Ə.Fətullayev, M.Mahmudov,
R.Fətullayev, R.Dərvişov
Rabitə və Yüksək Texnologiyalar Nazirliyi

NLP SİSTEMLƏRİNDƏ AZƏRBAYCAN DİLİNDƏ TABELİ MÜRƏKKƏB CÜMLƏLƏRİN FORMAL MEYARLARA GÖRƏ TƏSNİFİ VƏ TƏHLİLİ

Açar sözlər: Azərbaycan dili, sadə cümlə, mürəkkəb cümlə, tabesiz mürəkkəb cümlə, tabeli mürəkkəb cümlə, təhlil alqoritmi, formal əlamət

Ключевые слова: Азербайджанский язык, простое предложение, сложное предложение, сложноподчиненное предложение, сложносочиненное предложение, алгоритм анализа, формальный признак

Key words: The Azerbaijani language, simple sentences, complex sentences, no subordinate complex sentences, the subordinate complex sentences, algorithm of analysis, formal features

Xülasə. Məqalədə NLP sistemlərində Azərbaycan dilində tabeli mürəkkəb cümlələrin formal əlamətlərə görə təsnifi və təhlili məsələləri ilk dəfə olaraq yeni baxış bucağından araşdırılmış, formal təhlilin aparılması üçün linqvistik bazanın formalaşdırılması şərtləri müəyyənləşdirilmişdir. Formal prinsiplərə əsaslanan linqvistik baza formal linqvistik təhlil alqoritmlərini tərtib etməyə və proqramlaşdırma işləri aparmağa imkan yaradır.

NLP sistemləri və Azərbaycan dili. Son illər informasiya texnologiyalarının insan fəaliyyətinin müxtəlif sahələrinə dərindən daxil olması, bütün dünyada NLP (Natural Language Processing - təbii dillərin emalı) sistemlərinin geniş şəkildə inkişafına gətirib çıxarmışdır. Hal-hazırda NLP sistemləri dedikdə maşın tərcüməsi, orfoqrafiyanın və qrammatikanın yoxlanması, nitqin yazıya çevrilməsi və başa düşülməsi, sorğulara avtomatik cavab verilməsi, mətn üzrə insan nitqinin generasiyası və s. sistemlər başa düşülür ki, bu sistemlər də insanla kompüter arasında daha təbii interfeys yaradılmasına, insanın kompüterlərdən səmərəli şəkildə yararlanmasına və kompüterlərdən daha geniş istifadə edilməsinə şərait yaradır.

Azərbaycanda da bu sahədə xeyli işlər görülmüşdür. Keçən əsrin 70-ci illərindən başlayaraq NLP sistemlərin yaradılması sahəsində müəyyən tədqiqat işləri aparılmışdır [1-6]. Lakin praktik olaraq yaradılmış və geniş istifadə olunan ilk NLP sistemləri Rabitə və Yüksək Texnologiyalar Nazirliyinin dəstəyi ilə həyata keçirilən Dilmanc layihəsi* çərçivəsində hazırlanmışdır (www.dilmanc.az) [7]. Belə ki, bu layihə çərçivəsində elektron tərcümə vasitələri (www.dilmanc.az/az/dict), maşın tərcüməsi (www.dilmanc.az), nitqin tanınması (www.dilmanc.az/dilmanc-imla) və adekvat cavab verilməsi, mətnlərin səsləndirilməsi (www.dilmanc.az/az/tts), nitqin yazıya çevrilməsi və s. sistemlər yaradılmış və istifadəyə verilmişdir. Layihə çərçivəsində belə sistemlərin daha keyfiyyətli versiyalarının yaradılması istiqamətində tədqiqatlar davam etdirilir [7]-[15]. Bu məqalə də belə tədqiqatlardan biridir.

NLP sistemlərində formal təhlilin rolu. NLP sistemlərinin yaradılması dilin elementlərinin – xüsusilə soz-formaların və cümlələrin formal (kompüter vasitəsilə) təhlili sistemlərinin yaradılmasını tələb edir. NLP sistemlərinin keyfiyyəti formal təhlil (morfoloji, sintaktik, semantik və s.) alqoritmlərinin keyfiyyətindən ciddi surətdə asılıdır.

Azərbaycan dilində formal linqvistik təhlilin ilk mərhələsi morfoloji səviyyədə təhlil hesab olunur. Morfoloji təhlil ilkin pillə kimi vacibdir və formal linqvistik təhlilin sonrakı mərhələləri üçün zəruri olan linqvistik məlumatın toplanması və sistemləşdirilməsinə imkan yaradır. Sintaktik və semantik təhlil üçün zəruri olan linqvistik məlumatlar bazasının yaradılması morfoloji təhlil sayəsində mümkün olur. Morfoloji təhlil haqqında daha geniş məlumat bu sahə ilə bağlı tədqiqatlarda verilmişdir [6], [8], [9].

