıdakı şəkildə müəyyən olunur:

$$\mathbf{A = (Q, \Sigma, \delta, q_0, F)}$$

Burada Q – vəziyyətlər çoxluğu; Σ – giriş əlifbası (simvollar və ya morfemlər çoxluğu); δ– keçid funksiyası; q_0 – başlanğıc vəziyyət; F – son vəziyyətlər çoxluğudur [4, 58-59].

Morfoloji analizator qurularkən sözün quruluşu morfemlərin ardıcıl qoşulması kimi təsvir edilir. Azərbaycan dili aqqlütinativ dil olduğu üçün sözün quruluşu aşağıdakı ümumi model ilə ifadə edilə bilər: **SF = K + Sd_1 + … + Sdn + Sdf_1 + … + Sdfm**

Sonlu avtomat modelində bu struktur ardıcıl vəziyyətlər sistemi kimi təsvir olunur. Başlanğıc vəziyyətdə kök seçilir və daha sonra şəkilçilər ardıcıl keçidlər vasitəsilə əlavə olunur. Bu proses aşağıdakı sxemlə göstərilə bilər:

q_0 → K → Sd_1 → Sd_2 → … → Sdn → Sdf_1 → … → Sdfm → qf

Burada q_0 başlanğıc vəziyyət, qf isə son vəziyyəti göstərir.

Bu modeldə hər bir morfoloji element avtomatın keçidlərini təşkil edir. Məsələn, aşağıdakı sözün morfoloji quruluşuna baxaq: **dəmirçilikdən.** Bu sözün morfoloji strukturu belə təsvir edilə bilər: dəmir + çi + lik + dən. Bu struktur FSM modelində aşağıdakı keçidlər ardıcıllığı ilə ifadə olunur:

q_0 → dəmir → çi → lik → dən → qf; Burada dəmir – kök, çi – sözdüzəldici şəkilçi, lik – ikinci sözdüzəldici şəkilçi, dən – çıxışlıq hal şəkilçisidir.

Beləliklə, FSM modelində sözün morfoloji analizi kökdən başlayaraq şəkilçilərin ardıcıl şəkildə tanınması prosesi kimi həyata keçirilir.

Sonlu avtomat modelinin mühüm xüsusiyyətlərindən biri morfoloji təhlil və morfoloji sintez proseslərinin eyni formal sistem əsasında qurulmasıdır. Analiz zamanı söz forması morfemlərə parçalanır, sintez zamanı isə morfemlər birləşdirilərək yeni söz forması yaradılır.

FSM modelinin Azərbaycan dili üçün tətbiqi xüsusilə əlverişlidir. Bunun əsas səbəbləri aşağıdakılardır:

• Azərbaycan dilinin aqlütinativ quruluşu;
• şəkilçilərin ardıcıl və nisbətən sabit düzülüşü;
• morfoloji kateqoriyaların aydın formal göstəricilərə malik olması.

Bu xüsusiyyətlər sonlu avtomat əsaslı morfoloji analizatorun qurulmasına imkan verir. Belə analizatorlarda hər bir kök xüsusi lüğət bazasında saxlanılır, şəkilçilər isə ayrıca morfoloji qaydalar sistemi vasitəsilə təsvir olunur.

Sonlu avtomat modeli Azərbaycan dili üçün aşağıdakı NLP sistemlərində tətbiq oluna bilər:

• morfoloji analizatorların yaradılması;
• morfoloji omonimlik sistemləri;
• sintaktik analiz və qruplaşdırma alqoritmləri;
• maşın tərcüməsi sistemləri;
• milli korpusların avtomatik morfoloji annotasiyası.

Beləliklə, FSM əsaslı morfoloji modelləşdirmə Azərbaycan dilinin morfoloji quruluşunun formal təsvirini təmin edir və qoşmalı feili birləşmələr kimi mürəkkəb sintaktik strukturların avtomatik analizində ilkin mərhələ kimi mühüm rol oynayır.

