UDK: 519.711

32.973

R33

Rahimov A. Kompyuter lingvistikasi asoslari / A. Rahimov; mas’ul muharrir

A. Nurmonov; 0‘zR Oliy va o‘rta maxsus ta’lim vazirligi, Andijon davlat universiteti.

- Т.: Akademnashr, 2011. - 160 b.

BBK 32.973я73+81.2-4я73

Mazkur go ‘llanma oliy о ‘quv yurtlarining 5220100-filologiya (o'zbekfilologiyasi) ta’lim yo'nalishi 1-bosqich bakalavr talabalari, magistrantlar, tadqiqotchilar va

о ‘qituvchilarga mo 'Ijallangan. XJnda kompyuter lingvistikasining asosiy masalalari tahlil va talqin etilgan, shuningdek, kitob so ‘ngida kompyuter lingvistikasining tayanch terminlari lug'ati (glossariy) keltirilgan.

Данное пособие предназначено для студентов (5220100-филология (узбеская филология), магистрантов, исследователей и преподавателей ВУЗов. В этой книге обсуждены и интерпретированы основные вопросы компьютерной лингвистики, а также в конце книги приложен словарь (глоссарий) ключевых терминов по компьютерной лингвистики.

This manual (handbook) is referred to the first year bachelor-students (5220100- philology (Uzbek philology), masters, researchers and techers of universities. In this book basic questions of computational linguistics are analyzed and interpreted, as well as the dictionary (glossary) of key terms on computational linguistics is attached at the end of the book.

Mas’ul muharrir: A.Nurmonov, filologiya fanlari doktori, professor

Taqrizchilar: Sh.Shahobiddinova, filologiya fanlari doktori, dotsent,

D .Nabiyeva, filologiya fanlari doktori, dotsent

Andijon davlat universiteti Ilmiy kengashining 2010-yil 26-noyabrdagi 3-sonli yig'ilishi qarori bilan nashrga tavsiya etilgan.

ISBN 978-9943-373-92-1

© A.Rahimov «Kompyuter lingvistikasi asoslari». «Akademnashr» nashriyoti, 2011-yil.

Kompyuter lingvistikasi asoslari

3

SQ‘ZBOSHI

Jamiyatimizda hozirgi kunda avtomatlashgan informatsion texnologiyalar va kompyuter muhim ahamiyat kasb etmoqda. Kompyuter kundan kunga jamiyatning barcha sohalariga kirib bormoqda. Bu esa yangi yo‘nalishlarning yuzaga kelishiga sabab bolmoqda. Xususan, kompyuter texnologiyalarining tilshunoslik fani bilan ijobiy hamkorligi natijasida kompyuter lingvistikasi fani shakllandi. Bu fan XX asrning 50 - 60-yillarida yuzaga kelgan bo‘lib, u turlicha nomlar bilan atalgan: hisoblash lingvistikasi, matematik lingvistilca, kvantitativ lingvistika, injener lingvistikasi kabi. 0‘tgan yarim asrlik vaqt mobaynida kompyuter lingvistikasi sohasida bir qator ilmiy va amaliy nati- jalarga erishilgan: tabiiy tilda avtomatik tarjima tizimi yaratildi, matndagi ma’lumotlarning avtomatik qidiruv tizimi ishlab chiqildi, og‘zaki nutqning avtomatik analizi va sintezi tizimi yaratildi, bir qator lingvistik muammolar- ni hal etuvchi kompyuter dasturlari ishlab chiqildi, inson va mashina (komp­yuter) muloqoti optimallashtirildi, tabiiy tilni qayta ishlash (Natural Lan­guage Processing) tizimi shakllantirildi. Bu yo'nalishlarda sezilarli yutuqlar qo‘lga kiritilgan boisa-da, lekin sanalgan masalalar hali toiiq o‘z yechimini topgani yo‘q. Chunki inson tili va tafakkuri sirli mohiyatga ega, uning biz anglamagan qirralari juda ko‘p. Bu esa kompyuter lingvistikasi doirasida to‘xtovsiz izlanishlar olib borishni taqozo etadi.

Ma’lumki, kompyuter lingvistikasi fani doirasida xorijda ingliz, rus, ne- mis, fransuz, ispan tillarida yaratilgan monografiyalar, ilmiy jurnallar, dars- liklar miqdori ko‘p. 0‘zbek tilida kompyuter lingvistikasi bo‘yicha izlanishlarni kuchaytirish bugungi o‘zbek tilshunosligi oldidagi eng dolzarb vazifalardan biridir. Shu ma’noda ushbu kitobda kompyuter lingvistikasining asosiy masala- lari tahlilga tortildi. Muallif ushbu kitobni yozishda D.Jurafskiy, J.H.Martinlarning «Speech and Language Processing», R.Grishmanning «Computational linguistics», «Новое в зарубежной лингвистике» jurnali- ning 24-soni (Компьютерная лингвистика), G.G.Belonogovning «Компьютерная лингвистика и перспективные информационные технологии», Yu.I.Shemakinning «Начало компьютерной лингвистики»,

A.Nurmonovning «Struktur tilshunoslik: ildizlari va yo'nalishlari»,

A.Poiatov, S.Muhamedovalarning «Kompyuter lingvistikasi» kitoblari va Internet saytlaridan foydalanildi.

Mazkur o‘quv qo‘llanma kompyuter lingvistikasi asoslarini tashkil etuv­chi masalalarga bag‘ishlangan, shuningdek, unda kompyuter lingvistikasi­ning tayanch tushunchalari lug‘ati (glossariy) ham keltirilgan.

A.Nurmonov, filologiya fanlari doktori, professor

4

A.Rahimov

KOMPYUTER LINGVISTIKASIFANINING MAZMUN-MOHIY ATI

Reja:

1. Kompyuter lingvistikasi fanining yuzaga kelish asoslari.

2. Kompyuter lingvistikasi va matematik lingvistika.

3. Kompyuter lingvistikasi fanining tekshirish obyekti, maqsad va vazi- falari.

4. Kompyuter lingvistikasi va klassik tilshunoslik.

Til sistema sifatida aynan matematikaning о 'zidir.

Matematika anicj fan bo‘lib, formulalar munosabati orqali reallashadi. Til ham matematika singari aniqlikni talab qiladi va unda ikki yoki undan ortiq elementlarning

о ‘zaro munosabati bir butunlikni tashkil etadi.

Beryozin

Tayanch so‘z va iboralar: til algebrasi, matematik lingvistika, kompyuter lingvistikasi, mashina tarjimasi, metatil, sun ’iy til, deskriptiv, kvantitativ, muloqotning optimal modeli, analiz, sintez, NLP, injener lingvistikasi, gene- ratsiya, tilshunoslik va adabiyotshunoslik sohalarida kompyuterdan unumli foydalanish, tillarni o'qitish, bilimlarini baholash, matnlarni tahrir qilish, tabiiy tillarning matematik modeli, о ‘zbek tilining kompyuter uslubi, matn- lar korpusi, elektron lug‘at, tezaurus.

Ma’lumki, tilshunoslik fani XIX (1816-yilda) asrda mustaqil fan sifatida shakllandi. Shundan boshlab u turli aspektlarda, yo'nalishlarda rivojlanib kelmoqda. Keyingi yillarda barcha fanlardagi kabi tilshunoslikda ham ikki fanning «chorrahasida» (kesishuvida) yuzaga kelgan fanlar jadal rivojlan- moqda. Jumladan, ana shunday fanlar sirasiga sotsiolingvistika (sotsiologiya va tilshunoslik), psixolingvistika (psixologiya va tilshunoslik), etnolingvis- tika (etnografiya va lingvistika), neyrolingvistika (nevrologiya va tilshun­oslik), matematik lingvistika va kompyuter lingvistikasi fanlarini kiritish mumkin. Bunday holni boshqa fanlar doirasida ham kuzatish mumkin: biokimyo, astrofizika, matematik fizika, matematik logika kabi. Buni fan­lar tizimida bir necha fanlarning о‘zaro hamkorligi, integratsiyasi deb ba­holash lozim bo‘ladi. XX asrning 50-yillaridan boshlab tilshunoslikda

Kompyuter lingvistikasi usoslari

5

«mashina tarjimasi», «mashina tilshunosligi» atamalari qoilanila boshlan- di.1 Mazkur asrning buyuk kashfiyoti bo‘lgan kompyuter texnologiyalari tilshunoslikka ham kirib kclganining isboti edi.2 Mashina tarjimasi yoki avtomatik tarjima deyilganda bir tildagi matnni ikkinchi bir tilga EHM (kompyuter) vositasida, tez vaqt ichida tarjima qilish nazarda tutiladi. Mashi­na tarjimasining asoschilari kibernetika va matematika sohasi vakillari bo'lib, keyinchalik bu ishda tilshunoslar ham faol qatnasha boshlagan. Shu tariqa mashina tarjimasi g‘oyalari butun dunyoda nazariy va amaliy tilshu- noslikning rivojlanishida katta ahamiyat kasb etdi. Bu yo‘nalish bilan pa­rallel ravishda formal grammatika nazariyasi yuzaga kelib, til va uning alohida aspektlari modelini yaratishga e’tibor qaratildi. Tilning bu jihatlari matematik lingvistika fanida ishlab chiqildi, bu, o‘z navbatida, kompyuter lingvistikasi fanining yuzaga kelishi uchun poydevor bo‘ldi. Demak, shu asosda Lilshunoslikning yangi yo'nalishi - kompyuter lingvistikasi va tilshu- noslikning bir qator nazariy va amaliy yo‘nalishlari vujudga keldi.

Matematik lingvistika fani XX asrning 50-yillarida (1952-yilda) tilshu- noslikning alohida yo‘nalishi sifatida yuzaga keldi. Bu fanning shaklla- nishida Kopengagen struktural tilshunoslik maktabi (glossematika)ning asos- chisi Lui Yelmslevning g‘oyalari o‘ziga xos «turtki» vazifasini o‘tagan. U hatto til hodisalarini matematik bayonda tushuntiradigan fanning nomini ham taklif etgan. Olimning fikricha, bu fan «Til algebrasi» («IJngvistik algebra») deb atalishi lozim edi.3 Amerikalik tilshunos Noam Chomskiy- ning formal grammatika, transformatsion grammatika haqidagi qarashlari bevosita matematik lingvistilcaning alohida yo‘nalish sifatida yuzaga kelishi- ga sabab bo'lgan.4 Mana shunday qarashlar ta’sirida matematik lingvistika fani shakllandi. Matematik lingvistika - bu tabiiy tillarning matematik modellarini (bunday formallashgan til metatil deb ataladi) ishlab chiqish, xususan, sun’iy tillarni yaratish algoritmini tuzish bilan shug‘ullanuvehi fandir. Matematik lingvistika oldida turuvchi eng muhim masalalar quyi- dagilardir:

- tilning aksiomatik nazariyasini ishlab chiqish;

1 Grishman R. Computational linguistics // Cambridge University Press. 1994. -P.4.

2 Po'latov A., Muhamedova S. Kompyuter lingvistikasi. - Toshkent, 2007. - B.4-5.

3 Нурмонов А. Структур тилшунослик: илдизлари ва йуналишлари. - Тошкент, 2008. - Б.145.

4 Гладкий А.В., Мельчук И.А. Элементы математической лингвистики. - М.: Наука, 1969. -С.7.

6

A. Rahimov

- formal grammatika yaratish;

- tillarning matematik modellarini ishlab chiqish.

Har bir fanning o‘z maqsad va vazifalari boiadi. Matematik lingvistika fanining asosiy maqsadi tabiiy tillarning matematik modelini ishlab chi- qishdir. Ushbu maqsadga erishish uchun fan o‘z oldiga quyidagi vazifalarni qo‘yadi:

- tabiiy va sun’iy tillarning formal modellari algoritmini ishlab chiqish;

- lisoniy hodisalami matematik parametrlarda baholash;

- til hodisalarini matematik metodlar yordamida tahlil qilish (ehtimol- lar nazariyasi, statistika va kvantitativ metodlarni tatbiq etish).

Kompyuter lingvistikasi matematik lingvistikaning mantiqiy davomi bo‘lib, u amaliy tilshunoslikning eng muhim qismini tashkil etadi. Komp­yuter lingvistikasi 1954-yil AQSHda Jorjtaun universitetida mashina tarji- masi bo‘yicha dunyoda o'tkazilgan birinchi tajriba asnosida yo'nalish si- fatida shakllana boshladi, 1960-yilga kelib mustaqil fan sifatida shakllandi. Kompyuter lingvistikasi inglizcha «computational linguistics» so'zining kalkasidir. XX asrning 80-yillariga qadar bu fan turlicha nomlar bilan atal- gan: hisoblash lingvistikasi, matematik lingvistika, kvantitativ lingvistika, injener lingvistikasi kabi. Bu fanning asosiy maqsadi lingvistik masalalarni yechishning kompyuter dasturlarini ishlab chiqish, inson va mashina (komp­yuter) muloqotini optimallashtirishs, tabiiy tilni qayta ishlash (Natural Lan­guage Processing)dir.(i NLP kompyuter lingvistikasida tabiiy tillarning komp­yuter analizi va sintezini o‘z ichiga oladi. Bunda analiz tabiiy tilning komp- yuterda morfologik, sintaktik va semantik tahlil yordamida tushunilishiga nisbatan ishlatiladi, sintez esa kompyuterda matnning grammatik shakllan- tirilishi va generatsiyasi (hosil qilinishi) demakdir. NLP bo‘yicha yaratil- gan dasturiy ta’minotlar quyidagilar: AlchemyAPI, Expert System S.p.A., General Architecture for Text Engineering (GA ТЕ), Modular Audio Recog­nition Framework, Monty Lingua, Natural Language Toolkit (NLTK).

Kompyuter lingvistikasining asosiy vazifalariga quyidagilar kiradi:

- avtomatik o‘qitish tizimini ishlab chiqish;

- bilimlarni tekshirish;

- matnlarni turli jihatdan avtomatik tahrirlash;

- matnlarning avtomatik tarzda morfologik, sintaktik va semantik tah- lilini (inglizcha parsing) ta’minlovchi tizimlar yaratish (parser dasturlar);

5 Новое в зарубежной лингвистике. Вып.24. Компьютерная лингвистика. -М.: Прогресс, 1989.-С. 10.

6 Jurafsky D., Martin J.H. Speech and Language Processing. - New Jersey, 2000.-P.2-3.

Kompyuter lingvislikasi asoslari

7

- mashina tarjimasi uchun moMjallangan dasturlarni ishlab chiqish;

- lug‘atlarni va kompyuterdagi matnni statistik tahlil qilish;

- lingvistik muammolarni hal qilishga yo‘naltirilgan optimal dasturlar yaratish;

- muloqotning kompyuter modelini ishlab chiqish;

- matn strukturasining gipertekst texnologiyasini yaratish;

- elektron lug'atlar va tezauruslar yaratish;

- sun’iy intellekt tizimini ishlab chiqish;

- matnlar korpusini yaratish;

- matnni avtomatik referatlash;

- sujet strukturasini modeilashtirish;

- nutqni avtomatik tushunish tizimini ishlab chiqish (inglizcha ASR - Automatic Speech Recognition);

- matndan faktlarni avtomatik ajratib olish (inglizcha fact extraction, text mining).

Kompyuter lingvistikasida qoMlaniladigan va yaratiladigan lingvistik vositalarni shartli ravishda ilcki qismga bo‘lish mumkin: deklarativ hamda protsedura qismlari. Deklarativ qismga til va nutq birliklari lug'ati, gram- matik ma’lumotnomalar, matnlar korpusi kabilarni kiritish mumkin. Prot­sedura qismi esa yuqoridagi lingvistik ta’minot bazasini boshqarish vosi- talarini (algoritmlar tuzish, dasturlar yaratish, kompyuter analizi va sintezi kabilar) o‘z ichiga oladi.

0‘zbek tili materiallari bo‘yicha kompyuter lingvistikasiga oid tadqiqot olib borgan olimlar sifatida H.Arziqulov, S.Rizayev, S.Muhamedov,

A.Poiatov, S.Muhamedova, N.Jo‘rayevalarni ko'rsatish mumkin. Olimlar asosan statistik tahlil, algoritmlash, o‘zbek tilining aksiomatik nazariyasi, fe’llarning kompyuter analizi va sintezi kabi yo‘nalishlar bo‘yicha izlanishlar- ni amalga oshirganlar. 0‘zbek tilshunosligidagi kompyuter lingvistikasining boshqa yo'nalishlari esa o‘z tadqiqotchilarini kutmoqda. Shunga ko‘ra, bu- gungi kunda o‘zbek tilshunosligining kompyuter bilan bog‘liq holda hal eti- lishi lozim bolgan quyidagi vazifalarni ta’kidlab ko'rsatish mumkin:7

1. O'zbek tilining kompyuter uslubini yaratish.

2. Axborot matnlaridagi qoliplilik, qisqalik standartlarini ishlab chi­qish.

3. Kompyuter izohli va tarjima lug‘atlarini yaratish.

4. O'zbek tili va adabiyoti darsliklarining elektron versiyalarini ishlab chiqish.

7 Po'latov A., Muhametlova S. Kompyuter lingvistikasi. - Toshkent, 2007. - B.5-6.

8

A. Rahimov

5. Kompyuterda inglizcha-o'zbekcha tarjima dasturlarini ishlab chiqish.

6. Kompyuterdagi matnlarni avtomatik tahrirlash dasturlarini yaratish.

7. 0‘zbek tili grammatikasining kompyuter modelini yaratish.

8. «Inglizcha-o'zbekcha tarjimon» kompyuter dasturini ishlab chiqish.

9. 0‘zbek tilidagi matnlarni tahrirlash dasturlarini yaratish.

10. 0‘zbek tilidagi «Windows» operatsion tizimini yaratish.

11. Inglizcha-o'zbekcha, o'zbekcha-inglizcha kompyuter lug‘atlarini ishlab chiqish.

Kompyuter lingvistikasi amaliy tilshunoslikning tarkibiy qismi hisobla- nadi, u nazariy tilshunoslik yutuqlaridan foydalanadi, oziqlanadi. Komp­yuter lingvistikasi va nazariy tilshunoslik bir-birini to'ldiradi. Kompyuter lingvistikasida til insondan tashqarida mavhum sistema sifatida, L.Yelmslev ta’rifi bilan aytganda, «sof munosabatlar tizimi» sifatida tavsiflanadi. Ko­mpyuter lingvistikasi va klassik (mumtoz) tilshunoslik orasidagi farq qu- yidagi nuqtalarda ko'rinadi:

- mumtoz tilshunoslikda til inson bilan mutanosiblikda, juftlikda ko'rib chiqiladi. Ya’ni mumtoz tilshunoslik insonga yo'naltirilgan bo'ladi va uning faol ishtirokida qabul qilinadi. Kompyuter lingvistikasi esa tavsiflash jara- yonida insonni istisno qiladi va u ko‘proq kompyuterga moslashtiriladi;

- mumtoz tilshunoslik ko'proq tavsifiy (deskriptiv) xarakterga ega hisob- lanadi. Kompyuter lingvistikasi esa masalani miqdoriy xarakteristikalar va aniq parametrlar asosida hal etadi. Demak, mumtoz tilshunoslik ko‘proq tavsifiy bayonga asoslansa, kompyuter lingvistikasi miqdoriy (kvantitativ) tavsifga, algoritmlash, modellashtirish, statistik tahlilga asoslanadi;

- mumtoz tilshunoslik ko‘proq nazariy xarakterga ega bo‘lib, tilshunos- likning nazariy masalalari bilan shug'ullanadi. Kompyuter lingvistikasi esa amaliy xarakterga ega boiib, til bilan bogiiq muammolarning amaliy ji- hatlariga e’tibor qaratadi hamda uni aniq maqsadga yo'naltirilgan dasturlar, metodlar, tizimlar yordamida hal etish bilan shug‘ullanadi;

- mumtoz tilshunoslikning tahlil obyekti tabiiy til hisoblanadi va talilil- da uning mavjud barcha imkoniyatlari e’tiborga olinadi. Kompyuter lingvis­tikasida esa ko'proq sun’iy tillar (programmalashtirish tillari, algoritmik tillar)ga tayaniladi, tabiiy tillarning mavjud imkoniyatlari cheklanadi, bun- da tabiiy tilga ishlov berilib (NLP), kompyuterga moslashtiriladi.

Fanlar hamisha o'zaro uzviy bog‘liqlik va hamkorlikda ish ko'radi, ular bir-birisiz yashay olmaydi. Jumladan, kompyuter lingvistikasi fani ham bundan mustasno emas. Kompyuter lingvistikasi mantiq, informatika, statis- tika, semiotika, kibernetika, ehtimollar nazariyasi kabi turli fan sohalari bilan o‘zaro aloqadorlikda ish ko‘radi.

Kompyuter lingvistikasi asoslari

9

LINGVISTIKA FANI JENSKIY RODGA MANSUB, SHUNING UCHUN BU FANNISEVISH KERAK, SHUNDAG1NA UNI MUKAM- MAL O'RGANISH MUMKIN!

Topshiriqlar:

1. Matematik lingvistika va kompyuter lingvistikasi fanlarining amaliy ahamiyati nimadan iborat?

2. Tabiiy til va sun’iy til munosabatini qanday izohlaysiz?

3. Kompyuter lingvistikasining asosiy muammolari sifatida nimalarni ajratib ko‘rsatish mumkin?

4. Tilda aniqlikka erishish uchun nimalar qilish lozim deb o‘ylaysiz?

5. Kompyuter lingvistikasi fanining yuzaga kelish asoslarini tushunti- ring.

Adabiyotlar:

1. Шемакин Ю. И. Начало компьютерной лингвистики. - М.: Высшая школа, 1992.

2. Зубов А.В., Зубова И.И. Основы лингвистической информации. - Минск: МГПИИЯ, 1992.

3. Пиотровский Р.Г. Инженерная лингвистика и теория языка. - Л., 1979.

4. Нелюбим J1.J1. Компьютерная лингвистика и машинный перевод.

- М.: ВЦП, 1991.

5. Grishman R. Computational linguistics // Cambridge University Press, 1994.

6. Новое в зарубежной лингвистике. Вып.24. Компьютерная лингвистика. - М.: Прогресс, 1989.

7. Нурмонов А., Йулдошев Б. Тилшунослик ва табиий фанлар. - Тошкент: У^итувчи, 2001.

8. Po'latov A., Muhamedova S. Kompyuter lingvistikasi. - Toshkent, 2007.

9. Гладкий А.В., Мельчук И.А. Элементы математической лингвистики. - М.: Наука, 1969.

10. Jurafsky D., Martin J.H. Speech and Language Processing. - New Jersey, 2000.

10

А. Rahimov

MANTIQ, GRAMMATIKA VA ALGORITM Reja:

1. Mantiq - kompyuter lingvistikasining muhim tarkibiy qismi.

2. Mantiq va grammatika munosabati.

3. Semiotika va informatsiya nazariyasida tilning kod sifatidagi talqini.

4. Algoritm - dasturlash asosi.

Tavanch so‘z va iboralar: mantiq, grammatika, semiotika, algoritm, uchinchisi mustasno qonuni, oraliq uchinehi, matematik mantiq, dialektik mantiq, konyunksiya, dizyunksiya, implikatsiya, ekvivalensiya, kvantor, lingvistik protsessor, tabiiy tilli interfeys, semantika, sintaktika, semiotika, lingvosemiotika, pragmatilca, kod, kodlash, dekodlash, denotat, signifikat, algor it mik tillar, translyator dasturlar (kompilyatorlar).

Mantiq, grammatika, semiotika va algoritm kompyuter lingvistikasin­ing muhim unsurlari hisoblanadi. Ular til va nutqning funksionallashuv jara- yonini anglashda hamda lingvistik muammolarni avtomatik hal etish ti- zimini yaratishda poydevor boiib xizmat qiladi. Mantiq, grammatika, semi­otika va algoritm bir-biri bilan chambarchas bog'liq tushunchalar bo'lib, ular kompyuter lingvistikasida bir-birini to‘ldiradi, taqozo etadi hamda hamkorlikda ishlovchi qoidalar tizimini tashkil qiladi.

Tafakkur qonuniyatlarini, muhokama yuritish qonun-qoidalari, usullari va formalarini o‘rganuvchi fan mantiq sanaladi. Unda tushuncha, hukm, xulosa chiqarish kabi mantiqiy operatsiyalar o‘rganiladi. Mantiq dastlab Qadimgi Xitoy va Hindistonda miloddan avvalgi V asrda grammatika bi­lan uzviy bog'liq holda shakllana boshladi. Miloddan avvalgi IV asrda Qadimgi Yunonistonda mantiq mustaqil fan maqomiga ega bo‘ldi. Bu ho- disa yunon faylasufi Aristotel nomi bilan bog‘liq, u mantiq fani asoschisidir. Olim «Organon» (bu so‘z «qurol», «vosita» ma’nolarini anglatadi) asarini yozib, tafakkur qonuniyatlarini birinchi bo'lib tushuntirdi, deduksiya na- zariyasini - mantiqiy xulosa chiqarishning formal xarakterga ega ekanligi- ni ko‘rsatdi, mantiqni fikrlash quroli sifatida tavsifladi. Aristotel nominativ (otli) ifodalarning mantig‘ini ishlab chiqdi. Keyinroq Aristotelning shogirdi Teofrast va stoiklar maktabi vakillari jumla mantig‘ini yaratdilar.8

8 Белоногов Г.Г. Компьютерная лингвистика и перспективные информационные технологии. - М.: Русский мир, 2004. - С. 15.

Kompyuter lingvistikasi asoslari

11

Aristotel ta’limoti Farobiy, Beruniy, Gegel, Frege, Leybnis kabi buyuk olimlar tomonidan davom ettirildi. Shundan boshlab mantiq fani turli yo'nalishlarda rivojlana boshlanadi. Dastlab formal mantiq yuzaga keldi, unda hodisalar faqat formal asosda tushuntiriladi. Bu yo‘nalishda uchinchi- si mustasno qonuni ustuvorlik qiladi. Unga ko‘ra, bayon qilingan ikki zid fikrdan biri chin, boshqasi xato, lekin uchinchi holat boiishi mumkin emas. Bu qonun «А - В dir yoki В emasdir» formulasi orqali beriladi. Juda uzoq yillar ushbu qarash to‘g‘ri deb baholab kelindi.

XVIII - XIX asrlarda falsafada hodisalarni doimo rivojlanish va o'zaro ta’sirda o‘rganuvchi dialektik ta’limot paydo bo'lganidan so‘ng u mantiqqa ham ta’sir ko‘rsata boshladi. Natijada dialektik mantiq yo'nalishi yuzaga keldi. Bu yo‘nalishga nemis faylasufi Georg Vilgelm Fridrix Gegel asos soldi. U o‘z qarashlarini 4 jildli «Mantiq fani» nomli monografiyasida ba­yon qildi. Olim mantiqqa oraliq uchinchi qoidasini olib kirdi. Unga ko‘ra, hodisalarni baholashda ikkita bir-birini inkor etuvchi hodisa o'rtasida neyt- ral - «oraliq uchinchi» holat ham ajratiladi. Masalan:

|  |  |  |
| --- | --- | --- |
| issiq | iliq | sovuq |
| A | В | С |

Demak, dialektik mantiqqa asosan hodisalar ko‘p parametrlarda ba- holanadi. Keyinroq nemis olimi va matematigi Leybnis va Gilbertlar man­tiq faniga matematik simvollarni tatbiq etdi. Ushbu olimlarning buyuk xiz- matlari bilan matematik mantiq fani shakllandi.

Dastlab matematik mantiq faqat tabiiy fanlarda qo‘llanib kelindi, ke- yinchalik u gumanitar sohalarga ham tatbiq etilgan. Nemis faylasufi va matematigi G. Leybnis (1646 -- 1716) birinchilardan bo‘lib mantiqiy fikrlash- ning hisob xarakteriga ega ekanligini ko‘rsatdi. Uning fikricha, barcha il- miy tushuncha va mulohazalarning asosini mantiqiy elementlar tashkil qila­di. Mana shu mantiqiy elementlarni muayyan simvollar bilan belgilash mumkin boiadi. Leybnis g'oyalari faqat XIX asrdagina amalga oshdi. Ne­mis olimi G.Frege, ingliz olimi J.Bul, rus olimi V.Poretskiy ishlarida sim- volik mantiq asoslari yaratildi.9

Matematik mantiqda fikrlarning chin (to‘g‘ri) yoki yolg'onligi matema­tik yo‘l bilan isbotlanadi. Bunda x ni argument (o'zgaruvchi) deb belgilay- miz. Agar x = 1 bo‘lsa, mutlaq chin hukm, x = 0 bo'lsa, mutlaq yolg'on hukm deb belgilaymiz. Matematik mantiq turli simvollar bilan ishlaydi. Ular mantiqiy bog'lovchilar deb ham yuritiladi. Ushbu simvollar sun’iy

4 Po‘latov A., Muhamedova S. Kompyuter lingvistikasi. - Toshkent, 2007. -

B.10.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| A, В | A&B | AVB | A=>B | AsB |
| 00 | 0 | 0 | 1 | 1 |
| 01 | 0 | 1 | 1 | 0 |
| 10 | 0 | 1 | 0 | 0 |
| 1 1 | 1 | 1 | 1 | 1 |

12 A. Rahimov

programmalashtirish tillarida ishlatiladi. Shu o‘rinda biz ularni alohida ajra- tib ko'rsatishni lozim topdik:

1-jadval

1. A & В - konyunksiya. Bu o'zbek tilidagi «va» bog'lovchisiga to'g'ri keladi.

2. A V В - dizyunksiya. Bu o'zbek tilidagi «yoki» bog'lovchisiga to'g'ri keladi.

3. A => В - implikatsiya. Bu o'zbek tilidagi «agar... -sa, u holda ...» bog'lovchisiga to'g'ri keladi.

4. A = В - bu belgi mantiqiy teng kuchlilik (ekvivalensiya) uchun ishlatiladi.

5. AIВ - bu belgi «Sheffir tayoqchasi» deb nomlanadi, u «va ... emas» bog'lovchisiga to'g'ri keladi.

6. Ач1 В - bu belgi «Pirs strelkasi» deb nomlanadi, u «yoki ... emas» bog'lovchisiga to'g'ri keladi.

7. 3 = inglizcha exists so'zininq qisqartmasi boiib, mavjudlik kvantori sifatida

ishlatiladi.

8. ^ = summa, barcha narsaning jami.

9. V = inglizcha al| so'zining qisqartmasi bo'lib, hammasi uchun e kvantori sifatida ishlatiladi.

10. x, y, z - predmetlaming borligini belgilash uchun ishlatiladi.

11. X, Y, Z - predmetlaming sifatini (atributlarini) belgilash uchun ishlatiladi.

12. S-subyekt

13. P-predikat

14. A = 0 - mutlaqo noto'g'ri

15. A = 1 — mutlaqo to'g'ri

16. A - emas (inkor). Bir mulohaza ikkinchi bir mulohazaga zid kelsa, ishlatiladi.

Matematik mantiqda mulohazalarning chin yoki yolg'onligini formal nazariya asosida tekshirib ko‘rish mumkin. Bunda mantiqiy hukmlarni shartli ravishda simvolik belgiiar, mantiqiy bog'lovchilar yordamida belgilab ola- miz. U gapning propozitsional ko'rinishi deb nomlanadi. So‘ngjumiani for­mal nazariya asosida hisoblaymiz. Bunda bizga hukmlarning qiymatini ko'rsatuvchi jadval yordam beradi.

7-inHvn!

Kompyuter lingvistikasi asoslari

13

Formal mantiq va simvolik mantiq til strukturasini o‘rganishda amaliy ahamiyat kasb etadi, grammatikani formallashtirishda, sun’iy program- malashtirish tillarini yaratishda asos bo‘lib xizmat qiladi, kompyuter lingvistikasining o'ziga xos simvolik «tili» sifatida namoyon boiadi. Mantiq bizga kompyuter lingvistikasining murakkab va ko‘p aspektli muammosi hisoblangan tabiiy tilga ishlov berish (NLP) uchun optimal metodlarni taklif etadi. Jumladan, formallashtirish, analogiya, modellashtirish me- todlari kabi.

Antik davrda ham, o‘rta asrlarda ham grammatika va mantiq yonma- yon o‘qitilgan. Ayniqsa, o‘rta asrlarda grammatika va mantiq klassik ta’lim poydevori bo‘lgan «yetti erkin san’at» (artesliberales) tarkibiga kirgan. Bunda grammatika, ritorika (notiqlik san’ati) va mantiq «formal san’atlar» (artes- formates) sifatida trivia (uchlik)ga birlashgan. Arifmetika, geometriya, musiqa va astronomiya «real san’atlar» (artesreales) sifatida kvadrivia (to‘rtlik)ga birlshgan. Ushbu fanlarni o'zlashtirgan insongina yetuk va barka- mol hisoblangan.

Tabiiy tilning strukturasini o'rganishda mantiq, grammatika va semi- otika uzviy aloqadorlikda ish ko'radi. Mantiq va grammatika orasidagi aloqadorlikni til va tafakkur, so‘z va tushuncha, gap va hukm, ega va sub- yekt, kesim va predikat kabi kategoriyalarning dialektik munosabati yorda- mida yaqqol tasavvur etish mumkin.10 , . , .

|  |  |
| --- | --- |
| Mantiqiy kategoriyalar | Tilning grammatik unsurlari |
| Subyekt | Ega |
| Predikat | Kesim |
| Obyekt | To'ldiruvchi |
| Atribut | Aniqlovchi |
| Predmet, hodisa | Ot so'z turkumi |
| Jarayon (harakat, holat) | Fe’l so'z turkumi |
| Sifat | Sifat so‘z turkumi |
| Miqdor | Son so'z turkumi; son kategoriyasi |
| Bog'lovchilar; munosabatlar | Ko'makchilar, bog'lovchilar; kelishik kategoriyasi |

Til tafakkur bilan chambarchas bogiiq boiib, fikrni moddiylashtiradi- gan, kishilarning fikr almashishini ta’minlab beradigan quroldir. Til bilan

10 Лагута O.H. Логика и лингвистика. - Новосибирск, 2000. - С.7.

14

A. Rahimov

tafakkur bir-birini taqozo etadigan ajralmas hodisadir. Tilsiz tafakkur bo'lmaganidek, til ham tafakkursiz bo'lmaydi. Tafakkur inson oliy nerv sistemasining, ya’ni bosh miyaning faoliyati boiib, so‘z, so‘z birikmalari va gaplarda o‘z ifodasini topadi. Har qanday fikr so‘zlarda o‘z ifodasini topmaguncha, ya’ni so‘zlardan tarkib topgan ifoda qolipiga tushmaguncha yuzaga chiqmaydi. Fikrning voqeligi, uning mavjudligi tilda namoyon bo'ladi. Demak, fikr tilda voqelashadi, tilda mavjud bo‘ladi.

Tafakkurning reprezentatsiyasi til hisoblanadi. Ong amalda til formasi- da mavjud. Garchi til bilan tafakkur bir-biri bilan bevosita bog'liq ajralmas birlikni tashkil qilgan bo‘lsa-da, ammo ularni aynan bir hodisa deb tushu- nish noto‘g‘ri. Til bilan tafakkur murakkab, qarama-qarshiliklarga boy bo'lgan dialektik birlikni tashkil etadi. Til fonetika, grammatika qonunlari asosida tashkil topgan material hodisa boisa, tafakkur obyektiv borliqning inson miyasida aks etishining oliy formasi sifatida ideal hodisadir. Til qo­nunlari bilan tafakkur qonuniyatlarining tuzilishi va rivojlanishi bir xil emas. Tilning qurilish qonunlarini grammatika fani o‘rganadi. Tafakkur qonun­lari va birliklari esa mantiqda o‘rganiladi. Grammatik kategoriya va man- tiqiy kategoriya orasidagi munosabat ham juda murakkabdir. Grammatik kategoriyalarni mantiq, mantiqiy kategoriyalarni grammatik kategoriyalar- ga bo‘ysundirish yoki ularni qorishtirish mumkin emas. Bu kategoriyalar ba’zida mos kelishi, ba’zida mos kelmasligi ham mumkin.

Kompyuter lingvistikasida inson tafakkuri va tilini modellashtirish masalasi markaziy muammo hisoblanadi. Ya’ni sun’iy intellekt, tabiiy til- ga ishlov berish (NLP), lingvistik protsessorlar, kompyuter uchun tabiiy tilli interfeys bo‘yicha olib borilayotgan tadqiqotlar mantiq, grammatika, semi- otika, kompyuter texnologiyalari sohalarning ijodiy hamkorligi asosida amalga oshirilmoqda. Bu sohada juda ko‘p yutuqlarga erishildi, lekin hali- hanuz mo'jizakor tabiiy til ham, sirli inson tafakkuri ham to‘la modellash- tirilgani yo‘q. Ayni shu masala- til va tafakkurni modellashtirish kompyu­ter lingvistikasi fanining istiqbolli yo‘nalishi sanaladi.

Semiotika yoki semiologiya (yunoncha semiotike, ya’ni «belgilash», «tas- virlash», «ifodalash» degan ma’nolarni anglatadi) belgilar sistemasi ha- qidagi fan boiib, unda axborotni saqlovchi, uzatuvchi turli xil belgi va bel­gilar tizimlarining yaratilishi, vazifalarini, belgilar sistemasiga kiruvchi hodisalarni, ularning tabiati, mohiyati, turlari, ko‘rinishlari, amal qilishi kabilarni tekshiradi." Semiotika tilshunoslikda, xususan, kompyuter lingvis­tikasida ham muhim metodologik ahamiyat kasb etadi. Chunki til ham bel-

" Эко У. Отсутствующая структура. Введение в семиологию. - М.: Петрополис, 1998. - С.31.

Kompyuter lingvistikasi asoslari

15

gi sanaladi, semiotikada tabiiy tillar ham, sun’iy tillar ham, hatto yo‘l belgi- lari, audio va videosignallar ham tadqiq etiladi.

Semiotika fani XIX asr oxirlarida vujudga keldi, uning asoschisi amerikalik olim Charlz Sanders Pirs (1839 - 1914) hisoblanadi. Struk- turalizm asoschisi Ferdinand de Sossyur «Umumiy tilshunoslik kursi» asa- rida bu fanni «semiologiya» deb nomlagan. Ushbu fan yutuqlari hozirgi zamon kompyutershunosligida, telekommunikatsiya, uyali telefon tizim- lari, sun’iy yo‘ldoshlar, avtomatik boshqaruv tizimlari va boshqa sohalar- da keng foydalaniladi. Semiotikada belgilar tizimi to‘rt aspekt asosida o‘rganiladi: 1) sintaktika - belgilar strukturasini tekshiradi; 2) semantika

- belgilarning mazmuniy tomonini tekshiradi; 3) sigmatika - belgi va uni aks ettiruvchi obyekt o'rtasidagi munosabatni o‘rganadi; 4) pragmatika - belgilarning insonga bo‘lgan munosabati, foydaliligi, amaliy jihatlarini tadqiq etadi.

Semiotika belgilar haqidagi keng qamrovli nazariy-falsafiy fan bo‘lib, tilni belgilardan iborat butunlik sifatida o'rganadi. Tilshunoslik va semioti- kaning o‘zaro maqsadli aloqaga va munosabatga kirishuvidan lingvosemi- otika deb ataluvchi soha shakllandi. Lingvosemiotika tilning belgi tabiati- ga egaligini, informatsiya uzatuvchi til birliklarining xususiyatlarini tadqiq qiladi. Til birliklari, tilning tarkibini tashkil qiluvchi struktural elementlar

- morfema, so‘z, so‘z birikmasi, gap, matn va boshqalar muayyan infor- matsiyani berish, yetkazish - ijtimoiy vazifani bajarish uchun xizmat qiluv­chi belgilar sistemasi sifatida tadqiq qilinadi. Shunga ko‘ra til jamiyatda mavjud bo‘lgan boshqa belgilar bilan umumiylikka, o‘xshashlikka ega boiadi. Shunga ko‘ra, til semiotika fanining - lingvosemiotikaning o‘rganish obyektiga aylanadi. Tilning tabiatini, mohiyatini, jamiyatdagi o'rni va ro- lini bilish uchun uni faqat qaysidir til bilan qiyoslashning o'zigina yetarli bo'lmaydi. Tilni - til sistemasi tarkibidagi birliklarni, belgi deb olingan elementlarni boshqa belgilar - nolisoniy belgilar sistemasi bilan ham, jamiyatga faol xizmat qilayotgan ramzlar (znaklar) bilan ham, sun’iy tillar bilan ham qiyoslash lozim. Bu, shak-shubhasiz, tilning tabiatini mukammal yoritishga katta xizmat qiladi.

Ma’lumki, til - til (nutq) birliklari (so‘z, so‘z birikmasi, gap va boshqalar) muayyan belgilar sifatida ma’no, tushuncha, fikr ifodalaydi. Ular ma’no, tushuncha, fikr kabilarning simvoli (ramzi) yoki belgisi hisoblanadi. Shun­ga ko‘ra, Ferdinand de Sossyur til fikr ifoda etuvchi belgilar sistemasidir, deydi.12

12 Соссюр Ф. де. Курс общей лингвистики // Труды по языкознанию. - М.: Прогресс, 1977.-С.53.

16

A. Rahimov

F.de Sossyurning fikricha, lingvistik belgi ikki tomonlama psixik mo- hiyatdir. Buni quyidagi chizmada ko'rish mumkin.13

Demak, lisoniy belgi tushuncha va akustik obraz bilan bog‘liq. Bu ter- minlar «ifodalovchi» va «ifodalanmish» deb ham ataladi.14

Til birliklari o‘zaro zaruriy, mantiqiy aloqaga kirishib, til sistemasini shakllantiradi, uning obyektivligini, bir butunligini ta’minlaydi, tilning ijti- moiy vazifa bajarishi uchun imkoniyat beradi, shart-sharoit yaratadi. Ush- bu jarayonda tilning yuqorida qayd etilgan birliklarining har biri o‘ziga xos xususiyatlarga egaligi bilan, nisbiy mustaqilligi bilan, umumiy va farqli jihatlari bilan ajralib turadi.

Til sistemasining asosiy birliklari bo‘lgan fonema, morfema va so'zlarning ijtimoiy vazifa bajarishidagi eng muhim jihati ularning ikki to- monga egaligida ko'rinadi. Ushbu birliklar ifoda va mazmun jihatiga egali­gi bilan, shu nuqtayi nazardan umumiy hamda farqli xususiyatlari bilan o‘zaro ajralib turadi.

Tabiatda ham, jamiyatda ham mavjud har bir narsa, hodisa falsafiy ma’noda shakl (tashqi) va mazmun (ichki) tomonlariga ko‘ra baholanadi. Tilning eng kichik birligi bo'lgan tovush (fonema) shakl va mazmun jihati­ga ko'ra o‘ziga xos bo'lib, morfema va so‘zlardan farqlanadi. Tovush (fone­ma), qayd etilgan xususiyati nuqtayi nazaridan faqat bir tomonga - ifoda tomoniga egadek ko'rinadi. U morfema va so‘z kabi muayyan ma’no, tu­shuncha - mazmun bilan bog‘lanmaydi, ma’no ifodalamaydi. Shu nuqtayi nazardan tovush (fonema) bir tomonlama birlikka o‘xshaydi. Ammo tovush (fonema) faqat ifoda jihatiga, moddiylik xususiyatiga egaligidan tashqari, ayni vaqtda, u dastavval til birliklarining tashqi - ifoda jihatini tashkil qila- di va ularni shaklan - tovush tizimiga ko'ra farqlaydi. Shunga ko‘ra, til birliklarining ma’nosi o'zgaradi, o‘zaro farqlanadi. Demak, tovush (fone-

13 Шемакин Ю.И. Начало компьютерной лингвистики. - М.: Высшая школа, 1992. -С.7.

14 Нурмонов А. Лисоний белги хусусиятлари ^а^ида. - Андижон, 1992. - Б. 14.

Kompyuter lingvistikasi asoslari

17

ma) muayyan ma’noga ega bo‘lmagan, ayni vaqtda, muayyan ma’noga ega til birliklarining (morfema va so‘zlarning) differensiatsiyasi uchun, ularni o‘zaro farqlash uchun xizmat qiladi. Til birligi bo'lgan tovush (fonema) kommunikativ jarayonda ijtimoiy vazifa bajarishiga, tilning eng asosiy qurilish materiali, moddiy ashyosi ekanligiga, so‘z hosil qilish, so‘z yasash, ma’no farqlash kabi ijtimoiy-lisoniy xususiyatlarga egaligiga ko‘ra alohida baholanadi. Shunga ko‘ra, tovush (fonema) til birligi sifatida tilshunoslik­ning alohida sohasi bo‘lgan fonemalar haqidagi ta’limot - fonologiyaning tadqiq va tahlil abyekti hisoblanadi.

Morfema ifoda (moddiy) va mazmun (ruhiy) jihatlariningbir butunligidan iborat bo‘lib, tilning muayyan ma’noga ega eng kichik birligi hisoblanadi. Muayyan birlik bolgan so‘z (leksema) ikki planli belgi sifatida tovush va ma’no, ifoda va mazmun tomonlarining dialektik bog‘liqligidan iborat, nominativ birlik holatida ijtimoiy vazifa bajaradi. So‘zning ifoda plani uning tashqi - tovush tomoni bo'lib, bunda so'zning belgisi sifatida uning tovush tomoni xizmat qiladi. So‘z ifoda planiga ega deyilganida so'zni formal shakl- lantiruvchi, uning moddiy birlik sifatida sezgi a’zolarimizga ta’sir qilishini ta’minlovchi tovushiar nazarda tutiladi.

So‘zning mazmun plani uning ichki, ma’no tomoni (leksik ma’no, leksik tushuncha, ramziy axborot, signifikat, semema) bo‘iib, u obyektiv borliq elementlarining kishi ongida aks etishi, ayni elementlarga xos xususiyat- larning umumlashuvi natijasida yuzaga keladigan ideal (psixik) struktura- dir. Ma’no obyektiv borliq elementi yoki ushbu elementlar guruhining kishi ongidagi shartli ramzi (belgisi, ishorasi) emas, balki bu element (yoki ele­mentlar guruhi)ning kishi ongida mavhumlik kasb etgan obrazi, umumlash- gan in’ikosi hisoblanadi.

So'zning mazmun tomonini o‘zicha biron shaxsga ma’lum qilish mumkin bo'lmaganligi uchun uning tovush jihati, ramzi talaffuz qilinadi. Bu ramz (belgi) tinglovchi ongida tegishli tushuncha, tasavvur uyg'otadi. Demak, til sistemasida ramziy munosabat tovush va ma’no orasidagi shartli bog‘lanish tarzida yuzaga chiqadi. Ma’noda aks etadigan aniq narsa yoki hodisa tasav- vuri (denotat) ikkinchi planda bo‘ladi.

So‘z mazmun planida ikki hodisa farqlanadi: leksik ma’no (signifikat) va denotat.15 Denotat so‘z ma’nosi emas, balki so‘z yordamida ko‘rsatiladigan, nomlanadigan, ataladigan hodisadir. Ma’no esa, qayd etilganidek, obyektiv borliq elementlariga xos bo‘lgan xususiyatlarning inson ongida mavhumlik kasb etishi, umumlashuvi natijasida shakllanishidir. Nutqda so‘z denotat-

15 Нурмонов А. Лисоний белги хусусиятлари ^а^ида. - Андижон, 1992. - Б.17.

18

A. Rahimov

ning (obyektiv borliq elementining) lingvistik vakili vazifasini bajaradi. So‘zda tovush ifodalovchi (ramz, belgi) bo'lsa, ma’no ifodalanuvchi bo‘ladi.

So‘zdan yirik so‘z birikmalari va gaplar ham ifoda va mazmun planlari- ga ega, ularning mazkur ikki jihati ko‘proq grammatika va semantikaning aloqasi doirasida qaraladi. Ushbu nutq birliklari ham, aslida, ifoda va mazmun planiga ega so'zlarning birikuvidan, o'zaro bogManishidan yuza- ga keladi. Shu sababli ularni o‘rganish ularni hosil qilgan so'zlarni o‘rganishdan boshlanadi.

Informatsiya nazariyasi informatsiyani uzatish va saqlashning mazmun tomonini emas, balki uning statistik qurilishi bilan shug'ullanadi, ya’ni in­formatsiya nazariyasi axborot matnini statistik struktura sifatida o'rganadi. Statistik struktura axborotdagi signal (simvol) va signallar birlashuvining voqelanish chastotasi soni, miqdoridir. Signalning ko‘p marta takrorlanishida signal orqali uzatiladigan axborotning miqdorini aniqlash juda muhim. Demak, informatsiya nazariyasi muayyan axborotdagi ma’lumotlarni ush­bu ma’lumotlarning tabiatiga - mazmuniga bogiiq bo'lmagan holda miq­dorini oichash (tekshirish) yoilarini, usullarini ishlab chiqadi. Shunga ko‘ra, informatsiya nazariyasining muhim va asosiy xulosalari tilshunoslikda ham o‘z ifodasini topadi. Informatsiya nazariyasining asoschisi amerikalik olim K.Shennon informatsiyaning quyidagi aloqa kanallari bo‘yicha uzatilishini sxema asosida tushuntiradi:16

Manba —► uzatuvchi —►kanal—► priyomnik—► axborotni oluvchi

Ushbu holat til vositasida insonlarning kommunikatsiyaga kirishuv jara- yoniga ham to'la mos keladi. Aslida, ushbu model til, tafakkur va nutq mu- nosabatidan kelib chiqqan. Informatsiya tabiiy yoki sun’iy tillar orqali uza- tilar ekan, bunda til sistemasi kodlash va dekodlash jarayonlariga bog‘liq holda talqin etiladi.

Informatsiya nazariyasi ma’lumotlarini o‘rganish jarayonida maxsus terminlardan - tushunchalardan foydalaniladi. Bular quyidagilar:

1. Kod - axborotni yozib olish vositasi (usuli).

2. Matn (tekst) - muayyan axborotni berish, uzatish jarayonidagi belgi- larning (til birliklarining, simvollarning) izchilligi, ketma-ket kelishi.

3. Kanal - axborot beriluvchi, uzatiluvchi muhit, sharoit.

4. Informatsiya asosiy birlik bo‘lib, u bit (yoki binit inglizcha «binary unit» - «ikkilik birligi») deb ataladi.17

16 Эко У. Отсутствующая структура. Введение в семиологию. - М.: Петрополис, 1998. - С.45.

17 Реформатский А. А. Введение в языковедение. - М.: Просвещение, 1967. -

С.51.

Kompyuter lingvistikasi asoslari

19

5. Kodlash va dekodlash - informatsiyani mashina (kompyuter) tushuna- digan tilga o‘tkazish va bunga teskari bo‘lgan kompyuter tushunadigan til- dan inson tushunadigan tilga aylantirish jarayoni.

Tashqi olam haqidagi informatsiya insonning bosh miya qobig‘ida model va strukturalar ko‘rinishida saqlanadi, aks etadi va qayta ishlanadi. Ushbu holatning fiziologik asosi nerv-sensor hujayralar - aksonlar hisoblanadi. Bosh miya qobig'ida so‘zli strukturalar hamda semantik maydon shaklla- nadi. Funksionallashuv jarayonida (xabarning bir individuumdan boshqasi- ga yetkazilishi va qabul qilinishi) til diskret shaklda - nutq ko‘rinishida namoyon boiadi. Tovushli nutqning idrok qilinish jarayoni - bu nutq to- vushlaridan informatsiyaning ajralishi, tarqalishi demakdir. Eshitish organi havoning tebranishini qabul qiladi, signalni bosh miya qobig'iga jo'natish uchun uzluksiz impulslarni kodlashtiradi. Birinchi bosqichda bosh miyada tovushlar zaxirasidan «tovushlar modeli» (N.T.Jinkin buni «fonemalarning barqaror katakchalari» deb atagan) shakllanadi. Inson miyasidagi obrazlar va tushunchalar tovushlar kombinatsiyasi ko'rinishida moddiy qobiqqa ega bo‘ladi. Inson xotirasida so'zlarning modellari aniq hujayralar kompleksi- ga muvofiq ravishda saqlanadi. Informatsiyani solishtirish natijasida miya­da yangi kod- so‘z kodi kodlashtiriladi va uni qobiqning keyingi bosqichiga uzatadi. U yerda informatsiyaning yangi modeli shakllanadi, bunda oldin- giga nisbatan kodlashtirish oson kechadi, chunki o‘zida mavjud bo‘lmagan ramzlarnigina kodlashtiradi, keyingi so‘zni kodlashtirish oldingisidan ko‘ra ham oson kechadi, keyingi bosqichda - birikmalar kodlashtiriladi hamda informatsiya keyingi bosqichga uzatiladi. Bunda yuqoridagi jarayon qayta- riladi: informatsiya ajratiladi va mazmun modellashtiriladi.18

Tilni informatsiya nazariyasiga ko‘ra yuqorida qayd etilgan uchta tush- unchalardan kod bilan tenglashtirish, kod bilan parallel qo‘yish mumkin.19 Chunki kod, aytilganidek, axborotni yozib olish usuli, vositalaridan biridir. Demak, til - tilning har bir belgisi (til birliklari) kod sifatida muayyan ax- borot tashiydi, uzatadi, yetkazadi. Binobarin, ayni vaqtda, ushbu axborotni hisoblab chiqish, statistik ma’lumot to‘plash mumkin. Xullas, til semiotika va informatsiya nazariyasiga ko‘ra kod sifatida namoyon bo‘ladi.

Algoritm (lotincha algoritmus-«ibtido» ma’nosini anglatadi; bu so'zning etimologiyasi buyuk alloma Al-Xorazmiy nomi bilan bog‘liq. U Yevropada buzilgan shaklda «Algorithmi» tarzida ishlatilgan) - ma’lum bir tipga oid

18 Qarang: Амосов H.M. Моделирование мышления и психики. - Киев, 1965.

19 Шемакин Ю.И. Начало компьютерной лингвистики. - М.: Высшая школа, 1992.-С.7.

20

A. Rahimov

hamma masalalarni yechishda ishlatiladigan jarayonlar tizimining muay- yan tartibda bajarilishi haqidagi aniq qoida yoki qoidalar tizimi.20 O'rta asrlarda sanoqning o'nli sistemasi bo‘yicha to'rt arifmetik amal bajariladi- gan qoida algoritm deb atalgan. Bu qoidalarni fanga IX asrda o‘zbek mate- matigi Muhammad al-Xorazmiy kiritgan. Uning «Arifmetika» asarining birinchi so‘zi «Qola al-Xorazmiy» so'zlari lotin tiliga «Dixit Algorithmi» («Al-Xorazmiy dediki») deb tarjima qilingan. Bu yerda al-Xorazmiy so‘zi «Algorithmi» deb yozilgan. Shundan so'ng algoritm tushunchasi fanda e’tirof etilib, uning qo'llanilish doirasi kengaydi. XX asrda algoritmlar kiberneti- kaning nazariy va mantiqiy asosi sifatida tan olingan.

Qo'shuv, ayiruv, ko'paytiruv, bo'luv kabi arifmetik qoidalar, kvadrat ildiz chiqarish qoidalari, har qanday ikki natural son uchun eng umumiy bo'luvchini topish usuli arifmetikaga eng oddiy misollardir. Aslida, biror vazifani umumiy tarzda, ya’ni uning variatsiya qilinuvchi shartlarining butun turkumi uchun hal qilish vositalariga ega boiganimizda biz algoritm bilan ish ko‘ramiz. Oddiy hayotiy misol - choy yoki kofe tayyorlashning ham o‘z algoritmi mavjud.

Kompyuter texnologiyalari taraqqiyoti bilan bog‘liq holda odam bilan mashina orasida o'ziga xos «ko‘prik» vazifasini bajaruvchi sun’iy tillar yara- tildi, ular programmalashtirish jarayonida qoilaniladi. Sun’iy tillar buyruq- larning tizimli algoritmiga tayanib yaratilgani uchun ular algoritmik tillar (programmalashtirish tillari) deb ham ataladi. Dastlabki shunday til 1958- yilda Parijda yaratilgan va «Algol» deb atalgan (Uning toiiq nomi inglizcha

- «Algorithmic language»). Hozir jahonda 500 dan ortiq algoritmik tillar mavjud.

Algoritmik tillardagi matnlar xuddi tabiiy tildagidek aniq interpretator- ga - insonga yoki avtomatik qurilmaga yo‘nalgan bo‘lishi lozim, toki u axborotni qabul qilsin va tushunsin. Agar bu jarayon (qabul qilish va tushu- nish) amalga oshmasa, u holda algoritmik tillar va kompyuter orasida vosi- tachi sifatida translyator dasturlar (yoki kompilyatorlar)dan foydalaniladi.21 Algoritmik tillar uch asosiy komponentdan iborat bo'ladi: 1) alfavit - asosiy simvollar nabori; 2) sintaksis - simvollar kombinatsiyasi orqali jumla hosil qilish qoidalari; 3) semantika - algoritmik tillardagi birliklarning mazmun tomoni. Algoritmik tillar asosan imperativ xarakterga ega bo'ladi, ya’ni asosan buyruqlar tizimiga tayanadi (begin - boshla, end - tugalla kabi). V.A.Uspenskiy va A.L.Semyonovlar bu haqida shunday yozadi; «Algoritm-

211 Фалсафа, Кис^ача изошли лугат. - Тошкент: Шарк, 2004. - Б. 18.

21Кревский И.Г., Селиверстов М.Н., Григорьева К.В. Формальные языки, грамматики и основы построения трансляторов. - Пенза, 2003. - С.7.

Kompyuter lingvistikasi asoslari

21

ning mazmuni «buyruqlashtirilgan»: algoritm ijro etilishi, bajarilishi shart. Shunday qilib, aytish mumkinki, algoritmlarni о ‘rganuvchi nazariya buyruq gaplar lingvistikasi sifatida talqin etilishi mumkin».22

Algoritmlar asosan quyidagi shakllarda yoziladi: a) so‘z ko‘rinishida - tabiiy til yordamida; b) grafik ko'rinishida - tasvir, chizma va grafikli sim- vollar yordamida; c) psevdokod ko‘rinishida - algoritmlarning yarim for- mallashgan tavsifi bo'lib, u programmalashtirishning dastlabki bosqichlar- ida qo‘llaniladi. Bunda programmalashtirish tillari unsurlaridan ham, ta­biiy tildagi so‘z va birikmalardan ham foydalaniladi; d) programma ko'rinishida - programmalshtirish tillaridagi belgilar yordamida.