Formal sintaktik təhlil mövzusunda Azərbaycanda barmaqla sayılası qədər tədqiqat işlərinə rast

---

* Layihənin elmi rəhbəri akademik Əli Abbasovdur

gəlmək olar. Belə tədqiqat işləri yalnız yuxarıda qeyd etdiyimiz Dilmanc layihəsi çərçivəsində aparılmışdır və qeyd etdiyimiz kimi bu tədqiqatlar davam etdirilir [10]-[12].. Tədqiqatların aparılmasında əsas məqsədlərdən biri Azərbaycan dilinin formal sintaktik analizatorunun yaradılmasıdır ki, bu analizatorun yaradılması Azərbaycan dili üçün NLP sistemlərinin keyfiyyətinin artırılması üçün istifadə ediləcəkdir. Onu da qeyd edək ki, qeyd etdiyimiz layihə çərçivəsində sintaktik analizatorun ilk versiyası artıq yaradılmışdır və hal hazırda istifadə edilməkdədir.

İstənilən dil üçün yaradılmış formal sintaktik analizator nisbətən sadə (həm cümləyə daxil olan söz-formaların sayının azlığı, həm də qrammatik baxımdan) cümlələr üçün daha keyfiyyətli analiz aparır. Bu baxımdan mürəkkəb (tabeli və ya tabesiz) cümlələrin sadə cümlələrə bölünərək sintaktik təhlil olunması daha məqsədəuyğundur. Qeyd edək ki, mürəkkəb cümlənin sadə cümlələrə bölünərək təhlil olunması yeni ideya deyildir və Azərbaycan tədqiqatçıları da belə tədqiqatlar aparmışlar [13]. Bu məqalədə biz tabeli mürəkkəb cümlənin formal sintaktik təhlilinə dair daha bir yanaşmanı təsvir edəcəyik. Fikrimizcə bu yanaşma formal NLP sistemləri yaradıcılarına sintaktik təhlil prosesini avtomatlaşdırmaqda dəstək olacaqdır.

Cümlənin formal əlamətlərinin müəyyənləşdirilməsi, mürəkkəb cümlənin növlərinin təyini, tabeli mürəkkəb cümlədə budaq cümlə konstruksiyalarının diferensiasiyası vacib məsələdir.
Cümlə hər bir dilin daxili qanunları, özünəməxsusluqları ilə formalaşan sintaktik vahid, ünsiyyət vasitəsidir. Azərbaycan dilində (digər Türk dillərində də olduğu kimi) cümlənin formalaşması meyarları və strukturu digər - daha geniş tədqiqat işləri həsr olunmuş (məsələn, ingilis və rus dillərindən) köklü surətdə fərqlənir. Ona görə də, bu dillər üçün yaradılmış formal təhlil (sintaktik, morfoloji və s.) alqoritmlərini eyni ilə Azərbaycan dilinə (o cümlədən digər türk dillərinə) tətbiq etmək mümkün deyildir. Buna görə də, türk dillərinin, o cümlədən Azərbaycan dilinin formal təhlil sisteminə yanaşma üsulu fərqli olmalı, türk dillərinin sintaktik quruluşunun əsaslarına, sintaktik təbiətinə və mahiyyətinə söykənməlidir.

Cümlənin uzunluğu (cümlənin uzunluğu dedikdə cümlədə olan söz-formaların sayı nəzərdə tutulur) artdıqca onun sintaktik təhlili də mürəkkəbləşir. Belə ki, cümlə üzvlərinin sayı dəyişməz olduğundan cümlədə söz-formaların sayı artdıqca hər bir cümlə üzvünün uzunluğu (cümlə üzvünün təşkil olunduğu söz-formaların sayı) da orta hesabla artır. Bu da öz növbəsində təhlil prosesini çətinləşdirir. Dilmanc layihəsi çərçivəsində yaradılmış Azərbaycan dili korpusu (Bu korpus 61 milyondan bir qədər çox cümlədən ibarətdir) üzərində apardığımız tədqiqatlara görə müasir Azərbaycan dilində cümlələrin uzunluğu 1 ilə 100 arasında dəyişir (Korpus yaradılarkən uzunluğu 100-dən yuxarı olan cümlələr bu korpusa daxil edilməmişdir).

Aşağıdakı cədvəldə qeyd etdiyimiz korpusda cümlələrin uzunluğa görə paylanması göstərilmişdir (Cədvəl 1).

