

<https://doi.org/10.62837/2026.2.204>

**NATƏVAN CƏMALƏDDİN QIZI RƏHMANOVA**  
*AMEA Nəsimi adına Dilçilik İnstitutu*  
*natis-sdunina@mail.ru*

## **AZƏRBAYCAN DİLİNİN TƏYİNİ SÖZ BİRLƏŞMƏLƏRİNİN NLP SİSTEMİ ÜÇÜN MODELLEŞDİRİLMƏSİ**

**Açar sözlər:** təyini söz birləşməsi, morfoloji modelləşdirmə, sonlu avtomat morfologiyası, qaydalara əsaslanan NP, NLP, Azərbaycan dili.

**Key words:** attributive word combinations, morphological modeling, Finite State Morphology, Rule-based NP chunking, NLP, Azerbaijani language.

Azərbaycan dilində müxtəlif növ söz birləşmələri vardır ki, onların sistemli şəkildə öyrənilməsi əsasında müəyyən modellərin yaradılıb NLP-də, həmçinin məşin tərcüməsi sistemində, milli korpusların qurulmasında, sözlərin avtomatik sıralanmasının bir çox məsələlərinin həllində tətbiqi xüsusi maraq doğurur.

Azərbaycan dilindəki təyini söz birləşmələri ilə yanaşı, qoşmalı feili birləşmələrinin də modelləşdirməsi çətin görünür. Çünki bu cür birləşmələrdə ümumi strukturu qurmaq, belə strukturu formallaşdırmaq həm çox da mürəkkəb tədqiqat tələb etmir, həm də dilçiliyimizdə onlar haqqında kifayət qədər araşdırmalar aparılmışdır.

Q.Kazımov əsas tərəfin nitq hissələri ilə ifadəsinə görə söz birləşmələrini üç növə - ismi, feili və zərf birləşmələrinə ayırmışdır. Onun təsnifində əsas tərəfi adlar (isim, sifət, say və əvəzlilər) olanlar ismi söz birləşmələri qrupuna daxil edilmişdir [6, 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 [10; 4]. Fərqləndirmə meyarı birləşmənin tərəflərinin formal əlamət qəbul etməsi olmuşdur. Bu yanaşma strukturalist baxışa əsaslanır. Birləşmənin əsas və köməkçi tərəfinin formal əlamət və ya şəkilçi qəbul edib-etməməsi struktur fərqliliyinin göstəricisi rolunu oynamışdır. Əslində məsələyə bu şəkildə münasibət mətn tərkibində əlaqələnen adların analizini əhətə etmişdir. Analizin nəticəsi isə təyini söz birləşmələrinin konkret olaraq üç növünü ayırmağa imkan vermişdir. Belə təsnifdə analiz struktur, struktur komponentlərinin (tərəflərin) təkrar struktur bölgüsü zəminində getmiş və formallaşdırma üçün baza yaratmışdır [7, s.148-149; 8]. Formallaşdırma, struktur vahidlərinə parçalama

ardıcılığının mümkünlüyü kompüter dilçiliyinin ilk dövrləri üçün kifayət qədər geniş tətbiq edilmiş metod kimi həmişə üzdə olmuşdur.

Aqqlütinativ Azərbaycan dilində sözün maksimal mürəkkəb strukturunun ümumi modelini aşağıdakı kimi təsvir etmək, yaxud qurmaq mümkündür:

**Kök + sözdüzəldici affiks +...+ sözdüzəldici affiks+ sözdəyişdirici affiks+ ...+ sözdəyişdirici affiks**

Kök = K, sözdüzəldici affiks = Sd, sözdəyişdirici affiks = Sdf işarələməsini qəbul etsək yuxarıdakı modeli bir qədər kanonik şəkə sala bilərik:

$$\mathbf{K+Sd + \dots + Sd + Sdf + \dots + Sdf}$$

Sözün kökü sözdüzəldici şəkilçi qəbul edərək yeni mənalı sözə çevrilir. Məsələn, yazı + çı= yazıçı, dəmir + çi= dəmirçi; çəmən + lik = çəmənlik, dağ + lıq = dağlıq; ev+li= evli, can+lı = canlı; yapış+qan = yapışqan, sat + qın = satqın və s.

