

Natəvan Əzizli

AMEA Nəsimi adına Dilçilik İnstitutunun doktorantı

natis-sdunina@mail.ru

Maşın tərcüməsində obyekt dildə söz sırasının avtomatik qurulması

Açar sözlər: avtomatik tərcümə, alqoritm, cümlə, mübtəda, xəbər, komponent, üçkomponentli cümlə

Ключевые слова: автоматический перевод, алгоритм, существительное, сказуемое, компонент, трехкомпонентное предложение

Keywords: automatic translation, algorithm, sentence, subject, predicate, component, triple component sentence

Mürəkkəb sistem təşkil edən dilin informasiyanı ötürmə mexanizminin ümumi iş prinsipini öyrənmək üçün həmin sistemin mərkəzi elementini müəyyənləşdirmək lazım gəlir. İnformasiyanın tamlığı və ya nisbi tamlığının vacibliyi mərkəzi element olaraq cümlənin seçilməsini zəruriləşdirir. Cümlə sintaktik sistemin fəaliyyətinin son vahidi sayılır. Sintaktik vahid kimi cümlənin müxtəlif tərifləri vardır. O.Axmanova cümləni izah edərkən onun üç izahını vermişdir: «Cümlə -1. Qrammatika və intonasiyaya görə verilmiş dilin qanunlarına əsasən formalaşmış nitq vahidi olub müəyyən gerçəklik haqqında fikri və məlumatı, danışanın həmin hadisəyə münasibətini ifadə edən başlıca vasitədir. 2. Tərkibində iki baş üzvün, yaxud daha mürəkkəb hallarda mübtəda və xəbər qrupunun tam aşkar şəkildə ifadə olunduğu ikiüzvlü sintaktik kompleksdir. 3. Mürəkkəb cümlənin bir hissəsidir» [1, s. 347]. Cümlə sintaktik konstruksiya olsa da, yeganə sintaktik quruluş deyildir. Ona görə də cümləyə kommunikasiyanın minimal vahidi kimi baxmaq daha düzgündür. Sözlər və söz birləşmələri cümlənin daxili elementləridir. Cümlə sözdən və söz birləşməsindən fərqli olaraq, aktualaşdırılmış və gerçəkliyin müəyyən şəraitinə uyğun gələn hadisəni, nisbi bitmiş fikri ifadə edir. Sadə cümlə, mürəkkəb cümlə, adlıq cümlə, cümlə, cümlə, cümlə və s. növlərdir. Cümləni səciyyələndirən əsas əlamət onun quruluş, semantika və pragmatikaya görə formalaşmasını tələb edir. Bu əlamətlərdən birinin yoxluğu cümlənin formalaşmadığını göstərir. Bir sözdən və ya söz birləşməsindən ibarət olan adlıq cümlə, həmçinin söz-cümlə üçün formallaşdırma alqoritmini quramaq çətin deyildir. «Adlıq cümlələr iki əsas anlayış - əşya, hadisə və onun mövcud olması, varlığı haqqında anlayışları bildirir. Adlıq cümlə insan şüurunun

bilavasitə qavrayışı nəticəsində əmələ gələn əşya və hadisələrin obrazlarını verməyə xidmət edir. Belə obrazlar nəql zamanı söyləyən tərəfindən canlandırılır» [2, s.275].

Hər bir nisbi bitmiş məlumat mütləq iki komponentin fəal, dinamik birləşməsinə nəzərdə tutur. Cümlənin iki komponenti kimi onun subyektivi və obyektivi təyin etmək o qədər də çətin deyildir.

«Cümlənin qurulması və ya canlandırılması zamanı «predikativ münasibət» və ya sadəcə «predikativlik» adlandırılan sintaktik konstruksiyanın əhəmiyyəti, eləcə də vacibliyi danılmazdır. Beləliklə, predikativliyin hökmün subyektivi və obyektivi kimi iki komponentə bölünməsindən doğması aşkara çıxır. Hökmün subyektivi və predikativ hər bir cümlənin vacib komponentləridir» [3, s.37].

