



Azamatjon RAHIMOV

**KOMPYUTER
LINGVISTIKASI
ASOSLARI**



UDK: 519.711

32.973

R33

Rahimov A. Kompyuter lingvistikasi asoslari / A. Rahimov; mas'ul muharrir A. Nurmonov; O'zR Oliy va o'rta maxsus ta'lim vazirligi, Andijon davlat universiteti. – T.: Akademnashr, 2011. – 160 b.

BBK 32.973я73+81.2-4я73

Mazkur qo'llanma oliy o'quv yurtlarining 5220100-filologiya (o'zbek filologiyasi) ta'lim yo'nalishi 1-bosqich bakalavr talabalari, magistrantlar, tadqiqotchilar va o'qituvchilarga mo'ljallangan. Unda 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 (узбекская филология), магистрантов, исследователей и преподавателей ВУЗов. В этой книге обсуждены и интерпретированы основные вопросы компьютерной лингвистики, а также в конце книги приложен словарь (гlossарий) ключевых терминов по компьютерной лингвистики.

This manual (handbook) is referred to the first year bachelor-students (5220100-philology (Uzbek philology), masters, researchers and teachers 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.

SO'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 bo'lmoqda. 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 lingvistika, kvantitativ lingvistika, injener lingvistikasi* kabi. O'tgan yarim asrlik vaqt mobaynida kompyuter lingvistikasi sohasida bir qator ilmiy va amaliy natijalarga 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 muammolarini 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 bo'lsa-da, lekin sanalgan masalalar hali to'liq o'z yechimini topgani yo'q. Chunki inson tili va tafakkuri sirlari 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, nemis, fransuz, ispan tillarida yaratilgan monografiyalar, ilmiy jurnallar, darsliklar miqdori ko'p. O'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 masalalari tahlilga tortildi. Muallif ushbu kitobni yozishda D.Jurafskiy, J.H.Martinlarning «*Speech and Language Processing*», R.Grishmanning «*Computational linguistics*», «*Новое в зарубежной лингвистике*» jurnalining 24-soni (Компьютерная лингвистика), G.G.Belonogovning «*Компьютерная лингвистика и перспективные информационные технологии*», Yu.I.Shemakinning «*Начало компьютерной лингвистики*», A.Nurmonovning «*Структур tilshunoslik: ildizlari va yo'nalishlari*», A.Po'latov, 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

KOMPYUTER LINGVISTIKASI FANINING MAZMUN-MOHIIYATI

Reja:

1. Kompyuter lingvistikasi fanining yuzaga kelish asoslari.
2. Kompyuter lingvistikasi va matematik lingvistika.
3. Kompyuter lingvistikasi fanining tekshirish obyekti, maqsad va vazifalari.
4. Kompyuter lingvistikasi va klassik tilshunoslik.

Til sistema sifatida aynan matematikaning o'zidir. Matematika aniq fan bo'lib, formulalar munosabati orqali reallashadi. Til ham matematika singari aniqlikni talab qiladi va unda ikki yoki undan ortiq elementlarning o'zaro munosabati bir butunlikni tashkil etadi.

Beryozin

Tayanch so'z va iboralar: *til algebrasi, matematik lingvistika, kompyuter lingvistikasi, mashina tarjimai, metatil, sun'iy til, deskriptiv, kvantitativ, muloqotning optimal modeli, analiz, sintez, NLP, injener lingvistikasi, generatsiya, tilshunoslik va adabiyotshunoslik sohalarida kompyuterdan unumli foydalanish, tillarni o'qitish, bilimlarini baholash, matnlarni tahrir qilish, tabiiy tillarning matematik modeli, o'zbek tilining kompyuter uslubi, matnlar 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 rivojlanmoqda. Jumladan, ana shunday fanlar sirasiga sotsiolingvistika (sotsiologiya va tilshunoslik), psixolingvistika (psixologiya va tilshunoslik), etnolingvistika (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 o'zaro hamkorligi, integratsiyasi deb baholash lozim bo'ladi. XX asrning 50-yillaridan boshlab tilshunoslikda

«**mashina tarjimasi**», «**mashina tilshunosligi**» atamaları qo'llanila boshlandi.¹ Mazkur asrning buyuk kashfiyoti bo'lgan kompyuter texnologiyalari tilshunoslikka ham kirib kelganining 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 tilshunoslikning 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 tilshunoslikning yangi yo'nalishi – kompyuter lingvistikasi va tilshunoslikning bir qator nazariy va amaliy yo'nalishlari vujudga keldi.

Matematik lingvistika fani XX asrning 50-yillarida (1952-yilda) tilshunoslikning alohida yo'nalishi sifatida yuzaga keldi. Bu fanning shakllanishida Kopengagen struktural tilshunoslik maktabi (glossematika)ning asoschisi 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**» («**Lingvistik algebra**») deb atalishi lozim edi.³ Amerikalik tilshunos Noam Chomskiyning formal grammatika, transformatsion grammatika haqidagi qarashlari bevosita matematik lingvistikaning alohida yo'nalish sifatida yuzaga kelishiga 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'ullanuvchi fandır. Matematik lingvistika oldida turuvchi eng muhim masalalar quyidagilardir:

– 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 йўналишлари. – Тошкент, 2008. – Б.145.

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

- formal grammatika yaratish;
- tillarning matematik modellarini ishlab chiqish.

Har bir fanning o'z maqsad va vazifalari bo'ladi. Matematik lingvistika fanining asosiy maqsadi tabiiy tillarning matematik modelini ishlab chiqishdir. Ushbu maqsadga erishish uchun fan o'z oldiga quyidagi vazifalarni qo'yadi:

- tabiiy va sun'iy tillarning formal modellari algoritmini ishlab chiqish;
- lisoniy hodisalarni matematik parametrlarda baholash;
- til hodisalarini matematik metodlar yordamida tahlil qilish (*ehtimollar 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 tarjimai 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. Bu fanning asosiy maqsadi lingvistik masalalarni yechishning kompyuter dasturlarini ishlab chiqish, inson va mashina (kompyuter) muloqotini optimallashtirish⁵, tabiiy tilni qayta ishlash (**Natural Language Processing**)dir.⁶ NLP kompyuter lingvistikasida tabiiy tillarning kompyuter analizi va sintezini o'z ichiga oladi. Bunda *analiz* tabiiy tilning kompyuterda morfologik, sintaktik va semantik tahlil yordamida tushunilishiga nisbatan ishlatiladi, *sintez* esa kompyuterda matnning grammatik shakllantirilishi va *generatsiyasi* (hosil qilinishi) demakdir. NLP bo'yicha yaratilgan dasturiy ta'minotlar quyidagilar: *AlchemyAPI, Expert System S.p.A., General Architecture for Text Engineering (GATE), Modular Audio Recognition Framework, MontyLingua, 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 tahlilini (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.

- mashina tarjimasida uchun mo'ljallangan 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 strukturasi gipertekst texnologiyasini yaratish;
- elektron lug'atlar va tezauruslar yaratish;
- sun'iy intellekt tizimini ishlab chiqish;
- matnlar korpusini yaratish;
- matnni avtomatik referatlash;
- sujet strukturasi modellash;
- nutqni avtomatik tushunish tizimini ishlab chiqish (inglizcha **ASR – Automatic Speech Recognition**);
- matndan faktlarni avtomatik ajratib olish (inglizcha **fact extraction, text mining**).

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

O'zbek tili materiallari bo'yicha kompyuter lingvistikasiga oid tadqiqot olib borgan olimlar sifatida H.Arziqulov, S.Rizayev, S.Muhamedov, A.Po'latov, 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. O'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 etilishi lozim bo'lgan 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., Muhamedova S. Kompyuter lingvistikasi. – Toshkent, 2007. – B.5 – 6.

5. Kompyuterda inglizcha-o'zbekcha tarjima dasturlarini ishlab chiqish.
6. Kompyuterdagi matnlarni avtomatik tahrirlash dasturlarini yaratish.
7. O'zbek tili grammatikasining kompyuter modelini yaratish.
8. «Inglizcha-o'zbekcha tarjimon» kompyuter dasturini ishlab chiqish.
9. O'zbek tilidagi matnlarni tahrirlash dasturlarini yaratish.
10. O'zbek tilidagi «Windows» operatsion tizimini yaratish.
11. Inglizcha-o'zbekcha, o'zbekcha-inglizcha kompyuter lug'atlarini ishlab chiqish.

Kompyuter lingvistikasi amaliy tilshunoslikning tarkibiy qismi hisoblanadi, 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 quyidagi 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 jaryonida insonni istisno qiladi va u ko'proq kompyuterga moslashtiriladi;

- mumtoz tilshunoslik ko'proq tavsifiy (deskriptiv) xarakterga ega hisoblanadi. 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, tilshunoslikning nazariy masalalari bilan shug'ullanadi. Kompyuter lingvistikasi esa amaliy xarakterga ega bo'lib, til bilan bog'liq muammolarning amaliy jihatlari 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 tahlilida 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, bunda 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, statistika, semiotika, kibernetika, ehtimollar nazariyasi kabi turli fan sohalari bilan o'zaro aloqadorlikda ish ko'radi.

LINGVISTIKA FANI JENSKIY RODGA MANSUB, SHUNING UCHUN BU FANNI SEVISH KERAK, SHUNDAGINA UNI MUKAMMAL 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 tushuntiring.

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. Нелюбин Л.Л. 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., Йўлдошев Б. Тилшунослик ва табиий фанлар. – Тошкент: Ўқитувчи, 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.

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.

Tayanch so‘z va iboralar: *mantiq, grammatika, semiotika, algoritm, uchinchi mustasno qonuni, oraliq uchinchi, matematik mantiq, dialektik mantiq, konyunksiya, dizyunksiya, implikatsiya, ekvivalensiya, kvantor, lingvistik protsessor, tabiiy tilli interfeys, semantika, sintaktika, semiotika, lingvosemiotika, pragmatika, kod, kodlash, dekodlash, denotat, signifikat, algoritmik tillar, translyator dasturlar (kompilyatorlar).*

Mantiq, grammatika, semiotika va algoritm kompyuter lingvistikasining muhim unsurlari hisoblanadi. Ular til va nutqning funkcionallashuv jarayonini anglashda hamda lingvistik muammolarni avtomatik hal etish tizimini yaratishda poydevor bo‘lib 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 hodisa 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 nazariyasini – mantiqiy xulosa chiqarishning formal xarakterga ega ekanligini 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.

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 bo'lishi mumkin emas. Bu qonun «A – B dir yoki B 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 **oralik uchinchi** qoidasini olib kirdi. Unga ko'ra, hodisalarni baholashda ikkita bir-birini inkor etuvchi hodisa o'rtasida neytral – «oralik uchinchi» holat ham ajratiladi. Masalan:

issiq	iliq	sovuq
A	B	C

Demak, dialektik mantiqqa asosan hodisalar ko'p parametrlarda baholanadi. Keyinroq nemis olimi va matematigi Leybnis va Gilbertlar mantiq faniga matematik simvollarni tatbiq etdi. Ushbu olimlarning buyuk xizmatlari bilan matematik mantiq fani shakllandi.

Dastlab matematik mantiq faqat tabiiy fanlarda qo'llanib kelindi, keyinchalik u gumanitar sohalarga ham tatbiq etilgan. Nemis faylasufi va matematigi G.Leybnis (1646 – 1716) birinchilardan bo'lib mantiqiy fikrlashning hisob xarakteriga ega ekanligini ko'rsatdi. Uning fikricha, barcha ilmiy tushuncha va mulohazalarning asosini mantiqiy elementlar tashkil qiladi. Mana shu mantiqiy elementlarni muayyan simvollar bilan belgilash mumkin bo'ladi. Leybnis g'oyalari faqat XIX asrdagina amalga oshdi. Nemis olimi G.Frege, ingliz olimi J.Bul, rus olimi V.Poretskiy ishlarida simvolik 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 belgilaymiz. 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

⁹ Po'latov A., Muhamedova S. Kompyuter lingvistikasi. – Toshkent, 2007. – B.10.

programmalashtirish tillarida ishlatiladi. Shu o'rinda biz ularni alohida ajratib ko'rsatishni lozim topdik:

1-jadval

1. $A \& B$ – konyunksiya. Bu o'zbek tilidagi « va » bog'lovchisiga to'g'ri keladi.
2. $A \vee B$ – dzyunksiya. Bu o'zbek tilidagi « yoki » bog'lovchisiga to'g'ri keladi.
3. $A \Rightarrow B$ – implikatsiya. Bu o'zbek tilidagi « agar ... -sa, u holda ... » bog'lovchisiga to'g'ri keladi.
4. $A \equiv B$ – bu belgi mantiqiy teng kuchlilik (ekvivalensiya) uchun ishlatiladi.
5. A / B – bu belgi «Sheffir tayoqchasi» deb nomlanadi, u « va ... emas » bog'lovchisiga to'g'ri keladi.
6. $A \downarrow B$ – bu belgi «Pirs strelkasi» deb nomlanadi, u « yoki ... emas » bog'lovchisiga to'g'ri keladi.
7. \exists = inglizcha <u>exists</u> so'zining qisqartmasi bo'lib, mavjudlik kvantori sifatida ishlatiladi.
8. \sum = summa, barcha narsaning jami.
9. \forall = inglizcha <u>all</u> so'zining qisqartmasi bo'lib, hammasi uchun \in kvantori sifatida ishlatiladi.
10. x, y, z – predmetlarning bortigini belgilash uchun ishlatiladi.
11. X, Y, Z – predmetlarning 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. \bar{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 belgilar, mantiqiy bog'lovchilar yordamida belgilab olamiz. U gapning propozitsional ko'rinishi deb nomlanadi. So'ng jumlaning formal nazariya asosida hisoblaymiz. Bunda bizga hukmlarning qiymatini ko'rsatuvchi jadval yordam beradi.

2-jadval

A, B	$A \& B$	$A \vee B$	$A \Rightarrow B$	$A \equiv B$
00	0	0	1	1
01	0	1	1	0
10	0	1	0	0
11	1	1	1	1

Formal mantiq va simvolik mantiq til strukturasi o'rganishda amaliy ahamiyat kasb etadi, grammatikani formallashtirishda, sun'iy programmalashtirish tillarini yaratishda asos bo'lib xizmat qiladi, kompyuter lingvistikasining o'ziga xos simvolik «tili» sifatida namoyon bo'ladi. 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 metodlari 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-formales*) sifatida trivia (uchlik)ga birlashgan. Arifmetika, geometriya, musiqa va astronomiya «**real san'atlar**» (*artesreales*) sifatida kvadrivia (to'rtlik)ga birlashgan. Ushbu fanlarni o'zlashtirgan insongina yetuk va barkamol hisoblangan.

Tabiiy tilning strukturasi o'rganishda mantiq, grammatika va semiotika uzviy aloqadorlikda ish ko'radi. Mantiq va grammatika orasidagi aloqadorlikni til va tafakkur, so'z va tushuncha, gap va hukm, ega va subyekt, kesim va predikat kabi kategoriyalarning dialektik munosabati yordamida yaqqol tasavvur etish mumkin.¹⁰

3-jadval

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 bog'liq bo'lib, fikrni moddiylashtiradigan, kishilarning fikr almashishini ta'minlab beradigan quoldir. Til bilan

¹⁰ Лагуа О.Н. Логика и лингвистика. – Новосибирск, 2000. – С.7.

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 bo'lib, 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 tushunish 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 bo'lsa, 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 mantiqiy kategoriya orasidagi munosabat ham juda murakkabdir. Grammatik kategoriyalarni mantiq, mantiqiy kategoriyalarni grammatik kategoriyalarga 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 tilga ishlov berish (NLP), **lingvistik protsessorlar**, kompyuter uchun **tabiiy tili interfeys** bo'yicha olib borilayotgan tadqiqotlar mantiq, grammatika, semiotika, kompyuter texnologiyalari sohalarining 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 modellashtirilgani yo'q. Ayni shu masala – til va tafakkurni modellashtirish kompyuter lingvistikasi fanining istiqbolli yo'nalishi sanaladi.

Semiotika yoki **semiologiya** (yunoncha semiotike, ya'ni «belgilash», «tasvirlash», «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.¹¹ Semiotika tilshunoslikda, xususan, kompyuter lingvistikasida ham muhim metodologik ahamiyat kasb etadi. Chunki til ham bel-

¹¹ Эко У. Отсутствующая структура. Введение в семиологию. – М.: Петрополис, 1998. – С.31.

gi sanaladi, semiotikada tabiiy tillar ham, sun'iy tillar ham, hatto yo'l belgilari, audio va videosignallar ham tadqiq etiladi.

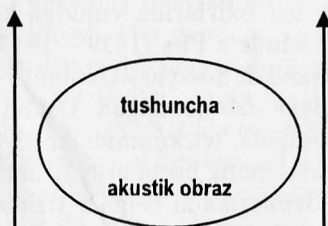
Semiotika fani XIX asr oxirlarida vujudga keldi, uning asoschisi amerikalik olim Charlz Sanders Pirs (1839 – 1914) hisoblanadi. Strukturalizm asoschisi Ferdinand de Sossyur «Umumiy tilshunoslik kursi» asarida bu fanni «**semiologiya**» deb nomlagan. Ushbu fan yutuqlari hozirgi zamon kompyutershunoslighida, telekommunikatsiya, uyali telefon tizimlari, sun'iy yo'ldoshlar, avtomatik boshqaruv tizimlari va boshqa sohalarda keng foydalaniladi. Semiotikada belgilar tizimi to'rt aspekt asosida o'rganiladi: 1) *sintaktika* – belgilar strukturasi tekshiradi; 2) *semantika* – belgilarning mazmuniy tomonini tekshiradi; 3) *igmatika* – belgi va uni aks ettiruvchi obyekt o'rtasidagi munosabatni o'rganadi; 4) *pragmatika* – belgilarning insonga bo'lgan munosabati, foydaliligi, amaliy jihatlarni tadqiq etadi.

Semiotika belgilar haqidagi keng qamrovli nazariy-falsafiy fan bo'lib, tilni belgilardan iborat butunlik sifatida o'rganadi. Tilshunoslik va semiotikaning o'zaro maqsadli aloqaga va munosabatga kirishuvidan lingvosemiotika deb ataluvchi soha shakllandi. Lingvosemiotika tilning belgi tabiatiga 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 informatsiyani 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 bo'ladi. Shunga ko'ra, til semiotika fanining – lingvosemiotikaning o'rganish obyektiga aylanadi. Tilning tabiatini, mohiyatini, jamiyatdagi o'rni va rolini 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.

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



Demak, lisoniy belgi tushuncha va akustik obraz bilan bog'liq. Bu terminlar «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 ijtimoiy vazifa bajarishi uchun imkoniyat beradi, shart-sharoit yaratadi. Ushbu 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 tomonga 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 nazardan 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 qiladi 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 ҳақида. – Андижон, 1992. – Б.14.

ma) muayyan ma'noga ega bo'lmagan, ayni vaqtda, muayyan ma'noga ega til birliklarining (morfema va so'zlarning) differentsiatsiyasi 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 obyekti hisoblanadi.

Morfema ifoda (moddiy) va mazmun (ruhiy) jihatlarining bir butunligidan iborat bo'lib, tilning muayyan ma'noga ega eng kichik birligi hisoblanadi. Muayyan birlik bo'lgan 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 shakllantiruvchi, uning moddiy birlik sifatida sezgi a'zolarimizga ta'sir qilishini ta'minlovchi tovushlar nazarda tutiladi.

So'zning mazmun plani uning ichki, ma'no tomoni (leksik ma'no, leksik tushuncha, ramziy axborot, signifikat, semema) bo'lib, u obyektiv borliq elementlarining kishi ongida aks etishi, ayni elementlarga xos xususiyatlarning 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, umumlashgan 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 ramzi (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 tasavvuri (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 ҳақида. – Андижон, 1992. – Б.17.

ning (obyektiv borliq elementining) lingvistik vakili vazifasini bajaradi. Soʻzda tovush ifodalovchi (ramz, belgi) boʻlsa, maʼno ifodalanuvchi boʻladi.

Soʻzdan yirik soʻz birkmalari va gaplar ham ifoda va mazmun planlariga ega, ularning mazkur ikki jihati koʻproq grammatika va semantikaning aloqasi doirasida qaraladi. Ushbu nutq birliklari ham, aslida, ifoda va mazmun planiga ega soʻzlarning birikuvidan, oʻzaro bogʻlanishidan yuzaga keladi. Shu sababli ularni oʻrganish ularni hosil qilgan soʻzlarni oʻrganishdan boshlanadi.

Informatsiya nazariyasi informatsiyani uzatish va saqlashning mazmun tomonini emas, balki uning statistik qurilishi bilan shugʻullanadi, yaʼni informatsiya nazariyasi axborot matnini statistik struktura sifatida oʻrganadi. Statistik struktura axborotdagi signal (simvol) va signallar birlashuvining voqelanish chastotasi soni, miqdoridir. Signalning koʻp marta takrorlanishida signal orqali uzatiladigan axborotning miqdorini aniqlash juda muhim. Demak, informatsiya nazariyasi muayyan axborotdagi maʼlumotlarni ushbu maʼlumotlarning tabiatiga – mazmuniga bogʻliq boʻlmagan holda miqdorini oʻlchash (tekshirish) yoʻllarini, usullarini ishlab chiqadi. Shunga koʻra, informatsiya nazariyasining muhim va asosiy xulosalari tilshunoslikda ham oʻz ifodasini topadi. Informatsiya nazariyasining asoschisi amerikalik olim K. Shennon informatsiyaning quyidagi aloqa kanallari boʻyicha uzatilishini sxema asosida tushuntiradi:¹⁶

Manba → uzatuvchi → kanal → priyomnik → axborotni oluvchi

Ushbu holat til vositasida insonlarning kommunikatsiyaga kirishuv jarayoniga ham toʻla mos keladi. Aslida, ushbu model til, tafakkur va nutq munosabatidan kelib chiqqan. Informatsiya tabiiy yoki sunʼiy tillar orqali uzatilar ekan, bunda til sistemasi kodlash va dekodlash jarayonlariga bogʻliq holda talqin etiladi.

Informatsiya nazariyasi maʼlumotlarini oʻrganish jarayonida maxsus terminlardan – tushunchalardan foydalaniladi. Bular quyidagilar:

1. **Kod** – axborotni yozib olish vositasi (usuli).
2. **Matn (tekst)** – muayyan axborotni berish, uzatish jarayonidagi belgilarning (til birliklarining, simvollarning) izchilligi, ketma-ket kelishi.
3. **Kanal** – axborot beriluvchi, uzatiluvchi muhit, sharoit.
4. **Informatsiya asosiy birlik boʻlib, 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.

5. Kodlash va dekodlash – informatsiyani mashina (kompyuter) tushunadigan 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 shakllanadi. Funksionallashuv jarayonida (xabarning bir individuumdan boshqasiga yetkazilishi va qabul qilinishi) til diskret shaklda – nutq ko‘rinishida namoyon bo‘ladi. Tovushli nutqning idrok qilinish jarayoni – bu nutq tovushlaridan 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.I.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 kompleksiga 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 oldingiga 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 qaytariladi: informatsiya ajratiladi va mazmun modellashtiriladi.¹⁸

Tilni informatsiya nazariyasiga ko‘ra yuqorida qayd etilgan uchta tushunchalardan 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 axborot 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 *algorithmus* – «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 Н.М. Моделирование мышления и психики. – Киев, 1965.

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

hamma masalalarni yechishda ishlatiladigan jarayonlar tizimining muayyan tartibda bajarilishi haqidagi aniq qoida yoki qoidalar tizimi.²⁰ O'rta asrlarda sanoqning o'nli sistemasi bo'yicha to'rt arifmetik amal bajariladigan qoida algoritm deb atalgan. Bu qoidalarni fanga IX asrda o'zbek matematigi 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 kibernetikaning 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 bo'lganimizda 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 yaratildi, ular programmashtirish jarayonida qo'llaniladi. Sun'iy tillar buyruqlarning tizimli algoritmiga tayanib yaratilgani uchun ular **algoritmik tillar (programmashtirish tillari)** deb ham ataladi. 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.

Algoritmik tillardagi matnlar xuddi tabiiy tildagidek aniq interpretatorga – insonga yoki avtomatik qurilmaga yo'nalgan bo'lishi lozim, toki u axborotni qabul qilsin va tushunsin. Agar bu jarayon (qabul qilish va tushunish) amalga oshmasa, u holda algoritmik tillar va kompyuter orasida vositachi sifatida **translyator dasturlar** (yoki **komplyatorlar**) 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-

²⁰ Фалсафа. Қисқача изоҳли луғат. – Тошкент: Шарқ, 2004. – Б.18.