Kök sözdüzəldici şəkilçi qəbul edərək yeni mənalı sözə çevrildikdə bu struktura malik söz əsas adlanır. Əsası Ə ilə işarə etsək, onun ilkin modeli belə olacaqdır: Ə = K + Sd. Dil materialının, daha dəqiq desək, Azərbaycan dilindəki əsasların analizi prosesində aydın olur ki, əsas sözdüzəldici şəkilçi qəbul edərək yeni mənalı sözə çevrilə bilər. Bu hal yeni düzəltmə sözə kökdən sonra iki fərqli sözdüzəldici şəkilçinin olmasına uyğun gəlir: Ə = Ə<sub>1</sub> + Sd = K + Sd + Sd.

Sd fərqli sözdüzəldici şəkilçini göstərə bilər. Yuxarıda əsasa ikinci sözdüzəldici şəkilçinin artırılmasının mümkünlüyünü göstərəkən bu affiksin özündən əvvəlki ilə eyni olub-olmaması barədə heç bir qeyd verilməmişdi. Nəzərə almaq lazımdır ki, iki eyni sözdüzəldici şəkilçi ardıcıl olaraq birləşmir. Ona görə də, bu fərqi modeldə ifadəsini tapması zəruridir. Deyilənə əsaslanaraq sözdüzəldici şəkilçi üçün qəbul etdiyimiz işarəni indeksli dəyişən forması ilə əvəz etmək məqsəduyğundur: Sd<sub>n</sub>. Burada “n” indeksi dəyişir. Yəni sözdüzəldici şəkilçi fərqli olduqda indeks müəyyən rəqəmlə əvəz olunacaqdır. Sözdəyişdirici şəkilçi üçün “m” indeksini seçmək olar. Məsələn, dəmir, dəmirçi, dəmirçilik sözləri aşağıdakı modellərə uyğun gələcəkdir: K= dəmir, K+ Sd<sub>1</sub>= dəmir + çi = dəmirçi, K+ Sd<sub>1</sub> + Sd<sub>2</sub> = dəmir + çi + lik = dəmirçilik.

Ümumi model bu halda aşağıdakı kimi olacaqdır:

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

Modelin ümumi formulundakı “n” və “m” əlavə olunan affiksin sayından asılı olaraq dəyişəcəkdir. Beləliklə, Azərbaycan dilinin söz-forması üçün ümumi struktur model SF = **K+Sd<sub>1</sub> + ... + Sd<sub>n</sub> + Sdf<sub>1</sub> + ... + Sdf<sub>m</sub>** olacaqdır. Xüsusi halda söz-forma kök və ya əsas şəklində özünü göstərə bilər. Bu hal üçün modelin formasını belə göstərmək mümkündür: S = **K+Sd<sub>1</sub> + ... + Sd<sub>n</sub>**. Ümumiləşdirmə aparsaq, Azərbaycan dilində sözün mümkün model formalarını belə bir modellər ardıcılığında göstərmək mümkündür:

I. **S = K**

**II.**  $S = K + Sd_n$

**III.**  $S = K + Sd_1 + \dots + Sd_n$

**IV.**  $SF = K + Sf_n$

**V.**  $SF = K + Sd_1 + \dots + Sd_n$

**VI.**  $SF = K + Sd_1 + \dots + Sd_n + Sdf_1 + \dots + Sdf_m$

Burada 6 əsas ümumi modelin əksini tapmışdır. “n”, “m” dəyişənlərinin sayından asılı olaraq modellərin xüsusi halları qeydə alınmışdır. Məsələn, çəmənlikdə ( $SF = K + Sd_1 + Sf_1$  : çəmən + lik + də), çəmənliklərin ( $SF = K + Sf_1 + Sf_2$  : çəmən + lər + in) və s.

Azərbaycan dilində söz-formanın strukturunun  $K + Sd_1 \dots + Sdf_m$  modeli ilə ifadə olunması morfoloji analiz üçün son dərəcə əlverişli formal baza yaradır. Bu cür struktur təsviri müasir kompüter dilçiliyində sonlu Avtomat morfoloqiyası (FSM) ilə tam uyğunluq təşkil edir [11, 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 [3]. Aqqlütinativ dillərdə şəkilçilərin ardıcılığının sabitliyi bu modelin tətbiqini xüsusilə effektiv edir.