Algoritm buyruqlar tizimi, qat’iy bajariladigan amallar ketma-ketligi hamda programmalashtirishning tayanch nuqtasi sanalgani bois uning vosi- tasida kompyuter dasturlarini ishlab chiqish va muammoning avtomatik tarz- da (mashina orqali) hal etilishini amalga oshirish mumkin. Algoritmni qam- rovli vazifalarni hal etishga yo‘naltirish, algoritm nazariyasini mukammal- lashtirish, ayniqsa, kompyuter texnologiyalari, shuningdek, kompyuter lingvistikasi rivojlanayotgan davrimizda dolzarb ahamiyat kasb etadi.

V.A.Uspenskiy va A.L.Semyonovlar algoritm haqida shunday yozadi: «Al­goritm tushunchasi nafaqat algoritmlar nazariyasi, matematika va informati- kaning tushunchasi, balki umuman zamonaviy fanlarning tayanch tushun- chalaridan biridir. Axborot asrining kirib kelishi bilan algoritmlar sivilizatsi- yaning eng muhim Jaktorlaridan biri bo'lib qoldi». Shu o'rinda ta’kidlash joizki, kompyuter lingvistikasida shunday masalalar ham borki, ularni al­goritmlar asosida hal etish mumkin emas. Jumladan, sun’iy intellekt tizimi doirasidagi neyron tarmoqlar, sifatli avtomatik tarjima kabilar. Ushbu muam- molarni hal etishda «algoritmga asoslangan yondashuv» («rule based approach»)dan tashqari, analogiya prinsipiga asoslanadigan «namunalarga asoslangan yondashuv» («example based approach»)ga ham ehtiyoj seziladi.

Buyuk faylasuf Gegelning fikricha, har qanday fan tadqiq etilgan man- tiqdir.

Topshiriqlar:

1. Mantiq va semiotikaning kompyuter lingvistikasidagi tatbiqini tu- shuntiring.

2. Konyunktiv, disyunktiv, implikativ hukmlarga o‘zbek tilidan misol- lar keltiring.

22 Успенский В.А., Семенов A.JT. Теория алгоритмов: основные открытия и приложения. - М.: Наука, 1987. - С. 13.

22

A. Rahimov

3. Matematik mantiq fanini qaysi olimlar rivojlantirgan?

4. Ushbu gapning propozitsional ko‘rinishini yozing: Agar Jons bu tun- da Smitni uchratmagan bo‘lsa, u holda Smit qotil yoki Jons aldayapti. Agar Smit qotil bo‘lmasa, u holda Jons bu tunda Smitni uchratmadi va qotillik yarim tundan keyin sodir boidi. Agar qotillik yarim tundan keyin sodir boigan boisa, u holda Smit qotil yoki Jons aldayapti. Yuqoridagilardan kelib chiqadiki, Smit qotil.

5. Qanday algoritmik tillarni bilasiz?

Adabiyotlar:

1. Yoqubov Т., Karimbekov S. Matematik mantiq elementlari. - Tosh- kent: 0‘qituvchi, 1996.

2. Шемакин Ю.И. Начало компьютерной лингвистики. - М.: МГОУ, 1992.

3. Лагута О.Н. Логика и лингвистика. - Новосибирск, 2000.

4. Нурмонов А. Лисоний белги хусусиятлари ^а^ида. - Андижон, 1992.

5. Успенский В.А., Семенов А.Л. Теория алгоритмов: основные открытия и приложения. - М.: Наука, 1987.

6. Кревский И.Г., Селиверстов М.Н., Григорьева К.В. Формальные языки, грамматики и основы построения трансляторов. - Пенза, 2003.

7. Эко У. Отсутствующая структура. Введение в семиологию. - М.: Петрополис, 1998.

8. Соссюр Ф. де. Курс общей лингвистики //Труды по языкознанию.

- М.: Прогресс, 1977.

9. Белоногов Г.Г. Компьютерная лингвистика и перспективные информационные технологии. - М.: Русский мир, 2004.

10. Po'latov A., Muhamedova S. Kompyuter lingvistikasi. -Toshkent, 2007.

Kompyuter lingvistikasi asoslari

23

AKSIOMATIK NAZARIYA - TILNI FORMAL TAVSIFLASH ASOSI

Reja:

1. Aksiomatik nazariya haqida umumiy ma’lumot.

2. Aksiomatik nazariyaning lingvistik tadqiqi.

3. Lingvistik universaliyalar g'oyasi.

4. Metagrammatika.

Tayanch so‘z va iboralar: aksioma, postulat, teorema, ziddiyatsizlik, to ‘lalik, bog ‘liqsizlik, aksiomatik metod, aksiomatik nazariya bosqichlari, formula, keltirib chiqarish qoidalari, rrtetagrammatika, metaqoidalar, lingvistik univer­saliyalar, frekventaliya va unikaliyalar, metaqoida, implikativ universaliya.

Aksiomatik nazariya miloddan avvalgi VII - VI asrlarda shakllangan. Uning yaratilishi matematik olim Pifagor nomi bilan bog‘liq. Lekin bu metodni birinchi marta miloddan avvalgi III asrda yashab ijod etgan yunon olimi Yevklid o‘zining «Negizlar» asarida qoMlagan. Aksiomatik nazariya bosqichlari quyidagilardan iborat:

1. Postulat - yunoncha «talab» ma’nosini anglatadi. Qadimgi dialektika o‘yinida har bir kishining ilgari surgan g‘oyasi, fikri. Uni tasdiqlash ham, inkor qilish ham mumkin.

2. Aksioma - yunoncha «hurmat qilaman» ma’nosini anglatadi. U ilmiy nazariyaning hech qanday mantiqiy isbotsiz to'g'ri deb tan olinuvchi va boshqa ilmiy xulosalar uchun asos bo‘luvchi maqbul deb topilgan haqiqat, go‘yo olimlar bu fikrni «hurmat qilib» inkor etmaydi. Arirstotel aksiomani «har qanday isbotsiz e’tirof etiladigan boshlang'ich hukmlar», «ilm-fanning poydevori, ibtidosi» sifatida talqin qilgan.

3. Teorema - aksiomadan farqli ravishda mantiqan keltirib chiqariladi- gan jumlalar, u shart, isbot va xulosa qismlariga ega.

Fanni aksiomatik asosda qurish uchun qabul qilingan aksiomalar tizimi quyidagi asosiy talablarni qanoatlantirishi kerak:23

- ziddiyatsizlik - aksiomalar dzimining shunday xossasiki, unga ko‘ra, sistemaning hech qanday ikkita holati bir-biriga zid kelmasligi kerak. Bun- da aksiomalar tizimining o‘zida rost va yolg‘on orasidagi farqni aks ettirishi kerakligi nazarda tutiladi;

23 <http://www.wikipedia.org/wiki/axiom.html> / Гильберт Д. Аксиоматическое мышление / Перевод с англ. А.Г.Барабашева.

24

A.Rahimov

- to‘lalik - aksiomalar sistemasiga qandaydir jumlani qo'shganimizda sistemada ziddiyat sodir boiadi;

- bogiiqsizlik - berilgan aksiomalar sistemasining u yoki bu sistemalari shu sistemaning boshqa sistemalari yordamida isbot qilinishi mumkin emas.

Aksiomatik metod - ilmiy nazariyaning shunday qurilishiki, unda bu nazariya asosida shunday boshlang‘ich postulat va aksiomalar qo‘yiladiki, qolgan barcha qoidalar (teoremalar) undan kelib chiqadi. Aksiomalar isbot talab qilmaydi, deymiz, lekin uning isboti uzoq asrlik kuzatishlar natijasida yuzaga kelgan va shuning uchun ham isbot shart emas. Aksiomatik nazariya masalalar modelini tuzishdagi universal til hisoblanadi. Bundan tashqari, mazkur til kompyuter tilidir. Formal nazariyada har bir formula umumiy ma’no nuqtayi nazaridan alohida tekshiriladi, aksiomatik nazariyada esa ko'rib chiqilayotgan formulaning umumiy ma’nosini tekshirishda isbot qilin- gan formulalarning barchasidan foydalanish mumkin boiadi. Aynan shu­ning uchun mazkur metod samarali hisoblanadi.

Aksiomatik nazariyada dastlabki boshlang‘ich tushunchalar, postulat, aksiomalar, teoremalar, keltirib chiqarish qoidalari asosiy o‘rin tutadi. Masalan: Aksioma: so‘zda nechta unli boisa, shuncha bo‘g‘in boiadi. Teo- rema: Agar ixtiyoriy X gap inversiyaga uchramasa, NSP - Normal Sen­tence Pattern (normal gap qurilishi)ga mos kelsa, u holda ega doim kesim- dan oldin keladi.

Aksomatik nazariya quyidagi holatlar aniqlanganda haqiqiy hisoblana­di:24

1. Nazariya ifodalarini tavsiflash uchun zaruriy boigan nazariya sim- vollari - harflar, belgilar yoki raqamlar.

2. Nazariya formulalari, ya’ni nazariyada fikrlangan barcha ifodalar- ning tavsifi.

3. Nazariya aksiomalari (boshlangich va mutlaq to‘g‘ri deb hisoblan- gan formulalar).

4. Nazariya xulosalarini chiqarish qoidalari, ya’ni mutlaq to‘g‘ri formu- lalardan boshqa mutlaq to‘g‘ri formulalarni keltirib chiqarish imkonini beruvchi barcha qoidalarni hisoblab chiqish.

5. Nazariyaning teoremalari - nazariyaning mutlaq to‘g‘ri, umum to- monidan qabul qilingan formulalari.

Til grammatikasining matematik modeli matematik mantiqning aksi­omatik nazariyasiga asoslanadi. Aksiomtik nazariya predmeti tabiiy tillarda- gi gaplardir (ya’ni ixtiyoriy so'zlar ketma-ketligi). Teorema sifatida tabiiy

24 Po'latov A., Muhamedova S. Kompyuter lingvistikasi. - Toshkent, 2007. -

B.15.

Kompyuter lingvistikasi asoslari

25

tillarda to‘g‘ri hisoblangan gap konstruksiyalari ko'rib chiqiladi. Dastlab tabiiy til aksiomatik nazariyasining bosqichlarini keltiramiz:

1. Aksiomatik nazariyaning simvollari: til harflari (alifbo), tinish belgi- lari va mantiqiy operatsiyalarning belgilari.

2. Aksiomatik nazariyaning formulalari: tildagi so'zlar, ular maxsus kengaytirilgan lug'atda keltiriladi.

3. Til grammatikasi aksiomalari: gaplarning dastlabki to‘g‘ri konstruk- siyalari. Masalan, ingliz.tili uchun ega+kes+toid+hol (kengaytirilgan tas- diq (darak) gap kostruksiyasi): Mother bought milk yesterday. 0‘zbek tili uchun ega+to‘ld+hol+kes (kengaytirilgan tasdiq (darak) gap kostruksiya­si): Nafisa ma’ruzani diqqat bilan tingladi.

4. Keltirib chiqarish qoidalari (berilgan konstruksiyalardan yangi gap konstruksiyalarini hosil qilish qoidalari):

G - ingliz tilidagi muayyan bir gap;

К Ch q - keltirib chiqarish qoidalari.

1. G (e,k,t,h)+%G‘ ("<?,&,), ya’niGgapidanyangi (kengaytirilmagan gap) hosil boiadi (unda gapning ikkinchi darajali boiaklari tushirib qoldiril- gan).

G: The car stopped at the gate - Mashina darvoza yonida to'xtadi.

G1: The car stopped - Mashina to'xtadi.

2. G (e,k,h) + %oG2 (e‘,k,h). Bunda G gapidan aniqlovchili (ya’ni har bir gap bo'lagining o‘z aniqlovchisi mavjud) gap hosil boiadi.

G: The car stopped at the gate - Mashina darvoza yonida to'xtadi.

G2: The blue car stopped at the gate - Ko‘k mashina darvoza yonida to‘xtadi.25

Aksiomatik nazariyaning rivoji natijasida barcha tillar uchun xos boigan umumiy qonuniyatlarni aniqlash masalasi ko‘ndalang qo‘yildi. Bu hodisa lingvistik universaliyalar deb ataladi.26 Ch.Xokket lingvistik universali- yalarni shunday ta’riflaydi: «Lingvistik universaliya - dunyodagi barcha tillarga xos bo ‘Igan yoki umuman tilga xos muayyan belgi, xususiyat».

B.Uspenskiy: «Lingvistik universaliyalar tushunchasi ostida barcha tillarga yoki mutlaq ko'pchilik qismiga xos umumiy qonuniyatlar tushuniladi». Lingvistik universaliyalar haqidagi dastlabki g‘oyalar XX asrning 50-yil- lariga to‘g‘ri keladi. 1941-yilda Aginskiy tomonidan «The Importance of language universals» sarlavhali maqola e’lon qilindi. 1961-yilda Nyu-York

:5 Po'latov A., Muhamedova S. Kompyuter lingvistikasi. - Toshkent, 2007. -

B.16.

26 Qarang: Новое в лингвистике. Вып.У. Языковые универсалии. - М.: Прогресс, 1970.

26

A. Rahimov

shahrida universaliyalarga bag‘ishlangan xalqaro konferensiya o'tkazildi. Unda J.Grinberg, J.Jenkins, Ch.Osgudlar «Lingvistik universaliyalar haqi- da memorandum» mavzusida chiqish qilishdi. Konferensiyada Ch.Osgud shunday yozadi: «Biz ushbu konferensiyada qonsiz inqilobning guvohi bo'lamiz». Haqiqatan ham shundan so‘ng universaliyalarning tadqiqi kuchaydi. Bunda dastlab universaliyalar, frekventaliyalar (tabiiy tillarda tez-tez kuzatiladigan hodisalar, belgilar) hamda unikaliyalar (muayyan tillarda uchraydigan o‘ziga xos hodisalar, jarayonlar va xususiyatlar) aniqlanib, metaqoidalar ishlab chiqiladi va natijada etalon til (metatil) yuza­ga keladi. Butun dunyo tillarining grammatik tavsifi yaratilishi jarayonida, ma’lumotlar bazasini tayyorlashda, formal grammatika yaratishda univer- saliyalarni aniqlash juda muhim.

Amerikalik olim Jozef Grinberg universaliyalarni aniqlashda yangicha bir yondashuvni taklif etdi. Unga ko‘ra, universaliyalar implikatsiya asosi­da aniqlanadi. Bu implikativ universaliya deb nomlanadi. Olim 45 ta imp- likativ universaliyani aniqlagan.27 Masalan:

1. Agar til uzluksiz affikslarga ega bo'lsa, unda yo prefiksatsiya yoki suffiksatsiya yaxshi rivojlangan bo‘ladi.

2. Agar til istisnoli tarzda suffiksal bo‘lsa, u posleloglarga, kelishiklarga boy bo'ladi.

3. Agar til istisnoli tarzda prefiksal bo'lsa, u predloglarga boy bo'ladi.

4. Agar til VSO modelida bo‘lsa, unda predloglar dominantlik qiladi.

Fonologiya, morfologiya, sintaksis, semantika sohalarida aniqlangan

lingvistik universaliyalar 2000-yilga kelib 1600 tani tashkil etdi. Germani- yaning Konstanz universitetida lingvistik universaliyalarning bazasi (Uni- versals Archive) yaratildi va u har yili yangi ma’lumotlar bilan boyitib bo- riladi. Uni obrazli qilib D.I.Mendeleyevning kimyoviy elementlar davriy jadvaliga o'xshatish mumkin. Davriyjadvalda keyinroq aniqlanishi mumkin bo'lgan elementlar uchun bo'sh kataklar qoldirilgan. Xuddi shunga o'xshab universaliyalar bazasini toldirib borish uchun imkoniyat mavjud.28 Maz- kur baza Internetda h\_ttp://[www.ling.uni-konstanz.de/pages/proj/](http://www.ling.uni-konstanz.de/pages/proj/) sprachbau.htm saytiga joylangan.

Lingvistik universaliyalar g‘oyasi ta’sirida metagrammatika haqida qarashlar, nazariy va amaliy tadqiqotlar yuzaga keldi. Ilmiy adabiyotlarda metagrammatika atamasiga turlicha ta’riflar mavjud:

27 Гринберг Дж. Некоторые грамматические универсалии, преимущественно касающиеся порядка значимых элементов // Новое в лингвистике. Вып.У. Языковые универсалии. - М.: Прогресс, 1970. - С.115 - 162.

28 htfrv/Лии™ lino i in i Unttan, Но/папм/пгл.'/гпгагЬК-л! htm

Kompyuter lingvistikasi asoslari

27

1. Metagrammatika (yunoncha «metha» - «keyin», «grammatike» - «o‘qish va yozish san’ati») - metaqoidalarga asoslangan grammatika.

2. Metagrammatika - tavsifiy grammatika, formal grammatika, trans- formatsion grammatika bosqichlaridan so‘ng yuzaga kelgan, qayta yozil- gan grammatika.

3. Metagrammatika - programmalash tillarini tavsiflash uchun tabiiy til grammatikasi modeli asosida ishlab chiqilgan metaqoidalar to'plami.

Metagrammatika haqidagi dastlabki g‘oyalar Dekart, Leybnis kabi matematiklar tomonidan ilgari surilgan ratsional (universal) grammatika yaratish, Klod Lanslo va Antoniy Arnoldlarning (1660, Parij) «Por-Royal grammatikasi» bilan bog‘liqdir. Algol 68 uchun ishlangan metaqoidalar Winjgaarden-grammar yoki W-grammar deb nomlangan. Keyinroq 1981- yil olimlar Gazdar, Pullman, Saglar ham metaqoidalar ishlab chiqdi. Uning nomi Generalized Phrase Structure Grammar (GPSG).29

Masalan:

For every production of the form

<VP> ::= <\*v> <NP> X

add a production of the form

< VP > :: = be < \*VEN > X

(where VEN is the word class for past participles). The ‘x’ in themetarule matches any sequence of zero or more symbols.

< VP > :: = < \*V > < NP > and

<VP> :: = < \*V> <NP> <PP>

the meta rule will add the productions

< VP > :: = be < \*VEN > and

< VP : : = be < \*VEN > < PP >

Mazkur metaqoida asosida «Lincoln was shot» («Linkoln otildi») va «А book was given to Mary» («Kitob Mariyaga berildi») kabi majhul nisbat shaklidagi jumlalar hosil qilinishi mumkin. Bu yerda majhul nisbatdagi gaplarning egasi aniq nisbatdagi vositali to'ldiruvchi vazifasini bajaradi.

0‘zbek tilida metaqoidaga misol sifatida formal-funksional tilshunoslik vakillarining WPmmodelini ko‘rsatish mumkin. Bunda W -inglizcha Word so'zidan olingan bo‘lib, har qanday nominativ birlikni anglatadi, P - predi­cate so‘zidan, m - marker so'zidan olingan, u predikativlik (kesimlik) ko‘rsatkichini bildiradi. Bu o'zbek tilidagi gapning minimal qolipidir.

29 Grishman R. Computational linguistics // Cambridge University Press. 1994.

- P.82 - 83.

28

A. Rahimov

Topshiriqlar:

1. Aksioma nima, unga misollar keltiring.

2. Boshlang'ich tushunchalar, postulatlarni belgilash qanday amalga oshiriladi?

3. Grammatikaning aksioma va teoremalariga o'zingiz misollar kelti­ring.

4. Keltirib chiqarish qoidalari deganda nimani tushunasiz?

5. Lingvistik universaliyalar naborini jurnal yoki Internetdan topib, daf- taringizga yozing.

Adabiyotlar:

1. Grishman R. Computational Linguistics // Cambridge University Press. 1994.

2. Шемакин Ю.И. Начало компьютерной лингвистики. - М.: МГОУ, 1992.

3. Good J. Linguistic Universals and Language Change // Oxford Uni­versity Press. 2008.

4. Gelbukh A. Computational Linguistics and Intelligent Text Process­ing. International Conference. - Mexico, 2003.

5. Гринберг Дж. Некоторые грамматические универсалии, преимущественно касающиеся порядка значимых элементов // Новое в лингвистике. Вып.У. Языковые универсалии. - М.: Прогресс, 1970. -

С. 115 - 162.

6. Новое в лингвистике. Вып.У. Языковые универсалии. - М.: Прогресс, 1970.

7. PoMatov A., Muhamedova S. Kompyuter lingvistikasi. - Toshkent,

2007.

8. <http://www.wikipedia.org/wiki/axiom.html> / Гильберт Д. Аксиоматическое мышление / Перевод с англ. А.Г.Барабашева.

Kompyuter lingvistikasi asoslari

29

KOMPYUTER LINGVISTIKASIDA MODELLASHTIRISH VA ANALOGIYA METODLARI

Reja:

1. Model tushunchasining mohiyati.

2. Modellashtirish va uning tilga tatbiqi.

3. Modellashtirishning pragmatik jihatlari va kamchiliklari.

4. Kompyuter lingvistikasida analogiya metodi.

Tayanch so‘z va iboralar: model, tabiiy model, matematik model, manti- qiy-matematik model, metamodel, deduktiv, lingvistik model, sintaktik model­lashtirish, so'z turkumlari va ularning simvollari, gap bo'laklari simvollari, analogiya metodi, induksiya, deduksiya.

Model (lotincha «modelus» so'zidan olingan boiib, «nusxa», «anda- za», «o'lchov», «me’yor» ma’nolarini anglatadi) tabiiy fanlar yoki umu- man fanda shunday moddiy qurilma, grafik, sxema, umuraan, bilish vosita- si sifatida tushuniladiki, u muayyan original-obyekt haqidagi ma’lumotlar majmui tarzida yuzaga kelgan hosila-obyekt demakdir. Boshqacharoq tushuntirganda, model tabiiy obyektlarning imitatsiyasidir (o‘xshashi, taq- lidiy ko‘rinishi), u o‘zbek tilidagi qolip, andaza so‘zlariga mos keladi. U hodisalarning yuzaga kelishi uchun asos vazifasini o‘taydi, bunda aniq yoki mavhum obyektlar kichraytirilgan obyektlar, sxemalar, chizmalar, fizikaviy konstruksiyalarda tadqiq etiladi. Buni oddiy hayotiy misol bilan tushuntira- digan bo‘lsak, olmani xarakterlovchi belgilar, atributlar - uning dumaloq- ligi, mevaligi, shirinligi o'sha tushunchaning fikriy modeli hisoblanadi. Agar biz olmani loydan yoki sun’iy bir materialdan yasasak, bu uning moddiy modeli hisoblanadi. Yoki globus yerning modeli, o‘yinchoq mashina haqi- qiy mashinaning modeli sifatida baholanishi mumkin.'0 Model quyidagi asoslarga ko‘ra bilishda muhim hisoblanadi:

- birinchidan, u o‘rganish obyektini soddalashtiradi;

- ikkinchidan, uni boshqa obyektlar ta’siridan ajratadi;

- uchinchidan, model obyektni ta’riflashni osonlashtiradi.

Model dastlab amaliy sohalarda ishlatilgan. Keyinchalik ilm-fanning ijtimoiy sohalariga ham kirib keldi. Bu matematika, kimyo fanlariga keng tatbiq qilingani yaqqol ko‘rinadi. «Model» tushunchasi fan va texnikada

,0 <http://www.wikipedia.org/wiki/model.html>

30

A. Rahimov

turli ma’nolarda ishlatilgani bois modellashtirish turlarining yagona tasnifi mavjud emas. Tasnif modelning xarakteriga ko‘ra, modellashtirilayotgan obyektning tabiatiga ko'ra, modellashtirish tatbiq qilinayotgan soha yoki yo'nalishga qarab amalga oshirilishi mumkin. Masalan, modellashtirish- ning quyidagi turlarini ajratib ko‘rsatish mumkin: informatsion modellash­tirish, lingvistik modellashtirish, kompyuterli modellashtirish, matematik modellashtirish, matematik-kartografik modellashtirish, molekulyar model­lashtirish, mantiqiy modellashtirish, pedagogik modellashtirish, psixologik modellashtirish, statistik modellashtirish, struktur modellashtirish, fizik modellashtirish, iqtisodiy-matematik modellashtirish, imitatsion modellash­tirish, evolutsion modellashtirish, kognitiv modellashtirish kabi.

Modellarni shartli ravishda quyidagicha tasniflash mumkin:

1. Tabiiy modellar - o‘rganilayotgan obyekt bilan bir turda bo‘ladi va undan faqat oMchamlari, jarayonlarining tezligi va ba’zi hollarda yasalgan materiali bilan farq qiladi.

2. Matematik modellar - prototipdan (asl nusxadan) jismoniy tuzilishi bilan farq qiladi, lekin prototip bilan bir xil matematik tasvirga ega boMadi.

3. Mantiqiy-matematik modellar - belgilardan iborat bo‘lib, abstrakt model hisoblanadi va tafakkur jarayonini o‘rganishda qo'llaniladi.

4. Kompyuterli modellar - matematik va mantiqiy modellashtirish metodlariga asoslanib kompyuterda algoritm va dasturlardan foydalanib yaratiladigan modellar.

Kompyuterli modellashtirish bugungi kunda barcha fanlarda, xususan, kompyuter lingvistikasida ham samarali metod hisoblanadi.31 Kompyuter modellashtirish quyidagi asosiy bosqichlardan iborat:

- masalaning qo'yilishi, modellashtirish obyektining aniqlashtirilishi;

- konseptual (tushunchaviy, fikriy) modelning ishlab chiqilishi, tizim aso­siy unsurlarining ajratib olinishi;

- formalizatsiya, ya’ni matematik model bosqichi; algoritm yaratilishi hamda dastur ishlanishi;

- kompyuter eksperimentlarini o‘tkazish;

- natijalar tahlili va talqini.

Sun’iy intellekt tizimi doirasidagi tabiiy tilli interfeys, ekspert tizimlari, eyron tarmoqlar, lingvoanalizatorlar, gapiruvchi avtomatlar - barchasi kompyuter modellashtirish natijasi hisoblanadi.

Modellashtirish jarayoni uchta unsurni o‘z ichiga oladi:

- subyekt (tadqiqotchi);

31 Grishman R. Computational linguistics // Cambridge University Press. 1994. - P.7 - 8.

Kompyuter lingvistikasi asoslari

31

- tadqiqot obyekti;

- o‘ rganuvchi subyekt va o‘ rganiluvchi obyekt munosabatini aks ettiruv- chi model.

Modellashtirish har bir fan obyektini soddalashtiruvchi metoddir. Lingvis­tik birliklarni modellashtirish bu belgilar tarkibidagi elementlarning bar- qaror munosabatlariga asoslanadi. Shuning uchun ham butunlik tarkibida­gi elementlar o'rtasida munosabatlarning barqaror va beqaror turlarga ajra- tilishi lingvistik modellashtirish uchun katta ahamiyatga ega. Modellashti­rish barcha fanlar uchun xos boigan umumilmiy metod hisoblanadi va u quyidagi tamoyillarga amal qiladi:

- deduktivlik - mantiqiy xulosa chiqarishga asoslangan boiadi, xusu- siylikdan umumiylikka tamoyilida boiadi;

- tafakkur eksperimentidan foydalanish;

- evristik funksiyaga ega boiishi. Ya’ni u yangi g'oyalar bera olishi va uni amaliyotda sinab ko'rish imkoniyati boiishi zarur;

- eksplanatorlik xususiyatiga, ya’ni tushuntirish kuchiga ega boiishi kerak. Shundagina model eski nazariya tushuntirib bera olmagan muam- moni hal qiladi, obyektning ilgari kuzatilmagan, ammo kelajakda amalga oshishi mumkin boigan tomonini kashf etadi;12

- modelni ideallashtirilgan obyekt sifatida talqin qilish.

Shu o'rinda aytib o'tish zarurki, modellashtirish obyektni umumlashti- rish darajasiga ko'ra quyidagicha boiadi:

1. Lingvistik faktni tavsiflashga qaratilgan analitik model.

2. Oraliq model yoki toidiruvchi model.

3. Maksimal umumlashtirishga asoslangan sintezlovchi model.

Lingvistik model tushunchasi struktur tilshunoslikning E.Sepir,

L.Blumfild, R.Yakobson, N.Chomskiy, Z.Harris, Ch.Hokket kabi namoyan- dalari tomonidan kirib kelgan. Uning taraqqiyoti esa XX asrning 60 - 70- yillariga (matematik va kibernetik lingvistika rivojlana boshlagan davrga) to‘g‘ri keladi. Lingvistik modelni quyidagi turlarga ajratish mumkin:

1. Inson nutqiy faoliyati modellari. Bu modellar konkret nutq jarayonini va hodisalarini aks ettiradi. Masalan, aniq bir tovushning talaffuz modeli yoki nutqning yuzaga chiqish modeli.

2. Lingvistik tadqiqot modellari. Bunda muayyan til hodisalari asosida olib borilgan tadqiqot jarayonini aks ettiradi. Masalan, o‘zbek tilida mor­fologik usul asosida so‘z yasalishining umumiy modeli: asos + so‘z yasovchi qo‘shimcha; xususiy modellari: asos + -chi; asos + -dosh; asos + -do‘z kabi.

Poiatov A., Muhamedova S. Kompyuter lingvistikasi. - Toshkent, 2007. --

B.23.

32

A. Rahimov

3. Metamodellar - bunda lingvistik modellar saralanadi, u gipotetik- deduktiv xarakterga ega, o‘ta abstraktlashgan va ratsionallashgan bo‘Iadi.

Modellashtirish metodi ayrim tillarga, jumladan, ingliz tiliga faol tatbiq qilingan. 0‘zbek tilida sodda gap qurilishi:

S + О + V : Men kitob o'qidim. Men xat yozdim.

S = ega, О - to'ldiruvchi, V = kesim.

Bundan kelib chiqib aytish mumkinki, o‘zbek tilida qo'shma gapning eng kichik modeli quyidagicha boiadi:

S, + V, S2 + V2: Bahor keldi, ishlar qizib ketdi.

Ingliz, rus tillarida sodda gap qurilishi quyidagicha belgilangan:

S + V + O:

Я пишу диссертацию. Он читает книгу.

I have read the book. I am writing a research work.

Modellashtirish tilshunoslikda strukturalizm yo‘nalishi ta’sirida faol tatbiq qilina boshlandi. Gap strukturasini modellashtirish g'oyasi XX asrning 50- yillarida amerikalik tilshunos Charlz Friz tomonidan olg'a surildi. Olim o‘z qarashlarini umumlashtirib 1952-yilda «The Structure of English» nomli tadqiqotini yaratdi: Ch.Friz o‘z konsepsiyasini distributiv model deb nom- lagan. Unga ko‘ra, gap muayyan so‘z turkumlariga oid bo‘lgan so'zlar zan- jiri hisoblanadi va tahlilda morfologiya bazasiga tayaniladi. Masalan, « The young man painted the door yesterday» jumlasi distributiv model asosida quyidagicha tahlil etiladi: D 3 Ia 2-d D Ib 4. Bu yerda D - otning aniqlovchisi (inglizcha determiner), 3 - sifat, Г - birlikdagi ot, 2-d - o'tgan zamon shak- lidagi fe’l, 4 - ravishni anglatadi. Demak, mazkur modelda turli so‘z tur­kumlariga mansub so‘z shakllarining nutq zanjiridagi distributsiyasi (tar- qalishi, qurshovi) gap strukturasini modellashtirishning asosiy mezoni sana- ladi.

Shunga ko‘ra, so‘z turkumlarini modellashtirishda quyidagi shartli bel­gilar keng e’tirof etilgan.

So‘z turkumlari

I. Ot predmetning (u keng ma’noda tushuniladi, ya’ni jonli va jonsiz narsalarni anglatadi) nomini bildiradi. U kim, nima, qayer so‘roqlaridan biriga javob bo'ladi. Ot uchun N, ko‘plik shaklidagi ot uchun Ns, turdosh otlar uchun N , atoqli otlar uchun N , otli birikma uchun NP simvollari-

com7 1 prop

dan foydalaniladi.

II. Sifat qanday, qanaqa so‘roqlariga javob boiib, predmetning belgisi-

33 Бархударов JI.C. Структура простого предложения английского языка.

- М.: Высшая школа, 1966. - С. 18.

Kompyuter lingvistikasi asoslari

33

ni bildiradi. Sifat uchun Adj yoki A, sifatli birikma uchun AP simvollaridan foydalaniladi.

III. Son qancha, nechanchi so'roqlariga javob bo'lib, predmetning miq- dori va tartibini bildiradi. Son so‘z turkumi uchun Q simvolidan foydala­niladi.

IV. Fe’l predmetning harakat yoki holatini bildirib, nima qilgan, nima qilyapti, nima qilmoqchi so‘roqlariga javob beradi. Fe’l uchun V, o‘timli fe’l uchun Vt o'timsiz fe’l uchun V, fe’lningpredikativ (shaxsli shakli-finite form) shakli uchun Vp fe’lning nopredikativ shakli (shaxssiz shakli - non- finite form) uchun Vnp hozirgi zamon sifatdoshi yoki gerundiy shaklidagi fe’l uchun V o‘tgan zamon sifatdoshi yoki majhul nisbatdagi fe’l uchun Vcn, fe’lli birikma uchun VP, yordamchi fe’llar uchun aux, modal fe’llar uchun mod simvollaridan foydalaniladi.

V. Ravish ifodalayotgan harakat yoki holatning belgisi yoki turlicha vaziyatlarini bildiradi. Ravish asosan fe’lga, yana sifat yoki ravishga ham bog‘lanib, qay tarzda, qayerda, qachon va h.k. so'roqlarga javob boiadi. Ravish uchun Adv yoki D simvolidan foydalaniladi.

VI. Olmosh gapda ot, sifat yoki ravish o‘rnida qo'llanuvchi so‘z turku- midir. Olmosh uchun Pron simvolidan foydalaniladi.

VII. Artikl yordamchi so‘z bo'lib, otning maxsus belgilovchisi hisobla- nadi. Artikl barcha tillarda uchrayvermaydi, Art simvolidan foydalaniladi.

VIII. Predlog (old ko‘makchi) yordamchi so‘z bo'lib, otning gapdagi boshqa so‘zlarga nisbatan bo'lgan munosabatini (ya’ni makon, vaqt va sa- bab kabi munosabatlarni) ifodalaydi. Predlog uchun Prep yoki p simvolidan foydalaniladi.

IX. Bog'lovchi yordamchi so‘z bo'lib, gap bo'laklari yoki gaplarni bog'Iash uchun xizmat qiladi. Bog'lovchi uchun Conj simvolidan foydala­niladi.

X. Yuklama uchun Part simvolidan foydalaniladi.

XI. Undov so'zlar uchun Interj simvolidan foydalaniladi.

XII. Taqlid so'zlar uchun Mim simvolidan foydalaniladi.

Gap bo‘laklari

Ixtiyoriy tildagi yozma matn yoki shaxsning nutqi jumla deb ataluvchi alohida scgmentlardan iborat bo'ladi. Ular sodda va qo'shma (ya’ni sodda gaplardan tashkil topgan) gaplarga bo'linadi. Har qanday tilning gramma- tikasini o'rganish sodda gaplar bilan tadqiq etishdan boshlanadi. Har qan­day tildagi sodda gaplar tarkibida ega, kesim, to'ldiruvchi, aniqlovchi, hoi kabi gapning bosh va ikkinchi darajali boiaklari mavjud bo'lishi mumkin.

34

A.Rahimov

I. Ega - gapning zaruriy bo‘lagi bo‘lib, boshqa boiaklar bilan gramma- tik bogianadi, u predmet yoki shaxsni bildiradi. Ega kim, nima, qayer so\* roqlariga javob boiadi. Ega uchun S simvolidan foydalaniladi.

II. Kesim - gapning ham semantik, ham grammatik markaziy boiagi boiib, gapning yadrosi hisoblanadi, kesimsiz gap shakllanmaydi, u harakat, holat, xususiyat kabilarni anglatadi. Kesim nima qilyapti, nima qilmoqchi, nimadan tashkil topgan kabi so'roqlarga javob beradi. Kesim uchun P (ba’zan V) simvolidan foydalaniladi.

III. Toidiruvchi deb predmetni bildiruvchi va ma’no jihatidan kesimni toidiruvchi gap boiagiga aytiladi. Vositasiz (obyektsiz) toidiruvchi kimni, nimani soioqlariga, vositali toidiruvchi esa kimga, nimaga, kimda, nima- da, kimdan, nimadan, kim haqida, nima haqida va h.k. so‘roqlarga javob boiadi. Toidiruvchi uchun О simvolidan foydalaniladi.

IV. Aniqlovchi otning belgilarini tushuntiruvchi, izohlovchi gap boiagidir. Aniqlovchi predmetning belgisini bildirgan holda qanday, kim- ning, nechanchi, qancha kabi so‘roqlarga javob boiadi. Aniqlovchi komp­yuter lingvistikasida gap boiagi sifatida ajratilmaydi, chunki u gapda boshqa boiaklarga tobelanib keladi, u gap uchun nozaruriy (fakultativ, obligator emas) boiak hisoblanadi.

V. Hoi kesimda ifodalanayotgan harakat yoki holatning o‘rni, payti, maqsadi, sababi va bajarilish sharti kabilarni bildiruvchi gap boiagidir. U qanday, qay tarzda, qayerda, qachon, nima sababdan, nima uchun kabi so'roqlarning biriga javob beradi. Hoi uchun M simvolidan foydalaniladi.

Modellashtirish metodi keyingi paytlarda tilshunoslikka faol tatbiq qili- na boshlandi. Bunda olimlar mazkur metodning bir qator afzalliklari va pragmatik jihatlarini nazarda tutadilar. Buni quyidagicha izohlash mumkin:

- birinchidan, modellashtirish amaliy metod hisoblanadi, tavsifiy emas;.

- ikkinchidan, modellashtirish metodi har qanday sharoitda optimal («eng qulay», «eng maqbul» degan ma’nolarni ifodalaydi) hisoblanadi;

- uchinchidan, modellashtirish metodi ekonomiya prinsipiga tayanadi. Bunda so'z, uzun ta’rif va tavsiflarga ehtiyoj boimaydi;

- to‘rtinchidan, obyektni tushuntirish va izohlashni osonlashtiradi va soddalashtiradi.

Borliqdagi hamma narsalarning ijobiy va salbiy tomonlari boiganidek, modellashtirish metodining ham pozitiv va pragmatik tomonlari bilan bir- ga salbiy tomonlari ham bor. Uning kamchiliklari, salbiy tomonlarini qu­yidagicha umumlashtirish mumkin:

-birinchidan, modellashtirishda faqat struktur va formal belgilarga asos- laniladi. Bunda mazmuniy tomon, semantik qirralar e’tibordan chetda qola-

Kompyuter lingvistikasi asoslari

35

di. Vaholanki, har qanday hodisaning, xususan, lisoniy hodisaning mohiyati shakl va mazmun birligida o‘z ifodasini topgan bo‘ladi;

- ikkinchidan, bizning obyekt haqidagi bilimlarimiz rivojlanib borgan sari modellar eskiradi. Avvalgi bilimlarni inkor qilish hisobiga fan rivojla- nadi. Mavjud modellar bilimlarimiz ufqini cheklaydi va xato tasavvurlarga olib kelishi mumkin;

- uchinchidan, tabiat va jamiyatdagi hech bir narsa boshqa obyektlar ta’siridan xoli boMmaydi. Ya’ni modellar nisbiylik xarakteriga ega, biz ularni mutlaq haqiqat sifatida qabul qila olmaymiz.

Kompyuter lingvistikasida eng ko‘p qo‘llanilayotgan metodlardan biri analogiya metodi hisoblanadi. Analogiya (yunoncha dva^oyia so‘zidan olingan, «moslik», «o‘xshashlik» ma’nosini anglatadi) - munosabatlarning tengligi, muvofiqligi; hodisa va jarayonlarning muayyan xossalarini qiyos- lash yo‘li bilan anglashda o‘xshash jihatlari. Qiyoslanayotgan obyektlar o‘rtasida o‘xshashlik va farqlanish kuzatiladi, bu qiyoslashning asosi sana- ladi. Bunda ma’lum obyektga xos bo‘lgan aniq belgining qiyoslanayotgan obyektlarda kuzatilishiga tayanib ularning identifikatsiyasi asosida mush- tarak jihati aniqlanadi. Masalan, sifat va ravish so‘z turkumiga mansub so‘zlar berildi. Ular bir qarashda bir-biriga juda yaqin, so‘roqlari o‘xshash, shuningdek, farqli tomonlari ham mavjud. Shunda analogiyaga tayanib berilgan so'zlarning mushtarak jihatlarini ajratish mumkin. Yoki ko‘p odam- lar «singan oyna - baxtsizlik, noxitsh voqealarni keltirib chiqaradi», degan irimga ishonadi. Bu irimning kelib chiqishi ham qandaydir mistik asosga ega emas, balki analogiyaga asoslangan. Ya’ni singan oyna insonni xunuk aks ettiradi, mazkur tashqi o‘xshashlik asosida tasvirning buzilishi undan foydalanuvchi uchun ham shunga mos yomon oqibatlarni keltirib chiqara­di, degan xulosaga olib kelgan. Shu o‘rinda ta’kidlash joizki, analogiya induksiya (umumiylikdan xususiylikka borish; qoidadan misolga borish) va deduksiya (xususiylikdan umumiylikka borish; misoldan qoidaga borish) metodlari bilan uzviy aloqadorlikda boladi. Bilish jarayonida ular o'rtasiga qat’iy chegara qo‘yib bo‘lmaydi. Analogiya deduksiya va induksiya o‘rtasida bogiovchi «halqa» vazifasini o‘taydi.

Bilishda analogiya metodining ahamiyatini mashhur matematik Dyord Poya shunday izohlaydi: «Analogiyaga asoslanmasdan na elementar, na oliy matematikada, na fanning boshqa sohalarida hech qanday kashfiyot qilib bo'lmaydi». Stefan Banax esa shunday yozadi: «Matematik isbotlar о ‘rtasidagi analogiyani topa olgan odam, eng yaxshi matematik - isbotlar о ‘rtasida analogiya о 'rnata olgan odam, bandan ham ко 'ra kuclili matema­tik - nazariyalar analogiyasini sezgan odam; biroq odam о ‘zini eng kuchli

36

A. Rahimov

hisoblashi uchun analogiyalar o'rtasidagi analogiyani ko'ra olishi kerak». Ushbu mulohazalardan kelib chiqqan holda aytish mumkinki, analogiya lingvistikadagi, xususan, kompyuter lingvistikasidagi tadqiqotlar uchun ham o‘ziga xos «kompas» vazifasini bajaradi.

Akademik L.V.Sherba o'zining bir ma’ruzasida yozuv taxtasiga shun- day jumlani yozadi: «Глокая куздра штеко будланула бокра и кудрячит бокренка». Talabalardan uning grammatik tahlilini talab qiladi. Mazkur jumla aslida hech qanday mazmunga ega emas, lekin u rus tilidagi mazmun- ga ega boigan qaysidir jumlani eslatadi. Bu jumlaning grammatik tahlilini amalga oshirish mumkin. Masalan, «куздра» so‘zi - birlik, bosh kelishik, jenskiy rod shaklidagi ot bilan ifodalangan ega, «будланула» so‘zi - o‘tgan zamon shaklidagi fe’l bilan ifodalangan kesimlardan biri, «кудрячит» so'zi

- noaniq shakldagi fe’l bilan ifodalangan boshqa bir kesim. Masalaga bun- day yondashish analogiya metodi asosida bo'ladi. Mazkur yondashuv, ay- niqsa, kompyuter lingvistikasidagi matnning avtomatik analizi yo‘nalishida muhim ahamiyat kasb etadi. Bunda so'zlarning grammatik tavsifi va so'zlarning oxiridagi harf tarkibi o'rtasidagi kuchli korrelatsion aloqaga tayaniladi.34 Ya’ni oxirgi harf tarkibi bir xil bo'lgan so‘zlarning derivatsion va relatsion modellari (so‘z yasalishi va shakl yasalishi) hamda grammatik ma’lumotlari (qaysi so‘z turkumiga mansubligi, qanday grammatik kate- goriyalarga egaligi) bir xil boiishi tabiiy. Masalan, ingliz tilida -er, -ness, - dom kabi, rus tilida -пость, -щик, -ник kabi, o'zbek tilida -shunos, -chi, -lik qo‘shimchalarga ega bo'lgan so‘zlar yuqori ehtimollik bilan ot (noun) so‘z turkumi hisoblanadi. Demak, tezaurusda (lug‘at bazasida) mavjud bo'lmagan so'zlarning avtomatik analizini mavjud bo'lganlarining harf tarkibidagi moslikdan, so‘zlar analogiyasidan kelib chiqqan holda amalga oshirish mumkin. Bu jarayonda tizimning lingvistik ta’minoti chappa (ters) lug'atlar bilan ham to‘ldirilgan bo'lishi lozim.

Topshiriqlar:

1. Lisoniy model va matematik modelning uzviy bog‘liqligi nimada?

2. Tilshunoslikning qaysi jabhalarini modellashtirib bo‘lmaydi va nega?

3. So‘z yasash modellariga turli tillardan misollar keltiring.

4. Gaplarni modellashtirishda nimalar e’tiborga olinishi lozim?

5. Aniqlovchi nega alohida gap boiagi maqomiga ega emas?

34 Белоногов Г.Г. Об использовании принципа аналогии при автоматической обработке текстовой информации // Проблемы кибернетики. 1974. №28.-С.ЗЗ.

Kompyuter lingvistikasi asoslari

37

Adabiyotlar:

1. Новое в зарубежной лингвистике. Вып.XXIV. Компьютерная лингвистика. - М., 1989.

2. Шемакин Ю.И. Начало компьютерной лингвистики. - М.: МГОУ, 1992.

3. Кодухов В.И. Методы лингвистического анализа. - JI., 1963.

4. Grishman R. Computational Linguistics // Cambridge University Press. 1994.

5. Gelbukh A. Computational Linguistics and Intelligent Text Process­ing. International Conference. - Mexico, 2003.

6. Гладкий А.В., Мельчук И.А. Элементы математической лингвистики. - М.: Наука, 1969.

7. Polatov A., Muhamedova S. Kompyuter lingvistikasi. - Toshkent, 2007.

8. Бархударов JI.C. Структура простого предложения английского языка. - М.: Высшая школа, 1966.

FORMAL GRAMMATIKA NAZARIYASI

Reja:

1. Grammatika haqidagi lingvistik qarashlar.

2. Grammatikaning turlari (tavsifiy, formal, transformatsion gram­matika).

3. Formal grammatika nazariyasi.

Tayanch so‘z va iboralar; grammatika, morfologik sistema, sintaktik siste- ma, germenevtika, preskriptiv grammatika, sinxroniya, diaxroniya, gram- matik modellar, ikki cho ‘qqili nazariya, bir cho 'qqili nazariya, verbosent- rizm, tobelik daraxti, bevosita ishtirokchilar modeli, tema, rema, formal gram­matika, transformatsion grammatika, operand, transforma, terminal va no­terminal simvollar, formallashtirish.

Grammatika tabiiy tillar strukturasi haqidagi fan bo‘lib, u morfologik kategoriya va shakllar, sintaktik ketegoriya va konstruksiyalar hamda so‘z

38

A.Rahimov

yasash usullari tizimini tashkil etadi. Til qurilish birliklarining paradigma- tik va sintagmatik munosabatga kirishuvi (vertikal va gorizontal yo‘nalishda) muayyan qoidalar asosida yuz beradi. Ushbu qoidalar yigindisi tilning gram- matik qurilishi va grammatik sistemasi deb yuritiladi. Grammatika ikki yirik sistemani o‘z ichiga oladi:

1. Morfologik sistema - so‘z formalari, grammatik shakllar, gramma­tik kategoriyalar tizimi.

2. Sintaktik sistema - sintaktik kategoriyalar va konstruksiyalar, so‘zlarning o'zaro birikish qonuniyatlari.

Grammatika yunoncha so‘z bo‘lib, «o‘qish va yozish san’ati» ma’nosini bildiradi. Bu tushuncha dastlab xat-savod, imloni o‘rgatuvchi fan ma’nosida ishlatilgan. Keyinroq grammatika tilning fonologik, morfologik, sintaktik va semantik belgilarini tavsiflovchi til qoidalari majmuasi ma’nosiga ko'chgan. Grammatika haqidagi ilk qarashlar falsafa, mantiq, germenevti- ka (diniy matnlarni sharhlash, talqin etish) qobig‘ida boigan. Bunda tilning paydo boiishi, til va tafakkur munosabati, lingvistik belgi motivatsiyasi, diniy matnlar mazmunini to‘g‘ri tushunish bilan bog‘liq masalalar o'rganilgan, til strukturasi esa yetarli darajada tadqiq etilmagan. Faqatgi- na miloddan avvalgi II - I asrlarda shakllangan Aleksandriya grammatika maktabining eng yirik vakili Dionisiy Traks (miloddan avvalgi 170 - 90- yillarda yashagan) o‘zigacha boigan tadqiqotchilarning ishlari va tajriba- larini o‘rganib, ulardan foydalanib rimliklar uchun «Grammatika san’ati» («Grammatike techne») nomli sistemalashtirilgan dastlabki yunon gramma- tikasini yaratdi. Shundan buyon grammatika falsafadan mustaqil soha si- fatida ajralib chiqdi.35

Antik davrda grammatika preskriptiv (to‘g‘ri gapirish, to‘g‘ri jumla tu- zishni ko‘rsatuvchi yo‘riqnoma) xarakterga ega boigan. XVIII asr oxirla- rida grammatikani tushunish keskin o‘zgardi. Bu davrda bobo til konsepsi- yasini ilgari surgan qiyosiy-tarixiy grammatika shakllandi. Unda yevropa tillarining qadimgi sanskrit tili bilan tarixiy aloqasini aniqlash yordamida hind-yevropa tillari shajarasini rekonstruksiya qilish bo‘yicha qiyosiy aspekt- da tadqiqotlar olib borildi.

XX asr boshlarida Ferdinand de Sossyur grammatikani tavsiflashda sinxroniya va diaxroniyani farqladi. Bunda grammatikani tarixiy (diaxron) o'rganishni emas, balki sinxron (muayyan vaqtdagi holati) tadqiq qilish lozim, deb hisobladi. XX asrning 50-yillarida Noam Chomskiy grammati­kani tavsiflashning yangi metodini tavsiya etdi. Bu tilshunoslikda «formal

35 Сусов И.П. История языкознания. - Тверь: Тверский гос. ун-т. 1999. -

С.24 - 25.

Kompyuter lingvistikasi asoslari

39

grammatika nazariyasi» deb ataladi. Olimning tadqiqot natijalari komp­yuter lingvistikasining shakllanishiga asos bo‘ldi. N.Chomskiyning formal grammatika haqidagi qarashlari «Syntactic Structures» (1957) hamda «As­pects of the Theory of Syntax» nomli kitoblarida yoritilgan.

Grammatikaga bo‘lgan yondashuvlar turlicha bo'lganligi bois tilshu- noslik tarixida turli grammatikalar yaratilgan. Jumladan, nazariy gram­matika va amaliy grammatika; an’anaviy (klassik) grammatika va noklassik grammatika; sinxron va diaxron grammatika; xususiy va umumiy (univer­sal) grammatika; kategorial grammatika, tobelik grammatikasi, leksik-funk- sional grammatika, konstruktiv grammatika, transformatsion grammatika, kontekstdan xoli grammatika (context-free grammar) kabi.

Grammatik tavsifning eng sodda (primitiv) formal modeli gap bo‘laklariga ajratish modeli sanaladi. Tabiiy tillar grammatikasida sodda gap ikki qismdan - ega va kesimdan tashkil topgan sintaktik qurilma sifati­da tavsiflanadi. Ega nutqning predmetini, kesim esa uning belgisini ko'rsatadi. Ega va kesim alohida so'zlar bilan ham, so'zlar guruhi bilan ham ifodalanishi mumkin. Agar ega va kesim so'zlar guruhi bilan ifodalan- sa, sodda gap ikki struktur elementga ajratiladi: ega guruhi (noun phrase) va kesim guruhi (verb phrase).

Ega guruhida asosiy mazmun ifodalovchi unsur ot so‘z turkumi bilan ifodalangan ega yoki otlashgan so'zlar (substantivatsiya), ega guruhidagi boshqa so‘zlar esa eganing aniqlashtiruvchilari vazifasida keladi. Ega- ning aniqlashtiruvchilari ot, sifat, son, olmosh, kvantifikatorlar (miqdor bildiruvchi so'zlar), determinatorlar (artikllar) orqali ifodalangan aniqlov- chi gap bo‘lagidir. Kesim guruhida asosiy mazmun ifodalovchi unsur fe’l- kesim hamda ot-kesim hisoblanadi, shuningdek, turli so‘z turkumlari bi­lan ifodalangan to'ldiruvchi va hoi kesimning aniqlashtiruvchilari bo‘lib keladi. Mazkur modelda ega va kesim gapning bosh boiaklari, gapning semantik-sintaktik yadrosi, aniqlovchi, to'ldiruvchi, hoi esa ikkinchi dara- jali bo'laklar deb nomlanadi. Ushbu talqin mumtoz sintaksis tamoyillari- ga muvofiq keladi.

XX asrning ikkinchi yarmida struktur sintaksis, semantik sintaksis yo'nalishlari paydo bo'lishi bilan gapning sintaktik tuzilishiga, gap bo'laklariga bo'lgan yondashuv ham o'zgardi. Yuqoridagi ikki cho‘qqili nazariyaning o'rniga bir cho‘qqili nazariya maydonga keldi. Unga ko'ra, faqat kesimgina gapning grammatik yadrosi sifatida e’tirof etildi. Ushbu yondashuv fransuz tilshunosi Lyusen Tenyerning verbosentrizm (lotincha «verbum» - «fe’l», «kesim» ma’nolarini anglatadi, «centrum» - «markaz», unga ko'ra, gapning markazi gap bo'lagi sifatida kesim, so'z turkumi sifati-

40

A. Rahimov

da fe’ldir) konsepsiyasi bilan bogiiqdir.-16 Ingliz tilshunosi O.Yespersen mazkur konsepsiyaga qarshi o'laroq gapdagi mutlaq hokim bo'lak sifatida egani e’tirof etgan.” Gap bo'laklari iyerarxiyasida egani markazga qo'yish tendensiyasi mantiq fanidan kelib chiqqan. Chunki mantiqda ism (subyekt) substansiyani (mohiyatni), fe’l (predikat) aksidensiyani (hodisani) ifodalay- di, degan qarash mavjud. Bunda ega gapning tub mohiyatini aks ettiruvchi birlamchi bo'lak, kesim esa egani to'ldiruvchi, eganing turli belgilarini ko'rsatuvchi ikkilamchi bo'lak sifatida talqin etiladi.

L.Tenyer g'oyalari ta’sirida shakllangan tobelik grammatikasi model- lari kompyuter lingvistikasining avtomatik analiz dasturlarida ishlatiladi. Bu model tobelik daraxti (shajara) deb ataladi. Bunda gap asosan daraxt ko'rinishida tasvirlanadi, jumla tarkibidagi so'zlar bir-biri bilan tobelik munosabatida bo'ladi (bevosita dominatsiya munosabati). Kesim daraxt- ning ildizida joylashgan bo'ladi va dominant bo'lak hisoblanadi, boshqa so'zlar (gap bo'laklari) unga bevosita yoki bilvosita (boshqa so'zlar yorda- mida) tobelangan bo'ladi.

Tilshunoslikda an’anaviy gap bo'laklariga ajratishdan tashqari, yana gapni kommunikativ nuqtayi nazardan qismlarga ajratish ham mavjud, u gapning aktual bo'linishi deb ataladi. Bunda gap tema va rema qismlariga ajratiladi. Bunda tema nutqiy muloqotning bizga ma’lum qismiga nisbatan ishlatiladi, rema esa muayyan narsa, voqea-hodisa, jarayon kabilar haqida- gi yangi informatsiyani anglatadi. Ko'pincha tema ega guruhi orqali, rema esa kesim guruhi orqali ifodalanadi. Ba’zan buning aksi bo'lishi ham mumkin. Masalan, «Botirjon universitetda ishlayapti» gapida ega vazifasi- da kelgan Botirjon tema hisoblanadi, chunki nutq vaziyati uchun Botirjon aniq va avvaldan tanish bo'lgan shaxsdir. Gapda kesim guruhini tashkil etgan universitetda ishlayapti remadir. Demak, bunda Botirjonning univer­sitetda ishlayotganligi fakti nutq vaziyati uchun yangi xabardir, shuning uchun u rema hisoblanadi.

Tobelik daraxti (shajara) modeliga eng yaqin model bevosita ishtirokchi- lar (immediate constituents) sintaktik modeli hisoblanadi. Bevosita ishtirokchi- lar metodiga ko'ra, tayanch nuqta konstruksiya hisoblanadi. Konstruksiya tarkibidan ishtirokchilar va bevosita ishtirokchilar ajratiladi.18 Konstruk-

\* Теньер Л. Основы структурного синтаксиса. - М.: Прогресс, 1988. -

С.12.

” Есперсен О. Философия грамматики. - М.: Иностранная литература, 1958.-С.164- 165.

18 Nurmonov A. Lingvistik tadqiqot metodologiyasi va metodlari. - Toshkent: Akademnashr, 2010. - B.88.

Kompyuter lingvistikasi asoslari

41

siya atamasi ostida ma’noli qismlarning ketma-ket munosabatidan tashkil topgan butunlik tushuniladi. Ishtirokchilar deb kattaroq konstruksiya tarki- biga kirgan so'z yoki konstruksiyaga aytiladi. Muayyan konstruksiyaning bevosita shakllanishida ishtirok etgan bir yoki bir nechta ishtirokchilarga bevosita ishtrokchilar deyiladi. Masalan, A 'lochi talabalar о 'z vazifalarini aniq va puxta bajaradilar. Bu jumla konstruksiya hisoblanadi, uning tarki- bidagi barcha ma’noli so'zlar («va» bog'lovchisidan tashqari) ishtirokchi­lar, bevosita aloqaga kirishgan so'zlar bevosita ishirokchilar deyiladi. Bun­da a’locki so‘zi talabalar so‘zi bilan, aniq, puxta so'zlari bajaradilar so‘zi bilan, o‘z so‘zi vazifalarini so‘zi bilan o'zaro zich bog'langan. Demak, be­vosita ishtirokchilar sintaktik modelida dastlab so'z birikmalari ajratiladi, so'ngra ular gap ichida birlashtiriladi.