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

ning mazmuni «buyruqlashtirilgan»: algoritm ijro etilishi, bajarilishi shart. Shunday qilib, aytish mumkinki, *algoritm*larni o'rganuvchi nazariya buyruq gaplar lingvistikasi sifatida talqin etilishi mumkin». ²²

Algoritm asosan quyidagi shakllarda yoziladi: a) so'z ko'rinishida – tabiiy til yordamida; b) grafik ko'rinishida – tasvir, chizma va grafikli simvollar yordamida; c) psevdokod ko'rinishida – algoritmning yarim formalashgan tavsifi bo'lib, u programmashtirishning dastlabki bosqichlarida qo'llaniladi. Bunda programmashtirish tillari unsurlaridan ham, tabiiy tildagi so'z va birikmalardan ham foydalaniladi; d) programma ko'rinishida – programmashtirish tillaridagi belgilar yordamida.

Algoritm buyruqlar tizimi, qat'iy bajariladigan amallar ketma-ketligi hamda programmashtirishning tayanch nuqtasi sanalgani bois uning vositasida kompyuter dasturlarini ishlab chiqish va muammoning avtomatik tarzda (mashina orqali) hal etilishini amalga oshirish mumkin. Algoritmni qamrovli vazifalarni hal etishga yo'naltirish, algoritm nazariyasini mukammallashtirish, 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 algoritm*lar nazariyasi, matematika va informatikaning tushunchasi, balki umuman zamonaviy fanlarning tayanch tushunchalaridan biridir. Axborot asrining kirib kelishi bilan algoritmlar sivilizatsiyaning eng muhim faktorlaridan biri bo'lib qoldi». Shu o'rinda ta'kidlash joizki, kompyuter lingvistikasida shunday masalalar ham borki, ularni algoritm asosida hal etish mumkin emas. Jumladan, sun'iy intellekt tizimi doirasidagi neyron tarmoqlar, sifatli avtomatik tarjima kabilar. Ushbu muammolarni 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 mantiqdir.

Topshiriqlar:

1. Mantiq va semiotikaning kompyuter lingvistikasidagi tatbiqini tushuntiring.
2. Konyunktiv, disyunktiv, implikativ hukmlarga o'zbek tilidan misollar keltiring.

²² Успенский В.А., Семенов А.Л. Теория алгоритмов: основные открытия и приложения. – М.: Наука, 1987. – С.13.

3. Matematik mantiq fanini qaysi olimlar rivojlantirgan?

4. Ushbu gapning propozitsional ko'rinishini yozing: Agar Jons bu tunda 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 bo'ldi. Agar qotillik yarim tundan keyin sodir bo'lgan bo'lsa, u holda Smit qotil yoki Jons aldayapti. Yuqoridagilardan kelib chiqadiki, Smit qotil.

5. Qanday algoritmik tillarni bilasiz?

Adabiyotlar:

1. Yoqubov T., Karimbekov S. Matematik mantiq elementlari. – Toshkent: O'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 ҳақида. – 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.

AKSIOMATIK NAZARIYA – TILNI FORMAL TAVSIFLASH ASOSI

Reja:

1. Aksiomatik nazariya haqida umumiy ma'lumot.
2. Aksiomatik nazariyaning lingvistik tadqiqi.
3. Lingvistik universalialar g'oyasi.
4. Metagrammatika.

Tayanch soʻz va iboralar: aksioma, postulat, teorema, ziddiyatsizlik, toʻlalik, bogʻliqsizlik, aksiomatik metod, aksiomatik nazariya bosqichlari, formula, keltirib chiqarish qoidalari, metagrammatika, metaqoidalar, lingvistik universalialar, frekventaliya va unikalialar, metaqoida, implikativ universalialar.

Aksiomatik nazariya miloddan avvalgi VII – VI asrlarda shakllangan. Uning yaratilishi matematik olim Pifagor nomi bilan bogʻliq. Lekin bu metodni birinchi marta miloddan avvalgi III asrda yashab ijod etgan yunon olimi Yevklid oʻzining «Negizlar» asarida qoʻllagan. Aksiomatik nazariya bosqichlari quyidagilardan iborat:

1. Postulat – yunoncha «talab» maʼnosini anglatadi. Qadimgi dialektika oʻyinida har bir kishining ilgari surgan gʻoyasi, fikri. Uni tasdiqlash ham, inkor qilish ham mumkin.

2. Aksioma – yunoncha «hurmat qilaman» maʼnosini anglatadi. U ilmiy nazariyaning hech qanday mantiqiy isbotsiz toʻgʻri deb tan olinuvchi va boshqa ilmiy xulosalar uchun asos boʻluvchi maqbul deb topilgan haqiqat, goʻyo olimlar bu fikrni «hurmat qilib» inkor etmaydi. Aristotel aksiomani «*har qanday isbotsiz eʼtirof etiladigan boshlangʻich hukmlar*», «*ilm-fanning poydevori, ibtidosi*» sifatida talqin qilgan.

3. Teorema – aksiomadan farqli ravishda mantiqan keltirib chiqariladigan jumlar, u shart, isbot va xulosa qismlariga ega.

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

– ziddiyatsizlik – aksiomalar tizimining shunday xossasiki, unga koʻra, sistemaning hech qanday ikkita holati bir-biriga zid kelmasligi kerak. Bunda aksiomalar tizimining oʻzida rost va yolgʻon orasidagi farqni aks ettirishi kerakligi nazarda tutiladi;

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

– to‘lallik – aksiomalar sistemasiga qandaydir jumlani qo‘shganimizda sistemada ziddiyat sodir bo‘ladi;

– bog‘liqsizlik – 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 (teoremlar) 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 qilingan formulalarning barchasidan foydalanish mumkin bo‘ladi. Aynan shuning uchun mazkur metod samarali hisoblanadi.

Aksiomatik nazariyada dastlabki boshlang‘ich tushunchalar, postulat, aksiomalar, teoremlar, keltirib chiqarish qoidalari asosiy o‘rin tutadi. Masalan: **Aksioma:** so‘zda nechta unli bo‘lsa, shuncha bo‘g‘in bo‘ladi. **Teorema:** Agar ixtiyoriy X gap inversiyaga uchramasa, NSP – Normal Sentence Pattern (normal gap qurilishi)ga mos kelsa, u holda ega doim kesimdan oldin keladi.

Aksiomatik nazariya quyidagi holatlar aniqlanganda haqiqiy hisoblanadi:²⁴

1. Nazariya ifodalarini tavsiflash uchun zaruriy bo‘lgan nazariya simvollarini – harflar, belgilar yoki raqamlar.

2. Nazariya formulalari, ya‘ni nazariyada fikrlangan barcha ifodalarning tavsifi.

3. Nazariya aksiomalari (boshlang‘ich va mutlaq to‘g‘ri deb hisoblangan formulalar).

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

5. Nazariyaning teoremlari – nazariyaning mutlaq to‘g‘ri, umum tomonidan qabul qilingan formulalari.

Til grammatikasining matematik modeli matematik mantiqning aksiomatik nazariyasiga asoslanadi. Aksiomatik nazariya predmeti tabiiy tillardagi gaplardir (ya‘ni ixtiyoriy so‘zlar ketma-ketligi). Teorema sifatida tabiiy

²⁴ Po‘latov A., Muhamedova S. Kompyuter lingvistikasi. – Toshkent, 2007. – B.15.

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

1. Aksiomatik nazariyaning simvollarini: til harflari (alfibo), tinish belgilari 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 konstruksiyalari. Masalan, ingliz tili uchun $ega+kes+to'ld+hol$ (kengaytirilgan tasdiq (darak) gap konstruksiyasi): Mother bought milk yesterday. O'zbek tili uchun $ega+to'ld+hol+kes$ (kengaytirilgan tasdiq (darak) gap konstruksiyasi): Nafisa ma'ruzani diqqat bilan tingladi.

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

G – ingliz tilidagi muayyan bir gap;

K Ch q – keltirib chiqarish qoidalari.

1. $G(e,k,t,h) + \%G^1(e,k)$, ya'ni G gapidan yangi (kengaytirilmagan gap) hosil bo'ladi (unda gapning ikkinchi darajali bo'laklari tushirib qoldirilgan).

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

G¹: The car stopped – Mashina to'xtadi.

2. $G(e,k,h) + \%G^2(e',k,h)$. Bunda G gapidan aniqlovchili (ya'ni har bir gap bo'lagining o'z aniqlovchisi mavjud) gap hosil bo'ladi.

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 bo'lgan umumiy qonuniyatlarni aniqlash masalasi ko'ndalang qo'yildi. Bu hodisa lingvistik universalialar deb ataladi.²⁶ Ch.Xokket lingvistik universalialarni shunday ta'riflaydi: «Lingvistik universalialar – dunyodagi barcha tillarga xos bo'lgan yoki umuman tilga xos muayyan belgi, xususiyat». B.Uspenskiy: «Lingvistik universalialar tushunchasi ostida barcha tillarga yoki mutlaq ko'pchilik qismiga xos umumiy qonuniyatlar tushuniladi». Lingvistik universalialar haqidagi dastlabki g'oyalar XX asrning 50-yillariga to'g'ri keladi. 1941-yilda Aginskiy tomonidan «The Importance of language universals» sarlavhali maqola e'lon qilindi. 1961-yilda Nyu-York

²⁵ Po'latov A., Muhamedova S. Kompyuter lingvistikasi. – Toshkent, 2007. – B.16.

²⁶ Qarang: Новое в лингвистике. Вып.V. Языковые универсалии. – М.: Прогресс, 1970.

shahrida universalialarga bag'ishlangan xalqaro konferensiya o'tkazildi. Unda J.Gringer, J.Jenkins, Ch.Osgudlar «**Lingvistik universalialar haqida memorandum**» mavzusida chiqish qilishdi. Konferensiyada Ch.Osgud shunday yozadi: «*Biz ushbu konferensiyada qonsiz inqilobning guvohi bo'lamiz*». Haqiqatan ham shundan so'ng universalialarning tadqiqi kuchaydi. Bunda dastlab **universalialar**, **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 universalialarni aniqlash juda muhim.

Amerikalik olim Jozef Gringer universalialarni aniqlashda yangicha bir yondashuvni taklif etdi. Unga ko'ra, universalialar implikasiya asosida aniqlanadi. Bu **implikativ universaliya** deb nomlanadi. Olim 45 ta implikativ universaliani 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 suffiksali bo'lsa, u posleloglarga, kelishiklarga boy bo'ladi.

3. Agar til istisnoli tarzda prefiksali 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 universalialar 2000-yilga kelib 1600 tani tashkil etdi. Germaniyaning Konstanz universitetida lingvistik universalialarning bazasi (**Universals Archive**) yaratildi va u har yili yangi ma'lumotlar bilan boyitib boriladi. Uni obrazli qilib D.I.Mendeleyevning kimyoviy elementlar davriy jadvaliga o'xshatish mumkin. Davriy jadvalda keyinroq aniqlanishi mumkin bo'lgan elementlar uchun bo'sh kataklar qoldirilgan. Xuddi shunga o'xshab universalialar bazasini to'ldirib borish uchun imkoniyat mavjud.²⁸ Mazkur baza Internetda <http://www.ling.uni-konstanz.de/pages/proj/sprachbau.htm> saytiga joylangan.

Lingvistik universalialar 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. V. Языковые универсалии. – М.: Прогресс, 1970. – С.115 – 162.

²⁸ <http://www.ling.uni-konstanz.de/pages/proj/sprachbau.htm>

1. Metagrammatika (yunoncha «metha» – «keyin», «grammatike» – «o'qish va yozish san'ati») – metaqoidalar asoslangan grammatika.

2. Metagrammatika – tavsifiy grammatika, formal grammatika, transformatsion grammatika bosqichlaridan so'ng yuzaga kelgan, qayta yozilgan 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 Winjaarden-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

$\langle VP \rangle ::= \langle *v \rangle \langle NP \rangle X$

add a production of the form

$\langle VP \rangle ::= be \langle *VEN \rangle X$

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

$\langle VP \rangle ::= \langle *V \rangle \langle NP \rangle$ and

$\langle VP \rangle ::= \langle *V \rangle \langle NP \rangle \langle PP \rangle$

the meta rule will add the productions

$\langle VP \rangle ::= be \langle *VEN \rangle$ and

$\langle VP \rangle ::= be \langle *VEN \rangle \langle PP \rangle$

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

O'zbek tilida metaqoidaga misol sifatida formal-funksional tilshunoslik vakillarining WP_m modelini 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.

Topshiriqlar:

1. Aksioma nima, unga misollar keltiring.
2. Boshlang'ich tushunchalar, postulatlarni belgilash qanday amalga oshiriladi?
3. Grammatikaning aksioma va teoremlariga o'zingiz misollar keltiring.
4. Keltirib chiqarish qoidalari deganda nimani tushunasiz?
5. Lingvistik universalialar 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.V. Языковые универсалии. – М.: Прогресс, 1970. – С.115 – 162.
6. Новое в лингвистике. Вып.V. Языковые универсалии. – М.: Прогресс, 1970.
7. Po'latov 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 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 soʻz va iboralar: *model, tabiiy model, matematik model, mantiqiy-matematik model, metamodel, deduktiv, lingvistik model, sintaktik model-lashtirish, soʻz turkumlari va ularning simvollari, gap boʻlaklari simvollari, analogiya metodi, induksiya, deduksiya.*

Model (lotincha «modelus» soʻzidan olingan boʻlib, «nusxa», «anda-za», «oʻlchov», «meʼyor» maʼnolarini anglatadi) tabiiy fanlar yoki umuman fanda shunday moddiy qurilma, grafik, sxema, umuman, bilish vositasi sifatida tushuniladiki, u muayyan original-obyekt haqidagi maʼlumotlar majmui tarzida yuzaga kelgan hosila-obyekt demakdir. Boshqacharoq tushuntirganda, model tabiiy obyektlarning imitatsiyasidir (oʻxshashi, taq-lidiy koʻrinishi), u oʻzbek tilidagi qolip, andaza soʻzlariga mos keladi. U hodisalarning yuzaga kelishi uchun asos vazifasini oʻtaydi, bunda aniq yoki mavhum obyektlar kichraytirilgan obyektlar, sxemalar, chizmalar, fizikaviy konstruksiyalarda tadqiq etiladi. Buni oddiy hayotiy misol bilan tushuntira-digan boʻlsak, olmani xarakterlovchi belgilar, atributlar – uning dumaloq-ligi, mevaligi, shirinligi oʻsha tushunchaning fikriy modeli hisoblanadi. Agar biz olmani loydan yoki sunʼiy bir materialdan yasasak, bu uning moddiy modeli hisoblanadi. Yoki globus yerning modeli, oʻyinchoq mashina haqi-qiy mashinaning modeli sifatida baholanishi mumkin.³⁰ Model quyidagi asoslarga koʻra bilishda muhim hisoblanadi:

- birinchidan, u oʻrganish obyektini soddalashtiradi;
- ikkinchidan, uni boshqa obyektlar taʼsiridan ajratadi;
- uchinchidan, model obyektini taʼriflashni osonlashtiradi.

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

³⁰ <http://www.wikipedia.org/wiki/model.html>

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, modellashtirishning 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 o'Ichamlari, 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 bo'ladi.

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 unurni o'z ichiga oladi:

- subyekt (tadqiqotchi);

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

– tadqiqot obyekti;
 – o‘rganuvchi subyekt va o‘rganiluvchi obyekt munosabatini aks ettiruvchi model.

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

– deduktivlik – mantiqiy xulosa chiqarishga asoslangan bo‘ladi, xususiylkdan umumiylikka 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;³²

– modelni ideallashtirilgan obyekt sifatida talqin qilish.

Shu o‘rinda aytib o‘tish zarurki, modellashtirish obyektini umumlashtirish darajasiga ko‘ra quyidagicha bo‘ladi:

1. Lingvistik faktni tavsiflashga qaratilgan analitik model.
2. Oraliq model yoki to‘ldiruvchi model.
3. Maksimal umumlashtirishga asoslangan sintezlovchi model.

Lingvistik model tushunchasi struktur tilshunoslikning E.Sepir, L.Blumfeld, R.Yakobson, N.Chomskiy, Z.Harris, Ch.Hokket kabi namoyandalari 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 yasaliishining umumiy modeli: **asos + so‘z yasovchi qo‘shimcha**; xususiy modellari: asos + -chi; asos + -dosh; asos + -do‘z kabi.

³² Po‘latov A., Muhamedova S. Kompyuter lingvistikasi. – Toshkent, 2007. – B.23.

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

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

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

S = ega, O = to'ldiruvchi, V = kesim.

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

S₁ + V₁, 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 strukturasi 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 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*» jumlasida distributiv model asosida quyidagicha tahlil etiladi: **D 3 I^a 2-d D I^b 4**. Bu yerda **D** – o'ning aniqlovchisi (inglizcha determiner), **3** – sifat, **I** – birlikdagi ot, **2-d** – o'tgan zamon shaklidagi fe'l, **4** – ravishni anglatadi. Demak, mazkur modelda turli so'z turkumlariga mansub so'z shakllarining nutq zanjiridagi distributsiyasi (tarqalishi, qurshovi) gap strukturasi modellashtirishning asosiy mezonini sanaladi.

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_{com}**, atoqli otlar uchun **N_{prop}**, otli birikma uchun **NP** simvollaridan foydalaniladi.

II. Sifat qanday, qanaqa so'roqlariga javob bo'lib, predmetning belgisi-

³³ Бархударов Л.С. Структура простого предложения английского языка. – М.: Высшая школа, 1966. – С.18.

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

III. Son qancha, nechanchi soʻroqlariga javob boʻlib, predmetning miqdori va tartibini bildiradi. Son soʻz turkumi uchun **Q** simvolidan foydalaniladi.

IV. Feʼl predmetning harakat yoki holatini bildirib, nima qilgan, nima qilyapti, nima qilmoqchi soʻroqlariga javob beradi. Feʼl uchun **V**, oʻtimli feʼl uchun **V_t**, oʻtimsiz feʼl uchun **V_i**, feʼlning predikativ (shaxsli shakli – finite form) shakli uchun **V_p**, feʼlning nopredikativ shakli (shaxssiz shakli – non-finite form) uchun **V_{np}**, hozirgi zamon sifatdoshi yoki gerundiy shaklidagi feʼl uchun **V_{ing}**, oʻtgan zamon sifatdoshi yoki majhul nisbatdagi feʼl uchun **V_{en}**, feʼlli birikma uchun **VP**, yordamchi feʼllar uchun **aux**, modal feʼllar uchun **mod** simvollaridan foydalaniladi.

V. Ravish ifodalayotgan harakat yoki holatning belgisi yoki turlicha vaziyatlarini bildiradi. Ravish asosan feʼlga, yana sifat yoki ravishga ham bogʻlanib, qay tarzda, qayerda, qachon va h.k. soʻroqlarga javob boʻladi. Ravish uchun **Adv** yoki **D** simvolidan foydalaniladi.

VI. Olmosh gapda ot, sifat yoki ravish oʻrnida qoʻllanuvchi soʻz turkumidir. Olmosh uchun **Pron** simvolidan foydalaniladi.

VII. Artikl yordamchi soʻz boʻlib, oʻtning maxsus belgilovchisi hisoblanadi. Artikl barcha tillarda uchrayvermaydi, **Art** simvolidan foydalaniladi.

VIII. Predlog (old koʻmakchi) yordamchi soʻz boʻlib, oʻtning gapdagi boshqa soʻzlarga nisbatan boʻlgan munosabatini (yaʼni maʼkon, vaqt va sabab kabi munosabatlarni) ifodalaydi. Predlog uchun **Prep** yoki **p** simvolidan foydalaniladi.

IX. Bogʻlovchi yordamchi soʻz boʻlib, gap boʻlaklari yoki gaplarni bogʻlash uchun xizmat qiladi. Bogʻlovchi uchun **Conj** simvolidan foydalaniladi.

X. Yuklama uchun **Part** simvolidan foydalaniladi.

XI. Undov soʻzlar uchun **Interj** simvolidan foydalaniladi.

XII. Taqlid soʻzlar uchun **Mim** simvolidan foydalaniladi.

Gap boʻlaklari

Ixtiyoriy tildagi yozma matn yoki shaxsning nutqi jumla deb ataluvchi alohida segmentlardan iborat boʻladi. Ular sodda va qoʻshma (yaʼni sodda gaplardan tashkil topgan) gaplarga boʻlinadi. Har qanday tilning grammatikasini oʻrganish sodda gaplar bilan tadqiq etishdan boshlanadi. Har qanday tildagi sodda gaplar tarkibida ega, kesim, toʻldiruvchi, aniqlovchi, hol kabi gapning bosh va ikkinchi darajali boʻlaklari mavjud boʻlishi mumkin.

I. Ega – gapning zaruriy bo‘lagi bo‘lib, boshqa bo‘laklar bilan grammatik bog‘lanadi, u predmet yoki shaxsni bildiradi. Ega kim, nima, qayer so‘roqlariga javob bo‘ladi. Ega uchun **S** simvolidan foydalaniladi.

II. Kesim – gapning ham semantik, ham grammatik markaziy bo‘lagi bo‘lib, 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. To‘ldiruvchi deb predmetni bildiruvchi va ma‘no jihatidan kesimni to‘ldiruvchi gap bo‘lagiga aytiladi. Vositasiz (obyektsiz) to‘ldiruvchi kimni, nimani so‘roqlariga, vositali to‘ldiruvchi esa kimga, nimaga, kimda, nimada, kimdan, nimadan, kim haqida, nima haqida va h.k. so‘roqlarga javob bo‘ladi. To‘ldiruvchi uchun **O** simvolidan foydalaniladi.

IV. Aniqlovchi otning belgilarini tushuntiruvchi, izohlovchi gap bo‘lagidir. Aniqlovchi predmetning belgisini bildirgan holda qanday, kimning, nechanchi, qancha kabi so‘roqlarga javob bo‘ladi. Aniqlovchi kompyuter lingvistikasida gap bo‘lagi sifatida ajratilmaydi, chunki u gapda boshqa bo‘laklarga tobelanib keladi, u gap uchun nozaruriy (fakultativ, obligator emas) bo‘lak hisoblanadi.

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

Modellashtirish metodi keyingi paytlarda tilshunoslikka faol tatbiq qilina 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 bo‘lmaydi;
- to‘rtinchidan, obyektни tushuntirish va izohlashni osonlashtiradi va soddalashtiradi.

Borliqdagi hamma narsalarning ijobiy va salbiy tomonlari bo‘lganidek, modellashtirish metodining ham pozitiv va pragmatik tomonlari bilan birga salbiy tomonlari ham bor. Uning kamchiliklari, salbiy tomonlarini quyidagicha umumlashtirish mumkin:

- birinchidan, modellashtirishda faqat struktur va formal belgilarga asoslaniladi. Bunda mazmuniy tomon, semantik qirralar e‘tibordan chetda qola-

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 rivojlanadi. Mavjud modellar bilimlarimiz ufqini cheklaydi va xato tasavvurlarga olib kelishi mumkin;

– uchinchidan, tabiat va jamiyatdagi hech bir narsa boshqa obyektlar ta'siridan xoli bo'lmaydi. 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 *ἀναλογία* so'zidan olingan, «moslik», «o'xshashlik» ma'nosini anglatadi) – munosabatlarning tengligi, muvofiqligi; hodisa va jarayonlarning muayyan xossalari qiyoslash yo'li bilan anglashda o'xshash jihatlari. Qiyoslanayotgan obyektlar o'rtasida o'xshashlik va farqlanish kuzatiladi, bu qiyoslashning asosi sanaladi. Bunda ma'lum obyektga xos bo'lgan aniq belgining qiyoslanayotgan obyektlarda kuzatilishiga tayanib ularning identifikatsiyasi asosida musharak jihatni 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 musharak jihatlari ajratish mumkin. Yoki ko'p odamlar «*singan oyna – baxtsizlik, noxush 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 bo'ladi. Bilish jarayonida ular o'rtasiga qat'iy chegara qo'yib bo'lmaydi. Analogiya deduksiya va induksiya o'rtasida bog'lovchi «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 o'rtasidagi analogiyani topa olgan odam, eng yaxshi matematik – isbotlar o'rtasida analogiya o'rnata olgan odam, bundan ham ko'ra kuchli matematik – nazariyalar analogiyasini sezgan odam; biroq odam o'zini eng kuchli*

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 shunday 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 mazmunga ega bo'lgan 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 bunday yondashish analogiya metodi asosida bo'ladi. Mazkur yondashuv, ayniqsa, 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 kategoriyalarga egaligi) bir xil bo'lishi tabiiy. Masalan, ingliz tilida *-er, -ness, -dom* kabi, rus tilida *-ность, -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 bo'lagi maqomiga ega emas?