Təqdim edilən  $SF = K + Sd_1 + \dots + Sd_n + Sdf_1 + \dots + Sdf_m$  modeli Azərbaycan dili üçün FSM əsaslı morfoloji analizatorun nəzəri təsviri funksiyasını yerinə yetirir və söz-formaların avtomatik morfoloji parçalanması üçün formal şablon yaradır.

Söz birləşməsinin (SB) ən sadə forması iki komponent kimi götürülə bilər. Bu, model kimi  $SB = SF_1 + SF_2$ . Modelin tərkibindəki “+” işarəsi söz-formaların bir-biri ilə birləşməsinə, yəni bitişik yazılmasını göstərmir. Söz mətnində kök, əsas və söz-forma şəklində ola bilər. Qrammatik formada sözdəyişdirici affiksin sıfır şəkilçi olması mümkündür. Modelləşdirmə zamanı qeyd edilən hal nəzərə alınmış və ikikomponentli söz birləşməsinin daha ümumi modeli qurulmuşdur. Bu modeli nümunələr əsasında şərh edək.

Qızıl qaşığı →  $SB = SF_1 + SF_2$  :  $SB = qızıl + qaşığı$ . Komponentlər struktur modelə ifadə olunduqda  $SB (qızıl qaşığı) = K_1 (qızıl) + K_2 (qaşığı)$ .

Ev heyvanı →  $SB = SF_1 + SF_2$  :  $SB (ev heyvanı) = K_1 (ev) + K_2 (heyvan) + Sf_1 (-i)$ .

Evin damı →  $SB = SF_1 + SF_2$  :  $SB (evin damı) = K_1 (ev) + Sf_1 (-in) + K_2 (dam) + Sf_2 (-i)$ .

Söz birləşmələrinin  $SB = SF_1 + SF_2$  modeli isə morfoloji analizdən keçmiş söz-formaların daha yüksək səviyyədə – sintaktik vahid kimi – birləşdirilməsinə imkan verir. Bu yanaşma kompüter dilçiliyində qaydalara əsaslanan Nisbi frazalara parçalama adlanır. İsmi frazalara parçalama prosesində sistem ardıcıl gələn söz-formaların morfoloji xüsusiyyətlərini yoxlayaraq onları isim qrupu (NP) kimi işarələyir.

Azərbaycan dilində təyini söz birləşmələrinin morfoloji əlamətlərlə dəqiq fərqlənməsi bu prosesi qayda əsaslı şəkildə həyata keçirməyə imkan verir. Beləliklə, FSM ilə aparılan morfoloji analiz ismi frazalara parçalama mərhələsi üçün giriş məlumatı rolunu oynayır [2, s. 934; 9, s. 154-155].

Azərbaycan dilində ismi birləşmələr adlanan qrupda təyini söz birləşmələri ayrıca lay əmələ gətirir. Təyini söz birləşmələrini tədqiqatçılar üç növə ayırırlar. Bu növlər, adətən, birinci, ikinci və üçüncü növ təyini söz birləşmələri terminlərlə ifadə edilir. Y.Seyidovun söz birləşmələri ilə bağlı tədqiqatında ismi birləşmələr geniş planda təqdim olunmuşdur. Müəllif onları iki qrupa bölmüş, birinci qrup təyini söz birləşmələri, ikinci qrup isə təyini söz birləşmələrinə daxil olmayan ismi birləşmələr adı altında tədqiqata cəlb etmişdir [10, s. 67].

Q. Kazımov bu məsələyə münasibət bildirərək yazır: “təyini söz birləşmələri” (“birinci növ təyini söz birləşmələri”, “ikinci növ təyini söz birləşmələri”, “üçüncü növ təyini söz birləşmələri”) və “təyini söz birləşmələrinə daxil olmayan ismi birləşmələr” terminləri həm həcmində, həm də anlayışı dəqiq ifadə edə bilmə imkanlarına görə o qədər də uğurlu deyildir” [6, s. 48]. Müəllifin bu fikri ilə razılaşmaq lazım gəlir və “təyini söz birləşmələrinə daxil olmayan ismi birləşmələr” qrupunu daha konkret terminin istifadə edilməsinin gərəkli olduğunu söyləmək lazım gəlir. Bununla belə, mətndə işlənən mümkün birləşmələrin bütün növlərinin müəyyənləşdirilməsi və onları imkan daxilində ayrıca qruplar şəklində birləşdirmək müasir kompüter diçiliyində, ümumiyyətlə NLP sistemlərində mətn elementlərinin avtomatik seqmentasiyası, onların modellər şəklində təqdimatı xüsusi əhəmiyyət daşıyır. Bu baxımdan Azərbaycan dilindəki təyini söz birləşmələrini ayrıca qrup, onun tərkibinə daxil olan birinci, ikinci və üçüncü növ təyini söz birləşmələrini isə üç alt qrup kimi təqdim etmək özünü doğruldur. Belə qənaətə gəlməyə təyini söz birləşmələrinin mətndə işlək olması da əsas verir.

