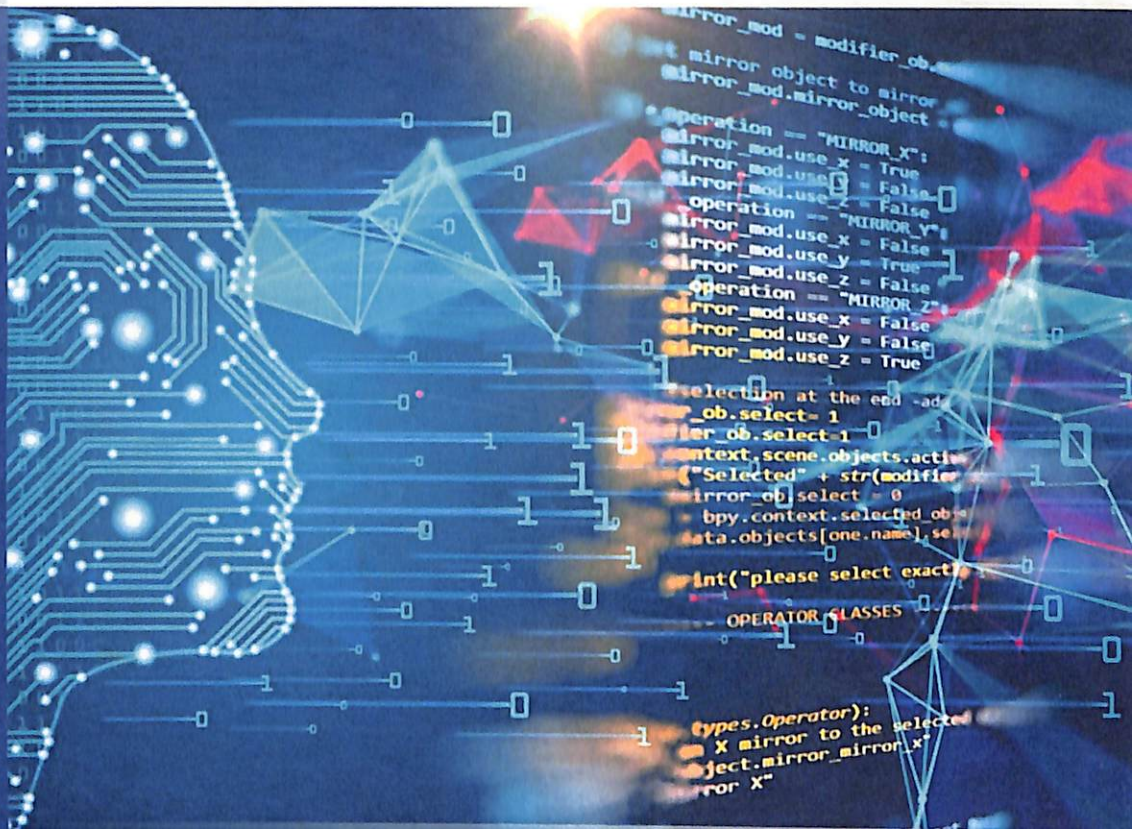


NILUFAR ABDURAXMONOVA

KOMPYUTER LINGVISTIKASI



O'ZBEKISTON RESPUBLIKASI
OLIV VA O'RTA MAXSUS TA'LIM VAZIRLIGI

ALISHER NAVOIY NOMIDAGI TOSHKENT DAVLAT
O'ZBEK TILI VA ADABIYOTI UNIVERSITETI

ABDURAXMONOVA NILUFAR ZAYNOBIDDIN QIZI

KOMPYUTER LINGVISTIKASI

(darslik)

“Nodirabegim” nashriyoti
Toshkent – 2021

iy
liq
ial
ng
jal
an
iro
til
ter
lar
us
iy
mi
ga

lar
bir
da
oir
iy
on
ha
da
ari
da
lab
pot
ing
lan

yil
tili
li,

TOH
Rial K
A
A

KBK: 74.268.1

A 15

UO'K: 37.016:811.512.133:004(075)

ISBN 978-9943-6941-2-5

Abduraxmonova Nilufar Zaynobiddin qizi

Kompyuter lingvistikasi [Matn] : darslik / N.Z. Abduraxmonova.
- Toshkent: Nodirabegim, 2021. – 394 b.

“Kompyuter lingvistikasi” darsligi kompyuter texnologiyalari va metodlari yordamida tabiiy tillarni qayta ishlashning lingvistik asoslarini yaratish hamda o‘zbek tili kompyuter lingvistikasiga tegishli nazariy va amaliy masalalarni o‘rganishga bag‘ishlangan.

Darslik 5111200 – O‘zbek tili va adabiyoti, 5120100 – Filologiya va tillarni o‘qitish (o‘zbek tili), 5120900 – O‘zbek-ingliz tarjima nazariyasi va amaliyoti yo‘nalishlari bakalavr talabalari, magistrlar, doktorantlar, tadqiqotchilar va soha mutaxassislari, shuningdek, malaka oshirish va pedagogik kadrlarni qayta tayyorlash kursi tinglovchilariga mo‘ljallangan.

Mas’ul muharrir:
f.f.d., prof. S. Muhamedova

Taqrizchilar:

tex.f.d., prof. M. Aripov (Mirzo Ulug‘bek nomidagi O‘zbekiston Milliy universiteti)

fil.f.d., prof. B. Mengliyev (Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti)

tex.f.n, dots. Ayrat Gatiatullin (Tatariston Respublikasi Fanlar Akademiyasi, Amaliy semiotika instituti)

Ushbu darslik O‘zbekiston Respublikasi Oliy va o‘rta maxsus ta’lim vazirligining 2019-yil 20-iyuldagi 654-sonli buyrug‘iga asosan chop etishga tavsiya qilingan.

© N.Z. Abduraxmonova.
© “Nodirabegim” nashriyoti, 2021.

KIRISH

Bugungi kunda tabiiy tilning lisoniy modellari va nutqiy imkoniyatlarini kompyuter tiliga o‘tkazish hamda til bilan bog‘liq masalalarni axborot texnologiyalari va metodlari yordamida hal qilish muhim masala hisoblanadi. Axborot asrida barcha tillarning rivojlanish barqarorligi, ularning milliy sofligini saqlab qolish global muammolardan biridir. Kompyuter texnologiyalari shiddat bilan rivojlanib borayotgan davrda har bir fan rivoji uchun sohalaro hamkorlik muhim sanaladi. Kompyuter texnologiyalarining til rivojiga va aksincha, til texnologiyasining kompyuter texnologiyalariga ijobiy ta’siri natijasida ko‘plab ilmiy o‘zgarishlar va yutuqlarga erishildi. Buning natijasida mashina tarjimasini, korpus lingvistikasi, kompyuter leksikografiyasi, tahrirlovchi dastur, nutqiy sintezator, lingvostatistik analiz qiluvchi dastur, matnlarni referatlash, matnlarni tasniflash kabi qator yo‘nalishlar vujudga keldi.

Mamlakatimizda jamiyatning barcha sohalarida izchil islohotlar amalga oshirilib, ilm-fan jadallik bilan takomillashtirilayotgan bir vaqtda kompyuter lingvistikasini dolzarb sohalardan biri sifatida qayd etish mumkin. O‘zbek kompyuter lingvistikasiga doir tadqiqotlarning yetarli darajada emasligi ushbu sohada jiddiy ilmiy tadqiqotlar olib borish zaruriyatini ko‘rsatadi. O‘zbekiston Respublikasini rivojlantirishning beshta ustuvor yo‘nalishi bo‘yicha 2017-2021-yillarga mo‘ljallangan Harakatlar strategiyasida “...ta’lim tizimini takomillashtirish, sifatli ta’lim xizmatlari imkoniyatlarini oshirish”ga alohida e’tibor qaratilgan¹. Bu borada kompyuter lingvistikasi fanini chuqurroq o‘rganish va ishlab chiqarishga yo‘naltirish, uning sohalar kesimidagi tadqiqot predmetlarini belgilashda nazariy va amaliy jihatdan fanning umummetodologik konsepsiyasini yaratish muhim masalalardan sanaladi.

Ushbu darslik O‘zbekiston Respublikasi Prezidentining 2016-yil 13-maydagi “Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universitetini tashkil etish to‘g‘risida”gi PF-4797-sonli,

¹ Ўзбекистон Республикаси Президентининг 2017 йил 7 февралдаги “Ўзбекистон Республикасини янада ривожлантириш бўйича Харажатлар стратегияси тўғрисида”ги ПФ-4947-сонли Фармони.

2017-yil 7-fevraldagi “O‘zbekiston Respublikasini yanada rivojlantirish bo‘yicha Harakatlar strategiyasi to‘g‘risida”gi PF-4947-sonli farmonlari, 2013-yil 27-iyundagi “O‘zbekiston Respublikasi Milliy axborot-kommunikatsiya tizimini yanada rivojlantirish chora-tadbirlari haqida”gi PQ-1989-sonli, 2017-yil 17-fevraldagi “Fanlar akademiyasi faoliyati, ilmiy tadqiqot ishlarini tashkil etish, boshqarish va moliyalashtirishni yanada takomillashtirish chora-tadbirlari to‘g‘risida”gi PQ-2789-sonli qarorlari hamda mazkur faoliyatga tegishli me‘yoriy-huquqiy hujjatlarda belgilangan vazifalarni amalga oshirishda muayyan darajada xizmat qiladi.

Jahonda kompyuter lingvistikasining turli yo‘nalishlari bo‘yicha ko‘plab olimlar tomonidan ilmiy izlanishlar olib borilgan². Endilikda til texnologiyasi, tabiiy tilni qayta ishlash (NLP), mashina ta‘limi (machine learning), korpus texnologiyalari kabi qator sohalar kesimida kompyuter lingvistikasiga doir qator muammolar tadqiqotchilar tomonidan hal etib kelinmoqda.

O‘zbekistonda kompyuter lingvistikasining shakllanishida dastlab tadqiqot olib borgan N. Yoqutova, M. Ayimbetov, S. Rizayev va S. Muhamedovlarning o‘zbek tili materiallari bo‘yicha lingvostatistikaga oid ilmiy izlanishlari diqqatga sazovor³. S. Muhamedov R. Piotrovskiy bilan hammualliflikda yozilgan

“Инженерная лингвистика и опыт системно – статистического исследования узбекских текстов” nomli kitobida lingvistik modellar, modellashtirish va uning umumiy tamoyillari o‘rganilib, o‘zbekcha matnlarning kvantativ modellari yaratilgan⁴. Shuni alohida qayd etish joizki, kompyuter lingvistikasining fan sifatida shakllanishida professor Abdumajid Po‘latov yaratgan ilmiy maktab katta ahamiyatga ega. Shuningdek, ushbu sohaning turli yo‘nalishlarida muayyan darajada ilmiy tadqiqotlar olib borilgan⁵.

Ushbu darslikni tayyorlashda kompyuter lingvistikasiga doir muallif tomonidan tayyorlangan o‘quv-uslubiy majmua, “Inglizcha matnlarni o‘zbek tiliga tarjima qilish dasturining lingvistik ta‘minoti (sodda gaplar misolida)” mavzusidagi (PhD) dissertatsiyasi, “Mashina tarjimasining lingvistik ta‘minoti” nomli monografiya, “Mashina tarjimasining lingvistik asoslari” nomli o‘quv qo‘llanmalardan⁶ zarur o‘rinlarda foydalanilgan.

Umid qilamizki, ushbu kitob o‘quvchilarga manzur bo‘ladi. Darslikning kengaytirilgan va yanada to‘ldirilgan keyingi nusxasi uchun sizning taklif va mulohazalaringizni kutib qolamiz.

² Jurafskiy D., James H. Martin Speech and language processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 2007 -P. 12-13– P. 140.; Karin Aijmer, Bengt Altenberg Advances corpus-based contrastive linguistics. USA: John Benjamins, 2013.; Koehn P., Och F.J., Marcu D. Statistical phrase based translation. In Proceedings of the Joint Conference on Human Language Technologies and the Annual Meeting of the North American Chapter of the Association of Computational Linguistics (HLT/NAACL). 2003.; Mitkov R. The Oxford handbook of Computational linguistics. Oxford university press, 2003.; Mohamed Zakaria Kurdi. Natural Language Processing and Computational Linguistics: Speech, Morphology and Syntax. —Great Britain, USA: Wiley-ISTE 2016. 300 p.

³ Ризаев С. Ўзбек тилининг лингвостатистик тадқиқи: филол. фан. д-ри. ... дисс. автореф. – Тошкент, 2008. – 50 б.; Мухамедов С.А. Статистический анализ лексико-морфологической структуры узбекских газетных текстов: автореф. дисс. ... канд. филол. наук. – Ташкент, 1980. – 25 с.; Бектаев К.Б., Пиотровский Р.Г. Математическая лингвистика. – М.: Высшая школа, 1997. – 420 с.; Айымбетов М.К. Проблемы и методы квантитативно-типологического измерения близости тюркских языков (на материалах каракалпакского, казахского и узбекского языков): Автореф. дисс. ... д-ра филол. наук. – Ташкент, 1997. – 47 с.

⁴ Мухамедов С.А., Пиотровский Г.Г. Инженерная лингвистика и опыт системно-статистического исследования узбекских текстов. – Ташкент: Фан, 1986; Махмудов М.А., Пиотровская А.А., Садыков Т. Система машинного анализа и синтеза тюркской словоформы // Переработка текста методами инженерной лингвистики. – Минск, 1982. 92 с.

⁵ Pulatov A.K., Muhamedova S. Kompyuter lingvistikasi (o‘quv qo‘llanma). –Toshkent, 2008. 80 b.; Pulatov A. Kompyuter lingvistikasi. –Toshkent: Akademnashr, 2011, 350b; Rahimov A. Kompyuter lingvistika asoslari. –Toshkent: Akademnashr, 2011, 160 b.

⁶ Абдурахмонова Н. Машина таржимасининг лингвистик таъминоти (содда гаплар мисолида): филол. фан. д-ри. ... дисс. автореф. – Тошкент, 2018. – 65 б.; Abduraxmonova N. Mashina tarjimasining lingvistik ta‘minoti (monografiya). – Tashkent: Muharrir, 2018. – 165 b.; Abduraxmonova N. Mashina tarjimasining lingvistik asoslari (o‘quv qo‘llanma). – Tashkent: Akademnashr, 2012. – 124 b.

1-MODUL KOMPYUTER LINGVISTIKASINING RIVOJLANISH TENDENSIYASI

1-mavzu. "Kompyuter lingvistikasi" faniga kirish

1.1. Fanning maqsadi va vazifasi

1.2. Kompyuter lingvistikasi fan sifatida rivojlanishida nazariy va amaliy tadqiqotlar

1.3. Turkologiyaning yetakchi maktablarida olib borilayotgan tadqiqotlar

1.4. Kompyuter lingvistikasining boshqa fanlar bilan aloqasi (matematika, psixologiya, informatika, dasturlash asoslari)

Tayanch soʻz va iboralar: *tilni avtomatik qayta ishlash, mashina tarjimasini, matnlarni ovozlashtirish, tilga oʻqitish, bilimlarni baholash, axborot qidiruvi, erkin kontekstli grammatika, ehtimollik nazariyasi, kiberfazo, raqamli insoniyat (digital humanity), savol-javob tizimi.*

1.1. Fanning maqsadi va vazifasi

Bugungi globallashtirish va integrallashtirish jarayonida kompyuter texnologiyalari kiberfazoning muhim va ajralmas qismiga aylanib ulgurdi. Sohalarda fanlarning yuzaga kelishida kompyuter taraqqiyotida erishilgan natijalar kuchli taʼsir qilganligi hech kimga sir emas. Endilikda filologik sohalarining axborot texnologiyalariga va aksincha, axborot texnologiyalarning filologik taʼlimga taʼsirlashuvi natijasida turli murakkab masalalar oʻz yechimini topib kelmoqda. Mamlakat hayotida yuz berayotgan jiddiy oʻzgarishlar va yangilanishlar, shubhasiz, til texnologiyasida amalga oshirilayotgan yangi qarashlarni oliy taʼlimda ham joriy etishni taqozo etmoqda.

Til ijtimoiy hodisa ekanligi undagi muammolarni yechishda axborot kommunikatsion texnologiyalarning oʻrni beqiyos. Bu borada Kompyuter lingvistikasi fani muhim ahamiyat kasb etadi. Lingvistik bilimlarni rivojlantirish boʻyicha dunyo tajribasiga tayanish hamda erishilgan yuksak natijalardan oqilona foydalanish talab etiladi.

