

İNFORMASIYA TEXNOLOGİYALARI VƏ LİNGVİSTİKA

Açar sözlər: riyazi lingvistikə, informasiya texnologiyaları, mətnin emalı, bilingva

Ключевые слова: математическая лексика, информационные технологии, производство текста, билингва

Key words: mathematical linguistics, information technology, text processing, bilingual

Dilçilik, dil, onun ictimai təbiəti və funksiyaları, daxili quruluşu, funksiyalarının qanunauyğunluqları, tarixi inkişafı və konkret dillərin təsnifatı haqqında elm lingvistikə adlanır. Elmi lingvistikə nəzəri, tətbiqi və praktik hissələrə bölünür. Nəzəri lingvistikə dil qanunlarını nəzəri cəhətdən tədqiq edir. Real nitqi təsvir edən empirik, deyiliş və yazılış göstərən normativ, bütün dillərin ümumi xüsusiyyətlərini öyrənən ümumi, ayrıca dil və dil qrupunu öyrənən xüsusi lingvistikaya bölünür. Tətbiqi lingvistikanın əsas məsələləri tərcümə, lingvodidaktika (xarici dili öyrənmə metodikası), interlingvistikə (süni dillərin yaradılması), informasiya lingvistikası (informasiyanın axtarılması, təsnifatı, unifikasiyası və s.)

Hazırda informasiya texnologiyalarının inkişafı ilə tətbiqi lingvistikə yeni avtomatlaşdırılmış istiqamətlərdə inkişaf edir:

- riyazi lingvistikə (dilin formal modellərinin işlənilib hazırlanması),
- kompüter lingvistikası (maşın mühitində bütün ənənəvi lingvistikə məsələləri),
- kompüter lingvodidaktikası,
- maşın tərcüməsi, lüğətlər,
- təbii dillərin avtomatik emalı bir nitqin tanınması və sintezi, informasiya işlərinin avtomatlaşdırılması, informasiya axtarışının avtomatik sistemləri,
- koqnitiv lingvistikə (mətnin tezlik analizi)

Tətbiqi lingvistikanın fundamental tətbiqinin əsas korpusu mətni fonetik, qrammatik, semantik strukturunun təsvirini və modelini verir. Dilçiliyin təsnifatı üsulları, məntiq, riyaziyyat və semiotika əsasında yeni struktur lingvistikası yarandı. Dilçiliklə riyaziyyatın əlaqəsindən riyazi lingvistikə yarandı. Bu elmin yaranması dilin strukturunun yaranması, dilə sistem kimi yanaşma, dil sisteminin mürəkkəbliyi və s. ilə əlaqədardır. Bu elmlərin əsas predmeti dilin təsvirinin analizinin struktur və formal üsullarını təkmilləşdirir.

Tətbiqi lingvistikanın əsas məsələləri süni intellekt sistemlərinin, avtomatik sistemlərin, mətnin yaradılması sistemlərinin, dil öyrətmə sistemlərin, şifahi nitqi başa düşən sistemlərin, avtomatlaşdırılmış informasiya axtarış sistemlərinin, mətnlərin şifrələnməsi sistemlərinin, müxtəlif verilənlər bazasının (lüğət, kataloq və s.), müxtəlif növ avtomatik lüğətlərin yaradılmasıdır.

Maşın tərcüməsi sistemi bir təbii dildə olan mətnin digər təbii dildə keyfiyyətli qaralama tərcüməsini almaq üçün yaradılan kompüter proqram sistemidir. Xarici dil bilməyən istifadəçilərə, hətta, aşağı keyfiyyətli maşın tərcüməsi belə çox əhəmiyyətli kömək göstərə bilər. Mütləq stilistik dəqiqliyin tələb olunmadığı və vaxt məhdudluğu başlıca meyar olduğu hallarda – belə hallar isə kompüter istifadəçiləri üçün böyük əksəriyyət təşkil edir (xüsusilə İnternetdən istifadə edərkən, məsələn, tapılmış mətnlərin sorğulara relevantliyini qiymətləndirərkən) – maşın tərcüməsi sistemləri istifadəçilər üçün əvəzedilməz köməkçiyə çevrilir.