Azərbaycan dilinin qrammatika kitablarının sintaksis hissələrində çox zaman təyini söz birləşmələri ayrıca paraqrafda verilir. A.Aslanov ismi söz birləşmələrinə aid işində birinci, ikinci və üçüncü növ ismi birləşmələrdən bəhs etmişdir. O, “təyini” sözündən istifadə etməmişdir. Müəllif birinci növ ismi söz birləşmələrinin morfoloji əlamətinin olmadığını, tərəflərin heç bir sözdəyişdirici şəkilçi qəbul etmədiyini qeyd etmiş, bundan sonra onları birinci tərəfin hansı nitq hissəsinə aid olmasına görə alt qruplara bölmüşdür. İkinci növün Azərbaycan dilində məhsuldarlığını qeyd edən A.Aslanov bu növdə birinci tərəfin sözdəyişdirici affiksiz, ikinci tərəfin mənsubiyyət şəkilçili olmasını yazmışdır. Nəhayət, müəllif üçüncü növdə hər iki tərəfin morfoloji əlamət malik olması qeyd etmişdir [1, s. 93-124].

Təyini söz birləşmələrinin hər üç növü söz birləşmələri üçün müəyyənləşdirilmiş ümumi  $SB = SF_1 + SF_2$  modelinə daxil olur. Hər növ ümumi modelin konkret alt modelləri kimi göstərilə bilər. Bu prosesdə hər növün konkret alt modelini müəyyənləşdirmək lazım gəlir. Qeyd edilən məsələni ardıcıl olaraq birinci,

ikinci və üçüncü növ təyini söz birləşmələri üçün həll etməyə çalışaq.

**Söz birləşməsinin tərəflərin strukturuna əsasən modelləşdirilməsi.**

**I növ təyini söz birləşmələri:**  $SB = SF_1 + SF_2$ . Bu növdə sözdəyişdirici affiks iştirak etmir. Ona görə də, həm  $SF_1$ , həm də  $SF_2$  ya kök, ya da əsas strukturunda olacaqdır. Mümkün alt modellərin sayı 9 olur (Bax: Cədvəl 1)

**Cədvəl 1**

Alt model	Nümunə
$SB = K + K$	Qızıl qaşığı
$SB = K + K + Sd_n$	Yaşıl çəmənlik
$SB = K + Sd_1 + \dots + Sd_n$	Yaxşı yenilikçi
$SB = K + Sd_n + K$	Dağlıq ərazi
$SB = K + Sd_n + K + Sd_n$	Sınıq güldan
$SB = K + Sd_n + K + Sd_1 + \dots + Sd_n$	Qapaqlı qutucuq
$SB = K + Sd_1 + \dots + Sd_n + K$	Yazılacaq kitab
$SB = K + Sd_1 + \dots + Sd_n + K + Sd_n$	Yayılacaq kitabça
$SB = K + Sd_1 + \dots + Sd_n + K + Sd_1 + \dots + Sd_n$	Yarımqıq papaqçılıq

Birinci növ təyini söz birləşmələrinin verilmiş 9 alt modeli bir-birindən işləklilik dərəcəsinə görə fərqlənir. İlkin təhlil göstərir ki, birinci iki alt model çox məhsuldardır. Üçüncü və dördüncü alt modellər nisbətən az işləkdir. Digər beş alt modelə aid birinci növ təyini söz birləşmələrinə dil materialında çox az rast gəlinir. Onu da qeyd etmək lazımdır ki, birinci növ təyini söz birləşmələrini başqa meyar, yaxud meyarlar əsasında da modelləşdirmək mümkündür. Tərəflərin nitq hissəsinə aidiyyəti, tərəfin birinin feili sifət olması, birinci növ təyini söz birləşməsində kəmiyyət kateqoriyasının qrammatik əlamətinin işlənməsi və s. belə meyarlar ola bilər. Nəhayət, çoxkomponentli birinci növ təyini söz birləşmələrini də modelləşdirməyə cəlb etmək mümkündür. Lakin qeyd edilən meyarlar adı çəkilən söz birləşmələrinin avtomatik emalını çətinləşdirir.