O'zbek tilshunosligi darg'alaridan biri, o'tkir nigohli olim, professor

A.Nurmonovning fikricha, mazkur modelda nutqning har bir mustaqil par- chasi ikki qismdan tashkil topadi. O'z navbatida, har qaysi qism yana o'z ichida ana shunday qismlardan iborat boiadi. Ko'rinadiki, har bir konst­ruksiya binar tamoyilga ko'ra qismlarga bo'linadi va bu bo'linish konstruk­siya doirasida graduallik (darajalanish) xususiyatiga ega boiadi.-19

Bevosita ishtirokchilarga ajratish modeli tanqidiy idrok etilib, undagi ayrim g'oyalar asosida birmuncha mukammalroq va formallashtirishning negizi hisoblangan generativ va transformatsion grammatika (generative and transformational grammar) maydonga keldi. Transformatsion gram­matika 1950-yillarda dastlab Z.Harrisning til strukturasidagi transfor- matsiya xususidagi qarashlari ta’sirida yuzaga kela boshladi, bu yo'nalish Z.Harrisning shogirdi N.Chomskiy tomonidan rivojlantirildi. Unga ko'ra, dastlabki yadro strukturalar birikma strukturalar qoidalar tizimi orqali tug'iladi. Ya’ni har qanday til grammatikasi yadro strukturalar (kernel struc­tures) yig'indisidir. Yadro gaplar deb sodda, yig'iq, darak gaplar tushu- niladi.40 Ingliz tilida yadro gaplar sifatida quyidagi konstruksiyalar ajratib ko'rsatiladi41 (4-jadval).

Mazkur konstruksiyalarda N (Ot) turli xil determinatorlar (masalan, ar- tikl, olmosh) hisobiga, V (Fe’l) ravishlar, adverbial (ravishli) birikmalar hisobiga kengayishi mumkin. Ushbu yadro konstruksiyalarning bir-biri bi-

,9 Nurmonov A. Struktur tilshunoslik: ildizlari va yo'nalishlari. - Toshkent, 2008.-B. 124.

40 Nurmonov A., Iskandarova Sh. Tilshunoslik nazariyasi. - Toshkent: Fan,

2008.-B.140- 141.

41 Херрис З.С. Совместная встречаемость и трансформация в языковой структуре // Новое в лингвистике. Вып.11. - М., 1962. - С.627 - 628.

42

A. Rahimov

4-jadval

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. | NV | Noun Verb | Otfe’l | John came |
| 2. | NVpN | Noun Verb Preposition Noun | Ot fe'l predlogli ot | John looked at Mary |
| 3. | NVN | Noun Verb Noun | Ot fe’l ot | John saw Mary |
| 4. | N is N | Noun is Noun | Ot is ot | John is a teacher |
| 5. | N is р N | Noun is Preposition Noun | Ot is predlogli ot | John is in bed |
| 6. | N is D | Noun is Adverb | Ot is ravish | John is out |
| 7. | NisA | Noun is Adjective | Ot is sifat | John is angry |

lan yoki turli transformatsiyalar asosida birikishi natijasida ingliz tilida boshqa turdagi gaplar hosil qilinishi mumkin. Masalan, The old man saw a black dog there jumlasi uchta yadro strukturaning yig'indisi sanaladi: 1) The man saw a dog there; 2) The man was old; 3) The dog was black.

Generativ grammatikada yadro strukturalardan (operandlar) turli shak- liy o'zgarishlar asosida asosiy ma’noni saqlagan holda ikkilamchi sintaktik qurilmalarning hosil qilinishi (transformatsiyalar = transformalar), bu jara- yondagi vositalar (operatorlar) va transformatsiya hosil qilish modellari tad- qiq etiladi.42 Masalan, Salim keldi - Salimning kelishi - Salim kelgach - Salim kelganda - Salim kelishi bilan...; Anvar xatni yozdi - Xat Anvar tomonidan yozildi. Ushbu misollarda transformatsiya uchun asos bo‘lgan gap (Salim keldi yoki Anvar xatni yozdi kabi gaplar) operand, transformatsiya natijasi bo'lgan hosila gap (Salim kelgach, Salim kelib yoki xat Anvar to­monidan yozildi kabi sintaktik strukturalar) transforma yoki transformand, operanddan transformandni hosil qiluvchi vosita (ravishdosh, sifatdosh, harakat nomi yoki majhul nisbat shakllari) transformatsiya operatori deb ataladi.

Transformatsion grammatika matematik lingvistika va kompyuter lingvistikasi fanining rivojlanishida muhim ahamiyat kasb etadi. Birinchidan, transformatsion grammatika tabiiy tillarning grammatikasini formallash- tirish nazariyasi negizlarini yaratishga asos bo‘ldi. Ikkinchidan, transfor­matsion grammatika avtomatik tarjima tizimining nazariy asosi sanaladi. Transformatsiya va derivatsiya jarayonlarining tabiiy tillarga faol tatbiq etilishi avtomatik tarjima samaradorligini ta’minlaydi. Uchinchidan, trans­formatsion grammatika til o‘qitishni modellashtirishda ham muhim aha-

42 Qarang: Хомский H. Синтаксические структуры // Новое в лингвистике.

Вып.II. - М., 1962. - С.412 - 527.

Kompyuter lingvistikasi asoslari

43

miyatga molik.43 To‘rtinchidan, transformatsion grammatika tabiiy tillarda- gi paradigmatik va sintagmatik munosabatlarning me’yoriyligi va gram- matik to‘g‘ri va noto‘g‘ri jumlalarni ajratishning aniq mezonlarini belgilaydi.

N.Chomskiy tavsiya etgan grammatikada gaplarni «tug'dirish» (hosil qilish) uchun o'rin almashtirish qoidalaridan foydalaniladi. Chap qismda gapning struktur komponenti simvollari joylashadi, o‘ng qism esa ularning simvollari o'rniga almashuvchilar yoki simvollar zanjiri (birikuvi)dan ibo- rat bo‘ladi. Quyidagi jadvalda ingliz tili uchun ayrim gapning struktur kom­ponenti simvollari ko‘rsatilgan:

5-jadval

|  |  |  |
| --- | --- | --- |
| № | Simvollar | Simvollarbirikuvi |
| 1 | S - boshlang'ich simvol, u inglizcha «Sentence» («gap») so‘zining bosh harfidan olingan. | S = NP + VP |
| 2 | VP - kesim guruhining simvoli, u inglizcha «Verb Phrase» birikmasidan olingan. | VP = Verb + NP |
| 3 | NP -ega guruhining simvoli, u inglizcha «Noun Phrase» birikmasidan olingan. | NP = Det + N |
| 4 | Verb - kesim gap bo'lagining simvoli, u inglizcha «Verb» so'zidan olingan. | V = Aux + V |
| 5 | Det - artiklning simvoli, u inglizcha «Determiner» («aniqlagich») so'zidan olingan. | Det = the, a, ... |
| 6 | N -ot so‘z turkumining simvoli, u inglizcha «Noun» so'zidan olingan. | N = book, computer... |
| 7 | Aux - yordamchi (ko'makchi) fe'llarning simvoli, u inglizcha «Auxiliary» so'zidan olingan. | Aux = will, can ... |
| 8 | V — fe’l so'z turkumining simvoli, u inglizcha «Verb» so'zidan olingan. | V = see, know ... |

Jadvaldagi 5 - 8-qatorlarda ko‘rsatilgan simvollar (Det, N, Aux, V) ter­minal simvollar deyiladi, chunki ular lug‘atdagi aniq so'zlardir. 1 - 4-qa- tordagi simvollar esa noterminal simvollar deb ataladi.

Gap hosil qilish jarayoni noterminal simvollarni o‘ng qismga davomiy o'rin almashtirish asosida yuz beradi. Bunda boshlang'ich simvol S o'ziga mos keluvchi simvollar birikuvi NP hamda VP ga almashtiriladi. So‘ng NP va VP simvollari ham, o‘z navbatida, Det + N hamda Verb + NP ga almash­tiriladi. Bu jarayon barcha simvollar noterminal bo‘lguncha davom eta-

45 Шаумян С.К. Теоретические основы трансформационной грамматики // Новое в лингвистике. Вып.II. - С.408.

44

A. Rahimov

veradi. Masalan, ingliz tilidagi quyidagi jumlani bevosita ishtirokchilar modeli asosida tahlil etaylik. Birinchi bosqichda jumla NP va VPga ajrati- ladi: The old man | saw a black dog there - Keksa odam и yerda qora itni ко ‘rdi. Keyingi bosqichda yuqoridagi ajralish, o‘z navbatida, yana ikki qism- ga ajratiladi: The || old man \ saw a black dog || there. So'nggi bosqichda jumla bevosita ishtirokchilar modeli asosida «daraxt» («sentence tree») ko‘rinishida tasvirlanishi mumkin:44

Sentence

NP

VP

NP’

The old

i

N

man

VP’

\_\_l

D

NP

NP’

 i

N

saw a black dog there

Context-free grammar - matndan xoli bo‘lgan grammatika, chunki un- dagi qoidalar universal hisoblanib, ma’lum matnga tayanib tahlil etishga asoslanmaydi, balki u avvaldan empirik tarzda belgilab olinadi. Hozirda ushbu grammatika sun’iy programmalash tillariga (Pascal, C++, Delphi, HTML) faol tatbiq etilgan, shuningdek, tabiiy tillarni formallashtirish jara- yonida ham grammatikani bayon qilishning eng optimal usuli hisoblanadi. Masalan:

< SENTENCE> : : = < SUBJECT> < VERBPHRASE >

<SUBJECT> :: = John I Mary

< VERB PHRASE > :: = < VERB> < OBJECT>

<VERB> :: = eats I drinks

<OBJECT> :: = wine I cheese

We could represent the derivation of the sentence «Mary eats cheese» as follows:

< SENTENCE> = < SUBJECT> < VERB PHRASE >

=Mary <VERBPHRASE>

44 Бархударов JI.C. Структура простого предложения английского языка.

- М.: Высшая школа, 1966. - С.21.

Kompyuter lingvistikasi asoslari

45

=Mary <VERB> <OBJECT>

=Mary eats <OBJECT>

=Mary eats cheese

< SENTENCE> =:> < SUBJECT> < VERB PHRASE >

=:> < SUBJECT> < VERB> < OBJECT>

=:> <SUBJECT> eats <OBJECT>

=:> Mary eats <OBJECT>

=:> Mary eats cheese.

Formallashtirish - bu bilimlar majmuasini, mazmunini muayyan shakl- lar yordamida sun’iy til belgilari orqali ifodalashdir. Grammatikani for­mallashtirish tendensiyasi XX asrning o'rtalarida kuchaydi. Bu bevosita tilning strukturasini yanada aniq parametrlarda ifodalashga intilish, uni kompyuterga moslashtirishga urinish bilan bog‘liqdir. Formal grammati- ka nazariyasi matematik mantiq, matematik linvistikaning rivojlanishi natijasida yuzaga kelgan. Grammatikani formallashtirish (lotincha «for­ma» - «shakl», «tashqi ko‘rinish») konsepsiyasining asoschisi amerikalik tilshunos Noam Chomskiydir. Formal grammatika abstrakt simvollar orqa­li bayon etiladigan grammatika hisoblanib, unda til qoidalari, so‘z yasa­lishi, birikma va gap qurilishi turli xil lingvistik modellar, struktur sxema- lar yordamida tavsiflanadi. Formal grammatika 3 ta birlik asosida ish ko'radi:

1. Word-classes - so‘z turkumlari (tub va yasama so'zlar).

2. Phrases - gap bo'laklari.

3. Sentences - gap konstruksiyalari sxemalari, modellari.

N.Chomskiy qoidalarning ifodalanish turiga qarab formal grammati­kani 4 tipga ajratgan va bu Chomskiy iyerarxiyasi deb nomlanadi:

1.0 grammatika (unrestricted phrase-structure grammar) -juda murakkab generativ grammatika, bunda G=(N,”, P, S) belgilari yordamida ish ko'riladi. N, “ - alifbo (N - noterminal simvollar, u grammatikadagi sim- vollarni o‘z ichiga oladi; “ - metatilning terminal simvollari, u tildagi so‘zlarni qamrab oladi); S - noterminal to'plamning boshlang‘ich simvoli, P - qayta ishlab chiqilgan qoidalar. Bunda terminal va noterninal simvol- larni birlashtiruvchi V belgisi qabul qilingan, u inglizcha Vocabulary (lug'at) so'zidan olingan (V=TWN). Bu grammatika o‘ta abstraktlashgan matema­tik modellarga, informatika nazariyasiga tatbiq etilgan.

2.1 grammatika (context-sensitive grammar) - kontekstga bogiiq gram­matika bo‘lib, bunda simvollar zanjiri kontekst bilan aniqlanishi mumkin. Bu grammatika tabiiy tillar unsurlarining generatsiyasida faol qo'llaniladi.

3.2 grammatika (context-free grammar) - kontekstga bog‘lanmagan gram-

46

A. Rahimov

matika bo'lib, u programmalashtirish tillari unsurlarining generatsiyasida (ifodalar, buyruqlar tizimida) ishlatiladi.

4. 3 grammatika (regular grammar) - regular grammatika deb ataladi, u juda sodda va cheklangan grammatika bo‘lib, dining sodda unsurlari uchun ishlatiladi (miqdor, konstantalar, o‘zgaruvchilar uchun).

Formal grammatikada 2 ta tushuncha asosiy: a) generation - tug‘dirish, hosil qilish, порождение; b) recognition - tanish, bilish, распознавание. Shunga ko‘ra, grammatika 2 ga bolinadi: 1) tug‘dirish grammatikasi (gene­rative grammar) - tog‘ri gaplarni qabul qilish protsedurasi; 2) tushunuvchi, tanuvchi grammatika (recognizing grammar) - to‘gri gaplarni tushunish protsedurasi.

Demak, grammatika tilning asosida yotuvchi shunday ichki strukturaki, uni har bir til tashuvchisi intuitiv ravishda his qiladi va undan g‘ayrishuuriy tarzda foydalanadi. Sir emaski, har bir inson o‘z ona tili grammatikasini bolaligidanoq o'zlashtiradi. Bu jihatlar esa grammatikani formallashtirish masalasi juda murakkab ekanligini tasdiqlaydi. Grammatikani ideal dara- jada tavsiflovchi qoidalar sistemasining mavjud emasligi, har qanday tavsiya etilayotgan model nomukammal ekanligi xususida mashhur tilshunos olim K.Fossler shunday yozadi: «Grammatikani fonetika, morfologiya va sintak- sisga bo ‘lish hech kimga sir emas. Tilni uning shakllanish jarayoni ichida emas, balki uning o‘z holati ichida o'rganish zarur. Uning ustida anatomik operatsiya о ‘tkaziladi. Jonli nutq gaplarga, gap bo ‘laklariga, so ‘zlarga, bo ‘g ‘inlarga va tovushlarga ajratib chiqiladi. Ushbu metod to ‘liq о ‘zini oqlay- di va ahamiyatga molik natijalarga olib borishi mumkin, biroq shu bilan birga xatolarning manbai hisoblanishi ham mumkin. Xatolar ajratilgan bo'linishning nutq organizmi uchun asos bo'lishini da’vo qilingan paytda boshlanadi, aslida bu bo'linish insonning ixtiyoriga bog'liq ravishda mexa- nik va majburiy tarzda parchalash edi. Mazkur holat anatomiyadagi holatni eslatadi: anatom о 'zining yorish, kesish ishlarini albatta о ‘zboshimchalik bilan asossiz amalga oshirmaydi, и o'ziga birmuncha qulaylik tug'diruvchi joylarni tanlab oladi. Xuddi sliuningdek grammatist ham nutqni tovush, bo 'g ‘in, asos, qo 'shimcha kabilarga ajratislida tabiiy bo ‘Imagan, lekin о ‘ziga qiday yo ‘Idan boradi».\*5

Yuqoridagilardan kelib chiqib aytish mumkinki, shu kungacha olimlar tomonidan ishlab chiqilgan barcha grammatik modellar o‘ta shartli va nis- biydir. Keltirilgan barcha modellar - gap boiaklari modeli ham, bevosita

45 Фосслер К. Позитивизм и идеализм в языкознании И Звегинцев В.А. История языкознания XIX - XX веков в очерках и извлечениях. Часть 1. - М.: Просвещение, 1964. - С.328 - 329.

Kompyuter lingvistikasi asoslari

47

ishtirokchilar modeli ham, tobelik daraxti modeli ham, transformatsion va generativ grammatika modeli ham mutlaq va benuqson emas.-Ularning hammasi til va nutq strukturasining qaysidir aspektlarini aks ettiradi, xolos. Murakkab sistema hisoblangan tilning barcha qirralarini formallashtirish- ning imkoni yo‘q, bunda muayyan cheklovlar mavjud (ayniqsa, tilning se- mantik aspektida). Bu xususda nemis olimi 1931-yilda K.Gedel shunday yozadi: «Yetarli darajada mazmungci boy nazariya formallashtirilsa, bu nazariyani formal tizimda to ‘la aks ettirish mumkin emas. Bunda aniqlanma- gan, formallashtirilmagan qoldiq qoladi».\*' Demak, har qanday formallash- gan model til grammatikasining butun qirralarini qamrab ololmaydi.

Topshiriqlar:

1. Grammatika deganda nimani tushunasiz?

2. Tilshunoslik tarixida grammatikaga boigan yondashuvlar haqida yozing.

3. Mumtoz sintaktik qarashlar haqida ma’lumotlar to'plang.

4. Transformatsion grammatikaning formal tavsif uchun ahamaiyatli jihatlari va cheklangan tomonlari haqida mulohaza yuriting.

5. Formal grammatika iyerarxiyasining mohiyatini tushuntiring.

Adabiyotlar:

1. Chomsky N. Aspects of the Theory of Syntax. - Cambridge, Mass.: MIT Press, 1965.

2. Хомский H. Синтаксические структуры // Новое в лингвистике. Выпуск II. - С.412 - 527.

3. Nurmonov A. Lingvistik tadqiqot metodologiyasi va metodlari. - Toshkent: Akademnashr, 2010.

4. Nurmonov A. Struktur tilshunoslik: ildizlari va yo‘nalishlari. - Tosh­kent, 2008.

5. ТеньерЛ. Основы структурного синтаксиса.-М.: Прогресс, 1988.

6. Есперсен О. Философия грамматики. - М.: Иностранная литература, 1958.

7. Фосслер К. Позитивизм и идеализм в языкознании // Звегинцев

В.А. История языкознания XIX - XX веков в очерках и извлечениях. Часть 1. - М.: Просвещение, 1964. - С.324 - 335.

8. Шаумян С.К. Теоретические основы трансформационной грамматики // Новое в лингвистике. Выпуск II. - С.391 - 411.

\* Анисимов А.В. Компьютерная лингвистика для всех: Мифы. Алгоритмы. Язык. - Киев: Наук, думка, 1991. - С.77.

48

A. Rahimov

KOMPYUTER LINGVISTIKASIDA MATEMATIK STATISTIKA VA KVANTITATIV METOD

Reja:

1. Matnning statistik tahlili.

2. Kvantitativ metod va uning tilga tatbiqi.

3. Chastotali lug'atlar.

4. Flektiv tillarda o'zakni aniqlash, polisemantik so‘zlarni va sinonimik qatorda dominanta so‘zni aniqlash.

Tayanch so‘z va iboralar: kvantitativ metod, matematik parametr, statis- tika, indeks, tanlama, tanlamaning hajmi, absolut chastota, nisbiy chastota, lisoniy tip, flektiv tillar, agglutinativ tillar, chastotali lug'atlar, matnning statistik tahlili, Zif qonuni, matn atributsiyasi, anonim va psevdonim mat fi­lar, idiostil, polisemantik so 'zlar, sinonimik qator, dominanta, to ‘plam, ling- vostatistika.

Statistik tahlil metodi boshqa fanlar qatori tilshunoslik fani uchun ham muhim ahamiyatga ega. Bunda tilda mavjud hodisalar aniq matematik parametrlarda baholanadi, tekshiriladi. Statistika lotincha «status» so'zidan olingan bo'lib, hodisalarning holati, ahvolini bildiradi. «Status» so‘zidan «stato» - davlat, «statusta» - davlatni biluvchi, «statustica» - davlat to'g'risida muayyan bilim, ma’lumotlar yig‘indisi ma’nosini bildiruvchi so'zlar kelib chiqqan.47 Bundan 3500 yil oldin Qadimgi Misrda o‘tkazilgan aholi hisobi, Rossiyada Pyotr I davridagi «aholi taftishi», Angliyadagi «Dahshatli sud kitobi» kabilar dastlabki statistik ma’lumotlardir.

XVII asrda Angliyada yuzaga kelgan «siyosiy arifmetika» (U.Petti va Jon Grauntlar nomi bilan bog‘liq) fani statistikaning shakllanishiga asos bo'lgan. Statistika ijtimoiy hodisalarning miqdoriy tomonlarini ularning sifat tomonlari bilan uzviy aloqadorlikda o'rganuvchi fandir. Hodisalar­ning sifat tomonlarini maxsus fanlar o'rganadi. Xususan, tovushlar, so'zlar, gaplarni lingvistika, qofiya, ritm, vazn, she’r tuzilishini adabiyotshunoslik, organik hayotni biologiya, yer qatlamlari va boyliklarini geologiya kabi fanlar o'rganadi. Statistika esa ushbu hodisalarni miqdoriy parametrlar aso- sida o'rganadi. Bunda statistika hodisalarni turli omillar ta’siridagi

47 Abdullayev Y. Statistika nazariyasi. - Toshkent, 2002. - B.4 - 5.

Kompyuter lingvistikasi asoslari

49

o'zgaruvchanlik xususiyatini - variatsiyasini e’tiborga oladi. Shu jihatdan statistik ma’lumotlar hodisalar rivojining dinamikasini aks ettira oladi. Lekin sliu o'rinda aytib o‘tish lozimki, statistikaga yagona mutlaq haqiqat deb yondashish, statistik ma’lumotlarga haddan tashqari ishonish ham to‘g‘ri emas. Ayrim yolg‘on faktik materiallarga asoslanib tayyorlangan statistik ma’lumotlar hodisalar haqida xato tushunchalar berishi mumkin. Shu ma’noda statistika yuzasidan shunday hazilomuz gap aytiladi: « Yolg ‘onning uclt turi bor: a) yolg‘on; b) g'irt yo'lg'on; c) statistika».

Statistikaning til hodisalariga tatbiq etilishi natijasida lingvostatistika sohasi shakllandi, unda til birliklari, lisoniy hodisalar, matnlar statistik tah- lil etiladi. Matnning statistik tahlili - tilni ehtimollik belgilariga ko‘ra tahlil qilish, til hodisalarining statistik xarakteristikalari asosida baholash de- makdir. Til birligining gapda yoki matnda qo‘llanish chastotasi (sur’ati) ehtimollik deyiladi. Til hodisalari haqidagi haqiqatlar lingvostatistik tad- qiqotlar yordamida aniqlanadi. Masalan, o‘zbek tili uchun qaysi bo‘g‘in strukturasi tipik ekanligini, gap modcllaridan qaysi biri eng asosiy konst- ruksiya ekanligini, bosh kelishikdagi otlarning asosan qaysi gap bo‘lagi vazifasida kelishini, tanlangan matnda so‘z turkumlarining qay darajada ishlatilishini aniqlash uchun lingvostatistik tahlillarga ehtiyoj seziladi.

Til hodisalarini statistik metod asosida tahlil qilish uchun doim ma’lum bir matn yoki matnlar to'plami olinadi. Ular badiiy adabiyotdan, gazeta va jurnallardan, publitsistik maqolalardan, ilmiy asarlardan, dialektologlar- ningjonli nutqidan yozib olgan materiallaridan tanlanadi. Statistik xusu- siyati o‘rganilayotgan muayyan matnlar «tanlama» deyiladi. Tanlamaning hajmi tekshirilayotgan hodisalarning umumiy sonidir, uni «Н» bilan belgi­lash mumkin. Masalan, 100 ta so‘z; 37 ta bosh kelishik shaklidagi ot kabi. Aniqlanayotgan hodisani ko‘rsatadigan miqdor (masalan, ot oldida aniqlov- chi vazifasida keladigan takrorlangan otlar miqdori) absolut chastota hisob­lanadi, uni «М» harfi bilan belgilash mumkin. 37 ta otdan 30 tasining aniqlovchi vazifasida kelishi absolut chastotadir. Absolut chastota (M)ning tanlama hajmi (H)ga nisbati (M/H)dan nisbiy chastota kelib chiqadi.48 Nis- biy chastota «Р» harfi bilan belgilanadi. Nisbiy chastota uch xil usul bilan aniqlanadi:

a) oddiy kasr hisobida: P = M/H

b) protsent asosida: P = МЧ100/Н

c) promil hisobida: Р = МЧ 1000/H

Matnning statistik tahlilida Zif qonunidan o‘rinli foydalaniladi. Katta

48 Усмонов С. Умумий тилшунослик. - Тошкент: У^итувчи, 1972. - Б.199.

50

A. Rahimov

hajmdagi matnlarda har bir so‘zning qanchalik tez takrorlanishini (chasto- tasini) va ularning takrorlanish tezligini hisoblashda Zif qonuni amal qila­di.49 Bunda so‘zning chastotasi (f), uning ro'yxatdagi holati qatori (r) deb belgilanadi. Zif qonuniga ko‘ra, matnda eng ko‘p ishlatilgan so‘z г = 1, undan pastroq chastotada kuzatilgan so‘z r = 2, undan keyingisi r = 3 tarzida belgilanadi. Bundan kelib chiqadiki, matndagi ixtiyoriy so‘z uchun (f) chas- totaga (r)ning ko'paytmasi natijasi o‘zgarmas miqdor- konstanta (C)ga teng bo‘ladi (r4f = C), bu yerda С matn hajmiga bog'liq. Masalan, amerikalik yozuvchi Mark Tvenning «Тош Soyerning sarguzashtlari» asari Zif qonuni asosida tahlil etilgan. Asarda jami 71 370 ta belgi mavjud bo‘lib, takror- lanmagan so‘z tiplari 8018 ta. Matnda ishlatilgan so‘zlarning o‘rtacha tak­rorlanish chastotasi 8,9 dir, ya’ni matndagi so‘zlar taqriban 9 marta takror- lanadi. Lekin bunda eng katta muammo shuki, so‘z tiplarining barchasi matn ichida qo'llanish darajasi bir xil emas. Ayrim so‘zlar 700 marta tak- rorlanadi, bunday so'zlar romanning 1% qismini tashkil etadi. Yana shun­day so'zlar ham borki, ular bir marta takrorlanadi. Bunday so'zlar hapas legomena (yunoncha so‘z «faqat bir marta o‘qi», degan ma’noni anglatadi) deb ataladi, ular asarning qariyb yarmini tashkil etadi. Qariyb 90% so‘zlar 10 marta yoki undan kam, 10% ga yaqin so‘z tiplari 3 marta yoki undan kam miqdorda takrorlanadi.

6-jadval

|  |  |  |  |
| --- | --- | --- | --- |
| Word(So\*\*) | Frequency(Chastota) | Rank(Qator) | f4r |
| the | 3332 | 1 | 3332 |
| а | 1775 | 3 | 5235 |
| and | 2972 | 2 | 5944 |
| he | 877 | 10 | 8770 |
| but | 410 | 20 | 8400 |
| be | 294 | 30 | 8820 |
| there | 222 | 40 | 8880 |
| one | 172 | 50 | 8600 |
| about | 158 | 60 | 9480 |
| more | 138 | 70 | 9660 |
| never | 124 | 80 | 9920 |
| oh | 116 | 90 | 10 440 |
| two | 104 | 100 | 10 400 |

49 Плат У. Математическая лингвистика // Новое в лингвистике. - М.: Прогресс, 1965. Вып.IV. - С.204.

Kompyuter lingvistikasi asoslari

51

|  |  |  |  |
| --- | --- | --- | --- |
| turned | 51 | 200 | 10 200 |
| You'll | 30 | 300 | 9000 |
| name | 21 | 400 | 8400 |
| comes | 16 | 500 | 8000 |
| group | 13 | 600 | 7800 |
| lead | 11 | 700 | 7700 |
| friends | 10 | 800 | 8000 |
| begin | 9 | 900 | 8100 |
| family | 8 | 1000 | 8000 |
| brushed | 4 | 2000 | 8000 |
| sins | 2 | 3000 | 6000 |
| could | 2 | 4000 | 6000 |
| applausive | 1 | 8000 | 8000 |

Matnning statistik tahlili matn atributsiyasi muammosiga ham oydinlik kiritishda muhim tadqiqot usuli hisoblanadi. Ayniqsa, badiiy matnlarning statistik tahlili natijasida o'sha matn muallifi ham aniqlanmoqda. Bunda matnda tez-tez ishlatiluvchi til birliklari (otlar, sifatlar, kalit so'zlar, fe’llar, grammatik shakllar, jumla qurilishi, bir so‘z bilan aytganda, yozuvchining idiostilini - o‘ziga xos uslubini ko‘rsatuvchi vositalai) lingvostatistik tahlil yordamida aniqlanadi. Turli matnlardan olingan dalillarning qiyosiy tah­lili bizga o'sha matnning mazmun-mundarijasini, matn tuzilgan davrni, da- lillarni argumentlash xarakterini va hatto mualliflikni aniqlashga imkon beradi. Badiiy asarlarning sujet qurilishiga, matn strukturasiga kvantitativ metodlarning tatbiq etilishi natijasida F.Dostoyevskiy, L.Tolstoy, M.Sholoxov kabi yozuvchilarning matnlari atributsiyasi amalga oshirilgan. Keyingi paytlarda anonim (muallifi noma’lum) va psevdonim (soxta muallifli) matnlarning aniqlanishi yuzasidan ham tadqiqotlar amalga oshirilmoqda.50 Bu aspekt kriminalistika sohasida muhim ahamiyat kasb etmoqda.

Amerikalik olim Jozef Grinberg kvantitativ metodni tilshunoslikka - morfologik tiplar nazariyasiga tatbiq etgan. Tadqiqot natijalarini «Quanti­tative approach to the morphological typology of languages» (1960) («Tillar morfologik tipologiyasiga kvantitativ yondashuv») nomli maqolasida umum- lashtirgan. Olim o‘zigacha amalga oshirilgan ishlar haqida shunday yoza- di: «Tillarning tipologik tasnifi mantiqiy-ratsional mezonlar asosida yuza­ga kelgan. Tasnif namunalari nazariy jihatdan asoslangan, ammo lingvisti/c fakt bilan uzviyligi yaqqol ко ‘zga tashlanmaydi. Bu esa tasnifni amaliyotga

50 Баранов A.H. Введение в прикладную лингвистику. - М.: Эдиториал УРСС, 2001.-С.24.

52

A. Rahimov

tatbiq etib tekshirishni taqozo etadi».5' Shu nuqtayi nazardan olim matema­tik aniqlikka erishish maqsadida til hodisalarini, mavjud lisoniy tiplarning uchrash darajasini miqdoriy parametrlar asosida qayta baholadi. J.Grinberg kvantitativ metodni 8 tilga (sanskrit, anglo-saks, yoqut, vyetnam, fors, ing- liz, suaxili, eskimos tillariga) tatbiq qilgan. Bunda olim o‘zi belgilagan for- mulalar asosida lisoniy tiplarning va tillarda kuzatiladigan tipologik belgi­larning miqdoriy tavsifini, statistik ma’lumotlarni ilmiy jamoatchilikka havola etdi.

Masalan, sintetiklikning dunyo tillaridagi darajasini aniqlash uchun М/ W, agglutinatsiyani aniqlash uchun A/J, kompozitsiyani aniqlash uchun RI W, derivatsiyani aniqlash uchun D/W, prefiksatsiyani aniqlash uchun P/W, suffiksatsiyani aniqlash uchun S/W formulalaridan foydalandi. Bu yerda M

- ma’lum tipdagi chekli (100 so‘zdan iborat) matn ichida uchragan morflar soni, A - agglutinatsiya, R (root) - o‘zak, D (derivational) - yasovchi qism, P - prefiks, S - suffiks, W (Word) - matndagi so'zlar sonidir. Hisoblash natijalariga ko‘ra, vyetnam tili uchun - 1,06 (ya’ni 100 so‘zda 106 morf uchraydi), ingliz tili uchun - 1,68, sanskrit tili uchun - 2,59, eskimos tili uchun - 3,72 dir. Natijalardan kelib chiqib olim tillarni baholaydi: «2 dan past ko‘rsatkichga ega boigan tillar (vyetnam, ingliz, fors, xitoy, italyan, nemis tillari) analitik tillar, 2 dan 3 gacha miqdoriy ko'rsatkichga ega bo‘lgan tillar (rus, sanskrit, qadimgi yunon, lotin, eski slavyan, chex, polyak, yoqut, suaxili tillari) sintetik tillar, 3 dan yuqori ko'rsatkichga ega bo'lgan tillar (eskimos, ayrim kavkaz tillari, Amerika hindulari tillari) polisintetik tillar hisoblanadi». Rus olimi V.Z.Demyankovning fikricha, Grinberg tavsiya et- gan metod tilshunoslikdu mavjud ho‘lgan kartotekalash usulidan ishonchli- roq va aniqroqdir. (7-жадвал)

Chastotali lug‘atlar matnda so‘zlarning ishlatilish chastotasi (qay dara- jada ishlatilishi, tez-tez ishlatilishi)dan kelib chiqqan holda tayyorlanadi- gan so'zliklar hisoblanadi. Chastotali lug'atlar statistik leksikografiyaning rivoji natijasida yuzaga kelgan. Bunday lug'atlar, ayniqsa, til o‘qitish jara- yonida leksik minimumlarni aniqlashda muhim ahamiyatga ega. So'zlarning chastotasini aniqlash lingvostatistik tahlillar yordamida amalga oshiriladi. Chastotali lug‘atlar kompyuterning lingvistik informatsion bazasi hisobla­nadi. Masalan, chastotali lug‘atdagi birinchi 10 va 50 eng ko‘p uchraydigan so‘z formalarining miqdori umumiy so‘z birliklariga nisbatan agglutinativ tillarda fiektiv tillarga qaraganda kamroq foizni tashkil etadi. Agglutinativ tillarda dastlabki 1000 so ‘z formasi umumiy so ‘z birliklarining 50 - 60 foizini

51 Qarang: Гринберг Дж. Квантитативный подход к морфологической типологии языков // Новое в лингвистике. - М.. 1963. Вып. III. - С.60 - 94.

Kompyuter lingvistikasi asoslari

53

7-jadval

Tillar morfologik tipologiyasining kvantitativ tadqiqi natijalari (J.Grinberg tadqiqoti)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Indekslar | Sanskrit | Anglo-saks | Fors | Yoqut | Suaxili | Vyetnam | Ingliz | Eskimos |
| Sintez | 2,59 | 2,12 | 1,52 | Г--cnT | 2,55 | COОт— | OOCO | CNJCO |
| Agglutinatsiya | CDCDCD | CD | 0,34 | LOo" | 0,67 | CDo\_o" | 0,30 | 0,03 |
| Qo'shma so'z yasash | CO | ОCD | COО | CNJО | ОCD | o\_ | CDCD | оo\_ |
| Derivatsiya (so'z yasash) | 0,62 | ОCNJcd“ | ОCD | 0,35 | 0,07 | 0,00 | 0,15 | LOCNJ\_ |
| Relatsiya (so'z o'zgartirish) | OOCD | 0,90 | 0,39 | CNJOOo" | CDOOCD | 0,00 | 0,53 | LO |
| Prefiksatsiya | COCD | 0,06 | ОCD | ооCD | CO | CDCDCD | CDCD | 0,00 |
| Suffiksatsiya | oo | COо | 0,49 | LO | CD | ОCDCD | COCD | 2,72 |
|  |  | Ю | CO | СП | О | О | LO | CNJ |
| Izolatsiya |  |  | LO | CNJ |  | О | I4» | CD |
|  | о | CD | CD | О | о | ч— | О | О |
|  |  | Г". |  | CD | О) | О |  | CO |
| Sof fleksiya | CD | CD | CNJo | lOCD | CD | CDCD | CD | CD |
| Moslashuv | OOCO | OOCO | 05 | CNJ |  | CDО | - | OOCO |
|  | CD | О | CD | CD | CD | О | О | CD |

8-jadval

Sintaktik tip (SOV va SVO)ning qit’alarda tarqalish darajasi kvantitativ tadqiqi

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Tip | Afrika | Yevroosiyo | Avstraliya | Shimoliy Amerika | Janubiy Amerika |
| SOV | 22 | 26 | 29 | 26 | 18 |
| svo | 21 | 19 | 6 | 6 | 5 |

54

A. Rahimov

tashkil etsa, bu ko‘rsatkich flektiv tillarda 62 - 87 foiz atrofida boiadi. Flektiv tillarda chastotali lug'atning dastlabki 1000 so‘zi umumiy har xil so'zlar sonining 69 - 89 foizini tashkil etsa, bu ko'rsatkich agglutinativ tillar­da 64 - 89 foiz atrofidadir. Flektiv tillarda dastlabki ko'p uchraydigan 100

- 200 so'z formasi (ko'pchiligi yordamchi so'zlar) jami har xil so'z forma- larining 50 foizini tashkil etsa, bu ko'rsatkich uchun agglutinativ konstruk- siyali turkiy tillarda 600 - 1000 ko'p uchraydigan so'z formalari zarur bo'lar ekan, buning asosiy sababi agglutinativ va flektiv tillar morfologik struktu- rasi va so'z formalarining tuzilishidadir.

Alfavit chastotali lug'at yordamida muayyan old qo'shimchalarning mahsuldorligini so'z turkumlari bo'yicha aniqlash mumkin. Masalan, rus tili chastotali lug'atini olib qaraylik. B3- old qo'shimchasi 74 har xil fei, 27 sifat, 14 ot va faqat 2 ravishda uchragan. БЕЗ- old qo'shimchasi asosan ot, sifat va ravish yasashda ishlatilar ekan, fe’lda esa atigi 4 martagina ishlatil- gan, xolos. Alfavit-chastotali lug'atlar tilda yoki qardosh tillarda va ular- ning sohalarida u yoki bu so'z paradigmasining amalga oshishini kuzatish imkonini beradi.

Chappa (ters) chastotali lug'atlar tilda qo'shimchalarning qanday qo'llanilishini aniqlab beradi. Chastotali va ters chastotali lug'atlarning ma’lumotlariga ko'ra, ot, olmosh, sifat, ravish va son tillardagi jami so'zlarning 70% ni tashkil qilgani holda, fei ulushi ularning 1/4 ga teng. Tillarda so'z turkumlarining ishlatilishi ham tadqiqotchilar e’tiborini o'ziga jalb etadi. Hind-yevropa tillarida yordamchi so'z turkumlarining ko'proq ishlatilishi kuzatiladi (28 - 34%), o'zbek tilida esa kam o'rin egallaydi (ati­gi 8%).

Chastotali Iug‘atlar yaratilishi

1. 1861-yilda bosmaxonada harf terish ishini osonlashtirish maqsadida dastlabki chastotali lug'at tuzildi.

2. 1963-yilda eng faol 2500 leksema kiritilgan chastotali lug'at Shteyn- feld tomonidan yaratildi.

3. 1970-yilda Leypsig universitetida prof. L.Gofman guruhi tomonidan tibbiyot, fizika, kimyo, qurilish ishi, matematika, veterinariya sohalarida eng ko'p qo'llaniluvchi so'zlarni jamlagan 3 tilli chastotali lug'at yaratildi.

4. 1972-yilda Kissen tomonidan «Словарь наиболее употребительных слов современного узбекского литературного языка» («Hozirgi o'zbek adabiy tilining ko'p ishlatiladigan so'zlar lug'ati») nomli chastotali lug'at yaratildi.

5. 1974-yilda harbiy chastotali lug'at-minimum yaratildi.

Kompyuter lingvistikasi asoslari

55

6. 1981-yilda R.Qo‘ng‘urov va S.Karimovlar tomonidan «Zulfiya poe- ziyasining lug'ati. Konkordans» tuzildi.

7. 1982-yilda matematika fani yuzasidan inglizcha-ruscha chastotali lug‘at ishlab chiqildi.

8. 1984-yil Leningrad davlat universitetida tarix, iqtisod fani yuzasidan inglizcha-ruscha chastotali lug'at ishlab chiqildi.

9. 1986-yilda S.Rizayev va N.Bo'ronovlar tomonidan «A.Qahhor «Sin- chalak» povesti tilining chastotali lug‘ati» yaratildi.

10.1987-yilda falsafa fanidan chastotali lug‘at yaratildi.

11,2006-yilda D.O'rinboyeva tomonidan «0‘zbek xalq dostonlari tilin­ing chastotali lug‘ati» ishlab chiqildi. Bu lug'at uchun «Alpomish», «Rav­shan», «Rustamxon» dostonlari matni statistik tadqiqot obyekti qilib olin- gan, unda 28 499 so‘z shakli statistik tahlil qilingan.

Chastotali lug‘atlarning kompyuter lingvistikasi uchun ahamiyati qu- yidagilarda ko‘rinadi:

- birinchidan, chastotali lug'atlar avtomatik qidiruv tizimi asosida ishlay- digan elektron lug‘atlar yaratish uchun, ularning sohaviy turlari uchun tayyor lingvistik ta’minot vazifasini o'taydi;

- ikkinchidan, chastotali lug‘atlar til o'qitishni avtomatlashtirish jara- yonida, kompyuter lingvodidaktikasi sohasida metodologik ahamiyat kasb etadi;

- uchinchidan, chastotali lug‘atlar chet tillarini o‘qitishning samarador- ligini oshirish uchun muhim omil bo‘lib, ular xorijiy tillarning leksik va leksik-morfologik minimumlarini yaratishga imkon beradi.

Agglutinativ tillarda o‘zak va qo'shimcha orasidagi chegara («chok») aniq sezilib turadi. Bu tillarda o'zakni aniqlash oson. Buni matematik asos- da quyidagicha aniqlash mumkin. Masalan: Y+o‘+l+d+a, y+o‘+l+l+a+r, y+o‘+l+n+i+n+g ...

3 (y+o‘+l)+d+2 a+2 n+i+g+r. Demak, bu so‘zlar uchun umumiy o‘zak eng ko‘p qo'llangan tovushlar (y+o‘+l) hisoblanadi.

Flektiv tillarda o‘zak va qo‘shimcha orasidagi chegara sezilmaydi, bu esa ularda o‘zakni aniqlash muammosini keltirib chiqaradi. Ingliz, arab tillari flektiv tillar guruhiga mansubdir. Bu tillarda o'zakni aniqlashda yu- qoridagi usuldan foydalanish mumkin. Bunda eng ko‘p takrorlanuvchi to­vushlar «o‘zak» («root») sifatida olinishi mumkin.

sing - sang - sung - song = s+n+g - o‘zak

ta’lim- olim - fJlc ilm -

Arabcha so'zlar uchun umumiy o‘zak - <J ^ ^ harflaridir.

Polisemantik so'zlar semalari (ma’nolari) orasida integral jihatlar kuza-

56

A. Rahimov

tiladi, mantiqiy aloqadorlik bo'ladi. Ularning polisemantik ekanligini to'plamlik belgilari asosida tahlil qilish yordamida, to'plam elementlari va belgilari orasidagi kesishuvni topish yordamida aniqlanadi.

Birikmalar:

A. Odamning ko'zi.

B. Uzukning ko'zi.

C. Buloqning ko‘zi.

D. Derazaning ko‘zi.

E. Ishning ko‘zi.

F. Yog‘ochning ko‘zi.

G. Qalb ko‘zi.

«Ko‘z»ning semalari:

1. Ko'rish organi.

2. Shaffoflik belgisi.

3. Cheti botganlik.

4. 0‘rtasi bo'rtganlik.

5. Ochilishi va yumilishi.

6. Asosiyligi.

7. Sezish (ko‘rish) organi

9-jcidval

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| № | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| A to‘plain | + | + | + | + | + | + | + |
| В to‘plam | - | + | + | + | - | - | - |
| С to‘plam | - | - | - | - | + | - | - |
| D to‘plam | - | + | - | - | + | - | - |
| E to‘plam | - | - | - | - | - | + | - |
| F to‘plam | - | - | + | + | - | - | - |
| G to'plam | - | - | - | - | - | - | + |

Sinonimik qatorda dominanta (bosh, asosiy) so‘zni aniqlash ham yuqo- rida berilgan to'plam metodi asosida amalga oshirilishi mumkin. Bunda turli nutq uslublaridagi matnlar tahlilga tortiladi va barcha uslublarda ishla- tilishi kuzatilgan so‘z o‘sha sinonimik qatorda dominanta so‘z hisoblanadi.

Topshiriqlar:

1. Statistik tahlil nima uchun zarur?

2. 0‘zingiz mustaqil statistik tadqiqot amalga oshiring.

3. Matematik statistikaga oid mustaqil ish tayyorlang.

4. Ma’lum tanlama matnni so‘z turkumlari bo‘yicha statistik tahlil aso­sida tekshiring.

5. Matn atributsiyasi deganda nimani tushunasiz?

Adabiyotlar:

1. Пиотровский Р.Г. Инженерная лингвистика и теория языка. - Л.: Наука, 1979.

Kompyuter lingvistikasi asoslari

57

2. Пиотровский Р.Г. Текст, машина, человек. - Jl.: Наука, 1975.

3. Гринберг Дж. Квантитативный подход к морфологической типологии языков // Новое в лингвистике. - М., 1963. Вып.Ш. - С.60 - 94.

4. Campbell L., Bubenik V., Saxon L. Word Order Universals: Refine­ments and Clarifications // Canadian Journal of Linguistics. №2. 1988. - P.209 - 230.

5. Abdullayev Y. Statistika nazariyasi. - Toshkent, 2002.

6. Плат У. Математическая лингвистика // Новое в лингвистике. Вып.ГУ. - М.: Прогресс, 1965.

7. Усмонов С. Умумий тилшунослик. - Тошкент: У^итувчи, 1972.

8. Баранов А.Н. Введение в прикладную лингвистику. - М.: Эдиториал УРСС, 2001.

KOMPYUTER LINGVISTIKASI TARIXI VA ZAMONAVIY HOLATI. KOMPYUTER LINGVISTIKASI BO‘YICHA TADQIQOTLAR

Reja:

1. Kompyuter lingvistikasi va uning asosiy yo'nalishlari.

2. Kompyuter lingvistikasiga oid nazariy va amaliy tadqiqotlar.

3. Lingvistik programmalar va ularning yo'nalishlari.

Tayanch so‘z va iboralar; kompyuter lingvistikasi, mashina tarjimasi, avtomatik tahrir, avtomatik о ‘qitish tizimi, matnning statistik tahlili, ling- vostatistika, injener lingvistikasi, chastotali lug‘at, o'qituvchi lingvistik av- tomat, lingvistik ta 'min, lingvistik programmalar.

Kompyuter lingvistikasi fani 1960-yilda AQSHda shakllandi. Mazkur fan XX asrning o'rtalarida paydo boigan matematik lingvistika fani asos- laridan kelib chiqqan. Tabiiy tillar grammatikasi formal modeli komp- yuterda tarjima qilish, tillarga o'rgatish, bilimlarni diagnostika qilish, matn­larni tahrir etish kabi amaliy dasturlarni ishlab chiqishga asos boigani hol- da kompyuter lingvistikasi yo‘nalishiga zamin yaratdi.52 Kompyuter lingvis-

52 Po‘latov A., Muhamedova S. Kompyuter lingvistikasi. - Toshkent, 2007. -B.37-38.

58

A. Rahimov

tikasi fanining asosiy maqsadi lingvistik masalalarni hal qilishning komp­yuter dasturlarini yaratishdir.53 Uning asosiy vazifalariga tillarga o'qitish, bilimlarni tekshirish, matnlarni turli jihatdan tahrirlash va mashina tarji- masi uchun mo'ljallangan dasturlarni ishlab chiqish kabilar kiradi. Komp­yuter lingvistikasi bir-biri bilan chambarchas bog'liq boigan uch yo‘nalishni o‘z ichiga oladi: 1) lingvostatistika; 2) injenerlik lingvistikasi; 3) algoritmik (dasturlash) lingvistika.

Kompyuter lingvistikasi amaliy tilshunoslikning tarkibiy qismi, mate­matik lingvistikaning mantiqiy davomi bo‘lib, u hozirda turli aspektlarda jadal rivojlanmoqda. Kompyuter lingvistikasi fanining asosiy yo'nalishlari sifatida quyidagilarni ajratish mumkin:

a) matn tahriri - kompyuter yordamida muayyan matnning xatolarini tog‘rilash;

b) avtomatik tarjima - berilgan matnni bir tildan boshqa tilga qisqa vaqt ichida tezkor tarjima qilish tizimi;

c) CALL (Computer Assisted Language Learning) - til o‘qitishni av- tomatlashtirish - xorijiy tillarni kompyuter yordamida о‘qitish tizimini ta’minlovchi amaliy dasturlar hamda bilimlarni avtomatik baholash me- xanizmlari;

d) kompyuter leksikografiyasi (elektron lug‘atlar tuzish) - muayyan til leksikonidagi so‘zlarning ma’lumotlar bazasi va maxsus programmalashti- rish tillari yordamida ishlash tizimiga ega bo‘lgan dasturlar (kompyuter lug‘atlari) yaratish bilan shug‘ullanuvchi soha;

e) sun’iy intellekt tizimini yaratish;

f) NLP - tabiiy tilning qayta ishlanishi;

g) gipertekst texnologiyasi;

h) on-line o'qitish tizimi (virtual sinf va laboratoriyalarda) - kompyuter lingvodidaktikasi;

i) informatsion qidiruv tizimi.

Xorijda kompyuter lingvistikasi bo'yicha olib borilgan tadqiqotlarning salmog'i va qamrovi juda keng. Ayniqsa, AQShda kompyuter lingvistikasi bo‘yicha tematik to'plamlar, «Computational Linguistics» («Kompyuter tilshunosligi») nomli jurnal nashr etiladi. AQShda kompyuter tilshunosligi assotsiatsiyasi mavjud bo‘lib (Association of Computational Lingistics), bu uyushma kompyuter tilshunosligi bilan bog‘liq ko‘plab tashkiliy va ilmiy ishlarni amalga oshiradi. AQShda ikki yilda bir marta COLING konferen- siyasi o‘tkaziladi. AQShning 10 ta universitetida kompyuter lingvistikasi

53 Новое в зарубежной лингвистике. Вып.24. Компьютерная лингвистика.

- М.: Прогресс, 1989. - С.10.

Kompyuter lingvistikasi asoslari

59

fakultetlari mavjud. Kompyuter tilshunosligining muammoli masalalari sun’iy intellekt bo'yicha o'tkaziladigan turli xalqaro konferensiyalarda ham ko‘rib chiqiladi. Rossiyada har yili kompyuter lingvistikasiga oid «ДИАЛОГ» xalqaro konferensiyasi 0‘tkaziladi.54 Shuningdek, Niderlan- diyada «Artificial Intelligence» («Sun’iy intellekt») jurnali ham chop etila- di. Mexiko shahrida har yili 10 yildan beri «Computational Linguistics and Intelligent Text Processing» nomli xalqaro konferensiya 0‘tkazib kelinadi.55

Rus kompyuter tilshunosligidagi tadqiqotlarning quyidagi yo‘nalishlarini alohida ajratib ko'rsatish mumkin:

- mashina tarjimasi (MT);

- avtomatik tahrirlash (AT);

- til o‘rgatish jarayonini kompyuterlashtirish (avtomatik til o‘qitish -

ATO‘);

- statistik tadqiqotlar (ST).

Kompyuter lingvistikasining mashina tarjimasi yo'nalishida amalga oshirilgan tadqiqotlar bugungi kunda barcha kompyuterlarda mavjud bolgan ruschadan boshqa tillarga tarjima qiluvchi dasturlarning yarati- lishiga asos boidi. Xususan, L.L.Nelyubin va uning shogirdlari tomonidan ishlab chiqilgan so‘zma-so‘z tarjima qilishning ilmiy strategiyasi (rasmiy- idoraviy uslub asosida) o‘sha davr jahon injener lingvistikasi talablari va ilg‘or tajribasiga monand bo‘lgan edi.56

L.L.Nelyubin tadqiqotlarining quyidagi o'ziga xos jihatlarini ta’kidlash lozim: birinchidan, ushbu ishlarda so‘z va iboralarning avtomatik lug'ati bilan birgalikda to‘liq tarjima qilish imkoniyatini beruvchi morfologik-sin- taktik algoritm-freym yaratishga harakat qilingan; ikkinchidan, vujudga keltirilgan tarjima dasturlari asosida o‘qituvchi lingvistik avtomat (ОЛА - Обучаюший лингвистический автомат) yaratish me’yorlari ishlab chiqil­gan. Ya’ni u o'quvchilarga ingliz tilidagi rasmiy hujjatlarni rus tiliga tarji­ma qilishni o'rgatish imkonini beradi. Professor L.L.Nelyubinning ilmiy faoliyatida mashina tarjimasi uchun (ingliz tili bo‘yicha) algoritmlar va dasturlar ishlab chiqish asosiy o'rinni egalladi. L.L.Nelyubin mashina tar- jimasida ingliz tili leksikasini butunligicha qamrab olmagan, u o‘z tadqiqot­larining obyekti sifatida AQSh harbiy hujjatlarini belgiladi va bu sohada ko‘plab ilmiy asarlar yaratdi. Jumladan, olimning «Перевод боевых документов армии США» (1971), «Частотный англо-русский военный

54 <http://www.dialog-21.ru/>

55 http: [www.CICLing.org](http://www.CICLing.org)

56 Нелюбин Л. Л. Перевод и прикладная лингвистика. - М.: Высшая школа, 1983.-С.6.

60

A. Rahimov

словарь-минимум» (1974), «Перевод и прикладная лингвистика» (1983), «Компьютерная лингвистика и машинный перевод» (1991) kabi monografik tadqiqotlari kompyuter lingvistikasi fani bo'yicha fundamental ishlar hisoblanadi.

Shuningdek, Yu.N.Marchuk tadqiqotlarida ham bir tabiiy tildan boshqa tilga bo'lgan tarjimani modellashtirish tamoyillari, tarjima birligi, tarjima jarayonining statikasi va dinamikasi hamda EHM ishtirokidagi tarjima- ning texnologik liniyalari tavsiflangan. Olimning nazariy qarashlari va g'oyalari quyidagi asarlarida o‘z ifodasini topgan: «Некоторые принципы автоматизации перевода с немецкого языка на русский» (1980), «Проблемы машинного перевода» (1983), Yu.V.Rojdestvenskiy bilan ham- mualiflikda yaratgan «Введение в прикладную филологию» (1987) monografiyalari, shuningdek, «Теория и практика машинного перевода», «Модель «тескт-тескт» и переводные соответствия в теории машинного перевода» nomli maqolalari. Yu.N.Marchuk ham tarjimani modellashtirish metodlari va uni avtomatlashtirish yo'llarini ishlab chiqqani holda bugungi rus kompyuter lingvistikasiga asos solgan olimlardan hisob­lanadi.

L.L.Nelyubin va Yu.N.Marchuklarning mashina tarjimasi bo'yicha olib borgan tadqiqotlaridan ta’sirlangan va ilhomlangan holda ko'plab rus va ukrain olimlari ham kompyuter lingvistikasiga oid qator izlanishlar olib bordilar. Natijada bir nechta monografiyalar yaratildi, o'nlab dissertatsiyalar himoya qilindi, ilmiy maqola va tezislar e’lon qilindi. Xususan, A.M.Kondratovning «Машинный перевод» (1963), I.I.Revzin, V.Yu.Rozensveyglarning «Основы общего и машинного перевода» (1964), V.Ingvening «Язык для программирования задач машинного перевода» (1967), E.I.Korolyovning «Промышленные системы машинного перево­да» (1991) nomli monografik tadqiqotlari; V.M.Kalininning «Развитие схемы Пуассона и ее применение для описания статистических свойств речи», I.L.Yeshanning «Опыт статистического описания научно-тех­нического стиля (на материале переводов с румынского языка)», A.V.Zubovning «Переработка текста естественного языка в системе «человек-машина»«, V.A.Chijakovskiyning «Фразеология и машинный перевод (опыт составления работы немецко-русского автоматическо­го словаря для публицистических и научных текстов)», A.D.Borisevichning «Англо-русский автоматический словарь оборотов (к проблеме идиоматичности при обращении текста в системе «чело- век-машина- человек»)», V.V.Goncharenkoning «Лексикографические, лингвостатистические и инженерно-лингвистические вопросы постро­

Kompyuter lingvistikasi asoslari

61

ения автоматического словаря (англо-русский автословарь по полу­проводникам)» mavzularidagi nomzodlik dissertatsiyalari; V.I.Gerasimovning «Современное состояние машинного перевода», G.G.Kotovning «Лингвистика и современное состояние машинного пере­вода», T.N.Moloshnayaning «Алгоритм перевода с английского языка на русский», Yu.V.Vannikovning «Виды адекватности и типологии пере­водов» sarlavhali maqolalarini alohida ko‘rsatish mumkin.

Mazkur tadqiqotlar rus tili bo'yicha jahon kompyuter tarmoqlarida ma’lumotlar bazasi yaratilishi va buning natijasi sifatida rus tilidagi matn- larni avtomatik tarzda boshqa tillarga tarjima qiluvchi dasturlarning vu- judga kelishi uchun asos bo'ldi. 0‘zbek tilida ham shu kabi ma’lumotlar bazasini yaratish bugungi kundagi eng dolzarb vazifadir.