³⁴ Белоногов Г.Г. Об использовании принципа аналогии при автоматической обработке текстовой информации // Проблемы кибернетики. 1974. №28. – С.33.

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. – Л., 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. Po'latov A., Muhamedova S. Kompyuter lingvistikasi. – Toshkent, 2007.
8. Бархударов Л.С. 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 soʻz va iboralar: *grammatika, morfologik sistema, sintaktik sistema, germeneytika, preskriptiv grammatika, sinxroniya, diaxroniya, grammatik modellar, ikki choʻqqili nazariya, bir choʻqqili nazariya, verbosent-rizm, tobelik daraxti, bevosita ishtirokchilar modeli, tema, rema, formal grammatika, transformatsion grammatika, operand, transforma, terminal va no-terminal simvollar, formallashtirish.*

Grammatika tabiiy tillar strukturasi haqidagi fan boʻlib, u morfologik kategoriya va shakllar, sintaktik ketegoriya va konstruksiyalar hamda soʻz

yasash usullari tizimini tashkil etadi. Til qurilish birliklarining paradigmatic va sintagmatik munosabatga kirishuvi (vertikal va gorizontal yoʻnalishda) muayyan qoidalar asosida yuz beradi. Ushbu qoidalar yigʻindisi tilning grammatik qurilishi va grammatik sistemasi deb yuritiladi. Grammatika ikki yirik sistemani oʻz ichiga oladi:

1. Morfologik sistema – soʻz formalari, grammatik shakllar, grammatik kategoriyalar tizimi.
2. Sintaktik sistema – sintaktik kategoriyalar va konstruksiyalar, soʻzlarning oʻzaro birikish qonuniyatlari.

Grammatika yunoncha soʻz boʻlib, «oʻqish va yozish sanʼati» maʼnosini bildiradi. Bu tushuncha dastlab xat-savod, imloni oʻrgatuvchi fan maʼnosida ishlatilgan. Keyinroq grammatika tilning fonologik, morfologik, sintaktik va semantik belgilarini tavsiflovchi til qoidalari majmuasi maʼnosiga koʻchgan. Grammatika haqidagi ilk qarashlar falsafa, mantiq, **germenevtika** (diniy matnlarni sharhlash, talqin etish) qobigʻida boʻlgan. Bunda tilning paydo boʻlishi, til va tafakkur munosabati, lingvistik belgi motivatsiyasi, diniy matnlar mazmunini toʻgʻri tushunish bilan bogʻliq masalalar oʻrganilgan, til strukturasi esa yetarli darajada tadqiq etilmagan. Faqatgina miloddan avvalgi II – I asrlarda shakllangan Aleksandriya grammatika maktabining eng yirik vakili Dionisiy Traks (miloddan avvalgi 170 – 90-yillarda yashagan) oʻzigacha boʻlgan tadqiqotchilarning ishlari va tajribalarini oʻrganib, ulardan foydalanib rimliklar uchun «**Grammatika sanʼati**» («**Grammatike techne**») nomli sistemalashtirilgan dastlabki yunon grammatikasini yaratdi. Shundan buyon grammatika falsafadan mustaqil soha sifatida ajralib chiqdi.³⁵

Antik davrda grammatika **preskriptiv** (toʻgʻri gapirish, toʻgʻri jumla tuzishni koʻrsatuvchi yoʻriqnoma) xarakterga ega boʻlgan. XVIII asr oxirlarida grammatikani tushunish keskin oʻzgardi. 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 rekonstruksiya qilish boʻyicha qiyosiy aspekt-da tadqiqotlar olib borildi.

XX asr boshlarida Ferdinand de Sossyur grammatikani tavsiflashda sinxroniya va diaxroniyani farqladi. Bunda grammatikani tarixiy (diaxron) oʻrganishni 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.

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.

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 ifodalansa, 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. Eganing aniqlashtiruvchilari ot, sifat, son, olmosh, kvantifikatorlar (miqdor bildiruvchi so'zlar), determinatorlar (artikllar) orqali ifodalangan aniqlovchi 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 hol kesimning aniqlashtiruvchilari bo'lib keladi. Mazkur modelda ega va kesim gapning bosh bo'laklari, gapning semantik-sintaktik yadrosi, aniqlovchi, to'ldiruvchi, hol esa ikkinchi darajali bo'laklar deb nomlanadi. Ushbu talqin mumtoz sintaksis tamoyillariga 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-

da fe'ldir) konsepsiyasi bilan bog'liqdir.³⁶ Ingliz tilshunosi O. Yespersen mazkur konsepsiyaga qarshi o'laroq gapdagi mutlaq hokim bo'lak sifatida egani e'tirof etgan.³⁷ Gap bo'laklari iyerarxiyasida egani markazga qo'yish tendensiyasi mantiq fanidan kelib chiqqan. Chunki mantiqda ism (subyekt) substansiyani (mohiyatni), fe'l (predikat) aksidensiyani (hodisani) ifodalaydi, 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 modellari 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 daraxtning ildizida joylashgan bo'ladi va dominant bo'lak hisoblanadi, boshqa so'zlar (gap bo'laklari) unga bevosita yoki bilvosita (boshqa so'zlar yordamida) 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 haqidagi 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 vazifasida 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 ishtirokchilar (immediate constituents)** sintaktik modeli hisoblanadi. Bevosita ishtirokchilar metodiga ko'ra, tayanch nuqta konstruksiya hisoblanadi. Konstruksiya tarkibidan ishtirokchilar va bevosita ishtirokchilar ajratiladi.³⁸ Konstruk-

³⁶ Теньер Л. Основы структурного синтаксиса. – М.: Прогресс, 1988. – С.12.

³⁷ Есперсен О. Философия грамматики. – М.: Иностранная литература, 1958. – С.164 – 165.

³⁸ Nurmonov A. Lingvistik tadqiqot metodologiyasi va metodlari. – Toshkent: Akadernashr, 2010. – B.88.

siya atamasi ostida ma'noli qismlarning ketma-ket munosabatidan tashkil topgan butunlik tushuniladi. Ishtirokchilar deb kattaroq konstruksiya tarkibiga 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 bajaradilar*. Bu jumla konstruksiya hisoblanadi, uning tarkibidagi barcha ma'noli so'zlar («va» bog'lovchisidan tashqari) ishtirokchilar, bevosita aloqaga kirishgan so'zlar bevosita ishtirokchilar 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 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 parchasi ikki qismdan tashkil topadi. O'z navbatida, har qaysi qism yana o'z ichida ana shunday qismlardan iborat bo'ladi. Ko'rinadiki, har bir konstruksiya binar tamoyilga ko'ra qismlarga bo'linadi va bu bo'linish konstruksiya doirasida graduallik (darajalanish) xususiyatiga ega bo'ladi.³⁹

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 strukturasiidagi transformatsiya 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 tushuniladi.⁴⁰ Ingliz tilida yadro gaplar sifatida quyidagi konstruksiyalar ajratib ko'rsatiladi⁴¹ (4-jadval).

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

³⁹ 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. II. – М., 1962. – С.627 – 628.

4-jadval

1.	NV	Noun Verb	Ot fe'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.	NisN	Noun is Noun	Ot is ot	John is a teacher
5.	NispN	Noun is Preposition Noun	Ot is predlogli ot	John is in bed
6.	NisD	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* jumlasida 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 shakliy o'zgarishlar asosida asosiy ma'noni saqlagan holda ikkilamchi sintaktik qurilmalarning hosil qilinishi (**transformatsiyalar = transformalar**), bu jarayondagi vositalar (**operatorlar**) va **transformatsiya hosil qilish modellari** tadqiq 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 formallashtirish 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 Н. Синтаксические структуры // Новое в лингвистике. Вып. II. – М., 1962. – С.412 – 527.

miyatga molik.⁴³ To'rtinchidan, transformatsion grammatika tabiiy tillardagi paradigmatic va sintagmatic munosabatlarning me'yoriyligi va grammatik to'g'ri va noto'g'ri jumllarni 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 iborat bo'ladi. Quyidagi jadvalda ingliz tili uchun ayrim gapning struktur komponenti simvollari ko'rsatilgan:

5-jadval

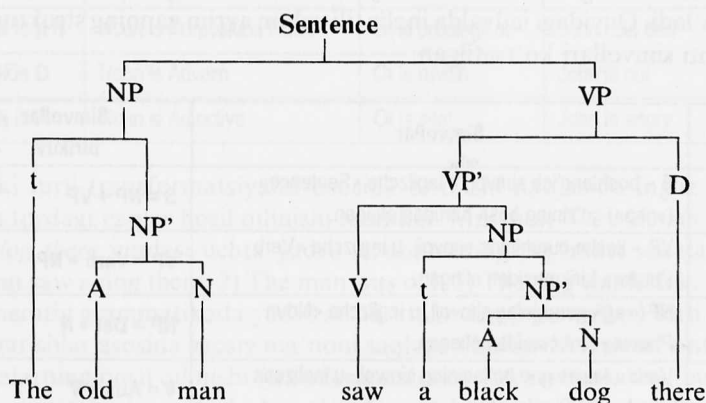
№	Simvollar	Simvollar birikuvi
1	S – boshlang'ich simvol, u inglizcha «Sentence» («gap») so'zining bosh harfidan olingan.	S = NP + VP
2	VP – kesim guruhining simvoli, u inglizcha «Verb Phrase» birikmasidan olingan.	VP = Verb + NP
3	NP – ega guruhining simvoli, u inglizcha «Noun Phrase» birikmasidan olingan.	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-qatordagi simvollar esa *noterminal simvollar* deb ataladi.

Gap hosil qilish jarayoni noterminal simvollarini 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-

⁴³ Шаумян С.К. Теоретические основы трансформационной грамматики // Новое в лингвистике. Вып. II. – С.408.

veradi. Masalan, ingliz tilidagi quyidagi jumlani bevosita ishtirokchilar modeli asosida tahlil etaylik. Birinchi bosqichda jumla NP va VPga ajratiladi: *The old man | saw a black dog there* – *Keksa odam u yerda qora itni ko'rdi*. Keyingi bosqichda yuqoridagi ajralish, o'z navbatida, yana ikki qismga 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



Context-free grammar – matndan 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 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 Л.С. Структура простого предложения английского языка. – М.: Высшая школа, 1966. – С.21.

=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 shakllar yordamida sun'iy til belgilari orqali ifodalashdir. Grammatikani formallashtirish tendensiyasi XX asrning o'rtalarida kuchaydi. Bu bevosita tilning strukturasi yanada aniq parametrlarda ifodalashga intilish, uni kompyuterga moslashtirishga urinish bilan bog'liqdir. Formal grammatika 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 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.

N.Chomskiy qoidalarining 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 $G=(N, \text{”}, P, S)$ belgilari yordamida ish ko'riladi. N, “ – alifbo (N – noterminal simvollar, u grammatikadagi simvollarini o'z ichiga oladi; “ – metatilning terminal simvollarini, u tildagi so'zlarni qamrab oladi); S – noterminal to'plamning boshlang'ich simvoli, P – qayta ishlab chiqilgan qoidalar. Bunda terminal va noterminal simvollarini birlashtiruvchi V belgisi qabul qilingan, u inglizcha Vocabulary (lug'at) so'zidan olingan ($V=TN$). Bu grammatika o'ta abstraktlashgan 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 gram-

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, tilning 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 bo'linadi: 1) tug'dirish grammatikasi (generative grammar) – tog'ri gaplarni qabul qilish protsedurasi; 2) tushunuvchi, tanuvchi grammatika (recognizing grammar) – to'g'ri gaplarni tushunish protsedurasi.

Demak, grammatika tilning asosida yotuvchi shunday ichki strukturaki, uni har bir til tashuvchisi intuitiv ravishda his qiladi va undan g'ayrihuuriy 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 darajada tavsiflovchi qoidalar sistemasining mavjud emasligi, har qanday tavsiya etilayotgan model nomukammal ekanligi xususida mashhur tilshunos olim K. Fossler shunday yozadi: *«Grammatikani fonetika, morfologiya va sintaksisga bo'lish hech kinga sir emas. Tilni uning shakllanish jarayoni ichida emas, balki uning o'z holati ichida o'rganish zarur. Uning ustida anatomik operatsiya o'tkaziladi. Jonli nutq gaplarga, gap bo'laklariga, so'zlarga, bo'g'inlarga va tovushlarga ajratib chiqiladi. Ushbu metod to'liq o'zini oqlaydi 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 mexanik va majburiy tarzda parchalash edi. Mazkur holat anatomiyadagi holatni eslatadi: anatom o'zining yorish, kesish ishlarini albatta o'z boshimchalik bilan asossiz amalga oshirmaydi, u o'ziga birmuncha qulaylik tug'diruvchi joylarni tanlab oladi. Xuddi shuningdek grammatist ham nutqni tovush, bo'g'in, asos, qo'shimcha kabilarga ajratishda tabiiy bo'lmagan, lekin o'ziga qulay yo'ldan boradi»*.⁴⁵

Yuqoridagilardan kelib chiqib aytish mumkinki, shu kungacha olimlar tomonidan ishlab chiqilgan barcha grammatik modellar o'ta shartli va nisbiydir. Keltirilgan barcha modellar – gap bo'laklari modeli ham, bevosita

⁴⁵ 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.328 – 329.

ishtirokchilar modeli ham, tobelik daraxti modeli ham, transformatsion va generativ grammatika modeli ham mutlaq va benuqson emas. Ularning hammasi til va nutq strukturasi qaysidir aspektlarini aks ettiradi, xolos. Murakkab sistema hisoblangan tilning barcha qirralarini formallashtirishning imkoni yo'q, bunda muayyan cheklavlar mavjud (ayniqsa, tilning semantik aspektida). Bu xususda nemis olimi 1931-yilda K.Gedel shunday yozadi: «*Yetarli darajada mazmunga boy nazariya formallashtirilsa, bu nazariyani formal tizimda to'la aks ettirish mumkin emas. Bunda aniqlanmagan, formallashtirilmagan qoldiq qoladi*».⁴⁶ Demak, har qanday formallashtirishgan model til grammatikasining butun qirralarini qamrab ololmaydi.

Topshiriqlar:

1. Grammatika deganda nimani tushunasiz?
2. Tilshunoslik tarixida grammatikaga bo'lgan yondashuvlar haqida yozing.
3. Mumtoz sintaktik qarashlar haqida ma'lumotlar to'plang.
4. Transformatsion grammatikaning formal tavsif uchun ahamiyatli jihatlari va cheklangan tomonlari haqida mulohaza yuring.
5. Formal grammatika iyerarxiyasining mohiyatini tushuntiring.

Adabiyotlar:

1. Chomsky N. Aspects of the Theory of Syntax. – Cambridge, Mass.: MIT Press, 1965.
2. Хомский Н. Синтаксические структуры // Новое в лингвистике. Выпуск II. – С.412 – 527.
3. Nurmonov A. Lingvistik tadqiqot metodologiyasi va metodlari. – Toshkent: Akademya, 2010.
4. Nurmonov A. Struktur tilshunoslik: ildizlari va yo'nalishlari. – Toshkent, 2008.
5. Теньер Л. 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.

⁴⁶ Анисимов А.В. Компьютерная лингвистика для всех: Мифы. Алгоритмы. Язык. – Киев: Наук. думка, 1991. – С.77.

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 dominant so'zni aniqlash.

Tayanch so'z va iboralar: *kvantitativ metod, matematik parametr, statistika, indeks, tanlama, tanlamaning hajmi, absolut chastota, nisbiy chastota, lisoniy tip, flektiv tillar, agglutinatativ tillar, chastotali lug'atlar, matnning statistik tahlili, Zif qonuni, matn atributsiyasi, anonim va psevdonim matnlar, idiosstil, polisemantik so'zlar, sinonimik qator, dominant, to'plam, lingvostatistika.*

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 fandır. 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 asosida o'rganadi. Bunda statistika hodisalarni turli omillar ta'siridagi

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

o'zgaruvchanlik xususiyatini – variatsiyasini e'tiborga oladi. Shu jihatdan statistik ma'lumotlar hodisalar rivojining dinamikasini aks ettira oladi. Lekin shu 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 uch 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 tahlil etiladi. Matnning statistik tahlili – tilni ehtimollik belgilariga ko'ra tahlil qilish, til hodisalarining statistik xarakteristikalari asosida baholash demakdir. Til birligining gapda 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 qaysi biri eng asosiy konstruksiya 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, dialektologlarning jonli nutqidan yozib olgan materiallaridan tanlanadi. Statistik xususiyati o'rganilayotgan muayyan matnlar «tanlama» deyiladi. **Tanlamaning hajmi** tekshirilayotgan hodisalarning umumiy sonidir, uni «**H**» bilan belgilash mumkin. Masalan, 100 ta so'z; 37 ta bosh kelishik shaklidagi ot kabi. Aniqlanayotgan hodisani ko'rsatadigan miqdor (masalan, ot oldida aniqlovchi vazifasida keladigan takrorlangan otlar miqdori) **absolut chastota** hisoblanadi, uni «**M**» 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 Nisbiy chastota «**P**» harfi bilan belgilanadi. Nisbiy chastota uch xil usul bilan aniqlanadi:

a) oddiy kasr hisobida: $P = M/H$

b) protsent asosida: $P = M \cdot 100/H$

c) promil hisobida: $P = M \cdot 1000/H$

Matnning statistik tahlilida Zif qonunidan o'rinli foydalaniladi. Katta

⁴⁸ Усмонов С. Умумий тилшунослик. – Тошкент: Ўқитувчи, 1972. – Б.199.

hajmdagi matnlarda har bir soʻzning qanchalik tez takrorlanishini (chastotasini) va ularning takrorlanish tezligini hisoblashda **Zif qonuni** amal qiladi.⁴⁹ Bunda soʻzning chastotasi (f), uning roʻyxatdagi holati qatori (r) deb belgilanadi. Zif qonuniga koʻra, matnda eng koʻp ishlatilgan soʻz $r = 1$, undan pastroq chastotada kuzatilgan soʻz $r = 2$, undan keyingisi $r = 3$ tarzida belgilanadi. Bundan kelib chiqadiki, matndagi ixtiyoriy soʻz uchun (f) chastotaga (r)ning koʻpaytmasi natijasi oʻzgarmas miqdor – konstanta (C)ga teng boʻladi ($r \cdot f = C$), bu yerda C matn hajmiga bogʻliq. Masalan, amerikalik yozuvchi Mark Tvenning «Tom Soyerning sarguzashtlari» asari Zif qonuni asosida tahlil etilgan. Asarda jami 71 370 ta belgi mavjud boʻlib, takrorlanmagan soʻz tiplari 8018 ta. Matnda ishlatilgan soʻzlarning oʻrtacha takrorlanish chastotasi 8,9 dir, yaʼni matndagi soʻzlar taqriban 9 marta takrorlanadi. Lekin bunda eng katta muammo shuki, soʻz tiplarining barchasi matn ichida qoʻllanish darajasi bir xil emas. Ayrim soʻzlar 700 marta takrorlanadi, bunday soʻzlar romanning 1% qismini tashkil etadi. Yana shunday soʻzlar ham borki, ular bir marta takrorlanadi. Bunday soʻzlar *hapas legomena* (yunoncha soʻz «faqat bir marta oʻqi», degan maʼnoni anglatadi) deb ataladi, ular asarning qariyb yarmini tashkil etadi. Qariyb 90% soʻzlar 10 marta yoki undan kam, 10% ga yaqin soʻz tiplari 3 marta yoki undan kam miqdorda takrorlanadi.

6-jadval

Word (Soʻz)	Frequency (Chastota)	Rank (Qator)	$f \cdot r$
the	3332	1	3332
a	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.

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 idiosstilini – o'ziga xos uslubini ko'rsatuvchi vositalar) lingvostatistik tahlil yordamida aniqlanadi. Turli matnlardan olingan dalillarning qiyosiy tahlili bizga o'sha matnning mazmun-mundarijasini, matn tuzilgan davrni, dalillarni argumentlash xarakterini va hatto mualliflikni aniqlashga imkon beradi. Badiiy asarlarning sujet qurilishiga, matn strukturasi 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 muallifi) 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 umumlashtirgan. 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 lingvistik fakt bilan uzviyligi yaqqol ko'zga tashlanmaydi. Bu esa tasnifni amaliyotga*

⁵⁰ Баранов А.Н. Введение в прикладную лингвистику. – М.: Эдиториал УРСС, 2001. – С.24.

tatbiq etib tekshirishni taqozo etadi».⁵¹ Shu nuqtayi nazardan olim matematika 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, ingliz, suaxili, eskimos tillariga) tatbiq qilgan. Bunda olim o'zi belgilagan formulalar asosida lisoniy tiplarning va tillarda kuzatiladigan tipologik belgilarning miqdoriy tavsifini, statistik ma'lumotlarni ilmiy jamoatchilikka havola etdi.

Masalan, sintetiklikning dunyo tillaridagi darajasini aniqlash uchun **M/W**, agglutinatsiyani aniqlash uchun **A/J**, kompozitsiyani aniqlash uchun **R/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 bo'lgan 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 etgan metod tilshunoslikda mavjud bo'lgan kartotekalash usulidan ishonchliroq va aniqroqdir.* (7-жадвал)

Chastotali lug'atlar matnda so'zlarning ishlatilish chastotasi (qay darajada ishlatilishi, tez-tez ishlatilishi)dan kelib chiqqan holda tayyorlanadigan so'zliklar hisoblanadi. 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 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 flektiv 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.

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	2,17	2,55	1,06	1,68	3,72
Agglutinatsiya	0,09	0,11	0,34	0,51	0,67	0,00	0,30	0,03
Qo'shma so'z yasash	1,13	1,00	1,03	1,02	1,00	1,07	1,00	1,00
Derivatsiya (so'z yasash)	0,62	0,20	0,10	0,35	0,07	0,00	0,15	1,25
Relatsiya (so'z o'zgartirish)	0,84	0,90	0,39	0,82	0,80	0,00	0,53	1,75
Prefiksatsiya	0,16	0,06	0,01	0,00	1,16	0,00	0,04	0,00
Suffiksatsiya	1,18	1,03	0,49	1,15	0,41	0,00	0,64	2,72
Izolatsiya	0,16	0,15	0,53	0,29	0,40	1,00	0,75	0,02
Sof fleksiya	0,46	0,47	0,29	0,59	0,19	0,00	0,14	0,46
Moslashuv	0,38	0,38	0,19	0,12	0,41	0,00	0,11	0,38

8-jadval

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

Tip	Afrika	Yevrosiyo	Avstraliya	Shimoliy Amerika	Janubiy Amerika
SOV	22	26	29	26	18
SVO	21	19	6	6	5

tashkil etsa, bu ko'rsatkich flektiv tillarda 62 – 87 foiz atrofida bo'ladi. 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 formalarining 50 foizini tashkil etsa, bu ko'rsatkich uchun agglutinativ konstruksiyali turkiy tillarda 600 – 1000 ko'p uchraydigan so'z formalari zarur bo'lar ekan, buning asosiy sababi agglutinativ va flektiv tillar morfologik strukturasini va so'z formalarining tuzilishidir.

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. **БЗ**- old qo'shimchasi 74 har xil fe'l, 27 sifati, 14 ot va faqat 2 ravishda uchragan. **БЕЗ**- old qo'shimchasi asosan ot, sifati va ravish yasashda ishlatilar ekan, fe'l da esa atigi 4 martagina ishlatilgan, xolos. Alfavit-chastotali lug'atlar tilda yoki qardosh tillarda va ularning sohalarida u yoki bu so'z paradigmasining amalga oshirishini 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, sifati, ravish va son tillardagi jami so'zlarning 70% ni tashkil qilgani holda, fe'l 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 lug'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 Shteynfeld 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 tilla 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.

6. 1981-yilda R.Qo'ng'urov va S.Karimovlar tomonidan «Zulfiya poeziyasining 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 «Sinchalal» povesti tilining chastotali lug'ati» yaratildi.

10. 1987-yilda falsafa fanidan chastotali lug'at yaratildi.

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

Chastotali lug'atlarning kompyuter lingvistikasi uchun ahamiyati quyidagilarda ko'rinadi:

– birinchidan, chastotali lug'atlar avtomatik qidiruv tizimi asosida ishlaydigan elektron lug'atlar yaratish uchun, ularning sohaviy turlari uchun tayyor lingvistik ta'minot vazifasini o'taydi;

– ikkinchidan, chastotali lug'atlar til o'qitishni avtomatlashtirish jarayonida, kompyuter lingvodidaktikasi sohasida metodologik ahamiyat kasb etadi;

– uchinchidan, chastotali lug'atlar chet tillarini o'qitishning samaradorligini 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 asosda 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 yuqoridagi 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 – تعلم olim – عالم ilm – علم

Arabcha so'zlar uchun umumiy o'zak – م ع ل harflaridir.