Cədvəl 1. Azərbaycan dilində cümlələrin uzunluğa görə paylanması

Sıra №-si	Cümlənin uzunluğu (söz-forma sayı ilə)	Korpusda rast gəlmə sayı	Faizlə
1	1-10	30 702 822	50,330%
2	11-20	21 680 233	35,539%
3	21-30	6 246 797	10,240%
4	31-40	1 665 406	2,730%
5	41-50	705 980	1,157%
6	51-100	2 153	0,004%
	Cəmi	**61 003 391**	**100%**

Göründüyü kimi yazıda uzunluğu 10-dan yuxarı cümlələr kifayət qədər geniş istifadə olunur (50%) və şübhəsiz ki, bu cümlələrin böyük bir hissəsini mürəkkəb cümlələr təşkil edir. Ona görə də mürəkkəb cümlələrin sadə cümlələrə formal olaraq bölünməsi alqoritmlərinin hazırlanması Təbii Dilin Emalı sistemlərinin vacib komponentlərindən biridir. Məsələn, Azərbaycan dilindən xarici dilə tərcümə edən maşın tərcüməsi sistemində daxil edilmiş cümlənin keyfiyyətli emal edilməsi üçün onun sadə cümlələrə ayrılması, hər sadə cümlədə isə cümlə üzvlərinin müəyyən edilməsi tərcümə prosesinin daha keyfiyyətli aparılması işinə kömək edir. Avtomatik sintez edilən xarici cümlədə bağlayıcıların, fel şəkilçilərinin, söz sırasının və s. düzgün seçilməsi üçün mürəkkəb cümlənin sadə cümlələrə ayrılması ilə yanaşı, budaq cümlələrin növü də müəyyən edilməlidir. Aşağıdakı misallar da bunu təsdiq edir.

Misal 1. Mətnin səsləndirilməsi sistemlərində cümləni səsləndirərkən onun ümumi intonasiyası, cümlədaxili pauzaların qoyulması üçün cümlənin tipinin müəyyən edilməsi vacibdir ▲

Misal 2. Təbii dilin başa düşülməsi sistemləri (NLU – Natural Language Understanding) istifadəçinin şifahi əmrlərini kompüterin başa düşəcəyi əmrlər siyahısına çevirən alqoritmlər əsasında işləyir. Bu sistemlər şifahi nitqi yazılı cümlələrə çevirdikdən sonra, cümlədən *niyyət* və *semantik slotlar* adlanan informasiyanı ayıra bilməlidirlər. Məsələn, "elə avtomobillərin siyahısını göstər ki, buraxılış ili 2005-dən yuxarı olsun". Bu cümlədə "niyyət = siyahının göstərilməsi", semantik slotlar isə "tip = avtomobil", "buraxılış ili > 2005" olmalıdır. Daxil edilən cümlə, məsələn, tabeli mürəkkəb cümlə olduqda əsas informasiya, yəni niyyət baş cümlədə, onun detalları isə budaq cümlədə yerləşdiyi üçün, tabeli mürəkkəb cümlənin tipi müəyyən olunmalıdır ▲

Formal təhlil sistemi cümləni və onu təşkil edən söz-formaları ayırmağa (cümləni mətndən, söz-formaları cümlədən) və linqvistik baxımdan formal təhlil etməyə imkan verir. Formal təhlilin ilk pillələrində söz-formalar lüğət formasına gətirilir və onun qrammatik statusu müəyyənləşdirilir. Əldə olunmuş informasiya bütövlükdə cümlənin sintaktik təhlili üçün zəruri məlumat kimi istifadə olunur.
Formal sintaktik təhlil sadə cümlə cümlə üçün daha keyfiyyətli aparıla bilir. Əgər sadə cümlənin tam sintaktik təhlili mümkündürsə, onda daha iri (geniş) sintaktik vahidləri də təhlil etmək olar. Başqa sözlə, sadə cümlənin təhlil üsulları mürəkkəb cümləyə də tətbiq oluna bilər. Mürəkkəb cümləni onu təşkil edən sadə cümlələrə ayırmaqla hər bir sadə cümləni həmin qayda ilə ayrı-ayrılıqda təhlil etmək mümkündür. Nəzərə almaq lazımdır ki, mürəkkəb cümlə (tabesiz və ya tabeli) sadə cümlələrin birləşməsi yolu ilə düzəlir. Sadə cümlələr müəyyən bağlayıcıların (tabesizlik və ya tabelilik) və qəlib sözlərin köməyi ilə mürəkkəb cümlə əmələ gətirir. Tabeli mürəkkəb cümlələrin növlərinin müəyyənləşdirilməsi prosesində həmin bağlayıcı və qəlib sözlər fərqləndirici əlamət kimi götürülür.