Qoşmalı feili birləşmələrin mətndən avtomatik seçilməsi tokenizasiyanın konkret qoşma əsasında aparılmasını aktuallaşdırır. Məsələn, tokenləmə **sonra** qoşması üzrə gedirsə, onda mətndə həmin söz axtarılır. Bu söz tapıldıqdan sonra onun qoşma və ya müstəqil mənalı söz olması yoxlanılır. Burada qoşmanın müəyyənləşdirilməsi üçün sözdən sonra durğu işarəsinin işlənməsinin, sözün böyük hərflə yazılmasının yoxlanması əgər deyilən markerlər olarsa, onda sonra müstəqil mənalı söz kimi götürülür və növbəti **sonra** sözü axtarılır. Qeyd edilən markerlər olmadıqda, qoşmadan əvvəlki söz götürülür və morfoloji analizatordan keçirilir. Növbəti addımda **sonra** sözündən sonrakı sözün yoxlanması həyata keçirilir. Əgər bu söz feil forması olarsa, onda qeyd edilən üç sözdən ibarət söz birləşməsi qoşmalı feili birləşmə kimi qəbul edilir. Təbii ki, verilmiş şərh üçkomponentli qoşmalı feili birləşmələrə aiddir.

Nəticə: Aparılmış tədqiqat Azərbaycan dilində qoşmalı feili birləşmələrin struktur xüsusiyyətlərinin sistemli şəkildə modelləşdirilməsinin mümkün olduğunu göstərir. Araşdırma zamanı Azərbaycan dilinin aqlütinativ quruluşu nəzərə alınaraq söz-formaların ümumi morfoloji modeli (SF = K + Sd_1 + … + Sdn + Sdf_1 + … + Sdfm) müəyyənləşdirilmiş və bu modelin sonlu

avtomat morfologiyası prinsipləri ilə uyğunluğu əsaslandırılmışdır. Bu yanaşma Azərbaycan dili üçün morfoloji analizatorların qurulmasına formal nəzəri baza yaradır.

Tədqiqatın mühüm nəticələrindən biri qoşmalı feili birləşmələrin struktur modellərinin müəyyənləşdirilməsi olmuşdur. Təhlil nəticəsində qoşmalı feili birləşmələrin ümumi modeli (QFB = MMS + Q + F) və onun müxtəlif qrammatik formalar əsasında yaranan alt modelləri müəyyən edilmişdir. Bu alt modellər müstəqil mənalı sözün müxtəlif hallarda işlənməsi və qoşmaların şəkilçiləşmiş formaları ilə əlaqədar olaraq formalaşır. Bundan əlavə, qoşma ilə feil arasında əlavə sözlərin daxil olması və birləşmənin birinci komponentinin təyini söz birləşməsi şəklində çıxış etməsi kimi hallar da nəzərə alınmışdır ki, bu da qoşmalı feili birləşmələrin struktur müxtəlifliyini göstərir.

Məqalədə göstərilmiş modellər təbii dilin avtomatik emalı sistemlərində sintaktik qruplaşdırma mərhələsində istifadə oluna biləcək formal qaydalar sisteminin qurulmasına imkan verir. Morfoloji analiz nəticəsində əldə olunan qrammatik etiketlər əsasında qoşmalı feili birləşmələrin avtomatik identifikasiyası mümkün olur. Bu isə sintaktik analiz sistemlərinin dəqiqliyinin artırılması üçün mühüm əhəmiyyət kəsb edir.

Araşdırmada həmçinin göstərilmişdir ki, qoşmalı feili birləşmələrin modelləri maşın tərcüməsi sistemlərində də vacib rol oynayır. Azərbaycan dilində qoşmalar vasitəsilə ifadə olunan münasibətlər digər dillərdə müxtəlif sintaktik vasitələrlə ifadə olunduğundan bu modellər sintaktik transformasiya qaydalarının qurulmasına imkan verir.

Nəticə: Qoşmalı feili birləşmələrin formal modelləşdirilməsi Azərbaycan dili üçün NLP sistemlərinin morfoloji analiz, sintaktik qruplaşdırma, avtomatik sintaktik təhlil, maşın tərcüməsi və milli korpusların sintaktik annotasiyası kimi komponentlərinin inkişafı üçün mühüm nəzəri və praktiki baza yaradır. Gələcək tədqiqatlarda bu modellərin geniş mətn korpusları üzərində tətbiqi və statistik üsullarla inteqrasiyası Azərbaycan dili üçün daha inkişaf etmiş NLP sistemlərinin yaradılmasına imkan verə bilər.