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

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: | «Ko'z»ning semalari: |
| A. Odamning ko'zi. | 1. Ko'rish organi. |
| B. Uzukning ko'zi. | 2. Shaffoflik belgisi. |
| C. Buloqning ko'zi. | 3. Cheti botganlik. |
| D. Derazaning ko'zi. | 4. O'rtasi bo'rtganlik. |
| E. Ishning ko'zi. | 5. Ochilishi va yumilishi. |
| F. Yog'ochning ko'zi. | 6. Asosiyligi. |
| G. Qalb ko'zi. | 7. Sezish (ko'rish) organi |

9-jadval

№	1	2	3	4	5	6	7
A to'plam	+	+	+	+	+	+	+
B to'plam	-	+	+	+	-	-	-
C to'plam	-	-	-	-	+	-	-
D to'plam	-	+	-	-	+	-	-
E to'plam	-	-	-	-	-	+	-
F to'plam	-	-	+	+	-	-	-
G to'plam	-	-	-	-	-	-	+

Sinonimik qatorda dominanta (bosh, asosiy) so'zni aniqlash ham yuqorida berilgan to'plam metodi asosida amalga oshirilishi mumkin. Bunda turli nutq uslublardagi matnlar tahlilga tortiladi va barcha uslublarda ishlatilishi kuzatilgan so'z o'sha sinonimik qatorda dominanta so'z hisoblanadi.

Topshiriqlar:

1. Statistik tahlil nima uchun zarur?
2. O'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.

2. Пиотровский Р.Г. Текст, машина, человек. – Л.: Наука, 1975.
3. Гринберг Дж. Квантитативный подход к морфологической типологии языков // Новое в лингвистике. – М., 1963. Вып. III. – 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. IV. – М.: Прогресс, 1965.
7. Усмонов С. Умумий тилшунослик. – Тошкент: Ўқитувчи, 1972.
8. Баранов А.Н. Введение в прикладную лингвистику. – М.: Эдиториал УРСС, 2001.

КОМПЬЮТЕР ЛИНГВИСТИКАСИ ТАРИХИ ВА ЗАМОНАВИЙ HOLATI. КОМПЬЮТЕР ЛИНГВИСТИКАСИ БО‘ЙИЧА 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 o‘qitish tizimi, matnning statistik tahlili, lingvostatistika, injener lingvistikasi, chastotali lug‘at, o‘qituvchi lingvistik avtomat, lingvistik ta‘min, lingvistik programmalar.*

Kompyuter lingvistikasi fani 1960-yilda AQSHda shakllandi. Mazkur fan XX asrning o‘rtalarida paydo bo‘lgan matematik lingvistika fani asoslaridan kelib chiqqan. Tabiiy tillar grammatikasi formal modeli kompyuterda tarjima qilish, tillarga o‘rgatish, bilimlarni diagnostika qilish, matnlarni tahrir etish kabi amaliy dasturlarni ishlab chiqishga asos bo‘lgani holda kompyuter lingvistikasi yo‘nalishiga zamin yaratdi.⁵² Kompyuter lingvis-

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

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 tarjima uchun mo'ljallangan dasturlarni ishlab chiqish kabilar kiradi. Kompyuter lingvistikasi bir-biri bilan chambarchas bog'liq bo'lgan 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 avtomatlashtirish – xorijiy tillarni kompyuter yordamida o'qitish tizimini ta'minlovchi amaliy dasturlar hamda bilimlarni avtomatik baholash mexanizmlari;

d) kompyuter leksikografiyasi (elektron lug'atlar tuzish) – muayyan til leksikonidagi so'zlarning ma'lumotlar bazasi va maxsus programmalashtirish 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 ishlaniishi;

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 Linguistics**), bu uyushma kompyuter tilshunosligi bilan bog'liq ko'plab tashkiliy va ilmiy ishlarni amalga oshiradi. AQShda ikki yilda bir marta COLING konferensiyasi o'tkaziladi. AQShning 10 ta universitetida kompyuter lingvistikasi

⁵³ Новое в зарубежной лингвистике. Вып.24. Компьютерная лингвистика. – М.: Прогресс, 1989. – С.10.

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 o'tkaziladi.⁵⁴ Shuningdek, Niderlandiyada «**Artificial Intelligence**» («Sun'iy intellekt») jurnali ham chop etiladi. Mexiko shahrida har yili 10 yildan beri «**Computational Linguistics and Intelligent Text Processing**» nomli xalqaro konferensiya o'tkazib kelinadi.⁵⁵

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

- mashina tarjimai (MT);
- avtomatik tahrirlash (AT);
- til o'rgatish jarayonini kompyuterlashtirish (avtomatik til o'qitish – АТО');
- statistik tadqiqotlar (ST).

Kompyuter lingvistikasining mashina tarjimai yo'nalishida amalga oshirilgan tadqiqotlar bugungi kunda barcha kompyuterlarda mavjud bo'lgan ruschadan boshqa tillarga tarjima qiluvchi dasturlarning yaratilishiga asos bo'ldi. Xususan, L.L.Nelyubin va uning shogirdlari tomonidan ishlab chiqilgan so'zma-so'z tarjima qilishning ilmiy strategiyasi (rasmidoraviy 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 jihatlari ta'kidlash lozim: birinchidan, ushbu ishlarda so'z va iboralarning avtomatik lug'ati bilan birgalikda to'liq tarjima qilish imkoniyatini beruvchi morfologik-sintaktik algoritm-freym yaratishga harakat qilingan; ikkinchidan, vujudga keltirilgan tarjima dasturlari asosida o'qituvchi lingvistik avtomat (ОЛА – Обучающий 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 tarjimai uchun (ingliz tili bo'yicha) algoritmlar va dasturlar ishlab chiqish asosiy o'rinni egalladi. L.L.Nelyubin mashina tarjimasida ingliz tili leksikasini butunligicha qamrab olmagan, u o'z tadqiqotlarining obykti 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>

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

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 tarjimaning 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 «текст-текст» 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 hana 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 «человек-машина- человек)», V.V.Goncharenkonging «Лексикографические, лингвостатистические и инженерно-лингвистические вопросы постро-

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 matnlarni avtomatik tarzda boshqa tillarga tarjima qiluvchi dasturlarning vujudga kelishi uchun asos bo'ldi. O'zbek tilida ham shu kabi ma'lumotlar bazasini yaratish bugungi kundagi eng dolzarb vazifadir.

Rus kompyuter tilshunosligida taraqqiy qilgan yana bir yo'nalish matnlarning 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 etilgan 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 «Тезаурус 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 maqolalari 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

noto'g'ri yozilgan so'zning tagiga qizil chiziq chiziladi, so'z to'g'ri yozilganidan so'ng chiziq yo'qoladi. Ko'rinadiki, bu dastur asosan imloni tekshirishga mo'ljallangan, unda uslubiy xatolarni aniqlash ko'zda tutilmagan. Kelajakda yaratiladigan dasturlar, umid qilamizki, avtomatik tahrirning bu jihatlarini ham e'tibordan chetda qoldirmaydi. Ta'kidlash lozimki, bunday dasturlarni ishlab chiqish uchun til birliklarining nafaqat fonetik, semantik, morfologik xususiyatlari, balki aynan sintaktik-valentlik xususiyatlari 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, injener-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 vazifalarni kompyuter yordamida yechish, til o'rgatish jarayonini optimallashtirish, o'qituvchi lingvistik avtomat (O'LA) uchun nutqiy materiallarni tayyorlash kabi masalalar tadqiq qilingan.⁵⁷

Tadqiqotlar natijasida ko'plab O'LAlar uchun lingvistik ta'min yaratilgan bo'lib, bu dasturlar rus tilini o'qitish jarayonini optimallashtirishga xizmat 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 avtomatlashtirilgan o'quv kurslarini vujudga keltirishga qaratiladi. AO'S esa o'quv jarayoni ishtirokchilariga nafaqat taqlid qilishi, balki ularning intellektual faoliyatini 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 turadigan so'zlarinigina o'zlashtirishga kirishiladi va muntazam ravishda bosqichma-bosqich so'z boyligi oshirib boriladi. Buning uchun esa leksikostatistik ma'lumotlar asosiy poydevor vazifasini o'taydi. L.N.Zasorinaning ta'kidlashicha, leksikostatistikaning markaziy muammosi jonli (funktional) tilning statistik qonuniyatlarini va matnning statistik strukturasini aniqlash-

⁵⁷ Пиотровский Г.Г. Компьютеризация преподавания языков. – Л.: ГГПИ, 1988. – С.74.

dir. Matnning statistik strukturasi deyilganda, shartli ravishda, ma'lum matndagi 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 lingvistikasida 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.Piotrovskiyning «*Основы статистической оптимизации преподавания иностранных языков*» (1974), N.D.Andreyevning «*Статистико-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 pomzodlik dissertatsiyalari; M.Ayimbetovning «*Проблемы и методы количественно-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.V.Bektayevning «*Статисти-*

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 tarjimasini, 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 «Абай 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 lingvistikasiga 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 bo'lib, 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.

O'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; Пиотровский 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. Опыт-промышленная система лексического машинного перевода. – МСМП, 1979.

ga oshirgan. H.Arziqulov o'zining «Информатика и переработка текста средствами вычислительной техники» nomli monografiyasida kompyuter yordamida katta hajmdagi matnlarga ishlov berishning avtomatik tizimi asoslarini chuqur tahlil qilgan. Eng muhimi, professor H.Arziqulov rahbarligida 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 tarjimai dasturini yaratish sohasida izchil ilmiy-tadqiqot ishlarini boshqargan. O'zbekcha matnlarga qayta ishlov beruvchi avtomatik tizim «Uzlington», lotin alifbosidagi matnlarni kirill yozuviga va lotin alifbosidan kirill alifbosiga o'tkazuvchi «Spellchecker», kompyuter ko'magida korrektura qiladigan, imloviy xatolarni tuzatadigan kompyuter dasturlarini yaratishga birinchilardan bo'lib qo'l urgan.

Hamdam Orziqulov o'zbek mashina tilini yaratish maqsadida so'z tarkibini nihoyatda chuqur tadqiq qildi: morfema va so'z yasalish strukturalari haqidagi fikrlari, matn statistikasi va undan chastotali lug'atlar tuzish prinsiplari, 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 asoschidir, deyish mumkin.

S.Muhamedov o'zining R.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 bo'lgan o'zbekcha matnlarning leksik – morfologik tuzilishi statistik tahlili natijalari keltirilgan. Ya'ni uning «Статистический анализ лексико-мор-

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

фологической структуры узбекских газетных текстов» (nomzodlik dissertasiyasi) va «O'zbek tilining alfavitli-chastotali lug'ati (gazeta tekstlari asosida)» asarlari yuqoridagi asarning yozilishida manba bo'lib 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.Mirtojiev, 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 leksikasi va grammatikasiga oid tadqiqotlariga suyanadi. Ulardagi ilmiy-nazariy fikrlarga asoslanadi, ulardan oziqlanadi.⁶⁰

So'nggi yillarda o'zbek kompyuter lingvistikasi asta-sekinlik bilan shakllanmoqda. 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 «Компьютер tilshunosligida avtomatik tahrir qiluvchi dasturning lingvistik ta'minotini yaratish asoslari» nomli maqolalari, shuningdek, A.Po'latov va S.Muhamedovalarning «Компьютер lingvistikasi» 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.

tarjima, analiz va sintez, tabiiy tilning qayta ishlanishi, kompyuter leksikografiyasiga oid dasturlarni o'z ichiga oladi.

10-jadval

1	<u>Link Grammar Parser</u> John Lafferty, Daniel Sleator, Davy Temperley Carnegi Melon University, USA.	Ingliz tilining sintaktik parser dasturi, uning bazasiga 60 000 so'z shakllari kiritilgan.
2	<u>C#237;bola Oleada loyihalari</u> Computing Research Laboratory (CLR) New Mexico State University, USA.	Mazkur dastur matnlarning lingvistik analizida ishlatiladi. Bu dastur yordamida 16 ta tildagi matnlarning avtomatik tarjimasini, statistik tahlilini amalga oshirish mumkin. Dastur tarkibida turli xil lug'atlar hamda tezauruslar ham mavjud.
3	<u>Russian Morphological Dictionary</u> Sergey Sikorsky	Ruscha morfologik lug'at rus tilidagi matnlarning sintaktik va morfologik analizi uchun ishlatiladi. Unga 120 000 so'zning morfologik lug'ati kiritilgan.
4	<u>Морфологический парсер Яndex</u>	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	<u>Лингвоанализатор</u> Д.В.Хмелев	Matn strukturasi 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 ularning matnga yaqin keladigan asarlari chiqarib beriladi.
6	<u>Система StarLing</u>	Ushbu tizim katta hajmdagi ko'p tilli matnlar bilan, transkripsion belgilar bilan

6	<p><u>Система StarLing</u></p> <p>С.А.Старостин СУБД</p>	<p>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.</p>
7	<p><u>Морфологический анализатор</u></p> <p>С.А.Старостин</p>	<p>Rus, ingliz tillaridagi so'zlarning morfologik analizi dasturi so'zlarning asosiy shakllari hamda morfologik ma'lumotini menyuga chiqaradi.</p>
8	<p><u>MonoConc</u></p> <p>Michael Barlow Dept of Linguistics, Rice university, Texas, USA.</p> <p>2 версии (MonoConc Pro 2.0 и MonoConc 1.5)</p>	<p>Tanlangan simvolli matnlar korpusi uchun konkordans yaratishda, so'zliklar va indekslar generatsiyasida foydalaniladi.</p>
9	<p><u>WordSmith Tools</u></p> <p>Mike Scott</p>	<p>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 paketti tahririni amalga oshiradi.</p>
10	<p><u>Concordance 2.0.0</u></p> <p>R.J.C. Watt, 2000</p>	<p>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.</p>
11	<p><u>TextAnalyst 2.0</u></p> <p>Научно-производственный инновационный центр «МикроСистемы»</p>	<p>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 analiz qilinadi, matn avtomatik referatlanadi.</p>

		<p><u>TextAnalyst SDK</u> – ingliz va rus tillari uchun lemmatizatsiya (so'zlarni me'yoriy shaklga keltirish), tushunchalarning chastotali ro'yxatini yaratish, kontekstda so'zlarni qidirish kabi ishlarni bajaradi.</p> <p><u>TextAnalyst Lib</u> – gipertekst texnologiyasi asosida elektron kitoblar yaratishda foydalaniladi.</p>
12	<u>Galaktika-ZOOM</u> корпорация Галактика, Москва	Informatsiyani analitik qayta ishlash va avtomatik qidiruvda ishlatiladi.
13	<u>Система Пропись 4.0</u> 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	<u>Textual Analysis Computing Tools (TACT)</u> 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 mo'ljallangan dastur sanaladi.
15	<u>Худломер</u> Леонид Делицын	Mazkur dastur rus tilidagi matnlarni uslubiga ko'ra avtomatik tasniflashga mo'ljallangan. Bunda 4 xil matnlar korpusi (badiiy asar, publitsistika, ilmiy maqolalar va dialoglar) tahlil etiladi, natijada stilga bog'liq 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

		СТАТЬЯ» (ILMIY MAQOLA) deb tasniflanadi.
16	«Штампомер» Леонид Делицын	Bu dastur mualliflik matnlarida tez-tez ishlatiladigan iboralar (shtampilar), o'sha muallif tomonidan ko'p ishlatiladigan so'z va ifodalarni aniqlashga yordam beradi.
17	<u>Свежий взгляд</u> Fresh Eye версия 1.21, 1995 Дмитрий Кирсанов	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, tavnologiyaning oldini oladi.
18	<u>RussianWord Constructor (RWC)</u> версия 1.0, 1992 Дмитрий Кирсанов	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 okkazonalizmlarni hosil qiladi.
19	<u>Scott Pakins automatic complaint-letter generator</u> Scott Pakin	Bu dastur muayyan shaxs yoki tashkilotga shikoyat matnini generatsiya qilishga yordam beradi. Dasturga 7000 leksik unsumni o'z ichiga olgan lug'at bazasi kiritilgan bo'lib, u ingliz tilida on-line rejimida ishlaydi.
20	<u>Cyrano ServerNando.net</u>	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	<u>Письмовник</u> MediaLingua Ltd., 1995 – 2002	Ish hujjatlari va yozishmalarni avtomatik tarzda shakllantiradi hamda ularni rus tilidan ingliz tiliga, ingliz tilidan rus tiliga tarjima qiladi.
22	<u>ALICE</u> Dr.Richard S.Wallace. ALICE AI Foundation, USA	Bu dastur gapiruvchi programma bo'lib, u 2000 va 2001-yillarda Lobner mukofotiga (Loebner prize) loyiq topilgan. Dastur AIML (sun'iy intellekt uchun maxsus til) texnologiyasiga asoslanadi.
23	<u>ANANOVA</u>	Buyuk Britaniyadagi virtual diktir asosida ishlaydigan axborot agentligi bo'lib, u

	Ananova Limited United Kingdom	orqali kunning yangiliklarini bilib olish mumkin.
24	<u>Sakrament Text-to-Speech Engine v2.0</u> компания «Сакрамент», 2002	Rus tilidagi nutq sintezi tizimi bo'lib, u Microsoft Speech API (SAPI) 5.1. standartiga to'liq mos keladi, erkak va ayol ovozida matnlarni talaffuz qiladi, foydalanuvchining xohishiga ko'ra intonatsion pauzalar qiladi, nutq tembri va tonini o'zgartiradi.
25	<u>Аки Росс</u> Альберт Товмасын, 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. Dasturning bilimlar bazasi juda katta emas, lekin uni AIML tili yordamida to'ldirib borish mumkin yoki dastur saytidan bilimlar bazasini yangilab turish mumkin.
26	<u>Programmable Artificial Intelligence (PAIv3.01) или Программируемый Искусственный интеллект</u> Claudio 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 materiallarini Internetdan yuklab oling.
2. Kompyuter lingvistikasi fani bilan shug'ullangan olimlar va ularning 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.

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. Po'latov A., Muhamedova S. Kompyuter lingvistikasi. – Toshkent, 2007.
6. Jurafsky D., Martin J.H. Speech and Language Processing. – New Jersey, 2000.
7. Арзиқулов Ҳ. Ўзбек лингвистик автмати // Мулоқот. 2002. №1. – Б.20 – 21.
8. Сафаров Ш., Йўлдошев Б. Компьютер лингвистикасини биласизми? // Моҳият. 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, ekvivalent moslik, transformatsion moslik, tarjima protsedurasi, generatsiya, trans-latsiya, bilingv bazasi, example-based translation (namunalarga asoslana-digan tarjima), rule-based translation (algoritimga asoslangan tarjima), translation memory (tarjimon xotirasi).*

Informatsiyaning geometrik progressiya asosida keskin ko'payishi ularning 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-

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 Bebbidj tomonidan 1836 – 1848-yillarda ilgari surilgan. Uning fikricha, mexanik-elektron mashinalar tillar o'rtasidagi struktur o'xshashliklar asosida kodlashtirilgan avtomatik tarjimani amalga oshirishi mumkin. Uning loyihasi bo'yicha 50 mingta so'z xotira kuchiga ega bo'lgan elektron mashinalar 100 000 ta so'zni avtomatik ravishda tarjima qilishi zarur bo'ladi.⁶¹

1949-yilda Ch.Bebbidjing g'oyalaridan ta'sirlangan amerikalik olim Uorren Uiver mashina tarjimasini yaratishning nazariy asoslarini ishlab chiqdi hamda lingvistika, deshifrovka, programmashtirish nazariyasi bo'yicha mutaxassislarni hamkorlikka chorladi.⁶² Shundan so'ng AQSH-da MT bo'yicha kollaboratsiyalar, hamkorlik loyihalari yaratildi. Mashina tarjimasini bo'yicha birinchi xalqaro konferensiya 1952-yilda Massachusetstexnologiya 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.⁶³

Rossiyada mashina tarjimasini 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 Ittifoqda mashina tarjimasini bo'yicha birinchi tajriba 1955-yilda o'tkazildi. Shundan so'ng O.S.Kulagina va I.A.Melchuk rahbarligi ostida Amaliy matematika institutida mashina tarjimasini 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 А.Н. Введение в прикладную лингвистику. – М.: Эдиториал УРСС, 2001. – С.89.

⁶³ Убин И.И. Современные средства автоматизации перевода: надежды, разочарования и реальность // Перевод в современном мире. – М.: ВЦП, 2001. – С.60 – 69.

⁶⁴ Панов Д.Ю., Ляпунов А.А., Мухин И.С. Автоматизация перевода с одного языка на другой. Сессия по научным проблемам автоматизации производства. – М.: Изд-во АН СССР, 1956.

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 tashkil qilindi.

1966-yilda AQShning fan masalalari bo'yicha kongress qo'mitasida avtomatik tarjimaga oid ma'ruza muhokamasi (**Automatic Language Processing Advisory Committee – ALPAC REPORT**) bo'lib o'tdi.⁶⁵ Unda avtomatik tarjima bo'yicha «*Language and Mashines: Computers in Translation and Linguistics*» («*Til va mashinalar: tilshunoslikda va tarjimada kompyuterlar*») deb nomlangan salbiy ma'nodagi ma'ruza qilindi va yuqori darajada sifatli avtomatik tarjimani ta'minlovchi universal tizim yaratish natijasizligi va imkonsizligi ta'kidlandi. Shuningdek, ma'ruzada ko'rsatilishicha, «tabiiy tarjima» (inson tarjimasi) mashina tarjimasidan arzonroq va sifatliroq. 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 izlanishlarni mablag' bilan ta'minlab turadi.

AQSH harbiy havo kuchlari buyurtmasi bo'yicha D.Tom rahbarligida SYSTRAN avtomatik tarjima sistemasi ishlab chiqildi. Mazkur tizim 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. 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**» loyahasini ishlab chiqdi. Uning tarkibiga kiruvchi 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 Слюкун 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.

mln ta soʻz xotira kuchiga ega **YEUROTRA** sistemasi – Yevropa olimlari tomonidan ishlab chiqilgan. U barcha yevropa tillari boʻyicha avtomatik tarjimani amalga oshira oladi. **SOKRAT** sistemasi 40 mln ta soʻz xotira kuchiga ega. Dunyodagi 40 ta tilda avtomatik tarjimani amalga oshiradi. Yaʼni qaysi tilda Windows operatsion tizimi versiyasi mavjud boʻlsa, shu tilda avtomatik tarjimani amalga oshira oladi. Lingvo sistemasi (lingvo kompyuter lugʻati) – bu lugʻatga ham dunyoning 40 tilidagi maʼlumotlar jamlangan. 1990-yillarda Amerikada **Global Link**, Rossiyada **PROMT** avtomatik tarjima tizimlari yaratildi.

Internet tizimining rivojlanishi natijasida avtomatik tarjima global tar-moqda ham tashkil etildi, yaʼni hozirgi kunda **online-tarjima tizimi** ham faol yoʻlga qoʻyilgan.

Avtomatik tarjimaning kamchiliklari:

1. Soʻzlar avtomatik tarjima jarayonida toʻgʻridan-toʻgʻri tarjima qilinadi. Shuning uchun tarjima grammatik qoidalarga deyarli rioya qilin-magan holda amalga oshiriladi.

2. Avtomatik tarjima qilingan matnni uslub jihatdan tahrirlab chiqish zarur boʻladi.

3. Hozirgi mavjud avtomatik tarjima qiluvchi sistemalar soʻzlar biriku-vini ham deyarli notoʻgʻri tarjima qiladi.

4. Avtomatik tarjima polisemiya, omonimiya, troplar, frazeologizmlar, idiomalarni eʼtiborga olmaydi.

Masalan, *The bat looks like a mouse with wings* gapini rus va oʻzbek tiliga tarjima qilishda har bir soʻzning polisemantik va omonimlik kabi se-mantik qirralarini eʼtiborga olish zarur. Aks holda tarjima muvaffaqiyatli chiqmaydi. Jumladagi *bat* soʻzi «koʻrshapalak», «belkurak», «toʻqmoq», «kichik ziyofat» kabi maʼnolarni, *look* soʻzi esa «qaramoq», «nigoh», «koʻrinish» maʼnolarini anglatadi. Bunda mazkur soʻzlar rus yoki oʻzbek tiliga tarjima qilinayotganda mos ekvivalentini tanlash uchun kontekst va uning distributsiyasidan (qurshovidan) kelib chiqish zarur. Ushbu holat, ayniqsa, idiomatik ifodalar qatnashgan matnlar tarjimasida murakkablik tugʻdiradi. Masalan, *You are an apple of my eyes* jumlasida «Siz koʻzlarimning olmasisiz» tarzida soʻzma-soʻz tarjima qilinadi, chunki «apple» soʻzining denotative maʼnosi «olma»dir. Lekin kontekstdan kelib chiqib tarjimaga yondashilsa, toʻgʻri tarjima yuzaga chiqadi. Bunda jumlaning «an apple of my eyes» qismi idioma ekanligini nazarda tutish lozim va u «Siz koʻzimning oq-u qorasisiz» tarzidagi gap hosil boʻladi.