Tabesiz mürəkkəb cümlələrin sadə cümlələrə bölünməsi alqoritmləri artıq müəyyən dərəcədə araşdırıldığından [13], biz burada əsas diqqəti tabeli mürəkkəb cümlənin sadə cümlələrə bölünməsi məsələsinin həllinə yönəldəcəyik.

Məlum olduğu kimi, cümlənin sintaktik vahid kimi formalaşmasında üç mühüm amil – predikativlik, modallıq və intonasiya mühüm rol oynayır [16]-[18].

Cümlənin bu sadalanan əlamətləri sırasında predikativlik ən mühümü hesab olunur. Formal təhlil baxımından predikativlik aydın, müşahidəsi mümkün olan, müəyyən əlaməti olan kateqoriyadır. Bu sadalanan xüsusiyyətləri modallıq və intonasiyaya aid etmək olmaz. Modallıq, danışanın söylədiyi hadisəyə, ifadə etdiyi fikrə olan münasibəti bildirir. Predikativlik cümlənin kommunikativliyini, modallıq isə emosional-ekspressiv funksiyasını ifadə edir. Predikativlikdən fərqli olaraq modallıq formal olaraq müəyyənləşdirilə bilmir. Çox vaxt intonasiyanın və modal sözlərin köməyi ilə cümlədə modallıq məzmunu yaradıla bilər. Bu baxımdan formal meyarlara əsaslanmaqla modallığı müəyyənləşdirmək, qeydə almaq və formal təhlil sisteminə daxil etmək mümkün deyil.

Modallıq kimi intonasiyanı da formal cəhətdən bu və ya digər əlamət və ya semantik xüsusiyyətinə görə ayırmaq mümkün deyil. Qeyd etmək lazımdır ki, intonasiya daha çox şifahi nitqə xas olan kateqoriyadır və fonetik səviyyədə araşdırma obyekti ola bilər. Cümlənin sintaktik vahid kimi formalaşmasında müəyyən rolu olan modallıq və intonasiya formal meyarlarla təyin oluna bilmədiyi

üçün formal təhlil sistemlərində nəzərə alınmır. Formal təhlil sistemlərində predikativlik ön plana keçir, modallığı və intonasiyanı da ehtiva edərək cümləni formal baxımdan təhlil, təsnif və tərcümə etməyə imkan verir. Deməli, formal sintaktik təhlildə əsas yük predekativliyin üzərinə düşür.

Cümlənin xəbəri zaman, şəxs şəkilçilərinin və modallığın köməyi ilə predikativliyin yaranmasına və formalaşmasına şərait yaradır. Dilçilikdə predikativliklə bağlı belə bir fikir formalaşıb ki, əslində cümləni təşkil edən hər bir sözdə predikativlik olub, sonralar xəbər cümlədə predikativliyin əsas ifadəçisinə çevrilmişdir. Beləliklə, xəbər cümlədə predikativliyin mərkəzi fiquru olmaqla, monopredikativliyin yaranmasını şərtləndirmişdir [16]-[18].

Predikativlik cümlənin mahiyyətində öz əksini tapır və bütün cümlə növlərində bu və ya digər şəkildə təzahür edir. Tədqiqatçılara görə, bütün cümlə konstruksiyalarında, təktərkibli cümlələrdə, hətta üzvlənməyən cümlələrdə belə predikativlik və predikativ münasibət olur. Lakin, onların təzahür formaları və dərəcələri fərqlidir [18, s.83].

Predikativliyin ifadə vasitələri – felin forma, zaman, şəxs şəkilçiləri, ədatlar, intonasiya və modal sözlərdir. Predikativlik cümlənin ən mühüm əlaməti və funksiyası olan kommunikativliyin əsasını təşkil edir.

Məhz bu cəhət nəzərə alınaraq mürəkkəb cümlənin təsnifi və təhlilində əsas meyar kimi predikativlik seçilmişdir.

Predikatlarının sayı iki və ya ikidən çox olan sintaktik vahidlər mürəkkəb cümlə kateqoriyasına aid edilir. Məlum olduğu kimi, mürəkkəb cümlə quruluşuna görə tabesiz və tabeli mürəkkəb cümlələrə ayrılır.