Rus kompyuter tilshunosligida taraqqiy qilgan yana bir yo‘nalish matn- larning avtomatik tahriri yo‘nalishidir. Mazkur yo'nalishda ilmiy tadqiqot­lar olib borgan olimlar sirasiga R.R.Kotov, V.E.Berzon, V.G.Britvin,

I.A.Melchuk, L.I.Belyayeva, V.A.Chijakovskiy, G.G.Belonogov, I.S.Duganova, A.B.Kuznetsov kabilarni kiritish mumkin. Nomi zikr etil- gan olimlarning quyidagi tadqiqotlarini alohida ajratib ko'rsatish mumkin: G.G.Belonogov va G.G.Kotovlarning «Автоматизированные информационно-поисковые системы» (1968), G.G.Kotovning «Лингвистические аспекты автоматизированных систем управления» (1977), «Лингвистические вопросы алгоритмической обработки сообщений» (1983), «Прикладная лингвистика и информационная технология» (1987), «Оптимизация речевого воздействия» (1990),

G.G.Kotov va B.V.Yakuninning «Язык информационных систем» (1989),

I.L.Belyayeva hamda V.A.Chijakovskiyning «Тезарус в системах автоматической переработки текста» (1983) nomli monografiyalari; V.G.Britvinning «Прикладное моделирование синтагматической семантики научно-технического текста (на примере автоматического индексирования)» mavzusidagi nomzodlik dissertatsiyasi; I.A.Melchukning «Порядок слов при автоматическом синтезе русского слова (предварительные сообщения)», G.G.Belonogov, I.S.Duganova,

A.B.Kuznetsovlarning «.Экспериментальная система автоматизиро­ванного обнаружения и исправления ошибок в тексте» sarlavhali maqola- lari avtomatik tahrir bo‘yicha muhim manbalar sanaladi.

Ushbu olimlarning asosiy xizmati shundaki, ular avtomatik tahrir qiluv­chi dasturlar uchun lingvistik ta’min yaratib berishgan. Buning natijasida kompyuterlardagi ruscha matnlarni tahrirlovchi dasturlar ishlab chiqilgan. Bu dasturlar bugungi kunda barcha kompyuterlarda mavjud. Ya’ni bunda

62

A. Rahimov

noto‘g‘ri yozilgan so'zning tagiga qizil chiziq chiziladi, so‘z to‘g‘ri yozil- ganidan so‘ng chiziq yo'qoladi. Ko'rinadiki, bu dastur asosan imloni tek- shirishga mo'Ijallangan, unda uslubiy xatolarni aniqlash ko'zda tutilma- gan. Kelajakda yaratiladigan dasturlar, umid qilamizki, avtomatik tah- rirning bu jihatlarini ham e’tibordan chetda qoldirmaydi. Ta’kidlash lozim- ki, bunday dasturlarni ishlab chiqish uchun til birliklarining nafaqat fone- tik, semantik, morfologik xususiyatlari, balki aynan sintaktik-valentlik xusu- siyatlari hisobga olingan lingvistik ta’min yaratish lozim bo'ladi.

Rus tilshunosligida avtomatik tarzda tilga o‘qitish yo‘nalishida ham izchil tadqiqotlar olib borilgan. Ular oliy va o‘rta maktabda til o'qitishni optimallashtirishning psixologik-kibernetik, semiotik, lingvo-statistik, in- jener-lingvistik va lingvodidaktik asoslarini ishlab chiqishga yo'naltirilgan. Xususan, G.G.Piotrovskiyning «Компьютеризация преподавания языков» o'quv qo‘llanmasida EHM vositasida o'qitish, ya’ni lingvodidaktik vazi- falarni kompyuter yordamida yechish, til o‘rgatish jarayonini optimallash- tirish, o'qituvchi lingvistik avtomat (OlLA) uchun nutqiy materiallarni tay- yorlash kabi masalalar tadqiq qilingan.57

Tadqiqotlar natijasida ko‘plab 0‘LAlar uchun lingvistik ta’min yaratil- gan bo'lib, bu dasturlar rus tilini o'qitish jarayonini optimallashtirishga xiz- mat qiladi. Eng muhimi, rus kompyuter tilshunosligida «tillarni o'qitishni kompyuterlashtirish» deyilganda o'qituvchi yoki o'quvchi harakatlariga taqlid qiluvchi ko'ngilochar dasturlarni yaratish tushunilmaydi. Bunda asosiy e’tibor AO'S (avtomatik o'qituvchi sistema)larga tayanuvchi avtomatlash- tirilgan o'quv kurslarini vujudga keltirishga qaratiladi. AO'S esa o'quv jara- yoni ishtirokchilariga nafaqat taqlid qilishi, balki ularning intellektual fa- oliyatini modellashtirishi ham zarur bo'ladi.

Rus tilshunosligidagi tillarga o'qitish yo'nalishida lingvistik statistikadan ham keng foydalanilgan. Ma’lumki, ona tilidan tashqari ikkinchi bir tilni o'rganayotganda, avvalo, ushbu tilning lug'at boyligiga murojaat qilinadi. Ammo har bir tilning lug'at boyligida ming-minglab so'zlar mavjud bo'lib, ularning hammasini eslab qolish mumkin emasligi tabiiy. Shu sababli o'rganilayotgan tilning dastlab eng asosiy hamda tez-tez qo'llanib turadi- gan so'zlarinigina o'zlashtirishga kirishiladi va muntazam ravishda bos- qichma-bosqich so'z boyligi oshirib boriladi. Buning uchun esa leksikostatis- tik ma’lumotlar asosiy poydevor vazifasini o'taydi. L.N.Zasorinaning ta’kidlashicha, leksikostatistikaning markaziy muammosi jonli (funksional) tilning statistik qonuniyatlarini va matnning statistik strukturasini aniqlash-

57 Пиотровский Г.Г. Компьютеризация преподавания языков. — JTГГПИ,

1988.-С.74.

Kompyuter lingvistikasi asoslari

63

dir. Matnning statistik strukturasi deyilganda, shartli ravishda, ma’lum matn- dagi turli so'zlar miqdori bilan shu matnda uning qaytarilish-qaytalanish chastotasi orasidagi munosabat tushuniladi. Shunga ko‘ra, statistik ma’lumotlarni to'plash, qayta ishlash kabi murakkab jarayondagi barcha ishlarni EHMga yuklash zaruriyati kelib chiqqan holda kompyuter lingvis- tikasida statistik yo‘nalish yuzaga keldi. U rus tilshunosligida avtomatik tarzda tilga o'qitish yo'nalishi bilan hamohang tarzda rivojlanib borgan. Bu yo'nalishda amalga oshirilgan eng asosiy tadqiqotlar quyidagilardir: P.M.Alekseyev, L.N.German-Prozorova, G.G.Piotrovskiylarning «Основы статистической оптимизации преподавания иностранных языков» (1974), N.D.Andreyevning «Статистике- комбинаторные методы в теоретическом и прикладном языковедении» (1967), P.M.Alekseyevning «Статистическая лексикография» (1975), Ye.A.Kalininaning «Изучение лексико-статистических закономерностей на основе вероятностной модели» (1985).

Matematik va injener lingvistikasi, ya’ni hozirgi kompyuter lingvistikasining yo'nalishlaridagi tadqiqot metodlari turkologiyada ham qo‘llanilgan. Turkologiyada bu yo'nalishda amalga oshirilgan tadqiqotlar sirasiga quyidagilarni kiritish mumkin: K.A.Veliyevaning «Формальное описание синтеза азербайджанского слова» (1970), V.Ya.Pinesning «Моделирование структуры азербайджанских глагольных форм в свя­зи с проблемой автоматического словаря» (1970), X.F.Isxakovaning «Исследования в области формальной морфологии тюркских языков (на материале татарского литературного языка в сопоставлении с турецким и узбекским)» (1972), A.Babanarovning «Разработка прин­ципов построения словарного обеспечения турецко-русского машин­ного перевода» (1981), M.A.Mahmudovning «Разработка системы формального морфологического анализа тюркской словоформы (на материале азербайджанского языка)» (1982), T.Sadikovning «Моде­лирование киргизской именной морфологии» (1982) mavzusidagi nomzodlik dissertatsiyalari; M.Ayimbetovning «Проблемы и методы квантитативно-типологического измерения близости тюркских язы­ков (на материалах каракалпакского, казахского и узбекского языков)» (1997) mavzusidagi doktorlik dissertatsiyasi; G.P.Melnikovning «Прин­ципы системной лингвистики в применении к проблемам тюркологии»,

A.Babanarovning «Автоматический анализ турецкой словоформы и турецко-русский машинный перевод» nomli maqolalari; M.S.Jikiyaning «Морфологическая структура слов в турецком языке (ранговая струк­тура аффиксальных морфем)» (1975), K.B.Bektayevning «Статисти-

64

A. Rahimov

ко-информационная типология тюркского текста» (1978) nomli monografiyalari mavjud.

Matematik va kompyuter lingvistikasi Qozog‘istonda izchil rivojlangan. Bu yerda dastlabki amalga oshirilgan tadqiqotlar, asosan, statistik yo‘nalishga oid bo‘lib, keyinchalik qozoq tilshunoslari mashina tarjimasi, avtomatik tahrir, tillarga o‘qitish kabi sohalarda ham keng ko'lamli ishlarni amalga oshirishdi. Bunda A.Axbayevning «Статистический анализ лек­сико-морфологической структуры языка казахской публицистики» (1971), A.X.Jubanovning «Статистическое исследование казахского текста с применением ЭВМ (на материале романа М.Ауэзова «А бай жолы») (1973), K.A.Maldibekovning «Частотный словарь казахской детской литературы» (1980), D.A.Baytanayevaning «Информационные характеристики казахского языка» (1985) kabi dissertatsion ishlarini alohida ta’kidlash joiz.

Qozoq kompyuter lingvistikasi shakllanishida prof. Q.B.Bektayevning ilmiy faoliyati muhim ahamiyatga ega. Olim matematik va injener lingvis- tikasiga oid o'nlab asarlar yaratgan.58 Q.B.Bektayev rahbarligidagi «Lingvostatistika va avtomatlashtirish» guruhi a’zolari M.Avezovning «Abay yo‘li» romani (4 kitob) tilining chastotali lug'ati so'zligini kompyu­ter yordamida 50 soat davomida tuzib chiqqanlar. Mazkur chastotali lug'atda 20 000 dan ziyodroq leksema va 60 000 so‘z shakli («glossema») mavjud boMib, u 466 000 marta qo‘llanilgan. Guruh a’zolari bu lug‘atni tuzish uchun atigi sakkiz oy vaqt sarflaganlar. Vaholanki, shu ish oddiy qo‘l kuchi bilan bajarilganda, mazkur guruh a’zolari tinimsiz 10 yil, bir tilchi esa kamida 100 yil ishlagan bo‘lar edi. Shu tariqa ushbu ulkan ish qozoq tilshunosligida kompyuter lingvistikasining rivojlanishiga asos bo'ldi.

0‘zbek tili materiallari bo‘yicha kompyuter tilshunosligiga oid dastlab­ki tadqiqotlar olib borgan olimlar H.Arziqulov, S.Rizayev va S.Muhamedovlardir. Mazkur olimlar ham asosan kompyuter lingvistika­sining statistik tahlil, avtomatik tahrir yo‘nalishlari bo'yicha ishlarni amal-

58 Бектаев К.Б., Лукьяненко К.Ф. О законах распределения единиц письменной речи. - СтРААТ, 1971; Бектаев К.Б., Джубанов А. Индексация и кодирование текстовой информации для ввода в ЭВМ. - Алма-Ата, 1973; Пиотровсикй Г.Г., Бектаев К.Б., Пиотровская А.А. Математическая лингвистика. - М.: Высш. шк, 1977; Бектаев К.Б., Кенесбаев С.К., Пиотровский Г.Г. О инженерной лингвистике // ВЯ. 1973. №2; Бектаев К.Б. Статистика - информационная типология тюркского текста. - Алма-Ата: Наука, 1978; Бектаев К.Б., Садчикова П.В. Опытно-промышленная система лексического машинного перевода. - МСМП, 1979.

Kompyuter lingvistikasi asoslari

65

ga oshirgan. H.Arziqulov o‘zining «Информатика и переработка текста средствами вычеслительной техники» nomli monografiyasida kompyu­ter yordamida katta hajmdagi matnlarga ishlov berishning avtomatik tizimi asoslarini chuqur tahlil qilgan. Eng muhimi, professor H.Arziqulov rahbar- ligida Samarqand davlat chet tillari institutida injener tilshunosligi markazi tashkil etilgan edi. Bu markaz davlat tilidagi matnlarga kompyuter yordamida qayta ishlov berish masalalari ustida ilmiy kuzatishlar olib borgan. Bunda matnlarga qayta ishlov beradigan kompyuter dasturlarini yaratish, o‘zbek tilining barcha sathlariga oid matnlarni yig‘ish, ularni muayyan bir tizimga keltirish, indeksatsiya qilish, algoritmlarini tuzish, ularga statistik jihatdan ishlov berish, inglizcha-o‘zbekcha mashina tarjimasi dasturini yaratish sohasida izchil ilmiy-tadqiqot ishlarini boshqargan. 0‘zbekcha matnlarga qayta ishlov beruvchi avtomatik tizim «Uzlington», lotin alifbosidagi matnlarni kirill yozuviga va lotin alifbosidan kirill alif- bosiga o'tkazuvchi «Spellchecker», kompyuter ko'magida korrektura qila- digan, imloviy xatolarni tuzatadigan kompyuter dasturlarini yaratishga bi- rinchilardan bo‘lib qo‘l urgan.

Hamdam Orziqulov o‘zbek mashina tilini yaratish maqsadida so‘z tarki- bini nihoyatda chuqur tadqiq qildi: morfema va so‘z yasalish strukturalari haqidagi fikrlari, matn statistikasi va undan chastotali lug‘atlar tuzish prin- siplari, o‘zbek matnlariga kompyuter yordamida qayta ishlov berilishi va ularning dasturlarini yaratish haqidagi jiddiy tadqiqotlari tom ma’noda o‘zbek tilshunosligida injener tilshunoslik sohasi yuzaga kelishiga zamin bo‘ldi, kompyuter tilshunosligi alohida tizim sifatida shakllanishi va o‘rganilishi ta’minlandi. Shuning uchun ham Hamdam Orziqulov o‘zbek tilshunosligining Samarqand maktabida injener tilshunosligi sohasi asoschi- sidir, deyish mumkin.

S.Muhamedov o‘ziningR.R.Piotrovskiy bilan hammualliflikda yozgan «Инженерная лингвистика и опыт системно-статистического ис­следования узбекских текстов» nomli kitobida lingvistik modellar, modellashtirish va uning umumiy tamoyillari haqida fikr yuritib, o‘zbekcha matnlarning kvantitativ modellarini keltiradi.59 Mazkur asarda sun’iy intellekt yaratish va injener lingvistikasi metodlari bilan o‘zbekcha nutqni avtomatik qayta ishlash jarayonida foydalanish uchun zaruriy boigan o‘zbekcha matnlarning leksik - morfologik tuzilishi statistik tahlili natijalari keltirilgan. Ya’ni uning «Статистический анализлексико-мор-

59 Мухамедов С.А., Пиотровский Г.Г. Инженерная лингвистика и опыт системно-статистического исследования узбекских текстов. - Ташкент: Фан, 1986.

66

A. Rahimov

фологической структуры узбекских газетных текстов» (nomzodlik dissertasiyasi) va «O'zbek tilining alfavitli-chastotali lug'ati (gazeta tekstlari asosida)» asarlari yuqoridagi asarning yozilishida manba boiib xizmat qilgan. S.Rizayev «Kibernetika va tilshunoslik» asarida tilshunoslikda aniq metodlarning qo'llanishi va bunda EHMdan foydalanish, til va nutq hodisalariga statistik yondashishning sabablari hamda matnlarni avtomatik qayta ishlash va mashina tarjimasi muammolari, shuningdek, harflar chastotasini aniqlashda EHMdan foydalanishga doir ma’lumotlar bergan.

S.Muhamedov va S.Rizayevlarning kompyuter tilshunosligiga oid asar­lari bugungi kun kompyuter tilshunosligiga asos hisoblanadi. Ammo o'zbek kompyuter tilshunosligini yangicha, zamonaviy texnologiyalar, dasturlar asosida rivojlantirish hamda samarali, ixcham, qulay va arzon kompyuter dasturlari uchun lingvistik ta’minot yaratib berish eng dolzarb vazifa hisob­lanadi.

Shuni ta’kidlash lozimki, mazkur vazifalarni amalga oshirishda komp­yuter lingvistikasi o'zbek tilshunosligida shu paytgacha yaratilgan birorta ham ilmiy asarni rad etmaydi. Aksincha, u A.G'ulomov, M.Asqarova, A.Kononov, F.Kamol, Z.Ma’rufov, N.Mamatov, Sh.Rahmatullayev, M.Mirtojiyev, I.Qo'chqortoyev, A.Nurmonov, N.Mahmudov, H.Ne’matov, R.Sayfullayeva, M.Sodiqova, R.Rasulov, E.Begmatov, I.Mirzayev,

H.Dadaboyev, Yo.Tojiyev kabi atoqli o'zbek olimlarining o'zbek tili lek- sikasi va grammatikasiga oid tadqiqotlariga suyanadi. Ulardagi ilmiy-na- zariy fikrlarga asoslanadi, ulardan oziqlanadi.60

So'nggi yillarda o'zbek kompyuter lingvistikasi asta-sekinlik bilan shakl- lanmoqda. Ayniqsa, bu sohada A.Po'latov, N.Jo'rayevalarning «Разработка формальной модели грамматики узбекского языка»,

A.Po'latov, B.Alixo'jayev, N.Jo'rayevalarning «Разработка программы компьютерного анализа и синтеза глаголов узбекского языка», A.Po'latov,

S.Muhamedovalarning «Kompyuter tilshunosligida avtomatiK tahrir qiluv- chi dasturning lingvistiK ta ’minotini yaratish asoslari» nomli maqolalari, shuningdek, A.Po'latov va S.Muhamedovalarning «Kompyuter lingvistika- si» nomli o'quv qo'llanmalari (2007) muhim ahamiyatga molik.

Kompyuter lingvistikasi sohasida turli xil lingvistik muammolarni hal qilishga yo'naltirilgan dastrurlar yaratilgan va hali-hanuz bunday amaliy tadqiqotlar amalga oshirilmoqda. Quyida ana shunday lingvistik dasturlar va ularning tavsifi haqida to'xtalamiz. Ular asosan matn tahriri, avtomatik

60 Po'latov A., Muhamedova S. Kompyuter lingvistikasi. - Toshkent, 2007. -

B.47-48.

Kompyuter lingvistikasi asoslari

67

tarjima, analiz va sintez, tabiiy tilning qayta ishlanishi, kompyuter lek- sikografiyasiga oid dasturlarni o‘z ichiga oladi.

10-jadval

|  |  |  |
| --- | --- | --- |
| 1 | Link Grammar Parser John Lafferty, Daniel Sleator, Davy TemperleyCamegi Melon University, USA. | Ingliz tilining sintaktik parser dasturi, uning bazasiga 60 000 so'z shakllari kiritilgan. |
| 2 | C&#237:bola Oleada loyihalariComputing Research Laboratory (CLR) New Mexico State University, USA. | Mazkur dastur matnlarning lingvistik analizida ishlatiladi. Bu dastur yordamida 16 ta tildagi matnlarning avtomatik tarjimasi, statistik tahlilini amalga oshirish mumkin. Dastur tarkibida turli xil lug'atlar hamda tezauruslar ham mavjud. |
| 3 | Russian Morpholoqical DictionarySergey Sikorsky | Ruscha morfologik lug'at rus tilidagi matnlarning sintaktik va morfologik analizi uchun ishlatiladi. Unga 120 000 so'zning morfologik lug'ati kiritilgan. |
| 4 | Морфологический парсер flndex | Ushbu morfologik parseming asosi A.A.Zaliznyak tomonidan yaratilgan «Rus tilining grammatik lug'ati» (110 ming so'zni o'z ichiga oladi) hisoblanadi. Tahlil natijalarini ko'rish uchun «разбор запроса» rejimi tanlanadi. |
| 5 | ЛингвоанализаторД.В.Хмелев | Matn strukturasining matematik tahlilini ta’minlovchi dastur bo'lib, u on-line rejimida ishlaydi. Bunda mualliflik etaloni matn strukturasi asosida aniqlanadi. Kiritilgan matn analiz qilinadi (Русская фантастика resursidan olingan matnlar analiziga tayaniladi), o'sha matnning muallifi bo'lishi mumkin bo'lgan 3 yozuvchining ismi hamda ulaming matnga yaqin keladigan asarlari chiqarib beriladi. |
| 6 | Система StarLina | Ushbu tizim katta hajmdagi ko'p tilli matnlar bilan, transkripsion belgilar bilan |

68

A. Rahimov

|  |  |  |
| --- | --- | --- |
| 6 | Система StarLina С.А.Старостин СУБД | Ushbu tizim katta hajmdagi ko‘p tilli matnlar bilan, transkripsion belgilar bilan ishlashga mo'ljallangan bo'lib, u DBF-formatida yuklanadigan Ojegovning izohli lug'ati va Zaliznyakning morfologik lug'ati asosida so‘z shakllarining analizi va sintezini, Myullerning lug'ati asosida avtomatik tarjimani ta'minlaydi. |
| 7 | МорфологическийанализаторС.А.Старостин | Rus, ingliz tillaridagi so'zlarning morfologik analizi dasturi so'zlarning asosiy shakllari hamda morfologik ma’lumotini menyuga chiqaradi. |
| 8 | МопоСопсMichael Barlow Dept of Linguistics, Rice university, Texas, USA.2 версии (MonoConc Pro 2.0 и MonoConc 1.5) | Tanlangan simvolli matnlar korpusi uchun konkordands yaratishda, so'zliklar va indekslar generatsiyasida foydalaniladi. |
| 9 | WordSmith Tools Mike Scott | WordSmith 3.0 turli alohida fayllar majmuidan tashkil topgan matnlar korpusi uchun konkordans yaratishda foydalaniladi. Bunda berilgan matnli fayllar majmui uchun so'zlar ro'yxatini generatsiya qiladi, katta hajmli matnlarni fragmentlar majmuiga ajratadi, matnlarning paketli tahririni amalga oshiradi. |
| 10 | Concordance 2.0.0 R.J.C. Watt, 2000 | Win9x/NT/2000 uchun chastotali ro'yxat va konkordanslar yaratish uchun ishlatiladi. Windows tizimida mavjud bo'lgan kodirovkalarda matnlarga ishlov beradi, natijalar HTML-fayllar ko'rinishida saqlanadi. |
| 11 | TextAnalvst 2.0 Научно-производственный инновационный центр «Микросистемы» | Matn analizi uchun ishlatiladi, bunda belgilangan matnda kontekstga mos keluvchi tushunchalarning semantik tarmog'i shakllantiriladi, matn fragmentlari bo'yicha mazmunli qidiruv amalga oshiriladi, iyerarxik daraxt ko'rinishida (tema-podtema) matn analizqilinadi, matn avtomatik referatlanadi. |

Kompyuter lingvistikasi asoslari

|  |  |  |
| --- | --- | --- |
|  |  | TextAnalvst SDK - inqliz va rus tillari uchun lemmatizatsiya (so'zlarni me'yoriy shaklga keltirish), tushunchalarning chastotali ro'yxatini yaratish, kontekstda so'zlarni qidirish kabi ishlarni bajaradi. |
|  |  | TextAnalvst Lib - aioertekst texnologiyasi asosida elektron kitoblar yaratishda foydalaniladi. |
| 12 | Galaktika-ZOOM корпорация Г алактика, Москва | Informatsiyani analitik qayta ishlash va avtomatik qidiruvda ishlatiladi. |
| 13 | Система Пропись 4.0 АО Агама | Rus tilidagi matnlarga ishlov berish uchun ishlatiladigan dasturlash paketi bo'lib, bunda orfografiya tekshiriladi, so'zlarning sinonimlar va antonimlar ro'yxati beriladi, matn grammatik va stilistik jihatdan tekshiriladi, izohli lug'at asosida so'zlarning izohi beriladi, matndagi so'zlarni qidirish va almashtirish hamda matnning statistik tahlilini amalga oshirish imkoniyati ham mavjud. |
| 14 | Textual Analysis ComDutina Tools (TACT)Library Electronic Text Resource Service Indiana University, USA | Matnni avtomatik joylashtirish, so'zlarning tartibli ro'yxatini yaratish, matndagi so'zlarning uzunligi va chastotasiga ko'ra taqsimlanishi tahlili, konkordans yaratish kabilar uchun moijallangan dastur sanaladi. |
| 15 | Худломер Леонид Делицын | Mazkur dastur rus tilidagi matnlarni uslubiga ko'ra avtomatik tasniflashga moijallangan. Bunda 4 xil matnlar korpusi (badiiy asar, publitsistika, ilmiy maqolalar va dialoglar) tahlil etiladi, natijada stilga bogiiq holda matndagi so'zlar uzunligining egri joylashuvi ko'rinadi. Ular tasniflash jarayonida mezon sifatida olinadi. Dastur yordamida kiritilgan matn «РАЗГОВОРНАЯ РЕЧЬ» (OG'ZAKI NUTQ), «ХУДЛО» (BADIIY ADABIYOT), «ГАЗЕТНАЯ СТАТЬЯ» (GAZETA MAQOLASI) yoki «НАУЧНАЯ |

70

A. Rahimov

|  |  |  |
| --- | --- | --- |
|  |  | СТАТЬЯ» (ILMIY MAQOLA) deb tasniflanadi. |
| 16 | «Штампомер» Леонид Делицын | Bu dastur mualliflik matnlarida tez-tez ishlatiladigan iboralar (shtamplar), o'sha muallif tomonidan ko'p ishlatiladigan so‘z va ifodalarni aniqlashga yordam beradi. |
| 17 | Свежий взглядFresh Eve версия 1.21. | Bu dastur rus tilidagi matnlarni stilistik jihatdan tekshiradi. Bunda matnda o'zaro yaqin joylashgan fonetik va morfologik jihatdan o'xshash bo'lgan so'zlarni topadi hamda stilistik g'alizlikni, tavtologiyaning oldini oladi. |
| 1995Дмитрий Кирсанов |
| 18 | RussianWord Constructor | Bu rus tilidagi she'rga o'xshash matnlarning generatsiyasi uchun yaratilgan tajriba dastur bo'lib, leksiko-statistik ma’lumotlarga va lug'atga tayanib ruscha neologizmlarni va okkazionalizmlarni hosil qiladi. |
| (Rwqверсия 1.0, 1992 Дмитрий Кирсанов |
| 19 | Scott Pakins automatic | Bu dastur muayyan shaxs yoki tashkilotga shikoyat matnini generatsiya qilishga yordam beradi. Dasturga 7000 leksik unsurni o'z ichiga olgan lug'at bazasi kiritilgan bo'lib, u ingliz tilida on-line rejimida ishlaydi. |
| complaint-letter qenerator Scott Pakin |
| 20 | Cvrano ServerNando.net | Bu dastur sevgi maktublarining generatsiyada qo'llaniladigan on-line rejimida ishlaydigan dastur bo'lib, bunda foydalanuvchi o'z ismini, yozish stilini, sevgilisiga murojaat shaklini, yana sevgilisining ayrim atributlarini tanlaydi, dastur ingliz tilida sevgi maktubini yaratadi. |
| 21 | Письмовник MediaLinqua Ltd., 1995-2002 | Ish hujjatlari va yozishmalarni avtomatik tarzda shakllantiradi hamda ularni rus tilidan ingliz tiliga, ingliz tilidan rus tiliga tarjima qiladi. |
| 22 | ALICEDr. Richard S. Wallace. ALICE Al Foundation, USA | Bu dastur gapiruvchi programme bo'lib, u 2000 va 2001-yillarda Lobner mukofotiga (Loebner prize) loyiq topilgan. Dastur AIML (sun'iy intellekt uchun maxsus til) texnologiyasiga asoslanadi. |
| 23 | ANANOVA | Buyuk Britaniyadagi virtual diktor asosida ishlaydigan axborot agentligi bo'lib, u |

Kompyuter lingvistikasi asoslari

71

|  |  |  |
| --- | --- | --- |
|  | Ananova Limited United Kingdom | orqali kunning yangiliklarini bilib olish mumkin. |
| 24 | Sakrament Text-to-Speech Enaine v2.0компания «Сакрамент», 2002 | Rus tilidagi nutq sintezi tizimi bo'lib, u Microsoft Speech API (SAPI) 5.1. standartiga toliq mos keladi, erkak va ayol ovozida matnlarni talaffuz qiladi, foydalanuvchining xohishiga ko'ra intonatsion pauzalar qiladi, nutq tembri va tonini o'zgartiradi. |
| 25 | Аки РоссАльберт Товмасян, 2002 | Bu dastur rus tilida muloqotni ta'minlaydi. Bunda maxsus plug-in Aki Ross ismli animatsiya ko'rinishidagi qiz-bola suhbatdosh bilan muloqot qilishga imkon beradi.Dastuming bilimlar bazasi juda katta emas, lekin uni AIML tili yordamida to'ldirib borish mumkin yoki dastur saytidan bilimlar bazasini yangilab turish mumkin. |
| 26 | Proqrammable Artificial Intelliaence (PAIv3.01) или Программируемый Искуственный интеллектClaudio Scordino, 2002 | Ushbu dastur foydalanuvchi bilan ingliz tilida muloqot qilishni ta’minlaydi, u C++ tilida Linux, FreeBSD va Windows tizimlarida ishlaydi. |

Topshiriqlar:

1. Kompyuter lingvistikasi bo'yicha xalqaro konferensiya materiallari- ni Internetdan yuklab oling.

2. Kompyuter lingvistikasi fani bilan shug‘ullangan olimlar va ular- ning asarlarini ko‘rsating.

3. Kompyuter lingvistikasi fanini rivojlantirish uchun nimalar qilish lozim?

4. Kompyuter lingvistikasida avtomatik tarjima muamosi bilan kimlar shug‘ullangan?

5. O'zbek tilshunosligida kompyuter lingvistikasiga oid qanday tadqiqot­lar mavjud?

Adabiyotlar:

1. Grishman R. Computational Linguistics // Cambridge University Press. 1994.

72

A. Rahimov

2. Шемакин Ю.И. Начало компьютерной лингвистики. - М.: МГОУ, 1992.

3. Gelbukh A. Computational Linguistics and Intelligent Text Process­ing. International Conference. - Mexico, 2003.

4. Новое в зарубежной лингвистике. Вып.24. Компьютерная лингвистика. - М.: Прогресс, 1989.

5. Poiatov A., Muhamedova S. Kompyuter lingvistikasi. - Toshkent, 2007.

6. Jurafsky D., Martin J.H. Speech and Language Processing. - New Jersey, 2000.

7. Арзи^улов Узбек лингвистик автомати // Муло^от. 2002. №1. - Б.20 - 21.

8. Сафаров LLL, Иулдошев Б. Компьютер лингвистикасини биласизми? // Мо^ият. 2004 йил 14 август.

KOMPYUTER LINGVISTIKASIDA AVTOMATIK TARJIMA YO‘NALISHI

Reja:

1. Kompyuter lingvistikasida avtomatik tarjima yo‘nalishi.

2. Avtomatik tarjima yo'nalishi tarixi. Tarjimon dasturlar.

3. Avtomatik tarjima yo‘nalishining zamonaviy holati va istiqboli.

Tayanch so‘z va iboralar: matn tahriri, avtomatik tarjima, til o'qitishni avtomatlashtirish, elektron lug'atlar, tezaurus lug'atlar, kompyuter lek- sikografiyasi, kompyuter lingvodidaktikasi, on-line tarjima tizimi, ekviva- lent moslik, transformatsion moslik, tarjima protsedurasi, generatsiya, trans- latsiya, bilingv bazasi, example-based translation (namunalarga asoslana- digan tarjima), rule-based translation (algoritmga asoslangan tarjima), trans­lation memory (tarjimon xotirasi).

Informatsiyaning geometrik progressiya asosida keskin ko‘payishi ular- ning boshqa tillarga tezkor tarjima qilinishini ta’minlovchi tizimlarga bo'lgan ehtiyojni keltirib chiqardi. Bunga javoban mashina tarjimasi yoki avtoma­tik tarjima yo'nalishida amaliy tadqiqotlar yaratila boshlandi. Eng aha-

Kompyuter lingvistikasi asoslari

73

miyatli tomoni shundaki, mashina tarjimasiga oid izlanishlar kompyuter lingvistikasi fanining yuzaga kelishida tayanch nuqta hisoblanadi.

Avtomatik tarjima bo‘yicha dastlabki g'oyalar ingliz olimi Charlz Beb- bidj tomonidan 1836 - 1848-yillarda ilgari surilgan. Uning fikricha, mexa- nik-elektron mashinalar tillar o'rtasidagi struktur o'xshashliklar asosida kodlashtirilgan avtomatik tarjimani amalga oshirishi mumkin. Uning loyi- hasi bo‘yicha 50 mingta so‘z xotira kuchiga ega bo‘lgan elektron mashi­nalar 100 000 ta so‘zni avtomatik ravishda tarjima qilishi zarur boiadi.61

1949-yilda Ch.Bebbidjing g'oyalaridan ta’sirlangan amerikalik olim Uorren Uiver mashina tarjimasi tizimini yaratishning nazariy asoslarini ishlab chiqdi hamda lingvistika, deshifrovka, programmalashtirish nazari- yasi bo'yicha mutaxassislarni hamkorlikka chorladi.62 Shundan so‘ng AQSH- da MT bo'yicha kollaboratsiyalar, hamkorlik loyihalari yaratildi. Mashi­na tarjimasi bo'yicha birinchi xalqaro konferensiya 1952-yilda Massachusets texnologiya institutida o'tkazildi.

1954-yilda AQShdagi Jorjtaun universitetida dunyoda birinchi marta avtomatik tarjima tajribadan o‘tkazildi. U GAT tizimi (inglizcha «George­town Automatic Translation» birikmasining qisqartmasi) deb ataladi. Bu tajribalar IBM-701 kompyuterida o'tkazildi. Bu tajribaning bazasi algoritm- lardan iborat bo'lib, 50 000 ta so‘z (60 ta gap)ni rus tilidan ingliz tiliga tarjima qildi.6?

Rossiyada mashina tarjimasi yo'nalishining pionerlari D.Yu.Panov,

A.A.Lyapunov, I.S.Muxinlar sanaladi.64 Shundan so‘ng O.S.Kulagina,

I.A.Melchuk, Yu.A.Motorin, Yu.N.Marchuk, R.G.Piotrovskiy kabi olim­lar ularning izdoshlari sifatida avtomatik tarjima tizimini rivojlantirdilar. Sobiq lttifoqda mashina tarjimasi bo'yicha birinchi tajriba 1955-yilda o'tkazildi. Shundan so'ng O.S.Kulagina va I.A.Melchuk rahbarligi ostida Amaliy matematika institutida mashina tarjimasi bo‘yicha uchta yirik tajri­ba tizimi yaratildi: fransuzchadan ruschaga (FR-1), inglizchadan ruschaga va fransuzchadan ruschaga (yangi variantda). 1959-yilda Moskva davlat

61 Po‘latov A., Muhamedova S. Kompyuter lingvistikasi. - Toshkent, 2007. -

B.52.

62 Баранов A.H. Введение в прикладную лингвистику. - М.: Эдиториал УРСС, 2001. -С.89.

м Убин И.И. Современные средства автоматизации перевода: надежды, разочарования и реальность // Перевод в современном мире. - М.: ВЦП, 2001.

- С.60 - 69.

м Панов Д.Ю., Ляпунов А.А., Мухнн И.С. Автоматизация перевода с одного языка на другой. Сессия по научным проблемам автоматизации производства. - М.: Изд-во АН СССР, 1956.

74

A. Rahimov

lingvistik universitetida mashina tarjimasi laboratoriyasi ochildi. Moskva- da 1974-yilda INFORM-ELEKTRO institutida fransuzcha-ruscha (ETAP-

1), inglizcha-ruscha (ETAP-2) hamda nemischa-ruscha (NERPA) avtoma­tik tarjima tizimlarini yaratish bo‘yicha amaliy tadqiqotlar jonlandi. 1960- yilda sobiq Ittifoq Fanlar akademiyasining avtomatik tarjimaga bag'ishlangan maxsus yig'ilishi bo'lib o‘tdi. Bu yig‘ilishda avtomatik tarji­ma bo‘yicha professor A.Belonogov rahbarligida maxsus laboratoriya tash- kil qilindi.

1966-yilda AQShning fan masalalari bo‘yicha kongress qo'mitasida av­tomatik tarjimaga oid ma’ruza muhokamasi (Automatic Language Process­ing Advisory Committee-ALPAC REPORT) boMib o'tdi.63 Unda avtomatik tarjima bo'yicha «Language and Mashines: Computers in Translation and Linguistics» («Til va mashinalar: tilshunoslikda va tarjimada kompyuter- lar») deb nomlangan salbiy ma’nodagi ma’ruza qilindi va yuqori darajada sifatli avtomatik tarjimani ta’minlovchi universal tizim yaratish natijasizli- gi va imkonsizligi ta’kidlandi. Shuningdek, ma’ruzada ko‘rsatilishicha, «tabiiy tarjima» (inson tarjimasi) mashina tarjimasidan arzonroq va sifatli- roq. Shundan keyin jahon miqyosida avtomatik tarjimaga oid tadqiqotlarni moliyalashtirish keskin kamayib ketdi. Ular faqatgina Yaponiyada davlat tomonidan moliyalashtirildi. «Generel Motors», «Siemens», «LG», «Phi­lips» kompaniyalari bugungi kunda dunyoda avtomatik tarjimaga oid izla- nishlarni mablag1 bilan ta’minlab turadi.

AQSH harbiy havo kuchlari buyurtmasi bo'yicha D.Tom rahbarligida SYSTRAN avtomatik tarjima sistemasi ishlab chiqildi. Mazkur tizim yor- damida bir yilda 2,5 mln sahifa matn ingliz tilidan nemis, fransuz, ispan, yunon, italyan tillariga hamda fransuz tilidan ingliz, ispan, italyan tillariga tarjima qilinadi. Avtomatik tarjima bo'yicha dunyodagi eng yirik loyiha 1991-yilda AQShda NASA mutaxassislari tomonidan yaratilgan: DIANA. Uning hajmi 10 mln ta so‘zni tashkil etadi. U dunyoning 6 ta yirik tiliga asoslanadi: ingliz, rus, fransuz, nemis, arab, ispan.

Rus olimlari ham professor Lyapunov rahbarligida «Vista Technology» kompaniyasi «Retrans Vista» loyihasini ishlab chiqdi. Uning tarkibiga kiruv- chi 3 mln 400 mingta so‘z xotiraga kiritildi. Shunda 1 mln 800 mingtasi ruscha-inglizcha, 1 mln 600 mingtasi inglizcha-ruscha so'zdan iborat.

Bugungi kunda avtomatik tarjimaning quyidagi sistemalari mavjud. 40

65 Слокум Дж. Обзор разработок по машинному переводу: история вопроса, современное состояние и перспективы развития // Новое в зарубежной лингвистике. Вып.XXIV. Компьютерная лингвистика. - М., 1989.

- С.358.

Kompyuter lingvistikasi asoslari

75

mln ta so‘z xotira kuchiga ega YEUROTRA sitemasi - Yevropa olimlari tomonidan ishlab chiqilgan. U barcha yevropa tillari bo'yicha avtomatik tarjimani amalga oshira oladi. SOKRAT sistemasi 40 mln ta so‘z xotira kuchiga ega. Dunyodagi 40 ta tilda avtomatik tarjimani amalga oshiradi. Ya’ni qaysi tilda Windows operatsion tizimi versiyasi mavjud boisa, shu tilda avtomatik tarjimani amalga oshira oladi. Lingvo sistemasi (lingvo kompyuter lug'ati) - bu lug'atga ham dunyoning 40 tilidagi ma’lumotlar jamlangan. 1990-yillarda Amerikada Global Link, Rossiyada PROMT avtomatik tarjima tizimlari yaratildi.

Internet tizimining rivojlanishi natijasida avtomatik tarjima global tar- moqda ham tashkil etildi, ya’ni hozirgi kunda online-tarjima tizimi ham faol yo‘lga qo‘yilgan.

Avtomatik tarjimaning kamchiliklari:

1. So'zlar avtomatik tarjima jarayonida to‘g‘ridan-to‘g‘ri tarjima qili- nadi. Shuning uchun tarjima grammatik qoidalarga deyarli rioya qilinma- gan holda amalga oshiriladi.

2. Avtomatik tarjima qilingan matnni uslub jihatdan tahrirlab chiqish zarur bo‘ladi.

3. Hozirgi mavjud avtomatik tarjima qiluvchi sistemalar so'zlar biriku- vini ham deyarli noto‘g‘ri tarjima qiladi.

4. Avtomatik tarjima polisemiya, omonimiya, troplar, frazeologizmlar, idiomalarni e’tiborga olmaydi.

Masalan, The bat looks like a mouse with wings gapini rus va o‘zbek tiliga tarjima qilishda har bir so'zning polisemantik va omonimlik kabi se- mantik qirralarini e’tiborga olish zarur. Aks holda tarjima muvaffaqiyatli chiqmaydi. Jumladagi bat so‘zi «ko‘rshapalak», «belkurak», «to'qmoq», «kichik ziyofat» kabi ma’nolarni, look so‘zi esa «qaramoq», «nigoh», «ko‘rinish» ma’nolarini anglatadi. Bunda mazkur so‘zlar rus yoki o‘zbek tiliga tarjima qilinayotganda mos ekvivalentini tanlash uchun kontekst va uning distributsiyasidan (qurshovidan) kelib chiqish zarur. Ushbu holat, ayniqsa, idiomatik ifodalar qatnashgan matnlar tarjimasida murakkablik tug'diradi. Masalan, You are an apple of my eyes jumlasi «Siz ko‘zlarimning olmasisiz» tarzida so‘zma-so‘z tarjima qilinadi, chunki «apple» so'zining denotative ma’nosi «olmawdir. Lekin kontekstdan kelib chiqib tarjimaga yondashilsa, to‘g‘ri tarjima yuzaga chiqadi. Bunda jumlaning «an apple of my eyes» qismi idioma ekanligini nazarda tutish lozim va u «Siz ko‘zimning oq-u qorasisiz» tarzidagi gap hosil boiadi.

Tarjimadagi moslik 3 xil bo‘ladi:

1) ekvivalent moslik - bunda bir tildagi so‘z boshqa bir tildagi so'zga

76

A. Rahimov

tarjimada to‘la muvofiq keladi. Masalan, rus tilidagi «книга» so‘ziga ing­lizcha «book» so'zi tarjimada to‘la mos keladi.

2) variantli moslik - bunda sinonimiya va polisemiya asosida so'zning ma’nosiga mos keluvchi variantlar ham chiqariladi. Masalan, rus tilidagi «добиться» so‘ziga ingliz tilida mos keluvchi tarjima variantlari quyidagi­lar:

добиться ГЛ COB

1. land (чего-либо)

2. win

3. gain (чего-либо расположения)

добиться гл сов (достигнуть, достичь, добиться) attain добиться гл сов (добиваться)

1. contest (премии, места в парламенте и т.п.)

2. court

добиться гл сов (выигрывать, добиваться, выиграть) gain добиться гл сов (достигать, добиваться, достичь) fetch, fetch up добиться гл сов (обеспечивать, добиваться, обеспечить) instate добиться гл сов (достигать, добиваться, достичь) achieve достичь цели - achieve one’s purpose достичь цели - achieve one’s aim добиться гл сов (добиваться, стремиться) seek добиться гл сов (добиваться) seek after, seek for иметь большой спрос - be much sought after иметь успех - be much sought after быть популярным - be much sought after добиться гл сов (выпросить, ухитриться получить) wangle ухитриться получить лишнюю неделю отпуска-wangle an extra week’s holiday

добиться гл сов (достигать, достичь, добиваться) obtain добиться гл сов (добиваться)

2. carry

отстоять свою позицию, добиться своего - carry one’s poin t

3. come at (чего-либо)

как вы это узнали? - how did you come at the information?

добиться гл сов (обрести, получить) find

добиться гл сов (достигать, достичь, добиваться) get

добиться славы - get glory

добиться гл сов (добиваться) secure

добиться гл сов (получать, получить, добиваться) have

Kompyuter lingvistikasi asoslari

77

мы получили известие - we had news ничего не добьешься - there is nothing to be had

3) transformatsion moslik - bu asosan birikmali ifodalar tarjimasida ishlatiladi, bunda mazmunga ta’sir ko'rsatmagan holda shakliy o'zgarishlar asosida tarjima amalga oshiriladi. Masalan, ingliz tilida «the decision of government» ifodasi «the government’s decision» tarzida transformatsiya qilinishi mumkin. Har ikkala holatda ham mazmun («hukumat qarori») o'zgarmaydi.

Shularni nazarda tutgan holda aytish mumkinki, avtomatik tarjima, ayniqsa, rasmiy til doirasida muvaffaqiyatli amalga oshiriladi. So‘zlashuv va badiiy uslubdagi matnlarning avtomatik tarjimasi hali-hanuz maksimal darajada hal etilgani yo‘q. Tipologik jihatdan o'xshash bo'lmagan, bir sistemaga mansub bo'lmagan tillarning avtomatik tarjimasi masalasi bir qator muammolarni keltirib chiqarmoqda. Demak, lingvistik tipologiyada- gi farqlarni izchil o'rganish, lingvistik universaliyalarni aniqlash, tillardagi frazeologizmlar, idiomalar, paremalar, sinonimlar, omonimlar naborini qat’iy belgilash asosida mexanik tarjimadagi muammolarni bartaraf etish mumkin.

Eng sifatli tarjima insonning aralashuvi bilan yuzaga chiqadi. Chunki tarjimon inson tabiiy tilning nozik qirralarini, semantik, grammatik va prag- matik jihatlarini nazarda tutgan holda tarjimaga yondashadi. Bunda yon­dashuv quyidagicha bo'lishi lozim: Mexanik tarjima yordamida tanlangan matnning tarjima qilingan dastlabki versiyasiga ega bo'lamiz, bundan so‘ng inson tarjima jarayoniga aralashadi, ya’ni u mavjud tarjimani tahrir etadi - qo'shimchalar qo'shilish tartibi, gap bo'laklarining joylashuvi, so'zlarning semalariga asosan to‘g‘ri tanlanganligi, polisemiya, omonimiya, sinonimi- yaning e’tiborga olinishi, jumlalarning stilistik jihatdan to‘g‘riligi kabi parametrlarni tekshiradi.

Rus olimi Yu.I.Shemakin o'zining «Начало компьютерной лингвис­тики» kitobida avtomatik tarjimaning 2 xil modelini va tarjima protsedurasini chizmalar asosida tushuntirib bergan. Birinchi yondashuv «matn-mazmun-matn» modeli va semantikaning formal tavsifiga tayanadi. Ikkinchi yondashuv esa «matn-matn» modeli va tarjima mosligiga asoslangan.

Birinchi yondashuvda tilning semantikasiga maksimal darajada e’tibor bergan holda avtomatik sifatli tarjimaga erishishga harakat qilingan. Tizi- mning lingvistik asosi «matndan uning mazmuniga qarab borish» («from surface structure to deep structure») hisoblanadi. Bunda tarjima tilning morfologik, sintaktik va semantik tahlili, lug'atlar, grammatik qoidalar,

78

A.Rahimov

matnlar korpusiga asoslanib amalga oshiriladi. Buni Yu.I.Shemakin qu­yidagi chizma bilan tushuntirgan.66

11-jadval

66 Шемакин Ю.И. Начало компьютерной лингвистики. - М.: МГОУ, 1992.

- С.68.

Kompyuter lingvistikasi asoslari

79

Ikkinchi yondashuvda matn ortiqchaliklarga ega ko‘p bosqichli tizim sifatida baholanadi, bunga ko‘ra matnning yuqori sathida ko‘proq infor­matsiya tashuvchi leksik birliklar, quyi sathida esa ozroq informatsiya beruv- chi birliklar joylashgan boiadi. Bunda so‘zma-so‘z tarjima amalga oshiri- lib, ikki tildagi gaplar strukturasi yagona superstrukturaga birlashtiriladi. Bunda avtomatik tarjima modeli ikki holatda boiadi: a) generatsiya holati tarjimaning ishlashini - lug'atlar jamlash, analiz algoritmlari, aniq bir maqsadga yo‘naltirilgan sintezni ta’minlaydi; b) translatsiya holati siste- maning mustaqil ravishda tarjima ustida ishlashini ta’minlaydi.

Ikkinchi yondashuvga ko‘ra, avtomatik tarjima quyidagi protseduralar asosida ishlaydi:

1-protsedura - qidiruv tizimi bilan bogliq bo lib, bunda lug‘at baza- sidan so‘zlar, birikmalar, iboralar izlanadi, so‘z-shakllarning morfologik analizi amalga oshiriladi.

2-protsedura - so‘zlarning sintaktik vazifasi, grammatik kategoriyalar - zamon, shaxs-son kabilar aniqlanadi.

3-protsedura - bir yoki ko‘p ma’noli so‘zlar tarjimasi bayon qilinadi.

4-protsedura - jumlani morfologik va sintaktik jihatdan shakllantirish jarayoni amalga oshiriladi.

5-protsedura - transformatsion o‘zgarishlar (matn yoki jumla struktu- rasini uning mazmuniga jiddiy ta’sir etmagan holda o'zgartirish; masalan, Kompyuter yoqildi va musiqa yangradi jumlasini «Kompyuter yoqilgach / kompyuter yoqilishi bilan / kompyuter yoqilib musiqa yangradi» tarzida o‘zgartirish) hosil qilinadi (agar zarurat tug‘ilsa).

6-protsedura yakuniy bosqich bolib, bunda murakkab grammatik for- malarning sintezi amalga oshiriladi.

12-jadval

Г Идентификация десигната J

Г Монограмматический анализ

; '1

[ Лексический перевод

[ Биграмматический анализ

Г Трансформационный анализ и синтез J

Г Синтез выходных словоформ и текстов j

80

A. Rahimov

Mashina tarjimasi strategiyasi sekin-astalik bilan o‘zgarib borgan. Ilk tajribalarda qoMlanilgan tarjima strategiyasi «tog 'ridan-to 'g ‘ri tarjima strate­giyasi» deb ataladi, bunda faqat tarjima jarayonida so‘zma-so‘z tarjima mufoviqligi e’tiborga olinadi. Ushbu strategiya tarjima sifati va samara- dorligiga salbiy ta’sir ko'rsatadi. Shuning uchun ushbu strategiyani tanqid qilgan holda «vositachi til yordamida amalga oshiriladigan tarjima strategi­yasi» ishlab chiqildi. Ayniqsa, bu strategiya grammatik strukturasi va ti- pologik xususiyatlari bir-biridan farqlanuvchi tillar o‘rtasidagi avtomatik tarjima tizimlariga faol tatbiq etiladi.

Avtomatik tarjimada inson hamda kompyuter ishtiroki quyidagi tarzda bo'ladi: a) postediting - matn (mashina-tarjimon) kompyuter yordamida tarjima qilinadi, so‘ng inson-muharrir (inson-tarjimon) uni tahrir qiladi; b) preediting - bunda inson matnni mashina-tarjimonga moslaydi, so‘ng uni kompyuterga havola etadi; c) interediting - bunda tarjimada murakkablik tug'diruvchi holatlar duch kelganda inson mashina-tarjimon ishiga arala- shadi.67

Matnni bir tabiiy tildan boshqasiga tarjima qilish jarayonidagi mu- rakkabliklar va muammolarga nazariy va amaliy jihatdan yechim topish bo‘yicha hali-hanuz izlanishlar davom etmoqda. 1984-yilda Yaponiya- ning Kioto universiteti professori, mashina tarjimasi bo'yicha yapon dav­lat dasturi rahbari Makoto Nagao bu yo‘nalishda yangi konsepsiya ishlab chiqdi.68 Ushbu konsepsiyaga ko'ra, matnlar ilgari inson qo‘li bilan tarji­ma qilingan matnlarga tayanib analogiya tamoyili asosida tarjima qili- nishi lozim. Buning uchun kompyuterga kiritilgan katta hajmdagi bilingv bazasi (matn va uning tarjimalari) shakllantirilishi zarur. Yangi matnlar tarjimasida bilingv massividan o'xshash matn fragmentlari tanlanadi va tarjimaning muvaffaqiyatli chiqishida foydalaniladi. M.Nagao mashina tarjimasiga bunday yondashuvni «Example Based Translation» («Namu- nalarga asoslanadigan tarjima») deb nomladi, ilgarigi an’anaviy yon- dashuv «Rule Based Translation» («Qoidalarga asoslanadigan tarjima») deb ataladi.

Professor M.Nagao o‘zi taklif etgan yondashuvning ham mutlaq emas- ligini ta’kidlaydi: «Mashina tarjimasi tizimi bo'yicha yaratilgan har qan­day ishlanmalar, yondashuvlar ertami, keehmi boshi berk ko'chaga (tupik-

67 Hutchins J. Retrospect and prospect in computer-based translation // Proceedings of MT Summit VII «МТ in the great translation era». - Tokyo: AAMT, 1996.-P.32.

“ Nagao M. A framework of a mechanical translation between Japanese and English by analogy principle, in Artificial and Human Intelligence, ed. A. Elithorn and R. Banerji. - North Holland, 1984. - P. 173 - 180.

Kompyuter lingvistikasi asoslari

81

ka) kiradi. Bizning yondashuvimiz ham bundan xoli emas, faqat biz bu jara- yonni kechiktirmoqchimiz».69

M.Nagaoning yondashuvi boshqa yangi yondashuvlarning yuzaga kelishiga turtki bo‘ldi. Jumladan, amerikalik olim Veb Laynning «Transla­tion Memory» («Tarjimon xotirasi») yoki «Sentence Memory» («Gap to'plagich») konsepsiyasi.70 Bu yondashuvga ko‘ra, muayyan matnni bir tildan ikkinchisiga tarjima qilish uchun dastlab yuqori malakali tarjimonga ko‘rsatiladi. So'ng asl matn va uning tarjimasi kompyuterga kiritiladi, matn fragmentlarga (alohida jumlalarga, birikmalarga) bo‘linadi, ushbu elemcnt- lardan ma’lumotlar bazasi tayyorlanadi. Ma’lumotlar bazasi qidiruv tizi- miga yuklanadi, u esa tarjimasi mavjud bo'lgan matn fragmentlarini va alohida jumlalarni sifatli tarjima qilishga imkon bcradi. Matnning ayrim notanish fragmentlari (bazada tarjimasi mavjud bo'lmagan matn qismlari) inson qo‘li bilan tarjima qilinadi. Bunda bazadagi bu fragmentlarga yaqin keladigan tarjimalardan foydalanish mumkin. Matnning inson tomonidan tarjima qilingan qismlari yana ma’lumotlar bazasiga kiritiladi. Mana shu yo‘l bilan «tarjimon xotirasi» asta-sekinlik bilan boyitib boriladi va tarjima samaradorligi yuksalib boradi. Lekin ushbu yo‘nalishning afzalliklari bilan birga ayrim kamchiliklari ham mavjud. Birinchidan, bu yo‘l bilan har qan- day turdagi matnni tarjima qilib boimaydi. U asosan muayyan tor yo‘nalishdagi bir turdagi matnlarga moMjallangan. Ikkinchidan, ma’lum matnning katta fragmenti ikkinchi bir matnning katta qismiga tarjimada mos kelishi yoki yaqin kelishi har doim ham kuzatilavermaydi. Uchinchidan, «tarjimon xotirasi» bazasini yaratish yoki ikki tilli matnlar (bilingv) mas- sivini toMdirib borish amaliy jihatdan ko‘p qiyinchiliklarni keltirib chiqaradi.

Rossiyada professor G.G.Belonogov frazeologik mashina tarjimasi kon- sepsiyasini asoslab berdi.71 Unga ko‘ra, matndagi tushunchalar ko‘pincha

69 Qarang: Сиокум Дж. Обзор разработок по машинному переводу: история вопроса, современное состояние и перспективы развития // Новое в зарубежной лингвистике. Вып.XXIV. Компьютерная лингвистика. - М.: Прогресс, 1989.

70 Webb Lynn Е. Advantages and Disadvantages of Translation Memory: a Cost/Benefit Analysis. San Francisco State University. 1992.

71 Qarang: Белоногов Г.Г., Зеленков Ю.Г., Кузнецов Б.А., Новоселов А.П., Хорошилов Ал-др А., Хорошилов Ал-сей А. Автоматизация составления и ведения словарей для систем фразеологического машинного перевода текстов с русского языка на английский и с английского на русский // Научно- техническая информация. Серия 2. №12. ВИНИТИ. 1993; Белоногов Г.Г., Кузнецов Б.А. Языковые средства автоматизированных информационных систем. - М.: Наука, 1983; Белоногов Г.Г., Зеленков Ю.Г., Новоселов А.П., Хорошилов Ал-др А., Хорошилов Ал-сей А. Системы фразеологического машинного перевода. Состояние и перспективы развития // Научно- техническая информация. Сер. 2. №12. ВИНИТИ. 1998.

82

A. Rahimov

alohida so‘zlar yordamida emas, balki so‘z birikmalari yordamida ifodala- nadi. O'z-o'zidan kelib chiqadiki, matnni bir tildan boshqasiga tarjima qi­lish jarayonida ham mazmunni yuzaga chiqaruvchi asosiy til birliklari ham alohida so‘zlar emas, balki tipik vaziyatlarni ifodalovchi frazeologik birik- malar (bu yerda frazeologik birlik keng ma’noda tushuniladi) hisoblanadi. Mazkur yondashuvga muvofiq matnda tez-tez uchraydigan so'zlar, iboralar, so‘z birikmalari, nutqiy etiket namunalarining tarjima ekvivalentlari bilan boyitilgan ma’lumotlar bazasi frazeologik mashina tarjimasi tizimining asosini tashkil etadi. Bunda so'zlarning distributsiyasiga eng ko‘p e’tibor beriladi. Bu metod polisemantik so'zlarning tarjimada mos keluvchi to‘g‘ri ekvivalentini tanlashga, omonim so'zlarning kontekstdagi holatini aniqlash- ga, iboralarning ham idiomatik xususiyatini belgilashga imkon beradi.