Tarjimadagi moslik 3 xil boʻladi:

1) ekvivalent moslik – bunda bir tildagi soʻz boshqa bir tildagi soʻzga

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 сов

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 point

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

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 shakli 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 tipologiyadagi farqlarni izchil o'rganish, lingvistik universalialarni aniqlash, tillardagi frazeologizmlar, idiomalar, paremlar, 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 pragmatik 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, sinonimiyaning e'tiborga olinishi, jumalarning 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. Tizimning 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,

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.

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 beruvchi birliklar joylashgan bo'ladi. Bunda so'zma-so'z tarjima amalga oshirilib, ikki tildagi gaplar strukturasi yagona superstrukturaga birlashtiriladi. Bunda avtomatik tarjima modeli ikki holatda bo'ladi: a) *generatsiya holati* tarjimaning ishlashini – lug'atlar jamlash, analiz algoritmlari, aniq bir maqsadga yo'naltirilgan sintezni ta'minlaydi; b) *translatsiya holati* sistemaning mustaqil ravishda tarjima ustida ishlashini ta'minlaydi.

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

1-protsedura – qidiruv tizimi bilan bog'liq bo'lib, bunda lug'at bazasidan 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.

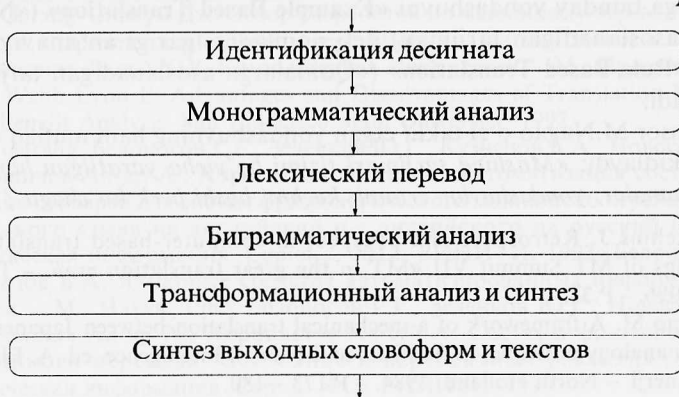
3-protsedura – bir yoki ko'p ma'noli so'zlar tarjimasi bayon qilinadi.

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

5-protsedura – transformatsion o'zgarishlar (matn yoki jumla strukturasi 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 bo'lib, bunda murakkab grammatik formalarning sintezi amalga oshiriladi.

12-jadval



Mashina tarjimasi strategiyasi sekin-astalik bilan o'zgarib borgan. Ilk tajribalarda qo'llanilgan 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 samaradorligiga 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 tipologik 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 aralashadi.⁶⁷

Matnni bir tabiiy tildan boshqasiga tarjima qilish jarayonidagi murakkabliklar 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 qilinishi lozim. Buning uchun kompyuterga kiritilgan katta hajmdagi **biliniv bazasi** (matn va uning tarjimalari) shakllantirilishi zarur. Yangi matnlar tarjimasida biliniv 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, ilgari an'anaviy yondashuv «**Rule Based Translation**» («Qoidalarga asoslanadigan tarjima») deb ataladi.

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

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

⁶⁸ 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.

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 elementlardan ma'lumotlar bazasi tayyorlanadi. Ma'lumotlar bazasi qidiruv tizimiga yuklanadi, u esa tarjimasi mavjud bo'lgan matn fragmentlarini va alohida jummalarni sifatli tarjima qilishga imkon beradi. 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 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.

Rossiyada professor G.G.Belonogov frazeologik mashina tarjimasi konsepsiyasini 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 E. 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. ВИНТИ. 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. ВИНТИ. 1998.

alohida soʻzlar yordamida emas, balki soʻz birikmalari yordamida ifodalangani. Oʻz-oʻzidan kelib chiqadiki, matnni bir tildan boshqasiga tarjima qilish jarayonida ham mazmuni yuzaga chiqaruvchi asosiy til birliklari ham alohida soʻzlar emas, balki tipik vaziyatlarni ifodalovchi frazeologik birikmalar (bu yerda frazeologik birlik keng maʼnoda tushuniladi) hisoblanadi. Mazkur yondashuvga muvofiq matnda tez-tez uchraydigan soʻzlar, iboralar, soʻz birikmalari, nutqiy etiket namunalari tarjima ekvivalentlari bilan boyitilgan maʼlumotlar bazasi frazeologik mashina tarjimasini tizimining asosini tashkil etadi. Bunda soʻzlarning distributsiyasiga eng koʻp eʼtibor beriladi. Bu metod polisemantik soʻzlarning tarjimada mos keluvchi toʻgʻri ekvivalentini tanlashga, omonim soʻzlarning kontekstdagi holatini aniqlashga, 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 moʻljallangan 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 koʻchirib ishlash)ga asoslanib ishlaydi. RETRANS tizimi tabiiy va texnika fanlari, iqtisodiyot, siyosat, qonunchilik hamda harbiy sohaga oid 4 mln lugʻat maqolasiga ega boʻlgan politematik soʻzlik bilan taʼminlangan. Kiritilgan til birliklarining 80% qismini soʻz birikmalari tashkil etadi, ular 2 ta soʻz uzunligidan tortib to 17 ta soʻz uzunligiga ega boʻlgan birikmalardir. Tizim tarkibida 400 ming lugʻat maqolasiga ega boʻlgan qoʻshimcha tematik lugʻat ham mavjud.

RETRANS tizimining ish tartibini koʻrib chiqaylik. Ruschadan inglizchaga tarjima qilish jarayonida dastlab matn avtomatik tarzda morfologik, soʻng semantik-sintaktik tahlil qilinadi, frazeologik birliklar aniqlanadi. Matnning barcha soʻzlari hamda frazeologik birliklari qidiruv obrazlari shakllanadi. Lugʻat bazasidagi joylashuv tartibiga koʻra soʻz va birikmalar qidiruv obrazlarining xeshkodlari topiladi, ularning dekodlanishi asosida qidiruv natijalari namoyon boʻladi. Keyingi bosqich matnning har bir fragmentiga mos yagona yoki bir nechta tarjima ekvivalentlari tanlanadi. Tarjima natijalarini quyidagi jadvalda yaqqol koʻrish mumkin.

Matnning rus tilidan ingliz tiliga qilingan tarjima oraliq natijalari fragmenti

Перевод 00001 00002 text translation
текстов 00002

	с 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 / translator					
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						

Shundan soʻng tarjima ikki rejimda amalga oshiriladi: 1) avtomatik tarjima rejimi; 2) oraliq tarjima natijalarini intellektual tahrirlash imkoniyati mavjud boʻlgan interaktiv rejim. Avtomatik tarjima rejimida ruscha matndagi har bir soʻz va birikma uchun faqat bitta (odatda, eng asosiy, sinonimik qatordagi dominantlik xususiyatiga ega boʻlgan soʻz) inglizcha ekvivalenti tanlanadi va matnning morfologik, semantik-sintaktik analizi asosida sintez jarayoni boshlanadi, soʻng gap strukturasi shakllanadi va tarjima yuzaga chiqadi. Interaktiv rejim ham xuddi avtomatik tarjima singari boʻladi, bunda faqat foydalanuvchi (tarjimon-muharrir) taklif etilgan tarjima variantlarining semantik jihatdan eng mos keluvchi ekvivalentini oʻzi tanlash imkoniga ega boʻladi, shuningdek, gap strukturasi, soʻz tartibini ham lingvistik bilimidan kelib chiqib oʻzi tahrir qilishi mumkin.

Shuni alohida taʼkidlash zarurki, yuqoridagi har uchala konsepsiya ham avtomatik tarjima sifati va samaradorligini oshirish, til bilan bogʻliq toʻsiqlarni yoʻqotish yoʻlidagi muvaffaqiyatli izlanishlar hisoblanadi.

Topshiriqlar:

1. Avtomatik tarjima boʻyicha qanday yondashuvlar mavjud?
2. Qanday tarjimon dasturlarni bilasiz?
3. Avtomatik tarjimaning ishonchlilik darajasi haqida mustaqil mulohaza yuriting.
4. Avtomatik tarjimaning toʻlaqonli boʻlishiga toʻsiq boʻluvchi omillarni sanang.
5. Avtomatik tarjima qanday protseduralar asosida ishlaydi?

Adabiyotlar:

1. Нелюбин Л. Л. 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. Poʻlatov 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 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, lemmatizatsiya, 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 tuzish 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 optimallasuvi 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 mashhur elektron lug'atlar sirasiga CONTEXT, ABBY LINGVO, MULTI-TRAN, POLYGLOSSUM, MULTILEKS kabilari 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 u 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 qulaylik tug'diradi. Bundan tashqari, boshqa maxsus dasturlar lug'atni tahrir qilish va chop etish borasidagi ancha mashaqqatli mehnatni bir qancha oson-

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 foydalanuvchi har bir xaridorga o'z lug'atini tuzish yoki mavjud lug'at so'z boyligini to'ldirib borish imkonini beradi. Ochiq turdagi Internet lug'atlari (**on-line lug'atlar**) ham ko'p hollarda barcha foydalanuvchilarga ushbu lug'atlarni to'ldirish imkonini beradi. Shuningdek, elektron lug'atlarning hajm jihatidan ixchamligi, kompaktligi, boshqa manbalarga (Internet tarmog'i orqali, gipermurojaatlar yordamida) ulanish imkoniyati, mavjud matnlar korpuslari yordamida ilustrativ 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'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 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. 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 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.⁷²

Kompyuterda yaratilgan birinchi matnlar korpusi **Braun korpusi** (БК, inglizcha *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. 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 А.Н. Введение в прикладную лингвистику. – М.: Эдиториал УРСС, 2001. – С.61.

mamlakatlarda XX asrning 80-yillaridan boshlab bunday korpuslar tuzila boshlandi. Ular turli maqsad va vazifalarga xizmat qiladi. Buyuk Britaniya da **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.⁷³ Masalan, Rus tilining milliy korpusi hajmi hozirgi kunda 149 mln soʻzdan iborat. Keyingi yillarda Internet tizimining rivojlanishi **virtual matnlar korpusi** yuzaga kelishiga olib keldi. Yaʼni Internetdagi qidiriv saytlari, elektron kutubxonalar, virtual ensiklopediyalar korpus vazifasini bajarimoqda. Korpusning janri va tematik rang-barangligi Internetdan foydalanuvchining qiziqishlariga bogʻliq. Masalan, ilm-fan doirasida **Wikipedia** katta hajmdagi matnlar korpusi sifatida foydalanilmoqda.⁷⁴

Korpus lingvistikasida parallel matnlar korpusi ham muhim ahamiyat kasb etadi. **Parallel matnlar korpusi** esa, oʻz navbatida, badiiy asar, qoʻllanma, ommaviy axborot vositalari, turli xil hujjatlarning ikki yoki undan koʻp tillardagi elektron holdagi koʻrinishlaridir. Masalan, Yevropa Ittifoqi oʻzining barcha qonun va hujjatlarini ingliz, fransuz, nemis, ispan va italyan tillarida nashr qiladi hamda ular Internet tizimiga barchaga ochiq arxiv sifatida qoʻyiladi. Bunday korpuslarning afzalligi shundaki, ular yordamida nafaqat biron bir soʻz 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 lugʻatlarini tuzish imkoniyati tugʻiladi. Ushbu imkoniyatlar kompyuter leksikografiyasi uchun ulkan ahamiyat kasb etadi.

Kompyuter leksikografiyasiga semantik maydon, semantik tarmoq, semantik toʻr hamda freym semantikasining faol tatbiq etilishi natijasida ulkan kompyuter leksikografiya resurslari yaratildi. Shunday yirik leksikografik resurslardan biri **FRAMENET** boʻlib, u Internet tizimida on-line rejimida ishlaydi.⁷⁵ Mazkur tizim Ch.Fillmorning «**Tools for Lexicon Building**» loyihasi asosida Kaliforniya shtati, Berkli shahridagi Xalqaro informatika institutida ishlab chiqilgan. Framenet resursining maʼlumotlar bazasida 10 000 ta leksik birlik mavjud, undan 6000 dan ortigʻi toʻliq annotatsiyaga ega. Bundan tashqari, maʼlumotlar bazasida 800 ta semantik freym koʻrsatilgan, 135 000 annotatsiyali gaplar keltirilgan.

Elektron lugʻatlar tuzish jarayonida maʼlumotlar ombori, matnlar kor-

⁷³ <http://www.corpus.leeds.ac.uk/list.html>

⁷⁴ http://www.wikipedia.org/wiki/corpus_linguistics

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

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 – 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 affiksalar 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:

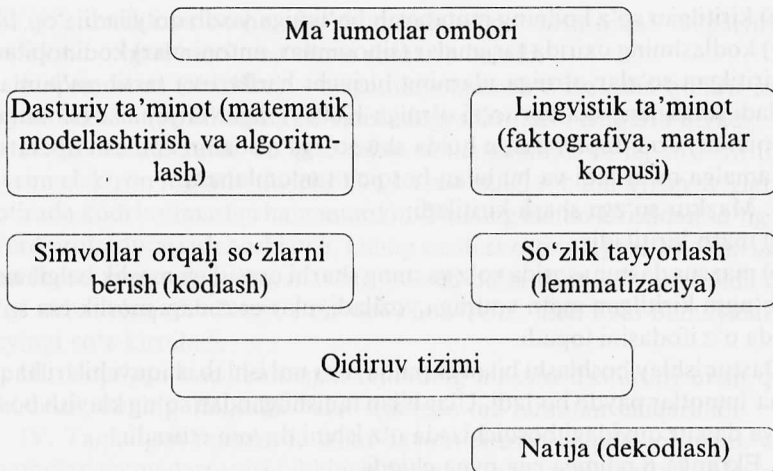
1. Kitobiy lug'atlarning tuzilishi: a) lug'atning so'zligi shakllantiriladi; b) misollar kartotekalari tuziladi; c) lug'at maqolalari yoziladi; d) lug'atning qolyozma varianti tayyorlanadi; e) qo'lyozma tahrir etiladi; f) muallif tuza-tishlar qiladi; g) nashr uchun teriladi; h) lug'at sahifalanadi; i)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

2. Elektron lug'atlarning tayyorlanishi: a) lug'atning so'zligi shakllanti-riladi; b) misollar korpusi (elektron kartotekalari) tuziladi; c) lug'at maqola-lari yoziladi; d) lug'at maqolalarini ma'lumotlar bazasiga ko'chiriladi; e) bevosita ma'lumotlar bazasida lug'at matni tahrir etiladi, korrektirovka qilinadi; f) 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 bo'ladi, 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 qo'llanadi, kodni qayta ishlash jarayonida zaruriy bo'lgan

⁷⁶ Po'latov A., Muhamedova S. Kompyuter lingvistikasi. – Toshkent, 2007. – B.58 – 61.



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, sinonimlari va antonimlarini ko'rsatish ularning birinchi harflarini kodirovkadagi 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 tegishligi 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 bo'lishi mumkin bo'lgan so'z so'raladi;

b) kiritilgan soʻz kodning mutanosib boʻlimiga yozib qoʻyiladi;

v) kodlashning oxirida tarjimalar (sinonimlar, antonimlar) kodi topiladi va kiritilgan soʻzlar oʻrniga ularning birinchi harflari va tartib raqamlari yoziladi (masalan, «katta» soʻzi oʻrniga K0083), agar tarjimalar (sinonim, antonimlar) kodi topilmasa, u holda shu soʻzga nisbatan kodlash operatsiyasi amalga oshiriladi va bu bilan bosqich tamomlanadi.

5. Mazkur soʻzga sharh kiritiladi:

a) matn kiritiladi;

b) maxsus dastur asosida soʻz va uning sharhi orasidagi moslik belgilanadi, soʻngra kiritilgan matn xotiraga yoziladi, ular orasidagi moslik esa soʻz kodida oʻz ifodasini topadi.

Dastur ishlay boshlashi bilan ekranda u va uni ishlab chiquvchilari haqida maʼlumotlar paydo boʻladi. Ular bilan tanishilgandan soʻng klavish bosiladi va dastur quyidagi bosqichlarda oʻz ishini davom ettiradi:

I. Ekranga 6 rejimga ega oyna chiqdi.

1-rejim. **Only translation (faqat tarjima)** – podstrochnikda foydalaniладigan tarjimalarni va mazkur soʻzning qaysi sohaga tegishliligi haqidagi maʼlumotlarni beradi. Mazkur rejimning asosiy funksiyasini Only-tran maxsus protsedurasi amalga oshiradi. Ushbu protsedura, avvalo, mazkur soʻz kodining adreslar boʻlimidan tarjimalar adresini izlab topadi. Soʻng uni qayta ishlashni boshlaydi: tarjimaning birinchi harfini olib, tekst faylini ochadi (shu harf bilan nomlangan) va berilgan tartib raqamiga koʻra tarjimani topib, uni ekranga chiqaradi.

2. **Grammatical information (grammatik maʼlumotlar)** – soʻzning barcha morfologik va sintaktik xususiyatlari va undan foydalanishdagi ayrim noaniqliklar haqidagi maʼlumotlarni beradi. Gram info maxsus protsedurasi ishlaydi. Grammatik maʼlumotlar kodi boʻlimida jamlangan maʼlumotlarni rasshifrovka qiladi.

3. **List of synonyms (sinonimlar roʻyxati).**

4. **List of antonyms (antonimlar roʻyxati).**

Ushbu rejimlar Syn-List va Ant-list maxsus protseduralar yordamida sinonimlar va antonimlar roʻyxatini beradi. Ular quyidagi tartibda ishlaydi: sinonim va antonimlar kiritilgan soʻz kodning «sinonimlar adresi» va «antonimlar adresi» boʻlimlaridan topiladi hamda ekranga chiqariladi.

5. **New word formation (yangi soʻzlarni qayta yasash)** – mazkur soʻz bilan bir xil oʻzakka ega boʻlgan barcha soʻzlarni chiqarib beradi. Asosiy ish NWF protsedurasi vositasida amalga oshiriladi:

a) soʻzning oʻzagi ajratiladi;

b) soʻz haqidagi barcha maʼlumotlarga ega boʻlgan holda ushbu oʻzakka

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. Ish 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 izlashdan iborat. Agar u topilsa, maxsus protseduralar uzatiladi, aks holda quyidagi xabar chiqadi «Sorry, I 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. «Ha» 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 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‘llanish imkoniyatlari, distributiv holatlari, birikma holiday 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 yozing.
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 Л. Л. 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. Баранов А.Н. Введение в прикладную лингвистику. – М.: Эдиториал УРСС, 2001.
6. Белоногов Г.Г. Компьютерная лингвистика и перспективные информационные технологии. – М.: Русский мир, 2004.

TIL O'RGATISH JARAYONINI AVTOMATLASHTIRISH

Reja:

1. Kompyuter lingvodidaktikasi va ta'lim jarayonida kompyuter texnologiyalaridan foydalanishning ahamiyati.
2. Til o'rgatishda avtomatlashtirish.
3. Til o'rgatuvchi o'quv dasturlari va ulardan foydalanish.

Tayanch so'z va iboralar: *til o'qitishni avtomatlashtirish, o'quv dasturlari, «elektron» o'qituvchi, multimedia, virtual kutubxona, CALL, avtomatizatsiya, bixevioristik yondashuv, kognitiv-intellektual yondashuv, avtomatik korrektorlar, speller, audio va vizual effektlar, 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 ko'effitsiyenti 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

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 harakatlanishga asoslangan, turli operatsiyalar bajarishga yo'naltirilgan va shu tamoyilda ishlashiga moslab programmalashtirilgan mexanizmlar yig'indisidir. Hammaga ma'lumki, yirik korporatsiyalarda, aytaylik, avtomobil 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 universitetlarida (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 bo'lib o'tdi. Mazkur konferensiyalarda kompyuter yordamida xorijiy tillarni o'rgatishning yangi metodlari taklif etildi, o'qituvchi lingvistik avtomatlar xususida qimmatli mulohazalar o'rta tashlandi, o'qitishda audio va video vositalardan foydalanish samarali ekanligi ta'kidlandi.

Kompyuter texnologiyalarining shiddat bilan rivojlanishi ta'lim jarayonini 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 muassasalarida yangi pedagogik texnologiyalarning ilmiy asoslarini yaratish, ularni tasniflash, metodik ahamiyatini belgilash, ta'lim jarayoniga tatbiq etish kabi dolzarb muammolar ustida ish olib borilmoqda. Ilg'or pedagogik texnologiyalarning kompyuter texnologiyalari bilan uyg'unlashtirib tashkil etilishi va shu asosda multimedia darslarining yaratilishi asosiy yo'nalishga aylanib qolmoqda. Multimedia («ko'p muhitlik» degan ma'noni bildiradi) zamonaviy kompyuterlarning deyarli barchasida mavjud bo'lgan imkoniyatlar 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>

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 mavzusining faqat to'rt dan bir qismini, ko'rgan materialining uchdan bir qisminigina 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

O'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 grammatikaning 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 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.

Til o'rgatishda quyidagi kompyuter dasturlarini qo'llash 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 kompyuterda o'qituvchining o'zi slaydlar yaratadi va ko'rgazmalar paketini tayyorlaydi. Elektron nusxa – bu o'zbek tili fani bo'yicha mavjud bo'lgan muayyan darslikning kompyuterdagi elektron versiyasi (nusxasi) bo'lib, unda multimedia vositalaridan faqat ovoz va matn qatnashadi.

2. *Elektron multimedia darsligi* – aynan kompyuter vositasida o'quv kursini yoki uning bitta katta bo'limini o'qituvchi yordamida yoki mustaqil o'zlashtirishni ta'minlovchi dasturiy-metodik kompleks.

⁷⁸ Po'latov A., Muhamedova S. Kompyuter lingvistikasi. – Toshkent, 2007. – B.64 – 65.

⁷⁹ Ko'rsatilgan asar. – B.66 – 67.

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 vositalari bilan boyitilgan dastur. Elektron lug'at turli yo'nalishda tuzilishi mumkin, undan talabaniq so'z boyligini oshirish, tarjima qilishda foydalaniladi.

5. *O'rgatuvchi audiodasturlar, elektron lingafon kurslari* umuman ikkinchi 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 eshitiriladi, talabaniq 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 mavzularda qo'llash mumkin.

7. *Elektron virtual kutubxona* – multimedia xonasining keng imkoniyatlaridan yana biri, elektron nusxa yoki elektron multimedia darsliklari mujassamlangan, 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 lingvodidaktikasida quyidagi turdagi topshiriqlardan foydalaniladi:

- a) **Close the gap** – Bo'sh kataklarni to'ldiring topshirig'i. Bunda semantika + grammatika sohalarini o'rganish ko'zda tutiladi;
- b) krossvordlar va leksik o'yinlar;
- c) elektron pochtdan 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:

TOEFL tizimi – ingliz tilini avtomatik o‘qitishga mo‘ljallangan eng rivojlangan xalqaro dastur. U dunyoning deyarli barcha mamlakatlarida amal qiladi, u inglizcha Test Of English as a Foreign Language so‘zlarining qisqartmasi bo‘lib, ingliz tilidan xorijiy til sifatida sinov standarti hisoblanadi. Bu tizim ingliz tilini o‘rgatish bilan birgalikda o‘zlashtirilgan bilimlarni diagnostika 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‘g‘ri jumla qurish kabilar yuzasidan bilimlarni baholaydi.

ENGLISH FOR ALL (Hamma uchun ingliz tili) – O‘zbekistonda yaratilgan avtomatik o‘qitishga asoslangan o‘quv dasturi. Unda o‘quvchiga ingliz tilidagi tovushlar va ularning talaffuzi audioeffektlar yordamida tushuntiriladi, qisqacha ingliz tili grammatikasi bayon etiladi, mashqlar va testlar yordamida bilimni tekshiradi. Ushbu o‘quv dasturining afzal tomoni shundaki, 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 necha yil davomida o‘qitishga mo‘ljallangan til o‘rgatish tizimi.