Bu borada olimlar tomonidan kompyuter lingvistikasi haqida turli munosabalar va mulohazalar bildiriladi. *Kompyuter lingvistikasi informatika uchun lingvistik taʼminotni yaratish*

*masalalari bilan shugʻullanadi. Informatika fan sifatida zamonaviy texnik vositalar yordamida yozuv qonuniyatlari, axborot toʻplash, qayta ishlash, uzatish va undan foydalanish masalalarini hal qiladi*⁷. Yana bir taʼrifda kompyuter lingvistikasi kompyuter nuqtayi nazaridan tabiiy tilni ilmiy tadqiq qiluvchi soha ekanligi, lingvistik qoidalarning kompyuter modellarini yaratilishiga doir fikrlar keltiriladi⁸.

Kompyuter lingvistikasi kompyuter injeneriya sohasi bilan aloqadorlikda tilning yozma va ogʻzaki nutqni tushunish va qayta ishlash jarayoniga tegishli masalalarni oʻrganadi. Kompyuter lingvistikasi tilning tafakkur mahsuli va muloqot vositasi sifatida uning nutqiy vaziyatdagi xoslanishini kompyuter orqali tahlil qilish hamda uning kompyuter modeli, shuningdek, lingvistik protsessor va unga mos dasturiy taʼminotlarini yaratish kabi vazifalarni hal qiladi.

Kompyuter lingvistikasi fanining maqsadi tabiiy til bilan bogʻliq masalalarni kompyuter texnologiyalari va metodlari orqali oʻrganadi, tegishli muammolar yuzasidan oʻz tavsiyalarini yaratadi, metodlarini ishlab chiqadi. Kompyuter lingvistikasi bir necha fanlarning integrallashtirish natijasida sohalararo hamkorlikni talab qiladi. Xususan, matnlarni ovozlashtirishda fizik akustika, mashina tarjimasida statistik va matematik modellashtirish, avtomatik tahrirda dasturiy taʼminot bilan bogʻliq boʻlgan metrikalar, kompyuter leksikografiyasida maʼlumot bazasini loyihalash kabilardan unumli foydalaniladi. Shuningdek, kompyuter lingvistikasi fani erishilgan ilmiy yutuqlar (texnologiyalar)dan foydalanib, yangi metodlar va lingvistik instrumentlar va resurslarni yaratadi.

⁷ Марчук Ю.Н. Компьютерная лингвистика (учеб. пособ.) –Москва: Восток Запад, 2007. – С. 13.

⁸ Sakthi Vel Applications of Computational Linguistics to Language Studies: An Overview // International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 4, Issue 3, March 2017.

Kompyuter lingvistikasining vujudga kelishi va rivojlanishi

Dastlab 1952-yili Massachusetts texnika universitetida Mashina tarjimai bo'yicha o'tkazilgan ilk ilmiy-amaliy konferensiya hamda 1954-yil 7-yanvarda Nyu-Yorkdagi Jorjtaun universiteti bilan IBM kompaniyasi hamkorligida birinchi IBM II mashina tarjimai dasturining amalda sinovdan o'tkazilganligini Kompyuter lingvistikasining yaratilishiga omil bo'lgan muhim voqealik sifatida baholash mumkin. Shu bois Martin Kay (2003) kompyuter lingvistikasining boshlanish davri 1949-yil mashina tarjimai bilan bog'liq ekanligini ta'kidlaydi.

Kompyuter lingvistikasi kompyuter texnologiyalari yordamida til bilan bog'liq masalalarni o'rganuvchi soha ekanligi yuqorida qayd etib o'tdik. *Kompyuter lingvistikasi* termini fanda turli nomlar bilan nomlanib kelingan. Ruslan Mitkovning e'tirof etishicha, "*Computational linguistics*" termini 1960-yillarda Milliy Fanlar akademiyasining Tilni avtomatik qayta ishlash maslahat komiteti a'zosi David Hays tomonidan kiritilgan⁹. Mashina tarjimasidan Kompyuter lingvistikasining ajralib chiqishi 1974-yilga to'g'ri keladi. Shundan so'ng "Mashina tarjimai va Kompyuter lingvistikasi" jurnalining o'rniga ilk bor "Kompyuter lingvistikasi" nomi ostida Amerika jurnali chop etiladi.

Kompyuter lingvistikasi mashina tarjimasiga doir tadqiqotlardan so'ng tilning kompyuterga oid formal modellarini rivojlantirish mahsuli sifatida nazariy va amaliy jihatdan rivojlanib bordi.

D.Jurafskiy Kompyuter lingvistikasining rivojlanish bosqichlarini quyidagicha davriy bosqichlarda tasniflaydi¹⁰. Dastlabki bosqich 1940-1950-yillarga to'g'ri keladi. 1940-yilning oxiri 50-yillarning boshida ikki asosiy paradigma avtomatlashtirish hamda ehtimollik yoki informatsion-nazariy modellar asosida tadqiqotlar olib borildi. Avtomatlashtirishning o'tgan asr 50-yillarida amalda tatbiq etilishi algoritmik hisoblashning Tyuring modelini yaratilishiga (1936) asos bo'ldi. Tyuring testi Mak Kuloch-Pitts

⁹The Oxford handbook of computational linguistics (edited by Mitkov R.) – Oxford, 2003. – P. 25

¹⁰Jurafskiy D., James H. Martin Speech and language processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 2007. – P. 9-14.

neyron tizimini vujudga kelishiga asosiy sabablardan biridir. Hisoblash elementi sifatida neyronning soddalashgan modeli mantiqiy mulohazalarning birligiga aylandi. Klinning ishlarida (1951, 1956) chekli avtomat (fine-state automaton) va o'zgarmas (regulyar) ifodalarga oid tadqiqotlar o'z aksini topdi. Shannon (1948) tillarni avtomatlashtirishda diskret Markov jarayonining amaliy ehtimollar modelidan foydalanib, uni amalga tatbiq etdi. Chomskiy (1956) mashinaning muayyan usuli sifatida avtomatlashgan grammatika yordamida tilni umumlashtirish orqali til haqidagi fikrlarni asoslab berdi. Ushbu modellar o'z navbatida formal til nazariyasining vujudga kelishiga turtki bo'ldi. Belgilarning ketma-ketligi sifatida algebra va ko'plik nazariyasi formal grammatikaning shakllanishida muhim omil vazifasini o'tadi. Bu Chomskiy tomonidan kiritilgan tabiiy tillar uchun yaratilgan erkin kontekstli grammatika (context-free grammar) nazariyasi bilan bog'liqdir. Biroq bu mustaqil holda Baskus (1959) hamda Naurlar (1960) tomonidan yaratilgan ALGOL dasturlash tilida o'z mohiyatini yanada ochib berdi.

Kompyuter lingvistikasi paydo bo'lishiga asos bo'lgan yana bir omillardan biri til va nutqni qayta ishlash uchun ehtimollik algoritmlarning rivojlantirilishidir. Shannonning ushbu masalalarga bag'ishlangan ilmiy ishlarida aloqa kanali va tilning ma'lumot hajmini o'lchash uchun termodinamikaning entropiya konsepti ustuvorlik qildi. Olim ehtimollik nazariyasidan foydalanib, ingliz tilining birinchi entropik hajmini aniqlab bera olgan. Ovozli spektrograf instrumental fonetika asosida rivojlantirildi. Bu esa dastlabki urinishlar sifatida nutqiy holatni avtomatik tushunish bosqichiga ko'tara oldi. 1952-yilda Bell laboratoriyasining tadqiqotchilari tomonidan inson nutqidagi 10 ta raqamdan birini aniqlay oladigan statistik tizimi yaratildi.

Keyingi davr 1957-1970-yillarni o'z ichiga oladi. O'tgan asrning 50-yillar oxiri va 70-yillar boshlarida Kompyuter lingvistikasi ramziy (simvolik) va ehtimollik nazariya kabi ikki paradigma asosida rivojlantirildi. Simvolik paradigma ikki yo'nalishda formal til nazariyasiga asoslangan Chomskiy ta'limoti, sintaktik analiz (parsing) algoritmi hamda dinamik dasturlash yordamida pastdan yuqoriga (bottom-up) va yuqoridan pastga (top-down) kabi metodlar

tadqiq etildi. Pensilvaniya universitetida 1958-59-yillar davomida to'liq sintaktik analiz tizimini joriy etish masalasi Zelig Xarrisning transformatsiya va diskurs analiziga bag'ishlangan loyihasida o'z ifodasini topdi. Ikkinchi yo'nalish sun'iy intellekt sifatida shakllandi. 1956-yil yoz oyida Mak Karti, Marvin Minskiy, Klauud Shennon va Nazaniel Rochesterlar ikki oylik seminarda ishtirok etishib, sun'iy intellekt terminini qo'llashga kelishib oladilar. Tadqiqotchilarning sun'iy intellekt sohasidagi ilmiy ishlarida ehtimollik nazariyasi va statistik algoritmgaga asoslangan metodlar faol qo'llanilmagan bo'lsa-da, Nivel va Simyonlar tomonidan *mantiqiy nazariyaga* asoslangan umumiy muammolarga asosiy e'tibor qaratiladi. Buning natijasida tabiiy tilni tushunuvchi avtomatik tizim yaratiladi. 1960-yillarning oxiriga qadar formal mantiqiy tizim izchil rivojlantirilgan.

Mazkur ehtimollik paradigmalari statistika va elektr muhandisligida keng qo'llanildi. 1950-yillarning oxirida optik xarakterdagi Bayesov metodi nutqni tushunish muammosini hal etish uchun qo'llanila boshlandi. Bledsov va Brovning tadqiqotlari natijasida (1959) katta hajmdagi lug'at va kompyuterlashtirilgan matnlarni o'qiy oladigan (text recognition) *Bayesov tizimi* yaratildi.

1960-yillarda dastlabki transformatsion metod asosida tabiiy tilni qayta ishlash jarayonini tekshiruvchi dasturning psixologik modeli tuzildi. Shu bilan birga, dastlabki internet tarmog'idagi korpus: turli janrga oid manbalar (gazeta, badiiy, ilmiy, roman va boshq.)dan olingan 500 ta yozma matn jamlanmasidan 1 mln.ta so'zning Amerika ingliz tilisi bo'yicha BROWN korpusi yaratildi.

Kompyuter lingvistikasining navbatdagi bosqichi 1970-1983-yillariga to'g'ri keladi. Bu davrda til va nutqni qayta ishlash va boshqa ko'plab tadqiqotlar soha rivojiga ijobiy ta'sir ko'rsatdi. Nutqni tanuvchi dastur algoritmlarini rivojlantirishda ehtimollik paradigmasi muhim rol o'ynadi. Ayniqsa, Yashirin Markov Modelidan foydalanish hamda Jelinek, Bal, Merseur va ularning IBMdagi hamkorlari Tomas J. Vatson tomonidan shovqinli kanal modeli va qayta kodlash tizimi rivojlantirildi. Kolmerauer va uning hamkasblari tomonidan Q-tizimli va transformatsion grammatika uchun mantiqqa asoslangan paradigmalardan foydalanildi.

Shu davrda tabiiy tilni qayta ishlash jarayoni Terri Vinogradning SHRDLU tizimi orqali qayta rivojlantirildi. Bu dastur tabiiy tildagi matnni qabul qila olish imkoniyatiga ega ekanligi bilan ahamiyatlidir. Xallidey tomonidan sistem grammatikaga asoslangan kengaytirilgan ingliz tili grammatikasi yaratildi. Vinogradov modeli semantik va diskurs modelning tahlil muammosini yechishga yordam berdi. Roger Shank va uning hamkasblari tomonidan inson konseptual bilimi va xotirasiga asoslangan tilni tushuna oladigan bir nechta dasturlar amalda joriy etildi.

LUNAR deb nomlangan savol-javob tizimida mantiqqa asoslangan tabiiy tilni qayta ishlash paradigmalari semantik ifoda sifatida qo'llanilib, predikativ mantiqqa birlashtirildi (Woods, 1967, 1973). Empirizm va avtomatlashgan modellarning vujudga kelish davri 1983-1993-yillarga to'g'ri keladi. 1950-yillar oxiri va 1960-yillar boshida Chomskiyning formal grammatikasidan farqli bo'lgan ikki model o'rganila boshlandi. Nutqni tanishda ehtimollik modellarni qo'llash Tomas Vetson tomonidan olib borildi. Ehtimollik metodlari va shunga o'xshash ma'lumotlarni boshqarish tizimlari semantika bilan bog'liq ko'p ma'nolik kabi masalalarni yechishda foydalanilgan. Shuningdek, ushbu davrda turli shakldagi matnlarning korpuslari yaratilgan. Bu resurslarning amalda qo'llanilishi semantik tahlil, mashina tarjimai, savol-javob, ma'lumotlarni boyitish kabi bir qator amaliy masalalarni yechishda yordam berdi.

Endilikda kompyuter lingvistikasi doirasida shug'ullanuvchi soha vakillarini ikki guruhga ajratish mumkin. Birinchi guruh lingvistika sohasi vakillari bo'lib, tilga kompyuter texnologiyalarining tatbiqi borasida izlanishlar olib boradi: ikkinchi guruh kompyuter texnologiyalari vakillari bo'lib, tabiiy tilni qayta ishlash texnologiyasi (NLP – natural language processing) bo'yicha tilga doir masalalar bilan shug'ullanadi. Har ikki guruh o'zaro ilmiy yutuqlaridan oziqlanadi va til texnologiyasining rivojlanishiga ko'maklashadi.

Bugungi kunda kompyuter lingvistikasi bilan bir qatorda tabiiy tilni tushunish (Natural language understanding) va mashinani o'qitish (machine learning) kabi yo'nalishlar ham mavjud bo'lib,

tadqiqot predmeti til va nutq hodisalariga doir masalalarni o'rganishga qaratilgan.

Til tafakkur bilan bog'liq ijtimoiy hodisa, shu bilan birga muayyan darajada modellashtirilgan va tizimlashtirilgan belgilar majmui hamdir.

Bugungi kunda kompyuter lingvistikasi bilan til texnologiyasiga aloqador boshqa soha mutaxassislarining hamkorligi natijasida ushbu fanning yo'nalishlari (mashina tarjimasi, matnlarni ovozashtirish, tilga o'qitish, bilimlarni baholash, axborot qidiruvi, avtomatik tahrirlash, sentiment analiz, korpus kabi) doirasida ijobiy yutuqlarga erishilmoqda.

1.2. Sohada olib borilayotgan tadqiqotlar

Jahon kompyuter lingvistikasining tadrijiy rivojlanishi natijasida turli ilmiy maktablarga asos solindi. Kompyuter lingvistikasi bo'yicha xalqaro tillar doirasida olib borilgan tadqiqotlarning ko'lami keng va rang-barang. Shu bois ushbu sohadagi barcha mutaxassislarni nomma-nom qayd etishning imkoni yo'q. Biroq dastlabki ishlar sifatida kompyuter lingvistikasining shakllanishiga turtki bergan soha vakillari sifatida kompyuter lingvistikasida rus tiliga doir matnlarni avtomatik tahrirlovchi dastur, mashina tarjimasi, formal til, tabiiy tilni modellashtirish, nutq sintezatori kabi yo'nalishlarda tadqiqotlar olib borgan 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. Kuznetsovlarning ishlari ahamiyatga ega. Shuningdek, mashina tarjimasi yo'nalishida tadqiqot olib borgan N.D. Andreyev, I.A. Melchuk, I.I. Revzin, V.Y. Rozensveyg, Y.N. Marchuk, R.G. Piotrovskiy, Y.A. Motorin, K.B. Bektayev, A.N. Belyayev, I.K. Belskiy, A.V. Zubov, G.E. Miram, L.L. Nelyubin, V.I. Perebinyos, V.A. Chijakovskiy, Y.A. Shingarev, G.G. Belonogov, R.G. Kotov, Babushkina N.V., Z. Shalyapina, O.Y. Mansurova, A.S. Panina, A.A. Xoroshilov hamda boshqa millat vakillari A. But, R. Richans, J. Xatchins, J. Allen, P. Braun (AQSh); M. Nagao (Yaponiya); A. Vaxer (Estoniya); Fransiyada J. Astrouni; R. Sinha, A. Jain (Hindiston); B. Bleyzer, U. Shvol, A. Storrer (Germaniya) kabi tadqiqotchilarning ishlarini alohida e'tirof etmoqchimiz.