Qeyd etmək lazımdır ki , tabesiz mürəkkəb cümlə və həmcins xəbərli sadə cümlələrdən fərqli olaraq tabeli mürəkkəb cümlələrdə hər bir budaq cümlənin müəyyən fərqləndirici formal əlaməti vardır. Həmin əlamətlər əsasında hər bir budaq cümləni müəyyənləşdirmək mümkündür. Bu isə tabeli mürəkkəb cümlələrin təsnifi, təhlili və başqa dillərə tərcüməsi baxımından olduqca əhəmiyyətlidir. Bu konkret əlamətlər baş cümlədə qəlib söz funksiyasında budaq cümləni əvəz etmiş olur. Baş cümlə budaq cümləyə həmin qəlib sözlərlə, ” ki” bağlayıcısı və vergüllə bağlanır:

1. Mübtəda budaq cümləsində baş cümlədə mübtəda vəzifəsində aşağıda göstərilən qəlib sözlər işlənir**: o(da), bu(da),orası(da), burası(da), eləsi, beləsi, elələri, belələri, bəziləri, bir neçələri, o cəhət, bu cəhət, bir cəhət, o şey, bu şey, bir şey, o məsələ, bu məsələ., kim, hər kim (ki), hər kəs (ki),nə, hər nə (ki)**. Bu sözlər, bəzən, buraxıla da bilər. O zaman xəbər funksiyasını—**məlumdur ki, məlum oldu ki, aydındır ki, bəllidir ki, bilinir ki** və s. sözlər yerinə yetirir.

- Kim ki işinə ciddi yanaşmır, çıxıb gedə bilər.

2.Xəbər budaq cümləsində baş cümlədə qəlib söz funksiyasını işarə əvəzlikləri yerinə yetirir**: budur (ki), odur(ki), bu idi, bu imiş, o imiş, oradadır, buradadır, orasındadır, burasındadır, ondadır, bundadır, burada imiş, burada idi, orada imiş, orada idi, burasında idi, burasında imiş, orasında idi, orasında imiş, ona görədir ki, ondan ötrüdür ki ,o deməkdir ki, bu deməkdir ki, ondan ibarətdir ki, bundan ibarətdir ki, o demək idi ki, bu demək idi, ondan ibarət idi, bundan ibarət idi, o demək imiş, ondan ibarət imiş, o olar ki, bu olar ki, bu olub ki, o olub ki, bu oldu ki, o oldu ki, o şeydir ki, bu şeydir ki, o kəsdir ki, bu kəsdir ki, o kəslərdir ki, bu kəslərdir ki və s.**

- Arzum budur ki, gələcəkdə həkim olum.

3. Tamamlıq budaq cümləsində baş cümlədə qəlib kimi aşağıdakı sözlər işlənir: **ona, buna, onu, bunu, onda, bunda, ondan , bundan, orasını, burasını, orasını da, burasını da, eləsini, beləsini, elələrini, belələrini, elələrinə, belələrinə, elələrində, belələrində, eləsinə, beləsinə, eləsində, beləsində, o cəhəti, bu cəhəti, o cəhətə, bu cəhətə, o cəhətdə, bu cəhətdə, o məsələyə, bu məsələyə, o məsələni, bu məsələni, o məsələdə, bu məsələdə, o məsələdən, bu məsələdən, o şeyi, bu şeyi,o şeyə, bu şeyə, o şeydə, bu şeydə, o şeydən, bu şeydən vəs.**

- Orasını da bilirlər ki, mən vicdanla xidmət göstərmişəm.

4. Təyin budaq cümləsi baş cümlədə işarə əvəzliyi ilə ifadə olunan qəlib sözün mənasını dəqiqləşdirir.Budaq cümlənin göstəricisi olan qəlib sözlər bunlardır: **elə, elə bir, belə, belə bir, o cür, bu cür, o, bu, həmin, haman, hansı, kim, nə və s.** Təyin budaq cümləsinin qəlib sözləri təyin etdikləri ad nitq hissəsindən əvvəl gəlir.

- Elə müəllimlər var ki, şagirdlər dinləməkdən yorulmurlar.

5.Tərz budaq cümləsi – baş cümlədəki hərəkətin icra tərzini, necə nə şəkildə icra olunduğunu bildirir və baş cümlə budaq cümləyə **elə, belə, o cür, o tövr, o yolla, bu yolla, bu halda, o halda, necə, nə cür, nə təhər qəlib sözləri, " ki "** bağlayıcısı ilə bağlanır.

- Elə get ki, heç kəs görməsin.

6. Dərəcə budaq cümləsi – baş cümlədə ifadə olunan hərəkət və ya əlamətin dərəcəsini bildirir və baş cümləyə **" ki "** bağlayıcısı, intonasiya və aşağıdakı qəlib sözlərlə bağlanır**: o dərəcədə, o qədər, elə, o yerə, bu yerə, oracan, o məqama, bir elə mərtəbəyə və s.**

- O qədər yorulmuşdum ki, nə vaxt yuxuya getdiyimi bilməmişəm.