Predikativ münasibət hər bir cümlənin əsasında duran və yalnız cümləyə münasib cümlədə mümkün olandır. Cümləni təşkil edən predikativliyin vəzifəsi cümlənin məzmununu ifadə etmək və onun gerçəkliklə əlaqələndirilməsini təmin etməkdən ibarətdir. Maşın tərcüməsində cümlə səviyyəsində seqmentasiya əhəmiyyətli və ilkin vacib şərtlərdən biridir. Mətni cümlələrə bölmək onu tərcümə vahidlərinə və ya elementlərinə ayırmaqdır. Cümlələr arasındakı semantik əlaqələrin ümumi informasiyaya əlavə edə biləcəyi mənanın təyini tərcümənin sonrakı mərhələsində reallaşdırıla bilər. Müasir avtomatik tərcümə proqramları ilkin tərcümənin yaxşılaşdırılması üçün müəyyən ekspert bloklarını nəzərdə tutur.

Cümlə üçün predikativ münasibəti (mübtəda ilə xəbərin münasibətini) xarakter əlamət hesab etmək olar. Bu münasibətin ifadəsində şəxs və kəmiyyətə görə uzlaşma da xüsusi rol oynayır.

Avtomatik tərcümə mənbə dildəki cümləni təşkil edən elementlər -söz və söz-formalar arasındakı əlaqələri, eləcə də strukturu, düzülüş ardıcılığını təyin etmək və bundan sonra obyekt dildə mənbə dilə müvafiq semantikanı saxlamaqla elementlərin düzülüş ardıcılığını yaratmaqdır. Şübhəsiz ki, avtomatik tərcümənin bu və ya digər dərəcədə mükəmməl alqoritmini qurmaq üçün sadədən mürəkkəbə doğru getmək məqsədəuyğundur. Bu baxımdan, birinci növbədə, iki elementdən ibarət cümlələrin avtomatik tərcüməsi sxemini işləmək lazım gəlir. Azərbaycan dilindən fərqli olaraq, ingilis dili nəqli cümlələrində mübtəda və xəbər həmişə ayrıca sözlə ifadə olunur. Ona görə də ingilis dilinin sadə nəqli cümlələrində iki element olduqda bunlardan biri mübtəda, digəri xəbərdir. Azərbaycan dili cümlələri üçün bu hökm doğru deyildir. Şəxs şəkilçiləri feilə əlavə olunaraq hərəkətin kim tərəfindən yerinə yetirildiyini göstərir.

Avtomatik morfoloji analizdə cümləni təşkil edən söz-formaların qrammatik əlamətləri təyin olunur. Bu cəhət leksik vahidi müəyyənləşdirmək və onun qarşılığını avtomatik lüğətdən seçmək imkanı yaradır. Məsələn, ingilis dilindəki “I go” cümləsinin birinci elementinin qarşılığı “Mən”, ikinci elementinin qarşılığı isə avtomatik morfoloji analizə istinadla “gedirəm” olacaqdır. Ən sadə hallardan biri olan bu halda mənbə və obyekt dilin söz sırası nəzərə alınmaqla “Mən gedirəm” cümləsini formalaşdırmaq mümkündür. Maşın birinci elementin ingilis dilindəki variantını onun Azərbaycan dili qarşılığı ilə əvəz edir və növbəti addımda eyni fəaliyyət ikinci element üçün yerinə yetirilir. Beləliklə müxtəsər cümlənin ən sadə növünün avtomatik tərcüməsi sadə əvəzetmə alqoritmi ilə reallaşır. İngilis dilində müəyyənlik kateqoriyası müəyyən və qeyri müəyyən artikllərin məhsuldar işlənməsini təmin edir. İngilis dilində sadə müxtəsər cümlələrdə cümlə üzvlərinin bir düzülüşü vardır.