Mashina tarjimasining qoidaga asoslangan, frazaga asoslangan hamda statistikaga asoslangan texnologiyalarni vujudga kelishida turli soha vakillarining o'zaro hamkorlikdagi mahsuli deb aytish mumkin. Ushbu soha ham amaliy, ham nazariy jihatdan qator nomzodlik va doktorlik dissertatsiyalarida o'z ifodasini topdi. Ayniqsa, o'tgan asrning 60-80-yillarida mashina tarjimasi bo'yicha ko'plab ilmiy va amaliy ishlar yuzaga keldi.

Turkologiyada ushbu sohada turk, tatar, o'zbek, qirg'iz, qozoq, uyg'ur, xakas, qrimtatar, boshqird kabi tillari doirasida ko'plab tadqiqotlar olib borilgan. Statistik analiz, mashina tarjimasi, avtomatik tahrir, korpus texnologiyasiga doir sohalarda ijobiy natijalarga erishilgan. Jumladan, turk tili bo'yicha kompyuter lingvistikasiga doir izlanishlar K. Oflazer¹¹, E. Adali, Gülsen Eryigit¹², M. Orhun¹³, A.C. Tantuğ¹⁴, Cüneyd Tantuğ kabi

¹¹ Oflazer K. (1994). Two-level description of Turkish morphology. *Literary and Linguistic Computing*, 9(2); Oflazer K. (1999). Dependency parsing with an extended finite state approach. In *Proceedings of ACL'99, the 37th Annual meeting of the Association for Computational Linguistics*; Oflazer K., Kuruoz I. (1994). Tagging and morphological disambiguation of Turkish text. In *Proceedings of the 4th Applied Natural Language Processing Conference*, p. 144-149. ACL; Oflazer K., Tur G. (1996). Combining hand-crafted rules and unsupervised learning in constraint-based morphological disambiguation.; In Brill E., Church K. (eds). *Proceedings of the ACL-SIGDAT Conference on Empirical Methods in Natural Language Processing*. Oflazer K., Tur G. (1997). Morphological disambiguation by voting constraints. In *Proceedings of ACL'97/EACL'97, The 35th Annual Meeting of the Association for Computational Linguistics*.

¹² Gül, sen Eryigit, Joakim Nivre, Kemal Oflazer. 2008. Dependency parsing of Turkish. *Computational Linguistics*, doi: 10.1162/coli.2008.07-017-R1-06-83, 34(3):357-389. (indexed in SCI and SSCI); Umut Sulubacak and Gül, sen Eryigit. 2018. Implementing universal dependency, morphology and multiword expression annotation standards for Turkish language processing. *Turkish Journal of Electrical Engineering & Computer Sciences*, 26(3):1662-1672.

¹³ Orhun M., Tantuğ A.C., Adalı E., Sönmez A.C., (2009a). Computational comparison of the Uyghur and Turkish Grammar. *The 2nd IEEE International Conference on Computer Science and Information Technology*, 3, 338-342, Beijing, China.; Orhun M., Tantuğ A.C., Adalı E., (2009b). Rule Based Tagging of the Uyghur Verbs. *Fourth International Conference on Intelligent Computing and Information Systems*. Faculty of Computer & Information Science, Ain Shams University, 811-816, Cairo, Egypt.

¹⁴ Shylov M., — Dilmaç: Turkish and Turkmen Morphological Analyzer and Machine Translation Program, 1 Master's thesis, Fatih University, İstanbul Turkey, 2008.; 91 Tantuğ, A. C., Adalı, E., and Oflazer, K. 2006. Computer analysis of the Turkmen language morphology. *Advances in natural language processing, proceedings (Lecture notes in artificial intelligence)*, 4139 . pp. 186-193.92 Hamzaoglu, G. 1993. Machine

olimlarning nomlari bilan bog'liq. Ushbu sohada morfologik analiz, mashina tarjimasi, sintaktik tahlil, korpus lingvistikasi, savol-javob tizimi kabi yo'nalishlarda ulkan natijalarga erishilgan.

Tatar tilining kompyuterli qayta ishlash tizimi bo'yicha Dj. Suleymanov¹⁵, A. Gatiatullin, R. Gilmullin, B. Hakimov kabi olimlar tomonidan olib borilgan ilmiy izlanishlari diqqatga sazovor.

2005-2010-yillarda Microsoft uchun Windows 95 OS uchun tatar tilining interfeysi hamda Tatariston Respublikasi Fanlar akademiyasi Amaliy semiotika institutida operatsion tizim uchun tatarcha interfeys yaratilgan¹⁶. Kompyuter analizi bo'yicha keng doiradagi ilmiy tadqiqotlar sifatida ABBY Lingvo lug'atiga tatar tilining kiritilganligi, Windows'95 '98 '2000 'XP, Vista, Win 7, Win 8 *WordCorr* deb nomlangan avtomatik tahrirlovchi dasturining yaratilganligi, 1990-yillarda tatar tili bo'yicha elektron o'qitish dasturlarining amalda qo'llanganligini keltirish mumkin¹⁷.

Qirg'iz tili bo'yicha kompyuter lingvistikasiga doir bir qator tadqiqotlar olib borilgan. 2000-yilgacha qirg'iz tilining til modellarini o'rganish va ularni yaratishda faqat amaliy jihatlarga

e'tibor berilib, dastlab ular kompyuter texnologiyalarsiz amalga oshirilgan¹⁸. 2007-yildan boshlab Qirg'izistonning bir nechta universitetlarida Facebook, Twitter, Google va Wikipedia saytlarini qirg'iz tiliga tarjima qiluvchi dasturiy ta'minot yaratilgan¹⁹. Xususan, E. D. Esanov tomonidan yaratilgan (2002) "Тамга-КИТ" dasturi 20 ta komponentdan iborat bo'lib, qirg'iz tilining ijtimoiy-kommunikativ rivojlantirishga mo'ljallangan. Shuningdek, 2011-yilda "KyrSpell 2.2 - Проверка орфографии кыргызского языка для MS Office 97-2013 uchun qirg'iz tili imlosini tekshiruvchi" til paketi yaratildi²⁰.

O'zbekistonda ham kompyuter lingvistikasiga doir muayyan darajada izlanishlar olib borilgan. Kompyuter lingvistikasi shakllanishida o'zbek tili materiallari bo'yicha lingvostatistik tadqiqotlar olib borgan S.Rizayev va S.Muhamedovlarning ishlarini qayd etish joiz. Mazkur olimlar tomonidan kompyuter lingvistikasining statistik tahlil yo'nalishi rivojlantirilgan. S.Muhamedov va R.R.Piotrovskiylar tomonidan «Инженерная лингвистика и опыт системно – статистического исследования узбекских текстов» nomli kitobida lingvistik modellar, modellashtirish va uning umumiy tamoyillari hamda o'zbekcha matnlarning kvantitativ modellari o'rganilgan.

Kompyuter lingvistikasi fani dastlab matematik lingvistika sifatida O'zbekistonda 2001-yil professor Abdumajid Po'latov tomonidan oliy ta'lim tizimiga fan sifatida kiritilgan bo'lib, ilmiy laboratoriyada izlanishlar olib borilgan. Ushbu laboratoriyada kompyuter lingvistikasi muammolariga qaratilgan tadqiqotlar ustida dasturchi, matematik va lingvistlardan tuzilgan ilmiy guruh faoliyat olib borgan. 2000-yillardan buyon Kompyuter lingvistikasining fan sifatida shakllanishida professor A.Po'latovning o'rni katta. 2005-2010-yillar mobaynida Kompyuter lingvistikasida lingvistik yondashuvlar asosida til bilan bog'liq masalalar bitiruv malakaviy

translation from Turkish to other Turkic languages and an implementation for the Azeri language. MSc Thesis, Bogazici University, Istanbul 93 Oner, M., 2007. (In Turkish) Tatar Turkcesi; Turk Lehceleri Grameri Ed., Ahmet Ercilasun, Akcag Publications, Ankara, Turkey

¹⁵ Сулейманов Д.Ш., Невзорова О.А., Соснин П.И., Беляева Л.Н., Лукашевич Н.В., Татевосов С.Г. Формальные модели и системы в вычислительной лингвистике. Казань, 2016. 187 с.; Тайсина Э. Галиева А. и др. Когнитивно-семиотические аспекты моделирования в гуманитарной сфере. Академия наук Татарстана, 2017. 345 с.; Alizade A., Guliyeva Z. Building an automatic dictionary in the expert system of teaching scientific vocabulary – P.5-17; A. Bakiyeva, T. Batura Application of the rhetorical structure theory in automatic text processing systems – P.18-29; R. Gilmullin, R. Gataullin Morphological disambiguation in tatar language using purepos –P.30-37; Ergesh B. Identifying the tonality of texts in the kazakh language on the basis of the dictionary of emotional lexis 62-67; Prokopiev N., Suleymanov D. Automated analysis of natural language question-answer texts in the e-testing system –P.92-98 va boshq. // Turklang. 2017 Kazan, 2017.

¹⁶ Сулейманов Д.Ш. К вопросу внедрения татарского языка в киберпространство / Turklang. 2013, -С. 20

¹⁷ O'sha joyda. -С. 22 («Татар Телле Заман» tatar-online multimediali elektron darsligi 2000 ta tatarcha so'z, 2500 ta rasm va fotosurat, 11 ta ko'ngilochar o'yin va bilimlarni tekshirishga yo'naltirilgan mashqlardan iborat tatarcha talaffuz qiluvchi dasturni o'z ichiga oladi. Ushbu dasturning interfeysi rus, tatar (kiril va lotin) hamda inglizcha tilga moslashtirilgan ko'p tilli platformaga ega.)

¹⁸ Мусаев С.Ж., Карабаева С.Ж., Иманалиева А.И. Проблемы и перспективы развития компьютерной лингвистики в Кыргызстане/ Turklang. 2013, -С. 30

¹⁹ O'sha joyda. – С. 32

²⁰ O'sha joyda. – С. 33

ishlar va magistrlik dissertatsiyalari darajasida o'rganilgan²¹. Xususan, ko'p tilli mashina tarjimasini yaratishga doir magistrlik ishi²², kompyuter lingvistikasining nazariy metodologiyasiga ilmiy materiallar, o'zbek tilini modellashtirishga oid tadqiqotlar²³, A.Po'latov²⁴, S.Muhamedova²⁵, A.Rahimov²⁶ mualliflar tomonidan o'quv adabiyotlari yaratilgan bo'lsa, kompyuter lingvistikasining turli yo'nalishlarda doktorlik dissertatsiyalari himoya qilingan²⁷. Bundan tashqari kompyuter lingvistikasining turli yo'nalishlarida bir qator soha mutaxassislari²⁸ tomonidan olib borilgan izlanishlarni alohida ta'kidlamochimiz.

2014-yildan beri YURIDA firmasi o'zbek tilining tahrirlovchi dasturi bo'yicha amaliy ishlar olib bormoqda. Ham kirill va lotin alifbosiga asoslangan ushbu dastur so'z shakllarini va imlosini tekshirish imkoniyatiga ega. Zamonaviy tahrirlovchi dasturlar singari sintaktik va semantik jihatdan tekshirish imkoniyatiga erishish uchun soha doirasida bir qator ilmiy-amaliy ishlar tadqiq qilinmoqda.

²¹ Abdurahmonova N. Z. Matnlarni kompyuterda ishlash muaommalari: Bitiruv malakaviy ishi. – Toshkent: O'zMU 2009; Abdurahmonova N.Z. Inglizcha sodda gaplarni o'zbek tiliga tarjima qilish dasturining lingvistik ta'minoti. Magistrlik dissertatsiyasi. – Toshkent: O'zMU, 2011

²² Axmedova X. MVC platformasida ko'p tillik tarjimon muhitini yaratish: Magistrlik dissertatsiyasi. – Toshkent: O'zMU, 2015.

²³ Пулатов А.К., Джураева Н. Разработка формальной модели грамматики узбекского языка // Узбекский математический журнал. – Ташкент, 2002. – №1. – С. 47-54; Пулатов А.К., Алиходжаев Б., Джураева Н. Разработка программы компьютерного анализа и синтеза глаголов узбекского языка // ЎзМУ хабарлари. – Ташкент, 2002. – № 2. – В. 17-19.

²⁴ Пулатов А. Дунёвий ўзбек тили. – Тошкент, 2008; Пулатов А. Компьютер лингвистикаси (дарслик). Тошкент: Академнашр, 2011.

²⁵ Мухамедова С. Компьютер лингвистикаси (методик қўлланма). – Т., 2007; Пулатов А, Мухамедова С. Компьютер лингвистикаси (ўқув қўлланма) – Т., 2014; Мухамедова С. Ўзбек тилидаги ҳаракат феъллари асосида компьютер дастурлари учун лингвистик таъмин яратиш. – Тошкент, 2006. – 143 б.

²⁶ Рахимов А. Компьютер лингвистикаси асослари. – Т.: Академнашр, 2011.

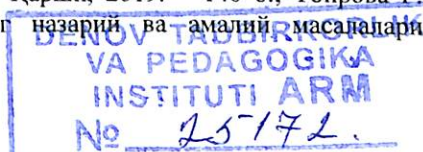
²⁷ Abdurahmonova N.Z. Inglizcha matnlarni o'zbek tiliga tarjima qilish dasturining lingvistik ta'minoti. Filol. fan.dok-ri.diss-yasi. (PhD) – Toshkent, 2018, 165 b.

²⁸ Qarang: Назиров Ш.А., Хомидов Х.Х., Алниязов А.И., Рахманов К.С., Махмудов А.З. Формализация конструкций предложений узбекского, турецкого и каракалпакского языков / Turklang. 2013. – Astana, 2013. – С. 34-47.

Shuningdek, mazkur sohaga aloqador yana bir qator ishlarni alohida qayd etish lozim²⁹. 2017-yil “Markaziy Osiyo universitetlarida sohalararo kompyuter lingvistikasi magistratura dasturini rivojlantirish” (“Development of the interdisciplinary master program on Computational Linguistics at Central Asian universities 585845-EPP-1-2017-1-ES-EPPKA2-CBHE-JP) nomli xalqaro Erasmus+ loyihasi qo'lga kiritilganligi O'zbekistonda ushbu fanga doir tadqiqotlarni umumlashtirish va hamkorlik aloqalarini o'rnatishda turtki bo'ldi. Ayniqsa, dunyo tajribasidan foydalanish borasida Yevropa davlatlaridan Ispaniyaning Santyago de Kompostella va Koruna universitetlari, Portugaliyaning Porto universiteti, Gretsiyaning Vest Attika universiteti, Polshaning Adam Miskiyevich universiteti, Qozog'istonning Al-Farobiy nomidagi Qozoq milliy universiteti, Gumilyov nomidagi Yevroosiyo universiteti, Kostanay universitetlari hamda O'zbekistonning bir qator nufuzli universitetlari, jumladan, O'zbekiston Milliy universiteti, Toshkent davlat o'zbek tili va adabiyoti universisteti, Urganch davlat universiteti, Samarqand chet tillari universitetlarining o'zaro hamkorligi yo'lga qo'yildi. Ushbu loyihada Karlos Gomez, Pablo Gamalo (Ispaniya), Belinda Maya (Portugaliya), Zigmunt Vitulani (Polsha), Altinbek Sharibbay, Diana Rahimova, Alisher Tukayev (Qozog'iston) kabi bugungi kompyuter lingvistikasining yetuk mutaxassislarining borligi loyiha doirasida kompyuter lingvistikasi rivojiga hizmat qildi.