Mazkur konsepsiyaga asoslangan frazeologik mashina tarjimasining dastlabki versiyasi 1993-yil RETRANS nomi bilan yaratildi. 1998-yilgacha RETRANS MS DOS operatsion tizimi sharoitida ishlashga moijallangan edi. 1998 - 2000-yillarda uning Windows va UNIX operatsion tizimida ham ishlaydigan versiyalari yaratildi. 2001-yilda RETRANS tizimining turli modifikatsiyadagi yangi versiyasi yaratildi: a) MS Word matn muharririda ishlaydi; b) MS Internet Explorerning Web-brauzerida ishlaydi; c) MS Note­pad, MS Wordpadda Clipboard (matnni ko'chirib ishlash)ga asoslanib ishlay­di. RETRANS tizimi tabiiy va texnika fanlari, iqtisodiyot, siyosat, qonun- chilik hamda harbiy sohaga oid 4 mln lug‘at maqolasiga ega bo‘lgan po- litematik so‘zlik bilan ta’minlangan. Kiritilgan til birliklarining 80% qismi- ni so‘z birikmalari tashkil etadi, ular 2 ta so‘z uzunligidan tortib to 17 ta so‘z uzunligiga ega boigan birikmalardir. Tizim tarkibida 400 ming lug'at maqo­lasiga ega boigan qo'shimcha tematik lug'at ham mavjud.

RETRANS tizimining ish tartibini ko'rib chiqaylik. Ruschadan inglizcha- ga tarjima qilish jarayonida dastlab matn avtomatik tarzda morfologik, so'ng semantik-sintaktik tahlil qilinadi, frazeologik birliklar aniqlanadi. Matnning barcha so'zlari hamda frazeologik birliklari qidiruv obrazlari shakllanadi. Lug'at bazasidagi joylashuv tartibga ko'ra so'z va birikmalar qidiruv obrazlarining xeshkodlari topiladi, ularning dekodlanishi asosida qidiruv natijalari namoyon bo'ladi. Keyingi bosqich matnning har bir frag- mentiga mos yagona yoki bir nechta tarjima ekvivalentlari tanlanadi. Tar­jima natijalarini quyidagi jadvalda yaqqol ko'rish mumkin.

Matnning rus tilidan ingliz tiliga qilingan tarjima oraliq natijalari

fragmenti Перевод 00001 00002 text translation текстов 00002

Kompyuter lingvistikasi asoslari

c 00003 00004 00005 00006 00007 00008 from some natural languages into others

одних 00004 естественных 00005 языков 00006 на 00007 другие 00008

- 00009 - / is

сложный 00010 complicated / complex / compound / difficult

творческий 00011 00012 creative process процесс 00012 ,00013

требующий 00014 requiring / demanding / claiming / making the demand от 00015 from переводчика 00016 interpreter / translater

не 00017 00018 not only / not just только 00018

соответствующей 00019 corresponding / appropriate /

respective / adequate лингвистической 00020 linguistic

подготовки 00021 training / preparation/ getting ready / preparing

,00022

но 00023 00024 but as well / but also и 00024

хороших 00025 good / nice / well знаний 00026 knowledge / knowing тех 00027 those предметных 00028 00029 subject field / object areas / application domain

областей 00029 , 00030

к 00031 00032 to which которым 00032 переводимые 00033 00034 translated texts тексты 00034 относятся 00035 relate / belong / concern .00036

84

A. Rahimov

Shundan so‘ng tarjima ikki rejimda amalga oshiriladi: 1) avtomatik tar­jima rejimi; 2) oraliq tarjima natijalarini intellektual tahrirlash imkoniyati mavjud boigan interaktiv rejim. Avtomatik tarjima rejimida ruscha matn- dagi har bir so‘z va birikma uchun faqat bitta (odatda, eng asosiy, sinonimik qatordagi dominantlik xususiyatiga ega bo‘lgan so‘z) inglizcha ekvivalenti tanlanadi va matnning morfologik, semantik-sintaktik analizi asosida sin- tez jarayoni boshlanadi, so‘ng gap strukturasi shakllanadi va tarjima yuza­ga chiqadi. Interaktiv rejim ham xuddi avtomatik tarjima singari boiadi, bunda faqat foydalanuvchi (tarjimon-muharrir) taklif etilgan tarjima vari- antlarining semantik jihatdan eng mos keluvchi ekvivalentini o‘zi tanlash imkoniga ega boiadi, shuningdek, gap strukturasi, so‘z tartibini ham lingvis- tik bilimidan kelib chiqib o‘zi tahrir qilishi mumkin.

Shuni alohida ta’kidlash zarurki, yuqoridagi har uchala konsepsiya ham avtomatik tarjima sifati va samaradorligini oshirish, til bilan bogiiq to'siqlarni yo'qotish yoiidagi muvaffaqiyatli izlanishlar hisoblanadi.

Topshiriqlar:

1. Avtomatik tarjima bo'yicha qanday yondashuvlar mavjud?

2. Qanday tarjimon dasturlarni bilasiz?

3. Avtomatik tarjimaning ishonchlilik darajasi haqida mustaqil muloha- za yuriting.

4. Avtomatik tarjimaning toiaqonli boiishiga to'siq bo'luvchi omillar- ni sanang.

5. Avtomatik tarjima qanday protseduralar asosida ishlaydi?

Adabiyotlar:

1. Нелюбим Jl. Jl. Компьютерная лингвистика и машинный перевод.

- М.: ВЦП, 1991.

2. Шемакин Ю.И. Начало компьютерной лингвистики. - М.: МГОУ, 1992.

3. Пиотровский Р. Г. Текст, машина, человек. - М.: Просвещение, 1989.

4. Марчук Ю.Н. Проблемы машинного перевода. - М.: Наука, 1983.

5. Баранов А.Н. Введение в прикладную лингвистику. - М.: Эдиториал УРСС, 2001.

6. PoMatov A., Muhamedova S. Kompyuter lingvistikasi. - Toshkent, 2007.

7. Nagao M. A framework of a mechanical translation between Japanese and English by analogy principle, in Artificial and Human Intelligence, ed.

A. Elithorn and R. Banerji. - North Holland, 1984. - P. 173 - 180.

Kompyuter lingvistikasi asoslari

85

KOMPYUTER LEKSIKOGRAFIYASI

Reja:

1. Kompyuter leksikografiyasi kompyuter lingvistikasi fanining alohida yo‘nalishi sifatida.

2. Kompyuter lug‘atlari va ularning ishlash tamoyili.

3. Kompyuter leksikografiyasi sohasining rivojlanishi.

Tayanch so‘z va iboralar: ma’lumotlar ombori, leksikografiya, kompyu­ter leksikografiyasi, kompyuter lug'atlari, kodlash va dekodlash, lemmati- zatsiya, matnlar korpusi, korpus lingvistikasi, virtual matnlar korpusi, paral­lel matnlar korpusi, on-line lug'atlar, kitobiy lug'atlar, elektron lug'atlar.

Ma’lumki, leksikografiya tilshunoslik fanining amaliy sohalaridan biri hisoblanadi. U yunoncha «lexikos» - «so‘z», «lug‘at», «grapho» - yozaman degan ma’nolarni anglatadi. Leksikografiya bo'limida lug'atlar, ularni tu- zish yo'llari va tamoyillari o'rganiladi. Avtomatizatsiya ijtimoiy hayotning barcha jabhalariga kirib kelishi natijasida lug'atlar ham elektron holatda yaratilish imkoniyati yuzaga keldi. Kompyuter yordamida lug'atlar bilan ishlashning optimallashuvi natijasida kompyuter leksikografiyasi yo'nalishi shakllandi. Kompyuter leksikografiyasi amaliy tilshunoslikning muhim tarkibiy qismi bo'lib, unda lingvistik va dasturiy ta’minot asosida ishlaydi­gan, kodlash va dekodlash prinsipi asosida yaratilgan kompyuter lug'atlari, ularni tuzish dasturlari, algoritmlari o'rganiladi. Mazkur sohaning rivojla­nishi natijasida turli nomlarda elektron lug'atlar yaratilmoqda. Eng mash- hur elektron lug'atlar sirasiga CONTEXT, ABBY LINGVO, MULTI- TRAN, POLYGLOSSUM, MULTILEKS kabilar kiradi.

Leksikografiya sohasiga kompyuter texnologiyalarining tatbiq etilishi lug'at tuzish ishlarini ancha yengillashtiradi. Lug'at tuzish jarayonining naqadar murakkabligini bir tilshunos olim hazilomuz tarzda quyidagicha izohlagan: «Agar biror kishi og'ir jinoyat sodir qilsa, uni katorgaga surgun qilish shart emas. U lug'at tuzsin, shunda и katorganing barcha azoblarini, qiyinchiliklarini tortgan bo'ladi». Axborot manbalarini yig'ish bo'yicha maxsus dasturlar (Database Software) lug'at asosini tashkil etuvchi barcha ma’lumot va misollarni jamlash va sistematik tarzda ishlov berishga qulay- lik tug'diradi. Bundan tashqari, boshqa maxsus dasturlar lug'atni tahrir qi­lish va chop etish borasidagi ancha mashaqqatli mehnatni bir qancha oson-

86

A. Rahimov

lashtiradi. Elektron lug‘atlar ham, o‘z navbatida, an’anaviy lug'atlarga qaraganda birmuncha afzalliklarga ega. Bugungi kunda, masalan, ABBY LINGVO kompaniyasining lug'at va tarjima dasturlari ulardan foydalanuv- chi har bir xaridorga o‘z lug'atini tuzish yoki mavjud lug'at so‘z boyligini toidirib borish imkonini beradi. Ochiq turdagi Internet lug'atlari (on-line lug‘atlar) ham ko‘p hollarda barcha foydalanuvchilarga ushbu lug‘atlarni toidirish imkonini beradi. Shuningdek, elektron lug'atlarning hajm jihatidan ixchamligi, kompaktligi, boshqa manbalarga (Internet tarmog'i orqali, giper- murojaatlar yordamida) ulanish imkoniyati, mavjud matnlar korpuslari yordamida illustrativ misollar tuza olish imkoniyati, multimediali misollar bilan boyitilganligi, so'zlarning sinonimik variantlari, omonimlik xusu- siyatlari, grammatik ma’lumotlar bilan ta’minlanganligi va ularning tezlik bilan foydalanuvchiga havola etilishi kabi jihatlari elektron lug‘atlarning optimalligini ta’minlaydi.

Kompyuter leksikografiyasini elektron matnlar korpusi yoki parallel matnlar korpuslarisiz tazavvur qilish mumkin emas. Matnlar korpusi («cor­pus» lotincha «tana» degan ma’noni anglatadi) - bu elektron holda saqla- nadigan ma’lum til birliklari bo‘lib, ular tilshunoslar uchun turli xil muam- molarni hal etish uchun tatbiq etishda va turli yo'nalishdagi tadqiqotlar uchun zaruriyatga qarab turli shakllarda tuziladi. Bular fonema, grafema, morfemalardan tortib undan kattaroq birliklar - leksema, gap va matnlar- dan (badiiy yoki ilmiy asar, gazeta va jurnal matnlari) tashkil topishi mumkin. Ularning qay tarzda saqlanishiga qarab maxsus dasturlar yorda­mida har bir kerakli so‘z yoki so‘z birikmasi uchun darhol uning qoilanishi bo‘yicha misollar topilishi, imlo bo‘yicha variantlari, sinonimik qatorlari topilishi mumkin. Matnlar korpusiga oid ilmiy tadqiqotlar salmog‘ining ko'payishi natijasida tilshunoslikda korpus lingvistikasi yo‘nalishi shakl- landi.72

Kompyuterda yaratilgan birinchi matnlar korpusi Braun korpusi (БК, inglizcha Brown Corpus, BC) hisoblanadi, u 1961-yilda Braun universiteti- da yaratilgan, har biri 2000 so'zli 500 ta matn fragmentini o‘z ichiga oladi. 1970-yillarda 1 mln so‘zni o‘z ichiga olgan matnlar korpusi asosida rus tilining chastotali lug'ati yaratildi. 1980-yillarda Shvetsiyaning Upsala uni- versitetida ham rus tilida matnlar korpusi yaratildi. Keyinchalik kompyuter leksikografiyasining rivojlanishi natijasida katta hajmli matnlar korpusiga ehtiyoj tug‘ildi. Ya’ni 1 mln ta so‘z elektron lug‘atlar bazasi uchun yetarli emas. Shu asosda yirik hajmli matnlar korpusi yaratila boshlandi. Ko'pgina

72 Баранов A.H. Введение в прикладную лингвистику. - М.: Эдиториал УРСС, 2001. -С.61.

Kompyuter lingvistikasi asoslari

87

mamlakatlarda XX asrning 80-yillaridan boshlab bunday korpuslar tuzila boshlandi. Ular turli maqsad va vazifalarga xizmat qiladi. Buyuk Britani- yada Ingliz tili Banki (Bank of English) hamda Britaniya Milliy Korpusi (British National Corpus. BMC), Rossiyada Rus tilining mashina fondi (Машинный фонд русского языка) hamda Rus tilining Milliy Korpusi (Национальный корпус русского языка) loyihalari ishlab chiqildi.73 Masalan, Rus tilining milliy korpusi hajmi hozirgi kunda 149 mln so‘zdan iborat. Keyingi yillarda Internet tizimining rivojlanishi virtual matnlar kor­pusi yuzaga kelishiga olib keldi. Ya’ni Internetdagi qidiriv saytlari, elek- tron kutubxonalar, virtual ensiklopediyalar korpus vazifasini bajarmoqda. Korpusning janri va tematik rang-barangligi Internetdan foydalanuvchi- ning qiziqishlariga bog‘liq. Masalan, ilm-fan doirasida Wikipedia katta hajm- dagi matnlar korpusi sifatida foydalanilmoqda.74

Korpus lingvistikasida parallel matnlar korpusi ham muhim ahamiyat kasb etadi. Parallel matnlar korpusi esa, o‘z navbatida, badiiy asar, qo‘llanma, ommaviy axborot vositalari, turli xil hujjatlarning ikki yoki undan ko‘p tillardagi elektron holdagi ko‘rinishlaridir. Masalan, Yevropa Ittifoqi o'zining barcha qonun va hujjatlarini ingliz, fransuz, nemis, ispan va ital- yan tillarida nashr qiladi hamda ular Internet tizimiga barchaga ochiq arxiv sifatida qo‘yiladi. Bunday korpuslarning afzalligi shundaki, ular yordami­da nafaqat biron bir so‘z yoki jumlaning, balki butun boshli matnlarning turli tillardagi variantlarini bilish imkoniyati mavjud. Xuddi mana shu im- koniyat tufayli maxsus konkordanser dasturlar ishlab chiqish orqali turli xil ixtisoslik lug‘atlari tuzish imkoniyati tug‘iladi. Ushbu imkoniyatlar komp­yuter leksikografiyasi uchun ulkan ahamiyat kasb etadi.

Kompyuter leksikografiyasiga semantik maydon, semantik tarmoq, se­mantik to‘r hamda freym semantikasining faol tatbiq etilishi natijasida ul­kan kompyuter leksikografiya resurslari yaratildi. Shunday yirik lek- sikografik resurslardan biri FRAMENET bo‘lib, u Internet tizimida on-line rejimida ishlaydi.75 Mazkur tizim Ch.Fillmorning «Tools for Lexicon Buil­ding» loyihasi asosida Kaliforniya shtati, Berkli shahridagi Xalqaro infor- matika institutida ishlab chiqilgan. Framenet resursining ma’lumotlar ba- zasida 10 000 ta leksik birlik mavjud, undan 6000 dan ortig'i toiiq anno- tatsiyaga ega. Bundan tashqari, ma’lumotlar bazasida 800 ta semantik freym ko‘rsatilgan, 135 000 annotatsiyali gaplar keltirilgan.

Elektron Iug‘atlar tuzish jarayonida ma’lumotlar ombori, matnlar kor-

73 http:[www.corpus.leeds.ac.uk/list.html](http://www.corpus.leeds.ac.uk/list.html)

74

linguistics

75 <http://framenet.icsi.berkeley.edu/>

88

A. Rahimov

pusini yaratish, qidiruv tizimi, kodlash, lingvistik va dasturiy ta’minot un- surlari qatorida lemmatizatsiya bosqichi ham mavjud. Lemmatizatsiya - bu so‘zning dastlabki, boshlang'ich formasini (lug‘atdagi shaklini - lemmasi- ni) tashkillashtirish texnikasi boiib, bu jarayon o‘sha so'zning boshqa so‘z- shakllaridan kelib chiqqan holda amalga oshiriladi. Lemmatizatsiya mor­fologik tahlil metodi tarkibiga kiradi, u ikki bosqichni o‘z ichiga oladi: 1) deklarativ bosqich - bunda muayyan so'zning mumkin boigan barcha shakl- lari (so‘z-shakllar) belgilanadi; 2) protsedura bosqichi - bunda so‘z asos va qo'shimchalarga, ya’ni leksemalarga yoki morfemalarga boiinadi. Lem­matizatsiya so'zlarning grammatik valentligi, qaysi affikslar bilan birika olish imkoniyatini ham belgilab beradi. Masalan, o'zbek tilida so‘zlarning lug‘atdagi shakli - lemmasi quyidagicha:

- ot so‘z turkumi uchun - bosh kelishik, birlik shakli;

- fe’l so‘z turkumi uchun - harakat nomi shakli;

- sifat so‘z turkumi uchun - oddiy daraja shakli.

Daftarlarni, daftarlarga, daftarlarning, daftarlardan, daftarlarda! daftar

Yugurdi, yugurgan, yuguryapti, yugurmoqchi ! yugurmoq

Ko‘kimtir, ko'kish, ko‘kroq ! ko‘k

Kitobiy lug‘atlar va elektron lug'atlarning farqi quyidagilarda ko'rinadi:

1. Kitobiy lug'atlarning tuzilishi: a) lug‘atning so'zligi shakllantiriladi; b) misollar kartotekalari tuziladi; c) lug'at maqolalari yoziladi; d) lug‘atning qolyozma varianti tayyorlanadi; e) qoiyozma tahrir etiladi; f) muallif tuza- tishlar qiladi; g) nashr uchun teriladi; h) lug‘at sahifalanadi; i) korrektirovka qilinadi; j) lug‘at nashr etiladi.

2. Elektron lug'atlarning tayyorlanishi: a) lug'atning so‘zligi shakllanti­riladi; b) misollar korpusi (elektron kartotekalari) tuziladi; c) lug'at maqola­lari yoziladi; d) lug‘at maqolalarini ma’lumotlar bazasiga ko'chiriladi;

e) bevosita ma’lumotlar bazasida lug'at matni tahrir etiladi, korrektirovka qilinadi; 0 lingvistik ta’minot dasturiy ta’minot bilan uyg'unlashtiriladi;

g) elektron lug‘at.

Kitobiy lug'atlar tuzilishi sahifalar ketma-ketligiga tayanadigan chiziq- lilik tamoyiliga bo‘ysunadi. Elektron lug‘atlar strukturasi gipertekst tex- nologiyasiga asoslangan boiadi, bu esa foydalanuvchiga lug'at maqola- larining ixtiyoriy qismiga tezkor murojaat qilish imkonini beradi.

Elektron lug'atlarning ishlash prinsiplarini umumlashtirib quyidagicha izohlash mumkin:76 Tilning har bir so'ziga mutanosib keluvchi kod ishlab chiqiladi va qoilanadi, kodni qayta ishlash jarayonida zaruriy boigan

76 Poiatov A.. Muhamedova S. Kompyuter lingvistikasi. -Toshkent, 2007. -

B.58-61.

Kompyuter lingvistikasi asoslari

89

Ma’lumotlar ombori

Dasturiy ta’minot (matematik modellashtirish va algoritm- lash)

Lingvistik ta’minot (faktografiya, matnlar korpusi)

Simvollar orqalj so'zlarni berish (kodlash)

So'zlik tayyorlash (lemmatizaciya)

Qidiruv tizimi

Natija (dekodlash)

ma’lumotlar, tarjimalar, sinonim, antonim va sharhlarga ega bo'lish mumkin.

So'zlarni kodlashtirish quyidagicha amalga oshiriladi. Ma’lumotlar tekst fayllarga joylashtiriladi, uning har bir elementi 3 qismdan iborat bo‘ladi: 1) so'zning tartib raqami; 2) so'z; 3) kod.

«So‘zning tartib raqami» (ya’ni uning adresi). So‘zlarning tarjimasi, si- nonimlari va antonimlarini ko‘rsatish ularning birinchi harflarini ko- dirovkadagi tartib raqamlari bilan birgalikda keltirish bilan amalga oshi­riladi.

«So'z» - mutanosib alfavit harflari bilan yozilgan oddiy so'z.

«Kod» - raqam va harflar ketma-ketligi bo'lib, unda so'z barcha zaruriy morfologik, sintaktik, leksik xususiyatlari hamda ushbu so'zning qaysi so'zga tegishliligi haqidagi ma’lumotlar jamlangan bo'ladi: grammatik ma’lumot, adreslar, tarjima, sinonim, antonim, mutanosib sharhlar.

Kodlarni yaratish CREATE va CREATE 1 dasturlari orqali amalga oshiriladi. Ular quyidagi tartibda ishlaydi:

So'z kiritiladi.

1. Grammatik ma’lumotlar yaratiladi - kompyuter so'z haqida mor­fologik va sintaktik ma’lumotlarni so'raydi va ular darhol aniq nisbat asosi­da shifrlanadi.

2. Tayyor shifr «Grammatik ma’lumotlar» bo'limiga yozib qo'yiladi.

3. Tarjimalar sinonimlar, antonimlar adreslari yaratiladi:

a) tarjimasi boiishi mumkin bo'lgan so'z so'raladi;

90

A. Rahimov

b) kiritilgan so‘z kodning mutanosib boiimiga yozib qo'yiladi;

v) kodlashning oxirida tarjimalar (sinonimlar, antonimlar) kodi topiladi va kiritilgan so‘zlar o‘rniga ularning birinchi harflari va tartib raqamlari yoziladi (masalan, «katta» so‘zi o‘rniga K0083), agar tarjimalar (sinonim, antonimlar) kodi topilmasa, u holda shu so'zga nisbatan kodlash operatsi- yasi amalga oshiriladi va bu bilan bosqich tamomlanadi.

5. Mazkur so'zga sharh kiritiladi:

a) matn kiritiladi;

b) maxsus dastur asosida so'z va unng sharhi orasidagi moslik belgilana- di, so'ngra kiritilgan matn xotiraga yoziladi, ular orasidagi moslik esa so‘z kodida o‘z ifodasini topadi.

Dastur ishlay boshlashi bilan ekranda u va uni ishlab chiquvchilari haqi- da ma’lumotlar paydo bo'ladi. Ular bilan tanishilgandan so'ng klavish bosi- ladi va dastur quyidagi bosqichlarda o‘z ishini davom ettiradi:

1. Ekranga 6 rejimga ega oyna chiqdi.

1-rejim. Only translation (faqat tarjima) - podstrochnikda foydalani- ladigan tarjimalarni va mazkur so'zning qaysi sohaga tegishliligi haqidagi ma’lumotlarni beradi. Mazkur rejimning asosiy funksiyasini Only-tran maxsus protsedurasi amalga oshiradi. Ushbu protsedura, avvalo, mazkur so‘z kodining adreslar bo'limidan tarjimalar adresini izlab topadi. So‘ng uni qayta ishlashni boshlaydi: tarjimaning birinchi harfini olib, tekst faylini ochadi (shu harf bilan nomlangan) va berilgan tartib raqamiga ko'ra tarji- mani topib, uni ekranga chiqaradi.

2. Grammatical information (grammatik ma’lumotlar) - so'zning barcha morfologik va sintaktik xususiyatlari va undan foydalanishdagi ayrim noa- niqliklar haqidagi ma’lumotlarni beradi. Gram info maxsus protsedurasi ishlaydi. Grammatik ma’lumotlar kodi bo'limida jamlangan ma’lumotlarni rasshifrovka qiladi.

3. List of synonyms (sinonimlar ro'yxati).

4. List of antonyms (antonimlar ro‘yxati).

Ushbu rejimlar Syn-List va Ant-list maxsus protseduralar yordamida sinonimlar va antonimlar ro'yxatini beradi. Ular quyidagi tartibda ishlay­di: sinonim va antonimlar kiritilgan so'z kodning «sinonimlar adresi» va «antonimlar adresi» bo'limlaridan topiladi hamda ekranga chiqariladi.

5. New word formation (yangi so‘zlarni qayta yasash) - mazkur so'z bi­lan bir xil o'zakka ega bo'lgan barcha so'zlarni chiqarib beradi. Asosiy ish NWF protsedurasi vositasida amalga oshiriladi:

a) so'zning o'zagi ajratiladi;

b) so'z haqidagi barcha ma’lumotlarga ega bo'lgan holda ushbu o'zakka

Kompyuter lingvistikasi asoslari

91

old qo'shimcha va boshqa qo‘shimchalarni qo‘shish orqali o'zgartirilishi mumkin bo‘lgan barcha so'zlarni chiqarib beradi.

6. Comments (sharhlar) - o‘rganilayotgan so‘z tez esda qolishi uchun «aytib berish» (podskazka) ko'rinishidagi sharhlarni chiqarib beradi.

II. lsh uchun zarur bo'lgan rejim tanlanganidan so‘ng so‘z kiritiladi. Ayrim elektron lug‘atlarda leksikon bazasi ozligi sababli ayrim so‘zlarning xotirada kodi bo'lmasligi ham mumkin. Shuning uchun kirishdan so'ng Find Word protsedurasi ish boshlaydi. Uning vazifasi xotiradagi so‘z kodini izlash- dan iborat. Agar u topilsa, maxsus protseduralar uzatiladi, aks holda quyi- dagi xabar chiqadi «Sorry, 1 do not know (Uzr, men buni bilmayman) va keyingi so'z kiritiladi.

III. Topilgan kod tanlangan rejimning maxsus dasturlari bilan qayta ishlanadi, buning natijasida talab qilingan ma’lumotlar chiqariladi.

IV. Tanlangan rejimlarda ishni davom ettirish haqida so'raladi. «На» javobidan so'ng dastur ishi ikkinchi bosqichda davom ettiriladi. Aks holda keyingi bosqichga o'tiladi.

V. Dastur ishini tamom qilish haqida so'raladi.

Kompyuter leksikografiyasi bugungi kunda turli qo'shimcha imkoniyat- lar asosida rivojlanib bormoqda. Xususan, dastlab kompyuter lug'atlari bir tilli, ikki tilli bo'lgan, hozirdako'p tilli (uch, to'rt, olti, o'n tilli) elektron lug'atlar yaratilmoqda. Shuningdek, keyingi paytlarda muayyan sohalarga ixtisoslashgan elektron lug'atlar ham yaratilmoqda. Ya’ni dastlab kompyu­ter lug'atlari faqat umumiy leksikon bilangina cheklangan edi, hozirda fanga oid bo'lgan, qurilish, aviatsiya, avtomobilsozlik, harbiy, diniy, yuridik so­halarga oid so'zlar bazasi ham e’tiborga olinmoqda. Hatto so'zlarning matn- da qo'llanish imkoniyatlari, distributiv holatlari, birikma holidagi ko'rinishlari ham nazarda tutilgan holda lug'atlar yaratilmoqda. Bu ko'rsatkichlar komp­yuter leksikografiyasi sohasining taraqqiy etayotganidan darak beradi.

Topshiriqlar:

1. Tezaurus lug'at deganda qanday lug'at tushuniladi?

2. Kompyuter leksikografiyasining rivojlanish tendensiyasi haqida yo- zing.

3. O'quv kompyuter lug'atlari haqida gapiring.

4. Kitobiy va elektron lug'atlar orasidagi mushtarak nuqtalar qaysilar?

5. Kitobiy va elektron lug'atlar orasidagi farqli jihatlar qaysilar?

Adabiyotlar:

1. Нелюбин JI. Л. Компьютерная лингвистика и машинный перевод.

- М.: ВЦП, 1991.

A. Rahimov

2. Шемакин Ю.И. Начало компьютерной лингвистики. - М.: МГОУ, 1992.

3. PoMatov A., Muhamedova S. Kompyuter lingvistikasi. - Toshkent, 2007.

4. Пиотровский P.Г. Текст, машина, человек. - М.: Просвещение, 1989.

5. Баранов А.Н. Введение в прикладную лингвистику. - М.: Эди гориал УРСС, 2001.

6. Белоногов Г.Г. Компьютерная лингвистика и перспективные информационные технологии. - М.: Русский мир, 2004.

TIL ORGATISH JARAYONINIAVTOMATLASHTIRISH

Reja:

1. Kompyuter lingvodidaktikasi va ta’lim jarayonida kompyuter tex- nologiyalaridan foydalanishning ahamiyati.

2. Til o‘rgatishda avtomatlashtirish.

3. Til o'rgatuvchi o‘quv dasturlari va ulardan foydalanish.

Tayanch so‘z va iboralar: til о 'qitishni avtomatlashtirish, о ‘quv dasturlari, «elektron»o'qituvchi, multimedia, virtualkutubxona, CALL, avtomatizatsiya, bixevioristik yondashuv, kognitiv-intellektual yondashuv, avtomatik korrek- torlar, speller, audio va vizual ejfektlar, interaktiv multimedia texnologiyasi, virtual ensiklopediya.

XX asr tarixga ilm-fan inqilobi davri, kompyuter, avtomatizatsiya asri sifatida kirdi. Bu davrda insoniyat o'zining og'ir mehnatini o'z aqli bilan osonlashtirishga erishdi. Ya’ni ishlab chiqarish jarayonida foydali ish koef- Htsiyenti bir necha barobarga oshdi, mehnat unumdorligi o'sdi. Insoniyat yirik zavod, fabrikalarni, korxonalarni avtomatlashtirishga erishdi. Bunda ilgari manufaktura (qo'l mehnati)ga asoslangan ishlab chiqarish tizimidan avtomatlashgan, kompyuterlashtirilgan tizimga o'tildi. Hozirgi kunda av­tomatizatsiya faqatgina ishlab chiqarish emas, balki jamiyatning deyarli barcha sohalariga kirib kelmoqda. Xususan, ta’lim jarayonida bilimlarni

Kompyuter lingvistikasi asoslari

93

diagnostika qilishning, xorijiy tillarni o‘rgatishning ham samarali avtoma­tik tizimi ishlab chiqilgan.

«Avtomatizatsiya» so'zi lotincha «o‘zi harakatlanadi» degan ma’noni bildiradi. So'zning ma’nosidan anglashilib turibdiki, tizim mustaqil hara- katlanishga asoslangan, turli operatsiyalar bajarishga yo‘naltirilgan va shu tamoyilda ishlashiga moslab programmalashtirilgan mexanizmlar yig'indisidir. Hammaga ma’lumki, yirik korporatsiyalarda, aytaylik, avto- mobil ishlab chiqaruvchi zavodlarda avtomatizatsiya tizimi keng yo'lga qo'yilgan. Ushbu tizim optimalligi uchun jamiyatning boshqa sohalariga ham asta-sekinlik bilan tatbiq etilmoqda. Hozirda til o'rgatishning avtoma­tik tizimlari ishlab chiqilgan. Til o'rgatishning avtomatlashganligi shu bi­lan izohlanadiki, u maxsus o‘quv dasturlari yordamida amalga oshiriladi.

Til o‘rgatishda kompyuterdan foydalanish bo‘yicha dastlabki tajribalar XX asrning 60-yillarida Stenford (Russian-Program) va Nyu-York univer- sitetlarida (Das Deutche Program) o'tkazildi. XX asrning 70 - 80-yillarida til o'qitishda kompyuterdan foydalanishga bo'lgan qiziqish kuchaydi va hatto alohida yo'nalishning yuzaga kelishiga olib keldi: Computer Assisted Lan­guage Learning (CALL).77 CALL bo‘yicha 1-xalqaro konferensiya 1985-yil 12 - 13-noyabrda Budapesht shahrida o'tkazildi. 2-xalqaro konferensiya 1989-yil 15 - 17-noyabrda boiib o'tdi. Mazkur konferensiyalarda komp­yuter yordamida xorijiy tillarni o'rgatishning yangi metodlari taklif etildi, o'qituvchi lingvistik avtomatlar xususida qimmatli mulohazalar o'rtaga tashlandi, o'qitishda audio va video vositalardan foydalanish samarali ekan- ligi ta’kidlandi.

Kompyuter texnologiyalarining shiddat bilan rivojlanishi ta’lim jarayo- nini yangi bosqichga ko'tardi, bu o'z o'rnida ta’lim mazmunini, metod va shakllarini qayta ko'rib chiqish, uni yangi bilim hamda ko'nikmalar bilan yanada boyitish zaruriyatini tug'dirdi. Hozirgi vaqtda oliy ta’lim muassa- salarida yangi pedagogik texnologiyalarning ilmiy asoslarini yaratish, ularni tasniflash, metodik ahamiyatini belgilash, ta’lim jarayoniga tatbiq etish kabi dolzarb muammolar ustida ish olib borilmoqda. Ug'or pedagogik texnologi­yalarning kompyuter texnologiyalari bilan uyg'unlashtirib tashkil etilishi va shu asosda multimedia darslarining yaratilishi asosiy yo'nalishga ay- lanib qolmoqda. Multimedia («ko'p muhitlik» degan ma’noni bildiradi) za- monaviy kompyuterlarning deyarli barchasida mavjud bo'lgan imkoniyat- lar majmuini, kompleks tushunchani anglatadi. Multimedia axborotning turli ko'rinishlari-matn, jadval, grafika, ovoz, animatsiya (multiplikatsiya), videotasvir, musiqa yordamida axborotni yig'ish, saqlash va qayta uzatish

77 http: // [www.yazykoznanie.ru](http://www.yazykoznanie.ru)

94

A. Rahimov

vazifalarini bajaradi. Multimedia «inson-kompyuter» interaktiv (dialogik) muloqotning yangi, takomillashgan pog'onasi bo'lib, bunda foydalanuvchi juda keng va har tomonlama axborot oladi. Bir qator olimlar, metodistlar tadqiqotlari shuni ko‘rsatadiki, ta’lim oluvchi birinchi marta eshitgan mav- zusining faqat to'rtdan bir qismini, ko‘rgan materialining uchdan bir qismini- gina eslab qoladi; ham ko'rib, ham eshitsa, axborotning ellik foizini esda saqlaydi. Interaktiv multimedia texnologiyalaridan foydalanilganda esa bu ko'rsatkich yetmish besh foizni tashkil etadi. Shu bois ta’lim jarayonida kompyuter texnologiyalaridan unumli foydalanish katta ahamiyatga ega.78

0‘qituvchi lingvistik avtomatlardan foydalanishda 2 xil yondashuv mavjud:

1) bixevioristik yondashuv - mavjud an’anaviy o‘qitish tizimining bar­cha jabhalarini e’tiborga oluvchi o‘qitish mexanizmlari, bunda grammati- kaning ma’lum bo'limi tugagandan so'ng uni topshiriqlar asosida baholash, shuningdek, leksik minimumlar bo‘yicha bilimni nazorat qilishga yo'naltirilgan dasturiy ta’minotga asoslangan o‘qitish tizimi. Bunda bixe- viorizmning «stimul-reaksiya-mustahkamlash» formulasiga asoslaniladi. O'qitish usullari «stimul» (turtki) vazifasini bajaradi, o‘rganuvchi bundan ta’sirlanadi va bilimni mustahkamlaydi.

2) kognitiv-intellektual yondashuv - CALL (CALL Software) uchun uni­versal dasturiy ta’minotlar yaratish, o‘qitish tizimini lug‘at bazasi, gram­matika bo‘yicha qo‘llanmalar, tizimni avtomatik korrektorlar, speller bilan ta’minlash, audio va vizual effektlar bilan boyitish kabi amaliy ishlar maj- muidir.

Til o'rgatishda quyidagi kompyuter dasturlarini qoilash mumkin:79

1. Taqdimot texnologiyasi, ya’ni mashg‘ulotni ko‘rgazma sifatida o'tkazish darsda kompyuterdan foydalanishning eng oson usulidir. Buning uchun o'qituvchiga birgina kompyuter va multimedia proyektori kerak bo'ladi. MS Power Point dasturi yordamida o‘tilajak dars uchun komp- yuterda o'qituvchining o‘zi slaydlar yaratadi va ko‘rgazmalar paketini tay- yorlaydi. Elektron nusxa - bu o‘zbek tili fani bo‘yicha mavjud bo‘lgan muayyan darslikning kompyuterdagi elektron versiyasi (nusxasi) boMib, unda multimedia vositalaridan faqat ovoz va matn qatnashadi.

2. Elektron multimedia darsligi - aynan kompyuter vositasida o‘quv kur- sini yoki uning bitta katta bo‘limini o‘qituvchi yordamida yoki mustaqil o‘zlashtirishni ta’minlovchi dasturiy-metodik kompleks.

78 PoMatov A., Muhamedova S. Kompyuter lingvistikasi. - Toshkent, 2007. -

B.64-65.

79 Ko'rsatilgan asar. - B.66 - 67.

Kompyuter lingvistikasi asoslari

95

3. Elektron ensiklopediya - kuchli qidiruv tizimiga ega bo‘lgan ma’lum yo‘nalishdagi illustratsiyali maqolalar, elektron xaritalar, fotosuratlar, giperhavola (gipermurojaat)lar, audio va videoizohlar, xronologik ro'yxat, foydalanilgan va tavsiya etiladigan adabiyotlar ro'yxati va hokazolardan iborat bo'lgan katta hajmdagi elektron dastur.

4. Elektron lug ‘at - odatdagi lug'atlar asosida tuzilgan, multimedia vosi- talari bilan boyitilgan dastur. Elektron lug'at turli yo‘nalishda tuzilishi mumkin, undan talabaning so‘z boyligini oshirish, tarjima qilishda foydala- niladi.

5. O'rgatuvchi audiodasturlar, elektron lingafon kurslari umuman ikkin- chi tilni o‘rganuvchilar uchun nihoyatda samarali ta’lim texnologiyalari- dir. To‘g‘ri talaffuzni o'rgatuvchi kompyuter dasturlarida animatsiya yor­damida tovushning artikulyatsiya o'rni, usuli ko'rsatiladi, etalon nutq eshit- tiriladi, talabaning o'z talaffuzi yozib kiritiladi, so'ngra u adabiy til me’yori bilan taqqoslanadi. Bundan tashqari, audiodasturlar qatoriga «Karaoke» tizimini kiritish mumkin. Undan hordiq chiqarishda yoki biron tadbirni o'tkazishda foydalanish juda qulaydir.

6. Universal test dasturi muayyan fanning biror yirik bo'limi bo'yicha test topshiriqlari va reyting tizimidan iborat bo'lgan kompyuter dasturidir. Uning qulayligi shundaki, birgina dasturiy ta’minot asosida turli xil test sinovlarini o'tkazish, ya’ni faqat matnni o'zgartirib, dasturni har xil mavzu- larda qo'llash mumkin.

7. Elektron virtual kutubxona - multimedia xonasining keng imkoniyat- laridan yana biri, elektron nusxa yoki elektron multimedia darsliklari mu- jassamlangan, tarmoqqa ulangan kutubxona.

8. Masofaviy ta’lim - bu masofadan Internet tarmog'i orqali o'qitish usuli bo'lib, uning asosida respublikamizda ma’lum sohalar bo'yicha test- tajribalar o'tkazilmoqda.

Tillarni avtomatik o'qitish tizimi hisoblangan kompyuter lingvodidak- tikasida quyidagi turdagi topshiriqlardan foydalaniladi:

a) Close the gap - Bo'sh kataklarni to'ldiring topshirig'i. Bunda seman- tika + grammatika sohalarini o'rganish ko'zda tutiladi;

b) krossvordlar va leksik o'yinlar;

c) elektron pochtadan foydalanib bajariladigan mashq (turli xatlar yoki elektron hikoyalar tuzish asosida - maze) word-hunting;

d) matnlarning avtomatik tarjimasi va tahriri bo'yicha mashqlar;

e) grammatikaning turli bo'limlari bo'yicha diagnostik testlar o'tkazish;

f) CD-ROMda standart o'qitish kurslaridan foydalanish.

Ingliz tilini o'rgatishga mo'ljallangan quyidagi oquv dasturlari mavjud:

96

A. Rahimov

TOEFL tizimi - ingliz tilini avtomatik o'qitishga mo‘ljallangan eng rivoj- langan xalqaro dastur. U dunyoning deyarli barcha mamlakatlarida amal qiladi, u inglizcha Test Of English as a Foreign Language so‘zlarining qisqart- masi bo'lib, ingliz tilidan xorijiy til sifatida sinov standarti hisoblanadi. Bu tizim ingliz tilini o‘rgatish bilan birgalikda o'zlashtirilgan bilimlarni diag- nostika qila olish imkoniyatiga ham ega.

IELTS - TOEFL tizimiga o‘xshaydigan avtomatik til o‘rgatish va bi­limlarni diagnostika qilish tizimi. U grammatika, lug'at, o‘qib va eshitib tushunish, to‘gri jumla qurish kabilar yuzasidan bilimlarni baholaydi.

ENGLISH FOR ALL (Hamma uchun ingliz tili) - 0‘zbekistonda yara- tilgan avtomatik o‘qitishga asoslangan o‘quv dasturi. Unda o‘quvchiga ing­liz tilidagi tovushlar va ularning talaffuzi audioeffektlar yordamida tushun- tiriladi, qisqacha ingliz tili grammatikasi bayon etiladi, mashqlar va testlar yordamida bilimni tekshiradi. Ushbu o‘quv dasturining afzal tomoni shun- daki, unda o‘qitish tizimi videoeffektlar bilan ham boyitilgan, ingliz tilida multfilmlar kiritilgan va inglizcha qo'shiqlar karaoke bilan ta’minlangan.

WAY TO PERFECTNESS (Mukammallik uchun yo‘l) - ingliz tilini avtomatik tarzda o‘qitishga mo'ljallangan o‘quv dasturi bo‘lib, bunda ham multimedia bilan ta’minlangan. Dasturning yuqoridagilardan farqli tomo­ni shundaki, unda dastur «elektron» («virtual») o‘qituvchi bilan ta’minlangan.

REWARD-bir nechayil davomida o'qitishga mo‘ljallangan til o‘rgatish tizimi.

PROFESSOR HIGGINS - АНГЛИЙСКИЙ KFT AKIIF.HTA - fone- tik trenirovka, ingliz tilida aksentsiz talaffuz me’yorlarini o'rgatuvchi ti­zim.

REPETITOR ENGLISH - Moskvadagi ingliz tilini o'rgatuvchi xalqa­ro markaz tashabbusi bilan yaratilgan. Mazkur tizim ingliz tilini turli xil zamonaviy metodlar va elektron metodik ishlanmalar paketi yordamida avtomatik tarzda o'rgatishga mo'ljallangan.

EURO PLUS - uch yil davomida ingliz tilini bosqichma-bosqich (Ele­mentary, Intermediate, Upper intermediate, Advanced English) tizimli tarz­da o‘rgatishga asoslangan avtomatik o‘qitish tizimi.

Bugungi kunda o‘qitishni kompyuterlashtirish nazariyasini ishlab chi­qish, haqiqatan ham, o'quv faoliyatining samaradorligini oshiradigan, o‘rganilayotgan mavzuga nisbatan ijobiy munosabat va qiziqishni shakl- lantiradigan ta’limiy-elektron ishlanmalarni yaratishning umumiy va xususiy mezonlarini belgilash zarurati paydo boidi. Hozirgi paytda Rossiya, Fran- siya, Yaponiya, Kanada va boshqa mamlakatlarda ta’limiy-elektron ishlan­malarni yaratish konsepsiyasini ishlab chiqish bo'yicha ilmiy-metodik tad-

Kompyuter lingvistikasi asoslari

97

qiqotlar olib borilmoqda. Konsepsiyada ishlanmalarning mohiyati, ularni yaratish, tekshirish va qo'llash tartibi belgilanadi. Mutaxassislarning fikricha, ta’limiy-elektron ishlanmalar asosan texnik, pedagogik, metodik talablargajavob berishi zarur. Ular shunday talablarga mos kelsagina, ta’lim jarayonida qoMlanilishi mumkin.80

Til o'qitishning avtomatik tizimi yildan yilga rivojlanib, yangi effektlar bilan boyib bormoqda. Bu tizimni yanada rivojlantirish, kompyuter lingvo- didaktikasini yangi dastruriy ta’minotlar va pedagogik texnologiya, bilim­lar diagnostikasi bilan boyitish kompyuter lingvistikasi fanining dolzarb vazifalaridan sanaladi.

Topshiriqlar:

1. Qanday til o‘rgatuvchi o‘quv dasturlarini bilasiz?

2. Til o‘rgatuvchi o'quv dasturlarining samaradorligi qanday bo'ladi?

3. Boshqa turdagi o'quv dasturlari haqida yozing.

4. Internet tizimi yordamida (on-line) til o‘qitish kurslaridan foydalan- ganmisiz?

5. CALL tizimining rivojlanishi va istiqbollari qanday deb o'ylaysiz?

Adabiyotlar:

1. Нелюбин JI.JI. Компьютерная лингвистика и машинный перевод.

- М.: ВЦП, 1991.

2. Шемакин Ю.И. Начало компьютерной лингвистики.-М.: МГОУ, 1992.

3. PoMatov A., Muhamedova S. Kompyuter lingvistikasi. - Toshkent, 2007.

4. Пиотровский Р.Г. Текст, машина, человек. - М.: Просвещение.

1989.

5. Белоногов Г.Г. Компьютерная лингвистика и перспективные информационные технологии. - М.: Русский мир, 2004.

“ PoMatov A., Muhamedova S. Kompyuter lingvistikasi. - Toshkent, 2007. - B.68.

98

A. Rahimov

KOMPYUTER LINGVISTIKASIDA AVTOMATIK TAHRIR, INFORMATSION QIDIRUV TIZIMLARI VA GIPERTEKST TEXN OLOGIY ASI

Reja:

1. Avtomatik tahrir yo‘nalishi.

2. Informatsion qidiruv tizimlari.

3. Gipertekst texnologiyasi.

Tayanch so‘z va iboralar: matn tahriri, mexanik xatolar, matn muhar- rirlari, avtomatik tahrir, tezaurus, deskriptor, gipertekst, kontekstli qidiruv, avtomatik referatlash, transliteratsiyalash, informatsion qidiruv tizimlari, manzilli qidiruv, semantik qidiruv, faktografik qidiruv, hujjatli qidiruv, indeks- lash, gipermedia.

Avtomatik tahrir kompyuter lingvistikasining asosiy yo‘nalishlaridan biri bo'lib, u kompyuterdagi matn muharrirlari (Microsoft Word, Excel, Wordpad, Lexicon) rivojlanishi bilan bogiiqdir. Bilamizki, avtomatik tahrir deganda matndagi mexanik xatolarning avtomatik tarzda to'g'rilanishi va xato ekanligi haqidagi signallarning userga taklif etilishi kabilar bilan bog'liq. Aniq misollar bilan tushuntiriladigan bo‘lsa, Microsoft Word das- turi bilan ishlar ekanmiz, unda so'zlarning yozilishi bilan bog'liq xatolar bo'lsa, ularning ostiga qizil yoki ko‘k rangda chizilishi, katta harf bilan yozilishi kerak bo'lgan so'zlarning avtomatik bosh harfga o'tkazilishi, huj- jatga avtomatik numeratsiya qo'yilishi kabilar avtomatik tahrir erishgan yutuqlar hisoblanadi.

Avtomatik tahrir yo'nalishi XX asr o'rtalariga kelib rivojlandi. U matn muharrir dasturlari bilan birgalikda yangicha imkoniyatlar bilan, qo'shimcha effektlar bilan rivojlanmoqda. Oddiy muharrirlardan farqi shun- daki, unda tahrir avtomatik tarzda qisqa vaqt birligi ichida hajman katta bo'lgan matnni tez tekshirish va xatolarni to'g'rilash imkoniyati bo'ladi. Avtomatik tahrirning pragmatik ahamiyati shundaki, u kelajakda taraqqiy etsa, yetarli ma’lumotlar bazasi yaratilsa va maxsus dasturlar ishlab chiqil- sa, muharrir kasbi uchun ehtiyoj qolmaydi. Bu esa matn bilan ishlashni osonlashtiradi.

Ma’lumki, kompyuter yaratilmasdan ilgari matnlar yozuv mashinkala- rida terilgan, bu jarayonda orfografik, punktuatsion, stilistik va grammatik

Kompyuter lingvistikasi asoslari

99

xatolarga yo‘l qo‘yilsa, matnni boshdan oxirigacha qayta yozib chiqishga to‘g‘ri kelgan, bu esa foydalanuvchiga noqulaylik tug'dirgan. Kompyuterda esa matnni tahrirlash birmuncha qulay boiib, unda foydalanuvchi dialogli rejimda ishlaydi va o‘z o‘rnida xatolarni to‘g‘rilash imkoniyatiga ega bo'ladi. Hatto matn bilan ishlash jarayonida biz yuqorida ta’kidlab o‘tgan avtomatik tahrirlash, xatolarni to'g'rilash imkoniyati ham mavjud. Shu o'rinda alohi- da ta’kidlash zarurki, avtomatik tahrirlash tizimi faqat matn muharrirlari dasturlari bazasiga qo‘shilgan tillarda amal qiladi, masalan, ingliz, rus, ne- mis, fransuz tillarida. Agar bazada mavjud bo'lmagan tilda matn kiritilsa, uni avtomatik tahrirlay olmay4i, ayrim internatsional so‘zlar bundan mus- tasno. Avtomatik tahrirlash jarayonida MS Word lug'atida mavjud boimagan so‘z uchrasa, ostiga qizil chiziq bilan ajratib ko'rsatiladi. Mazkur holat quyidagi sababga ko‘ra boiishi mumkin, bunda o‘sha so‘z imlosi haqi- da dastur variant taklif eta olmaydi, chunki Ms Word lug'atida bu so‘z mav­jud emas. Ehtimol u so‘z tog‘ri yozilgan bo‘lishi mumkin, bunda insonning ishtiroki zarur bo'ladi. Agar so'zning imlosi to‘g‘ri bo‘lsa, matnda yana tez- tez qayta ishlatilsa, uning avtokorreksiyada to‘g‘ri deb sanalishi uchun Ms Word lug‘atiga «kiritib qo'yish» (kontekst menyudagi «add» yoki «добавить» komandasi yordamida) mumkin. Matn kiritish davomida so'z yoki so'zlar birikmasi ko‘k chiziq bilan ajratib qo‘yiladi, bunda so‘zning imlosida xato mavjud. Bu holda foydalanuvchi xatoni o‘zi tog‘rilab qo'yishi zarur yoki ostiga chizilgan so'zda sichqonchaning o‘ng tomoni bosilsa, kontekst men- yusida so'zning to‘g‘ri varianti taklif etiladi va uni tanlashi zarur bo'ladi.

Avtomatik tahrirlash tizimi juda ko‘p amaliy vazifalarni o'z ichiga ola- di. Bu vazifalarning eng asosiylarini umumlashtirib sanab o'tmoqchimiz:

- orfografiyani tekshirish;

- grammatika va stilistikani tekshirish;

- kontekstli qidiruv;

- formatlash;

- avtomatik referatlash - matndagi kalit so'zlar asosida uning umumiy mundarijasini yaratish;

- matn ichiga jadval, rasm, turli figuralar, klip-artlar, grafik sxemalar, diagrammalar qo'shish, giperhavolalar bilan ta’minlash;

- matndagi abzaslar, belgilar, so'zlarning aniq miqdorini aniqlash (statis- tikasini ko'rsatish);

- tezaurus xizmati - matndagi muayyan so'zning sinonimik variantlari- ni aniqlash;

- transliteratsiyalash - bir alifbodan boshqa alifboga o'tkazish (masalan, kirill alifbosidan lotin alifbosidagi harfga almashtirish).

100

A. Rahimov

Avtomatik tahrirda uch tarkibli asosiy vazifa e’tiborga olinishi lozim:

1. Orfografiyani tekshirish - bu vazifa morfologik tahlil yordamida amalga oshiriladi, bunda etalon lug‘atdagi baza asosida tekshiriladi.

2. Sintaksisni tekshirish orfografiyani tekshirishdan ko‘ra murakkab- roq. Chunki orfografiyani tekshiruvchi dasturlar yaratilgan, lekin norasmiy matnlarning sintaktik jihatdan avtomatik tahriri masalasi haligacha hal etilgani yo‘q. Rasmiy matnlarning avtomatik tahriri bo'yicha muhim nati- jalarga erishilgan.

3. Semantik tahrirlash inson tafakkurini modellashtirishga asoslangan sun’iy intellekt tizimi bilan bog'liq ravishda o‘rganilmoqda. Bunda mazkur ko‘p aspektli muammoni semantik maydon, freym yondashuv, tezauruslar tizimini boyitish, so‘z va so‘z birikmalari o‘rtasidagi paradigmatik munosa- batlar (sinonimiya, gipo-giperonimik-tur-jins munosabati)ni modellashtirish­ga tatbiq etish asosida hal etish mumkin. Chunki inson bunday assotsiativ munosabatlarni o‘z intellektiga tayanib o'rnata oladi, biroq bu jarayon kompyuter uchun murakkabdir.

Ms Word dasturida orfografiyani tekshirish imkoniyati cheklanganligi bois hozirgi kunda maxsus Spell Checker dasturlar yaratilmoqda. Bunday dasturlar sirasiga quyidagilarni kiritish mumkin: SpellLink Multilingual ErQL\_(ingliz, ispan, italyan, nemis, fransuz, portugal, ukrain tillaridagi so'zlarning to‘g‘ri yozilishini tekshiradi), Qrfo SpellLink English, SpellLink French. SpellLink Italian. SpellLink Spanish, Ruta 1.0 (rus va ukrain tilla­ridagi so'zlarning imlosini tekshirishga yordam beradi), Asutvpe. Spellex, Spell Catcher kabi. Mazkur dasturlar Word dasturi ustiga qo'shimcha o'rnatiladi. Masalan, aniqroq tasavvur hosil qilish uchun bu dasturlarning orfografiya tekshirish jarayonini ko'rsatib o'tmoqchimiz.

I've just been Wormed Thar the sum of money, which was to be jaldto me prior to the date agreed beforehand, has been ШПВВ 1 bope you are not going to make me pay in your debts. I will have to solve this problem, using all my connections. 1 wouldn't like to choose tne way some people hurt their competitors' business, but it wholly depends on your good will, every week imports speed up the price decrease, so please send the money to our bank account as soon as possible.

NdcDMorwy gjjggj

O'™\*1" ; Iм\*™1

Kompyuter lingvistikasi asoslari

101

3 demo,\_se.rtf • MicrosoftWord - П X

He £<« »ew Insert F»m«\* ITable !tfndow «alp x

a \* ■ ■■■ у es'-'"»-.iinjn^. ;Ji.

|  |  |
| --- | --- |
| ®<ei | Spelling and Grammar: Spanish Spam I radi'Mcnal Sort \*? X |
| ГС] • • ■ 3S eti< эсот ecim os ar an re osas iterro1.,bie■}«» ш : P-jge 1 | Not In ИсИолагу i |
| maquinas que reconozcan las etiquetas cie 4' | ignore once J |
| los textos у apliquen el procedimiento mas , |
| traducclon manual cuando se topen con де addtoDfcuararv 1 |
| Suggest\*\*?: |
| :г-мКсюп 1:reducisn «а (Chen9eA\*. 1 |
| ■rj AutoCotrect { |
| Ctrttofwy language: jspanM Spetn-Tredit-ood Sort Г Check pammerOptions... j 'Ыо | Cancel | |
|  |

! [to 6» 1», №.. \*»« де e«ow rtfl. до» ВФ ДИЦК \_ .!

»■■'■•■■■ • О ’’ £5-' \*!: »

JEi\*)

;!(\_xf

ж1

few 1 ж

|  |  |  |
| --- | --- | --- |
| Nob to Dkttof^yi |  | няппнпш |
| Nous avin^ autons4 les membres d\*» conseil \_\*! | ftjnore | • d'admmistratlonl-rweAi jif""'' 1 |
|  |  |  |
| i ЧЛ1К »v ons wins |  | I f . СЫГФ\* 1 jJ \*ioCooert 1 |
| CKfcorwry language: ®l | |Fi«n<t> (Fiance) jj\_ i „ , \_ , |

J Uo'.umcntt Microsoft Word \_ □ X

|  |
| --- |
| I «в &» flew Insert Format look Tflbte Jtfndow уф X[? {i Йг ft j <§& v\*1, :: X £& # ' i Ш f J : IT |»0Й% \*j /' ? 'it |
|  | Spelling And Grammar: Italian (My ? X ; |
| ramom partic si svcIgrufcripteperccristreli] а в лРеве 1 | Not in Dictionary; |
| i visltator vengono accomagnati. Щ 1 ignore once j |
| gratoitamente, da guide della ioro lingua, i ma nei momenti di grande affluss^ con i irore | |
| gruppi numerosl deila stessa lingua, v да ю Dictionary j |
| SuggesttoQs: |
| vtjiurtori ^ —2225.—Jvis4et0 Change Af J |
| »j AutoCorrect j |
| Ocbonafylanguege: |nakanltaly 4Options... ( | Cancel { |
|  |

102

A.Rahimov

^3]Document! Microsoft Word

Me Ed\* Wew insert Pjjrm\* Took T«bte Window t\*\*>

D \ 'i. ' ■-j -: ffl Q \* 1!' {ioc>%

NutT

йгег

mei<

Este

dos

publ

NoHnDKttonaryj

tecnolPgica 6 permanente e atinge tod as as areas da sociedade, 6 necesArio que os museus utillzem efidentemente os meios que as

Suggestions:

v ftdd to Dictionary

■3

Change Af

\* | AutoCoprect

Gtcbonary language: (Portuguese Brazil

Г Owckgfsr/wiej

|  |  |  |
| --- | --- | --- |
| Spelling and Grammar Russian (Russia) |  |  |
| Grgnmatical error: |
| ОРФО Агент позволяет вам по ГОРЯЧЕЙ 1 КЛАВИШЕ проверять правописание в \ приложениях, не имеющих собственных средств проверки (NOTEPAD, UORDPAD и т.п.)/ включая почтовые программы. Н | | Ignore | Ignore Rule | Next Sentence j |
| Suggestion: | |
| I^Слишком иного идущих подряд согласных на стыке слов. | M | ' Change ■: , 1 |
| I Попробуйте построить фразу иначе. | г | ... ' U ■: |
| Ш |
|  |  |
| Dictionary language: (Russian (Russia) 17 Check grammar |
| ©j Options... j Updo | | dose | |
|  J |

£3 Document - WordPad j^jfDjfXi

Me £dtt yiew Insert Fermat Це!р

Q H §a F о %

:Mal v ID v. 'Western v В /

g—-—7-7-.-"--v;. 7 ;...4

With Asutype you dont need to do anything extra in order ; to have your typing spell checked.