Ədəbiyyat siyahısı

1. Beesley K. R., Karttunen, L. Finite-State Morphology. Stanford: CSLI Publications, 2003.
2. Faggella D. What is Machine Learning? // Emerj. URL:https://emerj.com/aiglossary-terms/what-is-machine-learning/
3. Hüseynzadə, M. Müasir Azərbaycan dili. Morfologiya. Bakı, Maarif, 1973, 358 s.
4. Jurafsky D., Martin, J. H. Speech and Language Processing**.** 2nd ed. Pearson, 2009.
5. Kazımov Q. Müasir Azərbaycan dili. Bakı: "Ünsiyyət", 2000, 496 s.
6. Mahmudov M. Türk dillərinin milli korpusları / M.Mahmudov. – Bakı: "Elm və Təhsil", 2018. – 392 s.
7. Seyidov, Y. Azərbaycan ədəbi dilində söz birləşməsi / Y.Seyidov. – Bakı: Maarif, 1966. – 340 s.
8. Xəlilov B. Müasir Azərbaycan dilinin sintaksisi. / B.Xəlilov. – Bakı: Elm, 2008. – 346 s.
9. Барашко Е.Н. Храпцов, Д.А. Современные системы машинного перевода / Е.Н.Барашко, Д.А.Храпцов. // Актуальные проблемы развития естественных и технических наук. – Саратов: Академия Бизнеса, 2018. – с. 5 – 11.

Натаван Рахманова

Моделирование послеложных глагольных словосочетаний азербайджанского языка для системы NLP

Аннотация

В статье исследуются структурные особенности послеложных глагольных словосочетаний в азербайджанском языке и вопросы их моделирования для систем автоматической обработки естественного языка (NLP). Основной целью исследования является определение формальной структуры послеложных глагольных словосочетаний,

являющихся одной из широко распространённых синтаксических конструкций в азербайджанском языке, а также предложение моделей для их автоматической идентификации в системах компьютерной лингвистики. В исследовании, с учётом агглютинативной структуры азербайджанского языка, определена общая морфологическая модель словоформ и показано её соответствие принципам морфологии конечных автоматов (Finite State Morphology). Исследование показывает, что предложенные модели могут быть использованы в различных NLP-приложениях для азербайджанского языка. К таким приложениям относятся системы морфологического анализа, алгоритмы синтаксической группировки, системы автоматического синтаксического анализа, модели машинного перевода, а также автоматическая морфологическая и синтаксическая аннотация национальных корпусов.

Таким образом, формальное моделирование послеложных глагольных словосочетаний позволяет осуществлять автоматический анализ сложных синтаксических конструкций в азербайджанском языке и вносит важный вклад в развитие технологий NLP для азербайджанского языка с точки зрения компьютерной лингвистики.

Ключевые слова: *словосочетания, постпозиционные глагольные конструкции, морфологическое моделирование, конечные автоматы, сегментация именных групп на основе правил, НЛП.*

Natavan Rahmanova

Modeling of postpositional verbal constructions in the Azerbaijani language for NLP systems

Abstract

The article investigates the structural characteristics of postpositional verbal constructions in the Azerbaijani language and their modeling for natural language processing (NLP) systems. Special attention is given to the role of such constructions within the syntactic structure of Azerbaijani and to the possibilities of their formal representation for computational processing of texts. Since Azerbaijani is an agglutinative language, its word forms are characterized by a clear morphological structure consisting of a root and a sequence of derivational and inflectional affixes. Based on this property, the study proposes a general morphological model of Azerbaijani word forms and demonstrates its compatibility with the principles of Finite State Morphology.

The research identifies the general structural pattern of postpositional verbal constructions and describes several sub-models depending on the grammatical case forms of the nominal component and the functional characteristics of postpositions. These constructions are formally represented by syntactic patterns that can be applied in rule-based NLP systems. The study also examines cases where additional lexical elements occur between the postposition and the verb, as well as constructions in which the nominal component itself is expressed by an attributive phrase, resulting in more complex syntactic structures.

Key words: *word combination, Postpositional Verbal Constructions, morphological modeling, Finite State Morphology, Rule-based NP chunking, NLP.*

Çapa tövsiyə edən: ***AMEA Nəsimi adına Dilçilik İnstitutu***

Rəyçilər: ***professor M.Mahmudov***
professor İ.Tahirov