PROFESSOR HIGGINS – АНГЛИЙСКИЙ БЕЗ АКЦЕНТА – fonetik 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 shakllantiradigan ta‘limiy-elektron ishlanmalarni yaratishning umumiy va xususiy mezonlarini belgilash zarurati paydo bo‘ldi. Hozirgi paytda Rossiya, Fransiya, Yaponiya, Kanada va boshqa mamlakatlarda ta‘limiy-elektron ishlanmalarni yaratish konsepsiyasini ishlab chiqish bo‘yicha ilmiy-metodik tad-

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

Til o'qitishning avtomatik tizimi yildan yilga rivojlanib, yangi effektlar bilan boyib bormoqda. Bu tizimni yanada rivojlantirish, kompyuter lingvodidaktikasini 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 Л.Л. 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. Белоногов Г.Г. Компьютерная лингвистика и перспективные информационные технологии. – М.: Русский мир, 2004.

⁸⁰ Po'latov A., Muhamedova S. Kompyuter lingvistikasi. – Toshkent, 2007. – B.68.

KOMPYUTER LINGVISTIKASIDA AVTOMATIK TAHRIR, INFORMATSION QIDIRUV TIZIMLARI VA GIPERTEKST TEKNOLOGIYASI

Reja:

1. Avtomatik tahrir yo'nalishi.
2. Informatsion qidiruv tizimlari.
3. Gipertekst texnologiyasi.

Tayanch so'z va iboralar: *matn tahriri, mexanik xatolar, matn muharrirlari, avtomatik tahrir, tezaurus, deskriptor, gipertekst, kontekstli qidiruv, avtomatik referatlash, transliteratsiyalash, informatsion qidiruv tizimlari, manzilli qidiruv, semantik qidiruv, faktografik qidiruv, hujjatli qidiruv, indekslash, gipermedia.*

Avtomatik tahrir kompyuter lingvistikasining asosiy yo'nalishlaridan biri bo'lib, u kompyuterdagi matn muharrirlari (**Microsoft Word, Excel, Wordpad, Lexicon**) rivojlanishi bilan bog'liqdir. 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 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 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 shundaki, 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 chiqilsa, muharrir kasbi uchun ehtiyoj qolmaydi. Bu esa matn bilan ishlashni osonlashtiradi.

Ma'lumki, kompyuter yaratilmasdan ilgari matnlar yozuv mashinkalarida terilgan, bu jarayonda orfografik, punktuatsion, stilistik va grammatik

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 bo'lib, 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 alohida ta'kidlash zarurki, avtomatik tahrirlash tizimi faqat matn muharrirlari dasturlari bazasiga qo'shilgan tillarda amal qiladi, masalan, ingliz, rus, nemis, fransuz tillarida. Agar bazada mavjud bo'lmagan tilda matn kiritilsa, uni avtomatik tahrirlay olmaydi, ayrim internatsional so'zlar bundan mustasno. Avtomatik tahrirlash jarayonida MS Word lug'atida mavjud bo'lmagan so'z uchrasa, ostiga qizil chiziq bilan ajratib ko'rsatiladi. Mazkur holat quyidagi sababga ko'ra bo'lishi mumkin, bunda o'sha so'z imlosi haqida 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 sichqonchani o'ng tomoni bosilsa, kontekst menyusida 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 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, diagrammalar qo'shish, giperhavolalar bilan ta'minlash;
- matndagi abzaslar, belgilar, so'zlarning aniq miqdorini aniqlash (statistikasini ko'rsatish);
- tezaurus xizmati – matndagi muayyan so'zning sinonimik variantlarini aniqlash;
- transliteratsiyalash – bir alifbodan boshqa alifboga o'tkazish (masalan, kirill alifbosidan lotin alifbosidagi harfga almashtirish).

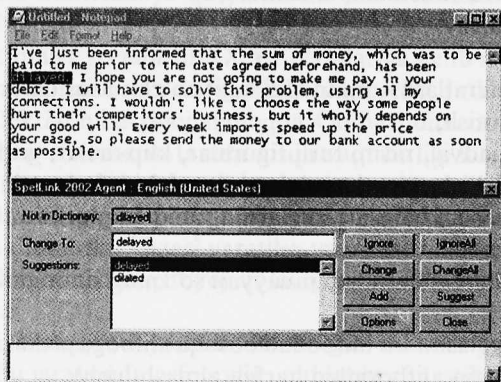
Avtomatik tahrirda uch tarkibli asosiy vazifa e'tiborga olinishi lozim:

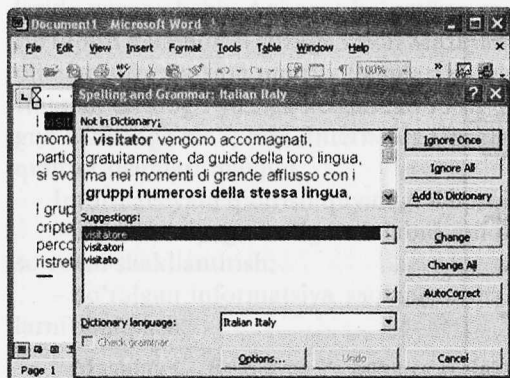
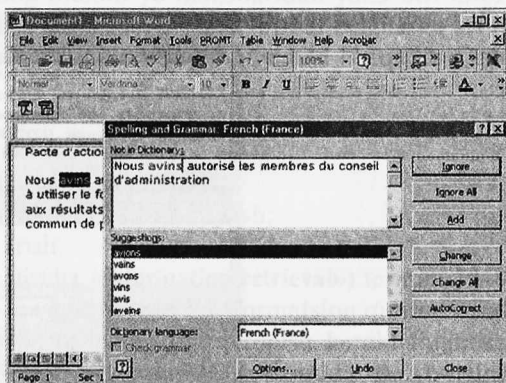
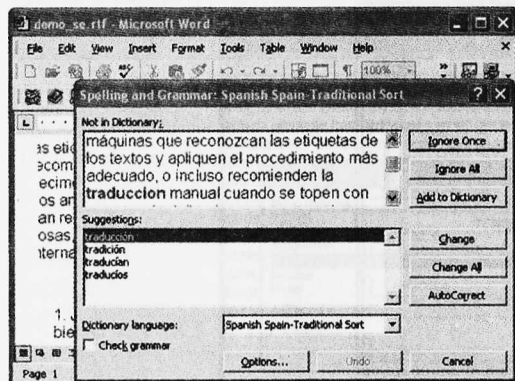
1. Orfografiyani tekshirish – bu vazifa morfologik tahlil yordamida amalga oshiriladi, bunda etalon lug'atdagi baza asosida tekshiriladi.

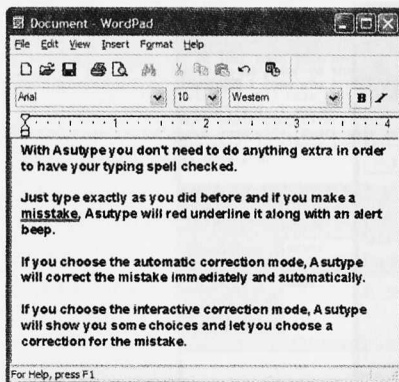
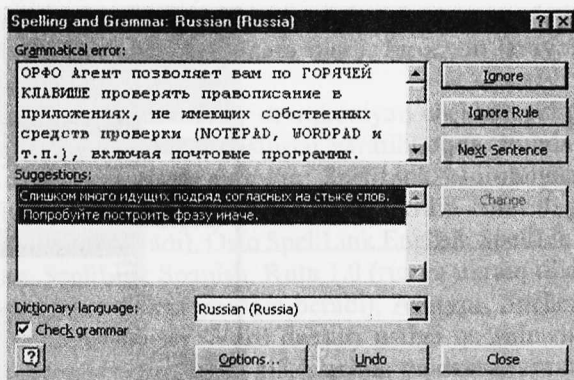
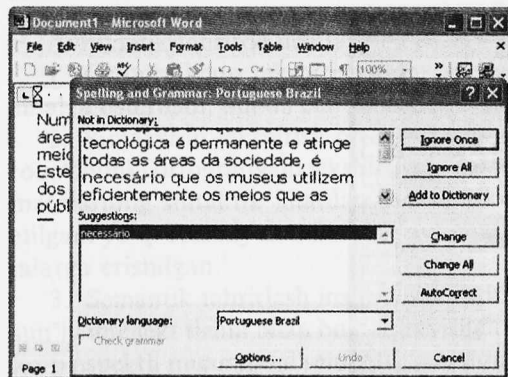
2. Sintaksisni tekshirish orfografiyani tekshirishdan ko'ra murakkabroq. Chunki orfografiyani tekshiruvchi dasturlar yaratilgan, lekin norasmiy matnlarning sintaktik jihatdan avtomatik tahriri masalasi haligacha hal etilgani yo'q. Rasmiy matnlarning avtomatik tahriri bo'yicha muhim natijalarga erishilgan.

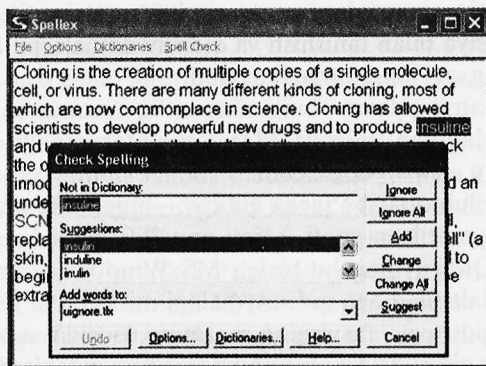
3. Semantik tahrirlash inson tafakkurini modellashtirishga asoslangan sun'iy intellekt tizimi bilan bog'liq ravishda o'rganilmoqda. Bunda mazkur ko'p aspektli muammoni semantik maydon, freym yondashuv, tezauruslar tizimini boyitish, so'z va so'z birikmalari o'rtasidagi paradigmatic munosabatlar (sinonimiya, gipo-giperonimik-tur-jins munosabati)ni modellashtirishga tatbiq etish asosida hal etish mumkin. Chunki inson bunday assotsiativ munosabatlarni o'z intellektiga tayanib o'rnatadi, 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 Pro (ingliz, ispan, italyan, nemis, fransuz, portugal, ukrain tillaridagi so'zlarning to'g'ri yozilishini tekshiradi), Orfo SpellLink English, SpellLink French, SpellLink Italian, SpellLink Spanish, Ruta 1.0 (rus va ukrain tillaridagi so'zlarning imlosini tekshirishga yordam beradi), Asutype, 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.









Kompyuter lingvistikasining muhim yo‘nalishlaridan yana biri informatsion 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 (HQQ)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 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 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 massivlarni) aniqlash;

⁸¹ Manning C., Raghavan P., Schutze H. **Introduction to Information Retrieval** // Cambridge University Press. 2008. – P. 5.

- informatsion massivlardan zaruriy informatsiyani ajratib olish;
- informatsiya bilan tanishish va qidiruv natijalarini baholash.

Qidiruvning quyidagi turlari mavjud:

1. To'liq matnli qidiruv – hujjatning to'liq mazmun-mundariyasi asosida amalga oshiriladigan qidiruv. Masalan, Internetdagi qidiruv tizimlari: www.yandex.ru, www.google.com.

2. Metama'lumotlar bo'yicha qidiruv – hujjatning muayyan atributlari (hujjat nomi, yaratilgan vaqti, 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 shunga 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 bo'lib, bunda hujjatning aniq manzili ko'rsatilishi lozim. Hujjatning manzili web-server va web-sahifa manzili yoki bibliografik qayd elementlari va hujjat saqlangan joy sanaladi. Semantik qidiruv hujjatlarlarining mazmuni asosida qidirish usuli bo'lib, 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 ma'lumotga mos keluvchi faktlarni (matnli hujjatlar tarkibidan ajratilgan ma'lumotlar) qidirish jarayonidir. Yuqoridagilardan kelib chiqib aytish mumkinki, qidiruv tizimi relatsion ma'lumotlar bazasida hamda gipertekstli ma'lumotlar bazasida mavjud bo'lgan hujjatlar, hujjatlar haqidagi metama'lumotlar, matnlar, rasmlar, video va audiodaylarni qidirishni o'z ichiga oladi.

Informatsiyaning eksponensial tarzda ko'payib ketayotganligi, shuningdek, Internet tizimida ham kundan kunga ma'lumotlar hajmining o'sib borayotganligi kompyuter lingvistikasi oldiga bir qator vazifalarni ko'ndalang qo'yimoqda:

- qidiruv tizimining optimal modellarini yaratish;
- hujjatlarni tasniflash (klassifikatorlar tizimini yaratish);
- hujjatlarni klasterlash;

– foydalanuvchi interfeyslari va qidiruv tizimlari arxitekturasini loyihalash;

– 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 bo'linadi: 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. Indeksash natijasida har bir hujjatga IQTdagi ta'rif shakli yoziladi. Mazkur ko'p bosqichli jarayonda deskriptorli informatsion 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.

Matnning asosiy mazmun-mundarijasini aks ettiradigan kalit so'zlar, terminlar, tayanch tushunchalar bazasi asosida kompyuterda lug'at tuziladi. 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 darajasiga 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 (turjins)**, **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. Θησαυρός – сокровище) в современной лингвистике – особая разновидность словарей общей или специальной лексики, в которых указаны семантические отношения (синонимы, антонимы, паронимы, гипонимы, гиперонимы и т.п.) между лексическими единицами.

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.

Категории: Морфология (лингвистика) | Синтаксис | Разделы лингвистики | 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 15 000 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. отношений^[1])

– РуТез – 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 – Тезаурус английского языка
- Захаров Н.В., Луков А.В. Школа тезаурусного анализа // Знание. Понимание. Умение. 2006. № 1. – С.231 – 233.
- Лесников С.В., Лесникова Д.С. Информационно-поисковый тезаурус

– *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» soʻziga bogʻishlangan lugʻat maqolasida semantik munosabatlar, assotsiativ bogʻlanishlar hamda iyerarxiya munosabati yaqqol koʻzga tashlanadi. Shu oʻrinda taʼkidlash zarurki, keyingi paytlarda qidiruv tizimlari bazasidagi tezauruslar giperhavolalar bilan ham taʼminlanmoqda, bu esa userga bir-muncha qulayliklar yaratadi. Yaʼni qidirilayotgan terminlarga aloqador boʻlgan boshqa yaqin tushunchalar bilan bogʻliq maʼlumotlarni ham oson topish imkonini beradi.

Ilm-fan, texnika taraqqiyoti intensiv ravishda rivojlanyotgan bu davrda terminlar dinamikasini eʼtiborga olgan holda tezauruslar bazasini muttasil toʻldirib, boyitib borish zarur. Chunki til ochiq dinamik tizim hamda jamiyat taraqqiyotining «barometri» hisoblanadi, yaʼni ayrim tushunchalar eskiradi, ularning oʻrniga yangi tushuncha va terminlar yuzaga keladi. Informat-sion qidiruv tezauruslar yaratishda mana shuni doimo eʼtiborga olish lozim. Xalqaro **INFOTERM (Xalqaro terminologik markaz, Avstriya)** tashkiloti maʼlumotiga koʻra, hozirgi kunda tillardagi terminlar miqdori 50 mln, mahsulot nomlari miqdori 100 mln ni tashkil etadi. Lekin ushbu statistik maʼlumot turgʻun emas, u jamiyat taraqqiyoti bilan yonma-yon oʻzgarib turadi.

Kompyuter lingvistikasida inqilobiy oʻzgarish yasagan yangilik gipertekst texnologiyasi hisoblanadi. Chunki bu tizim Guttenberg asos solgan kitob bosish texnologiyasiga zid ravishda yangi turdagi matn strukturasini vujudga keltirdi. Gipertekst gʻoyasi AQSH prezidenti F.Ruzveltning fan va taʼlim boʻyicha maslahatchisi Vannevar Bush nomi bilan bogʻliq boʻlib, u foydalanuvchiga matn va uning fragmentlarini turli assotsiativ munosabatlar asosida bogʻlashga imkon beruvchi «**MEMEKS**» tizimini nazariy jihatdan asoslab berdi. U 1945-yilda «The Atlantic Monthly» jurnalida «Memeks» deb atalgan kelajakdagi xayoliy qurilma haqida «**As We May Think**» («**Biz oʻylaganimiz kabi**») nomli maqolasini eʼlon qildi. V.Bush oʻsha qurilmani kutubxonadagi har qanday hujjatlar, yozishmalar, kitoblarni displeyda koʻrsatib bera oladigan, cheklanmagan mikrofilmlarga ulangan elektromexanik stol sifatida tasvirlaydi. Kompyuter texnikasi yoʻqligi bois loyihaning amalda qoʻllanishi bir oz keyinga surildi.

Bushning gʻoyasi gipertekst haqidagi nazariy va amaliy izlanishlar juda katta taʼsir koʻrsatdi. Bushning kelajak loyihasi 1960-yilda T.Nelsonning

«Ksanadu» tizimi orqali qayta ko'tarildi. Bu texnik tizim kiritilgan matnlarni va uning qismlarini turli xil usullarda, har xil qatorda, ixtiyoriy ketma-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 nomlangan.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 gipermedia terminlari T.Nelson tomonidan fanga kiritildi. Olim bu haqda o'zining «**Litarary Machines**» kitobida shunday yozadi: «*Gipertekst termini gipermedia terminiga qaraganda keng ommalashib ketdi. Aslida har ikki termin bir tushunchani anglatmaydi. Ya'ni gipertekst faqat matnlarning tarmoqlanishiga nisbatan ishlatiladi, gipermedia termini esa grafiklar, audio va videofayllar, shuningdek, matnlarning shaxobchalanishini ham o'z ichiga oladi. Bir muddat har ikki termin o'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 giperhavolalar yordamida bog'lashga imkon beradi. Bunda u o'quvchiga faqat matnlar ustida 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 beruvchi tizim gipertekst g'oyasidir. Internet saytlari manzilini ko'rsatishda HTTP dan foydalanamiz (Hyper Text Transfer Protocol – Giper matnni o'tkazish yo'lagi).

⁸² <http://www.sigweb.org/conferences/ht-cover.shtml>

⁸³ Субботин М.М. Гипертекст. Новая форма письменной коммуникации. – ВИНТИ. Сер. Информатика. 1994. - С.18

⁸⁴ <http://www.iw3c2.org>

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 strukturasi ko'ra **iyerarxik (pog'onali)** yoki **tarmoqli** bo'lishi mumkin. Daraxt ko'rinishidagi iyerarxik gipertekstlar foydalanuvchining navigatsiya (uzellar bo'ylab bir qismdan ikkinchi qismga o'tish) imkoniyatini cheklaydi, gipertekst texnologiyasining barcha imkoniyatlarini yuzaga chiqarmaydi. Bunday gipertekstda komponentlar o'rtasidagi aloqalar jins-tur munosabatiga asoslangan tezaurus strukturasi eslatadi. Tarmoqli gipertekst esa birmuncha keng tarqalgan va imkoniyati kengroq tizim hisoblanadi, u jins-tur munosabatidan tashqari boshqa assotsiativ munosabatlarni 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 kitoblarni 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. Haqiqatan ham gipertekst tizimi kompyuter lingvistikasida, Internet tizimida o'ziga xos inqilob yasadi, deyish mumkin.

Topshiriqlar:

1. Avtomatik tahrirning pragmatik jihatlarini tushuntiring.
2. Avtomatik tahrirning ishonchlilik darajasi nimalarga bog'liq?
3. Avtomatik tahrirda ma'lum salbiy jihatlarini tushuntiring.
4. Gipertekst texnologiyasi va an'anaviy kitob bosish texnologiyasini solishtiring.
5. Qanday informatsion qidiruv tizimlarini bilasiz?
6. Deskriptor nima?

Adabiyotlar:

1. Нелюбин Л. Л. 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. – ВИНТИ. Сер. Информатика, 1994.
6. Белоногов Г.Г. Компьютерная лингвистика и перспективные информационные технологии. – М.: Русский мир, 2004.

**KOMPYUTER ANALIZ VA SINTEZI.
INGLIZ TILIDAGI GAPLARNI O'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 dasturlar, formal model, morfologik, semantik-sintaktik va konseptual analiz, lingvoanalizatorlar, parser dasturlar.*

Matnning formal ajraladigan eng kichik birligi soʻz hisoblanadi, lekin u maʼno anglatuvchi eng kichik birlik morfemalardan tashkil topadi. Soʻz tarkibi oʻzak morfemalar, prefikslar (old qoʻshimchalar) va suffikslar (oʻzakdan keyin qoʻshimchalar)dan iborat boʻladi. Masalan, *besabrlikdandir* soʻzida 5 morfema mavjud. Soʻzlarning morfologik analizi uchun tildagi prefiks va suffikslar bazasi hamda soʻzlarning grammatik shakllarda oʻzgarishi (qanday qoʻshimchalarni qabul qilishi)ni koʻrsatuvchi lingvistik taʼmin yaratilishi lozim. Bunday baza avtomatik tarjima tizimi yaratilgan rus, ingliz, fransuz tillarida yaratilgan. Shulardan birini (rus tili grammatikasi asosida) berib oʻtmoqchimiz.⁸⁵

14-jadval

№ п/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.

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»

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	

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	мой, твой, свой	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

berish jarayonida yoki uning soʻnggi bosqichida (natijalarni inson hukmiga havola qilishda) soʻzlarning turli shakllarini chiqarish, analiz natijalarini umumlashtirishdan iborat. Morfologik analiz va sintez jarayoni kompyuterda kodlash-dekodlash tamoyili asosida ishlaydi. Bunda soʻz turkumlari maxsus kodlar yordamida belgilanadi. Quyidagi jadvalda RETRANS (inglizcha-ruscha, ruscha-inglizcha frazeologik mashina tarjimasiz tizimi) avtomatik tarjima tizimida morfologik analiz jarayonida foydalaniladigan inglizcha soʻzlarning grammatik tasnifini berib oʻtamiz:

15-jadval

Inglizcha soʻzlarning 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

Артикли	
71	Определенный артикль
72	Неопределенный артикль
Частицы	
81	Частица to
82	Частицы not и no
83	Остальные частицы

Ushbu jadvalda «1» simvoli otlarni, «2» simvoli sifatlarni, «3» simvoli fe'llarni, «4» simvoli ravishlarni, «5» simvoli predloglarni, «6» simvoli bog'lovchilarni, «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 анализа	Классы 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	embroided 0/2433	э 2433
23	in 0/52	И 52
ZZZ6v T э5дэИ = 11		
16	No 0/82	6 82

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

6vTэ5 д эИЗ»v = 32

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
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	A	11

vTэ5д э ИЗ»vA = 24

Bu yerda ayrim soʻzlar grammatik omonimiyaning hisobiga dastlab («tahlil natijalari»da) ikki xil raqam bilan berilgan, soʻnggi bosqichda ularning kontekstdagi holatidan kelib chiqib aniq tasnifi keltiriladi. Yuqoridagi matn fragmentida shunday soʻzlar sirasiga «state» soʻzini misol qilish mumkin. «State» soʻzi dastlabki tahlilda 0/1132 kodi orqali beriladi, bunda «0» raqami soʻzda qoʻshimchani yoʻqligini (nol morfemani), «11» raqami ot soʻz turkumini, «32» raqami esa hozirgi zamon shaklidagi feʼl yoki infinitivni bildiradi. Bunga oʻxshash yana bir soʻz «wished» soʻzining dastlabki tahlilida bir qiymatli boʻlmagan maʼlumotlarga ega boʻlamiz. «Wished» soʻzi 0/2433 kodi bilan beriladi, bunda «0» raqami yuqoridagidek nol morfemani, «24» feʼl asosida hosil boʻlgan sifat turkumini, «33» oʻtgan zamon shaklidagi feʼlni anglatadi.

Soʻzlarning qaysi turkumga mansubligini aniqlashda chappa (ters) lugʻatlardan foydalaniladi. Bunga turli xil turkumga mansub soʻzlarning morfologik analizini koʻrsatuvchi inglizcha soʻz shakllarining grammatik chappa lugʻati yaqqol misol boʻla oladi.