Just type exactly as you did before and if you make a misstake. Asutype will red underline it along with an alert beep.

If you choose the automatic correction mode, Asutype will correct the mistake immediately and automatically.

I If you choose the interactive correction m ode, A sutype will show you some choices and let you choose a | correction for the mistake.

For Heip, press F i

Kompyuter lingvistikasi asoslari

103

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| I SSpeHex |  |  |  | i'P X. |
| 19\* fiptfow 6 | fcSonarie | »'s Check |  | ,, . J |

Cloning is the creation of multiple copies of a single molecule, cell, or vims. There are many different kinds of cloning, most of which are now commonplace in science. Cloning has allowed^ scientists to develop powerful new drugs and to produce BflBBIS

 :k

if (a I to e

|  |  |  |  |
| --- | --- | --- | --- |
| andithee | Check Spelling |  |  |
| innO( | Not m Оюйопаде |  | Ifltwie | |
| undeSCNrepleskin. | Suggestions: | ИИИГл" | tflnoieAl j Add | |
| jindulrte |  | | |
| begiextra | |auinAddiWtdJ to: | Ш | а»п9»д| | |
|  | |uig’Kxe 1Ы | d | Suwesl | |
|  | Undo j Option»... I fiictionaf«$...] | и\*.. 1 | Cancel | |

Kompyuter lingvistikasining muhim yo‘nalishlaridan yana biri infor- matsion qidiruv tizimlari (IQT) hisoblanadi. Ular formal belgilarga ko‘ra informatsiyani qidirishga mo‘ljallangan. IQTlar quyidagi bosqichlar aso­sida ishlaydi:

- informatsiyani jamlash;

- informatsiyani klassifikatsiya qilish;

- hujjatning qidiruv obrazi (HQO)ni yaratish;

- hujjatning qidiruv obrazi va hujjatlarni saqlash;

- qidirish va natijani chiqarish.

«Informatsion qidiruv» (inglizcha «information retrieval») termini 1948- yilda Kelvin Mur tomonidan fanga kiritilgan.81 Informatsion qidiruv tizimi

- bu ma’lumotlarni qisqa vaqt birligi ichida qidirib topish, kerakli hujjatlar­ni ma’lumotlar bazasidan topishni optimallashtirish tizimidir. Bu yo‘nalish ilmiy-texnikaviy axborotlarning ko'payib ketishiga javoban 1960-yil oxir- larida yuzaga keldi. Axborot ko‘paygan sari uni saralash, keraklisini qidi­rib topish muammosi yuzaga keldi. Mazkur muammoni bartaraf etish uchun olimlar informatsion qidiruv tizimlari g'oyasini taklif etdilar. Bu yo'nalish Internet tarmog'i yuzaga kelgandan so‘ng yanada rivojlandi. Masalan, bu- gun hammaga ma’lumki, Internetda Google, Yandex, Rambler, Yahoo kabi qidiruv tizimlari mavjud.

Informatsiyani qidirish quyidagi 4 bosqichni o‘z ichiga oladi:

- informatsiyaga bo'lgan talabni aniqlashtirish hamda informatsion so‘rovni shakllantirish;

- so'ralgan informatsiya saqlangan manbalarni (informatsion massiv- larni) aniqlash;

81 Manning C., Raghavan P., Schutze H. // Cambridge University Press. 2008. - P. 5.

104

А. Rahimov

- informatsion massivlardan zaruriy informatsiyani ajratib olish;

- informatsiya bilan tanishish va qidiruv natijalarini baholash.

Qidiruvning quyidagi turlari mavjud:

1. Toiiq matnli qidiruv - hujjatning to‘liq mazmun-mundarijasi asosida amalga oshiriladigan qidiruv. Masalan, Internetdagi qidiruv tizimlari:

2. Metama’lumotlar bo'yicha qidiruv - hujjatning muayyan atributlari (hujjat nomi, yaratilgan vfiqti, hajmi, muallifi) asosida amalga oshiriladi­gan qidiruv. Qidiruvning bu turiga MS Windows tizimida fayllar bilan ishlashda foydalaniladigan qidiruv dialogi misol bo'la oladi.

3. Tasvir qidiruvi - tasvirning mazmuni asosida amalga oshiriladigan qidiruv. Bunda qidiruv tizimi rasmning mazmunini taniydi va natijada shun- ga o'xshash rasm topiladi. Polar Rose, Picollator qidiruv tizimlari mana shu tamoyil asosida ishlaydi.

4. Multimediali informatsiya qidiruvi - bunda audio va videofayllar bo'yicha qidiruv tushuniladi.

Qidirish jarayoni turli metodlar asosida yuz beradi: manzilli qidiruv, semantik qidiruv, hujjatli qidiruv, faktografik qidiruv. Manzilli qidiruv so'rovda ko'rsatiladigan sof formal belgilar bo'yicha hujjatlarni qidirish jarayoni boiib, bunda hujjatning aniq manzili ko'rsatilishi lozim. Hujjat­ning manzili web-server va web-sahifa manzili yoki bibliografik qayd ele- mentlari va hujjat saqlangan joy sanaladi. Semantik qidiruv hujjatlarlar- ning mazmuni asosida qidirish usuli boiib, bunda hujjatning umumlashma mazmuni, sarlavhasi deskriptor ko'rinishida shakllantirilishi lozim. Hujjat­li qidiruv matnli ma’lumotlar hamda ular haqidagi bibliografik qaydlarni qidirish jarayonidir. Faktografik qidiruv so'rovda nazarda tutilgan maiumotga mos keluvchi faktlarni (matnli hujjatlar tarkibidan ajratilgan maiumotlar) qidirish jarayonidir. Yuqoridagilardan kelib chiqib aytish mumkinki, qidiruv tizimi relatsion maiumotlar bazasida hamda gipertekst- li maiumotlar bazasida mavjud bo'lgan hujjatlar, hujjatlar haqidagi metama’lumotlar, matnlar, rasmlar, video va audiofayllarni qidirishni o'z ichiga oladi.

Informatsiyaning eksponensial tarzda ko'payib ketayotganligi, shu- ningdek, Internet tizimida ham kundan kunga maiumotlar hajmining o'sib borayotganligi kompyuter lingvistikasi oldiga bir qator vazifalarni ko'ndalang qo'ymoqda:

- qidiruv tizimining optimal modellarini yaratish;

- hujjatlarni tasniflash (klassifikatorlar tizimini yaratish);

- hujjatlarni klasterlash;

Kompyuter lingvistikasi asoslari

105

- foydalanuvchi interfeyslari va qidiruv tizimlari arxitekturasini loyi- halash;

- zaruriy informatsiyani ajratib olish, hujjatlarni avtomatik referatlash va annotatsiyalsh;

- deskriptorli tillar va tezauruslarni mukammallashtirish, boyitish.

Saqlangan ma’lumotning hamda qidirish xususiyatiga ko‘ra informat­sion qidiruv tizimlari 2 katta guruhga boiinadi: hujjatli va faktografik IQT. Hujjatli IQTlarda matn, tavsif va ta’riflar saqlanadi, faktografik IQTlar esa jadvallar, formulalar, grafik ko'rinishidagi ma’lumotlarni o‘z ichiga oladi. Shuningdek, aralash tabiatli IQTlar ham mavjud bo'lib, unda ham hujjatli, ham faktografik ma’lumotlar saqlanadi. IQTda qidiruvni ta’minlash uchun maxsus informatsion qidiruv tillari mavjud. Informatsion qidiruv tillari bu formal til bo'lib, u IQTda saqlangan hujjatlarning mazmunini tavsiflash va so'rov uchun mo'ljallangan. Informatsion qidiruv tilida hujjatlarga ishlov berish protsedurasi indekslash deb ataladi. Indekslash natijasida har bir huj- jatga IQTdagi ta’rif shakli yoziladi. Mazkur ko'p bosqichli jarayonda deskriptorli informatsion qidiruv tillari ishlatiladi. Hujjat mavzusi deskrip- torlarning birlashuvi bilan yoziladi. Deskriptorlar sifatida so'zlar, termin- lar, muayyan sohadagi boshlang'ich tushuncha va kategoriyalardan foy­dalaniladi. Deskriptorlarni kalit so'zlar deb nomlash ham mumkin. Deskrip­torli til bilan ishlaydigan tizimlardan biri M.Taube tomonidan yaratilgan UNITERM tizimidir. Bu tizimda deskriptorlar sifatida hujjatning kalit so'zlari - unitermlar ishlatiladi. Demak, muayyan sohaning terminologiya- si, ya’ni deskriptorlar qidiruv tizimi va tezauruslarning asosini tashkil etadi.

Matnning asosiy mazmun-mundarijasini aks ettiradigan kalit so'zlar, terminlar, tayanch tushunchalar bazasi asosida kompyuterda lug'at tuzila- di. Bunday lug'at informatsion qidiruv tezauruslari deb ataladi. Tezauruslar ensiklopedik va izohli lug'atlardan farqli ravishda til birliklarining matnda qo'llanish chastotasi va matnning predmet mundarijasini aks ettirish dara- jasiga qarab tuziladi. Bunda terminning mazmuniy tavsifi orqali bazadagi ma’lumotlar qidirilishiga mo'ljallangan bo'ladi. Terminlar tezaurusga qat’iy semantik prinsiplar asosida kiritiladi, bunda ularning gipo-geperonimik (tur- jins), holo-meronik (butun-bo'lak), iyerarxik (pog'onali) munosabatlari hamda assotsiativ mazmuniy bog'lanishlari e’tiborga olinishi lozim. Masalan, Wikipedia sahifasidan «tezaurus» terminining tavsifini olaylik:

Тезаурус (от греч QtjffavpoQ - сокровище) в современной лингвистике

- особая разновидность словарей общей или специальной лексики, в ко-

ронимы. гипонимы, гиперонимы и т.п.) между лексическими единицами.

106

A.Rahimov

Таким образом, тезаурусы, особенно в электронном формате, являют­ся одним из действенных инструментов для описания отдельных пред­метных областей.

В отличие от толкового словаря, тезаурус позволяет выявить смысл не только с помощью определения, но и посредством соотнесения слова с другими понятиями и их группами, благодаря чему может использо­ваться в системах искусственного интеллекта.

В прошлом термином тезаурус обозначались по преимуществу сло­вари, с максимальной полнотой представлявшие лексику языка с приме­рами её употребления в текстах.

Также термин тезаурус употребляется в теории информации для обозначения совокупности всех сведений, которыми обладает субъект.

вис/пики | Уровни языка

Тезаурусы

- Амара-коша (Словарь Амары) - первый и наиболее важный санск­ритский тезаурус (III - IV век)

- Первым тезаурусом в современном смысле слова стал английский Тезаурус Роже (XIX век, около 15 ООО синсетов).

- Викисловарь: многоязычный свободно пополняемый словарь и теза­урус (2004-наст.вр., проект фонда Викимедиа, в Русском Викисловаре (дамп - январь 2009) содержится 250 тыс. понятий, 67 тыс. отноше­ний111)

- ЕуТ.сз - тезаурус русского языка (был создан Центром информаци­онных исследований как инструмент для автоматического индексиро­вания; разрабатывается с 1997 года по настоящее время, 45 тыс. поня­тий, 107 тыс. слов и выражений, 177 тыс. отношений)

См. также

- Викисловарь

- Идеографический словарь

- Русский семантический словарь

- Семантический анализ

- Семантическая сеть

Ссылки

- [www.thesaurus.com](http://www.thesaurus.com) - Тезаурус английского языка

- Захаров Н.В., Луков А.В. Школа тезаурусного анализа // Знание. Понимание. Умение. 2006. № 1. - С.231 - 233.

- Лесников С.Б.. Лесникова Л. С. Информационно-поисковый тезау- дас

Kompyuter lingvistikasi asoslari

107

- Луков А.В. Фрактальность тезаурусов (в статье делается заклю­чение о фрактальности тезауруса лишь на основании перевода слова «фрактал» )

- Луков В.А.. Луков В.А. Тезаурусный подход в гуманитарных на­уках

Virtual ensiklopedia sanalgan Wikipediada keltirilgan «tezaurus» so'ziga bag'ishlangan lug‘at maqolasida semantik munosabatlar, assotsiativ bogianishlar hamda iyerarxiya munosabati yaqqol ko'zga tashlanadi. Shu o'rinda ta’kidlash zarurki, keyingi paytlarda qidiruv tizimlari bazasidagi tezauruslar giperhavolalar bilan ham ta’minlanmoqda, bu esa userga bir- muncha qulayliklar yaratadi. Ya’ni qidirilayotgan terminlarga aloqador boigan boshqa yaqin tushunchalar bilan bogiiq ma’lumotlarni ham oson topish imkonini beradi.

Ilm-fan, texnika taraqqiyoti intensiv ravishda rivojlanyotgan bu davrda terminlar dinamikasini e’tiborga olgan holda tezauruslar bazasini muttasil toidirib, boyitib borish zarur. Chunki til ochiq dinamik tizim hamda jamiyat taraqqiyotining «barometri» hisoblanadi, ya’ni ayrim tushunchalar eskira- di, ularning o‘rniga yangi tushuncha va terminlar yuzaga keladi. lnformat- sion qidiruv tezauruslar yaratishda mana shuni doimo e’tiborga olish lozim. Xalqaro INFOTERM (Xalqaro terminologik markaz, Avstriya) tashkiloti maiumotiga ko'ra, hozirgi kunda tillardagi terminlar miqdori 50 mln, mah- sulot nomlari miqdori 100 mln ni tashkil etadi. Lekin ushbu statistik maiumot turg'un emas, u jamiyat taraqqiyoti bilan yonma-yon o'zgarib turadi.

Kompyuter lingvistikasida inqilobiy o‘zgarish yasagan yangilik gipertekst texnologiyasi hisoblanadi. Chunki bu tizim Guttenberg asos sol- gan kitob bosish texnologiyasiga zid ravishda yangi turdagi matn struktu- rasini vujudga keltirdi. Gipertekst g'oyasi AQSH prezidenti F.Ruzveltning fan va ta’lim bo'yicha maslahatchisi Vannevar Bush nomi bilan bogiiq bo'lib, u foydalanuvchiga matn va uning fragmentlarini turli assotsiativ munosabatlar asosida bogiashga imkon beruvchi «MEMEKS» tizimini nazariy jihatdan asoslab berdi. U 1945-yilda «The Atlantic Monthly» jur- nalida «Memeks» deb atalgan kelajakdagi xayoliy qurilma haqida «As We May Think» («Biz o‘ylaganimiz kabi») nomli maqolasini e’lon qildi. V.Bush o'sha qurilmani kutubxonadagi har qanday hujjatlar, yozishmalar, kitob- larni displeyda ko'rsatib bera oladigan, cheklanmagan mikrofilmlarga ulan- gan elektromexanik stol sifatida tasvirlaydi. Kompyuter texnikasi yo'qligi bois loyihaning amalda qoilanishi bir oz keyinga surildi.

Bushning g'oyasi gipertekst haqidagi nazariy va amaliy izlanishlar juda katta ta’sir ko'rsatdi. Bushning kelajak loyihasi 1960-yilda T.Nelsonning

108

A.Rahimov

«Ksanadu» tizimi orqali qayta ko‘tarildi. Bu texnik tizim kiritilgan matn- larni va uning qismlarini turli xil usullarda, har xil qatorda, ixtiyoriy ket- ma-ketlikda o'qishga imkon beradi. Bu esa ko'rilgan matnlarning o'rnini, ketma-ketligini eslab qolishga va xohlagan vaqtda o‘zingiz uchun kerakli matnni tanlab olishga va tezkor murojaat etishga yordam beradi. Bunday xususiyatga ega bo'lgan matnlarni T.Nelson gipertekst deb nomladi. 1968- yil dekabrda amerikalik olim Duglas Engelbart ham o‘zi yaratgan gipertekst interfeysini e’lon qildi. Olim uni «The Mother of All Demos» deb atadi. 1983-yilda Ben Shneyderman ham o'z gipertekst loyihasini yaratdi, ushbu loyiha The Interactive Encyclopedia System (TIES) deb nomlan- gan.82 1992-yilda birinchi marta gipertekst World Wide Web (xalqaro to'r)ga tatbiq etildi.

Gipertekst yunoncha «hyper» - «ostida», «orqasida», «ortidan» ma’nolarini anglatuvchi old qo'shimcha hamda lotincha «tekst»- «to'qima» ma’nosini anglatuvchi so'zlardan olingan.83 1965-yilda gipertekst va giper- media terminlari T.Nelson tomonidan fanga kiritildi. Olim bu haqda o'zining «Litarary Machines» kitobida shunday yozadi: «Gipertekst termini giper- media terminiga qaraganda keng onxmalashib ketdi. Aslida har ikki termin bir tushunchani anglatmaydi. Ya’ni gipertekst faqat matnlarning tarmoqla- nishiga nisbatan ishlatiladi, gipermedia termini esa grafiklar, audio va video- fayllar, shuningdek, matnlarning shaxobchalanishini ham o'z ichiga oladi. Bir maddat har ikki termin о ‘rniga «interaktiv multimedia» atamasi ham ishlatildi. Lekin negadir gipertekst termini hammasidan ko'ra keng e'tirof etildi va ommalashdi» ,84

Gipertekst bu shunday matnki, u kompyuterda boshqa matnlarga giper- havolalar yordamida bog'lashga imkon beradi. Bunda u o'quvchiga faqat matnlar ustida emas, balki turli xil jadvallar, sxemalar, rasmlar, videoro- liklarda ham erkin navigatsiya qilish imkonini beradi. Asosan, gipertekst World Wide Web (Xalqaro o'rgimchak to‘ri) strukturasini ko'rsatuvchi tushuncha bo'lib, u orqali Internet tizimida informatsiya tarqalishida qulay- lik, osonlik, optimallik ta’minlanadi. Internet olami virtual olam sanalgani uchun biz undagi birinchi olamdan ikkinchi olamga hech qanday to'siqlarsiz o'ta olamiz. Mana shu virtuallikni bizga ta’minlab beruvchi tizim gipertekst g'oyasidir. Internet saytlari manzilini ko'rsatishda HTTP dan foydalana- miz (Hyper Text Transfer Protocol - Giper matnni o'tkazish yo'lagi).

82 http: // [www.sigweb.org/conferences/ht-cover.shtml](http://www.sigweb.org/conferences/ht-cover.shtml)

83 Субботин M.M. Гипертекст. Новая форма письменной коммуникации.

- ВИНИТИ. Сер. Информатика. 1994. - С.18

84 http: // [www.iw3c2.org](http://www.iw3c2.org)

Kompyuter lingvistikasi asoslari

109

Tekst va gipertekst tushunchalari o‘rtasidagi farqni quyidagicha umumlashtirish mumkin:

13-jadval

|  |  |
| --- | --- |
| TEKST | GIPERTEKST |
| Tugallanganlik | Tugallanmaganlik |
| Chiziqlilik | Nochiziqlilik |
| Aniq muallifning mavjudligi | Aniq muallifning mavjud emasligi |
| Bir tomonlama yo'nalganlik | Ko‘p tomonlama yo'nalganlik |
| Bir jinslilik (bir xilda tuzilganlik) | Bir jinsda tuzilmaganlik (turfa xillilik) |
| Yopiq strukturaga egalik | Ochiq strukturaga egalik |

Gipertekst strukturasiga ko‘ra iyerarxik (pog‘onali) yoki tarmoqli bo'lishi mumkin. Daraxt ko'rinishidgi iyerarxik gipertekstlar foydalanuvchining navigatsiya (uzellar bo'ylab bir qismdan ikkinchi qismga o'tish) imkoniya- tini cheklaydi, gipertekst texnologiyasining barcha imkoniyatlarini yuzaga chiqarmaydi. Bunday gipertekstda komponentlar o'rtasidagi aloqalar jins- tur munosabatiga asoslangan tezaurus strukturasini eslatadi. Tarmoqli gipertekst esa birmuncha keng tarqalgan va imkoniyati kengroq tizim hisob­lanadi, u jins-tur munosabatidan tashqari boshqa assotsiativ munosabatlar- ni ham o‘z ichiga oladi. Shuning uchun tarmoqli gipertekstlar Internet tizimida faol ishlatiladi.

Gipertekst ikki xil holatda bo'lishi mumkin: statik va dinamik gipertekst­lar. Statik gipertekstlar oldindan tayyorlangan va saqlangan bo'lib, ular turg'un holatda bo'ladi. Masalan, gipertekst texnologiyasi asosida CDga yozilgan kitoblar statik gipertekstga misol bo'la oladi. Dinamik gipertekst­lar murakkab va tizimli bog'lanishlarga tayangan bo'lib, u ko'proq tarmoq tizimida ishlaydi. Ayniqsa, Internet tizimida dinamik gipertekstlar faol qo'llaniladi. Masalan, Arizona axborot tizimi (AAIS) dinamik gipertekst sanaladi, uning bazasiga har oyda 300 - 500 referat ko'rinishida ma’lumotlar kiritib turiladi.

Demak, gipertekst texnologiyasi bizga informatsion massivni tezkor ko'rib chiqish (brauzing), muayyan informatsiya yoki matn fragmentlariga tezkor havola qilish, gipertekst bo'yicha erkin navigatsiya qilish, gipertekstni yan­gi informatsiyalar bilan boyitib borish kabi optimal imkoniyatlar yaratib beradi.

Ko'pgina tadqiqotchilar gipertekst tizimini yangi informatsion eraning boshlanishi, virtual olamga dastlabki qadam sifatida baholadilar. Haqiqa- tan ham gipertekst tizimi kompyuter lingvistikasida, Internet tizimida o'ziga xos inqilob yasadi, deyish mumkin.

lto

A. Rahimov

Topshiriqlar:

1. Avtomatik tahrirning pragmatik jihatlarini tushuntiring.

2. Avtomatik tahrirning ishonchlilik darajasi nimalarga bogiiq?

3. Avtomatik tahrirda ma’lum salbiy jihatlarni tushuntiring.

4. Gipertekst texnologiyasi va an’anaviy kitob bosish texnologiyasini solishtiring.

5. Qanday informatsion qidiruv tizimlarini bilasiz?

6. Deskriptor nima?

Adabiyotlar:

1. Нелюбин JT. JI. Компьютерная лингвистика и машинный перевод. -М.: ВЦП, 1991.

2. Шемакин Ю.И. Начало компьютерной лингвистики. - М.: МГОУ, 1992.

3. Po‘latov A., Muhamedova S. Kompyuter lingvistikasi. - Toshkent,

2007.

4. Пиотровский Р.Г. Текст, машина, человек. - М.: Просвещение, 1989.

5. Субботин М.М. Гипертекст. Новая форма письменной коммуникации. - ВИНИТИ. Сер. Информатика, 1994.

6. Белоногов Г.Г. Компьютерная лингвистика и перспективные информационные технологии. - М.: Русский мир, 2004.

KOMPYUTER ANALIZ V A SINTEZI.

INGLIZ TILIDAGI GAPLARNI 0‘ZBEK TILIGA TARJIMA QILISH ALGORITMI

Reja:

1. Morfologik analiz va sintez.

2. Sintaktik analiz va sintez.

3. Tarjima algoritmi.

Tayanch so‘z va iboralar: kompyuter analizi va sintezi, avtomatik tarji­ma, lingvistik ta’minot, dasturiy ta'minot, dastur algoritmi, tarjimon das- turlar, .formal model, morfologik, semantik-sintaktik va konseptual analiz, lingvoanalizatorlar, parser dasturlar.

Kompyuter lingvistikasi asoslari

111

Matnning formal ajraladigan eng kichik birligi so‘z hisoblanadi, lekin u ma’no anglatuvchi eng kichik birlik morfemalardan tashkil topadi. So‘z tarki­bi o‘zak morfemalar, prefikslar (old qo‘shimchalar) va suffikslar (o‘zakdan keyin qo‘shimchalar)dan iborat bo‘ladi. Masalan, besabrlikdandir so'zida 5 morfema mavjud. So'zlarning morfologik analizi uchun tildagi prefiks va suffikslar bazasi hamda so'zlarning grammatik shakllarda o'zgarishi (qan­day qo'shimchalarni qabul qilishi)ni ko'rsatuvchi lingvistik ta’min yarati- lishi lozim. Bunday baza avtomatik tarjima tizimi yaratilgan rus, ingliz, fransuz tillarida yaratilgan. Shulardan birini (rus tili grammatikasi asosida- gi) berib o'tmoqchimiz.85

14-jadval

|  |  |  |  |
| --- | --- | --- | --- |
| №л/п | Слово-представитель | Окончания:1) им.п., ед.ч.2) тв.п., ед.ч.3) им.п., мн.ч.4) род.п., мн.ч. | Примечание |
| Существительные мужского рода неодушевленные |
| 001 | телефон | +\*), ом, ы, ов |  |
| 002 | тираж | +, ом, и, ей |  |
| 003 | огонь | ь, ем, и, ей | См. п. 012 |
| 004 | перебой | й, ем, и, ев | См. п. 005 |
| 005 | санаторий | й, ем, и, ев | Основы на букву «и» |
| 006 | бланк | +, ом, и, ов |  |
| 007 | сапог | +, ом, и, + |  |
| 010 | лес | +, ом, а, ов |  |
| 011 | колодец | +, ем, ы, ев |  |
| 012 | путь (класс состоит из одного слова) |  |  |
| 013 | край | й, ем, я, ев |  |
| 014 | брус | +, ом, я, ев |  |
| 015 | глаз | +, ом, а, + |  |

85 Белоногов Г.Г. Компьютерная лингвистика и перспективные информационные технологии. - М.: Русский мир, 2004. - С.68 - 90.

112

A. Rahimov

|  |  |  |  |
| --- | --- | --- | --- |
| 016 | зародыш | +, ем, и, ей |  |
| 017 | волос | +, ом, ы, + |  |
| 020 | лагерь | ь, ем, я, ей |  |
| Существительные мужского рода одушевленные |
| 021 | кузнец | +, ом, ы, ов |  |
| 022 | солдат | +, ом, ы, + |  |
| 023 | сосед | +, ом, и, ей | См. п. 024 |
| 024 | врач | +, ом, и, ей | Основы на ж, ч, ш, щ |
| 025 | пролетарий | й, ем, и, ев |  |
| 026 | воробей | ей, ем, и, ев |  |
| 027 | конь | ь, ем, и, ей |  |
| 030 | учитель | ь, ем, я, ей |  |
| 031 | сапожник | +, ом, и, ов |  |
| 032 | испанец | +, ем, ы, ев |  |
| 033 | юноша | а, ей, и, ей |  |
| 034 | мужчина | а, ой, ы, + |  |
| 035 | судья | я, ей, и, ей |  |
| 036 | товарищ | +, ем, и, ей |  |
| 037 | гражданин | +, ом, е, + |  |
| 040 | профессор | +, ом, а, ов |  |
| 041 | муж | +, ем, я, ей |  |
| 042 | Иванов | +, ым, ы, ых |  |
| 043 | сын | +, ом, я, ей |  |
| Существительные женского рода одушевленные |
| 044 | женщина | а, ой,ы, + |  |
| 045 | переводчица | а, ей, ы, + |  |
| 046 | нутрия | я, ей, и, й | Основы на букву «и» |

Kompyuter lingvistikasi asoslari

113

|  |  |  |  |
| --- | --- | --- | --- |
| 047 | швея | я, ей, и, й | см. п. 046 |
| 050 | цапля | я, ей, и, ь |  |
| 051 | санитарка | а, ой, и, + |  |
| 052 | мышь | ь, ю, и, ей |  |
| 053 | Иванова | а, ой, ы, ых |  |
| Существительные женского рода неодушевленные |
| 054 | речь | ь, ю, и, ей | Основы на ж, ч, ш, щ |
| 055 | грань | ь, ю, и, ей | см. п. 054 |
| 056 | колба | а, ой, ы, + |  |
| 057 | задача | а, ей, и, + |  |
| 060 | заготовка | а, ой, и, + |  |
| 061 | линия | я, ей, и, й | Основы на букву «и» |
| 062 | галерея | я, ей, и, й | см. п. 061 |
| 063 | земля | я, ей, и, ь |  |
| 064 | эскадрилья | я, ей, и, ий |  |
| 065 | статья | я, ей, и, ей |  |
| 066 | башня | я, ей, и, + |  |
| 067 | улица | а, ей,ы, + |  |
| Существительные среднего рода |
| 070 | место | о, ом, а, + |  |
| 071 | облако | о, ом, а, ов |  |
| 072 | поле | е, ем, я, ей |  |
| 073 | сомнение | е, ем, я, и |  |
| 074 | жилище | е, ем, а, + |  |
| 075 | перо | о, ом, я, ев |  |
| 076 | время | я, ем, а, + |  |
| 077 | побережье | е, ем, я, ий |  |

114

A. Rahimov

|  |  |  |  |
| --- | --- | --- | --- |
| 100 | колено | о, ом, и, ей |  |
| 101 | очко | о, ом, и, ов |  |
| 102 | ружье | е, ем, я, ей |  |
| \*) Символ + обозначает нулевое окончание. |

Прилагательные

|  |  |  |  |
| --- | --- | --- | --- |
| № п/п | Слово-представитель | Окончания:1) им. п., муж. р., ед.ч.,2) им. п., жен. р., ед.ч.,3) род. п., муж. р., ед.ч.,4) им. п., мн.ч. | Примечание |
| 103 | главный | ый, ая, ого, ые |  |
| 104 | передний | ий, яя, его, ие |  |
| 105 | хороший | ий, ая, его, ие |  |
| 106 | легкий | ий, ая, ого, ие |  |
| 107 | годовой | ой, ая, ого, ые |  |
| 110 | плохой | ой, ая, ого, ие |  |
| 111 | третий | ий, я, его, и |  |
| 112 | этот, сам | +, а, ого, и |  |
| 113 | мои, твои, свои | и, я, его, и |  |
| 114 | наш,ваш | +, а, его, и |  |
| 115 | весь | ь, я, его, е |  |

Глаголы в личной форме

|  |  |  |  |
| --- | --- | --- | --- |
| № п/п | Слово-представитель | Окончания 1,2 и 3-го лица ед. ч. и 3-го лица мн. ч. | Примечание |
| 116 | делать | ю, ешь, ет, ют |  |
| 117 | строить | ю, ишь, ит, ят |  |
| 120 | писать | у, ешь, ет, ут |  |
| 121 | стучать | у, ишь, ит, ат |  |
| 122 | бежать | у, ишь, ит, ут |  |
| 123 | хотеть | у, ешь, ет, ят |  |
| 124 | зависеть | у, ишь, ит, ят |  |

Morfologik analiz morfemalarga ajratish hamda grammatik ma’lumotlar tahlilini o‘z ichiga oladi. Morfologik sintez esa matnga avtomatik ishlov

Kompyuter lingvistikasi asoslari

115

berish jarayonida yoki uning so‘nggi bosqichida (natijalarni inson hukmiga havola qilishda) so‘zlarning turli shakllarini chiqarish, analiz natijalarini umumlashtirishdan iborat. Morfologik analiz va sintez jarayoni kompyuter- da kodlash-dekodlash tamoyili asosida ishlaydi. Bunda so'z turkumlari maxsus kodlar yordamida belgilanadi. Quyidagi jadvalda RETRANS (ing- lizcha-ruscha, ruscha-inglizcha frazeologik mashina tarjimasi tizimi) av­tomatik tarjima tizimida morfologik analiz jarayonida foydalaniladigan inglizcha so'zlarning grammatik tasnifmi berib o'tamiz:

15-jadval

Inglizcha so‘zlarning grammatik tasnifi

|  |
| --- |
| Существительные |
| 11 | Собственно существительные |
| 12 | Местоименные существительные |
| 13 | Количественные числительные |
| 14 | Отглагольные существительные |
| Прилагательные |
| 21 | Собственно прилагательные |
| 22 | Местоименные прилагательные |
| 23 | Порядковые числительные |
| 24 | Отглагольные прилагательные |
| Глаголы |
| 31 | Вспомогательные глаголы |
| 32 | Глаголы настоящего времени, инфинитив |
| 33 | Глаголы прошедшего времени |
| 34 | Модальные глаголы |
| Наречия |
| 41 | Собственно наречия |
| 42 | Местоименные наречия |
| 43 | Наречия сравнительной степени, прилагательные сравнительной степени |
| 44 | Отглагольные наречия (инговая форма) |
| Предлоги |
| 51 | Предлог of |
| 52 | Остальные предлоги |
| Союзы |
| 61 | Сочинительные союзы |
| 62 | Подчинительные союзы |

116

A. Rahimov

|  |
| --- |
| Артикли |
| 71 | Определенный артикль |
| 72 | Неопределенный артикль |
| Частицы |
| 81 | Частица to |
| 82 | Частицы not и по |
| 83 | Остальные частицы |

Ushbu jadvalda «1» simvoli otlarni, «2» simvoli sifatlarni, «3» simvoli fe’llarni, «4» simvoli ravishlarni, «5» simvoli predloglarni, «6» simvoli bogiovchilarni, «7» simvoli artiklni, «8» simvoli yuklamalarni bildiradi. Morfologik analiz jarayonida ushbu simvollar orqali so'zlar kodlanadi. Masalan, inglizcha matn fragmentining RETRANS avtomatik tarjima tizi­mida morfologik analiz namunasiga e’tibor beraylik.

Inglizcha matn fragmentining morfologik analiz namunasi

Language 0/11 engineering 0/11 is 0/31 the 0/71 core 0/11 of 0/51 information 0/11 technology 1/11 and 0/61 information 0/11 technology 1/ 11 will 0/31 be 0/31 the 0/71 key 0/1121 industry 1/11 of 0/51 the 0/71 next 0/21 decades 1/11 .

Inglizcha matn fragmentining RETRANS avtomatik tarjima tizimida morfologik analiz namunasi

|  |  |  |
| --- | --- | --- |
| NNo | Результаты | Классы |
| п/п | анализа | словоформ |
| 16 | No 0/82 | 6 | 82 |
| 17 | Balkan 0/21 | • v | 21 |
| 18 | state 0/1132 | T | 1132 = 11 |
| 19 | wished 0/2433 | Э | 2433 |
| 20 | to 0/81 | 5 | 81 |
| 21 | become 0/2432 | Д | 2432 |
| 22 | embroiled 0/2433 | Э | 2433 |
| 232ггбуТэ5дэИ= 11 | in 0/52 | И | 52 |
| 16 | No 0/82 | 6 | 82 |

Kompyuter lingvistikasi asoslari

|  |  |  |  |
| --- | --- | --- | --- |
| 17 | Balkan 0/21 | V | 21 |
| 18 | state 0/1132 | T | 1132 |
| 19 | wished 0/2433 | Э | 2433 = 33 |
| 20 | to 0/81 | 5 | 81 |
| 21 | become 0/2432 | д | 2432 |
| 22 | embroiled 0/2433 | Э | 2433 |
| 23 | in 0/52 | И | 52 |
| 24 | the 0/71 | 3 | 71 |
| ZZ6vT э 5дэИЗ = 33 |  |  |  |
| 16 | No 0/82 | 6 | 82 |
| 17 | Balkan 0/21 | V | 21 |
| 18 | state 0/1132 | T | 1132 |
| 19 | wished 0/2433 | э | 2433 |
| 20 | to 0/81 | 5 | 81 |
| 21 | become 0/2432 | д | 2432 = 32 |
| 22 | embroiled 0/2433 | э | 2433 |
| 23 | in 0/52 | И | 52 |
| 24 | the 0/71 | 3 | 71 |
| 25 | « | « |  |
| 26 | Great 0/21 | V | 21 |
| 6уТэ5 д эИЗ»у = 32 |  |  |  |
| 17 | Balkan 0/21 | V | 21 |
| 18 | state 0/1132 | Т | 1132 |
| 19 | wished 0/2433 | э | 2433 |
| 20 | to 0/81 | 5 | 81 |
| 21 | become 0/2432 | д | 2432 |
| 22 | embroiled 0/2433 | э | 2433 = 24 |
| 23 | in 0/52 | И | 52 |
| 24 | the 0/71 | 3 | 71 |
| 25 | « | « |  |
| 26 | Great 0/21 | V | 21 |
| 27 | War 0/11 | А | 11 |

уТэ5д э H3»vA = 24

118

A. Rahimov

Bu yerda ayrim so'zlar grammatik omonimiyaning hisobiga dastlab («tahlil natijalari»da) ikki xil raqam bilan berilgan, so'nggi bosqichda ular­ning kontekstdagi holatidan kelib chiqib aniq tasnifi keltiriladi. Yuqoridagi matn fragmentida shunday so'zlar sirasiga «state» so'zini misol qilish mumkin. «State» so'zi dastlabki tahlilda 0/1132 kodi orqali beriladi, bunda «0» raqami so'zda qo'shimchaning yo'qligini (nol morfemani), «11» raqami ot so'z turkumini, «32» raqami esa hozirgi zamon shaklidagi fe’l yoki infini- tivni bildiradi. Bunga o'xshash yana bir so'z «wished» so'zining dastlabki tahlilida bir qiymatli bo'lmagan ma’lumotlarga ega bo'lamiz. «Wished» so'zi 0/2433 kodi bilan beriladi, bunda «0» raqami yuqoridagidek nol mor­femani, «24» fe’l asosida hosil bo'lgan sifat turkumini, «33» o'tgan zamon shaklidagi fe’lni anglatadi.

So'zlarning qaysi turkumga mansubligini aniqlashda chappa (ters) lug'atlardan foydalaniladi. Bunga turli xil turkumga mansub so'zlarning morfologik analizini ko'rsatuvchi inglizcha so'z shakllarining grammatik chappa lug'ati yaqqol misol bo'la oladi.

16-jadval

Inglizcha so‘z shakllarining grammatik chappa lug‘ati fragment! (ushbu lug‘at 54 000 leksik birlikni o‘z chiga oladi)

|  |
| --- |
| So‘zGrammatik ma’lumot |
| Ot (Noun) | Sifat (Adjective) | Fe’l (Verb) |
| America 0/11 | morphologic 0/21 | tightened 0/2433 |
| Africa 0/11 | epidemiologic 0/21 | shortened 0/2433 |
| Corsica 0/11 | physiologic 0/21 | fastened 0/2433 |
| Arctica 0/11 | immunologic 0/21 | moistened 0/2433 |
| Antarctica 0/11 | hydrologic 0/21 | flattened 0/2433 |
| Casablanca 0/11 | geohydrologic 0/21 | aligned 0/2433 |
| Canada 0/11 | paleohydrologic 0/21 | standardize 0/32 |
| transcanada 0/11 | micrometeorologic 0/21 | focalize 0/32 |
| Granada 0/11 | metrologic 0/21 | realize 0/32 |
| Nevada 0/11 | neurologic 0/21 | specialize 0/32 |
| Andromeda 0/11 | paleontologic 0/21 | commercialize 0/32 |
| Florida 0/11 | allergic 0/21 | rationalize 0/32 |
| Propaganda 0/11 | readable 0/21 | neutralize 0/32 |
| addenda 1/11 | achine-readable 0/21 | visualize 0/32 |
| agenda 0/11 | degradable 0/21 | conceptualize 0/32 |
| thermobalance 0/11 | biodegradable 0/21 | stabilize 0/32 |
| microbalance 0/11 | upgradable 0/21 | utilize 0/32 |
| heatbalance 0/11 | decidable 0/21 | straightening 0/142444 |

Kompyuter lingvistikasi asoslari

119

|  |  |  |
| --- | --- | --- |
| resemblance 0/11 | undecidable 0/21 | heightening 0/142444 |
| surveillance 0/11 | formidable 0/21 | brightening 0/142444 |
| ambulance 0/11 | avoidable 0/21 | whitening 0/142444 |
| conformance 0/11 | artifical 0/21 | fastening 0/142444 |
| performance 0/11 | strategical 0/21 | listening 0/142444 |
| high-performance 0/11 | pedagogical 0/21 | peculiarities 3/11 |
| maintenance 0/11 | logical 0/21 | similarities 3/11 |
| standartization 0/11 | mineralogical 0/21 | polarities 3/11 |
| stochastization 0/11 | pharmacological 0/21 | multipolarities 3/11 |
| analyzation 0/11 | gynecological 0/21 | micropolarities 3/11 |
| radioabsorbtion 0/11 | silent 0/21 | particularities 3/11 |
| action 0/11 | excellent 0/21 | regularities 3/11 |
| co-action 0/11 | repellent 0/21 | singularities 3/11 |
| reaction 0/11 | violent 0/21 | stabilizes 1/32 |
| prereaction 0/11 | turbulent 0/21 | utilizes 1/32 |
| photoreaction 0/11 | Ravish (Adverb) | minimizes 1/32 |
| trigger 0/11 | profoundly 0/41 | optimizes 1/32 |
| debugger 0/11 | upwardly 0/41 | maximizes 1/32 |
| plugger 0/11 | outwardly 0/41 | customizes 1/32 |
| tiqer 0/11 | nicely 0/41 | reorganizes 1/32 |
| anger 0/11 | largely 0/41 | recognizes 1/32 |
| changer 0/11 | likely 0/41 | specify 1/32 |
| histler 0/11 | unlikely 0/41 | modify 1/32 |
| hustler 0/11 | extremely 0/41 | gualify 1/32 |
| settler 0/11 | timely 0/41 | implify 1/32 |
| bottler 0/11 | germetically 0/41 | unify 1/32 |
| transmitter 0/11 | genetically 0/41 | clarify 1/32 |
| knitter 0/11 | diamagnetically 0/41 | ОТ (Noun) |
| plotter 0/11 | geomagnetically 0/41 | contractions 1/11 |
| slotter 0/11 | ОТ (Noun) | abstractions 1/11 |
| relevancy 1/11 | endings 1/11 | attractions 1/11 |
| conservancy 1/11 | findings 1/11 | extractions 1/11 |
| adjacency 1/11 | disbondings 1/11 | transactions 1/11 |
| complacency 1/11 | surroundings 1/11 | mperfections 1/11 |
| dependency 1/11 | soundings 1/11 | objections 1/11 |
| interdependency 1/11 | codings 1/11 | isolators 1/11 |
| tendency 1/11 | encodings 1/11 | translators 1/11 |
| agency 1/11 | oxidizer 0/11 | turbulators 1/11 |
| interagency 1/11 | localizer 0/11 | calculators 1/11 |
| analogy 1/11 | demineralizer 0/11 | circulators 1/11 |
| trilogy 1/11 | neutralizer 0/11 |  |
| ecology 1/11 | equalizer 0/11 |  |
| oncology 0/11 | stabilizer 0/11 |  |

120

A. Rahimov

Morfologik analizning mantiqiy davomi sintaktik analiz hisoblanadi. Sintaktik analiz morfologik analiz ustiga quriladi. Bunda sintaktik analiz turli xil modellar asosida amalga oshirilishi mumkin. Formal grammatika asoschisi bo'lgan Noam Chomskiy belgilagan algoritm o'zining optimalligi bilan boshqa sintaktik modellardan ajralib turadi. Matnning avtomatik sin­taktik analizi masalasi XX asrning 50-yillarida bir tildan boshqa tilga komp­yuter yordamida tarjimani amalga oshiruvchi tizimlarga oid tadqiqotlar bilan uzviylikda o'rganilgan. Zero, matn tarjimasida gaplarning sintaktik strukturasi, sintaktik aloqasi hal qiluvchi ahamiyat kasb etadi. Dastlabki sintaktik analiz protseduralari ancha cheklangan boiib, unda faqat gap- ning strukturasi analizi e’tiborga olingan, bunda har qanday gap so'zlarning muayyan ketma-ketligi asosida tashkil topgan qurilma sifatida baholangan. Gap strukturasining formallashgan modeli sifatida, odatda, tobelik daraxti (sentence tree)dan foydalanilgan.86 Shuni alohida ta'kidlash zarurki, sintak­tik analizning samaradorligi hozirgi kunda juda yuqori saviyada, ular faqat avtomatik tarjima sistemasi ishlab chiqilgan tillarda maksimal darajada ishlatilmoqda. Lekin haligacha gapning to'laqonli semantik analizi yuza- sidan sezilarli natijalarga erishilgani yo'q. Bu muammo kompyuter lingvis- tikasidagi ko'plab muammolar bilan tutashib ketadi. Jumladan, inson tafak- kurini modellashtirish, sun’iy intellekt tizimi, semantik freym, Word-net kabilar. Agar sanab o'tilgan muammolar o'z yechimini topsa, semantik va konseptual analizning avtomatik tizimi ham to'laqonli ishlaydi.

Kompyuterdagi avtomatik tarjima kompyuter lingvistikasining eng za- monaviy yo'nalishlaridan biri hisoblanadi. Hozirgi kunda dunyoda keng tarqalgan jahon tillari uchun avtomatik tarjima dasturlari ishlab chiqilgan. Jumladan, ruschadan inglizchaga, inglizchadan ruschaga, ruschdan fran- suzchaga yoki aksi tarzida dasturlar yaratilgan. Bunday dasturlar sirasiga SOCRAT, PROMPT, MAGIC GOODDY kabilar kiradi. Ular 2 xil ta'minotga asoslanib ishlaydi: 1) lingvistik ta’minot - bunda mavjud tillarning umumiy leksikoni kiritiladi, uning grammatikasi joylashtiriladi, bir so'z bilan aytganda, maiumotlar bazasi yaratiladi; 2) programmaviy ta’minot

- bunda bir tildan ikkinchi tilga tarjima qilish algoritmi va kompyuter ana­lizi ishlab chiqiladi (maxsus programmalashtirilgan tillar yordamida).

Avtomatik tarjimaning ishlash algoritmi quyidagicha bo'ladi:

1. Boshlanish - dasturlar paneli va menyu satri ko'rinadi.

2. Ishlash rejimi tanlanadi (tarjimaning yo'nalishi - inglizchadan rus­chaga yoki ruschadan inglizchaga).

1,6 Jurafsky D., Martin J.H. Speech and Language Processing. - New Jersey, 2000.-P.353.

Kompyuter lingvistikasi asoslari

121

3. Jumla kiritiladi va xotiradagi kodi qidirib topiladi.

4. Topilgan kod maxsus dasturlar yordamida qayta ishlanadi va tarjima oynada paydo boiadi.

5. Joriy rejimda ishlashni davom ettirish haqida so'raladi.

6. Dasturni yopish haqida savol so'raladi.

7. Dastur yopiladi.

Avtomatik tarjimada samaradorlikning yuqori bo'lishiga tabiiy tillar- ning ayrim jihatlari - polisemiya (ko'p ma’nolilik), omonimlik, frazeolo- gizmlar, idiomalar, paremalar, troplar to'sqinlik qiladi. Shularni istisno qilganda, hozirda avtomatik tarjima yo'nalishida yuqori samaradorlikka erishilgan. Hatto hozirda faqat avtomatik tarjima yo'nalishiga bag'ishlangan saytlar ham yaratilgan. Bu avtomatik tarjimaning keng qamrovda taraqqiy etayotgan sohalardan ekanligini ko'rsatib turibdi.

Hozirgi kunda inglizcha-o'zbekcha va o'zbekcha-inglizcha kompyuter tarjimasi dasturlari juda katta ahamiyatga ega. Ma’lumki, ingliz va o'zbek tillari leksik-grammatik xususiyatlariga ko'ra bir-biridan tubdan farq qila­di. O'zbek tilida gap qurilishi SOV (ega+to'ld+kes) qolipida bo'lsa, ingliz tilida SVO (ega+kes+to'ld) qolipidadir; o'zbek tili agglutinativ tillar sirasi- ga kiradi, ingliz tili esa flektiv til hisoblanadi. Shuning uchun ingliz tilidan o'zbekchaga va o'zbekchadan inglizchaga kompyuter tarjimasini yaratish o'ziga xos qiyinchliklarni keltirib chiqaradi. Bugungi kunda rus tili vositasi- da ingliz tilidan o'zbek tiliga avtomatik tarjima qiluvchi dasturlarning ver- siyalari e’lon qilingan.

Ammo biz taklif qilayotgan ingliz tilidan o'zbek tiliga kompyuterda tar­jima qilish dasturi vositachi tilsiz amalga oshiriladi. Ta’kidlash lozimki, mazkur ish gaplarning sintaktik analiz qilish algoritmlari va formal model- lari bazasi asosida amalga oshiriladi.

Ingliz tilidan o'zbek tiliga kompyuterda tarjima qilish algoritmi quyida- gi vazifalarni hal etish uchun mo'ljallangan:

1.1. Analiz - bunda ingliz tilidagi gap quyidagi soddalashtirilgan model ramkasida sintaktik tahlil qilinadi.

1.1. Ushbu model faqat sodda gaplarni qamrab oladi.

2. Gapning har bir bo'lagi bitta so'zdan iborat bo'ladi.

3. Gaplarda aniqlovchilar bo'lmaydi.

4. Gaplarning standart tiplari ko'rib chiqiladi (darak gap (ega+kes+to'ld+hol), so'roq, inkor va so'roq-inkor gaplar).

5. Fe’lning quyidagi tuslanishli shakllarini qamrab oluvchi kesimli gap­lar ko'rib chiqiladi:

a) shaxs (1,11, III shaxs);

122

A. Rahimov

b) son (birlik va ko'plik);

v) zamon (o‘tgan, hozirgi, kelasi);

g) harakat tipiga ko‘ra (Simple)

d) maylga ko'ra (Indicative)

e) nisbatga ko'ra (Active)

II. Tarjimada gaplar ingliz tilidan o‘zbek tiliga o‘giriladi.

Algoritm quyidagi bosqichlardan tashkil topadi:

1) gap kiritiladi;

2) gapning har bir so‘zi massivining elementlariga qo'shiladi;

3) massivining elementlari yordamida lug‘at elementlari bilan taqqos- lanadi, bu lug‘atda olmosh, ko‘makchilar, ko‘makchi va modal fe’llar, ar- tikllar va noto'g'ri fe’llar ro'yxati mavjud bo'ladi;

4) agar so'zlar yordamchi lug'atda topilmasa, unda taqqoslash maxsus lug‘at yordamida davom ettiriladi;

5) topilgan so'zlar yordamchi lug‘atga beriladi, bu yerda so‘zga ushbu so‘zni va uning tarjimasini saqlovchi kod beriladi;

6) bunday so‘z lug'atlarda mavjud boMmasa, so‘z shakl yasovchi affiks- lardan ajratib olinadi va 5-ish bajariladi;

7) agar so‘zlar yordamchi va maxsus lug'atlardan topilmasa, ushbu so'zning yo‘qligi haqida ma’lumot kiritiladi;

8) gap 2 guruhga bo'linadi: kesimgacha bo'lgan so‘zlar ega guruhiga kiradi (ega guruhiga ega);

9) kesimdan boshlanib gapning oxirigacha bolgan so'zlar kesim guruhi hisoblanadi (kesim guruhiga: kesim, to'ldiruvchi, hoi);

10)kesim guruhidan kesim ajratib olinadi;

11)so‘ngra to‘ldiruvchi ajratiladi;

12)gapning qolgan qismi hoi hisoblanadi;

13)gapning har bir bo‘lagi shakl yasovchi qo‘shimchalarsiz tarjima qili- nadi;

14)tarjima qilingan gap boiaklaridan o‘zbek tilidagi gap tuziladi, u al- batta ingliz tilidagi gap konstruksiyasiga mutanosib bo‘ladi;

15)o‘zbek tilidagi so‘zlarga ingliz tilidagi so'zlarga mutanosib ravishda affiks va qo‘shimchalar qo‘yib chiqiladi;

16)tarjima chiqarib beriladi («tarjima» rejimida);

17)analiz chiqarib beriladi («analiz» rejimida).

Topshiriqlar:

1. Morfologik analiz va sintez qanday protseduralarni o‘z ichiga oladi?

2. Sintaktik analiz va sintezning mohiyati nimadan iborat?

Kompyuter lingvistikasi asoslari

123

3. Ingliz tilida gaplarning dastur orqali tarjimasi sifati qanday boiadi?

4. Algoritmni hayotiy misollar bilan tushuntiring.

5. Tarjima algoritmi nima?

Adabiyotlar:

1. Нелюбин J1. J1. Компьютерная лингвистика и машинный перевод.

- М.: ВЦП, 1991.

2. Шемакин Ю.И. Начало компьютерной лингвистики. - М.: МГОУ, 1992.

3. Пиотровский Р.Г. Текст, машина, человек. - М.: Просвещение. 1989.

4. Poiatov A., Muhamedova S. Kompyuter lingvistikasi. - Toshkent, 2007.

5. Белоногов Г.Г. Компьютерная лингвистика и перспективные информационные технологии. - М.: Русский мир, 2004.

KOMPYUTER LINGVISTIKASI FANINING RIVOJLANISHISTIQBOLLARI

Reja:

1. Informatsion jamiyat va kompyuter lingvistikasi.

2. Sun’iy intellekt g'oyasi.

3. Ekspert tizimlar va neyron tarmoqlar.

Tayanch so‘z va iboralar: informatsion jamiyat, sivilizatsiyaning uch to ‘Iqini, istiqbolli yo ‘nalishlar, sun ’iy intellekt, neyron tarmoqlar, ekspert tizimlari, lingvistik protsessorlar, tabiiy tilli interfeys, biokompyuter, semi- otik va biologik yondashuv, mashina ijodi, intellektual (ratsional) agent, texnologik singularlik, Tyuring testi, bilimlar injenerligi.

Insoniyat uchinchi ming yillikning boshida axborot asriga qadam qo‘ydi. Bu bilan bogiiq holda informatsion jamiyat konsepsiyasi ishlab chiqildi. Bu konsepsiyaning asoschisi amerikalik futurolog olim E.Toffler hisoblana­di. U jamiyatni turli bosqichlarga boiib o'rganishni taklif etib, sivilizatsi­yaning uch toiqinini ajratdi: a) birinchi toiqin - industrial jamiyatgacha

124

A. Rahimov

bo‘lgan davr (qishloq xo‘jaligi sivilizatsiyasi). Bunda asosiy boylik yer va tabiatdan olingan zaxiralar hisoblanadi; b) ikkinchi to‘lqin - industrial sivi- lizatsiya. Bunda asosiy boylik ishlab chiqarish vositalari sanaladi; c) uchin- chi to‘lqin - postindustrial sivilizatsiya (informatsion jamiyat). Bunda kapi- tal va mehnat o‘z o‘rnini informatsiya va bilimga bo‘shatib beradi.

Minglab yillar davomida rivojlangan inson tafakkuri uchun bugungi kunda dastlabki manba, ya’ni axborot olish hayotiy zaruratga aylandi. Shu- ning uchun jahon bozorida ma’lumot oltindan ham qimmatroq baholanadi. Mazkur axborotni topish, saqlash, qayta ishlash va boshqalarga yetkazish- ning qulay usullariga bo‘lgan ehtiyoj kun sayin oshib bormoqda. Demak, kimki mazkur qulay usullarni ishlab chiqsa, ularga egalik qilsa, dunyodagi eng boy odam boiadi. Dunyodagi eng badavlat kishi Bill Geyts kompyuter, ya’ni axborotni qabul qilish, saqlash va qayta ishlash sohasi vakili. Bill Geytsning o‘zi shunday deydi: «Kim axborotga ega bo'lsa, и hamma narsa- ga ega bo ‘ladi», uning yana bir mushohadasi dunyoga mashhur: «Axborot dunyodagi eng katta, eng qimmat moddiy boyliklardan ham ustun turadi, chunki ushbu boyliklar axborot vositasida qo'lga kiritiladi»}1

2003-yil dekabrda Jenevada (Shveytsariya) informatsion jamiyat qurish muammosiga bag‘ishlangan Butunjahon Sammiti bo‘lib o'tdi. Unda shun­day shior o‘rtaga tashlandi: «Informatsion jamiyat qurish - yangi mingyil- likning global chaqirig ‘idir». Sammitda ikki muhim hujjat qabul qilindi: Informatsion jamiyat qurish prinsiplari to‘g‘risidagi Deklaratsiya88 hamda information jamiyat qurish bo‘yicha ish Rejasi.89 Mazkur hujjatlarda belgi- lanishicha, endilikda jamiyatning taraqqiyoti kompyuter texnologiyalari bilan chambarchas bogiiq bo'lib qoladi. Shu o'rinda alohida ta’kidlash lo- zimki, kompyuter lingvistikasi informatsion texnologiyalarning istiqbolini belgilovchi eng muhim yo'nalish sanaladi hamda u informatsion jamiyat taraqqiyotida hal qiluvchi ahamiyatga ega bo'ladi.

Olimlar kompyuterlarni intellektual jarayonlar avtomatizatsiyasini yuza- ga chiqaruvchi muhim vosita sifatida baholadilar. Qisqa vaqt ichida komp- yuterlardan teoremalarni isbotlashda, rasm chizishda, shaxmat o‘yinida foydalanilishi sun’iy intellekt bo'yicha izlanishlarning ilk samarasidir. Asta-

87 Po'latov A., Muhamedova S. Kompyuter lingvistikasi. - Toshkent, 2007. - B.75 - 76.

88 World summit on the information society. Declaration of Principles. Building the Information Society: a global challenge in the new Millennium. Document WSIS-03/GENEVA/DOC/4-E. 12 December 2003.

89 World summit on the information society. Plan of Action. Document WSIS- 03/ GENEVA/DOC/5-E. 12 December 2003.

Kompyuter lingvistikasi asoslari

125

sekinlik bilan insonning aqliy faoliyatini modellashtiruvchi dasturlarning yaratilishi muayyan sohalardagi murakkab masalalarni yechuvchi ekspert tizimlarning yuzaga kelishiga olib keldi. Lekin shuni ta’kidlash zarurki, hali-hanuz inson miyasining analogi boia oladigan dastur yoki tizim yara- tilgani yo‘q.