7. Zaman budaq cümləsi – baş cümlədəki hərəkət və ya əlamətin zamanını bildirir və baş cümləyə **" ki"** bağlayıcısı , intonasiya , **eləcə də o zaman, o vədə, o vaxt, elə bir vaxtda, onda, elə ki, o zaman ki, o gün ki, o vaxt ki, nə qədər, nə qədər ki, necə ki** və s. bağlayıcı- qəlib sözlərlə bağlanır.

- O zaman gəldi ki, artıq gec idi.

8. Yer budaq cümləsi- baş cümlədəki hərəkət və ya əlamətin yerini bildirir və budaq cümləyə **orada, oradan, o yerə, o yerdə, o yerdən, həmin yerə, həmin yerdən, hansı yerə, hansı yerdən, hansı yerdə, haranın ki, haraya ki, haranı ki, harada ki, haradan ki, o yerə ki, o yerdə ki, o yerdən ki, o yerin ki, bir yerə ki, bir yeri ki, bir yerdə ki, bir yerdən ki** və s. kimi qəlib-bağlayıcı sözlərlə bağlanır.

- O yerdə ki, böyük-kiçik yoxdur, orada bərəkət olmaz.

9. Kəmiyyət budaq cümləsi – baş cümlədəki hərəkət və ya əlamətin kəmiyyətini bildirir və budaq cümləyə **nə qədər, hər nə qədər, o qədər, o kəmiyyətdə, o miqdarda** qəlib sözləri ilə bağlanır.

- O qədər yazdım ki, tamam əldən düşdüm.

10. Səbəb budaq cümləsi – baş cümlədə ifadə olunan hərəkət və ya əlamətin səbəbini bildirir və budaq cümləyə **ona görə, ondan ötrü, onun üçün, o səbəbə, bu səbəbə, buna görə, bundan ötrü, ona görə ki, ondan ötrü ki, o səbəbə ki, çünki, ki** bağlayıcı və qəlib sözlərlə bağlanır.

- Sən haqlı deyilsən, o səbəbə ki, məsələdən xəbərin yoxdur.

11. Məqsəd budaq cümləsi – baş cümlədə ifadə olunan hərəkətin məqsədini bildirir və baş cümləyə " ki" bağlayıcısı və intonasiya, **eləcə də, ona görə, ondan ötrü, o məqsədlə** sözləri ilə bağlanır.

- Ondan ötrü dedim ki, xəbərin olsun.

12. Şərt budaq cümləsi – baş cümlədə ifadə olunan hərəkət və ya əlamətin şərtini bildirir və baş cümləyə intonasiya **, " ki"** bağlayıcısı**, -sa,-sə(isə)** köməkçi sözü və şəkilçisi**,** eləcə də **bir şərtlə, bu şərtlə, o şərtlə, həmin şərtlə** qəlib sözləri**, əgər(ki), madam ki, indi ki, vaxta ki, bir halda ki** bağlayıcı sözləri və ya **ki** ədatı ilə bağlanır.

- Madam ki, etiraf etdin, onda sənə güzəşt etmək olar.

13. Qarşılaşdırma budaq cümləsində baş və budaq cümlənin məzmunu bir-biri ilə **nə qədər, hər nə qədər, hərçənd** bağlayıcı sözləri**,-sa,-sə** şəkilçisi**, da(də) və belə** ədatları ilə qarşılaşdırılır, müqayisə edilir.

- Nə qədər çalışdım, xeyri olmadı.

14. Nəticə budaq cümləsi baş cümlədən doğan nəticəni bildirir və baş cümləyə intonasiya, **" ki"** bağlayıcısı**,** eləcə də **odur ki, ona görə (də), buna görə (də), onun üçün (də), bunun üçün (də)** bağlayıcıları ilə bağlanır.Bəzən baş cümlənin daxilində **elə, elə bir, o qədər** sözləri də işlənə bilər.

- Ona görə inciyib ki, dərslərini yaxşı öyrənməmisən.

15. Qoşulma budaq cümləsi -- baş cümləyə qoşularaq, ya baş cümlədəki fikirlə bağlı əlavə məlumat verir, yaxud da həmin fikrə münasibət bildirir.Baş cümlə budaq cümlədən əvvəl işlənir ona intonasiya və **“ ki”** bağlayıcısı ilə bağlanır. Budaq cümlənin əvvəlində ismin müxtəlif hallarında **bu, o** əvəzlikləri və **da (də)** ədatı işlənir.