Sohani yanada rivojlantirishga qaratilgan “Turkiy tillarning kompyuterda qayta ishlovi – Turklang” nomli xalqaro konferensiya

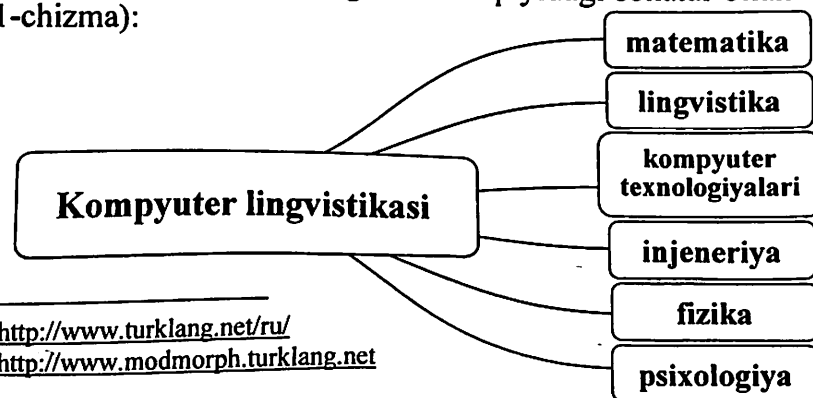
²⁹ Norov A. Kompyuter lingvistikasi asoslari. – Qarshi, 2017; Abduraxmonova N.Z. Inglizcha matnlarni o'zbek tiliga tarjima qilish dasturining lingvistik ta'minoti: filol. fan. bo'yicha falsafa dok. (PhD)...diss. – Toshkent, 2018. Abduraxmonova N.Z. Mashina tarjimasining dasturiy ta'minoti (monografiya). Toshkent: Muharrir, 2018, 175 b; Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: филол. фан. бўйича фалсафа доктори (PhD)...диссер. – Қарши, 2018. – 250 б.; Абжалова М.А. Ўзбек тилидаги матнларни тахрир ва таҳлил қилувчи дастурнинг лингвистик модуллари (Расмий ва илмий услубдаги матнлар тахрири дастури учун): филол.фан.бўйича фалсафа доктори (PhD)...дисс. – Фарғона, 2019. – 164 б.; Эшмўминов А.А. Ўзбек тили миллий корпусининг синоним сўзлар базаси: филол. фан. бўйича фалсафа доктори (PhD)...дисс. – Қарши, 2019. – 140 б.; Тоирова Г. Ўзбек тили миллий корпусини яратишнинг назарий ва амалий масалалари / Globedit, 2020. 168 б.



har yili an'anaviy tarzda o'tkaziladi. Konferensiya dastlab 2013-yili Qozog'istonning Ostona shahrida, so'ng an'anaviy tarzda Istanbul (Turkiya, 2014), Qozon (Tatariston, 2015, 2017), Bishkek (Qirg'iziston, 2016)da bo'lib o'tgan. Kompyuter lingvistikasining mashina tarjimasi, til texnologiyasi, lingvistik ma'lumotlar bazasi, nutq texnologiyasi, elektron lug'atshunoslik (tezaurus, ontologiya) va korpusga doir tadqiqotlar uchun umumiy platforma yaratilgan³⁰. Konferensiyaning O'zbekistonda ilk bor o'tkazilishi (2018-yil, 18-20-oktabr, Toshkent davlat o'zbek tili va adabiyoti universiteti) sohadagi olimlar bilan hamkorlikni yo'lga qo'yish va ilmiy maktablarni yanada kengaytirish uchun muhim qadamlardan biri sanaladi. Negaki soha mutaxassislari Eshref Adali, Murat Orhun (Turkiya), Altinbek Sharibbay, Diana Rahimova (Qozog'iston), Ayrat Gatiatullin, Rinat Gilmullin, Aydar Xusainov (Tatariston), Belinda Maia (Portugaliya), Olimjon Zokirov (Qirg'iziston), Lenara Kubedimova (Qrim), Muftah Hakimov, Mirsaid Aripov (O'zbekiston) kabilarning ishtiroki hamkorlik munosabatlarini o'rnatishda katta turtki bo'ldi. Endilikda Uniturk seminari doirasida kompyuter lingvistikasiga doirasida ilmiy tadqiqotlar hamkorligi yo'lga qo'yilgan bo'lib, turkiy morfem portal³¹ yaratildi va turli yo'nalishlarda xalqaro loyihalar davom ettirilmoqda.

1.3. Kompyuter lingvistikasining boshqa fanlar bilan aloqasi (matematika, psixologiya, informatika, dasturlash asoslari)

Til axborotni saqlash, uzatish va aloqa munosabat shaklidir. Tilning tabiiy imkoniyatlarini hisobga olganda kompyuter lingvistikasi boshqa sohalar bilan ham o'zaro bog'liqlikda rivojlanadi. Kompyuter lingvistikasi quyidagi sohalar bilan aloqador (1-chizma):



³⁰ <http://www.turklang.net/ru/>

³¹ <http://www.modmorph.turklang.net>

Manbalarda qayd etilishicha, kompyuter lingvistikasi uchun kompyuter bilimlari va lingvistikani psixologiya va mantiq fanlari birlashtiradi³². Ayniqsa, kompyuter lingvistikasining boshqa sohalariga ta'siri, va aksincha, boshqa sohalarning ushbu fanga ta'siriga doir ko'plab tadqiqotlari misol sifatida keltirish mumkin³³.

Til kiberfazoning muhim va ajralmas aloqa vositasi sifatida "raqamli insoniyat" (*digital humanity*) uchun muhim ahamiyatga ega.

Lingvistik texnologiya deyilganda, matn generatsiyasi, interpretatsiyasi, matn tahlili, nutqni tanish kabi vazifalarga mo'ljallangan kompyuter tizimlarini rivojlantirish uchun lingvistik bilimlarning dasturiy ilovalari nazarda tutiladi³⁴.

Ma'lumotlarni qidirish, ularni tasniflash, savol-javob tizimi kabi dasturiy ta'minotlar foydalanuvchi va kompyuter o'rtasidagi aloqani bog'lashga xizmat qiladi. Mashina va foydalanuvchilar o'rtasida turli tillar uchun yaratilgan tizimlar sirasiga mashina tarjimasi, kompyuter yordamida tilga o'qitish hamda lingvistik vazifalarni bajaruvchi tizimlar korpus, orfografik tekshiruvchi dasturlar kiradi.

Kompyuter lingvistikasining boshqa sohalar kesimida bog'liqligi kompyuter lingvistikasining nutq sintezatori (fizika, muhandislik), mashina tarjimasi (matematika, dasturlash, tarjima nazariyasi va amaliyoti), sentiment analiz va diolog tizimi (psixologiya), axborot qidiruvi informatika kabi yo'nalishlarida o'z ifodasini topadi.

Y.Karaulov ta'kidlaganidek³⁵, tilga oid bilimlar bazasi ikki asosiy konseptdan iborat bo'ladi: 1) til haqidagi istalgan ma'lumot leksikografik shaklda berilishi; 2) til haqidagi leksikografik

³² Qarang: Eman M. Muslim An Introduction to Computational Linguistics Advantages&Disadvantages <https://www.researchgate.net/publication/318116302>

³³ Emiel Krahmer What computational linguists can learn from psychologists (and Vice versa) <https://www.mitpressjournals.org/doi/pdf/10.1162/coli.2010.36.2.36201>; Patrik Blekburn, Marc Dymetman, Alain Lecomte and others. Logical aspects of Computational linguistics: an introduction, https://www.researchgate.net/publication/2679919_Logical_Aspects_of_Computational_Linguistics_an_introduction

³⁴ Moodle.udc.es/course

³⁵ Караулов Ю. Методология лингвистического исследование и машинный фонд русского языка. // машинный фонд русского языка: идея и суждения. – М.: Наука, 1986. – С. 14

ma'lumotlar mashina formatiga keltirilgan algoritmik tizimlashtirilishi lozim.

Kompyuter lingvistikasi kognitiv fanlar va sun'iy intellekt o'rtasidagi soha hisoblanadi. Fanning nazariy jihatlarini lingvistika va kognitiv sohalar bilan bog'liq. Dunyo tajribasida formal lingvistik nazariyalar uchun kompyuter modellarini rivojlantirishda kognitiv psixologiya yordamida lingvistlar dasturchilar bilan hamkorlik qiladilar. Amaliy komponentlar sun'iy intellektga bog'liq bo'lgan amaliy modellarni rivojlantiradi. Shu jihatdan "Til injeneriyasi" (Language Engineering) yoki "(Inson) Til texnologiyasi" ((Human) Language Technology) kompyuter lingvistikasi terminiga sinonim sifatida qaraladi. Amaliy kompyuter lingvistikasining asosiy vazifalaridan biri inson va mashina o'rtasidagi muloqot jarayonini osonlashtirishga yordam berish va til texnologiyasini yaratish hisoblanadi.

Savol va topshiriqlar

1. Kompyuter lingvistikasi qachondan boshlab fan sifatida o'rganila boshlandi?
2. Kompyuter lingvistikasi fanining paydo bo'lishiga asos bo'lgan omillarni sanang?
3. Fan bo'yicha olib borilgan tadqiqotlar qanday dasturlarning yaratilishiga asos bo'ldi?
4. Dunyo tilshunosligida mashina tarjimai bo'yicha kimlar izlanish olib borgan?
5. Kompyuter lingvistikasi fani O'zbekistonda qay darajada o'rganilgan?
6. Til industriyasi nima?
7. Jamiyatda tilning amaliy ahamiyati hamda til texnologiyasiga doir fikrlaringizni o'zaro muhokama qiling.

Adabiyotlar:

1. Пулатов А.К., Джураева Н. Разработка формальной модели грамматики узбекского языка // Узбекский математический журнал. – Т., 2002. – №1. – С.47-54.
2. Абдурахмонова Н.З. Машина таржимасининг лингвистик асослари. – Тошкент: Академнашр, 2012.

3. Мухамедова С. Компьютер лингвистикаси (методик қўлланма). – Т., 2007.

4. Пулатов А, Мухамедова С. Компьютер лингвистикаси (ўқув қўлланма) – Т., 2014.

5. Рахимов А. Компьютер лингвистикаси асослари. – Т.: Академнашр, 2011.

6. Jurafskiy D., James H. Martin Speech and language processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 2007 – P. 12-13.

1-mavzu yuzasidan test

1. Rus kompyuter tilshunosligida avtomatik tahrirlash sohasi bo'yicha ilmiy tadqiqot olib borgan olimlar qatorini aniqlang.

- a) I.A. Melchuk, G.G. Belonogov, V.E. Berzon
- b) R.R. Kotov, V.E. Berzon, L.L. Nelyubin
- c) I.S. Duganova, V.G. Britvin, N.N. Kirov
- d) L.L. Nelyubin, N.N. Kirov, A.B. Kuznetsov

2. Avtomatik tarjima bo'yicha dastlabki tajriba qayerda o'tkazilgan?

- a) AQSHdagi Jorjtaun universitetida
- b) Rossiya Fanlar Akademiyasida
- c) Germaniya Fanlar Akademiyasida
- d) AQSHning Fan masalalari bo'yicha kongress qo'mitasida

3. Transformatsion grammatikaning asoschisi kim?

- a) Harris
- b) Chomskiy
- c) Klodnyajnyaya
- d) Bar Hillel

4. Kompyuter lingvistikasi fanining asosiy maqsadi nima?

a) lingvistik masalalarni yechishning kompyuter dasturini yaratish

- b) tillarni o'qitish
- c) bilimlarni tekshirish
- d) matnlarni turli jihatdan avtomatik tahrirlash

5. Kompyuter lingvistikasining 1940-1950-yillar orasidagi rivojlanish bosqichida qanday tadqiqotlar olib borilgan?

a) avtomatlashtirish hamda ehtimollik yoki informatsion-nazariy

modellar asosida tadqiqotlar

b) Chomskiy ta'limoti hamda sintaktik analiz (parsing) algoritmi va dinamik dasturlashga oid izlanishlar

c) transformatsion metod asosida tabiiy tilni qayta ishlashning jiddiy tekshiruvchi psixologik metodlarning o'rganilishi

d) a va b

6. Kim grammatikani xarakterlashda avtomatlashgan mashinani muayyan usuli sifatida hisoblab, avtomatlashgan grammatika yordamida tilni umumlashtirish orqali avtomatlashgan tilni tavsiflab berdi?

- a) Chomskiy
- b) Shannon
- c) Mitkov
- d) Klin

7. Qaysi olimning tadqiqotlari chekli avtomat (fine-state automaton) va o'zgarmas ifodalarga (regular expression) oid?

- a) Shannon
- b) Klin
- c) Chomskiy
- d) Mitkov

8. Katta hajmdagi lug'at va kompyuterlashtirilgan matnlarni tanish (text recognition) uchun Bayesov tizimi qaysi olimlarning nomi bilan bog'liq?

- a) Bledsov va Brov
- b) Nivel va Simyon
- c) Vatson va Kolmerauer
- d) Klaud Shennon va Nazaniel Rochesterlar

9. Vinogradov modeli qanday muammoni hal etishga yo'naltirilgan?

- a) semantik va diskurs modelni yaxshi tushuna oladigan tahlil
- b) tabiiy tilni qayta ishlash jarayoni
- c) shovqinli kanal modeli va dekodlash tizimi
- d) ehtimollik paradigmasi

10. Nutqni tanishda ehtimollik modellarini qo'llash kim tomonidan olib borildi?

- a) Tomas Vetson
- b) Vinogradov
- c) Klaud Shennon
- d) Kolmerauer

11. Mantiqqa asoslangan va tabiiy tilni qayta ishlash paradigmalari nima deb nomlanadi?

- a) LUNAR
- b) SHRDLU
- c) CALL
- d) BROWN

12. Kolmerauer va uning hamkasblari tomonidan olib borilgan tadqiqotlari mazmuni nimadan iborat?

- a) Q-tizimli va transformatsion grammatikada mantiqqa asoslangan paradigmalardan foydalanila boshlandi
b) shovqinli kanal modeli va dekodlash tizimi rivojlantiriladi
c) semantika va diskurs modelni yaxshi tushuna oladigan tahlil
d) BROWN korpusi bo'yicha tadqiqotlar

13. Simvolik paradigma necha yo'nalishda tadqiq etilgan?

a) ikki

b) uch

c) to'rt

d) besh

14. 1970-1983-yillarida qanday tadqiqotlar olib borilgan?

a) til va nutqni qayta ishlash bo'yicha

b) transformatsion metod asosida tabiiy tilni qayta ishlashning jiddiy tekshiruvchi psixologik modeli bo'yicha

c) lug'at va kompyuterlashtirilgan matnlar bo'yicha

d) paradigma statistikasi va elektr muhandisligi bo'yicha

2-mavzu. Kompyuter lingvistikasining asosiy yo'nalishlari

2.1. Kompyuter lingvistikasi yo'nalishlarining nazariy mazmuni

2.2. NLP texnologiyasi va lingvistik instrumentlar

Tayanch so'z va iboralar: tabiiy tilni qayta ishlash, savol-javob, mashina tarjimasi, avtomatik annotatsiyalash, referatlash, nutqiy sintezator, tabiiy tilni qayta ishlash, sentiment analiz, tabiiy tilni tushunish, matnlarini mavzuli qisqartirish, matn tahlili, Brat dasturi, Paracon texnologiyasi, tarjima xotirasi, terminologik boshqaruv, BlackLab, SketchEngine, BootCat

Tabiiy tilni qayta ishlash yoki til haqidagi bilimlardan iborat dasturni qayta ishlash o'rtasida quyidagicha tafovut mavjud. *Microsoft office* dasturi lingvistik bilimlardan foydalanadi, ya'ni har bir birlikni so'z sifatida o'qisa-da, uning imkoniyati nisbatan cheklangan. Tabiiy tilni qayta ishlash jarayoni esa *Microsoft office* imkoniyatlari bilan solishtirilganda ancha murakkab jarayon ekanligi amalda isbotini topgan. Ularga savol-javob tizimi, mashina tarjimasi, avtomatik annotatsiyalash, referatlash, nutqiy sintezator kabi tizimlar kiradi.

Tabiiy tilni qayta ishlash bo'yicha ma'lumotlar³⁶:

- Amerika havo yo'llaridagi *Amtrak* deb nomlangan avtomatik sayohat-provayderi sayohat joyi va manzilga yetib kelish vaqtini avtomatik tarzda boshqarib turadi.

- Mercedes-benz zamonaviy mashinalarda nutqni tanish va matnni nutqqa o'tkazuvchi texnologiya mavjud. Bu tizim haydovchilarga muhit, navigatsion tizimlar, ko'ngilochar mavzularda ovoz orqali boshqarish imkoniyatini beradi.

- *Blinkx* va boshqa video qidiruv kompaniyalari nutqni tanish texnologiyalari orqali million soatli veb-videolarni yozuvga aylantira oladi.

³⁶ Jurafskiy D., James H. Martin Speech and language processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 2007 – P. 8.

- Google tarjimon ovozli so'rov orqali boshqa tildagi so'ralgan ma'lumotni tarjima qilib (ovozli holatdan), so'ngra qidiruv tizimiga uzatadi.

- Pearsonga o'xshash ma'rifiy nashriyotlarda ETSga o'xshash tekshiruv tizimi mavjud bo'lib, avtomatlashgan tizim orqali minglab talabalarning insholarini tahlil qila oladi va ularni kompetensiyaviy darajalay oladi.

- animatsion xarakterdagi interaktiv tyutorlar afaziya va Parkinson kasalligiga uchragan bolalarni terapevt sifatida o'qita oladi.