Müasir ingilis dilində bəzi sözlərin isim olması ancaq onların qarşısında artiklin işlədilməsi ilə müəyyən edilir, məsələn: question, answer, hand, sentence və s. qəbildən olan sözlərin qarşısında artıkl işlənmədikdə, onların isim olmasını göstərəcək heç bir morfoloji əlamət olmur. Həmin sözlərin qarşısında artıkl işləndikdə isə onların isim olduğu qəti müəyyənləşir. Buna görə də müasir ingilis dilində artıkl isim əlaməti hesab edilir.

Artıklın nitqdə başlıca vəzifəsi ismə xas olan qeyri-müəyyənlik və müəyyənlik kateqoriyalarını ifadə etməkdən ibarətdir. Bununla əlaqədar olaraq da müasir ingilis dilində iki artıkl vardır:

- 1) qeyri-müəyyənlik artıklı: a;
- 2) müəyyənlik artıklı: the.

Qeyri-müəyyənlik artıklı qədim ingilis dilində mövcud olmuş an (bir) sayından əmələ gəlmişdir. Bu artıklə tarixən xas olan say mənası bu gün də bəzi hallarda aydınca hiss olunmaqdadır.

Qeyri-müəyyənlik artıklının iki forması vardır: a və an. Bu formalardan hansının nə zaman işlədilməsi təsadüfi olmayıb, ondan (artıklədən) sonra gələn sözün (isim, yaxud onun təyini rolunda işlənən sifətin) sait, yaxud samitlə başlanması ilə şərtlənir: isim yaxud onun təyini rolunda çıxış edən sifət (1) samitlə başlanarsa, qeyri-müəyyənlik artıklının a, (2) saitlə başlanarsa, an forması işlənilir.

Qeyri-müəyyənlik artıklı bir qayda olaraq sayıla bilən tək isimlərlə işlənilir; cəmdə ismin qarşısında qeyri-müəyyənlik artıklı işlənmir.

Cümlədə müəyyən və ya qeyri-müəyyən artıkl olduqda cümlə üç elementli olur. “The bird flies” tipli cümlələrin obyekt dildə verilməsi üçün cümlənin birinci elementinin artıklə görə yoxlanması tələb olunur.

İngilis dilində seqmentasiyadan sonra alınmış cümlə üç komponentə malikdirsə, birinci komponentin “the” və ya “a”, yaxud “an” olması yoxlanılır. Alqoritm aşağıdakı şəkildə qurulur.

1. Birinci komponenti “the” ilə müqayisə et.
2. Əgər “the” olarsa, keç 7, olmazsa keç 3
3. Birinci komponenti “a” ilə müqayisə et.
4. Əgər “a” olarsa, keç 7, olmazsa keç 4
5. Birinci komponenti “an” ilə müqayisə et.
6. Əgər “an” olarsa, keç 7, olmazsa keç 11
7. İkinci komponenti götür və onun qarşılığını avtomatik lüğətdən tap.
8. İkinci komponentin qarşılığını birinci mövqedə yaz.
9. Üçüncü komponenti götür və onun qarşılığını tap.
10. Üçüncü komponentin qarşılığını ikinci mövqedə yaz.
11. Son

Yuxarıda verilmiş “The bird flies” cümləsinin bu alqoritm əsasında emalı nəticəsində “the” leksik mənalı söz kimi nəzərə alınmır. Alqoritmə ikinci komandadan sonra 7-ci komandaya keçid alınır. Bundan sonra digər komandalar ardıcıl yerinə yetirilir və “The bird flies” cümləsinin obyekt dildəki qarşılığı “quş uçuş” alınır. Təbii ki, ingilis dilində söz sırasına görə müəyyənləşən zaman və hərəkətin üçüncü şəxsə aidliyi avtomatik morfoloji təhlil blokunda təyin olunur və “uçuş” söz-forması formalaşdırılır.

İngilis dilində tərkibində üç komponent olan və Azərbaycan dilində ikikomponentli cümlə formasına uyğun gələn cümlələr də vardır. Onlardan biri zamanın ifadəsi ilə bağlıdır. *Shall* köməkçi feili əsas məna daşıyan feillə bir yerdə birinci şəxsin tək və cəmi ilə gələcək zaman formasını düzəltmək üçün işlənir. Məsələn, *I shall go*.