16-jadval

**Inglizcha soʻz shakllarining grammatik chappa lugʻati fragmenti
(ushbu lugʻat 54 000 leksik birlikni oʻz chiga oladi)**

Soʻz Grammatik maʼlumot		
Ot (Noun)	Sifat (Adjective)	Feʼl (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

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	artificial 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
tiger 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	qualify 1/32
settler 0/11	timely 0/41	imply 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	OT (Noun)
plotter 0/11	geomagnetically 0/41	contractions 1/11
slotter 0/11	OT (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	

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 tarjimoni 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 bo'lib, unda faqat gapning strukturasi analizi e'tiborga olingan, bunda har qanday gap so'zlarning muayyan ketma-ketligi asosida tashkil topgan qurilma sifatida baholangan. Gap strukturasi 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 yuzasidan sezilarli natijalarga erishilgani yo'q. Bu muammo kompyuter lingvistikasidagi ko'plab muammolar bilan tutashib ketadi. Jumladan, inson tafakkurini 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 zamonaviy yo'nalishlaridan biri hisoblanadi. Hozirgi kunda dunyoda keng tarqalgan jahon tillari uchun avtomatik tarjima dasturlari ishlab chiqilgan. Jumladan, ruschadan inglizchaga, inglizchadan ruschaga, ruschadan fransuzchaga 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, ma'lumotlar 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).

⁸⁶ Jurafsky D., Martin J.H. *Speech and Language Processing*. – New Jersey, 2000. – P.353.

3. Jumla kiritiladi va xotiradagi kodi qidirib topiladi.
4. Topilgan kod maxsus dasturlar yordamida qayta ishlanadi va tarjima oynada paydo bo'ladi.
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 tillarning ayrim jihatlari – polisemiya (ko'p ma'nolilik), omonimlik, frazeologizmlar, idiomalar, paremlar, 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 sirasiga kiradi, ingliz tili esa flektiv til hisoblanadi. Shuning uchun ingliz tilidan o'zbekchaga va o'zbekchadan inglizchaga kompyuter tarjimasini yaratish o'ziga xos qiyinchiliklarni keltirib chiqaradi. Bugungi kunda rus tili vositasida ingliz tilidan o'zbek tiliga avtomatik tarjima qiluvchi dasturlarning versiyalari 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 quyidagi vazifalarni hal etish uchun mo'ljallangan:

I.I. 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'ning quyidagi tuslanishli shakllarini qamrab oluvchi kesimli gaplar ko'rib chiqiladi:
 - a) shaxs (I, II, III shaxs);

- 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) massivning elementlari yordamida lug'at elementlari bilan taqqoslanadi, bu lug'atda olmosh, ko'makchilar, ko'makchi va modal fe'llar, artikllar 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 bo'lmasa, so'z shakl yasovchi affikslardan 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 bo'lgan so'zlar kesim guruhi hisoblanadi (kesim guruhiga: kesim, to'ldiruvchi, hol);
- 10) kesim guruhidan kesim ajratib olinadi;
- 11) so'ngra to'ldiruvchi ajratiladi;
- 12) gapning qolgan qismi hol hisoblanadi;
- 13) gapning har bir bo'lagi shakl yasovchi qo'shimchalarsiz tarjima qilinadi;
- 14) tarjima qilingan gap bo'laklaridan o'zbek tilidagi gap tuziladi, u albatta 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?

3. Ingliz tilida gaplarning dastur orqali tarjimasi sifati qanday bo'ladi?
4. Algoritmni hayotiy misollar bilan tushuntiring.
5. Tarjima algoritmi nima?

Adabiyotlar:

1. Нелюбин Л. Л. 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. Po'latov A., Muhamedova S. Kompyuter lingvistikasi. – Toshkent, 2007.
5. Белоногов Г.Г. Компьютерная лингвистика и перспективные информационные технологии. – М.: Русский мир, 2004.

KOMPYUTER LINGVISTIKASI FANINING RIVOJLANISH ISTIQBOLLARI

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'liqini, 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 bog'liq holda informatsion jamiyat konsepsiyasi ishlab chiqildi. Bu konsepsiyaning asoschisi amerikalik futurolog olim E. Toffler hisoblana- di. U jamiyatni turli bosqichlarga bo'lib o'rganishni taklif etib, **sivilizatsi- yaning uch to'liqini** ajratdi: a) birinchi to'liqin – industrial jamiyatgacha

bo'lgan davr (qishloq xo'jaligi sivilizatsiyasi). Bunda asosiy boylik yer va tabiatdan olingan zaxiralar hisoblanadi; b) ikkinchi to'lqin – industrial sivilizatsiya. Bunda asosiy boylik ishlab chiqarish vositalari sanaladi; c) uchinchi to'lqin – postindustrial sivilizatsiya (informatsion jamiyat). Bunda kapital 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. Shuning uchun jahon bozorida ma'lumot oltindan ham qimmatroq baholanadi. Mazkur axborotni topish, saqlash, qayta ishlash va boshqalarga yetkazishning qulay usullariga bo'lgan ehtiyoj kun sayin oshib bormoqda. Demak, kimki mazkur qulay usullarni ishlab chiqsa, ularga egalik qilsa, dunyodagi eng boy odam bo'ladi. Dunyodagi eng badavlat kishi Bill Geys kompyuter, ya'ni axborotni qabul qilish, saqlash va qayta ishlash sohasi vakili. Bill Geysning o'zi shunday deydi: «*Kim axborotga ega bo'lsa, u hamma narsaga 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*».⁸⁷

2003-yil dekabrda Jenevada (Shveysariya) informatsion jamiyat qurish muammosiga bag'ishlangan Butunjahon Sammiti bo'lib o'tdi. Unda shunday shior o'rta tashlandi: «*Informatsion jamiyat qurish – yangi mingyillikning 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 belgilanishicha, endilikda jamiyatning taraqqiyoti kompyuter texnologiyalari bilan chambarchas bog'liq bo'lib qoladi. Shu o'rinda alohida ta'kidlash lozimki, 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 yuzaga chiqaruvchi muhim vosita sifatida baholadilar. Qisqa vaqt ichida kompyuterlardan teoremlarni 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.

sekinlik bilan insonning aqliy faoliyatini modellashtiruvchi dasturlarning yaratilishi muayyan sohalaridagi murakkab masalalarni yechuvchi ekspert tizimlarning yuzaga kelishiga olib keldi. Lekin shuni ta'kidlash zarurki, hali-hanuz inson miyasining analogi bo'la oladigan dastur yoki tizim yaratilgani yo'q.

Kompyuter lingvistikasi fanining eng istiqbolli yo'nalishlaridan biri sun'iy intellekt hisoblanadi. Virtual qomus hisoblangan Wikipedia sahifalarida 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 bo'lgan masalalarni, xususan, fikrlash, nutq mazmunini anglash, ma'lumotlarni umumlashtirish kabilarni yechishga yo'naltirilgan bo'ladi. Bu termin mazkur xususiyatlarga ega sistemalarni ishlab chiqish bilan bog'liq kompyuter texnologiyalarining bir bo'limiga 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 u haqidagi fan. Intellektual sistema strukturasi 3 asosiy blokni o'z ichiga oladi: bilimlar bazasi, masalani yechuvchi tizim va intellektual interfeys*».⁹¹

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 operatsiyalarni bajara olishiga bog'liq. Intellektni bunday talqin qilishdan ko'rinib turibdiki, sun'iy intellekt insondagi intellekt tushunchasining aynan imitatsiyasi 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-yilda** Darmut universitetidagi konferensiyada qo'lladi va shundan buyon tanqidlarga qaramay ishlatib kelinmoqda. Rossiya sun'iy intellekt assotsiatsiyasi Peterburg bo'limi boshlig'i T.A.Gavrilovning yozishicha, ingliz tilidagi *artificial intelligence* birikmasi fantastik antropomorfi 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 kibernetika hamda informatika doirasida olib borilgan. XX asrning 80 – 90-yillaridan boshlab sun'iy intellekt juda ko'plab fanlarning o'rganish obyektiga aylandi. Jumladan, neyrolingvistika, psixologiya, informatika, neyrofi-

⁹⁰ http://www.wikipedia.org/wiki/artificial_intelligence.html

⁹¹ <http://www-formal.stanford.edu/jmc/whatsai.html>

ziologiya, epistemologiya (falsafadagi bilish haqidagi ta'limot), kognitologiya, 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 bo'lgan tizimlarni yaratish;

– yuqori (inglizcha *Bottom-Up AI*), biologik yondashuv – neyron tarmoqlarini o'rganish, biologik unsurlarga asoslangan holda intellektual salohiyatni modellashirish va natijada **neyrokompyuter** yoki **biokompyuterlar**ni yaratish.

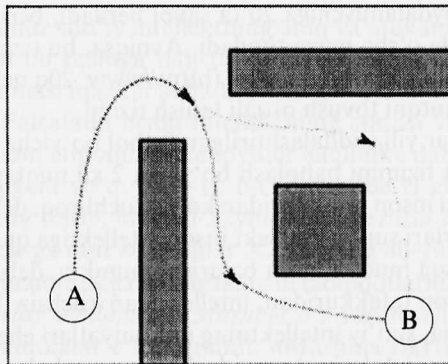
– semiotik va biologik yondashuvning sintezi bo'lgan umumlashtiruvchi gibridd tizimli yondashuv – sun'iy intellektning yakuniy maqsadi bo'lishi lozim.

Mazkur yondashuvlardan tashqari, yana simvulli, mantiqiy, agentga yo'naltirilgan yondashuvlar ham mavjud. Simvulli yondashuv, asosan, simvolik mantiq qoidalarini tizimga tatbiq etish bilan bog'liq bo'lib, bu ko'proq tizim qoidalarini 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 yozilgan 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 to'laqonli yechish uchun atrofda 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 aloqador tushunchalar bo'lib, 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_intelligence.html



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 yuzaga keldi, bunda san'at asarlarini avtomatik tarzda yaratish jarayoni tushuniladi. Bugungi kunda ijodni to'la modellashtirishga erishilgani yo'q, lekin bu sohada sezilarli yutuqlar qo'lga kiritilgan. Jumaladan, kompyuter yordamida musiqalar yaratilishi, treker-dasturlar, sintezatorlar, sekvensorlar yordamida bir turdagi melodiylar 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) 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 shaxmat 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 bo'lib, u sun'iy intellekt bo'yicha o'tkazilgan dastlabki tajribalardan biridir. Bu o'yin muallifi Robin Byorgenerdir. Bunda foydalanuvchidan biror obyektini topish

soʻraladi, soʻng foydalanuvchiga 20 ta savol beriladi, berilgan javoblardan kelib chiqqan holda oʻsha narsa topiladi. Ayniqsa, bu tizim Internet paydo boʻlgandan soʻng juda mashhur boʻldi (<http://www.20q.net> sayti mavjud).

– **Via Voice** – nutqni tovush orqali tanish tizimi.

– **RoboCup** – har yili soddalashtirilgan futbol boʻyicha robotlar turniri.

Sunʼiy intellekt tizimini baholash boʻyicha 2 xil nuqtayi nazar mavjud: 1) sunʼiy intellektni inson tafakkuridan koʻra kuchliroq, deb baholash. Ushbu qarash tarafdorlari sunʼiy intellekt inson intellektiga qaraganda koʻproq operatsiyalarni qisqa muddatlarda bajarishi mumkin, deb hisoblaydilar; 2) sunʼiy intellekt inson tafakkuridan, intellektidan kuchsiz. Bu yoʻnalishdagi olimlarning fikricha, sunʼiy intellektning imkoniyatlari cheklangan boʻladi, u faqat programmaviy taʼminot asosida muayyan sohalarida ishlaydi, lekin u insonning bilish qobiliyatining, tafakkurining barcha qirralarini qamrab ololmaydi. Buni Jon Syorl tomonidan oʻtkazilgan «Xitoy uyi» nomi bilan mashhur boʻlgan semantik eksperiment ham tasdiqlaydi. Olim bu tajribasini 1980-yilda «*Minds, Brains and Programs*» («Tafakkur, miya va programmlar») nomli maqolasida eʼlon qildi. J.Syorlning fikricha, *kompyuter sintaktik qoidalarni algoritim nazariyasi orqali formallashtira oladi, lekin semantika, tafakkurni toʻla modellashtirishga dasturlar oʻzlik qiladi.*⁹⁴ Xuddi shu pozitsiyada turgan olimlardan yana biri Rodjer Penrouz oʻzining «*Qirolning yangi aqli*» kitobida tafakkurni formal tizimlar asosida mutlaq modellashtirishning imkonsizligini bayon qilgan.⁹⁵

Ayrim olimlar sunʼiy intellekt boʻyicha erishilayotgan yutuqlarning keyingi oqibatlaridan qoʻrqmoqdalar. AQSHda Sunʼiy intellekt singular instituti (SIAT) tadqiqotchisi E.Yudkovskiy insoniyatni sunʼiy intellektlar bilan bogʻliq global xavf kutayotganidan ogohlantiradi. Olimning fikricha, agar kelajakdagi sunʼiy intellekt tizimiga insonlarga nisbatan doʻstlona munosabatda boʻlishini taʼminlovchi dastur oʻrnatilmasa, butun insoniyat uchun katta xavf paydo boʻladi.⁹⁶ Ilmiy fantastik asarlar muallifi Veror Vinj ham sunʼiy intellektlar tizimi toʻla yaratib boʻlingan paytda jamiyatda keskin oʻzgarishlar boʻlishi haqida yozgan. Bu davr ilmiy adabiyotlarda **texnologik singularlik** deb atalmoqda. Hatto futurologlarning fikricha, bu davr 2030-yilda boshlanadi.⁹⁷

⁹⁴Searle J.R. Minds, brains, and programs // Behavioral and Brain Sciences. 1980. – P.417 – 437.

⁹⁵ Пенроуз Р. Новый ум короля. О компьютерах, мышлении и законах физики. – М.: УРСС, 2005. – P.14.

⁹⁶ <http://www.thetech.org/robotics/etics/index.html>

⁹⁷ Kurzweil R. The Singularity Is Near. – N.Y.: Viking, 2005. – P.3.

Fanda hali-hanuz sun'iy intellektning aniq va mukammal ta'rifi mavjud emas, chunki fan bu tizimga hali to'liq 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 mezonini 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 davrida «*Mashina fikrlay oladimi?*», «*Mashina o'z yaratuvchisidan aqlliroqmi?*» sarlavhalarida maqolalar e'lon qilindi. «*Mashina fikrlay oladimi*» (1950) sarlavhali maqolaning muallifi ingliz olimi Alan Tyuring mashina imkoniyatini 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) gaplashayotganini aniqlashi lozim bo'ladi. Kompyuter dasturining vazifasi esa insonni xatoga yo'l qo'yishiga, noto'g'ri tanlov qabul qilishiga majbur qildirish sanaladi. 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*».⁹⁸

Avtomatlashtirilgan ekspert tizimlar mantiqiy masalalarni kompyuter yordamida yechishga imkon beradi. Mazkur tizimlar ma'lum sohadagi bilimlar 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 tashxis 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.

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 beradigan dasturlar (**lingvistik protessorlar**) 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» korporatsiyasi tomonidan **Q&A** dasturi, «Karnegi grupp» kompaniyasi tomonidan **Language Craft** dasturiy ta'minoti ishlab chiqildi.⁹⁹ Bu turdagi dasturlar hozirda ingliz tilining standartlashgan so'zlashuv uslubida uncha katta bo'lmagan lu'g'at zaxirasi bilan ishlaymoqda. Tabiiy tilni tushunishga imkon beruvchi dasturlarning qamrovini kengaytirish kompyuter lingvistikasining dolzarb vazifalaridan biridir. Bunday dasturlarni rivojlantirish inson kabi muloqot qila oladigan, erkin kommunikatsiyaga kirisha oladigan 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 modellash-tirish 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 imkonsiz qolayotganligini ta'kidlaymoqdalar. Mazkur jihatlar sirasiga insonning ichki dunyosi, intuitsiyasi, ong osti jarayonlari, ilhom kabi sirli fenomenlarni aytish mumkin.

Inson miyasi murakkab biologik to'r hisoblanadi, u yuz mlrd nerv to-laridan – neyronlardan tarkib topgan. Neyronlarning bir-biriga uzviy aloqadorligi, 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.

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 strukturasi korrekcirovka 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 strukturasi korrekcirovka amal-ga oshirilmaydi. Faqatgina ma'lumotlar bir-biriga muvofiq kelmaganda korrekcirovka yuz beradi. Shundan so'ng neyron tarmoqlarida muayyan amaliy vazifani hal etish tizimi to'la shakllanadi.¹⁰⁰

Sun'iy intellekt tizimiga doir amaliy ishlardan biri ekspert tizimlari hisob-lanadi. 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 va manipulyatsiya tizimigagina ega bo'ladi. Ekspert tizimlari mazkur jihatlardan tashqari, xuddi insonlarda bo'lgani kabi dalillarga tayanib muhokama yuritish imkoniyatiga ham egadir. Eks-pert tizimlarining asosiy ikkita tarkibiy qismi mavjud: ma'lumotlar ombori va mantiqiy mashina.

Ekspert tizimlari konsepsiyasi Stenford universiteti professori Eduard Feygenbaum nomi bilan bog'liq bo'lib, 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.

zasiga bog'liq ekanligini asoslab berdi.¹⁰¹ Dastlab ekspert tizimlari diagnostika va kasallikni davolash sohasiga tatbiq etildi. Ekspert tizimlari keyinchalik kimyo, biologiya, geologiya sohalarida ham qo'llanila boshlandi.

Ekspert tizimlariga bilimlar bazasini kiritish uchun simvolli ma'lumotlarga ishlov berishga xizmat qiladigan LISP, REFAL, PROLOG kabi sun'iy tillardan foydalanilmoqda. Bu tillar evristik bilimlar va mantiqiy xulosalarni yuzaga chiqarishda eng qulay programmashtirish 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 programmlar kiradi:

- savol-javobli dastur foydalanuvchi bilan tabiiy tilda munosabatga kirishishni ta'minlab beradi;
- bilimlar ombori dasturi foydalanuvchiga bilimlar bazasiga o'zgartirishlar kiritish hamda to'ldirib, boyitib borishga imkon beradi;
- tushuntiruvchi dastur – javoblar tizimini tekshiradi;
- interpretator-dastur muayyan sohadagi terminlarni sharhlash, izohlashga yordam beradi;
- ijro etuvchi dastur barcha tizimlar ishini ta'minlab beradi.

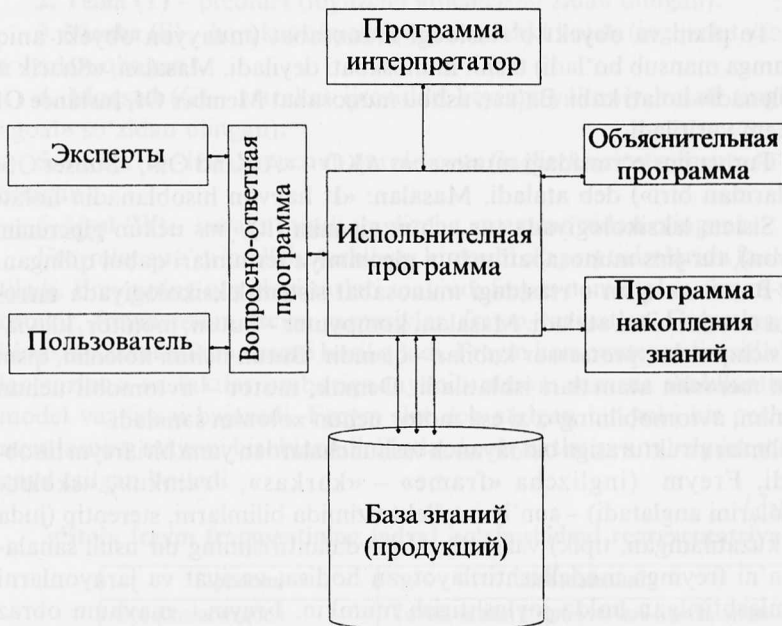
Yuqorida aytilganidek, ekspert tizimlari deklarativ va protsedurali bosqichlar asosida ishlaydi. Deklarativ bosqich – bu ekspert tizimiga kiritilgan bilimlar bazasi bo'lsa, protsedurali bosqich ekspert tizimlarining mavjud bilimlar bazasiga tayanib hukm chiqarish va aniq qaror qabul qilish bosqichini 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 quyidagilar: klapaning tiqilishi, filtrning axlatga to'lganligi, klapan tepa qismining ifloslanganligi, nasos membranasining zararlanganligi. Agar benzonasosdan benzin ajralishi yetarli bo'lmasa, u holda kiritish shlangi axlatga to'lib, tiqilib qolgan bo'lishi mumkin kabi.

Informatsiya shiddat bilan ko'payib borayotgan bugungi davrda ekspert tizimlarining bilimlar bazasini uzluksiz to'ldirib, yangi bilimlar bazasi bilan boyitib borish birlamchi vazifa bo'lib qolmoqda. Chunki vaqt o'tgan sari mavjud bilimlar eskirib, talabga javob bermay qolishi mumkin, shunda ularni yangi va qo'shimcha ma'lumotlar bilan to'ldirish joiz bo'ladi. Bilim-

¹⁰¹ Белоногов Г.Г. Компьютерная лингвистика и перспективные информационные технологии. – М.: Русский мир, 2004. – С.226.

larning mavjud holatlar bilan muvofiqligi hamda bilimlar bazasining boyitilishi 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 strukturasi o'ld bo'lgan **semantik tarmoq**, **freym**, **senariy**, **Word-net** kabi tushunchalar va ularning tatbiqi muhim hisoblanadi. 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 munosabatlarning 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 obyektarga, muayyan sohaga oid tushunchalarga, dugalar obyektlar o'rtasidagi semantik munosabatlarga aloqador tushunchalardir. Semantik

tarmoq ikki qismdan tashkil topadi: 1) mavhum tarmoq; 2) aniq tarmoq. Mavhum tarmoq umumiy tushunchalar orqali ifodalanadigan (zamon, makon, 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** – «A 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 **meronimiya** termini bilan ataladi. Masalan, kompyuter – butun; monitor, klaviatura, sichqoncha, protsessor kabilar – qismdir. Butun uchun **xolonim**, qism uchun **meronim** atamaları ishlatiladi. Demak, motor – avtomobil uchun meronim, avtomobilning o'zi esa motor uchun xolonim sanaladi.

Bilimlar strukturasi o'ld 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 gavdalandirishning 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 bo'lish, 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 strukturasi nisbatan ishlatilgan.¹⁰² Shundan so'ng freym semantikasi konsepsiyasi (**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 bo'ladi. Bir freym boshqa bir freymning davomi yoki tarkibiy qismi bo'lishi mumkin. Ch.Fillmorning qarashlariga L.Tenyening aktantlar (fe'lining bo'sh

¹⁰² Minsky M. A Framework for Representing Knowledge / The Psychology of Computer Vision. McGraw-Hill, New York (U.S.A.), 1975.

o'rinlarini to'ldiruvchilar) haqidagi g'oyasi katta ta'sir ko'rsatgan. Fillmorning **Case (kelishiklar)** g'oyasiga asosan aktantlar (fe'lining hamrohlari) fe'l yordamida ifodalanadigan holat yoki harakat munosabatlaridagi bajariladigan «rollar» sifatida talqin qilinadi. Ch.Fillmor ajratgan kelishiklar quyidagilar:

1. **Agent (A)** – harakat bajaruvchisi (inglizcha «agent» so'zidan olingan).
2. **Tema (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. **Usul (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-strukturalar, freym-rollar, freym-senariylar, freym-vaziyatlar. Freymlar tizimi birlashib, **semantik tarmoqni** hosil qiladi. Freym kompyuter xotirasida bilimlar xazinasini va faktlar omboriga tayanib olamni maxsus shakllantiruvchi model vazifasini bajaradi. Freym **slot** deb atalgan bir yoki bir necha elementlarning majmui hisoblanadi. Slotlarda obyektarga 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 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 kausal bog'liqlik tamoyili ishlaydi, ya'ni har bir harakat natijasi shartli sanaladi,

uning ta'siri va oqibatida navbatdagi ish-harakat sodir bo'lishi 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 tarmoqlari 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 gibridd 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 «qoida 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-

da) asosan neyron tarmoqlaridan, obraz aniqlangandan keyin bajariladigan mantiqiy amallarda esa ekspert tizimlaridan foydalanilsa, yuqori samaradorlikka 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 kiritish mumkin?
5. Sun'iy intellekt tizimini yaratish yuzasidan qanday yondashuvlar mavjud?

Adabiyotlar:

1. Нелюбин Л. Л. Компьютерная лингвистика и машинный перевод. – М.: ВЦП, 1991.
2. Шемакин Ю.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.

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. ВИНТИ. 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. V. 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. Ёқубов Т., Каримбеков С. Математик мантиқ элементлари. – Тошкент: Ўқитувчи, 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.