- matn tahlili bilan shug'ullanuvchi kompaniyalar, ya'ni Nilson Buzmetriks, Umria kabi tizimlar veb-sahifalardagi ijtimoiy fikrlar, muhokamalar va munosabatlarni tahlil qiladi.

Endilikda kompyuter lingvistikasining yo'nalishlari kengayib bormoqda. Ularga quyidagilar kiradi:

- mashina tarjimasi;
- matn generatsiyasi;
- savol-javob tizimi;
- nutqni ovozashtirish;
- matn korpusi;
- ma'lumot qidiruvi;
- kompyuter leksikografiyasi;
- sentiment analiz;

- matn tayyorlash kabi;

- a) tabiiy til matnlardagi so'zlarni avtomatik ajratish;
- b) nutqiy vaziyatni tekshirish;
- s) grammatikani tekshirish;
- d) uslubni tekshirish.

- ma'lumotlar bazasi va boshqa tizimlar uchun tabiiy til interfeyslarini yaratish;

- matnlarni yangilab, boyitib borish;

- tabiiy tilni tushunish;

- matnlarini mavzuli qisqartirish va h.k.

Makkeri (Macquarie) universiteti professori Mark Jonson kompyuter texnologiyalari yordamida til bilan shug'ullanish sohani ikkiga ajratadi:

I. Kompyuter lingvistikasi (CL):

- 1) tilni tushunish (language comprehension);
- 2) tilga oid vazifalarni tayyorlash;
- 3) tildan foydalanish;

II. Tabiiy tilni qayta ishlash (NLP):

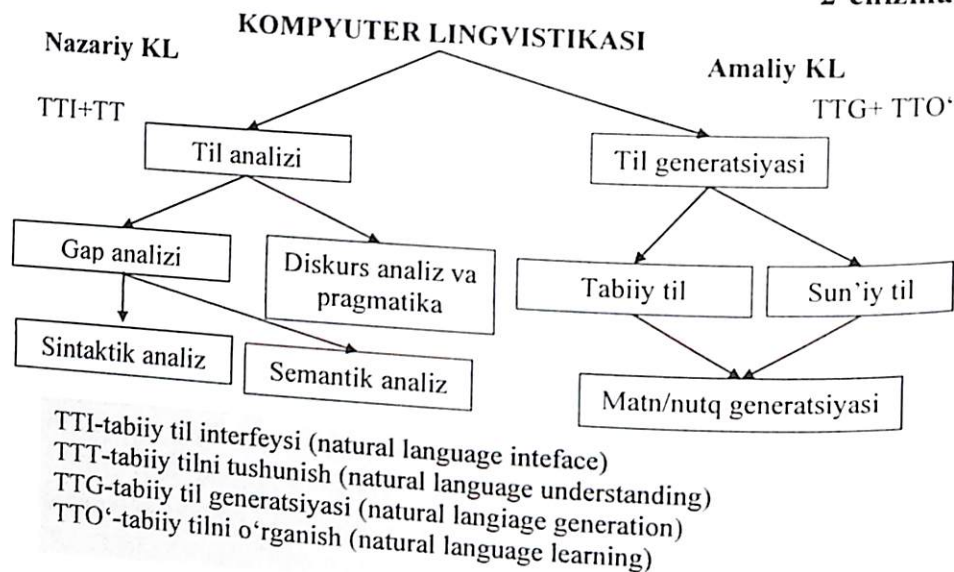
- 1) ma'lumot qidirish;
- 2) mavzuni aniqlash va hujjatlarni tasniflash;
- 3) hujjatlarni qisqartirish;
- 4) sentiment analiz (idrokiy analiz – qidiruv so'rovining bog'liqligiga ko'ra matnlarni tsniflash);
- 5) mashina tarjimasi;
- 6) nutqni tanish;
- 7) matnning avtomatik tahlili³⁷:
 - a) ma'lumotlarni ajratib olish (information extraction);
 - b) matnlarni tokenlash, stemplarga ajratish va matnlarni teglash (kodlash);
 - c) hujjatlar jamlanmasini guruhlash (klasterlash);
 - d) guruhlangan to'plamlarni (klasterlarni) tasniflash.

Yana bir Sakthi Vel kompyuter lingvistikasining yo'nalishlarini tasniflashda quyidagi chizma orqali izoh beradi³⁸(2-chizma). Lekin ushbu chizmada negadir morfologik analiz e'tibordan chetda qolgan. Morfologik analiz sintaktik va semantik analiz uchun asos vazifasini o'taydi. Diqqatimizni tortgan narsa tadqiqotchi tomonidan TTI-tabiiy til interfeysi (natural language inteface), TTT-tabiiy tilni tushunish (natural language understanding), TTG-tabiiy til generatsiyasi (natural langiage generation), TTO'-tabiiy tilni o'rganish (natural language learning) o'zaro bog'lanishini muayyan darajada tushuntirishga harakat qilingan.

³⁷ text mining, глубоинный анализ текста, раскопка текста глубоинный анализ текста, раскопка текста

³⁸ Sakthi Vel S. Applications of Computational Linguistics to Language Studies: An Overview // International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 4, Issue 3, March 2017

2-chizma



2.2. NLP Texnologiyasi va lingvistik instrumentlar

Lingvistika tor ma'noda tilga oid bilimlar sohasi bo'lib, unga boshqa intellektual sohalar kesimida turlicha yondashish mumkin. Lingvistik uskunalar (instrumentlar) – lingvistlar uchun til bo'yicha tadqiqot qilishda o'rganish vositasi hisoblanadi. Bunday lingvistik instrumentlar til va nutq hodisalari uchun mo'ljallangan bo'ladi.

Lingvistik resurslarni (ma'lumotlarni) analiz qilish va ularni boshqarish uchun bir nechta bosqichli tizim yaratiladi. Resurslar uch turda bo'lishi mumkin³⁹:

1. Qo'shma resurslar: bu kabi resurslar tarkibiga kirish metodi, chiqish metodi, standart ifodalar kiradi.

2. Lingvistik resurslar: tilga oid resurslar faqat ma'lumotlar bazasidan iborat bo'lib, leksikon, korpuslar, tezaurus, lug'at, lingvistik analizator kiradi. Lingvistik resurslar onlayn yoki avtonom shaklda bo'lishi mumkin.

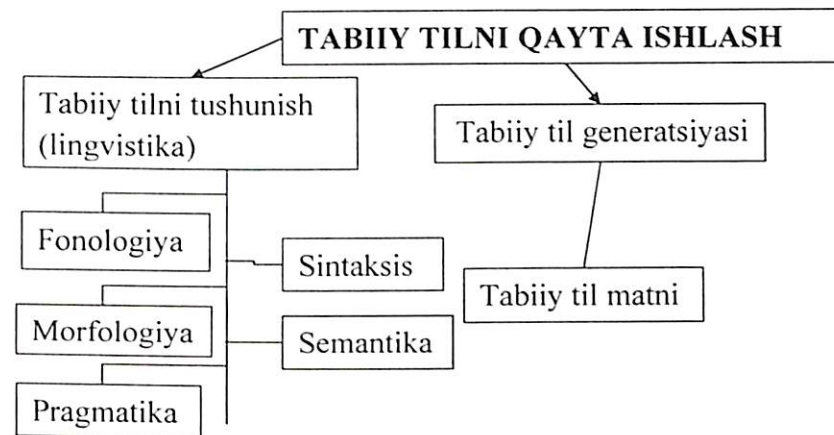
3. Qayta ishlangan resurslar: bunday resurslarga qidiruv tizimi, nutqni o'quvchi tizim, tarjimon dasturlar, grammatikani tekshiruvchi vositalar, so'zlarni oldindan topish dasturi, imloni tekshiruvchi

³⁹ http://tdil-de.in/index.php?option=com_vertical&parentid=8&Itemid=553&lang=en

vositalar, sintaktik analizator kiradi.

Kompyuter lingvistikasi va NLP, ya'ni tabiiy tilni qayta ishlash texnologiyasi haqida olimlarning turli qarashlari mavjud: 3-chizma⁴⁰

3-chizma



Tayyor lingvistik instrumentlardan foydalanishda quyidagi maqsadlar ko'zda tutiladi:

- lingvistik tahlil uchun;
- dasturlash tillari uchun;
- korpus lingvistikasi tahlili uchun;
- avtomatik tarjima jarayoni uchun.

Uskunalar tasnifiga e'tibor qaratsak, ular quyidagicha bo'lishi mumkin:

-matnni tahrirlovchi uskunalar: Notepad++, Sublime text, Language quality insurance, Checkmate, Xbench;

matnlarni qiyoslovchi uskunalar: Beyond Compar; CAT uskunolari MetaCat, Omega, Memoq;

fonetik tahli va annotatsiya bilan shug'ullanuvchi uskunalar: Elan, Praat;

matnlarni modellashtiruvchi uskuna: MALLET (machine learnig for language toolkit);

⁴⁰ Diksha Khurana, Aditya Koli Natural Language Processing: State of The Art, Current Trends and Challenges / <https://arxiv.org/ftp/arxiv/papers/1708/1708.05148.pdf>

matn analizi va qidiruv tizimga mo'ljallangan uskunalar: AntConc, CasualConc, Crossref Text va tadqiqotchilar uchun matn qidiruvchi uskuna, Juxta, WoarHoard, WordSeer;

veb-sahifalar bilan ishlovchi uskunalar: import.io, Tapor va boshq. Shulardan ayrimlari haqida to'xtalib o'tamiz.

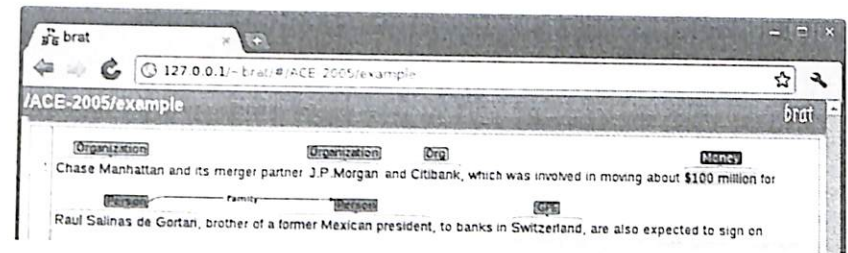
CLARIN loyihasi 2012-yilda tashkil topgan bo'lib, u umumiy til resurslari va texnologiyasi infrastrukturasi yaratishga mo'ljallangan. U turli tilga oid resurslarni qayta ishlash va kompyuter lingvistikasi uchun foydalanishga mo'ljallangan ochiq platforma hisoblanadi. U tilga oid barcha resurslarni raqamli holga o'tkazib, onlayn muhitida foydalanish imkonini beradi. Ushbu resursda yozma, og'zaki va ko'p modelli formada raqamli tilga oid ma'lumotlar kiritilgan. CLARIN ko'plab turli lingvistik instrumentlarni qo'llab-quvvatlaydi.

WebLicht webga asoslangan lingvistik annotatsiya instrumentlarining majmuasi bo'lib, ushbu web ilova lokal kompyuterga biror dasturiy ta'minotni yuklamay yoki o'rnatmay turib ham internet yordamida foydalanish imkonini beradi⁴¹. Shuningdek, Gabmap web ilovasi so'zlarni turli lingvistik aspektlarda analiz qilishga mo'ljallangan⁴² bepul foydalanish imkoniyatini beradi.

Bugungi kunda ko'plab lingvistik uskunalar tabiiy tilni qayta ishlashda samarali foydalanib kelinmoqda. Masalan, korpus lingvistikasi, mashina tarjimasi, matnni annotatsiyalash kabi maqsadlarda tayyor bepul va ochiq platformali lingvistik resurslar har jihatdan til texnologiyasini yaratishda asosiy omillardan biri hisoblanadi. Matn annotatsiyasi uchun vebga asoslangan *brat* uskunasi mavjud hujjatlarga avtomatik tarzda qo'shimchalar qo'shish va strukturaviy annotatsiyalash vazifasini bajaradi. Quyidagi rasmda rasmda annotatsiyaning ikki asosiy kategoriyasi ifodalangan (1-rasm):

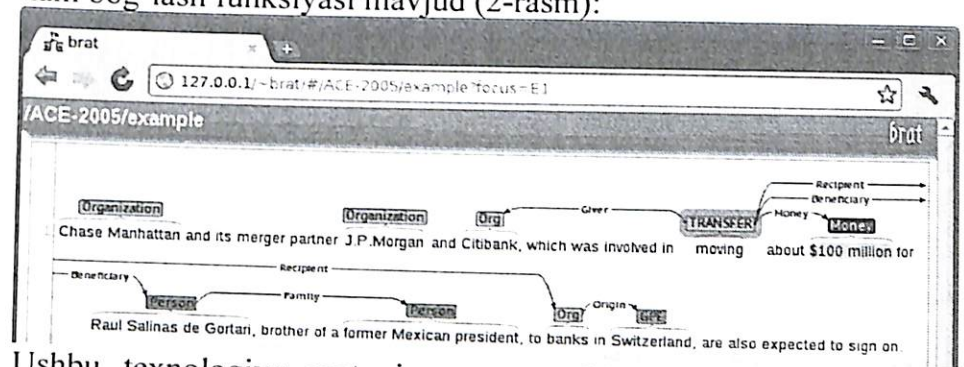
⁴¹ https://weblight.sfs.uni-tuebingen.de/weblight/demo/?samplename=fr_ExampleFood

⁴² <https://www.sciencedirect.com/journal/lingua/vol/178>

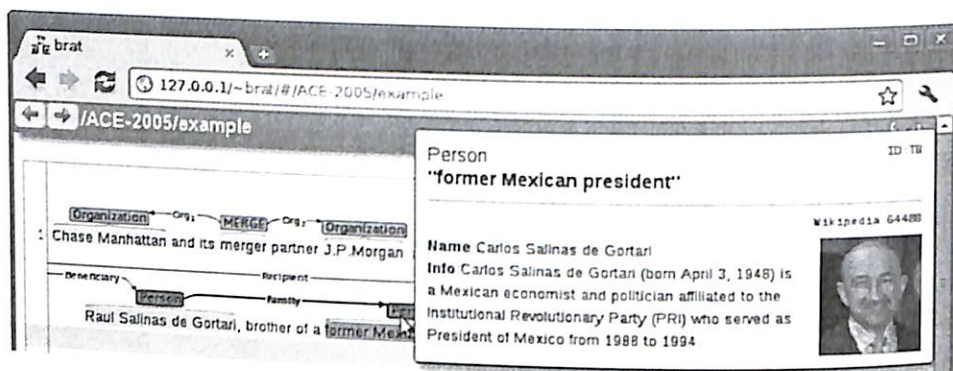


Ushbu jihatdan quyidagi usullar bajariladi:

- matndagi shaxs va tashkilot nomlarini belgilash
 - bir-biriga aloqador so'zlarni ajratish
 - matndagi so'zlarning kategoriyalari bo'yicha annotatsiya uchun kerakli birliklarini aniqlash
 - Ma'lumotlarni kengaytirishda binar munosabatga tayanish.
- Brat* uskunasi orqali boshqa annotatsiyalarning *n*-assotsiatsiyani ham bog'lash funksiyasi mavjud (2-rasm):



Ushbu texnologiya matnni muayyan fakt va atributlarga ko'ra annotatsiyalashda foydalaniladi. Shuningdek, *brat* texnologiyasi vikipediya dagi resurslarning annotatsiyasini normallashtirish hamda inson tomonidan annotatsiyaga yordamchi sifatida tabiiy tilni qayta ishlash kabi funksional imkoniyatlarga ham ega (3-rasm).



TARJIMON INSTRUMENTLARI

Tarjimonlik sohasida ham bir qancha lingvistik instrumentlar mavjud. Ular parallel tarzda tarjima muqobillarini serverga to'plash hamda bilingval terminologik ma'lumotlar bazasini yig'ishga mo'ljallangan. Parallel korpusga asoslangan bilingval terminlarni boyitib borish o'tgan asrning 80-yillariga borib taqaladi. Huddi shu davrda yaponiyalik olim Makato Nagao namunaga asoslangan (example based translation) mashina tarjimasi texnologiyasini ilgari surgan. Endilikda ushbu turdagi texnologiyalar tarjima xotirasi (*TM-translation memory*), deb nomlanmoqda.