Bu məsələlər aşağıdakı problemlərlə bağlıdır: lüğətlərin yaradılması, sözün morfoloji analizi, çoxmənalı sözlərin mənasının təyin olunması, cümlənin sintaktik analizi, lüğətdə sözün axtarışı və cümlələrin yaradılması.

Kompüter lingvistikası tətbiqi lingvistikanın bir qoludur və dil haqqında elmin məsələsini həll etmək üçün istifadə olunur. Kompüter lingvistikası elmi istiqamət kimi 1960-cı illərdə formalaşmış. Kompüter lingvistikası ilə həm dilin modelini yaradan filoloq-lingvist, həm də bu model əsasında riyaziyyatçı-lingvist məşğul olur. Lingvistin maraq sahəsi nəzəriyyə deyil, bu nəzəriyyənin tətbiqinin nəticəsidir. Kompüter lingvistikası dilin kompüter modelləşməsi, struktur modelləşdirmə, mətni hipermətn texnologiyaları, maşın tərcüməsi, kompüter leksikoqrafiyası kimi istiqamətləri var. Kompüter lingvistikası aşağıdakı istiqamətlərdə inkişaf edir: maşın tərcüməsi, simvolların avtomatik tanınması (OCR), nitqin avtomatik tanınması (ASR), verilənlərin avtomatik əldə edilməsi (Data Mining), biliklərin idarə edilməsi sistemlərin yaradılması, lüğətlərin yaradılması, korpus lingvistikası və s. Təbii dilin ənənəvi emal üsullarından kompüter lingvistikası onunla fərqlənir ki, ənənəvi yanaşmada lingvistikanın tədqiqat sahəsinin modelləşdirilməsinə baxılır, kompüter lingvistikasında isə həm nəzəri lingvistikanın korrekt olunmasına həm də təklif olunan modellərin adekvatlığına diqqət yetirilir.

Riyazi lingvistikə süni intellektin bir istiqamətidir. Süni intellekt yeni informasiya texnologiyalarının əsasını təşkil edir. Süni intellekt intellektual kompüter proqramları, intellektual maşınların yaranması elmi və texnologiyasıdır. Kompüter lingvistikası qismən təbii dillərin emalı ilə də məşğul olur. Kompüter lingvistikasının fəaliyyət sahəsi dil informasiyasının emalı üçün tətbiqi proqramların və alqoritmlərin hazırlanmasıdır. Kompüter lingvistikasının aşağıdakı istiqamətləri var: təbii dillərin emalı, simvolların optik tanınması, nitqin avtomatik tanınması və nitqin avtomatik sintezi.

Lingvistik prosessorlar qrammatik, leksik və semantik prosessorlara bölünür. Bu prosessorların əsas vəzifəsi mətnin avtomatik yaxud avtomatlaşdırılmış emalıdır. Onlar mətnin emalının müxtəlif problemlərini həll edir, avtomatik

tərcümə edir, orfoqrafiyanı, qrammatikanı və stili korrektə edir, mətni kodlaşdırır və dekodlaşdırır, mətni yaradır və tanıyır, əlaqəli mətnləri analiz edir.

Linqvistikada informasiya texnologiyaları dedikdə kompüter vasitəsi ilə dil və onun qanunları haqqında informasiyanın alınması, saxlanması, ötürülməsi, yayılması, dəyişdirilməsi üçün qanun, metod və vasitələr toplusu nəzərdə tutulur.

Kompüter linqvistikası süni intellektin əsas problemi- biliklərin elektron təqdimatı problemi ilə sıx bağlıdır. Kompüter linqvistikasının əsas məsələsi məntiqi-linqvistik modellərin onlara uyğun alqoritm və proqramların qurulmasıdır.

Bilik–insan tərəfindən varlığın adekvat inikasının nəticəsidir (anlayış, nəzəriyyə, təqdimat şəklində). Kompüter sistemlərinə nəzərən bilik konkret sahənin təbii yaxud süni dildə adekvat təsviridir.