**II növ təyini söz birləşmələri:**  $SB = SF_1 + SF_2$ . Burada birinci tərəf sözdəyişdirici şəkilçi qəbul etmir. Alt modellərdə  $SF_1$  birinci növ təyini söz birləşmələrində olduğu kimi dəyişir. İkinci tərəf isə mənsubiyyət şəkilçili olur.

Birinci tərəf  $SF_1$  üç alt modelə malikdir:  $K$ ,  $K + Sd_n$ ,  $K + Sd_1 + Sd_n$ . İkinci tərəf bu üç alt modeldən hər hansı birinə mənsubiyyət şəkilçisinin artırılması ilə düzəlir. Mənsubiyyət şəkilçisini ümumi şəkildə  $Mş$  ilə işarə edək. Bu halda ikinci növ təyini söz birləşmələrinin alt modellərinin birinci növ təyini söz birləşmələrinin ikinci tərəfinə  $Mş$  artırılması ilə formalaşdığını görürük. Məsələn,  $SB = K + K + Mş$ ,  $K + Sd + K + Mş$ ,  $SB = K + Sd_1 + \dots + Sd_n + K + Mş$  və s.

İş paltar (K (iş) + K (paltar) +  $Mş$  (-ı; mənsubiyyət şəkilçisi), aqlıq dövrü [K (ac) + Sd (-lıq) + K (dövr) +  $Mş$  (-ü; mənsubiyyət şəkilçisi)] və s.

**III növ təyini söz birləşmələri:**  $SB = SF_1 + SF_2$ . Söz birləşməsinin hər iki tərəfi sözdəyişdirici şəkilçi qəbul edir. Birinci tərəf yiyəlik hal şəkilçisi qəbul edir, ikinci tərəf isə ikinci növ təyini söz birləşmələrində olduğu kimi, mənsubiyyət şəkilçisi ilə işlənir. Yiyəlik hal şəkilçisini “Yhş” ilə işarə edib III növ təyini söz birləşməsi üçün alt modelləri müəyyənləşdirmək olar.

Əgər III növ təyini söz birləşməsinin birinci tərəfi kök + yiyəlik hal şəkilçisi, ikinci tərəfi kök + mənsubiyyət şəkilçisi formasındadırsa bu tip üçün alt model  $SB = K + Yhş + K + Mş$  olacaqdır. Uşağın paltarı, evin pəncərəsi, qapının dəstəyi, kitabın cildi, qonşunun toyuğu, səsin gücü bu model əsasında formalaşmış III növ təyini söz birləşmələridir. Oxşar qayda ilə digər alt modelləri qurmaq mümkündür. Beləliklə, iki komponenti olan, yaxud iki tərəfdən ibarət olan təyini söz birləşmələrinin həm ümumi modeli, həm də hər bir növün alt modellərinin qurulması məsələsi həllini tapmış olur. Bu modellərin qurulmasında söz-formanın kök, əsas, sözdüzəldici şəkilçi və sözdəyişdirici şəkilçi kimi struktur hissələrindən istifadə edilmişdir.

#### **Təyini söz birləşmələrinin avtomatik tanınması üçün qaydalar sistemi**

Təyini söz birləşmələrinin modelləri morfoloji əlamətlərə əsaslanan qaydalar sistemi vasitəsilə kompüter tərəfindən avtomatik tanına bilər. Bunun üçün hər söz-forma əvvəlcə FSM əsaslı morfoloji analizdən keçirilir və aşağıdakı əlamətlər müəyyənləşdirilir: 1) kök (K); 2) sözdüzəldici şəkilçi (Sd); 3) sözdəyişdirici şəkilçi (Sdf); 4) yiyəlik hal şəkilçisi (Yhş); 5) mənsubiyyət şəkilçisi (Mş)

Bu əlamətlər əsasında söz birləşməsinin növü aşağıdakı qaydalarla təyin edilir.