Bu tipli müxtəsər cümlənin Azərbaycan dilinə avtomatik tərcüməsi “shall” komponentinin cümlədə qrammatik funksiyanı yerinə yetirməsini aydınlaşdırmaq tələb edir. Cümlə üç komponentli olduqda onun ikinci komponentinin “shall” olması yoxlanmalıdır. Bu hal üçün qurulan alqoritm bundan əvvəl artiklin təyini üçün qurulmuş alqoritmlə müştərək işləməlidir. Yəni alqoritm mümkün vəziyyətlərin hər birini əhatə etməlidir.

1. Birinci komponenti “the” ilə müqayisə et.
2. Əgər “the” olarsa, keç 7, olmazsa keç 3
3. Birinci komponenti “a” ilə müqayisə et.
4. Əgər “a” olarsa, keç 7, olmazsa keç 4
5. Birinci komponenti “an” ilə müqayisə et.
6. Əgər “an” olarsa, keç 7, olmazsa keç 11

7. İkinci komponenti götür və onun qarşılığını avtomatik lüğətdən tap.
8. İkinci komponentin qarşılığını birinci mövqedə yaz.
9. Üçüncü komponenti götür və onun qarşılığını tap.
10. Üçüncü komponentin qarşılığını ikinci mövqedə yaz.
11. Birinci komponentin qarşılığını avtomatik lüğətdən tap
12. Birinci komponentin qarşılığını birinci mövqedə yaz.
13. İkinci komponenti “shall” sözünə görə yoxla.
14. Əgər “shall” olarsa, onda keç 15, olmazca keç 17
15. Üçüncü komponentin qarşılığını morfoloji təhlili nəzərə almaqla tap
16. Üçüncü komponentin qarşılığını ikinci mövqedə yaz.
17. Cümlənin bitməsini yoxla.
18. Əgər bitibsə keç 1.

Qeyd etmək lazımdır ki, ingilis dilində modal “will” feili də ikinci mövqedə və üçüncü mövqedə işləyə bilər. Eyni zamanda bu modal feilin keçmiş zaman forması “would” da eyni mövqələrdə işləyə bilər. Deməli, alqoritm normal işləməsi üçün bu komponentlər üzrə də yoxlama tələb olunur. Buradan belə nəticə çıxır ki, alqoritmə ya əlavə addımlar daxil edilməli, ya da iki komponentdən ibarət qrammatik formalar üzrə yoxlamalar üçün əlavə blok formalaşdırılmalıdır. Avtomatik tərcümə prosesinin sadələşdirilməsi və təkmilləşdirilməsi məqsədilə iki istiqamətdə təhlilin aparılması sınaqdan çıxarılmalıdır. Birinci istiqamət cümlədə sözlərin və ya komponentlərin sayına əsasən alqoritm qurmaqdırsa, ikinci istiqamət mümkün formaların hər birini nəzərə alan ümumi alqoritm tərtibidir. Mətnin cümlələrə avtomatik seqmentasiyası ayrıca alqoritm əsasında reallaşırsa, fikrimizcə, cümlədə sözlərin sayına istinad etməklə avtomatik tərcüməni həyata keçirmək daha sadə təsir bağışlayır. Bunun üçün ingilis dilində üç komponentdən ibarət bütün mümkün cümlə formaları müəyyənləşdirilməli və onların avtomatik tərcüməsi üçün müvafiq alqoritm qurulmalıdır.