Korpusga asoslangan terminologik ma'lumotlar bazasini yaratish internet tizimidagi ma'lumotlarni parallel ravishda berilishi hamda mehnat resursining tejamlorligi jihatdan yuqori bo'lganligi bois g'arbda ushbu texnologiya an'anaga aylanib, ko'p tadqiqotlar olib borilgan. Ushbu texnologiya mashina tarjimasi uchun samaradorligi va uni izchil ravishda boyitib borish imkoniyati bilan ham ahamiyatlidir. Mana shunday instrumentlardan biri *Paraconc* texnologiyasidir. U bilingval va ko'p tilli dastur bo'lib, quyidagi imkoniyatlarga ega⁴³:

- turli usullarda so'zlarni saralash va hisoblash;
- tarjima qilingan parallel matnlarni moslashtirish;

⁴³Ketiwe Ndhlovu Using ParaConc to extract bilingual terminology from parallel corpora: A case of English and Ndebele Literator / Journal of Literary Criticism, Comparative Linguistics and Literary Studies ISSN: (Online) 2219-8237, (Print) 0258-2279 <http://www.literator.org.za>

- konteksdagi mavjud birliklarni oson topish va ularni ko'rsatish;
- terminlarni qidirish;
- tarjima ekvivalentlarini aniqlash;
- muhim bo'lgan tarjimalardagi o'rinlarni aniqlash;
- so'z birikmasi sifatida so'zlar o'ramini tahlil qilish;
- so'zlar chastotasi va boshq.

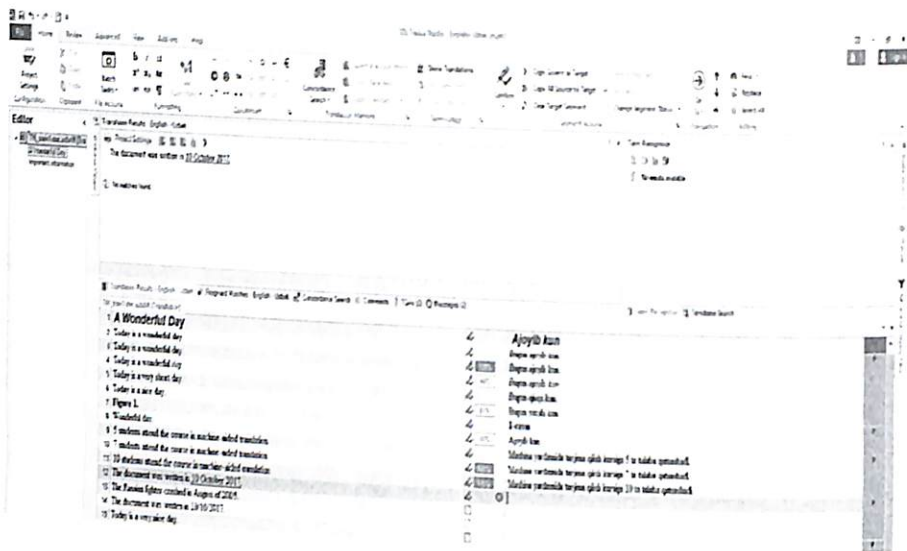
Kompyuter yordamidagi mashina tarjimasi texnologiyasi (CAT –computer assisted translation) ikki qismga ajralgan:

- *Tarjima xotirasi* (Translation Memory)
- *Terminologik boshqaruvi* (Terminology Management)

Tarjima instrumentlarga tarjima xotirasi texnologiyasiga (translation memory) asoslangan Trados dasturi hamda WinAlign tarjima texnologiyasi kiradi (CAT –computer assisted translation). Ushbu dasturlar *Translator's workbench* va *MultiTerm* bazasi asosida yaratilgan. Bunda jumlar sifatida nuqta yoki gap uchun ishlatiladigan boshqa ishoraviy belgilargacha bo'lgan qator gap sifatida segmentlarga ajratiladi. Professional tarjimonlar orqali qilingan matnlar parallel ravishda kompyuter xotirasiga yuklanadi. Keyingi safar avvalgi qilingan tarjimaga o'xshash matn yoki ibora uchrasa, kompyuter foiz ko'rsatkichida segment o'xshashlikni ilova qiladi. Bu esa tarjimonga muayyan yo'nalishda tarjima qilinayotgan asarni qayta tarjima qilmay, bazada mavjud ekvivalentlikni topish ixtiyoriga ega bo'ladi. Bunda ma'lumotlarni foydalanuvchilar tomonidan boyitib borilgan tarjimalar hisobiga kompyuter xotirasidagi asos va tarjima til uchun yaratilgan tayyor tarjima qoliplari hosil bo'ladi. Manbalarda qayd etishicha, ushbu tarjima uskunlaridan foydalanib, inglizcha ilmiy matnlarni nemis tiliga tarjima qilinsa 28% tezroq mehnat samaradorligiga erishish mumkin⁴⁴.

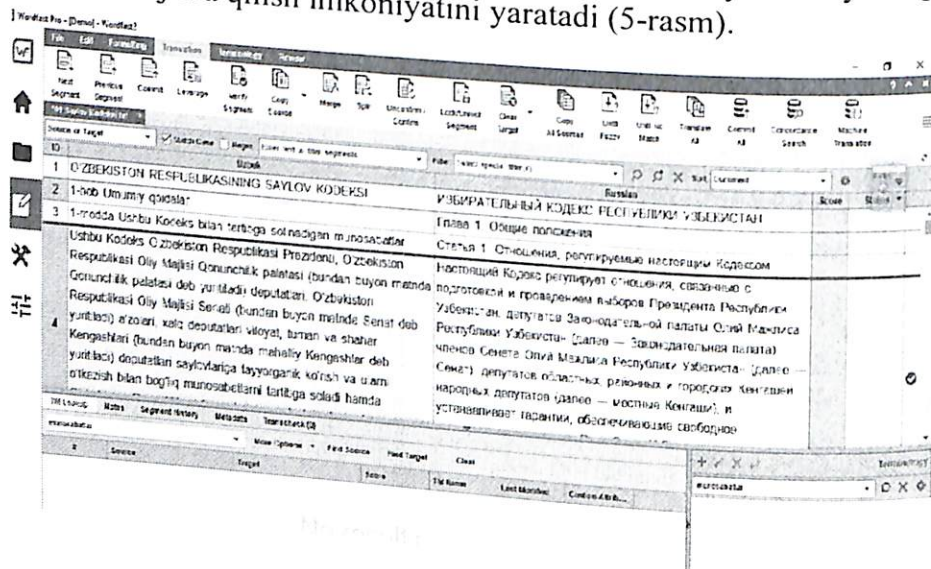
Quyida Trados dasturi orqali inglizcha matnni o'zbek tiliga tarjima qilish namunasi ko'rsatilgan (4-rasm). Tarjima xotiraning yana bir afzallik jihati mavjudki, bir vaqtning o'zida tarjima qilingan so'zlar bilinval lug'at sifatida xotirada saqlanadi. So'ngra undan

⁴⁴ Marcos Zampieri, Mihaela Vela Quantifying the Influence of MT Output in the Translators' Performance: A Case Study in Technical Translation /Workshop on Humans and Computer-assisted Translation, pages 93–98, Gothenburg, Sweden, 26 April 2014. c 2014 Association for Computational Linguistics



4-rasm

kontekstga asoslangan tarjima lug'at sifatida foydalanish mumkin. Tradosning yana bir ahamiyatli jihati tarjima qilinayotgan matn .doc hamda HTML, SGML, XML, ASP, JSP, Tagged RTF, TRADOSTag kabi formatlarni ham o'qishga mo'ljallangan *TagEditor* dasturining mavjudligi bilan ahamiyatlidir. Bu esa asarni ayni asliyatdagi shaklida tarjima qilish imkoniyatini yaratadi (5-rasm).



5-rasm

Bunday tarjima instrumentlari ma'lumotlar bilan ishlashda sohaviy kengayib borayotgan terminologik muammolarni ham barataraf etadi. Chunki tarjimon bir vaqtning o'zida bir nechta mutaxassis bo'la olmaydi va buning imkoni ham yo'q. Zero sohalarning yangi tarmoqlari va integratsiyasi natijasida sohalarning ixtisoslashuvi jamiyatning barcha yo'nalishlar uchun tarjimondan sifat va tezkorlikni talab qiladi. Bunda kompyuter texnologiyalardan foydalanish samarali usullardan biridir. Shu o'rinda o'zbek tili uchun ham mashina tarjimasining lingvistik ta'minotini yaratish muhim ishlardan sanaladi. Ochiq platformali parallel matnlar muhitini yaratish uchun CAT texnologiyasi yordamida dastlabki bosqich amalga oshiriladi. Mashina tarjimasining CAT (computer assisted translation – kompyuter yordamidagi tarjima) texnologiyasi parallel matnlar asosida lingvistik resurs yaratishning eng maqbul usuli sifatida dunyo olimlari tomonidan tan olinib kelinmoqda. Tarjima xotirasi muhiti yordamida ikki tildagi gaplarning o'zaro muqobilik (alignment) holati aniqlanadi. Tarjima uskunasi SDL sifatida foydalaniladi. Uskunalar yordamida tarjimon "tarjima xotirasi" bankini yaratadi natijada, professional tarjimon muayyan sohada tarjimini amalga oshirishda tarjima qilingan matnlardan qayta foydalanish, avval tarjima qilingan segment birliklarni mos tartibga ajratish avtomatik tarzda amalga oshirilishi, o'xshash tarjimalar yoki terminlarni ilova qilish imkoniyatiga ega bo'ladi. Tarjima uskunolari onlayn va kompyuter platformasiga moslashgan bo'lib, ularning xizmatlari ayrimlari pulli va ayrimlari bepul. Kompyuter platformasi uchun pulli foydalanishga mo'ljallangan bepul xizmatlardan Omega T, Café TranEspresso, pulli onlayn xizmatlardan Memsources, Wordbee, XTM Cloud, tekin foydalanish imkoniyatiga ega Smartcat, Matecat va Wordfast Enywhere kabilarni keltirish mumkin.

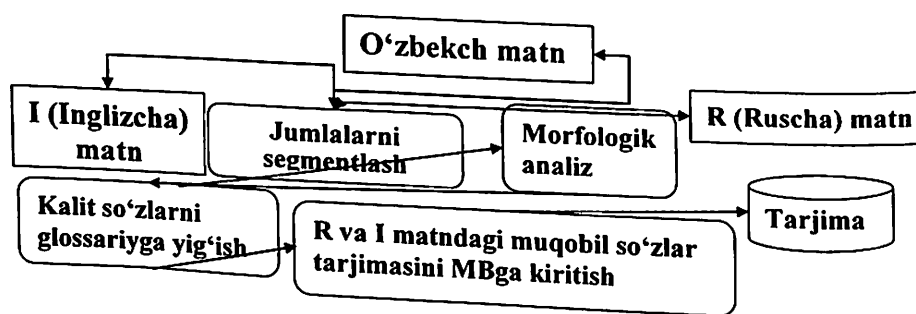
Wordfast Prodan foydalanib, professional tarjimonlar tomonidan qilingan matnlar elektron shakldagi .txt formatda yuzdan ortiq o'zbek tilidagi ilmiy (muayyan sohalarga doir) matnlarning ruscha va ingliz tiliga professional tarjimonlar tomonidan tarjima qilingan juft matnlari jamlandi (5-rasm). Yuqoridagi ilmiy uslubdagi matnlar tibbiyot, matematika, axborot xavfsizli, tarjima, siyosat, filologiya

kabi turli sohani o'z ichiga qamrab olgan. Ma'lumot o'rnida shuni aytish mumkinki, internetning imkoniyatlaridan unumli foydalanib, Webcrawler yordamida ko'p tilli saytlar yoki vikipediyalarni ham tarjima xotirasiga yuklash mumkin. Manbalarda qayd etilishicha, yuklab olingan bilingval saytdagi muqobil parallel matnlar avtomatik generator Betextor yordamida TMX faylga to'planadi [Parallel Corpora Preparation for Machine Translation of Low-Resource Languages: Turkish to English Cardiology Corpora Gökhan Doğru, Adrià Martín-Mor, Anna Aguilar-Amat / http://rec-conf.org/workshops/lrec2018/W3/pdf/5_W3.pdf].

Biroq bu generator faqat juda ayrim tillar uchungina amal qilib, turkiy tillarning boshqa til juftliklari uchun xos emas.

Tarjimaning nechog'lik sifatli tarjima qilinishi so'z, so'z birikmasi yoki iboralarning uslub jihatdan to'g'ri tanlanganligiga ham bog'liq. Shu bois matn qayta tahlil qilinib, segmentlash yoki tarjima jarayonida yo'l qo'yilgan xatolar tarjimon va tilshunoslar tomonidan tekshiriladi. Segmentlash jarayoni to'liq yakuniga yetkanidan so'ng barcha parallel matnlar tarjima xotirasiga yuboriladi (4-chizma).

Tarjima xotirasi muhitidan foydalanish inson tomonidan bajariladigan tarjima jarayonini tezlashtiribgina qolmay, uning sifatiga ham ijobiy ta'sir ko'rsatadi. Chunki matnda yo'l qo'yilgan xatolarni ko'rsatish va to'g'irlash, qolaversa, turli uslubga tegishli matnlardan ma'nodosh so'zlarning muqobilini tanlashda ushbu dasturning imkoniyatlaridan foydalanish o'rinli.



4-chizma

Kompyuter yordamidagi tarjima (CAT) muhitida tarjima xotirasiga yuklangan parallel korpuslar tarjimon va lingvistlar

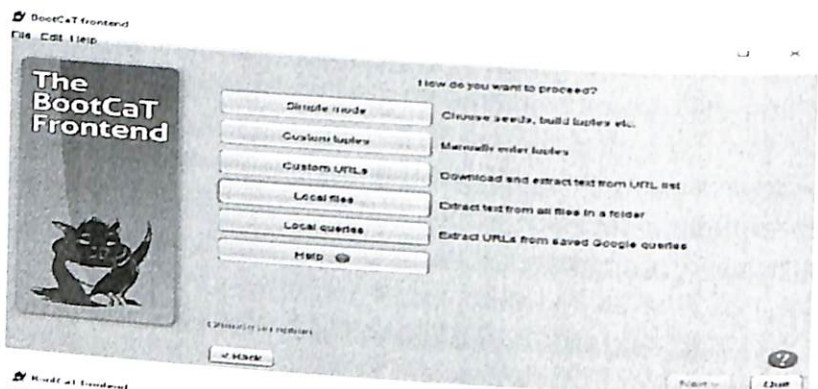
tomonidan qayta tekshiriladi. 4-chizmadagi algoritmning morfologik analiz muhim bosqich sanaladi. Chunki o'zbek tilining agglyutinativ xususiyati matnda token sifatida qatnashuvchi so'z shakllarini morfologik jihatdan analiz qilishda namoyon bo'ladi. Shu bois o'zbek tilidagi token va lemmalar ajratilgandan so'ng ular mutaxassislar tomonidan tarjima juftliklari bilan bazaga kiritiladi.

KORPUSGA MO'LJALLANGAN INSTRUMENTLAR

XXI asr globallashuv va tezkor axborot almashish davridir. O'tgan asrning 80-90-yillarida Kompyuter lingvistikasi va tabiiy tilni qayta ishlash bo'yicha ilmiy yutuqlarga erishildi. Korpus lingvistikasida matnlarni to'plash va ularning ustida muayyan amallarni bajarish uchun tayyor dasturiy uskunalar (tools) yaratilgan. Shular jumlasiga veb sahifalar korpusiga mo'ljallangan **BlackLab**, **SketchEngine**, **BootCat**, **Webcrawler** va tayyor matnlar bilan ishlashga mo'ljallangan **AntConc** kabi dasturiy ilovalarni keltirish mumkin.