**I növ üçün qayda:** Şəkilçisiz + şəkilçisiz → I növ təyini söz birləşməsi

**II növ üçün qayda:** Şəkilçisiz + mənsubiyyət şəkilçisi → II növ təyini söz birləşməsi

**III növ üçün qayda:** Yiyəlik hal şəkilçisi + mənsubiyyət şəkilçisi → III növ təyini söz birləşməsi

#### **Alqoritmik təsvir**

for each (SF1, SF2) in text:

analyze\_morphology(SF1)

analyze\_morphology(SF2)

if no\_inflection(SF1) and no\_possessive(SF2):

label = "Type I"

elif no\_inflection(SF1) and has\_possessive(SF2):

label = "Type II"

elif has\_genitive(SF1) and has\_possessive(SF2):

label = "Type III"

Bu alqoritm NP chunking, dependency parsing və rule-based sintaktik analiz sistemlərinə birbaşa inteqrasiya oluna bilər [Imamura].

NLP sistemlərində modelləşdirmənin uğurlu olması konkret proqramlar

əsasında kompüterdə mətnin avtomatik emalının sadələşməsi ilə ölçülməlidir.

Təyini söz birləşmələrinin avtomatik tanınması zamanı əsas çətinlik leksik-qrammatik omonimlikdir. Məsələn, *qızıl qaşığı* birinci növ təyini söz birləşməsidir, lakin “qızıl” həm isim, həm də sifət funksiyasında çıxış edə bilər. Bu kimi hallarda morfoloji analiz nəticələri ilə yanaşı, kontekstual POS işarələmə və leksik baza məlumatlarından istifadə etmək tələb olunur. Beləliklə, təqdim edilən model morfoloji analizator və leksik resurslarla birlikdə işlədikdə yüksək dəqiqlik təmin edir.

Tədqiqat nəticəsində Azərbaycan dilində təyini söz birləşmələrinin klassik sintaktik təsnifatının formal struktur modelləri vasitəsilə kompüter tərəfindən tanına bilən qaydalara çevrilmişdir. Söz-formanın  $K + Sd_1 + \dots + Sd_m$  modeli Azərbaycan dili üçün sonlu avtomat morfoloqiyası əsaslı morfoloji analizatorun nəzəri bazasını yaradır. Bu analiz nəticəsində əldə edilən morfoloji informasiya  $SB = SF_1 + SF_2$  modeli əsasında qaydalara əsaslanan NP-yə parçalama mexanizmi ilə isim qruplarının avtomatik aşkarlanmasına imkan verir.

Beləliklə, təyini söz birləşmələrinin I, II və III növləri morfoloji şərtlərlə dəqiq fərqləndirilərək alqoritmik qaydalara çevrilmiş olur. Bu yanaşma Azərbaycan dili üçün qayda əsaslı sintaktik analiz sistemlərinin qurulması, milli korpusların avtomatik işarələnməsi, maşın tərcüməsi, informasiya axtarışı və terminoloji vahidlərin tanınması kimi NLP tətbiqlərində praktik əhəmiyyət daşıyır.

Təklif olunan modellər Azərbaycan dilçiliyində mövcud sintaktik biliklərin kompüter dilçiliyi müstəvisinə keçirilməsini təmin edir və aqqlütinativ quruluşun üstünlüklərindən istifadə etməklə formal, alqoritmik və tətbiq oluna bilən dil modeli təqdim edir.

## ƏDƏBİYYAT

1. Aslanov, A.Ə. İsmi söz birləşmələri // Müasir Azərbaycan dili. III c. Sintaksis. Bakı: “Elm”, 1981, s. 89-124
2. Bengio, Y., Ducharme, R., and Vincent, P. A neural probabilistic language model / Y. Bengio [et al.] // In Advances in Neural Information Processing Systems. – 2001. – P. 932-938.
3. Faggella D. What is Machine Learning? // Emerj. URL: <https://emerj.com/aiglossary-terms/what-is-machine-learning/>(дата обращения: 03.05.2020).
4. Xəlilov, B. Müasir Azərbaycan dilinin sintaksisi. / B.Xəlilov. – Bakı: Elm, 2008. – 346 s.
5. Imamura, K. Doctor's Thesis Automatic Construction of Translation Knowledge for Corpus-based Machine Translation / K. Imamura. 2004.; Elektron mənbə: chrome-extension://efaidnbmnnnibpccajpcglclefind mkaj