Kompyuter lingvistikasi fanining eng istiqbolli yo'nalishlaridan biri sun’iy intellekt hisoblanadi. Virtual qomus hisoblangan Wikipedia sahifala- rida sun’iy intellektga (artificial intelligence) shunday ta’rif berilgan: «Sun’iy intellekt - bu kompyuter yoki robotning shunday xususiyatiki, u asosan in­sonning aqliy faoliyati bilan bog‘liq boigan masalalarni, xususan, fikrlash, nutq mazmunini anglash, ma’lumotlarni umumlashtirish kabilarni yechish- ga yo‘naltirilgan boiadi. Bu termin mazkur xususiyatlarga ega sistemalar- ni ishlab chiqish bilan bogliq kompyuter texnologiyalarining bir bolimiga nisbatan ham ishlatilmoqda».90 Sun’iy intellekt bo‘yicha tadqiqotlar olib borgan olim Jon Makkarti shunday yozadi: «Sun’iy intellekt - intellektual mashinalar, intellektual kompyuter dasturlar yaratish texnologiyasi va и haqidagi fan. Intellektual sistema strukturasi 3 asosiy blokni о ‘z ichiga ola- di: bilimlar bazasi, masalani yechuvchi tizim va intellektual interfeys».9[

Ekspert tizimlari modelida, intellektual informatsion tizimlarda, ba’zi mashinalarda, hayvonlarda va odamlarda intellektning turlicha darajalari va ko‘rinishlari mavjud. Mazkur turlichalik bilimlar bazasi va qanday ope- ratsiyalarni bajara olishiga bogliq. lntellektni bunday talqin qilishdan ko‘rinib turibdiki, sun’iy intellekt insondagi intellekt tushunchasining aynan immitatsiyasi emas, balki inson intellektual faoliyatining ko'plab qirralari- ni o‘z ichiga olgan, modellashtirilgan tizimlar texnologiyasi demakdir. Ayni shu ma’noda «artificial intelligence» terminini birinchi marta Jon Makkarti 1956-vilda Darmut universitetidagi konferensiyada qolladi va shundan buyon tanqidlarga qaramay ishlatib kelinmoqda. Rossiya sun’iy intellekt assotsiatsiyasi Peterburg bolimi boshlig‘i T.A.Gavrilovning yozishicha, ingliz tilidagi artificial intelligence birikmasi fantastik antropomorf bo‘yog‘iga ega emas, u rus tiliga muvaffaqiyatsiz tarjima qilingan. Intelli­gence so‘zi «aql bilan hukm chiqarish, fikr yuritish qobiliyati» ma’nosini anglatadi, u rus tiliga «интеллект» deb noto‘g‘ri tarjima qilingan.

XX asrning 70-yillariga qadar sun’iy intellekt bo‘yicha tadqiqotlar kiber- netika hamda informatika doirasida olib borilgan. XX asrning 80 - 90-yil- laridan boshlab sun’iy intellekt juda ko‘plab fanlarning o‘rganish obyektiga aylandi. Jumladan, neyrolingvistika, psixologiya, informatika, neyrofi-

40 http: //[www.wikipedia.org/wiki/artificial](http://www.wikipedia.org/wiki/artificial) intelligence.html

41 <http://www-formal.stanford.edu/jmc/whatisai.html>

126

A. Rahimov

ziologiya, epistemologiya (falsafadagi bilish haqidagi ta’limot), kogni- tologiya, kognitiv lingvistika, kompyuter lingvistikasi kabi fanlar ham o‘z doirasida sun’iy intellekt muammosi bilan shug‘ullanadilar.

Sun’iy intellekt tizimlarni yaratish muammosiga uch xil asosiy yondashuvni ajratib ko‘rsatish mumkin:92

- quyi (inglizcha Top-Down AI). semiotik yondashuv - ekspert tizimlari, bilimlar bazasi, mantiqiy hukm chiqarish tizimi, yuqori darajadagi psixik jarayonlar (tafakkur, nutq, emotsiya, ijod kabi)ning imitatsiyasi va analogi boigan tizimlarni yaratish;

- yuqori (inglizcha Bottom-Up AI). biologik yondashuv - neyron tarmoq- larini o‘rganish, biologik unsurlarga asoslangan holda intellektual salohiyatni modellashtirish va natijada nevrokompvuter yoki hiokompvuterlarni yara­tish.

- semiotik va biologik yondashuvning sintezi boigan umumlashtiruvchi gibrid tizimli yondashuv - sun’iy intellektning yakuniy maqsadi boiishi lozim.

Mazkur yondashuvlardan tashqari, yana simvolli, mantiqiy, agentga yo'naltirilgan yondashuvlar ham mavjud. Simvolli yondashuv, asosan, sim- volik mantiq qoidalarini tizimga tatbiq etish bilan bogiiq boiib, bu ko‘proq tizim qoidalarni ishlab chiqishda, hisoblashda (hukmning chin yoki yolg'onligini baholashda) samara beradi. Mantiqiy yondashuv predikatlar tilidan foydalanish asosida sun’iy intellekt tizimini bilimlar bazasining man­tiqiy modellari bilan ta’minlash uchun ximat qiladi. 1980-yilda Prolog nomli mantiqiy programmalashtirish tizimi va tili yaratildi. Prolog tilida yozil- gan bilimlar bazasi faktlar naborini hamda mantiqiy xulosa chiqarish qoi­dalarini belgilaydi. Agentga yo'naltirilgan yondashuv (yoki intellektual, ratsional agentlardan foydalanishga asoslangan yondashuv) 1990-yillarda rivojlandi. Unga ko'ra, intellekt - bu qo'yilgan masalani hisoblash qismi (rejalashtirish qismi), xolos. Bunda intellekt masalani toiaqonli yechish uchun atrofdagi tizimlar munosabatga kirisha oladigan qo‘shimcha intellek­tual (ratsional) agentga ehtiyoj sezadi. Buni 1-chizmada yaqqol ko‘rish mumkin:93

Umuman olganda, robototexnika va sun’iy intellekt bir-biri bilan aloqa- dor tushunchalar boiib, kelajakda ushbu ikki yo‘nalishning integratsiyasi natijasida intellektual robotlar yaratilishi nazarda tutilmoqda, bu esa sun’iy

92 Рассел С.. Норвиг П. Искусственный интеллект: современный подход / Artificial Intelligence: a Modern Approach / Пер. с англ. и ред. К.А.Птицына. 2- е изд. - М.: Вильямс. 2006. - С.8.

93 http: //[www.wikipedia.org/wiki/artificial](http://www.wikipedia.org/wiki/artificial) intelligence.html

Kompyuter lingvistikasi asoslari

127

1-chizma

intellektning yana bir yangi yo'nalishi hisoblanishi mumkin. PLEO, AIBO, QRIO kabi o‘yinchoq-robotlar intellektual robototexnikaga misol bo‘la oladi.

Sun’iy intellekt tizimi insonning ijodiy tafakkurini modellashtirishni ham o‘z ichiga oladi. Shu bilan bog'liq ravishda mashina ijodi tushunchasi yuza- ga keldi, bunda san’at asarlarini avtomatik tarzda yaratish jarayoni tushu- niladi. Bugungi kunda ijodni toia modellashtirishga erishilgani yo‘q, lekin bu sohada sezilarli yutuqlar qoiga kiritilgan. Jumaladan, kompyuter yor­damida musiqalar yaratilishi, treker-dasturlar, sintezatorlar, sekvensorlar yordamida bir turdagi melodiyalar va tovushlar generatsiyasi oqibatida elektron musiqalarning yaratilishi, matnning avtomatik generatsiyasi yor­damida she’rlar, ertaklar - umuman matnlar yaratilishi, rasmlar generatsi­yasi va kompyuter grafikasi yordamida kino va o‘yinlarning yaratilayot- ganligi (ayniqsa, strategiya va harbiy manyovrlarga asoslangan kompyuter o'yinlari) mashina ijodiga misol bo'la oladi.

Sun’iy intellekt tizimi doirasida tatbiq etilgan loyihalar (eksperimentlar) sifatida quyidagilarni ajratib ko'rsatish mumkin:

- Deep Blue tizimi - IBM kompaniyasi tomonidan ishlab chiqilgan shax- mat bo'yicha superkompyuter. Deep Blue 1997-yil 11-mayda shaxmat bo‘yicha jahon chempioni Garri Kasparov ustidan g‘alaba qozondi. «Deep Blue» nomi «Deep Thought» («Chuqur mulohaza, tafakkur» ma’nosini ang­latadi) loyihasiga aloqadordir.

- MYCIN tizimi - bir qator kasalliklarga shifokor singari to‘g‘ri tashxis qo‘yishni ta’minlaydi.

- 20Q - «20 savol» klassik o‘yini asosida yaratilgan loyiha boiib, u sun’iy intellekt bo‘yicha o'tkazilgan dastlabki tajribalardan biridir. Bu o'yin muallifi Robin Byorgenerdir. Bunda foydalanuvchidan biror obyektni topish

128

A. Rahimov

so'raladi, so'ng foydalanuvchiga 20 ta savol beriladi, berilgan javoblardan kelib chiqqan holda o'sha narsa topiladi. Ayniqsa, bu tizim Internet paydo bo'lgandan so'ng juda mashhur bo'ldi (<http://www>. 20q.net sayti mavjud).

- Via Voice - nutqni tovush orqali tanish tizimi.

- RoboCup - har yili soddalashtirilgan futbol bo'yicha robotlar turniri.

Sun’iy intellekt tizimini baholash bo'yicha 2 xil nuqtayi nazar mavjud:

1) sun’iy intellektni inson tafakkuridan ko'ra kuchliroq, deb baholash. Ush­bu qarash tarafdorlari sun’iy intellekt inson intellektiga qaraganda ko'proq operatsiyalarni qisqa muddatlarda bajarishi mumkin, deb hisoblaydilar; 2) sun’iy intellekt inson tafakkuridan, intellektidan kuchsiz. Bu yo'nalishdagi olimlarning fikricha, sun’iy intellektning imkoniyatlari cheklangan bo'ladi, u faqat programmaviy ta’minot asosida muayyan sohalarda ishlaydi, lekin u insonning bilish qobiliyatining, tafakkurining barcha qirralarini qamrab ololmaydi. Buni Jon Syorl tomonidan o'tkazilgan «Xitoy uyi» nomi bilan mashhur bo'lgan semantik eksperiment ham tasdiqlaydi. Olim bu tajribasi- ni 1980-yilda «Minds, Brains and Programs» («Tafakkur, miya va program- malar») nomli maqolasida e’lon qildi. J.Syorlning fikricha, kompyuter sin­taktik qoidalarni algoritm nazariyasi orqali formallashtira oladi, lekin se- mantika, tafakkurni to ‘la modellashtirishga dasturlar ojizlik qiladi,94 Xuddi shu pozitsiyada turgan olimlardan yana biri Rodjer Penrouz o'zining «Qirol- ning yangi aqli» kitobida tafakkurni formal tizimlar asosida mutlaq model- lashtirishning imkonsizligini bayon qilgan.95

Ayrim olimlar sun’iy intellekt bo'yicha erishilayotgan yutuqlarning ke­yingi oqibatlaridan qo'rqmoqdalar. AQSHda Sun’iy intellekt singular insti- tuti (SIAI) tadqiqotchisi E.Yudkovskiy insoniyatni sun’iy intellektlar bilan bog'liq global xavf kutayotganidan ogohlantiradi. Olimning fikricha, agar kelajakdagi sun’iy intellekt tizimiga insonlarga nisbatan do'stlona munosa- batda bo'lishini ta’minlovchi dastur o'rnatilmasa, butun insoniyat uchun katta xavf paydo bo'ladi.96 Ilmiy fantastik asarlar muallifi Veror Vinj ham sun’iy intellektlar tizimi to'la yaratib bo'lingan paytda jamiyatda keskin o'zgarishlar bo'lishi haqida yozgan. Bu davr ilmiy adabiyotlarda texnologik singularlik deb atalmoqda. Hatto futurologlarning fikricha, bu davr 2030- yilda boshlanadi.97

,4Searle J.R. Minds, brains, and programs // Behavioral and Brain Sciences. 1980. -P.417- 437.

95 Пенроуз P. Новый ум короля. О компьютерах, мышлении и законах физики. - М.: УРСС, 2005. - Р. 14.

96 <http://www.thetech.org/robotics/etics/index.html>

97 Kurzweil R. The Singularity Is Near. - N.Y.: Viking, 2005. - P.3.

Kompyuter lingvistikasi asoslari

129

Fanda hali-hanuz sun’iy intellektning aniq va mukammal ta’rifi mavjud emas, chunki fan bu tizimga hali toMiq erishgani yo‘q. Mavjud ta’rif va tavsiflarning hammasi mazkur sohada erishilgan yutuqlarga asoslanib kelti­rib chiqarilgan. Falsafada inson intellektining statusi va tabiati haqidaga munozaralar davom etmoqda. Kompyuter «aqllilik» darajasining aniq me- zoni ishlab chiqilgani yo‘q. To‘g‘ri, bu sohada ba’zi gipotezalar mavjud. Xususan, «Tyuring testi», «Nyuel-Saymon gipotezasi» kabi.

Sun’iy intellektga doir izlanishlar XX asrning 40-yillarida boshlangan. Sun’iy intellekt muammosiga bag‘ishlangan tadqiqotlarning eyforiyasi davri- da «Mashina fikrlay oladimi?», «Mashina о ‘z yaratuvchisidan aqlliroqmi?» sarlavhalarida maqolalar e’lon qilindi. «Mashina fikrlay oladimi» (1950) sarlavhali maqolaning muallifi ingliz olimi Alan Tyuring mashina imkoniya- tini insonning aqllilik darajasi bilan qiyoslash protsedurasini ishlab chiqdi, u «Tyuring testi» deb nomlanadi. Ushbu empirik test g'oyasi 1950-yilda «Mind» falsafiy jurnalida «Hisoblash mashinalari va tafakkur» (Computing Machinery and Intelligence) nomli maqolasida e’lon qilindi. Bu testdan ko'zlangan maqsad sun’iy tafakkurning inson tafakkuriga yaqin kelishi mumkinligini aniqlash hisoblanadi. Unga ko‘ra, inson bir kompyuter va bir inson bilan o‘zaro munosabatga kirishadi. Savollarga javob berish asosida u kim bilan (inson bilanmi yoki kompyuter dasturi bilanmi) gaplashayotga- nini aniqlashi lozim bo‘ladi. Kompyuter dasturining vazifasi esa insonni xatoga yo‘l qo‘yishiga, noto‘g‘ri tanlov qabul qilishiga majbur qildirish sana- ladi. Barcha ishtirokchilar bir-birini ko‘rmaydilar.

TURING TESTidagi savol-javob namunasi:

Q: Please write me a sonnet on the topic of the Forth Bridge.

A: Count me out on this one. I never could write poetry

Q: Add 34957 to 70764

A: (Pause about 30 seconds and then give answer as) 105621.

Test natijasida A.Tyuring shunday xulosaga keladi: «Mashina (komp­yuter) kelajakda 10 gegabaytli xotira bilan insonni 5 minut ichida 30% ga chalg'itishi mumkin bo‘ladi».w

Avtomatlashtirilgan ekspert tizimlar mantiqiy masalalarni kompyuter yordamida yechishga imkon beradi. Mazkur tizimlar ma’lum sohadagi bi- limlar majmuiga asoslangan implikativ qoidalar- «agar ...sa, u holda ...» qolipidagi mantiqiy qoidalarni o‘z ichiga oladi. Masalan, MYCIN ekspert tizimi shifokorlarga qon tarkibidagi bir qancha bakterial infeksiyalarga tash- xis qo‘yishda va ularga mos davolash metodini tanlashda yordam beradi.

98 Jurafsky D., Martin J.H. Speech and language processing. - New Jersey, 2000.-P.7.

130

A. Rahimov

MYCIN tizimining kompyuter dasturi dastlab simptomlar asosida kasalning holati haqida fikr yuritadi, so'ng mikroorganizmning kasallik qo‘zg‘otish tabiatiga mos keluvchi tashxisni aniqlaydi, bundan so'ng infeksiya sababini aniqlab, bir yoki bir nechta muqobil davolash usullarini tavsiya etadi.

Tabiiy tilda kompyuter bilan muloqot qilishni ta’minlashga yordam be- radigan dasturlar (lingvistik protsessorlar) ham yaratilmoqda. XX asrning 70-yillarida AQSHda kompyuter bilan cheklangan muloqotni ta’minlaydigan tabiiy tilli interfeys tizimi (LIFER - Language Interface Facility with Ellipsis and Recursion) yaratildi. 1985-yilda «Semantek» kor- poratsiyasi tomonidan Q&A dasturi, «Karnegi gruppa» kompaniyasi to­monidan Language Craft dasturiy ta’minoti ishlab chiqildi." Bu turdagi dasturlar hozirda ingliz tilining standartlashgan so'zlashuv uslubida uncha katta bo'lmagan lu’g'at zaxirasi bilan ishlamoqda. Tabiiy tilni tushunishga imkon beruvchi dasturlarning qamrovini kengaytirish kompyuter lingvis­tikasining dolzarb vazifalaridan biridir. Bunday dasturlarni rivojlantirish inson kabi muloqot qila oladigan, erkin kommunikatsiyaga kirisha oladi- gan sun’iy intellekt g'oyasining istiqbolini belgilaydi.

Sun’iy intellekt yo'nalishidagi tadqiqotlar sirasiga «ekspert tizimlar» yoki «neyron tarmoqlar» bo'yicha olib borilgan amaliy tadqiqotlar ham kiradi. Mazkur ikki yo'nalishga nisbatan «kompyuter intellekti» («computer intel­ligence») termini ishlatilmoqda, bunda inson tafakkurini global modellash­tirish jarayoni hamda murakkab masalalarni operativ tarzda yechishga mo'ljallangan avtomatlarning yaratilishi tushuniladi.

Sun’iy intellekt tizimi bo'yicha izlanayotgan mutaxassislar inson ta­fakkuri, ong-u shuurining barcha qirralarini modellashtirish masalasi im- konsiz qolayotganligini ta’kidlamoqdalar. Mazkur jihatlar sirasiga inson- ning ichki dunyosi, intuitsiyasi, ong osti jarayonlari, ilhom kabi sirli feno- menlarni aytish mumkin.

Inson miyasi murakkab biologik to‘r hisoblanadi, u yuz mlrd nerv tola- laridan - neyronlardan tarkib topgan. Neyronlarning bir-biriga uzviy alo- qadorligi, zanjirli bog'lanishi natijasida miya faoliyati me’yoriy darajada ishlaydi. Zamonaviy sun’iy neyron tarmoqlar inson miyasidagi neyronlar modeli asosida yaratilmoqda. Hozirgi kunda neyron tarmoqlar bir qator amaliy masalalarni hal etishda keng qo'llanilmoqda. Jumladan, ob-havo prognozini aniqlashda, obrazlarni tanishda (xususan, og‘zaki va yozma nutqni aniqlashda), robotlarni boshqarishda, tibbiyotda tashxis qo‘yishda ishlatilmoda. Amerikaning NASA kosmik agentligida neyron tarmoqlar-

99 Анисимов А.В. Компьютерная лингвистика для всех: мифы. Алгоритмы. Язык. - Киев: Наукова думка, 1991. -С.10 - 11.

Kompyuter lingvistikasi asoslari

131

dan fazoda robotlarni boshqarish, hududga tasodifan kelib tushgan obyekt- larni tutishda foydalanilmoqda. «General Dynamics» kompaniyasi esa dvi- gatel shovqiniga qarab kemalarni tanishga imkon beruvchi tizim ishlab chi- qishdi. AQSH Pensilvaniya universitetida uchuvchi obyektlarni (NUJ) aniqlash uchun neyron tarmoqlar tizimi yaratildi.

Neyron tarmoqlarining ishlash tizimi matematik ifodalar yoki mantiqiy qoidalarga asoslanmaydi, balki namuna (example based approach) asosida o‘qitishga asoslangan. Bunda muayyan masalani yechish uchun namunalar (misollar), ya’ni o'qituvchi tanlamalar tayyorlanadi va tarmoq shu asosda tekshiradi. Bunda turli. xil metodlardan foydalaniladi. Eng keng tarqalgan metod - bu dastlabki yechim va sinov natijalarini e’tiborga olgan holda neyron tarmoqlari strukturasini korrektirovka qilishga asoslangan metod (back propagation method) bo‘lib, unda oldindan ko'plab o‘qitish tanlama- lari tayyorlab olinadi, u o‘z bazasidagi mavjud ma’lumotlarni yangi kiriti- layotgan o‘qituvchi namuna-tanlamalar bilan solishtiradi. Agar ma’lumotlar bir-biriga mos kelsa, neyron tarmoqlari strukturasida korrektirovka amal­ga oshirilmaydi. Faqatgina ma’lumotlar bir-biriga muvofiq kelmaganda korrektirovka yuz beradi. Shundan so‘ng neyron tarmoqlarida muayyan amaliy vazifani hal etish tizimi tola shakllanadi.100

Sun’iy intellekt tizimiga doir amaliy ishlardan biri ekspert tizimlari hisob­lanadi. Ekspert tizimi - bu muayyan sohadagi ekspertlik yoki mavjud bilim- larning imitatsiyasi tarzida yaratilgan kompyuter dasturi. Boshqacharoq aytganda, ekspert tizimi maxsus sohaga oid barcha bilimlar jamlangan qo- mus yoki o‘ta professional ekspertlarning kompyuter modelidir. Ekspert tizimlari odatiy kompyuter dasturlaridan farq qiladi. Odatiy kompyuter das- turlari qidiruv tizimi, hisoblash vamanipulyatsiya tizimigagina ega boiadi. Ekspert tizimlari mazkur jihatlardan tashqari, xuddi insonlarda bolgani kabi dalillarga tayanib muhokama yuritish imkoniyatiga ham egadir. Eks­pert tizimlarining asosiy ikkita tarkibiy qismi mavjud: ma’lumotlar ombori va mantiqiy mashina.

Ekspert tizimlari konsepsiyasi Stenford universiteti professori Eduard Feygenbaum nomi bilan bogliq bolib, u 1977-yilda murakkab mantiqiy masalalarni yechishda kompyuter dasturlarining samaradorligi formallash- tirish va dasturlash texnikasidan ko‘ra muammoli sohaga doir bilimlar ba-

100 Anderson J. and Rosenfeld E. Neurocomputing Foundations of Research.

- Cambridge: MIT Press, 1998; Fukushima K. A Neural Network for Visual Pattern Recognition. IEEE Computer, 1988; Hopfield J.J. Neural Networks and Physical Systems with Emergent Collective Computational Abilities. Proceedings of the National Academy of Sciences, 1982.

132

A.Rahimov

zasiga bog‘liq ekanligini asoslab berdi.101 Dastlab ekspert tizimlari diagnos- tika va kasallikni davolash sohasiga tatbiq etildi. Ekspert tizimlari keyin- chalik kimyo, biologiya, geologiya sohalarida ham qoilanila boshlandi.

Ekspert tizimlariga bilimlar bazasini kiritish uchun simvolli maiumotlarga ishlov berishga xizmat qiladigan LISP, REFAL, PROLOG kabi sun’iy tillardan foydalanilmoqda. Bu tillar evristik bilimlar va manti­qiy xulosalarni yuzaga chiqarishda eng qulay programmalashtirish tillari hisoblanadi. Bunda yagona qabul qilingan bilimlar bazasi uchun deklarativ formalizm hamda qaror qabul qilish tizimi uchun protsedura tizimi qo‘llaniladi.

Ekspert tizimlari ichiga ekspertlar hamda bir qator yordamchi program- malar kiradi:

- savol-javobli dastur foydalanuvchi bilan tabiiy tilda munosabatga ki- rishishni ta’minlab beradi;

- bilimlar ombori dasturi foydalanuvchiga bilimlar bazasiga o'zgartirishlar kiritish hamda toidirib, boyitib borishga imkon beradi;

- tushuntiruvchi dastur - javoblar tizimini tekshiradi;

- interpretator-dastur muayyan sohadagi terminlarni sharhlash, izohlash- ga yordam beradi;

- ijro etuvchi dastur barcha tizimlar ishini ta’minlab beradi.

Yuqorida aytilganidek, ekspert tizimlari deklarativ va protsedurali bos-

qichlar asosida ishlaydi. Deklarativ bosqich - bu ekspert tizimiga kiritilgan bilimlar bazasi bo‘lsa, protsedurali bosqich ekpert tizimlarining mavjud bi­limlar bazasiga tayanib hukm chiqarish va aniq qaror qabul qilish bosqichi- ni o‘z ichiga oladi. Ekspert tizimlarining ishlash prinsipiga doir bir misol keltiraylik. Masalan, qanday sababga ko‘ra avtomobildagi benzonasosdan benzin ajralishi qisqarmoqda?

Bu holat bo'yicha ekspert tizimlariga kiritilgan deklarativ bilimlar qu- yidagilar: klapanning tiqilishi, filtrning axlatga toiganligi, klapan tepa qismining ifloslanganlgi, nasos membranasining zararlanganligi. Agar ben­zonasosdan benzin ajralishi yetarli boimasa, u holda kiritish shlangi axlat­ga to‘lib, tiqilib qolgan boiishi mumkin kabi.

Informatsiya shiddat bilan ko'payib borayotgan bugungi davrda ekspert tizimlarining bilimlar bazasini uzluksiz toidirib, yangi bilimlar bazasi bi­lan boyitib borish birlamchi vazifa boiib qolmoqda. Chunki vaqt o‘tgan sari mavjud bilimlar eskirib, talabga javob bermay qolishi mumkin, shunda ularni yangi va qo‘shimcha ma’lumotlar bilan toidirish joiz boiadi. Bilim-

101 Белоногов Г.Г. Компьютерная лингвистика и перспективные информационные технологии. - М.: Русский мир, 2004. - С.226.

Kompyuter lingvistikasi asoslari

133

larning mavjud holatlar bilan muvofiqligi hamda bilimlar bazasining bo- yitilishi ekspert tizimlarining tushuntiruvchi programmalari asosida amal­ga oshiriladi. Hatto keyingi paytlarda bilimlar bazasi bilan shug'ullanuvchi maxsus yo‘nalish ham shakllandi. Bu yo'nalish «bilimlar injenerligi» («Knowledge Engineering») deb ataladi.

Bilimlar injenerligi kognitiv lingvistika, kompyuter semantikasi va komp­yuter lingvistikasining yutuqlariga tayanadi. Bunda informatsiya va bilim­lar strukturasiga oid boigan semantik tarmoq, freym, senariy, Word-net kabi tushunchalar va ularning tatbiqi muhim hisoblanadi. Semantik tarmoq borliqni so'zlar va ularning semantik munosabatlari yordamida modellash- tirishga yo‘naltirilgan tizim hisoblanadi. Semantik tarmoq aniq belgilan- gan uzellar (obyektlar kesishadigan joy) va dugalar (semantik munosabat- larning biri ikkinchisiga ta’sir ko‘rsatadigan nuqta) munosabatdorligi asosida modellashtiriladi. Semantik tarmoq g‘oyasi kompyuter texnologiyalarida- gi lokal va global tarmoq nazariyasi ta’sirida kelib chiqqan. Uzellar muay­yan obyektlarga, muayyan sohaga oid tushunchalarga, dugalar obyektlar o‘rtasidagi semantik munosabatlarga aloqador tushunchalardir. Semantik

134

A. Rahimov

tarmoq ikki qismdan tashkil topadi: 1) mavhum tarmoq; 2) aniq tarmoq. Mavhum tarmoq umumiy tushunchalar orqali ifodalanadigan (zamon, ma- kon, miqdor, sifat kabi) bilimlar fondini o‘z ichiga oladi; 2) aniq tarmoq mavhum tarmoq tarkibiga kiruvchi tushunchalarning real obyektlarini va konkret bilimlar fondini o‘z ichiga oladi.

Semantik tarmoq tizimi quyidagi munosabatlarni o'rnatish orqali yara­tiladi:

1. To‘plam va obyekt o‘rtasidagi munosabat (muayyan obyekt aniq to'plamga mansub bo‘ladi) tasnif munosabati deyiladi. Masalan, «Sharik it hisoblanadi» holati kabi. Ba’zan ushbu munosabat Member Of, Instance Of deb ham yuritiladi.

2. Tur va jins o'rtasidagi munosabat AKO - «А Kind Of», «Subset Of» («turlaridan biri») deb ataladi. Masalan: «It hayvon hisoblanadi» holati kabi. Sistem leksikologiyada tur uchun giponim (it), jins uchun giperonim (hayvon), tur-jins munosabati uchun giponimiya terminlari qabul qilingan.

3. Butun va qism o‘rtasidagi munosabat sistem leksikologiyada mero- nimiva termini bilan ataladi. Masalan, kompyuter - butun; monitor, klavia- tura, sichqoncha, protsessor kabilar - qismdir. Butun uchun xolonim, qism uchun meronim atamalari ishlatiladi. Demak, motor - avtomobil uchun meronim, avtomobilning o‘zi esa motor uchun xolonim sanaladi.

Bilimlar strukturasiga oid tayanch tushunchalardan yana biri freym hisob­lanadi. Freym (inglizcha «frame» - «karkas», «ramka», «skelet» ma’nolarini anglatadi) - sun’iy intellekt tizimida bilimlarni, stereotip (juda ko‘p kuzatiladigan, tipik) vaziyatlarni gavdalantirishning bir usuli sanala­di. Ya’ni freymga modellashtirilayotgan hodisa, vaziyat va jarayonlarni umumlashtirilgan holda joylashtirish mumkin. Freym - mavhum obraz modeli bo‘lib, muayyan obyekt, voqea-hodisa, vaziyat va jarayonlarning formal tavsifi demakdir (masalan, restoranga borish, shifokor qabulida boiish, ma’ruzalarga qatnashish, tug‘ilgan kun, futbol o‘yini kabi tipik vaziyatlar). Dastlab «freym» termini XX asrning 70-yillarida Marvin Mins­kiy tomonidan olamning kompyuter modelini o‘z ichiga oluvchi bilimlar strukturasiga nisbatan ishlatilgan.102 Shundan so‘ng freym semantikasi kon- sepsiyasi (frame semantics) amerikalik tilshunos Charlz Fillmor tomonidan rivojlantirildi. Mazkur konsepsiyaga ko'ra, so‘zlar, so‘z birikmalari, gap­lar, matn ma’nosi sahnalar bilan uyg'unlashtirilgan, shartlangan boiadi. Bir freym boshqa bir freymning davomi yoki tarkibiy qismi boiishi mumkin. Ch.Fillmorning qarashlariga L.Tenyening aktantlar (fe’lning bo‘sh

102 Minsky M. A Framework for Representing Knowledge / The Psychology of Computer Vision. McGraw-Hill, New York (U.S.A.), 1975.

Kompyuter lingvistikasi asoslari

135

o'rinlarini toMdiruvchilar) haqidagi g'oyasi katta ta’sir ko'rsatgan. Fillmor- ning Case (kelishiklar) g'oyasiga asosan aktantlar (fe’lning hamrohlari) fe’l yordamida ifodalanadigan holat yoki harakat munosabatlaridagi bajarila- digan «rollar» sifatida talqin qilinadi. Ch.Fillmor ajratgan kelishiklar quyi- dagilar:

1. Agent (A) - harakat bajaruvchisi (inglizcha «agent» so‘zidan olingan).

2. Теша (T) - predmet (inglizcha «theme» so'zidan olingan).

3. Manba (S) - harakat obyektining dastlabki holati (inglizcha «source» so‘zidan olingan).

4. Maqsad (G) - harakat ijrosidan keyingi yakuniy holati (inglizcha «goal» so‘zidan olingan).

5. Vosita (I) - harakat uchun qurol-vosita (inglizcha «instrument» so'zidan olingan).

6. Usui (W) - harakat usuli (inglizcha «way» so‘zidan olingan).

Freymlar sun’iy intellekt tarkibiga kiruvchi ekspert tizimlarida faol ishla­tiladi. Freymning quyidagi turlari mavjud: freym-namunalar, freym-struk- turalar, freym-rollar, freym-ssenariylar, freym-vaziyatlar. Freymlar tizimi birlashib, semantik tarmoqni hosil qiladi. Freym kompyuter xotirasida bilim- lar xazinasi va faktlar omboriga tayanib olamni maxsus shakllantiruvchi model vazifasini bajaradi. Freym slot deb atalgan bir yoki bir necha ele- mentlarning majmui hisoblanadi. Slotlarda obyektlarga xos belgilar ma’nosi qayd etilgan bo'ladi.

17-jadval

«Stol» freym fragmentining jadval ko‘rinishidagi reprezentatsiyasi

|  |  |
| --- | --- |
| Slot nomi | Slot mazmuni |
| Oyoqchalar miqdori | To rtta, undan ko'proq bo'lishi ham mumkin |
| Materiali | Yog'och, plastmassa, shisha |
| Yuzasi, sirti | Oval, kvadrat, aylana |
| Tumbaning (tagkursining) mavjudligi | Ixtiyoriy tarzda |
| Vazifalari | Ovqatlanish uchun, dars tayyorlash uchun, ish uchun va hokazo |
| Boshqalar |  |

Freym-senariy ba’zi harakat, voqea-hodisalarning xarakterli unsurlari- ni o‘z ichiga olgan tipik strukturalar hamda vaziyatlarning rivojlanishi, ketma-ketligi (algoritm bosqichlari) demakdir. Ssenariy bu oddiy voqea- hodisalar zanjiri emas, balki ushbu harakatlar ketma-ketligida kauzal bogiiqlik tamoyili ishlaydi, ya’ni har bir harakat natijasi shartli sanaladi,

136

A. Rahimov

uning ta’siri va oqibatida navbatdagi ish-harakat sodir boiishi mumkin. Har bir ssenariyda rol bajaruvchilar mavjud. Ssenariylar real hayotda uchraydigan vaziyatlarni aks ettiruvchi holatlar asosida yuzaga keladi. Masalan, restoranga tashrif buyurish ssenariysini olib qaraylik.

Ssenariy: restoran

Rollar: tashrifchi, ofitsiantka, xo'jayin, kassir.

Maqsad: Qorinni to‘yg‘izish uchun ovqat olish.

Sahna I. Kirish

Restoranga kiriladi.

Bo‘sh o‘rindiqlar bor joyga nazar tashlanadi.

O'tirish uchun joy tanlanadi.

Tanlangan joyga boriladi.

O'tiriladi.

Sahna II. Buyurtma

Taomnoma (menyu) olinadi

Menyu o‘qiladi

Nimani buyurtma qilish haqida kelishiladi.

Ofitsiantkadan tanlangan taomni keltirish so‘raladi.

Sahna III. Taom

Ovqat olinadi.

Ovqat yeyiladi.

Sahna IV. Chiqish

Hisob so'raladi.

Chek olinadi.

Kassirning oldiga boriladi.

Pul to‘lanadi.

Restorandan chiqiladi.

Ko‘pchilik mutaxassislarning fikricha, ekspert tizimlari va neyron tar- moqlari sun’iy intellekt bo‘yicha izlanishlarda o‘zaro musobaqalashuvchi ikki muqobil yo‘nalishlardir. Har ikkala yo‘nalishning o‘ziga xos afzal va kamchilik tomonlari mavjud. Shuni e’tiborga olgan holda aytish mumkin- ki, neyron tarmoqlari va ekspert tizimlarining birgalikda ishlashini ta’minlovchi gibrid tizimlar ishlab chiqilishi lozim. Shunda qo'yilayotgan masalaning xarakteridan kelib chiqqan holda mos tizim tanlash va muam- moga optimal yechim topish imkoniyati yuzaga keladi. Agar masala «qoi- da bo'yicha» (implikatsiya - if..., then... - agar.. ,-sa, u holda...) hal etilsa, ekspert tizimlariga murojaat qilish zarur. Agar masalada ko'pgina empirik ma’lumotlar mavjud bo'lsa, neyron tarmoqlaridan foydalanish maqsadga muvofiq. Yoki obrazlarni tanishda (masalan, moliyaviy holatni aniqlash-

Kompyuter lingvistikasi asoslari

137

da) asosan neyron tarmoqlaridan, obraz aniqlangandan keyin bajariladi- gan mantiqiy amallarda esa ekspert tizimlaridan foydalanilsa, yuqori sa- maradorlikka erishish mumkin.

Topshiriqlar:

1. Kompyuter lingvistikasi fanining zamonaviy yo‘nalishlari haqida In­ternet ma’lumotlarini toping.

2. Lingvistik kompyuter dasturlarining ishlash samaradorligini oshirish uchun nimalar qilish lozim?

3. Ekspert tizimlari qaysi sohalarda ishlatilmoqda?

4. Sun’iy intellekt tizimi bo'yicha erishilgan yutuqlarga nimalarni kiri- tish mumkin?

5. Sun’iy intellekt tizimini yaratish yuzasidan qanday yondashuvlar mavjud?

Adabiyotlar:

1. Нелюбин Jl. Jl. Компьютерная лингвистика и машинный перевод. - М.: ВЦП, 1991.

2. ШемакинЮ.И. Начало компьютерной лингвистики.-М.: МГОУ, 1992.

3. Белоногов Г.Г. Компьютерная лингвистика и перспективные информационные технологии. - М.: Русский мир, 2004.

4. Рассел С.. Норвиг П. Искусственный интеллект: современный подход // Artificial Intelligence: a Modern Approach / Пер. с англ. и ред. К.А.Птицына. 2-е изд. - М.: Вильямс. 2006.

5. Searle J.R. Minds, brains, and programs // Behavioral and Brain Sci­ences. 1980.

6. Jurafsky D., Martin J.H. Speech and language processing. - New Jersey, 2000.

7. Minsky M.A. Framework for Representing Knowledge // The Psy­chology of Computer Vision. McGraw-Hill, New York (U.S.A.), 1975.

8. Анисимов А.В. Компьютерная лингвистика для всех: Мифы. Алгоритмы. Язык. - Киев: Наукова думка, 1991.

9. Po'latov A., Muhamedova S. Kompyuter lingvistikasi. - Toshkent, 2007.

138

A. Rahimov

FOYDALANILGAN ADABIYOTLAR RO‘YXATI

1. Abdullayev Y. Statistika nazariyasi. - Toshkent, 2002.

2. Анисимов А.В. Компьютерная лингвистика для всех: Мифы. Алгоритмы. Язык. - Киев: Наукова думка, 1991.

3. Апресян Ю.Д. Идеи и методы современной структурной лингвистики. -М.: Просвещение, 1966.

4. Ахманова О.С. Словарь лингвистических терминов. - М.: СЭ, 1966.

5. Баранов А.Н. Введение в прикладную лингвистику. - М.: Эдиториал УРСС, 2001.

6. Бархударов Л.С. Структура простого предложения английского языка. -М.: Высшая школа, 1966.

7. Белоногов Г.Г. Компьютерная лингвистика и перспективные информационные технологии. - М.: Русский мир, 2004.

8. Белоногов Г.Г., Зеленков Ю.Г., Новоселов А.П., Хорошилов А.А., Хорошилов А.А. Метод аналогии в компьютерной лингвистике // Научно- техническая информация. Сер.2. № 1. ВИНИТИ. 2000.

9. Буранов Дж. Сравнительная типология английского и тюркских языков. - М.: Высшая школа, 1983.

10. Гладкий А.В., Мельчук И. А. Элементы математической лингвистики. - М.: Наука, 1969.

11. Гринберг Дж. Квантитативный подход к морфологической типологии языков // Новое в лингвистике. - М., 1963. Вып.III. - С.60 - 94.

12. Гринберг Дж. Некоторые грамматические универсалии, преимущественно касающиеся порядка значимых элементов // Новое в лингвистике. Вып.У. Языковые универсалии. - М.: Прогресс, 1970. - С.115 - 162.

13. Диалектика и логика. Законы мышления. - М.: Наука, 1962.

14. Есперсен О. Философия грамматики. - М.: Иностранная литература, 1958.

15. Е^убовТ., Каримбеков С. Математик манти^ элементлари. -Тошкент: У^итувчи, 1996.

16. Жуков Д.А. Мы переводчики. - М.: Знание, 1975.

17. Звегинцев В.А. История языкознания XIX - XX веков в очерках и извлечениях. В 2-х частях. - М.: Просвещение, 1964 - 1965.

18. Зубов А.В., Зубова И.И. Основы лингвистической информации. - Минск: МГПИИЯ, 1992.

19. Кондрашов Н.А. История лингвистических учений.-М.: Просвещение,

1979.

20. Кревский И.Г., Селиверстов М.Н., Григорьева К.В. Формальные языки, грамматики и основы построения трансляторов. - Пенза, 2003.

21. Лагута О.Н. Логика и лингвистика. - Новосибирск, 2000.

22. Лингвистический энциклопедический словарь. - М.: СЭ, 1990.

Kompyuter lingvistikasi asoslari

139

23. Лурия A.P. Язык и сознание. - М.: МГУ, 1979.

24. Марчук Ю.Н. Проблемы машинного перевода. - М.: Наука, 1983.

25. Набиева Д. Узбек тилининг турли сатхларида умумийлик-хусусийлик диалектикасининг намоён булиши. - Тошкент: Шар^, 2005.

26. Нелюбин Л.Л. Компьютерная лингвистика и машинный перевод. - М.: ВЦП, 1991.

27. Неъматов Бозоров О. Тил ва нут^. - Тошкент: У^итувчи, 1993.

28. Новое в лингвистике. Вып.У. Языковые универсалии. - М.: Прогресс, 1970.

29. Новое в зарубежной лингвистике. Вып.24. Компьютерная лингвистика.

- М.: Прогресс, 1989.

30. Нурмонов А. Лисоний белги хусусиятлари ^акида. - Андижон, 1992.

31. Нурмонов А., Йулдошев Б. Тилшунослик ва табиий фанлар. -Тошкент: Укитувчи, 2001.

32. Nurmonov A. Lingvistik tadqiqot metodologiyasi va metodlari. - Tosh- kent: Akademnashr, 2010.

33. Nurmonov A. Struktur tilshunoslik: ildizlari va yo‘nalishlari. - Toshkent,

2008.

34. Пенроуз P. Новый ум короля. О компьютерах, мышлении и законах физики. - М.: УРСС, 2005.

35. Пиотровский Р.Г. Инженерная лингвистика и теория языка. - Л., 1979.

36. Пиотровский Р.Г. Текст, машина, человек. - Л.: Наука, 1975.

37. Плат У. Математическая лингвистика // Новое в лингвистике. Вып.IV.

- М.: Прогресс, 1965.

38. Po'latov A., Muhamedova S. Kompyuter lingvistikasi. - Toshkent, 2007.

39. Рассел С.. Норвиг П. Искусственный интеллект: современный подход / / Artificial Intelligence: a Modern Approach / Пер. с англ. и ред. К.А.Птицына. 2- е изд. - М.: Вильямс. 2006.

40. Реформатский А.А. Введение в языковедение. - М.: Просвещение, 1967.

41.Сиокум Дж. Обзор разработок по машинному переводу: история вопроса, современное состояние и перспективы развития // Новое в зарубежной лингвистике. Вып.XXIV. Компьютерная лингвистика. - М.: Прогресс, 1989.

42. Собиров А. Узбек тилининг лексик сат^ини системалар системаси тамойили асосида тадкиь; этиш. - Тошкент: Маънавият, 2004.

43. Соссюр Ф. де. Курс общей лингвистики // Труды по языкознанию. - М.: Прогресс, 1977.

44. Степанов Ю.С. Основы общего языкознания. - М.: Просвещение, 1975.

45. Субботин М.М. Гипертекст. Новая форма письменной коммуникации. ВИНИТИ. Сер. Информатика. 1994.

46. Сусов И.П. История языкознания. - Тверь: Тверской гос. унт, 1999.

47. Теньер Л. Основы структурного синтаксиса. - М.: Прогресс, 1988.

48. Убин И.И. Современные средства автоматизации перевода: надежды, разочарования и реальность // Перевод в современном мире. -М.: ВЦП, 2001.

140

A. Rahimov

49. Усмонов С. Умумий тилшунослик. - Тошкент: Укитувчи, 1972.

50. Успенский В.А., Семенов A.JI. Теория алгоритмов: основные открытия и приложения. - М.: Наука, 1987.

51. Фалсафа. Кискача изошли лугат. - Тошкент: Шарк, 2004.

52. Фосслер К. Позитивизм и идеализм в языкознании // Звегинцев В.А. История языкознания XIX - XX веков в очерках и извлечениях. Часть 1. - М.: Просвещение, 1964. - С.324- 335.

53. Херрис З.С. Совместная встречаемость и трансформация в языковой структуре II Новое в лингвистике. Выпуск II. - М., 1962.

54. Шаумян С.К. Теоретические основы трансформационной грамматики // Новое в лингвистике. Выпуск II. - С.391 - 411.

55. Шемакин Ю.И. Начало компьютерной лингвистики. - М.: Высшая школа, 1992.

56. Шерму^аммедова Н. Фалсафа ва фан методологияси. - Тошкент: Университет, 2005.

57. Эко У. Отсутствующая структура. Введение в семиологию. - М.: Петрополис, 1998.

58. Anderson J., Rosenfeld Е. Neurocomputing Foundations of Research. - Cambridge: MIT Press, 1998.

59. Belonogov G., Khoroshilov A., Kuznetsov B., Novoselov A., Zelenkov Yu. Systems of Phraseological Machine Translation of Polythematic Texts from Russian into English and from English into Russian (RETRANS and ERTRANS Systems). International Forum on Information and Documentation. Vol. 20. №2.

1995.

60. Campbell L., Bubenik V., Saxon L. Word Order Universals: Refinements and Clarifications // Canadian Journal of Linguistics. №2. 1988. - P.209 - 230.

61. Chomsky N. Aspects of the Theory of Syntax. - Cambridge, Mass.: MIT Press, 1965.

62. Fukushima K. A Neural Network for Visual Pattern Recognition. IEEE Computer, 1988.

63. Grishman R. Computational linguistics//Cambridge University Press. 1994.

64. Gelbukh A. Computational Linguistics and Intelligent Text Processing. International Conference. - Mexico, 2003.

65. Good J. Linguistic Universals and Language Change // Oxford University Press. 2008.

66. Hopfield J.J. Neural Networks and Physical Systems with Emergent Col­lective Computational Abilities // Proceedings of the National Academy of Scienc­es. 1982.

67. Hutchins J. Retrospect and prospect in computer-based translation // Pro­ceedings of MT Summit VII «МТ in the great translation era». - Tokyo: AAMT,

1996.

68. Jurafsky D., Martin J.H. Speech and Language Processing. - New Jersey, 2000.

Kompyuter lingvistikasi asoslari

141

69. Kurzweil R. The Singularity Is Near. - N. Y.: Viking, 2005.

70. Manning C., Raghavan P., Schbtze H. Introduction to Information Re­trieval // Cambridge University Press. 2008.

71. Minsky M. A Framework for Representing Knowledge // The Psychology of Computer Vision. McGraw-Hill. - New York (U.S.A.), 1975.

72. Nagao M. A framework of a mechanical translation between Japanese and English by analogy principle, in Artificial and Human Intelligence, ed. A. Elithorn and R. Banerji. - North Holland, 1984. - P.173 - 180,

73. Ricardo М., Juana G. Linguistic Universals. - Cambridge: Cambridge University Press, 2006.

74. Searle J.R. Minds, brains, and programs // Behavioral and Brain Sciences.

1980.

75. Thomas S. Kuhn. The Structure of Scientific Revolution. - Chicago: Uni­versity of Chicago Press, 1970.

76. Webb Lynn E. Advantages and Disadvantages of Translation Memory: a Cost // Benefit Analysis. San Francisco State University. 1992.

77. World summit on the information society. Declaration of Principles. Building the Information Society: a global challenge in the new Millennium. Document WSIS-Q3/GENEVA/DOC/4-E. 12 December 2003.

INTERNET SAYTLARI

1.

2.

3.

4.

5.

6.

7.

8.

9.

10. 11. 12.

13.

14.

15.

16.

17.

18.

19.

20.

<http://www.vaal.ru>

<http://www.rusf.ru/books/analysis>

<http://www.ruscorpora.ru> (corpora.yandex.ru) <http://www.philol.msu.ru/~lex/corpus>

<http://www.philol.msu.ru/~humlang/articles/PolystylCorp.htm>

<http://www.bokrcorpora.narod.ru/frqlist/frqlist.html>

<http://www.info.ox.ac.uk/bnc>

http: //[www.wikipedia.org/wiki/artificial](http://www.wikipedia.org/wiki/artificial) intelligence.html http: // [www.sigweh.org/conferences/ht-cover.shtml](http://www.sigweh.org/conferences/ht-cover.shtml)

http: // [www.yazvkoznanie.ru](http://www.yazvkoznanie.ru) http:[www.corpus.leeds.ac.uk/list.html](http://www.corpus.leeds.ac.uk/list.html)

http: [www.CICLing.org](http://www.CICLing.org)

142

A. Rahimov

KOMPYUTER LINGVISTIKASI TERMINLARI LUG‘ATI (GLOSSARIY)

Aksiomatik nazariya - ilmiy nazariya qurilishi, bu nazariya asosida shunday boshlang'ich postulat va aksiomalar qo‘yiladiki, qolgan barcha qoidalar (teore- malar) undan kelib chiqadi. U miloddan avvalgi VII - VI asrlarda shakllangan, asoschisi matematik olim Pifagordir. Aksiomatik nazariyaning quyidagi asosiy tamoyillari mavjud:

- ziddiyatsizlik - aksiomalar tizimining shunday xossasiki, unga ko'ra, siste- maning hech qanday ikkita holati bir-biriga zid kelmasligi kerak. Bunda aksiomalar tizimi o‘zida rost va yolg'on orasidagi farqni aks ettirishi kerakligi nazarda tutiladi;

- to'lalik - aksiomalar sistemasiga qandaydir jumlani qo'shganimizda siste- mada ziddiyat sodir bo'ladi;

- bog'liqsizlik - berilgan aksiomalar sistemasining u yoki bu sistemalari shu sistemaning boshqa sistemalari yordamida isbot qilinishi mumkin emas.

Algoritm (lotincha «algoritmus» - «ibtido» ma’nosini anglatadi; bu so'zning etimologiyasi buyuk alloma Al-Xorazmiy nomi bilan bogiiq. U Yevropada buzil- gan shaklda «Algorithmi» tarzida ishlatilgan) - ma’lum bir tipga oid hamma masala- larni yechishda ishlatiladigan jarayonlar tizimining muayyan tartibda bajarilishi haqidagi aniq qoida yoki qoidalar tizimi.

Algoritmik tillar - programmalashtirishga asoslangan sun’iy tillar boiib, un- dagi matnlar xuddi tabiiy tildagidek aniq interpretatorga - insonga yoki avtoma­tik qurilmaga yo'nalgan bo'lishi lozim, toki u axborotni qabul qilsin va tushun- sin. Agar bu jarayon (qabul qilish va tushunish) amalga oshmasa, u holda algorit­mik tillar va kompyuter orasida vositachi sifatida translator-dasturlar (yoki kompilatorlar)dan foydalamiadi. Algoritmik tillar imperativ xarakterga ega bo'ladi, ya’ni asosan buyruqlar tizimiga tayanadi (begin - boshla, end - tugalla kabi). Dastlabki shunday til 1958-yilda Parijda yaratilgan va «Algol» deb atalgan (Uning to'liq nomi inglizcha - «Algorithmic language»). Hozir jahonda 500 dan ortiq algoritmik tillar mavjud.

Analiz (tabiiy tilning kompyuterda morfologik, sintaktik va semantik tahlil yordamida tushunilishi) va sintez (kompyuterda matnning grammatik shakllanti- rilishi va generatsiyasi (hosil qilinishi) - kompyuter lingvistikasining tayanch tushun- chalari. Morfologik analiz morfemalarga ajratish hamda grammatik maiumotlar tahlilini o'z ichiga oladi. Morfologik sintez esa matnga avtomatik ishlov berish jarayonida yoki uning so'nggi bosqichida (natijalarni inson hukmiga havola qi- lishda) so'zlarning turli shakllarini chiqarish, analiz natijalarini umumlashtirish- dan iborat. Morfologik analiz va sintez jarayoni kompyuterda kodlash-dekod- lash tamoyili asosida ishlaydi. Morfologik analizning mantiqiy davomi sintaktik analiz hisoblanadi. Sintaktik analiz morfologik analiz ustiga quriladi, bunda gap strukturasi turli formal modellar asosida tahlil etiladi.

Avtomatik tarjima (mashina tarjimasi) - bir tildagi matnni ikkinchi bir tilga EHM (kompyuter) vositasida tez vaqt ichida tarjima qilish tizimi. Mashina tarji­masi bo'yicha birinchi xalqaro konferensiya 1952-yilda Massachusets texnologiya

Kompyuter lingvistikasi asoslari

143

institutida o'tkazildi. 1954-yilda AQShdagi Jorjtaun universitetida dunyoda birin- chi marta avtomatik tarjima tajribadan o‘tkazildi. U GAT tizimi (inglizcha «Geor­getown Automatic Translation» birikmasining qisqartmasi) deb ataladi. Bu tajri- balar IBM-701 kompyuterida o'tkazildi. Tajriba bazasi algoritmlardan iborat bo'lib, 50 000 ta so'z (60 ta gap)ni rus tilidan ingliz tiliga tarjima qildi.

SYSTRAN avtomatik tarjima tizimi yordamida bir yilda 2,5 mln sahifa matn ingliz tilidan nemis, fransuz, ispan, yunon, italyan tillariga hamda fransuz tilidan ingliz, ispan, italyan tillariga tarjima qilinadi. DIANA avtomatik tarjima tizimi 10 mln ta so'z bazasiga asoslanib dunyoning 6 ta yirik tilida avtomatik tarjimani ta’minlaydi. 40 mln ta so'z xotira kuchiga ega YEUROTRA sitemasi barcha yevro- pa tillari bo'yicha avtomatik tarjimani amalga oshira oladi. SOKRAT sistemasi ham 40 mln ta so'z xotira kuchiga ega bo'lib, dunyodagi 40 ta tilda avtomatik tarjimani amalga oshiradi.

Internet tizimining rivojlanishi natijasida avtomatik tarjima global tarmoqda ham tashkil etildi, ya’ni hozirgi kunda online-tarjima tizimi ham faol yo'lga qo'yilgan.

Avtomatik tahrir - kompyuter lingvistikasining asosiy yo'nalishlaridan biri bo'lib, u kompyuterdagi matn muharrirlari (Microsoft Word, Excel, Wordpad, Lexicon) rivojlanishi bilan bog'liqdir. Avtomatik tahrir matndagi mexanik xato- larning avtomatik tarzda to'g'rilanishi va xato ckanligi haqidagi signallarning userga taklif etilishi kabilar bilan bog'liq. Aniq misollar bilan tushuntiriladigan bo'lsa, Microsoft Word dasturi bilan ishlar ekanmiz, unda so'zlarning yozilishi bilan bog'liq xatolar bo'lsa, ularning ostiga qizil yoki ko'k rangda chizilishi, katta harf bilan yozilishi kerak bo'lgan so'zlarning avtomatik bosh harfga o'tkazilishi, hujjatga avtomatik numeratsiya qo'yilishi kabilar avtomatik tahrir erishgan yutuq- lar hisoblanadi.

Avtomatik tahrir yo'nalishi XX asr o'rtalariga kelib rivojlandi. U matn mu­harrir dasturlari bilan birgalikda yangicha imkoniyatlar bilan, qo'shimcha effekt- lar bilan boyimoqda. Shu o'rinda alohida ta’kidlash zarurki, avtomatik tahrirlash tizimi faqat matn muharrirlari dasturlari bazasiga qo'shilgan tillarda amal qiladi, masalan, ingliz, rus, nemis, fransuz tillarida.

Avtomatik tahrirlash tizimi juda ko'p amaliy vazifalarni o'z ichiga oladi. Bu vazifalarning eng asosiylarini umumlashtirib sanab o'tmoqchimiz:

- orfografiyani tekshirish;

- grammatika va stilistikani tekshirish;

- kontekstli qidiruv;

- formatlash;

- avtomatik referatlash - matndagi kalit so'zlar asosida uning umumiy mundarijasini yaratish;

- matn ichiga jadval, rasm, turli figuralar, klip-artlar, grafik sxemalar, dia- grammalar qo'shish, giperhavolalar bilan ta’minlash;

- matndagi abzaslar, belgilar, so'zlarning aniq miqdorini aniqlash (statistika- sini ko'rsatish);

144

A. Rahimov

- tezaurus xizmati - matndagi muayyan so'zning sinonimik variantlarini aniqlash;

- transliteratsiyalash - bir alifbodan boshqa alifboga o'tkazish (masalan, kirill alifbosidan lotin alifbosidagi harfga almashtirish).

Ms Word dasturida orfografiyani tekshirish imkoniyati cheklanganligi bois hozirgi kunda maxsus Spell Checker dasturlar yaratilmoqda. Bunday dasturlar sirasiga quyidagilarni kiritish mumkin: SpellLink Multilingual Pro (ingliz. ispan, italyan, nemis, fransuz, portugal, ukrain tillaridagi so'zlarning to'g'ri yozilishini

Spanish, Ruta 1.0 (rus va ukrain tillaridagi so‘zlarning imlosini tekshirishga yor- dam beradi), Asutype, Spellex, Spell Catcher kabi.

Bevosita ishtirokchilar (immediate constituents) - sintaktik model hisoblanadi. Bevosita ishtirokchilar metodiga muvofiq tayanch nuqta konstruksiya hisoblana­di. Konstruksiya tarkibidan ishtirokchilar va bevosita ishtirokchilar ajratiladi.103 Konstruksiya atamasi ostida ma’noli qismlarning ketma-ket munosabatidan tash­kil topgan butunlik tushuniladi. Ishtirokchilar deb kattaroq konstruksiya tarki- biga kirgan so'z yoki konstruksiyaga aytiladi. Muayyan konstruksiyaning bevos­ita shakllanishida ishtirok etgan bir yoki bir nechta ishtirokchilarga bevosita ish­tirokchilar deyiladi. Masalan, A’lochi talabalar o'z vazifalarini aniq va puxta ba- jaradilar. Bu jumla konstruksiya hisoblanadi, uning tarkibidagi barcha ma’noli so'zlar («va» bog'lovchisidan tashqari) ishtirokchilar, bevosita aloqaga kirishgan so'zlar bevosita ishirokchilar deyiladi. Bunda a’lochi so‘zi talabalar so'zi bilan, aniq, puxta so'zlari bajaradilar so‘zi bilan, o'z so'zi vazifalarini so'zi bilan o'zaro zich bog'langan. Demak, bevosita ishtirokchilar sintaktik modelda dastlab so'z birikmalari ajratiladi, so'ngra ular gap ichida birlashtiriladi.

Bilimlar injenerligi (Knowledge Engineering) - bilimlar bazasi bilan shug'ullanuvchi maxsus yo'nalish bo'lib, u sun’iy intellekt tizimining tarkibiy qismi sanaladi. Bilimlar injenerligi kognitiv lingvistika, kompyuter semantikasi va komp­yuter lingvistikasining yutuqlariga tayanadi. Bunda informatsya va bilimlar struk- turasiga oid bo'lgan semantik tarmoq, freym, senariy, Word-net kabi tushunchalar va ularning tatbiqi muhim hisoblanadi.