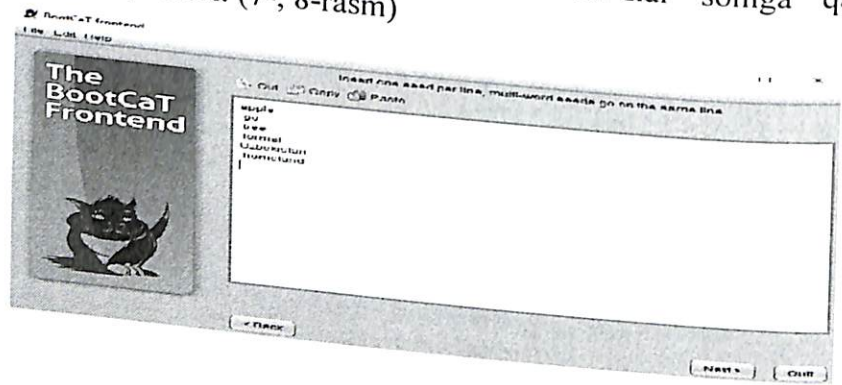
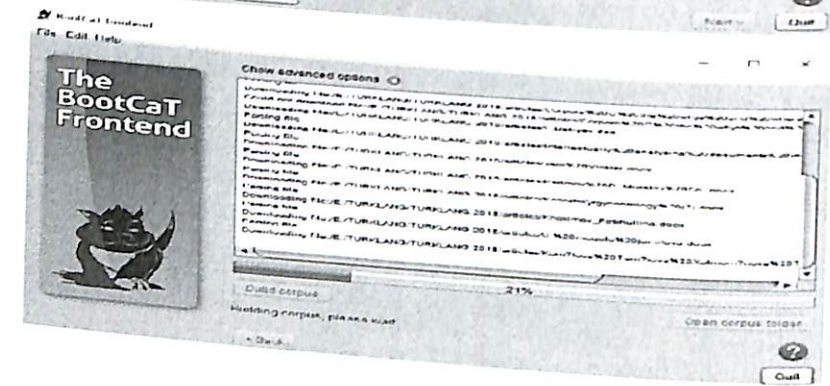
BootCat instrumenti matnlarni elektron shaklda yig'ishda qulay uskuna hisoblanadi. Chunki unda foydalanuvchining imkoniyatlari hisobga olinadi. Masalan, matnlarni global internet tarmog'idan yoki kompyuter bazasiga saqlangan fayllarni baza sifatida yuklab olish imkoniyatiga ega (6-rasm).

Internet tizimi endilikda ma'lumot olishning asosiy manbasi hisoblanadi. Internet korpus lingvistikasi va tarjimonlik sohasida terminologik ma'lumotlar bazasini kompilyatsiya qilish yoki dasturiy ta'minot uchun sohaviy resurs yaratishda muhim omil sanaladi. Google qidiruv tizimi orqali BootCat instrumenti internetdagi ma'lumotlarni qisqa muddat ichida to'plash imkoniyatiga ega bo'lib, uning yordamida muayyan termin orqali sohaviy terminologik bazani yaratish mumkin.

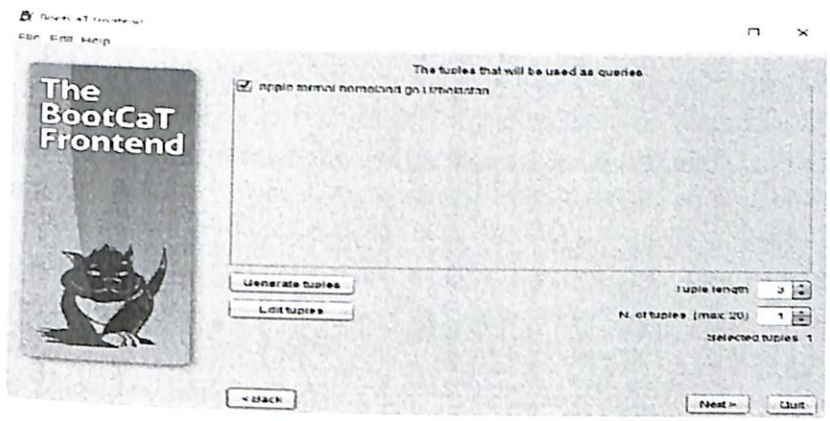


6-rasm

Ushbu uskuna yordamida veb-sahifalari orqali matn korpusi yaratilsa, dastlab har bir qatordan kalit so'z sifatida beshtadan kam bo'lmagan so'z kiritiladi, keyin ushbu so'zlar soniga qarab generatsiya qilinadi. (7-, 8-rasm)

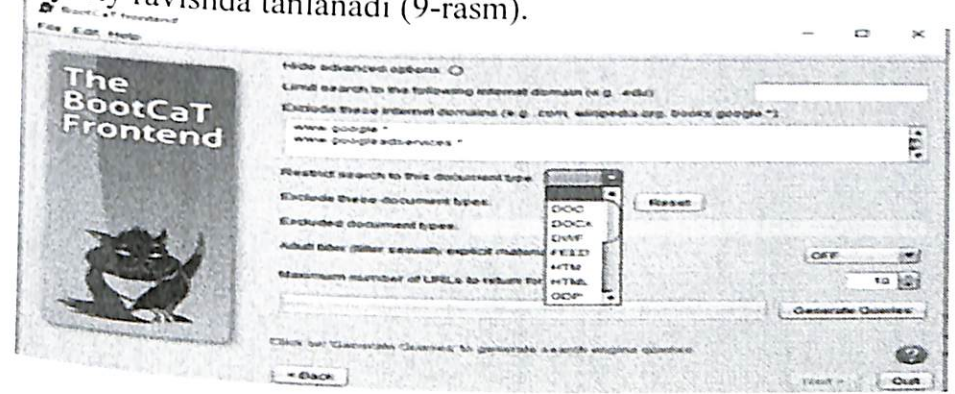


7-rasm



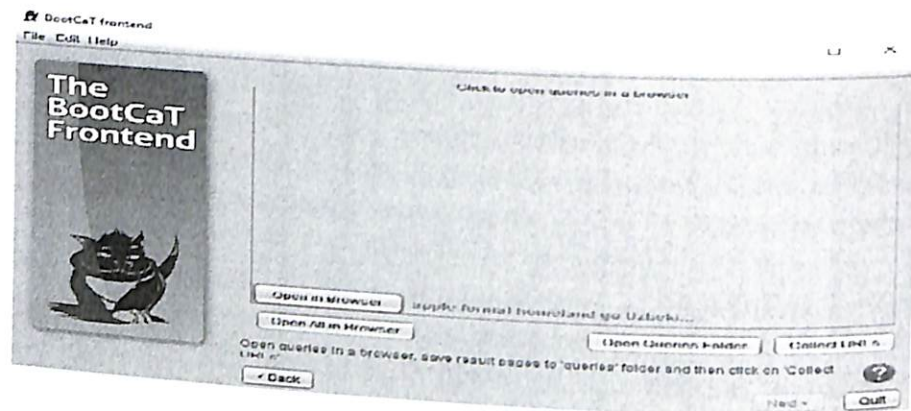
8-rasm

Global tarmoqdan ma'lumotlar qidirilganda, ularning qaysi formatda bo'lishi yoki bo'lmisligi foydalanuvchi tomonidan ixtiyoriy ravishda tanlanadi (9-rasm).



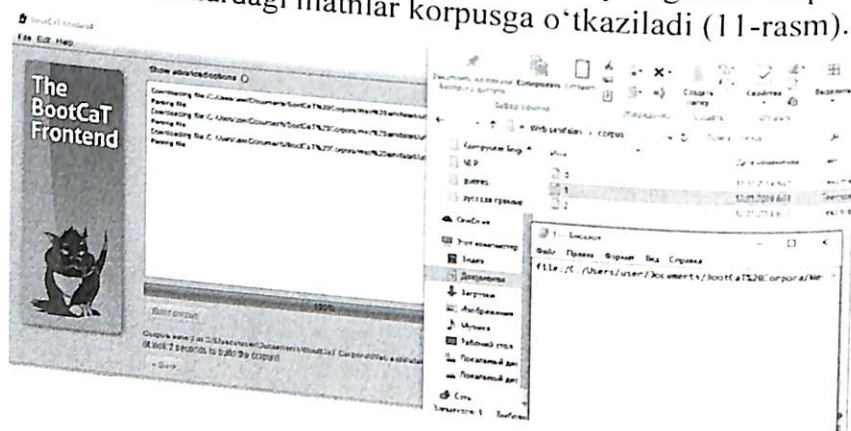
9-rasm

Bundan-da kengroq bazadan so'rov yaratishda kalit so'zlarga tegishli ma'lumotlar sayt sifatida foydalanuvchiga havola etiladi (10-rasm).



10-rasm

Bunda brauzerdan olingan .doc formatidagi ma'lumotlar oldin saqlab qo'yilgan hujjatga so'rovnomaga degan manzilda saqlanishi kerak. *Collect URLS* orqali ma'lumotlar to'planganda saqlangan barcha veb-sahifalardagi matnlar korpusga o'tkaziladi (11-rasm).



11-rasm

Shuningdek, ingliz tili uchun Britaniya milliy korpusi doirasida quyidagi lingvistik instrumentlar mavjud:

- word sketch – so'zlarning grammatik munosabatlariga ko'ra ingliz tilidagi so'z birikmalarini tasniflash;
- thesaurus – har bir so'z uchun sinonim va o'xshatish so'zlarni qidirish;
- word lists – ingliz tilidagi ot, fe'l, sifat so'z turkumiga tegishli so'zlarni chastotasiga ko'ra tartiblash;
- n-grams – ko'p tarkibli so'zlarning chastotasini aniqlash;

- concordance – kontekstdan namunalarni topish;
- trends – iste'moldagi neologizmlar va ularning o'zgarishi avtomatik ravishda aniqlash kabilar.

Britaniya milliy korpusidagi⁴⁵ so'z turkumlarini teglash algoritmi Sketch Engine texnologiyasi orqali quyidagi funksional imkoniyatlarga ega (part-of-speech tagsets):

1. Penn TreeBank teg jamlanmasi
2. CLAWS teggiri

Savol va topshiriqlar:

1. Tabiiy tilni qayta ishlash bo'yicha qanday ishlar olib borilmoqda?
2. Kompyuter lingvistikasi yo'nalishlari qaysilar?
3. NLP ning qaysi yo'nalishi tabiiy til matni bilan ishlaydi?
4. Paraconc texnologiyasining qanday imkoniyatlari mavjud?
5. Tarjima xotirasi texnologiyasi qaysi dastur uchun asos sanaladi?
6. AntConc qanday dastur va uning korpus yaratishdagi o'rni qanday?
7. Korpus yaratish texnologiyasining ketma-ketligini sanang.
8. Tarjima xotirasi (TM) texnologiyasining ishlash prinsipini tushuntiring.
9. Mavzuda qayd qilingan saytlar bilan tanishib, lingvistik va tarjima instrumentlarining imkoniyatlari haqida BBB (bilmas edim, bilib oldim, bilishni xohlayman) texnologiyasi asosida daftaringizga qayd qiling.
10. Mavzuda ko'rsatilgan instrumentlardan foydalanib, tarjima uchun parallel matn segmentlariga ajrating. Uning asosida tarjima lug'atlarini yarating.
11. <http://kazcorpus.kz/klcweb/en/> ushbu sayt orqali qozoq tilida amalga oshirilgan tadqiqotlar bilan tanishing.
12. O'zbek tilining korpusini yaratish bo'yicha lingvistik muammolar haqida muhokama qiling.
13. Sketch engine texnologiyasi bo'yicha quyidagi tartibda amallarni bajaring.

⁴⁵ <https://ru.wikipedia.org/wiki/>

2-mavzu yuzasidan test

1. Kompyuter lingvistikasi termini fanga kim tomonidan kiritilgan?
 - a) David Hays
 - b) Allan Krucher
 - c) Ruslan Mitkov
 - d) Daniel Jurafskiy
2. Algoritmik hisoblashning Tyuring modelini yaratilishiga asos bo'lgan voqea qaysi?
 - a) avtomatlashtirishning amalda tatbiq etilishi
 - b) mantiqiy modellashtirish
 - c) sistemaviy yondashuv
 - d) matematik hisoblash
3. Tillarni avtomatlashtirishda diskret Markov jarayonining amaliy ehtimollar modelidan foydalanib amalga tadbiq etgan olim...?
 - a) Shannon
 - b) Norbert Viner
 - c) Loic Boizou
 - d) Aleksandr Gelbuk
4. Ehtimollik nazariyasidan foydalanib ingliz tilining birinchi entropik hajmini aniqlab bera olgan olim...
 - a) Shannon
 - b) Shemakin
 - c) Marchuk
 - d) Melchuk
5. Kim nutqni tanishda ehtimollik modellarini qo'llagan?
 - a) Tomas Vetson
 - b) Lee Kim
 - c) Bolshakon
 - d) Panina
6. S.Rizayev va S.Muhamedovlar kompyuter lingvistikasining qaysi yo'nalishi bo'yicha ishlarni amalga oshirgan?
 - a) statistik tahlil
 - b) ehtimollik nazariyasi
 - c) matematik modellashtirish
 - d) tizimli tahlil

7. Kompyuter lingvistikasi qaysi fanlar bilan aloqador?

- a) psixologiya
 - b) matematika
 - c) tarjima
 - d) barchasi
8. Qaysi instrument mavjud hujjatlarga avtomatik tarzda qo'shimchalar qo'shish va strukturaviy annotatsiya uchun mo'ljallangan?
 - a) brat insrumenti
 - b) python dasturi
 - c) tarjimon instrumenti
 - d) korpus instrumenti
 9. Trados dasturi qaysi texnologiyaga asoslangan?
 - a) tarjima xotirasi texnologiyasi
 - b) terminologik boshqaruv texnologiyasi
 - c) LUNAR tizimi
 - d) brat texnologiyasi
 10. Tayyor lingvistik instrumentlarga qaysilar kiradi?
 - a) lingvistik tahlil uchun
 - b) python dasturi uchun va tarjimon jarayoni uchun
 - c) korpus lingvistikasi tahlili uchun
 - d) barcha javoblar to'g'ri

2-MODUL. KOMPYUTER LINGVISTIKASIDA LINGVISTIK MASALALAR

3-mavzu. Grammatik nazariyalar

3.1. Formal til nazariyasi

3.2. Chomskiy iyerarxiyasi

3.3. Tobelanish nazariyasi

Tayanch soʻz va iboralar: *til, alifbo, grammatika, kategorial grammatika, kognitiv grammatika, ishoraviy belgilarga asoslangan konstruktiv grammatika, formal til nazariyasi, konkentenatsiya, oʻzgaras ifodalar, chekli avtomat, formallashtirish, Chomskiy iyerarxiyasi, tobelanish nazariyasi, universal tobelanish, teg, lemma*

3.1. Formal til nazariyasi

Til (L) – simvollarning chekli toʻplamidan qurilgan chekli uzunlikdagi gaplar majmuidir. Grammatika – tildagi gaplarni vosita sifatida qaraydi. L ning grammatikasi L aniqlikdagi funksiyaning tartibi hisoblanadi. Grammatikaning turlari quyidagicha: 1) tavsifiy (til uchun lingvistik meʼyorlarini oʻrganadi); 2) deskriptiv (tilning faol qoʻllanish doirasini oʻrganadi); 3) formal (aniq qoidalarga asoslangan grammatika); 4) generativ (tabiiy til ifodalarini shakl usulda oʻrganadi).

Formal grammatikaning generativ va analitik kategoriyalari mavjud. Generativ grammatika tildagi mavjud satrlarni yaratish algoritmini formallashtiradi. Analitik grammatika esa qoidalar toʻplami boʻlib, kiritilgan satrni tildagi mavjud satrlarni ifodalovchi bulean natijaga kamaytiradi. Generativ grammatika tilni qanday yozishni, analitik grammatika esa uni qanday oʻqishni oʻrganadi.

Formallashtirish – bu bilimlar majmuasini, mazmunini muayyan ifodalar yordamida sunʼiy til belgilari orqali ifodalashdir. Grammatikani formallashtirish XX asrning oʻrtalarida rivojlangan. Bu bevosita tilning strukturasi yanada aniq parametrlarda ifodalab, lingvistik imkoniyatlar kompyuterga moslashtirilgan. Formal grammatika nazariyasi matematik mantiq, matematik lingvistikaning rivojlanishi natijasida yuzaga kelgan. Grammatikani formallashtirish (lotincha “forma” – “shakl”, “tashqi koʻrinish”) konsepsiyasining

asoschisi amerikalik tilshunos Noam Chomskiydir. Formal grammatika abstrakt simvollar orqali bayon etiladigan grammatika boʻlib, unda til qoidalari, soʻz yasalishi, birikma va gap qurilishi turli xil lingvistik modellar, struktur sxemalar yordamida tavsiflanadi. Formal grammatika 3 ta birlik asosida ish koʻradi:

1. Word-classes – soʻz turkumlari (tub va yasama soʻzlar).
2. Phrases – gap boʻlaklari.
3. Sentences – gap konstruksiyalari sxemalari, modellari.

Formal grammatikaning ikki asosiy tushunchasi mavjud: a) generatsiya – tugʻdirish, hosil qilish; b) tanish. Shunga koʻra, grammatika ikkiga boʻlinadi: 1) tugʻdiruvchi grammatika (generative grammar) – toʻgʻri gaplarni qabul qilish protsedurasi; 2) tushunuvchi, tanuvchi grammatika (recognizing grammar) – toʻgʻri gaplarni tushunish protsedurasi (ravishdosh, sifatdosh, harakat nomi yoki majhul nisbat shakllari) transformatsiyaning operatori deb ataladi.