- [/https://aclanthology.org/E03-1029.pdf](https://aclanthology.org/E03-1029.pdf)
6. Kazımov, Q. Müasir Azərbaycan dili. Bakı: “Ünsiyyət”, 2000, 496 s.
  7. Mahmudov M. Türk dillərinin milli korpusları / M.Mahmudov. – Bakı: “Elm və Təhsil”, 2018. -392 s.
  8. R. Van Gog and T. M. Van Engers. Modeling legislation using natural language processing // 2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and eMan for Cybernetics in Cyberspace (Cat.No.01CH37236), Tucson, AZ, USA, 2001
  9. Saarikoski J. et al. On the Influence of Training Data Quality on Text Document Classification Using Machine Learning Methods // Int. J. Knowl. Eng. Data Min. 2015. № 2 (3). pp. 143-169.
  10. Seyidov, Y. Azərbaycan ədəbi dilində söz birləşməsi / Y.Seyidov. – Bakı: Maarif, 1966. – 340 s.
  11. 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.

## Natəvan Cəmaləddin qızı RƏHMANOVA

### **RƏHMANOVA N.C. Azərbaycan dilinin təyini söz birləşmələrinin NLP sistemi üçün modelləşdirilməsi**

#### **Xülasə**

Bu məqalədə Azərbaycan dilində təyini söz birləşmələrinin formal struktur modelləri qurularaq onların NLP (Natural Language Processing) sistemlərində avtomatik tanınması imkanları araşdırılır. Tədqiqatın əsas məqsədi ənənəvi sintaktik təsnifatı kompüter tərəfindən emal oluna bilən formal qaydalara çevirməkdir. Aqqlütinativ quruluşa malik olan Azərbaycan dilində söz-formanın kök + sözdüzəldici şəkilçilər + sözdəyişdirici şəkilçilər ardıcılığı ilə qurulması sonlu avtomat morfoloqiyası yanaşması üçün əlverişli baza yaradır. Bu morfoloji model əsasında söz birləşmələrinin  $SB = SF_1 + SF_2$  formulu ilə ifadəsi isim qruplarının avtomatik aşkarlanmasına, yəni qaydalara əsaslanan ismi frazalara parçalama mexanizminin tətbiqinə imkan verir. Məqalədə təyini söz birləşmələrinin I, II və III növləri morfoloji əlamətlərə əsaslanan qaydalar sistemi ilə modelləşdirilmiş, onların alqoritmik tanınma prinsipləri göstərilmişdir. Təklif edilən modellər morfoloji analiz, sintaktik seqmentasiya, milli korpusların işarələnməsi, maşın tərcüməsi və informasiya çıxarışı kimi NLP istiqamətlərində tətbiq oluna bilər. Tədqiqat nəticəsində Azərbaycan dilçiliyində mövcud sintaktik biliklər hesablama dilçiliyi

müstəvisinə keçirilmiş və aqqlütinativ quruluşun üstünlüklərindən istifadə etməklə formal, alqoritmik və praktik tətbiq imkanlarına malik model təqdim edilmişdir.

**Natavan J. Rahmanova**

### **Modeling of Azerbaijani language word combinations for NLP system**

#### **Abstract**

This article presents formal structural models of attributive word combinations in the Azerbaijani language and explores their applicability in Natural Language Processing (NLP) systems for automatic recognition. The main objective of the study is to transform traditional syntactic classification into formal rules that can be processed computationally. Due to its agglutinative nature, the Azerbaijani word-form structure, consisting of root + derivational suffixes + inflectional suffixes, provides a suitable basis for the Finite State Morphology (FSM) approach. Based on this morphological model, the representation of word combinations as  $SB = SF_1 + SF_2$  enables automatic identification of noun phrases through a rule-based NP chunking mechanism. The study models the three types of attributive word combinations using morphological criteria and demonstrates their algorithmic recognition principles. The proposed models can be applied in morphological parsing, syntactic segmentation, national corpus annotation, machine translation, and information extraction tasks. As a result, traditional syntactic knowledge in Azerbaijani linguistics is transferred into the domain of computational linguistics, providing a formal, algorithmic, and practically applicable language model that leverages the advantages of the agglutinative structure.

**Rəyçi: professor Məsud Mahmudov**