- Tarlada bircə adam var idi, o da heç bir iş görmürdü.

Burada daha bir mühüm cəhət diqqətdən yayınmamalıdır: mürəkkəb cümlə ilə budaq cümlənin sərhədlərini necə müəyyənləşdirməli? Baş cümlə harada qurtarır və budaq cümlənin başlanğıcı və sonu necə təyin olunur? Formal olaraq baş cümlə budaq cümləyə konkret qəlib sözlər, “ki” bağlayıcısı və vergüllə bağlanır. Elə həmin əlamətlərə görə də budaq cümlənin növünü müəyyənləşdirmək mümkündür. Məsələn, aşağıdakı cümləni nəzərdən keçirək:

Arzum budur ki, müəllim olum.

Bu nümunədə “budur” işarə əvəzliyi ilə ifadə olunan qəlib söz cümlənin xəbər budaq cümləsi olduğunu göstərir. Qəlib söz, “ki” bağlayıcısı və vergül baş cümləni səciyyələndirən əlamətlərdir. Budaq cümlə(xəbər budaq cümləsi) icə “müəllim olum” hissəsidir.

Deməli, yuxarıda qeyd etdiyimiz fərqləndirici əlamətlər olduqda baş cümlə ilə budaq cümlənin sərhədinin müəyyənləşdirilməsində problem yaranmır. Problem onda yaranır ki, bu əlamətlər tam və ya qismən buraxılır. Məsələn, qarşılaşdırma budaq cümləsində baş və budaq cümlələr bir-birinə “nə qədər”, “hər nə qədər”, “hərçənd” bağlayıcı sözləri, “sa/ sə” şəkilçiləri, “da/də”, “belə” ədatları ilə bağlana bilər:

Nə qədər(hər nə qədər) dedimsə, sözündən dönmədi. Bu tipli cümlələrdə formal əlamətlərin bəziləri(bəzən hamısı) buraxıla bilər:

Nə qədər dedim, sözündən dönmədi/Dedim, sözündən dönmədi və s.

Bu qəbildən olan cümlələr , fikrimizcə, formal olaraq tabeli cümlə olmaq əlamətini itirir, tabesiz mürəkkəb cümləyə çevrilmiş olur. Onları tabesiz cümlə kimi təhlilə cəlb etmək olar, baxmayaraq ki, semantikasına görə bu cümləni qarşılaşdırma budaq cümləsi hesab etmək olar.

Fikrimizcə, bu ideya formal əlamətləri qismən və ya tamamilə buraxılmış bütün tabeli mürəkkəb cümlə tiplərinə aid edilə bilər. Deməli, tabeli mürəkkəb cümlənin formal əlamətləri buraxılmışsa (yoxdursa), semantika nəzərə alınmır (nəzərə alınması mümkün olmur) və həmin cümlələr ənənəvi dilçilikdən fərqli olaraq tabesiz mürəkkəb cümlə tipinə aid edilir.

Təqdim olunan iş mətnlərin formal təhlili sahəsində aparılan araşdırmalar çərçivəsində linqvistik yanaşma üsullarından biri kimi səciyyələndirilə bilər. Hər şeydən əvvəl, NLP sistemləri üçün sintaktik analizatorun işlənib hazırlanması işlərində bu araşdırmalar ilkin hazırlıq, baza hesab oluna bilər. Gələcəkdə Azərbaycan dilinin formal qrammatikasının hazırlanması prosesində də cümlənin formal təhlili sahəsindəki tədqiqatlardan faydalanmaq olar.

Nəzərə almaq lazımdır ki, formal təhlil açıq sistemdir: təhlil prosesində meydana çıxan səhvlər, düzgün alınmayan nəticələr diqqətlə araşdırılır, zəruri təkmilləşdirmələr aparılır, düzəlişlər edilir və düzgün nəticə almağa imkan verən versiyalar mütəmadi olaraq sistemə daxil edilir. Məqsəd Azərbaycan dilində formal əlamətlər əsasında cümləni təhlil və təsnif etmək, alınan nəticələrin İKT sahəsində aparılan tədqiqatlara uğurlu tətbiqini təmin etməkdir.

Bu sahədə toplanmış təcrübə belə bir nəticəyə gəlməyə imkan verir ki, Azərbaycan dili üçün tərtib olunmuş sintaktik təhlil alqoritmləri digər türk dillərinə də şamil oluna bilər.

İstifadə olunmuş ədəbiyyat

1. Велиева К.А. Формальное описание синтеза азербайджанского слово. АКД, Москва, 1971,20 с.