Mətnin cümlələrə seqmentasiyası üzrə apardığımız araşdırmalar göstərir ki, cümlənin sonu “nöqtə” [.] , “üç nöqtə” [...], “nida işarəsi” [!], sual işarəsi [?], nida+sual işarəsi [!?] , sual+ nida işarəsi [?!], nida+ iki nöqtə [!..], qoşa nöqtə [:] ilə işarələnə bilər. Təhlil göstərir ki, cümlənin sonunun işarələnməsində istifadə olunan əsas simvollar qeyd edilənlərdir. Mətnə cümlələri ayırma alqoritmni bu simvollarla istinadla aşağıdakı kimi qurmaq mümkündür:

I addım: mətnin birinci simvolunu götürmək və onun “abzas” işarəsi olmasını yoxla. Əgər abzas işarəsidirsə keç ikinci addıma, abzas işarəsi deyilsə, keç III addıma.

II addım: Mətndəki sonrakı simvoldan başlayaraq [·], [?], [...], [!], [!·], [?!], [!..], [·] işarələrindən hər hansı birinə qədərki hissəni ayır və bu hissəni cümlə blokuna daxil et.

III addım: Əgər işarə [·], [?], [...], [!], [!·], [?!], [!..], [·] işarələrindən hər hansı biridirsə ondan sonrakı işarənin probel olmasını yoxla.

IV addım: Əgər probeldirsə, keç V addıma, probel deyilsə, keç VI addım addım.

V addım: mətndəki sonrakı simvoldan başlayaraq [·], [?], [...], [!], [!·], [?!], [!..], [·], [abzas] işarələrindən hər hansı birinə qədər olan hissəni ayır və bu hissəni cümlə blokuna daxil et.

VI addım: Mətnin sonrakı işarəsini yoxla, əgər abzasdırsa, keç I addım, abzas deyilsə keç VII addım.

VII addım: Qalan hissədə birinci simvol [·], [?], [...], [!], [!·], [?!], [!..], [·] hər hansı biridirsə, sonrakı işarənin probel olmasını yoxla.

VIII addım: əgər probeldirsə, mətndəki sonrakı simvoldan başlayaraq [·], [?], [...], [!], [!·], [?!], [!..], [·], [abzas] işarələrindən hər hansı birinə qədər olan hissəni ayır və bu hissəni cümlə blokuna daxil et.

IX addım: Əgər probel deyilsə, mətnin sonunu göstərən işarəyə görə yoxla

X addım: Əgər mətnin sonunu göstərən işarədirsə işi dayandır. Əgər deyilsə, keç VI.

Verilmiş alqoritm mətnin cümlələrə seqmentasiyası məsələsini ümumən həll edir. Belə alqoritmın reallaşmasından sonra cümlənin komponentlər sayını da avtomatik təyin etmək mümkündür. Bundan sonra iki və üç komponentli cümlələrin obyekt dildəki söz sırasını müəyyənləşdirmək üçün alqoritmi təkmilləşdirmək olar.

ƏDƏBİYYAT:

1. Ахманова О.С. Словарь лингвистических терминов. –Москва: Советская энциклопедия, 1966. -606с.
2. Budaqova Z. Adlıq cümlə //Müasir Azərbaycan dili III h. -Bakı: Elm, 1980. - S.275-284
3. Лайонз Дж. Введение в теоретическую лингвистику. – Москва: Прогресс, 1978. -543с.

Натаван Азизли

**Автоматическое построение слов в предложении при
машинном переводе на целевой язык**

Резюме

В статье решаются вопросы автоматизации построения слов в целевом языке на основе группирования компонентов по их числу. С этой целью проводится автоматическая сегментация предложений текста, затем определяется число компонентов предложений. Так как двухкомпонентные предложения английского языка передаются на азербайджанском языке в таком же порядке, то появляется необходимость формирования алгоритма в целевом языке для компонентов трехкомпонентных предложений.

Natavan Azizli

**Automatic compartmentalization of word order in automatic
translation into a target language**

Resume

In the article, solution of automation of word ordering in a target language is considered on the base of grouping of sentences in accordance with number of components. With this aim, automatic segmentation of a text into sentences is carried out and then number of components of the sentences is defined. As two-component English sentences being translated into Azerbaijani language with the same word order, need is revealed for formation of algorithm of triple component sentences in a target language.