Linqvistikanın tətbiq sahələrindən biri də maşın tərcüməsidir. Bu problemə iki yanaşma var: deduktiv və induktiv. Deduktiv üsulda "mətn –məna –mətn" modelindən istifadə olunmaqla semantikanın formal təsviri verilir, induktiv üsulda "mətn-mətn" modelindən istifadə olunmaqla uyğunluğa baxılır.

Birinci yanaşmada tam avtomatlaşdırılmış dilin semantikasından istifadə olunmaqla yüksək səviyyəli maşın tərcüməsi almağa cəhd edilir.

Süni intellektin mühüm istiqamətlərindən biri olan kompüter linqvistikası sahəsində Azərbaycan dilinin terminoloji verilənlər bankının arxitekturası işlənmiş, istifadəçi və informasiya sistemləri üçün nəzərdə tutulmuş xüsusi təyinatlı izahlı, terminoloji, avtomatik lüğətlərin, informasiya-axtarış tezauruslarının tərtibi, təhlili və korreksiyası modelləri və metodları təklif edilmiş, mətnin avtomatlaşdırılması emalı və maşın tərcüməsi sistemində Azərbaycan dilinin formal modellərinin qurulması və linqvistik prosessorun yaradılması üçün biliklər bazasına söykənən məntiqi üsullar, müxtəlif dil səviyyələrində (morfoloji, sintaktik, semantik) ekspert sisteminin sintezi metodları, süni intellekt texnologiyasının sosial-iqtisadi problemlərin həllinə tətbiqi çərçivəsində demoqrafik proqnoz, əmək bazarında tələb və təklifin idarə olunması, ailə gəlirlərinin qiymətləndirilməsi və digər zəif strukturlu məsələlərinin həlli üçün fəzi məntiq üsullarına əsaslanan orijinal yanaşmalar, metod və modellər təklif edilmişdir.

Linqvistik məsələlərdə bütün verilənlər məsələnin şərtində olur, onların həlli üçün xüsusi hazırlığa və biliyə ehtiyac olmur. Linqvistik məsələni həll etmək üçün həll edən dilçilik intuisiyası və məntiqi düşünmə qabiliyyəti olmalıdır. Ən çox yayılmış linqvistik məsələlərdən biri bilinmə adlanır. Bilinmə tanış olmayan dildən doğma dila yaxud əksinə nümunə əsasında tərcümə etməyi tələb edir.

Məsələ. Suaxili dilində bəzi cümlələr və onların tərcüməsi verilib:

Akupenda – O səni sevir.

Avapiqa – O onları döyür.

Nikupiqa – Mən səni döyürəm.

Avupenda – O bizi sevir.

Suaxili dilində necə deyərsən "Mən onları sevirəm"?

Həlli: Şərtə verilən cümlələri araşdırsaq, görərik ki, a -o, ku- səni, va- onları, ni –mən, piqa – döymək, penda – sevmək kimi tərcümə oluna bilər. Onda "Mən onları sevirəm"- "Nivapenda" olar.

Ədəbiyyat:

1. Баранов А.Н. Введение в прикладную лингвистику, 1973
2. Вейжедл Р.М. Представление знаний и обработка естественных. языков. М., 1986.
3. Лайонз Дж. Введение в теоретическую лингвистику. М., 1979.
4. Попов Э.В. Общение с ЭВМ на естественном языке. М., 1989.
5. Пospelов Г.С. Искусственный интеллект - основа новой информационной технологии. М., 1988.

Information technologies and linguistics

Summary

The article discusses the possibilities of computer technology in linguistics. The new concepts as linguodidactics, interlingualistics and information linguistics are explained. Examples of machine translation programs, automatic character and speech recognition programs, knowledge management programs, etc. are given.

Информационные технологии и лингвистика

Резюме

В статье рассматриваются возможности компьютерных технологий в лингвистике. Появляются новые понятия как лингводидактика, интерлингвистика и информационная лингвистика. Приводятся примеры программ машинного перевода, программ автоматического распознавания символов и речи, программ управления знаниями и т. д.

Rəyçi: dos. X.Novruzova