2. Пинес В. Я. Моделирование структуры азербайджанских глагольных форм в связи с проблемой автоматического словаря. АКД, Москва, 1970, 19 с.
3. Mahmudov M. Maşın tərcüməsi sistemində avtomatik morfoloji analizin yeri. Azərbaycan Elmlər Akademiyasının Xəbərləri. Ədəbiyyat, dil və incəsənət seriyası, №1, 1981, s. 68-71
4. Pines V,Mahmudov M. Elektron hesablayıcı maşınlar dilçilikdə. Bakı, AzETETMİ, 1977, 32 s.
5. Махмудов М. А. Система автоматической переработки тюркского текста на лексико-морфологическом уровне. Баку, «Елм», 1991, 143 с.
6. Mahmudov M. Mətnin formal təhlili sistemi. Bakı, “ Elm”, 2002, 244 s.
7. Fatullayev R, Abbasov A, Fatullayev A. 2008. Dilmanc is the 1st MT system for Azerbaijani. In: Proc. of SLTC-08, Stockholm, Sweden, pp.63-64
8. Fatullayev R, Abbasov A, Fatullayev A. 2008a. Set of active suffix chains and its role in development of MT system for Azerbaycani. In: Proc. of IMCSIT-08, Wisla, Poland, pp.363-368
9. Fatullayev R., Abbasov A., Fatullayev A. Overcoming Agglutination Difficulties in the Development of an MT system from the Azerbaijani Language/ Speech and Language Technology. International journal of Polish phonetics association. Poznan 2008, vol. 11, pp. 215-228
10. Fətullayev Ə. Tərcümə prosesinin modelinin qurulması və bu model əsasında Azərbaycan-ingilis maşın tərcüməsi sisteminin tərkib hissələrinin müəyyən edilməsi. Azərbaycan Milli Elmlər Akademiyasinin Xəbərləri(FRT), XXV cild, № 2, 2005, səh. 181-186
11. Fətullayev Ə. Azərbaycan dilində cümlədə feli birləşmələrin formal təyin olunması alqoritmi. AMEA-nın Məruzələri. Bakı, Elm nəşriyyatı 2004, № 5-6, səh. 38-45
12. Abbasov A., Fatullayev A. The use of syntactical and semantic valences of the verb for formal delimitation of verb word phrases/ In Proceedings of the 3rd Language & Technology Conference (L&TC’07), 2007, Poznan, Poland, pp. 468-472
13. Fatullayev R., Mammadova S., Fatullayev A. Translating Composite Sentences in Azerbaijani-English MT System/ International Symposium on Data and Sense Mining Machine Translation and Controlled Languages and their application to emergencies and safety critical domains, University of Franche-Comté, Besanchon, France 2009, pp. 104-113
14. Fətullayev Ə. və b. Mətnin səsləndirilməsi sistemi və Azərbaycan dilinin fonemlərinin fono-akustik bloku. Terminologiya məsələləri. Bakı, Elm nəşriyyatı, 2015, №1, səh. 33-49
15. Fatullayev R, Abbasov A, Fatullayev A. 2008. Peculiarities of the development of the dictionary for the MT System from Azerbaijani. In: Proc. of EAMT-08, Hamburg, Germany, pp.35-40
16. Abdullayev Ə. Müasir Azərbaycan dilində tabeli mürəkkəb cümlələr. Bakı, “Maarif”, 1974, 390s.
17. Abdullayev Ə., Seyidov Y.,Həsənov A. Müasir Azərbaycan dili (sintaksis), Bakı,” Maarif”, 1972, 464 s.
18. Kazımov Q.Ş. Müasir Azərbaycan dili. Bakı,”Elm və təhsil”, 2010, 500 s.

А. Фатуллаев, М. Махмудов,
Р. Фатуллаев, Р. Дервишов

Классификация и анализ сложноподчиненных предложений азербайджанского языка по формальным критериям в системах NLP

Резюме

В статье впервые исследованы вопросы классификации и анализа сложноподчиненных предложений азербайджанского языка по формальным критериям в системах NLP, даны принципы формирования лингвистической базы для проведения формального лингвистического анализа.

Лингвистическая база основанная на формальных признаках дает возможность составления алгоритма формального лингвистичеслого анализа и программирования.

A.Fatullaev, M. Mahmudov,
R. Fatullaev, R. Dervishov

Classification and analysis of azerbaijani language's subordinate complex sentences by formal criteria in NLP systems.

Summary

This article researches for the first time problems of classification and analysis of subordinate complex sentences in azerbaijani language by using formal criteria in NLP. It provides principles for creation of the linguistic base that enables to conduct formal linguistic analysis. Linguistic base created on formal attributes gives opportunity to design the algorithm for formal linguistic analysis and programming it.