23. Лурия А.Р. Язык и сознание. – М.: МГУ, 1979.
24. Марчук Ю.Н. Проблемы машинного перевода. – М.: Наука, 1983.
25. Набиева Д. Ўзбек тилининг турли сатҳларида умумийлик-хусусийлик диалектикасининг намоён бўлиши. – Тошкент: Шарқ, 2005.
26. Нелюбин Л.Л. Компьютерная лингвистика и машинный перевод. – М.: ВЦП, 1991.
27. Незматов Ҳ., Бозоров О. Тил ва нутқ. – Тошкент: Ўқитувчи, 1993.
28. Новое в лингвистике. Вып.V. Языковые универсалии. – М.: Прогресс, 1970.
29. Новое в зарубежной лингвистике. Вып.24. Компьютерная лингвистика. – М.: Прогресс, 1989.
30. Нурмонов А. Лисоний белги хусусиятлари ҳақида. – Андижон, 1992.
31. Нурмонов А., Йўлдошев Б. Тилшунослик ва табиий фанлар. – Тошкент: Ўқитувчи, 2001.
32. Nurmonov A. Lingvistik tadqiqot metodologiyasi va metodlari. – Toshkent: Akademnashr, 2010.
33. Nurmonov A. Struktur tilshunoslik: ildizlari va yo'nalishlari. – Toshkent, 2008.
34. Пенроуз Р. 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. Ўзбек тилининг лексик сатҳини системалар системаси тамойили асосида тадқиқ 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. ВИНТИ. 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.

49. Усмонов С. Умумий тилшунослик. – Тошкент: Ўқитувчи, 1972.
50. Успенский В.А., Семенов А.Л. Теория алгоритмов: основные открытия и приложения. – М.: Наука, 1987.
51. Фалсафа. Қисқача изоҳли луғат. – Тошкент: Шарқ, 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 // 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. Шермухаммедова Н. Фалсафа ва фан методологияси. – Тошкент: Университет, 2005.
57. Эко У. Отсутствующая структура. Введение в семиологию. – М.: Петрополис, 1998.
58. Anderson J., Rosenfeld E. 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 «MT in the great translation era». – Tokyo: AAMT, 1996.
68. Jurafsky D., Martin J.H. Speech and Language Processing. – New Jersey, 2000.

69. Kurzweil R. The Singularity Is Near. – N. Y.: Viking, 2005.
70. Manning C., Raghavan P., Schütze 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 M., 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-03/GENEVA/DOC/4-E. 12 December 2003.

INTERNET SAYTLARI

1. <http://www.vaal.ru>
2. <http://www.rusf.ru/books/analysis>
3. <http://www.starling.rinet.ru>
4. <http://www.ruscorpora.ru> (corpora.yandex.ru)
5. <http://www.philol.msu.ru/~lex/corpus>
6. <http://www.philol.msu.ru/~humlang/articles/PolystylCorp.htm>
7. <http://www.bokrcorpora.narod.ru/frqlist/frqlist.html>
8. <http://www.info.ox.ac.uk/bnc>
9. <http://www.wikipedia.org/wiki/axiom.html>/Гильберт
10. <http://www.thetech.org/robotics/etics/index.html>
11. <http://www-formal.stanford.edu/jmc/whatisai.html>
12. http://www.wikipedia.org/wiki/artificial_intelligence.html
13. <http://www.sigweb.org/conferences/ht-cover.shtml>
14. <http://www.iw3c2.org>
15. <http://www.yazykoznanie.ru>
16. <http://www.corpus.leeds.ac.uk/list.html>
17. <http://framenet.icsi.berkeley.edu>
18. <http://www.dialog-21.ru>
19. <http://www.CICLing.org>
20. <http://www.ling.uni-konstanz.de/pages/proj/sprachbau.htm>

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 (teoremlar) 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, sistemaning 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'lalilik – aksiomalar sistemasiga qandaydir jumlaning qo'shganimizda sistemada 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 bog'liq. U Yevropada buzilgan shaklda «Algorithmi» tarzida ishlatilgan) – ma'lum bir tipga oid hamma masalalarni yechishda ishlatiladigan jarayonlar tizimining muayyan tartibda bajarilishi haqidagi aniq qoida yoki qoidalar tizimi.

Algoritmik tillar – programmalashtirishga asoslangan sun'iy tillar bo'lib, undagi qurilmalar xuddi tabiiy tildagidek aniq interpretatorga – insonga yoki avtomatik qurilmaga yo'nalgan bo'lishi lozim, toki u axborotni qabul qilsin va tushunsin. Agar bu jarayon (qabul qilish va tushunish) amalga oshmasa, u holda algoritmik tillar va kompyuter orasida vositachi sifatida **translator-dasturlar** (yoki **kompilatorlar**)dan foydalaniladi. 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 shakllantirilishi va generatsiyasi (hosil qilinishi) – kompyuter lingvistikasining tayanch tushunchalari. Morfologik analiz morfemalarga ajratish hamda grammatik ma'lumotlar tahlilini o'z ichiga oladi. Morfologik sintez esa matnga avtomatik ishlov 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 kompyuterda kodlash-dekodlash 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

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. 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 yevropa 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 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 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 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 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 asosiyalarini 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 (statistikasini ko'rsatish);

– tezaurus xizmati – matndagi muayyan soʻzning sinonimik variantlarini aniqlash;

– transliteratsiyalash – bir alifbodan boshqa alifboga oʻtkazish (masalan, kiril 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 soʻzlarning toʻgʻri yozilishini tekshiradi), **Orfo SpellLink English**, **SpellLink French**, **SpellLink Italian**, **SpellLink Spanish**, **Ruta 1.0** (rus va ukrain tillaridagi soʻzlarning imlosini tekshirishga yordam 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 maʼnoli qismlarning ketma-ket munosabatidan tashkil topgan butunlik tushuniladi. Ishtirokchilar deb kattaroq konstruksiya tarkibiga kirgan soʻz yoki konstruksiyaga aytiladi. Muayyan konstruksiyaning bevosita shakllanishida ishtirok etgan bir yoki bir nechta ishtirokchilarga bevosita ishtirokchilar deyiladi. Masalan, *Aʼlochi talabalar oʻz vazifalarini aniq va puxta bajaradilar*. Bu jumla konstruksiya hisoblanadi, uning tarkibidagi barcha maʼnoli soʻzlar («va» bogʻlovchisidan tashqari) ishtirokchilar, bevosita aloqaga kirishgan soʻzlar bevosita ishtirokchilar deyiladi. Bunda *aʼlochi soʻzi talabalar soʻzi bilan, aniq, puxta* soʻzlari *bajaradilar* soʻzi bilan, *oʻz soʻzi vazifalarini* soʻzi bilan oʻzaro zich bogʻlangan. Demak, bevosita ishtirokchilar sintaktik modelda dastlab soʻz birkimlari ajratiladi, soʻngra ular gap ichida birlashtiriladi.

Bilimlar injenerligi (Knowledge Engineering) – bilimlar bazasi bilan shugʻullanuvchi maxsus yoʻnalish boʻlib, u sunʼiy intellekt tizimining tarkibiy qismi sanaladi. Bilimlar injenerligi kognitiv lingvistika, kompyuter semantikasi va kompyuter lingvistikasining yutuqlariga tayanadi. Bunda informatsiya va bilimlar strukturasi oʻziga oid boʻlgan **semantik tarmoq**, **freym**, **senariy**, **Word-net** kabi tushunchalar va ularning tatbiqi muhim hisoblanadi.

CALL (Computer Assisted Language Learning) – til oʻqitishni avtomatlash-tirish – xorijiy tillarni kompyuter yordamida oʻqitish tizimini taʼminlovchi amaliy dasturlar hamda bilimlarni avtomatik baholash mexanizmlari. Til oʻrgatishda kompyuterdan foydalanish boʻyicha dastlabki tajribalar XX asrning 60-yillarida Stenford (Russian-Program) va Nyu-York universitetlarida (Das Deutche Program) oʻtkazildi. XX asrning 70 – 80-yillarida til oʻqitishda kompyuterdan foydalanishga boʻlgan qiziqish kuchaydi va hatto alohida yoʻnalish – **Computer Assisted Language Learning (CALL)**ning yuzaga kelishiga olib keldi. CALL boʻyicha 1-xalqaro konferensiya 1985-yil 12 – 13-noyabrda Budapesht shahrida oʻtkazildi.

Oʻqituvchi lingvistik avtomatlardan foydalanishda 2 xil yondashuv mavjud:

1) bixevioristik yondashuv – mavjud anʼanaviy oʻqitish tizimining barcha jabhalarini eʼtiborga oluvchi oʻqitish mexanizmlari, bunda grammatikaning maʼlum

bo'limi tugagach, uni topshiriqlar asosida baholash, shuningdek, leksik minimumlar bo'yicha bilimni nazorat qilishga yo'naltirilgan dasturiy ta'minotga asoslangan 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 immittatsiyasi 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 konsepsiyasi Stenford universiteti professori Eduard Feygenbaum nomi bilan bog'liq bo'lib, u 1977-yilda murakkab mantiqiy masalalarni yechishda kompyuter dasturlarining samaradorligi formallashtirish va dasturlash texnikasidan ko'ra muamlo sohaga doir bilimlar bazasiga bog'liq ekanligini asoslab berdi.

Elektron ensiklopediya – kuchli qidiruv tizimiga ega bo'lgan ma'lum yo'nalishdagi ilustratsiyali 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 hajmdagi elektron dastur.

Elektron lug'atlar – leksikografiya sohasiga kompyuter texnologiyalarining tatbiq etilishi natijasida yaratilgan lug'atlar, ular odatdagi lug'atlar asosida tuzilgan, multimedia vositalari bilan boyitilgan dastur. Eng mashhur elektron lug'atlar sirasiga CONTEXT, ABBY LINGVO, MULTITRAN, POLYGLOSSUM, MULTILEKS kabilar kiradi.

Elektron lug'atlarning tayyorlanishi quyidagicha: a) lug'atning so'zligi shakllantiriladi; b) misollar korpusi (elektron kartotekalari) tuziladi; c) lug'at maqolalari yoziladi; d) lug'at maqolalari ma'lumotlar bazasiga ko'chiriladi; e) bevosita ma'lumotlar bazasida lug'at matni tahrir etiladi, korrektirovka qilinadi; f) lingvistik ta'minot dasturiy ta'minot bilan uyg'unlashtiriladi; g) elektron lug'at.

Elektron lug'atlar an'anaviy kitobiy lug'atlarga qaraganda birmuncha afzalliklarga ega. Elektron lug'atlarning hajm jihatidan ixchamligi, kompaktligi, boshqa manbalarga (Internet tarmog'i orqali, gipermurojaatlar yordamida) ulanish imkoniyati, mavjud matnlar korpuslari yordamida ilustrativ 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'atlarning 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

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 matnlarga tayanib analogiya tamoyili asosida tarjima qilinishi lozim. Buning uchun kompyuterga kiritilgan katta hajmdagi **biliniv bazasi** (matn va uning tarjimalari) shakllantirilishi zarur. Yangi matnlar tarjimasida biliniv 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, ilgari 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 tendensiyasi XX asr o'rtalarida kuchaydi. Bu bevosita tilning strukturasi 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 generativ grammatika, bunda $G=(N, P, S)$ belgilari yordamida ish ko'riladi. N, " – alifbo (N – noterminal simvollar, u grammatikadagi simvollarini o'z ichiga oladi; " – metatilning terminal simvollar, u tildagi so'zlarni qamrab oladi); S – noterminal to'planing boshlang'ich simvoli, P – qayta ishlab chiqilgan qoidalar. Bunda terminal va noterminal simvollarini birlashtiruvchi V belgisi qabul qilingan, u inglizcha Vocabulary (lug'at) so'zidan olingan ($V=TVN$). Bu grammatika o'ta abstraktlashgan 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 programallashtirish tillari unsurlarining generatsiyasida (ifodalar, buyruqlar tizimida) ishlatiladi.

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 namunalarning 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 aniqlashga, iboralarning ham idiomatik xususiyatini belgilashga imkon beradi.

Freym – bilimlar strukturasi o‘id tayanch tushunchalardan biri. Freym – (inglizcha «frame» – «karkas», «ramka», «skelet» ma‘nolarini anglatadi) sun‘iy intellekt tizimida bilimlarni, stereotip (juda ko‘p kuzatiladigan, tipik) vaziyatlarni gavdalanitirishning 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, restoran borish, shifokor qabulida bo‘lish, 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 strukturasi nisbatan ishlatilgan. Shundan so‘ng freym semantikasi konsepsiyasi (**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 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 xazinasini va faktlar omboriga tayanib olamni maxsus shakllantiruvchi model vazifasini bajaradi. Freym **slot** deb atalgan bir yoki bir necha elementlarning majmui hisoblanadi. Slotlarda obyektarga 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

har bir harakat natijasi shartli sanaladi, uning ta'siri va oqibatida navbatdagi ish-harakat sodir bo'lishi mumkin. Har bir ssenariyda rol bajaruvchilar mavjud. Ssenariylar real hayotda uchraydigan vaziyatlarni aks ettiruvchi holatlar asosida yuzaga keladi. Masalan, restoranga tashrif buyurish ssenariysi.

Generativ va transformatsion grammatika – formallashtirishning negizi hisoblangan grammatika, 1950-yillarda dastlab Z.Harrisning til strukturasiidagi transformatsiya 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 shakli o'zgarishlar asosida asosiy ma'noni saqlagan holda ikkilamchi sintaktik qurilmalarning 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 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.

Gipertekst – kompyuter lingvistikasida inqilobiy o'zgarish yasagan texnologiya. Gipertekst g'oyasi AQSH prezidenti F.Ruzveltning fan va ta'lim bo'yicha maslahatchisi Vannevar Bush nomi bilan bog'liq bo'lib, u foydalanuvchiga matn va uning fragmentlarini turli assotsiativ munosabatlar asosida bog'lashga imkon beruvchi «MEMEKS» tizimini nazariy jihatdan asoslab berdi. U 1945-yilda «The Atlantic Monthly» jurnalida «Memeks» deb atalgan kelajakdagi 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, cheklanmagan 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 kiritilgan. 1992-yilda birinchi marta gipertekst World Wide Web (xalqaro to'rga)ga tatbiq etildi.

Gipertekst bu shunday matnki, u kompyuterda boshqa matnlarga giperhavolar yordamida bog'lashga imkon beradi. Bunda u o'quvchiga faqat matnlar ustida emas, balki turli xil jadvallar, sxemalar, rasmlar, videoroliklarda ham erkin navigatsiya qilish imkonini beradi. Asosan, gipertekst World Wide Web (Xalqaro o'rgimchak to'ri) strukturasi 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

beruvchi tizim gipertekst tizimidir. 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.

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, germeneytika (diniy matnlarni sharhlash, talqin etish) qobig'ida bo'lgan. Bunda tilning paydo bo'lishi, 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. 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**» («**Grammatike techné**») 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 grammatikani 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 rekonstruksiya 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 metodini 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 universaliy – Amerikalik olim Jozef Grinberg tomonidan implikatsiyalar (mantiqiy shartli hukmlar) asosida aniqlangan universalialar. Olim 45 ta implikativ universalini aniqlagan. Masalan:

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

2. Agar til istisnoli tarzda suffiksali bo'lsa, u holda bu til poslelolarga, keli-shiklarga boy bo'ladi.

3. Agar til istisnoli tarzda prefiksali bo'lsa, u holda bu til predloglarga boy bo'ladi.

4. Agar til VSO modelida bo'lsa, bu tilda predloglar dominantlik qiladi.

Informatsion qidiruv tizimlari – formal belgilarga ko'ra informatsiyani qidirish-ga mo'ljallangan avtomatik tizim. 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. 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 bo'lib, u IQTda saqlangan hujjatlarning mazmunini tavsiflash va so'rov uchun mo'ljallangan. 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 informatsion qidiruv tillari ishlatiladi. Hujjat mavzusi deskriptorlarning 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 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 qo'llanish chastotasi va matnning predmet mundarijasini aks ettirish darajasiga 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. Shu o'rinda ta'kidlash zarurki, keyingi paytlarda qidiruv tizimlari bazasidagi tezaurlar giperhavolalar bilan ham ta'minlanmoqda, bu esa usurga 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 leksikografiyasi (elektron lug'atlar tuzish) – muayyan til leksikonidagi 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, distributiv holatlari, birikma holdidagi 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 tarjimai bo'yicha dunyoda o'tkazilgan birinchi tajriba asosida 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 usullarining 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, neyron tarmoqlar, lingvoanalizatorlar, gapiruvchi avtomatlar – barchasi kompyuter model-lashtirish natijasi hisoblanadi.

Kontekstga bog'liq bo'lmagan 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 mor-fologik tahlil metodi tarkibiga kiradi, u ikki bosqichni o'z ichiga oladi: 1) dekla-rativ 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 gram-matik 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'kimitir, ko'kish, ko'kroq ! ko'k

Lingvistik protsessorlar – tabiiy tilda kompyuter bilan muloqot qilishni ta'minlashga yordam beradigan dasturlar. XX asrning 70-yillarida AQSHda kom-pyuter bilan cheklangan muloqotni ta'minlaydigan tabiiy tilli interfeys tizimi (**LI-FER** – Language Interface Facility with Ellipsis and Recursion) yaratildi. 1985-yilda «Semantek» korporatsiyasi tomonidan **Q&A** dasturi, «Karnegi grupp» kompaniyasi tomonidan **Language Craft** dasturiy ta'minoti ishlab chiqildi. Bu turdagi dasturlar hozirda ingliz tilining standartlashgan so'zlashuv uslubida un-chakatta bo'lmagan lug'at zaxirasi bilan ishlamoqda.

Lingvistik universaliali – barcha tillarga yoki dunyo tillarining mutlaq ko'pchilik qismiga xos umumiy qonuniyatlar. 1941-yilda Aginskiy tomonidan «**The Impor-tance of language universals**» sarlavhali maqola e'lon qilindi. 1961-yilda Nyu-York shahrida universalialarga bag'ishlangan xalqaro konferensiya o'tkazildi. Unda J.Grinberg, J.Jenkins, Ch.Osgudlar «**Lingvistik universalialar haqida memoran-dum**» mavzusida chiqish qildilar. Konferensiyada Ch.Osgud shunday yozadi: «*Biz ushbu konferensiyada qonsiz inqilobning guvohi bo'lamiz*».

Fonologiya, morfologiya, sintaksis, semantika sohalarida aniqlangan lingvis-tik universalialar 2000-yilga kelib 1600 tani tashkil etdi. Germaniyaning Konstanz universitetida lingvistik universalialar bazasi (**Universals Archive**) yaratildi

va u har yili yangi ma'lumotlar bilan boyitib boriladi. Mazkur baza Internetda <http://www.ling.uni-konstanz.de/pages/proj/sprachbau.htm> saytiga joylangan.

Matematik lingvistika – XX asrning 50-yillarida (1952-yilda) yuzaga kelgan amaliy tilshunoslikning alohida yo'nalishi bo'lib, unda tabiiy tillarning matematik modellarini (bunday formallashtirilgan til **metatil** deb ataladi) ishlab chiqish, xususan, sun'iy tillarni yaratish algoritmini tuzish bilan shug'ullanuvchi fandır. Bu fanning shakllanishida Kopengagen struktural tilshunoslik maktabi (glossematika)ning asoschisi Lui Yelmslevning g'oyalari o'ziga xos «turtki» vazifasini 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'nalishi 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 birikmasi 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, inglizcha *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 Internetdagi qidiriv saytlari, elektron kutubxonalar, virtual ensiklopediyalar korpus vazifasini bajarmoqda. Korpusning janri va tematik rang-barangligi Internetdan foydalanuvchining qiziqishlariga bog'liq. Masalan, ilmfan doirasida **Wikipedia** katta hajmdagi matnlar korpusi sifatida ishlatilmoqda.

Matnning statistik tahlili – tilni ehtimollik belgilariga ko'ra tahlil qilish, til hodisalarini statistik xarakteristikalarini asosida baholash demakdir. Til birligining gapda yoki matnda qo'llanish chastotasi (sur'ati) ehtimollik deyiladi. Til hodisalarini haqidagi haqiqatlar lingvostatistik tadqiqotlar yordamida aniqlanadi. Masalan, o'zbek tili uchun qaysi bo'g'in strukturasi tipik ekanligini, gap modellaridan qaysi biri eng asosiy konstruksiya ekanligini, bosh kelishikdagi otlarning asosan qay-

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 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 idiositilini – o'ziga xos uslubini ko'rsatuvchi vositalar) lingvostatistik tahlil yordamida aniqlanadi. Turli matnlardan olingan dalillarning qiyosiy tahlili bizga o'sha matnning mazmun-mundarijasini, matn tuzilgan davrni, dalillarni argumentlash xarakterini va hatto mualliflikni aniqlashga imkon beradi. Badiiy asarlarning sujet qurilishiga, matn strukturasi qantitativ 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.

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 musiqa yaratilishi, treker-dasturlar, sintezatorlar, sekvensorlar yordamida bir turdagi melodiya 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 universalialar g'oyasi ta'sirida metagrammatika haqida qarashlar, nazariy va amaliy tadqiqotlar yuzaga keldi. 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-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 demakdir. Boshqacharoq tushuntirganda, model tabiiy obyektlarning imitatsiyasidir (o'xshashi, taqlidiy ko'rinishi), u o'zbek tilidagi qolip, andaza so'zlariga mos

keladi. Masalan, globus yerning modeli, o'yinchoq mashina haqiqiy mashinaning modeli sifatida baholanishi mumkin.

Modellarni shartli ravishda 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 material bilan farq qiladi; 2) matematik modellar – prototiptan (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 jarayonini o'rganishda qo'llaniladi; 4) kompyuterli modellar – matematik va mantiqiy modellashtirish metodlariga asoslangan holda kompyuterda algoritm va dasturlardan 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 umumiylikka 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 unurni o'z ichiga oladi:

– subyekt (tadqiqotchi);

– tadqiqot objekti;

– o'rganuvchi subyekt va o'rganiluvchi obyekt munosabatini aks ettiruvchi model.

Modellashtirish tilshunoslikka strukturalizm yo'nalishi ta'sirida faol tatbiq qilina boshlandi. Gap strukturasi modellashtirish g'oyasi XX asrning 50-yillarida 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*» jumlasini distributiv model asosida quyidagicha tahlil etiladi: **D 3 I^a 2-d D I^b 4**. Bu yerda **D** – o'ning aniqlovchisi (inglizcha determiner), **3** – sifat, **I** – 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 strukturasi modellashtirishning asosiy mezonini sanaladi.

Neyron tarmoqlar – inson miyasidagi neyronlar modeli asosida yaratilgan tizim bo'lib, u sun'iy intellektning muhim tarkibiy qismi hisoblanadi. Hozirgi kunda neyron tarmoqlar bir qator amaliy masalalarni hal etishda keng qo'llanilmoqda.

Jumladan, ob-havo ma'lumotini aniqlashda, obrazlarni tanishda (xususan, og'zaki va yozma 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 namunalar (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 jumlangan, 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», «tasvirlash», «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 tilshunoslikda, 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 strukturasi 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 jihatlarni 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 munosabatlarning 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 obyektlarga, muayyan sohaga oid tushunchalarga, dugalar obyektlar o'rtasidagi semantik munosabatlarga aloqador tushunchalardir.

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 imitatsiyasi 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-yilda Darmut universitetidagi konferensiyada qo'lladi.

Tobelik daraxti (shajara) – L.Tenyer g'oyalari ta'sirida shakllangan tobelik grammatikasi modeli bo'lib, u kompyuter lingvistikasining avtomatik analiz dasturlarida 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 bo'laklari) 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 jumalarga, 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 jumalarni sifatli tarjima qilishga imkon beradi. 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 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 protsedurasini ishlab chiqdi. Ushbu g'oya 1950-yilda «Mind» falsafiy jurnalida «*Hisoblash mashinalari 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-

ter va bir inson bilan o'zaro munosabatga kirishadi. Savollarga javob berish asosida u kim bilan (inson yoki kompyuter dasturi bilanmi) gaplashayotganini aniqlashi lozim bo'ladi. Kompyuter dasturining vazifasi esa insonni xatoga yo'l 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 kuzatilgan so'z $r = 2$, undan keyingisi $r = 3$ tarzida belgilanadi. Bundan kelib chiqadiki, matndagi ixtiyoriy so'z uchun (f) chastotaga (r)ning ko'paytmasi natijasi o'zgarmas miqdor – konstanta (C)ga teng bo'ladi ($r \cdot f = C$), bu yerda C 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
Kompyuter lingvistikasi terminlari lug'ati (Glossariy)	142

Ilmiy nashr

RAHIMOV AZAMATJON USMANJANOVICH

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'oz. Qog'oz bichimi: 60x84 $\frac{1}{16}$. LexTimes garnituras. 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, 20^A-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.