CALL (Computer Assisted Language Learning) - til o'qitishni avtomatlash- tirish - xorijiy tillarni kompyuter yordamida o'qitish tizimini ta’minlovchi amaliy dasturlar hamda bilimlarni avtomatik baholash mexanizmlari. Til o'rgatishda kompyuterdan foydalanish bo'yicha dastlabki tajribalar XX asrning 60-yillarida Stenford (Russian-Program) va Nyu-York universitetlarida (Das Deutche Prog­ram) o'tkazildi. XX asrning 70 - 80-yillarida til o'qitishda kompyuterdan foy- dalanishga bo'lgan qiziqish kuchaydi va hatto alohida yo'nalish - Computer As­sisted Language Learning (CALL)ning yuzaga kelishiga olib keldi. CALL bo'yicha 1-xalqaro konferensiya 1985-yil 12 - 13-noyabrda Budapesht shahrida o'tkazildi.

O'qituvchi lingvistik avtomatlardan foydalanishda 2 xil yondashuv mavjud:

1) bixevioristik yondashuv - mavjud an’anaviy o'qitish tizimining barcha jab- halarini e’tiborga oluvchi o'qitish mexanizmlari, bunda grammatikaning ma’lum

Kompyuter lingvistikasi asoslari

145

bo'limi tugagach, uni topshiriqlar asosida baholash, shuningdek, leksik mini- mumlar bo'yicha bilimni nazorat qilishga yo'naltirilgan dasturiy ta’minotga asos- langan o'qitish tizimi. Bunda bixeviorizmning «stimul-reaksiya-mustahkamlash» formulasiga asoslaniladi. O'qitish usullari «stimul» (turtki) vazifasini bajaradi, o'rganuvchi bundan ta’sirlanadi va bilimni mustahkamlaydi.

2) kognitiv-intellektual yondashuv - CALL (CALL Software) uchun universal dasturiy ta’minotlar yaratish, o'qitish tizimini lug'at bazasi, grammatika bo'yicha qo'llanmalar, tizimni avtomatik korrektorlar, speller bilan ta’minlash, audio va vizual effektlar bilan boyitish kabi amaliy ishlar majmuidir.

Ekspert tizimi - muayyan sohadagi ekspertlik yoki mavjud bilimlarning immi- tatsiyasi tarzida yaratilgan kompyuter dasturi. Boshqacharoq aytganda, ekspert tizimi maxsus sohaga oid barcha bilimlar jamlangan qomus yoki o'ta professional ekspertlarning kompyuter modelidir. Ekspert tizimlarining asosiy ikkita tarkibiy qismi mavjud: ma’lumotlar ombori va mantiqiy mashina. Ekspert tizimlari kon- sepsiyasi Stenford universiteti professori Eduard Feygenbaum nomi bilan bogiiq bo'lib, u 1977-yilda murakkab mantiqiy masalalarni yechishda kompyuter das- turlarining samaradorligi formallashtirish va dasturlash texnikasidan ko‘ra muam- moli sohaga doir bilimlar bazasiga bog'liq ekanligini asoslab berdi.

Elektron ensiklopediya - kuchli qidiruv tizimiga ega bo'lgan ma’lum yo‘nalishdagi illustrasiyali maqolalar, elektron xaritalar, fotosuratlar, giperhavo- la (gipermurojaat)lar, audio va videoizohlar, xronologik ro‘yxat, foydalanilgan va tavsiya etiladigan adabiyotlar ro'yxati va hokazolardan iborat bo'lgan katta hajm- dagi elektron dastur.

Elektron lug‘atlar - leksikografiya sohasiga kompyuter texnologiyalarining tatbiq etilishi natijasida yaratilgan lug‘atlar, ular odatdagi lug‘atlar asosida tuzil- gan, multimedia vositalari bilan boyitilgan dastur. Eng mashhur elektron lug‘atlar sirasiga CONTEXT, ABBY LINGVO, MULTITRAN, POLYGLOSSUM, MUL- TILEKS kabilar kiradi.

Elektron lug'atlaming tayyorlanishi quyidagicha: a) lug'atning so'zligi shakl- lantiriladi; b) misollar korpusi (elektron kartotekalari) tuziladi; c) lug'at maqola- lari yoziladi; d) lug'at maqolalari ma’lumotlar bazasiga ko'chiriladi; e) bevosita ma’lumotlar bazasida lug'at matni tahrir etiladi, korrektirovka qilinadi; 0 lingvis­tik ta’minot dasturiy ta’minot bilan uyg'unlashtiriladi; g) elektron lug'at.

Elektron lug'atlar an’anaviy kitobiy lug'atlarga qaraganda birmuncha afzal- liklarga ega. Elektron lug'atlaming hajm jihatidan ixchamligi, kompaktligi, boshqa manbalarga (Internet tarmog'i orqali, gipermurojaatlar yordamida) ulanish im­koniyati, mavjud matnlar korpuslari yordamida illustrativ misollar tuza olish imkoniyati, multimediali misollar bilan boyitilganligi, so'zlarning sinonimik vari­antlari, omonimlik xususiyatlari, grammatik ma’lumotlar bilan ta’minlanganligi va ularning tezlik bilan foydalanuvchiga havola etilishi kabi jihatlari elektron lug'atlaming optimalligini ta’minlaydi.

Elektron multimedia darsligi - aynan kompyuter vositasida o'quv kursini yoki uning bitta katta bo'limini o'qituvchi yordamida yoki mustaqil o'zlashtirishni

146

A. Rahimov

ta’minlovchi dasturiy-metodik kompleks bo'lib, bunda tizim matnli fayllar, turli xil animatsiyalar, audio va videoeffektlar bilan boyitilgan bo'ladi.

Example Based Translation («Namunalarga asoslanadigan tarjima») - 1984- yilda Yaponiyaning Kioto universiteti professori, mashina tarjimasi bo'yicha yapon davlat dasturi rahbari Makoto Nagao tomonidan yaratilgan yangi konsepsiya. Ushbu konsepsiyaga ko‘ra, matnlar ilgari inson qo‘li bilan tarjima qilingan matn- larga tayanib analogiya tamoyili asosida tarjima qilinishi lozim. Buning uchun kompyuterga kiritilgan katta hajmdagi bilingv bazasi (matn va uning tarjimalari) shakllantirilishi zarur. Yangi matnlar tarjimasida bilingv massividan o‘xshash matn fragmentlari tanlanadi va tarjimaning muvaffaqiyatli chiqishida foydalaniladi. M.Nagao mashina tarjimasiga bunday yondashuvni «Example Based Transla­tion» («Namunalarga asoslanadigan tarjima») deb nomladi, ilgarigi an’anaviy yon­dashuv «Rule Based Translation» («Qoidalarga asoslanadigan tarjima») deb ataladi.

Formallashtirish - bilimlar majmuasini, mazmunini muayyan shakllar yorda­mida sun’iy til belgilari orqali ifodalash. Grammatikani formallashtirish tendensi- yasi XX asr o'rtalarida kuchaydi. Bu bevosita tilning strukturasini yanada aniq parametrlarda ifodalashga intilish, uni kompyuterga moslashtirishga urinish bi­lan bog‘liqdir. Formal grammatika nazariyasi matematik mantiq, matematik lingvistikaning rivojlanishi natijasida yuzaga kelgan. Grammatikani formallash­tirish (lotincha «forma» - «shakl», «tashqi ko‘rinish») konsepsiyasining asoschisi amerikalik tilshunos Noam Chomskiydir. Formal grammatika abstrakt simvollar orqali bayon etiladigan grammatika hisoblanib, unda til qoidalari, so‘z yasalishi, birikma va gap qurilishi turli xil lingvistik modellar, struktur sxemalar yordamida tavsiflanadi. Formal grammatika 3 ta birlik asosida ish ko'radi:

1. Word-classes - so‘z turkumlari (tub va yasama so'zlar).

2. Phrases - gap bo'laklari.

3. Sentences - gap konstruksiyalari sxemalari, modellari.

Formal grammatika turlari - N.Chomskiy qoidalarning ifodalanish turiga qarab formal grammatikani 4 tipga ajratgan va bu Chomskiy iyerarxiyasi deb nomlanadi:

1. 0 grammatika (unrestricted phrase-structure grammar) -juda murakkab gene- rativ grammatika, bunda G=(N,’\ P, S) belgilari yordamida ish ko'riladi. N, “ - alifbo (N - noterminal simvollar. u grammatikadagi simvollarni o‘z ichiga oladi; “

- metatilning terminal simvollari, u tildagi so'zlarni qamrab oladi); S - noterminal to'plamning boshlang‘ich simvoli, P - qayta ishlab chiqilgan qoidalar. Bunda terminal va noterminal simvollarni birlashtiruvchi V belgisi qabul qilingan, u ing­lizcha Vocabulary (lug‘at) so'zidan olingan (V=TVUS). Bu grammatika o‘ta abs- traktlashgan matematik modellarga, informatika nazariyasiga tatbiq etilgan.

2. 1 grammatika (context-sensitive grammar) - kontekstga bog‘liq grammatika bo'lib, bunda simvollar zanjiri kontekst bilan aniqlanishi mumkin. Bu grammati­ka tabiiy tillar unsurlarining generatsiyasida faol qo'llaniladi.

3. 2 grammatika (context-free grammar) - kontekstga bog'lanmagan gramma­tika bo'lib, u programmalashtirish tillari unsurlarining generatsiyasida (ifodalar, buyruqlar tizimida) ishlatiladi.

Kompyuter lingvistikasi asoslari

147

4. 3 grammatika (regular grammar) - regular grammatika deb ataladi, u juda sodda va cheklangan grammatika bo'lib, tilning sodda unsurlari uchun ishlatiladi (miqdor, konstantalar, o'zgaruvchilar uchun).

Frazeologik mashina tarjimasi - avtomatik tarjimaning sifatini oshirish uchun Rossiyada professor G.G.Belonogov tomonidan ishlab chiqilgan konsepsiya. Unga ko'ra, matndagi tushunchalar ko'pincha alohida so'zlar yordamida emas, balki so'z birikmalari yordamida ifodalanadi. O'z-o'zidan kelib chiqadiki, matnni bir tildan boshqasiga tarjima qilish jarayonida ham mazmunni yuzaga chiqaruvchi asosiy til birliklari alohida so'zlar emas, balki tipik vaziyatlarni ifodalovchi frazeologik birikmalar (bu yerda frazeologik birlik keng ma’noda tushuniladi) hisoblanadi. Mazkur yondashuvga muvofiq matnda tez-tez uchraydigan so'zlar, iboralar, so'z birikmalari, nutqiy etiket namunalarining tarjima ekvivalentlari bi­lan boyitilgan ma’lumotlar bazasi frazeologik mashina tarjimasi tizimining asosini tashkil etadi. Bunda so'zlarning distributsiyasiga eng ko'p e’tibor beriladi. Bu metod polisemantik so'zlarning tarjimada mos keluvchi to'g'ri ekvivalentini tan- lashga, omonim so'zlarning kontekstdagi holatini aniqlashga, iboralarning ham idiomatik xususiyatini belgilashga imkon beradi.

Freym - bilimlar strukturasiga oid tayanch tushunchalardan biri. Freym - (inglizcha «frame» - «karkas», «ramka», «skelet» ma’nolarini anglatadi) sun’iy intellekt tizimida bilimlarni, stereotip (juda ko'p kuzatiladigan, tipik) vaziyatlarni gavdalantirishning bir usuli sanaladi. Ya’ni freymga modellashtirilayotgan hodi- sa, vaziyat va jarayonlarni umumlashtirilgan holda joylashtirish mumkin. Freym

- mavhum obraz modeli bo'lib, muayyan obyekt, voqea-hodisa, vaziyat va jara- yonlarning formal tavsifi demakdir (masalan, restoranga borish, shifokor qabu- lida bo'lish, ma’ruzalarga qatnashish, tug'ilgan kun, futbol o'yini kabi tipik vaziyat- lar). Dastlab «freym» termini XX asrning 70-yillarida Marvin Minskiy tomonidan olamning kompyuter modelini o'z ichiga oluvchi bilimlar strukturasiga nisbatan ishlatilgan. Shundan so'ng freym semantikasi konsepsiyasi (frame semantics) amerikalik tilshunos Charlz Fillmor tomonidan rivojlantirildi. Mazkur konsepsi- yaga ko'ra, so'zlar, so'z birikmalari, gaplar, matn ma’nosi sahnalar bilan uyg'unlashtirilgan, shartlangan bo'ladi. Bir freym boshqa bir freymning davomi yoki tarkibiy qismi bo'lishi mumkin.

Freymlar sun’iy intellekt tarkibiga kiruvchi ekspert tizimlarida faol ishlatiladi. Freymning quyidagi turlari mavjud: freym-namunalar, freym-strukturalar, freym- rollar, freym-ssenariylar, freym-vaziyatlar. Freymlar tizimi birlashib, semantik tar- moqni hosil qiladi. Freym kompyuter xotirasida bilimlar xazinasi va faktlar om- boriga tayanib olamni maxsus shakllantiruvchi model vazifasini bajaradi. Freym slot deb atalgan bir yoki bir necha elementlarning majmui hisoblanadi. Slotlarda obyektlarga xos belgilar ma’nosi qayd etilgan bo'ladi.

Freym-ssenariy - ba’zi harakat, voqea-hodisalarning xarakterli unsurlarini o'z ichiga olgan tipik strukturalar hamda vaziyatlarning rivojlanishi, ketma-ketligi (algoritm bosqichlari) demakdir. Ssenariy - bu oddiy voqea-hodisalar zanjiri emas, balki ushbu harakatlar ketma-ketligida kauzal bog'liqlik tamoyili ishlaydi, ya’ni

148

A. Rahimov

har bir harakat natijasi shartli sanaladi, uning ta’siri va oqibatida navbatdagi ish- harakat sodir bo'lishi mumkin. Har bir ssenariyda rol bajaruvchilar mavjud. Sse- nariylar real hayotda uchraydigan vaziyatlarni aks ettiruvchi holatlar asosida yuzaga keladi. Masalan, restoranga tashrif buyurish ssenariysi.

Generativ va transformatsion grammatika - formallashtirishning negizi hisob- langan grammatika, 1950-yillarda dastlab Z.Harrisning til strukturasidagi trans- formatsiya xususidagi qarashlari ta’sirida yuzaga kela boshladi, bu yo'nalish Z.Harrisning shogirdi N.Chomskiy tomonidan rivojlantirildi. Unga ko‘ra, dast­labki yadro strukturalar birikma strukturalar qoidalar tizimi orqali tug'iladi. Ya’ni har qanday til grammatikasi yadro strukturalar (kernel structures) yig'indisidir.

Generativ grammatikada yadro strukturalardan (operandlar) turli shakliy o‘zgarishlar asosida asosiy ma’noni saqlagan holda ikkilamchi sintaktik qurilma- larning hosil qilinishi (transformatsiyalar = transformalar), bu jarayondagi vositalar (operatorlar) va transformatsiya hosil qilish modellari tadqiq etiladi. Masalan, Salim keldi - Salimning kelishi - Salim kelgach - Salim kelganda - Salim kelishi bilan...; Anvar xatni yozdi - Xat Anvar tomonidan yozildi. Ushbu misollarda transfor­matsiya uchun asos bo‘lgan gap (Salim keldi yoki Anvar xatni yozdi kabi gaplar) operand, transformatsiya natijasi boigan hosila gap (Salim kelgach, Salim kelib yoki xat Anvar tomonidan yozildi kabi sintaktik strukturalar) transforma yoki transfor- mand, operanddan transformandni hosil qiluvchi vosita (ravishdosh, sifatdosh, harakat nomi yoki majhul nisbat shakllari) transformatsiya operatori deb ataladi.

Gipertekst - kompyuter lingvistikasida inqilobiy o'zgarish yasagan tex- nologiya. Gipertekst g'oyasi AQSH prezidenti F.Ruzveltning fan va ta’lim bo'yicha maslahatchisi Vannevar Bush nomi bilan bogiiq bo'lib, u foydalanuv- chiga matn va uning fragmentlarini turli assotsiativ munosabatlar asosida bog'lashga imkon beruvchi «MEMEKS» tizimini nazariy jihatdan asoslab ber- di. U 1945-yilda «The Atlantic Monthly» jurnalida «Memeks» deb atalgan kela- jakdagi xayoliy qurilma haqida «As We May Think» («Biz o‘ylaganimiz kabi») nomli maqola e’lon qildi. V.Bush o'sha qurilmani kutubxonadagi har qanday hujjatlar, yozishmalar, kitoblarni displeyda ko'rsatib bera oladigan, cheklanma- gan mikrofilmlarga ulangan elektromexanik stol sifatida tasvirlaydi. Kompyu­ter texnikasining yo'qligi bois loyihaning amalda qo'llanishi bir oz keyinga surildi. «Gipertekst» termini 1965-yil T.Nelson tomonidan ilmiy iste’molga olib kiril- gan. 1992-yilda birinchi marta gipertekst World Wide Web (xalqaro to‘r)ga tat­biq etildi.

Gipertekst bu shunday matnki, u kompyuterda boshqa matnlarga giperhavola- lar yordamida bog'lashga imkon beradi. Bunda u o'quvchiga faqat matnlar usti- da emas, balki turli xil jadvallar, sxemalar, rasmlar, videoroliklarda ham erkin navigatsiya qilish imkonini beradi. Asosan, gipertekst World Wide Web (Xalqaro o'rgimchak to'ri) strukturasini ko'rsatuvchi tushuncha bo'lib, u orqali Internet tizimida informatsiya tarqalishida qulaylik, osonlik, optimallik ta’minlanadi. In­ternet olami virtual olam sanalgani uchun biz undagi birinchi olamdan ikkinchi olamga hech qanday to'siqlarsiz o'ta olamiz. Mana shu virtuallikni bizga ta’minlab

Kompyuter lingvistikasi asoslari

149

beruvchi tizim gipertekst tizimidir. Demak, gipertekst texnologiyasi bizga infor- matsion massivni tezkor ko‘rib chiqish (brauzing), muayyan informatsiya yoki matn fragmentlariga tezkor havola qilish, gipertekst bo'yicha erkin navigatsiya qilish, gipertekstni yangi informatsiyalar bilan boyitib borish kabi optimal im- koniyatlar yaratib beradi.

Grammatika - yunoncha so'z bo'lib, «o'qish va yozish san’ati» ma’nosini bildiradi. Bu tushuncha dastlab xat-savod, imloni o'rgatuvchi fan ma’nosida ishla- tilgan. Keyinroq grammatika tilning fonologik, morfologik, sintaktik va semantik belgilarini tavsiflovchi til qoidalari majmuasi ma’nosiga ko'chgan. Grammatika haqidagi ilk qarashlar falsafa, mantiq, germenevtika (diniy matnlarni sharhlash, talqin etish) qobig'ida bo'lgan. Bunda tilning paydo bo'lishi, til va tafakkur mu- nosabati, lingvistik belgi motivatsiyasi, diniy matnlar mazmunini to'g'ri tushu- nish bilan bog'liq masalalar o'rganilgan, til strukturasi esa yetarli darajada tadqiq etilmagan. Faqatgina miloddan avvalgi II - I asrlarda shakllangan Aleksandriya grammatika maktabining eng yirik vakil Dionisiy Traks (miloddan avvalgi 170 - 90-yillarda yashagan) o'zigacha bo'lgan tadqiqotchilarning ishlari va tajribalarini o'rganib, ulardan foydalanib rimliklar uchun «Grammatika san’ati» («Gramma- tike techne») nomli sistemalashtirilgan dastlabki yunon grammatikasini yaratdi. Shundan buyon grammatika falsafadan mustaqil soha sifatida ajralib chiqdi.

Antik davrda grammatika preskriptiv (to'g'ri gapirish, to'g'ri jumla tuzishni ko'rsatuvchi yo'riqnoma) xarakterga ega bo'lgan. XVIII asr oxirlarida gramma- tikani tushunish keskin o'zgardi. Bu davrda bobo til konsepsiyasini ilgari surgan qiyosiy-tarixiy grammatika shakllandi. Unda yevropa tillarining qadimgi sanskrit tili bilan tarixiy aloqasini aniqlash yordamida hind-yevropa tillari shajarasini re- konstruksiya qilish bo'yicha qiyosiy aspektda tadqiqotlar olib borildi.

XX asr boshlarida Ferdinand de Sossyur grammatikani tavsiflashda sinxroniya va diaxroniyani farqladi. Bunda grammatikani tarixiy (diaxron) o'rganishni emas, balki sinxron (muayyan vaqtdagi holati) tadqiq qilish lozim, deb hisobladi. XX asrning 50-yillarida Noam Chomskiy grammatikani tavsiflashning yangi rnetodi- ni tavsiya etdi. Bu tilshunoslikda «formal grammatika nazariyasi» deb ataladi. Olimning tadqiqot natijalari kompyuter lingvistikasining shakllanishiga asos bo'ldi. N.Chomskiyning formal grammatika haqidagi qarashlari «Syntactic Structures» (1957) hamda «Aspects of the Theory of Syntax» nomli kitoblarida yoritilgan.

Grammatikaga bo'lgan yondashuvlar turlicha bo'lganligi bois tilshunoslik tarixida turli grammatikalar yaratilgan. Jumladan, nazariy grammatika va amaliy grammatika; an’anaviy (klassik) grammatika va noklassik grammatika; sinxron va diaxron grammatika; xususiy va umumiy (universal) grammatika; kategorial grammatika, tobelik grammatikasi, leksik-funksional grammatika, konstruktiv grammatika, transformatsion grammatika, kontekstdan xoli grammatika (con­text-free grammar) kabi.

Implikativ universaliya - Amerikalik olim Jozef Grinberg tomonidan implikatsi- yalar (mantiqiy shartli hukmlar) asosida aniqlangan universaliyalar. Olim 45 ta implikativ universaliyani aniqlagan. Masalan:

150

A.Rahimov

1. Agar til uzluksiz affikslarga ega boisa, unda yo prefiksatsiya yoki suffik- satsiya yaxshi rivojlangan bo'ladi.

2. Agar til istisnoli tarzda suffiksal boisa, u holda bu til posleloglarga, keli- shiklarga boy boiadi.

3. Agar til istisnoli tarzda prefiksal boisa, u holda bu til predloglarga boy bo'ladi.

4. Agar til VSO modelida boisa, bu tilda predloglar dominantlik qiladi.

Informatsion qidiruv tizimlari - formal belgilarga ko‘ra informatsiyani qidirish-

ga moijallangan avtomatik tizim. lQTlar quyidagi bosqichlar asosida ishlaydi: informatsiyani jamlash, informatsiyani klassifikatsiya qilish, hujjatning qidiruv obrazi (HQO)ni yaratish, hujjatning qidiruv obrazi va hujjatlarni saqlash, qidi­rish va natijani chiqarish.

«Informatsion qidiruv» (inglizcha «information retrieval») termini 1948-yilda Kelvin Mur tomonidan fanga kiritilgan. Informatsion qidiruv tizimi - bu ma’lumotlarni qisqa vaqt birligi ichida qidirib topish, kerakli hujjatlarni ma’lumotlar bazasidan topishni optimallashtirish tizimidir. Bu yo'nalish ilmiy- texnikaviy axborotlarning ko'payib ketishiga javoban 1960-yil oxirlarida yuzaga keldi. Axborot ko‘paygan sari uni saralash, keraklisini qidirib topish muammosi yuzaga keldi. Mazkur muammoni bartaraf etish uchun olimlar informatsion qidiruv tizimlari g'oyasini taklif etdilar. Bu yo'nalish Internet tarmog'i yuzaga kelgandan so‘ng yanada rivojlandi. Masalan, bugun Internetda Google, Yandex, Rambler. Yahoo kabi qidiruv tizimlari mavjud.

Informatsion qidiruv tillari - IQTda qidiruvni ta’minlash uchun yaratilgan maxsus tillar. Informatsion qidiruv tillari bu formal til boiib, u IQTda saqlangan hujjatlarning mazmunini tavsiflash va so'rov uchun moijallangan. Informatsion qidiruv tilida hujjatlarga ishlov berish protsedurasi indekslash deb ataladi. Indeks- lash natijasida har bir hujjatga IQTdagi ta’rif shakli yoziladi. Mazkur ko‘p bos- qichli jarayonda deskriptorli information qidiruv tillari ishlatiladi. Hujjat mavzusi deskriptorlarning birlashuvi bilan yoziladi. Deskriptorlar sifatida so'zlar, termin­lar, muayyan sohadagi boshlang'ich tushuncha va kategoriyalardan foydalaniladi. Deskriptorlarni kalit so'zlar deb nomlash ham mumkin. Deskriptorli til bilan ishlaydigan tizimlardan biri M.Taube tomonidan yaratilgan UNITERM tizimidir. Bu tizimda deskriptorlar sifatida hujjatning kalit so'zlari - unitermlar ishlatiladi. Demak, muayyan sohaning terminologiyasi, ya’ni deskriptorlar qidiruv tizimi va tezauruslarning asosini tashkil etadi.

Informatsion qidiruv tezauruslari (lotincha «thesaurus» - «xazina» degan ma’noni anglatadi) - matnning asosiy mazmun-mundarijasini aks ettiradigan kalit so'zlar, terminlar, tayanch tushunchalar bazasi asosida kompyuterda tuzil- gan lug'at. Tezauruslar ensiklopedik va izohli lug'atlardan farqli ravishda til birliklarining matnda qoilanish chastotasi va matnning predmet mundarijasini aks ettirish darajasiga qarab tuziladi. Bunda terminning mazmuniy tavsifi orqali bazadagi ma’lumotlar qidirilishiga moijallangan boiadi. Terminlar tezaurusga qat’iy semantik prinsiplar asosida kiritiladi, bunda ularning gipo-gepeionimik

Kompyuter lingvistikasi asoslari

151

(tur-jins), holo-meronik (butun-bo'lak), iyerarxik (pog‘onali) munosabatlari hamda assotsiativ mazmuniy bogManishlari e’tiborga olinishi lozim. Shu o'rinda ta’kidlash zarurki, keyingi paytlarda qidiruv tizimlari bazasidagi tezauruslar giperhavolalar bilan ham ta’minlanmoqda, bu esa userga birmuncha qulayliklar yaratadi. Ya’ni qidirilayotgan terminlarga aloqador bo'lgan boshqa yaqin tushunchalar bilan bog'liq ma’lumotlarni ham oson topish imkonini beradi.

Kodlash va dekodlash - informatsiyani mashina (kompyuter) tushunadigan tilga o'tkazish va bunga teskari bo'lgan kompyuter tushunadigan tildan inson tushunadigan tilga aylantirish jarayoni.

Kompyuter leksikograftyasi (elektron lug'atlar tuzish) - muayyan til leksikonida- gi so'zlarning ma’lumotlar bazasi va maxsus programmalashtirish tillari yorda­mida ishlash tizimiga ega bo'lgan dasturlar (kompyuter lug'atlari) yaratish bilan shug'ullanuvchi soha. Kompyuter leksikografiyasi bugungi kunda turli qo'shimcha imkoniyatlar asosida rivojlanib bormoqda. Xususan, dastlab kompyuter lug'atlari bir tilli, ikki tilli bo'lgan, hozirda ko'p tilli (uch, to'rt, olti, o'n tilli) elektron lug'atlar yaratilmoqda. Shuningdek, keyingi paytlarda muayyan sohalarga ixtisoslashgan elektron lug'atlar ham yaratilmoqda. Ya’ni dastlab kompyuter lug'atlari faqat umumiy leksikon bilangina cheklangan edi, hozirda fanga oid bo'lgan, qurilish, aviatsiya, avtomobilsozlik, harbiy, diniy, yuridik sohalarga oid so'zlar bazasi ham e’tiborga olinmoqda. Hatto so'zlarning matnda qo'llanilish imkoniyatlari, dist- ributiv holatlari, birikma holidagi ko'rinishlari ham nazarda tutilgan holda lug'atlar yaratilmoqda.

Kompyuter lingvistikasi - matematik lingvistikaning mantiqiy davomi bo'lib, u amaliy tilshunoslikning eng muhim qismini tashkil etadi. Kompyuter lingvistikasi 1954-yil AQSHda Jorjtaun universitetida mashina tarjimasi bo'yicha dunyoda o'tkazilgan birinchi tajriba asnosida yo'nalish sifatida shakllana boshladi, 1960- yilga kelib mustaqil fan sifatida shakllandi. Kompyuter lingvistikasi inglizcha «com­putational linguistics» so'zining kalkasidir. XX asrning 80-yillariga qadar bu fan turlicha nomlar bilan atalgan: hisoblash lingvistikasi, matematik lingvistika, kvanti­tativ lingvistika, injener lingvistikasi kabi. Mazkur fanning asosiy maqsadi lingvistik masalalarni yechishning kompyuter dasturlarini ishlab chiqish, inson va mashina (kompyuter) muloqotini optimallashtirish, tabiiy tilni qayta ishlash (Natural Lan­guage Processing)dir.

Kompyuterli modellashtirish - bugungi kunda barcha fanlarda, xususan, komp­yuter lingvistikasida ham samarali metod hisoblanadi. Kompyuter modellashti­rish quyidagi asosiy bosqichlardan iborat:

- masalaning qo'yilishi, modellashtirish obyektining aniqlashtirilishi;

- konseptual (tushunchaviy, fikriy) modelning ishlab chiqilishi, tizim asosiy unsurlarining ajratib olinishi;

- formalizatsiya, ya’ni matematik model bosqichi; algoritm yaratilishi hamda dastur ishlanishi;

- kompyuter eksperimentlarini o'tkazish;

- natijalar tahlili va talqini.

152

A. Rahimov

Sun’iy intellekt tizimi doirasidagi tabiiy tilli interfeys, ekspert tizimlari, neyron tarmoqlar, lingvoanalizatorlar, gapiruvchi avtomatlar - barchasi kompyuter model- lashtirish natijasi hisoblanadi.

Kontekstga bog‘Iiq bo‘)magan grammatika (Context-free grammar) - matn- dan xoli bo'lgan grammatika, chunki undagi qoidalar universal hisoblanib, ma’lum matnga tayanib tahlil etishga asoslanmaydi, balki u avvaldan empirik tarzda belgilab olinadi. Hozirda ushbu grammatika sun’iy programmalash tillari- ga (Pascal, C++, Delphi, HTML) faol tatbiq etilgan, shuningdek, tabiiy tillarni formallashtirish jarayonida ham grammatikani bayon qilishning eng optimal usuli hisoblanadi.

Lemmatizatsiya - so'zning dastlabki, boshlang‘ich formasini (lug'atdagi shak- lini - lemmasini) tashkillashtirish texnikasi bo'lib, bu jarayon o'sha so'zning boshqa so'z-shakllaridan kelib chiqqan holda amalga oshiriladi. Lemmatizatsiya mor­fologik tahlil metodi tarkibiga kiradi, u ikki bosqichni o'z ichiga oladi: 1) dekla­rativ bosqich - bunda muayyan so'zning mumkin bo'lgan barcha shakllari (so'z- shakllar) belgilanadi; 2) protsedura bosqichi - bunda so'z asos va qo'shimchalarga, ya’ni leksemalarga yoki morfemalarga bo'linadi. Lemmatizatsiya so'zlarning gram­matik valentligi, qaysi affikslar bilan birika olish imkoniyatini ham belgilab bera­di. Masalan, o'zbek tilida so'zlarning lug'atdagi shakli - lemmasi quyidagicha:

- ot so'z turkumi uchun - bosh kelishik, birlik shakli;

-- fe’l so'z turkumi uchun - harakat nomi shakli;

- sifat so'z turkumi uchun - oddiy daraja shakli

Daftarlarni, daftarlarga, daftarlarning, daftarlardan, daftarlarda ! daftar

Yugurdi, yugurgan, yuguryapti, yugurmoqchi ! yugurmoq

Ko'kimtir, ko'kish, ko'kroq ! ko'k

Lingvistik protsessorlar - tabiiy tilda kompyuter bilan muloqot qilishni ta’minlashga yordam beradigan dasturlar. XX asrning 70-yillarida AQSHda komp­yuter bilan cheklangan muloqotni ta’minlaydigan tabiiy tilli interfeys tizimi (LI­FER - Language Interface Facility with Ellipsis and Recursion) yaratildi. 1985- yilda «Semantek» korporatsiyasi tomonidan Q&A dasturi, «Karnegi gruppa» kompaniyasi tomonidan Language Craft dasturiy ta’minoti ishlab chiqildi. Bu turdagi dasturlar hozirda ingliz tilining standartlashgan so'zlashuv uslubida un- cha katta bo'lmagan lug'at zaxirasi bilan ishlamoqda.

Lingvistik universaliya - barcha tillarga yoki dunyo tillarining mutlaq ko'pchilik qismiga xos umumiy qonuniyatlar. 1941-yilda Aginskiy tomonidan «The Impor­tance of language tiniversals» sarlavhali maqola e’lon qilindi. 1961-yilda Nyu-York shahrida universaliyalarga bag'ishlangan xalqaro konferensiya o'tkazildi. Unda J.Grinberg, J.Jenkins, Ch.Osgudlar «Lingvistik universaliyalar haqida memoran­dum» mavzusida chiqish qildilar. Konferensiyada Ch.Osgud shunday yozadi: «Biz ushbu konferensiyada qonsi: inqilobning guvohi bo'lamiz».

Fonologiya, morfologiya, sintaksis, semantika sohalarida aniqlangan lingvis­tik universaliyalar 2000-yilga kelib 1600 tani tashkil etdi. Germaniyaning Kons­tanz universitetida lingvistik universaliyalar bazasi (Universals Archive) yaratildi

Kompyuter lingvistikasi asoslari

153

va u har yili yangi ma’lumotlar bilan boyitib boriladi. Mazkur baza Internetda

Matematik lingvistika - XX asrning 50-yillarida (1952-yilda) yuzaga kelgan amaliy tilshunoslikning alohida yo‘nalishi bo‘lib, unda tabiiy tillarning matema­tik modellarini (bunday formallashgan til metatil deb ataladi) ishlab chiqish, xusu- san, sun’iy tillarni yaratish algoritmini tuzish bilan shug'ullanuvchi fandir. Bu fanning shakllanishida Kopengagen struktural tilshunoslik maktabi (glossematika)ning asoschisi Lui Yelmslevning g‘oyalari o‘ziga xos «turtki» vazi- fasini o'tagan. U hatto til hosidalarini matematik bayonda tushuntiradigan fan­ning nomini ham taklif etgan. Olimning fikricha, bu fan «til algebrasi» («lingvistik algebra») deb atalishi lozim edi. Amerikalik tilshunos Noam Chomskiyning for­mal grammatika, transformatsion grammatika haqidagi qarashlari bevosita mate­matik lingvistikaning alohida yo'nalish sifatida yuzaga kelishiga sabab bo‘lgan.

Matnlar korpusi («corpus» - lotincha «tana» degan ma’noni anglatadi) - elektron holda saqlanadigan ma’lum til birliklari bo'lib, ular tilshunoslar uchun turli xil muammolarni hal etish uchun tatbiq etishda va turli yo‘nalishdagi tadqiqotlar uchun zaruriyatga qarab turli shakllarda tuziladi. Til birliklarining qay tarzda saqlanishiga qarab maxsus dasturlar yordamida har bir kerakli so‘z yoki so‘z birikma^i uchun darhol uning qo'llanishi bo‘yicha misollar topilishi, imlo bo‘yicha variantlari, sinonimik qatorlari topilishi mumkin. Matnlar korpusiga oid ilmiy tadqiqotlar salmog‘ining ko‘payishi natijasida tilshunoslikda korpus lingvistikasi yo‘nalishi shakllandi.

Kompyuterda yaratilgan birinchi matnlar korpusi Braun korpusi (БК, ing- lizcha Brown Corpus, BC) hisoblanadi, u 1961-yilda Braun universitetida yaratil­gan, har biri 2000 so‘zli 500 ta matn fragmentini o'z ichiga oladi. 1970-yillarda 1 mln so‘zni o‘z ichiga olgan matnlar korpusi asosida rus tilining chastotali lug'ati yaratildi. 1980-yillarda Shvetsiyaning Upsala universitetida ham rus tilida matnlar korpusi yaratildi. Buyuk Britaniyada Ingliz tili Banki (Bank of English) hamda Britaniya Milliy Korpusi (British National Corpus, BNC), Rossiyada Rus tilining mashina fondi (Машинный фонд русского языка) hamda Rus tilining Milliy korpu­si (Национальный корпус русского языка) loyihalari ishlab chiqildi. Masalan, Rus tilining Milliy korpusi hajmi hozirgi kunda 149 mln so'zdan iborat. Keyingi yillarda Internet tizimining rivojlanishi virtual matnlar korpusining yuzaga keli­shiga olib keldi. Ya’ni lnternetdagi qidiriv saytlari, elektron kutubxonalar, virtual ensiklopediyalar korpus vazifasini bajarmoqda. Korpusning janri va tematik rang- barangligi Internetdan foydalanuvchining qiziqishlariga bog'liq. Masalan, ilm- fan doirasida Wikipedia katta hajmdagi matnlar korpusi sifatida ishlatilmoqda.

Matnning statistik tahlili - tilni ehtimollik belgilariga ko'ra tahlil qilish, til hodi- salarini statistik xarakteristikalari asosida baholash demakdir. Til birligining gap- da yoki matnda qo'llanish chastotasi (sur’ati) ehtimollik deyiladi. Til hodisalari haqidagi haqiqatlar lingvostatistik tadqiqotlar yordamida aniqlanadi. Masalan, o'zbek tili uchun qaysi bo'g'in strukturasi tipik ekanligini, gap modellaridan qay- si biri eng asosiy konstruksiya ekanligini, bosh kelishikdagi otlarning asosan qay-

154

A. Rahimov

si gap bo'lagi vazifasida kelishini, tanlangan matnda so'z turkumlarining qay darajada ishlatilishini aniqlash uchun lingvostatistik tahlillarga ehtiyoj seziladi.

Matnning statistik tahlili matn atributsiyasi muammosiga ham oydinlik kiri- tishda muhim tadqiqot usuli hisoblanadi. Ayniqsa, badiiy matnlarning statistik tahlili natijasida o‘sha matn muallifi ham aniqlanmoqda. Bunda matnda tez-tez ishlatiluvchi til birliklari (otlar, sifatlar, kalit so'zlar, fe’llar, grammatik shakllar, junila qurilishi, bir so'z bilan aytganda, yozuvchining idiostilini - o'ziga xos uslu- bini ko'rsatuvchi vositalar) lingvostatistik tahlil yordamida aniqlanadi. Turli matn- lardan olingan dalillarning qiyosiy tahlili bizga o‘sha matnning mazmun-mundari- jasini, matn tuzilgan davrni, dalillarni argumentlash xarakterini va hatto muallif- likni aniqlashga imkon beradi. Badiiy asarlarning sujet qurilishiga, matn struk- turasiga kvantitativ metodlarning tatbiq etilishi natijasida F.Dostoyevskiy, L.Tolstoy. M.Sholoxov kabi yozuvchilarning matnlari atributsiyasi amalga oshiril­gan. Keyingi paytlarda anonim (muallifi noma’Ium) va psevdonim (soxta muallif- li) matnlarning aniqlanishi yuzasidan ham tadqiqotlar amalga oshirilmoqda.

Mashina ijodi - san’at asarlarini avtomatik tarzda yaratish jarayoni. Bugungi kunda ijodni to‘la modellashtirishga erishilgani yo‘q, lekin bu sohada sezilarli yutuqlar qo'lga kiritilgan. Jumladan, kompyuter yordamida musiqalar yaratili- shi, treker-dasturlar, sintezatorlar, sekvensorlar yordamida bir turdagi melodiya- lar va tovushlar generatsiyasi oqibatida elektron musiqalarning yaratilishi, matn­ning avtomatik generatsiyasi yordamida she’rlar, ertaklar - umuman matnlar yaratilishi, rasmlar generatsiyasi va kompyuter grafikasi yordamida kino va o'yinlarning yaratilayotganligi (ayniqsa, strategiya va harbiy manyovrlarga asos­langan kompyuter o'yinlari) kabilar mashina ijodiga misol bo‘la oladi.

Metagrammatika - yunoncha «metha» - «keyin», «grammatike» - «o'qish va yozish san’ati», metaqoidalarga asoslangan grammatika; tavsifiy grammatika, formal grammatika, transformatsion grammatika bosqichlaridan so‘ng yuzaga kelgan, qayta yozilgan grammatika; programmalash tillarini tavsiflash uchun tabiiy til grammatikasi modeli asosida ishlab chiqilgan metaqoidalar to'plami. Lingvis­tik universaliyalar g'oyasi ta’sirida metagrammatika haqida qarashlar, nazariy va amaliy tadqiqotlar yuzaga keldi. Metagrammatika haqidagi dastlabki g‘oya!ar Dekart, Leybnis kabi matematiklar tomonidan ilgari surilgan ratsional (univer­sal) grammatika yaratish, Klod Lanslo va Antoniy Arnoldlarning (1660, Parij) «Por-Royal grammatikasi» bilan bog'liqdir. Algol 68 uchun ishlangan metaqoi­dalar Winjgaarden-grammar yoki W-grammar deb nomlangan. Keyinroq 1981- yilda olimlar Gazdar, Pullman, Saglar ham metaqoidalar ishlab chiqdi, u Genera­lized Phrase Structure Grammar (GPSG) deb ataladi.

Model (lotincha «modelus» so'zidan olingan bo‘lib, «nusxa», «andaza», «o'lchov», «me’yor» ma’nolarini anglatadi) - fanda shunday moddiy qurilma, grafik, sxema, umuman, bilish vositasi sifatida tushuniladiki, u muayyan original- obyekt haqidagi ma’lumotlar majmui tarzida yuzaga kelgan hosila-obyekt de- makdir. Boshqacharoq tushuntirganda, model tabiiy obyektlarning imitatsiya- sidir (o'xshashi, taqlidiy ko'rinishi), u o'zbek tilidagi qolip, andaza so'zlariga mos

Kompyuter lingvistikasi asoslari

155

keladi. Masalan, globus yerning modeli, o'yinchoq mashina haqiqiy mashinaning modeli sifatida baholanishi mumkin.

Modellarni shartli ravisda quyidagicha tasniflash mumkin: 1) tabiiy modellar - o'rganilayotgan obyekt bilan bir turda bo'ladi va undan faqat o‘lchamlari, jarayonlarining tezligi va ba’zi hollarda yasalgan materiali bilan farq qiladi; 2) matematik modellar - prototipdan (asl nusxadan) jismoniy tuzilishi bilan farq qiladi, lekin u bilan bir xil matematik tasvirga ega bo'ladi; 3) mantiqiy-matematik modellar - belgilardan iborat bo'lib, abstrakt model hisoblanadi va tafakkur jara- yonini o‘rganishda qo'llaniladi; 4) kompyuterli modellar - matematik va mantiqiy modellashtirish metodlariga asoslangan holda kompyuterda algoritm va dasturlar- dan foydalanib yaratiladigan modellar.

Modellashtirish - barcha fanlar uchun xos bo'lgan umumilmiy metod hisobla­nadi va u quyidagi tamoyillarga amal qiladi:

- deduktivlik - mantiqiy xulosa chiqarishga asoslangan, xususiylikdan umumiy- likka tamoyilida bo'ladi;

- tafakkur eksperimentidan foydalanish;

- evristik funksiyaga ega bo'lishi. Ya’ni u yangi g'oyalar bera olishi va uni amaliyotda sinab ko'rish imkoniyati bo'lishi zarur;

- eksplanatorlik xususiyatiga, ya’ni tushuntirish kuchiga ega bo'lishi kerak. Shundagina model eski nazariya tushuntirib bera olmagan muammoni hal qiladi, obyektning ilgari kuzatilmagan, ammo kelajakda amalga oshishi mumkin bo'lgan tomonini kashf etadi;

- modelni ideallashtirilgan obyekt sifatida talqin qilish.

Modellashtirish jarayoni uchta unsurni o'z ichiga oladi:

- subyekt (tadqiqotchi);

- tadqiqot obyekti;

- o'rganuvchi subyekt va o'rganiluvchi obyekt munosabatini aks ettiruvchi model.

Modellashtirish tilshunoslikka strukturalizm yo'nalishi ta’sirida faol tatbiq qilina boshlandi. Gap strukturasini modellashtirish g'oyasi XX asrning 50-yillari- da amerikalik tilshunos Charlz Friz tomonidan olg'a surildi (1952-yilda «The Structure of English» nomli tadqiqotini yaratdi). Ch.Friz o'z konsepsiyasini dist­ributiv model deb nomlagan. Unga ko'ra, gap muayyan so'z turkumlariga oid bo'lgan so'zlar zanjiri hisoblanadi va tahlilda morfologiya bazasiga tayaniladi. Masalan, «The young man painted the door yesterday» jumlasi distributiv model asosida quyidagicha tahlil etiladi: D 3 Iя 2-d D Ib 4. Bu yerda D - otning aniqlovchisi (inglizcha determiner), 3 - sifat, 1 - birlikdagi ot, 2-d - o'tgan zamon shaklidagi fe’l, 4 - ravishni anglatadi. Demak, bu modelda turli so'z turkumlariga mansub so'z shakllarining nutq zanjiridagi distributsiyasi (tarqalishi, qurshovi) gap struktura­sini modellashtirishning asosiy mezoni sanaladi.

Neyron tarmoqlar - inson miyasidagi neyronlar modeli asosida yaratilgan ti­zim bo'lib, u sun’iy intellektning muhim tarkibiy qismi hisoblanadi. Hozirgi kun- da neyron tarmoqlar bir qator amaliy masalalarni hal etishda keng qo'llanilmoqda.

156

A. Rahimov

Jumladan, ob-havo ma’lumotini aniqlashda, obrazlarni tanishda (xususan, og'zaki va yoztna nutqni aniqlashda), robotlarni boshqarishda, tibbiyotda tashxis qo'yishda ishlatilmoda. Neyron tarmoqlarining ishlash tizimi matematik ifodalar yoki mantiqiy qoidalarga asoslanmaydi, balki namuna (example based approach) asosida o‘qitishga asoslangan. Bunda muayyan masalani yechish uchun namu- nalar (misollar), ya’ni o'qituvchi tanlamalar tayyorlanadi va tarmoq shu asosda tekshiradi.

On-line lug'atlar - ochiq turdagi Internet tizimida ishlaydigan lug'atlar. Bu lug'atlarning aktivlashishi uchun albatta Internetga ulangan bo'lishimiz lozim. Ularning afzalligi shundaki, foydalanuvchilarga ushbu lug'atlar bazasini to'ldirib borish, boyitish imkonini beradi.

Parallel matnlar korpusi - badiiy asar, qo'llanma, ommaviy axborot vositasi nashrlari, turli xil hujjatlarning ikki yoki undan ko'p tillardagi elektron holdagi ko'rinishlari. Masalan, Yevropa Ittifoqi o'zining barcha qonun va hujjatlarini ingliz, fransuz. nemis, ispan va italyan tillarida nashr qiladi hamda ular Internet tizimiga barchaga ochiq arxiv sifatida qo'yiladi. Bunday korpuslarning afzalligi shundaki, ular yordamida nafaqat biron bir so'z yoki jumlaning, balki butun boshli matnlarning turli tillardagi variantlarini bilish imkoniyati mavjud. Xuddi mana shu imkoniyat tufayli maxsus konkordanser dasturlar ishlab chiqish orqali turli xil ixtisoslik lug'atlari tuzish imkoniyati tug'iladi. Ushbu imkoniyatlar komp­yuter leksikografiyasi uchun ulkan ahamiyat kasb etadi.

Semiotika yoki semiologiya (yunoncha «semiotike», ya’ni «belgilash», «tas- virlash», «ifodalash» degan ma’nolarni anglatadi) belgilar sistemasi haqidagi fan bo'lib, unda axborotni saqlovchi, uzatuvchi turli xil belgi va belgilar tizimlarining yaratilishi, vazifalarini, belgilar sistemasiga kiruvchi hodisalarni, ularning tabiati, mohiyati, turlari, ko'rinishlari, amal qilishi kabilarni tekshiradi. Semiotika tilshu- noslikda, xususan, kompyuter lingvistikasida ham muhim metodologik ahamiyat kasb etadi. Chunki til ham belgi sanaladi, semiotikada tabiiy tillar ham, sun’iy tillar ham, hatto yo'l belgilari, audio va videosignallar ham tadqiq etiladi. Semioti­ka fani XIX asr oxirlarida vujudga keldi, uning asoschisi amerikalik olim Charlz Sanders Pirs (1839 - 1914) hisoblanadi. Semiotikada belgilar tizimi to'rt aspekt asosida o'rganiladi: 1) sintaktika - belgilar strukturasini tekshiradi; 2) semantika - belgilarning mazmuniy tomonini tekshiradi; 3) sigmatika - belgi va uni aks ettiruv- chi obyekt o'rtasidagi munosabatni o'rganadi; 4) pragmatika - belgilarning inson- ga bo'lgan munosabati, foydaliligi, amaliy jihatlarini tadqiq etadi.

Semantik tarmoq - borliqni so'zlar va ularning semantik munosabatlari yor­damida modellashtirishga yo'naltirilgan tizim hisoblanadi. Semantik tarmoq aniq belgilangan uzellar (obyektlar kesishadigan joy) va dugalar (semantik munosa- batlarning biri ikkinchisiga ta’sir ko'rsatadigan nuqta) munosabatdorligi asosida modellashtiriladi. Semantik tarmoq g'oyasi kompyuter texnologiyalaridagi lokal va global tarmoq nazariyasi ta’sirida kelib chiqqan. Uzellar muayyan obyektlar- ga, muayyan sohaga oid tushunchalarga, dugalar obyektlar o'rtasidagi semantik munosabatlarga aloqador tushunchalardir.

Kompyuter lingvistikasi asoslari

157

Sun’iy intellekt - intellektual mashinalar, intellektual kompyuter dasturlar yara­tish texnologiyasi va bu haqidagi fan. Intellektual sistema strukturasi 3 asosiy blokni o‘z ichiga oladi: bilimlar bazasi, masalani yechuvchi tizim va intellektual interfeys. Sun’iy intellekt insondagi intellekt tushunchasining aynan immitatsiyasi emas, balki inson intellektual faoliyatining ko'plab qirralarini o‘z ichiga olgan, modellashtirilgan tizimlar texnologiyasi demakdir. Ayni shu ma’noda «artificial intelligence» terminini birinchi marta Jon Makkarti 1956-vilda Darmut univer- sitetidagi konferensiyada qoiladi.

Tobelik daraxti (shajara) - L.Tenyer g‘oyalari ta’sirida shakllangan tobelik grammatikasi modeli bo'lib, u kompyuter lingvistikasining avtomatik analiz das- turlarida ishlatiladi. Bunda gap asosan daraxt ko'rinishida tasvirlanadi, jumla tarkibidagi so'zlar bir-biri bilan tobelik munosabatida bo'ladi (bevosita dominatsiya munosabati). Kesim daraxtning ildizida joylashadi va dominant bo‘lak hisoblana­di, boshqa so'zlar (gap boiaklari) unga bevosita yoki bilvosita (boshqa so'zlar yordamida) tobelangan bo'ladi.

Translation Memory («Tarjimon xotirasi») yoki Sentence Memory («Gap to'plagich») - avtomatik tarjima samaradorligini oshirish maqsadida amerikalik olim Veb Layn tomonidan ishlab chiqilgan konsepsiya. Bu yondashuvga ko'ra, muayyan matnni bir tildan ikkinchisiga tarjima qilish uchun dastlab u yuqori malakali tarjimonga ko'rsatiladi. So'ng asl matn va uning tarjimasi kompyuterga kiritiladi, matn fragmentlarga (alohida jumlalarga, birikmalarga) bo'linadi, ush­bu elementlardan ma’lumotlar bazasi tayyorlanadi. Ma’lumotlar bazasi qidiruv tizimiga yuklanadi, u esa tarjimasi mavjud bo'lgan matn fragmentlarini va alohida jumlalarni sifatli tarjima qilishga imkon beradi. Matnning ayrim notanish frag- mentlari (bazada tarjimasi mavjud bo'lmagan matn qismlari) inson qo'li bilan tarjima qilinadi. Bunda bazadagi bu fragmentlarga yaqin keladigan tarjimalardan foydalanish mumkin. Matnning inson tomonidan tarjima qilingan qismlari yana ma’lumotlar bazasiga kiritiladi. Mana shu yo‘l bilan «tarjimon xotirasi» asta- sekinlik bilan boyitib boriladi va tarjima samaradorligi yuksalib boradi. Lekin ushbu yo'nalishning afzalliklari bilan birga ayrim kamchiliklari ham mavjud. Birin- chidan, bu yo‘1 bilan har qanday turdagi matnni tarjima qilib bo'lmaydi. U aso­san muayyan tor yo'nalishdagi bir turdagi matnlarga mo'ljallangan. Ikkinchidan, ma’lum matnning katta fragmenti ikkinchi bir matnning katta qismiga tarjimada mos kelishi yoki yaqin kelishi har doim ham kuzatilavermaydi. Uchinchidan, «tar­jimon xotirasi» bazasini yaratish yoki ikki tilli matnlar (bilingv) massivini to'ldirib borish amaliy jihatdan ko'p qiyinchiliklarni keltirib chiqaradi.

«Tyuring testi» - «Mashina fikrlay oladimi» (1950) sarlavhali maqolaning muallifi ingliz olimi Alan Tyuring tomonidan ishlab chiqilgan empirik test. Olim mashina imkoniyatini insonning aqllilik darajasi bilan qiyoslash prolsedurasini ishlab chiqdi. Ushbu g'oya 1950-yilda «Mind» falsafiy jurnalida «Hisoblash mashi- nalari va tafakkur» (Computing Machinery and Intelligence) nomli maqolada e’lon qilindi. Bu testdan ko'zlangan maqsad sun’iy tafakkurning inson tafakkuriga yaqin kelishi mumkinligini aniqlash hisoblanadi. Unga ko'ra, inson bir kompyu-

158

A. Rahimov

ter va bir inson bilan o‘zaro munosabatga kirishadi. Savollarga javob berish asosi­da u kirn bilan (inson yoki kompyuter dasturi bilanmi) gaplashayotganini aniqlashi lozim bo'ladi. Kompyuter dasturining vazifasi esa insonni xatoga yo‘1 qo'yishga, noto‘g‘ri tanlov qabul qilishga majbur etish sanaladi.

Zif qonuni — matnning statistik tahlilida keng foydalaniladi, ayniqsa, katta hajmdagi matnlarda har bir so‘zning qanchalik tez takrorlanishini (chastotasini) va ularning takrorlanish tezligini hisoblashda Zif qonuni amal qiladi. Bunda so'zning chastotasi (f), uning ro'yxatdagi holati qatori (r) deb belgilanadi. Zif qonuniga ko'ra, matnda eng ko'p ishlatilgan so‘z r = 1, undan pastroq chastotada kuzatil- gan so‘z r = 2, undan keyingisi r = 3 tarzida belgilanadi. Bundan kelib chiqadiki, matndagi ixtiyoriy so'z uchun (0 chastotaga (r)ning ko'paytmasi natijasi o'zgarmas miqdor - konstanta (C)ga teng boiadi (r4f = C), bu yerda С matn hajmiga bog'liq.

Chastotali lug‘atlar - matnda so'zlarning ishlatilish chastotasi (qay darajada ishlatilishi, tez-tez ishlatilishi)dan kelib chiqqan holda tayyorlanadigan so'zliklar. Chastotali lug'atlar statistik leksikografiyaning rivoji natijasida yuzaga kelgan. Bunday lug'atlar, ayniqsa, til o'qitish jarayonida leksik minimumlarni aniqlashda muhim ahamiyatga ega. So'zlarning chastotasini aniqlash linvostatistik tahlillar yordamida amalga oshiriladi. Chastotali lug'atlar kompyuterning lingvistik in­formatsion bazasi hisoblanadi.

MUNDARIJA

So'zboshi 3

Kompyuter lingvistikasi fanining mazmun-mohiyati 4

Mantiq, grammatika va algoritm 10

Aksiomatik nazariya - tilni formal tavsiflash asosi 23

Kompyuter lingvistikasida modellashtirish va analogiya metodlari 29

Formal grammatika nazariyasi 37

Kompyuter lingvistikasida matematik

statistika va kvantitativ metod 48

Kompyuter lingvistikasi tarixi va zamonaviy holati.

Kompyuter lingvistikasi bo'yicha tadqiqotlar 57

Kompyuter lingvistikasida avtomatik

tarjima yo‘nalishi 72

Kompyuter leksikografiyasi 85

Til o'rgatish jarayonini avtomatlashtirish 92

Kompyuter lingvistikasida avtomatik tahrir,

informatsion qidiruv tizimlari va gipertekst texnologiyasi 98

Kompyuter analiz va sintezi. Ingliz tilidagi gaplarni

o‘zbek tiliga tarjima qilish algoritmi 110

Kompyuter lingvistikasi fanining

rivojlanish istiqbollari 123

[Foydalanilgan adabiyotlar ro'yxati 138](#bookmark12)

[Kompyuter lingvistikasi terminlari lug'ati (Glossariy) 142](#bookmark13)

Ilmiy tiashr

RAHIMOV AZAMAT JON USMAN JANOVICH

KOMPYUTER LINGVISTIKASI ASOSLARI

Muharrir: Abdulla SHAROPOV Musahhih: Xurshid IBROHIMOV Badiiy muharrir: Feruza NAZAROVA Texnik muharrir: Dilshod NAZAROV

Nashriyot litsenziyasi: AI №134, 27.04.2009.

Terishga 01.01.2011-yilda berildi. Bosishga ruxsat etildi: 09.02.2011 -y. Ofset qog'ozi. Qog'oz bichimi: 60x84 7l6. LexTimes garniturasi. Ofset bosma. Hisob-nashriyot t.: 8,5. Shartli b.t.: 9,3. Adadi: 1000 nusxa. Buyurtma № 42

«AKADEMNASHR» nashriyotida nashrga tayyorlandi. 100156, Toshkent shahri, Chilonzor tumani, 20A-mavze, 42-uy. Tel: (+998 97) 331-56-22 e-mail: akademnashr@mail.ru

“KO‘HI NUR” MCHJ bosmaxonasida chop etildi. Toshkent shahri, Mashinasozlar mavzesi, 4-uy.