Formal grammatikaning quyidagi turlari mavjud:

1. Regulyar grammatika
2. Erkin kontekstli grammatika
3. Bevosita tarkib toptiruvchilar grammatikasi
4. Leksik-funksional grammatika
5. Montegyu grammatikasi

Erkin kontekstli grammatikaning formal tavsifi quyidagilardan iborat: G -grammatika, N , Σ , R , $S - N$ (noterminal belgilar toʻplami), Σ -terminal belgilar toʻplami, R -qoidalar yoki hosilalar toʻplami.

Dastlab bevosita ishtirokchilarga ajratish modeli tanqidiy qaralib, undagi ayrim gʻoyalar 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 taʼsirida yuzaga kelgan, soʻng bu yoʻnalish Z.Harrisning shogirdi N.Chomskiy tomonidan rivojlantirildi. Unga koʻra, dastlabki yadro strukturalar birikmalarning struktur qoidalar tizimi orqali tugʻiladi. Yaʼni har qanday til grammatikasi yadro strukturalar (kernel structures) yigʻindisidir. Yadro gaplar deb sodda, yigʻiq, darak

gaplarga aytiladi. Ingliz tilida yadro gaplar sifatida quyidagi konstruksiyalar ajratib ko'rsatiladi (4-jadval).

Mazkur konstruksiyalarda N (Ot) turli xil determinantlar (masalan, artikl, olmosh) hisobiga, V (fe'l) ravishlar, adverbial (ravishli) birikmalar hisobiga kengayishi mumkin. Ushbu yadro konstruksiyalarning bir-biri bilan 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 bor: 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. Masalan, *Kitob sotib oldim* – *kitobning sotib olinishi* – *kitob sotib olingach* – *kitob sotib olinganda* – *kitob olishi bilan* – *kitob men tomonimdan sotib olindi*. 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.

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. Transformatsion va derivatsion jarayonlarining tabiiy tillarga faol tatbiq etilishi avtomatik tarjima samaradorligini ta'minlaydi. Uchinchidan, transformatsion grammatika til o'qitishni modellashtirishda ham muhim ahamiyatga ega. To'rtinchidan, transformatsion grammatika tabiiy tillardagi paradigmik va sintagmatik munosabatlarning me'yoriyligi va grammatik to'g'ri va noto'g'ri jumlalarga ajratishning aniq mezonlarini belgilaydi. N.Chomskiy tavsiya etgan 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 simvollarini joylashadi, o'ng qism esa ularning simvollarini o'rniga almashuvchilar yoki simvollar zanjiri (birikuvini)dan iborat bo'ladi.

Bundan tashqari lingvistik modellashtirishda maksimum entropiya (Maximum Entropy) modeli samarali usullardan biridir. Ushbu model Berger va Della Pietra tomonidan NLP texnologiyasiga tatbiq qilinadi⁴⁶.

3.2. Chomskiy iyerarxiyasi

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

1.0 grammatika (unrestricted phrase – structure grammar) – murakkab generativ grammatika, bunda $G=(N, \Sigma, P, S)$ belgilaridan foydalaniladi. N – alifbo (N – noterminal simvollar, u grammatikadagi simvollarini o'z ichiga oladi); Σ – metatilning terminal simvollarini, u tildagi so'zlarni qamrab oladi); S – noterminal to'planning 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 ($V=TWN$). 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) – kontekstdan holi grammatika bo'lib, u programmashtirish tillari unsurlarining generatsiyasida (ifodalar, buyruqlar tizimida) ishlatiladi.

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

⁴⁶Gulila Altenbek Identification Of The Kazakh basic Phrases Based On The Maximum Entropy Model / Turklang. 2013, Astana, 2013. – C. 49.

Kontekstdan holi grammatika (Context-Free Grammar-CFG) frazaga asoslangan struktur grammatika deb ham nomlanadi. Kontekstdan holi grammatika qoidalar to'plami yoki tildagi simvollarning guruhi leksikon, so'z yoki simvollarning leksikonini aks ettiruvchi hosilalardan iborat bo'ladi. Iyerarhik tuzilmani quyidagicha tushuntirish mumkin:

sinf	Grammatika	Til	Avtomat
0-TUR	cheklanmagan	rekursiv (Tyuring)	Tyuring mashinasi
1-TUR	kontekstga bog'liq	kontekstga bog'liq	chiziqli-chekli
2-TUR	kontekstdan holi	kontekstdan holi	pushdown (itaruvchi)
3-TUR	regulyar (o'zgarimas)	regulyar	chekli

Noam Chomskiyning ilmiy qarashlari tilning formal tavsifini yozishda asos vazifasini o'tadi. U lingvistik modelning ikki turini keltiradi: 1-minimal lingvistik nazariyaga moslashgan Markov jarayoni sifatida til konsepsiyasiga asoslangan nazariy model hamda tarkib toptiruvchi analiziga asoslangan frazali strukturaviy model. N.Chomskiy strukturalarni analiz qilishning ikki yondashuvi mavjud ekanligi, bu o'z navbatida har bir sintaktik qolipning tuzilishidan kelib chiqishini ta'kidlaydi: *chuqur struktura (deep structure)* va *yuzaki struktura (surface (shallow) structure)*. Chuqur (ichki) struktura bu semantik munosabatga qurilgan tuzilma, yuzaki struktura esa uning sintaktik (shakliy) ifodasi.

So'z va alifbo

So'z yoki satr muayyan alifboga tegishli bo'lgan simvollar yoki harflarning muayyan ketma-ketligidan iborat bo'ladi. So'z uzunligi simvol bilan belgilanadi. Hech qaysi simvolga ega bo'lmagan so'zlar *bo'sh so'z* deb nomlanadi. So'z W simvoli bilan, bo'sh so'zlar Λ (yoki λ yoki ε yoki 1) bilan belgilanadi. Bo'sh so'zning uzunligi 0 bilan belgilanadi. Agar alifbodagi k simvollarni anglatsa, u holda k^n so'zlarning uzunligi n bilan belgilanadi.

So'zlarning asosiy operatsiyasi konkatenatsiya hisoblanib, ular qo'shma so'z hisoblanadi. $w1$ va $w2$ so'zlarning konkatenatsiyasi $w1w2$ tarzida ifodalanadi. Konkatenatsiya agglyutinatv tillar uchun mos keladi⁴⁷. Masalan, Inuktitut shevasida morfemalarning morfotakrikasi hatto gap darajasida bo'lishi mumkin:

Inuktitut **parimunnngaujumaniralausimannngittunga**
shevasi

morfotaktikasi Paris+mut+nngau+juma+niraq+lau+sima+nngit+junga
Ma'nosi Men Parijga borishni xohlashimni hech qachon aytmaganman.

Biroq ayrim tillar uchun bu hodisa to'g'ri kelmaydi. Masalan, tagalog shevasining morfemalarini ajratishning imkoni mavjud emas. Chunki so'zlarning birinchi CV bo'g'ini redublikatsiyaga uchraydi:

O'zak	CV+o'zak	Ma'nosi
pili	pipili	tanlamoq

Shuningdek, ayrim tillar uchun to'liq redublikatsiya hodisasi kuzatiladi. Jumladan, malay tilida quyidagicha bo'lishi mumkin:

O'zak	Ma'nosi	Redublikatsiya	Ma'nosi
anak	bola	anak-anak	bolalar

Agar alifbo $\{a,b,c\}$ dan iborat bo'lsa, undan hosil bo'luvchi qo'shma so'zlar yig'indisi quyidagicha bo'lishi mumkin:

$$w1 = aacbba, w2 = caac, w1w2 = aacbbacaac$$

$$w1 = aacbba, w2 = \Lambda, w1w2 = w1 = aacbba$$

$$w1 = \Lambda, w2 = caac, w1w2 = w2 = caac$$

Konkatenatsiya quyidagi assotsiativ holatda bo'ladi:

$$w1(w2w3) = (w1w2)w3.$$

Noam Chomskiy *til* terminiga quyidagicha ta'rif keltirgan: "Til bu (cheklangan yoki cheklanmagan) gaplar yig'indisidir. Barcha tabiiy tillarning yozma va og'zaki shakllari mazmunga ega bo'lib, har bir tilning cheklangan va tartiblangan fonemalari (yoki harflari)

⁴⁷ Kenneth R. Beesley, Lauri Kartunen Finite state morphology, USA, 2003. – P. 396.

bo'ladi. Shunga o'xshash matematikada formallashtirilgan tizimlarning "gaplar" to'plami *til* deb nomlanadi. Tilni L lingvistik analiz qilishning asosiy maqsadi L dagi grammatik ketma-ketlikni gap bo'lmagan boshqa nutqiy unsurlardan ajratish hamda grammatik ketma-ketlikning strukturasi o'rganishdan iborat⁴⁸.

Tillarning maxsus namunalarini aniq miqdordagi so'zlardan iborat bo'ladi. Bo'sh tilda esa hech qanday so'z bo'lmaydi va u \emptyset belgisi bilan belgilanadi. Quyidagi \subseteq (qo'shish), \subset (tegishli), \cup (to'plam), \cap (kesishma), $-$ (farq) and (alifbo ustida yozilgan barcha so'zlarni to'ldiradi) operatorlardan foydalaniladi.

O'zgarmas (regular) ifodalar

Matnlarni qidirishda kompyuter uchun standart tilning o'zgarmas ifodalari yaratilgan. Bunday ifodalarning eng sodda shakli bu tartiblangan sodda simvollar sanaladi. Jurafskiyning qayd qilishicha⁴⁹, har bir ifoda yagona birlikka ega. Agar ifodalar $/a/$ tarzda berilsa, faqat va faqat shu qiymatga ega simvolni bildiradi. Bosh harfdagi $/A/$ kichik belgining qiymatiga mos emas. Agar to'rtburchak qavs berilsa, $/[aA]/$ degan ma'noni beradi. Masalan, $/[a,b,c]/$ berilgan bo'lsin, u holda qatordagi *bahor* so'zida ikki element mos kelishini; agar ifoda $/-/$ holatda berilsa, ikki element o'rtasidagi simvollarini: $/[4-8]/$ bunda 4, 5, 6, 7 yoki 8ni nazarda tutadi. $/[^{A}]/$ bu belgi bo'lishi mumkin emas, degan ma'noga ega: $/^{A-D}/$ bu A bosh harfidan D gacha bo'lgan harflardan tashqari belgilarni bildiradi. Ixtiyoriy belgilar uchun esa $/[?]/$ belgisi qo'yiladi. Bu undan oldingi yoki keyingi belgining kelishi yoki hech qanday element kelmasligi mumkin: $/[buru?n]/$ - *burun* yoki *burni*. Agar $/a^*/$ belgisi bilan kelsa, o'sha element yoki bir necha marta takroriy qiymatni hisobga oladi: $/[ab]^*/$ bo'lsa, satr sifatida *aaaa*, *baba*, *abab* yoki *bbbb* bo'lib kelishi mumkin. Agar $/./$ kelsa, berilgan simvollar o'rtasidagi biror belgi borligini ko'rsatadi: $/u.ga/-$ *unga*. Muayyan so'zning istalgan satrini aniqlamoqchi bo'lsak, $/abc.^*abc/$ tarzida yoziladi. $^$ belgisi satrning boshini ko'rsatadi. Agar biz $/^{A}Ona/$ tarzida bersak, demak *ona* satrning faqat boshida kelishini

ifodalaydi. Satrning oxirini esa $\$$ dollar belgisi bilan beriladi. $\backslash b$ belgisi so'z chegarasini bildiradi. Agar qidirilayotgan ikki element haqidagi ma'lumot kerak bo'lsa, u holda yoki $|$ belgisi yoziladi. $/gul|daraxt/$ bunda *gul* yoki *daraxt* turkumiga kiruvchi so'zlarni ilova qilinadi. Shuningdek, maxsus belgilar ham mavjud:

- $\backslash n$ -yangi qator uchun
- $\backslash t$ -tab uchun
- $\backslash ?$ -so'roq belgisi uchun
- $\backslash .$ -nuqta uchun
- $\backslash *$ -ajratib yozish uchun

O'zgarmas ifoda + ishorasi orqali konkatenatsiya va konkatenatsiya hosilasidan iborat tilni aniqlovchi formula hisoblanadi. Bu operatsiyalar zaruriyatiga ko'ra $()$ qoidalar to'plamiga birlashadi. Formulaning tarkibi bu alifbo Σ , bo'sh til \emptyset , bo'sh so'z Λ va tushirilib qoldirilgan to'plam $\{ \}$ bilan belgilanadi.

L_1 va L_2 tillarning konkatenatsiyasi $L_1L_2 = \{w_1w_2 \mid w_1 \in L_1 \text{ and } w_2 \in L_2\}$.

L_1 va L_2 tillarning n -konkatenatsiyasi $L_n = \{W_1W_2 \dots W_n \mid W_1, W_2, \dots, W_n \in L\}$

Chekli avtomat (Finite automata) tildagi so'zlarni aniqlash uchun ishlatiladi.

N.Chomskiy chekli avtomat terminiga quyidagicha izoh beradi: cheklangan grammatika terminiga tasavvurimizdagi mashina muayyan tartiblangan chekli simvollarini so'z sifatida, boshlanish va tugallanish nuqtasigacha bo'lgan oraliqni gap sifatida aniqlasa, gaplarning to'plami ham shu tarzda yig'iladi. Aniqlangan chekli til (finite state language), mashinani esa aniqlangan chekli grammatika deb nomlanadi⁵⁰. Olim ushbu grammatikani grafik shaklda quyidagicha ifodalaydi (13-rasm):



13-rasm

⁴⁸ Noam Chomsky Syntactic structures. - Paris: Mouton, 1957. - P. 13.

⁴⁹ Jurafskiy D., James H. Martin Speech and language processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 2007 - P. 12-13.

⁵⁰ O'sha joyda -P. 19

Ya'ni shakl nuqtai nazaridan ushbu gap muayyan siklda birlashadi. Ushbu gapni ixtiyoriy ravishda kengaytirish mumkin.

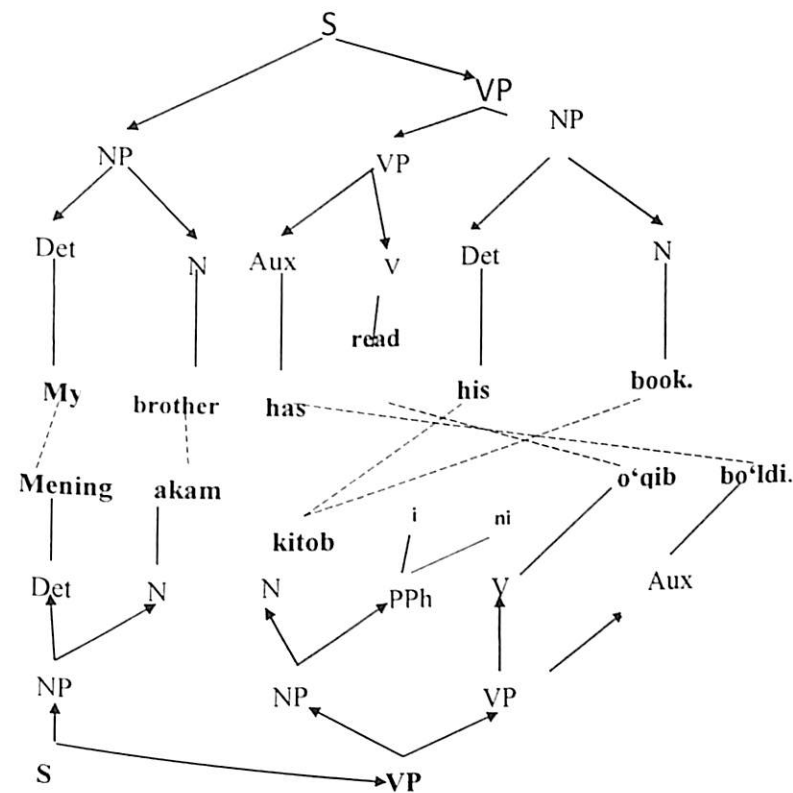
Hech qaysi sintaktik shakl barcha gaplar uchun yagona model bo'la olmaydi. Chunki yadro gaplardan qo'shimcha birliklar orqali uni istalgan kattalikda kengaytirish mumkin. N.Chomskiy sintaktik darajada tobelik analizi (parsing)ni quyidagi lingvistik tavsif orqali aks ettiradi⁵¹:

- (i) Gap → NP + VP
- (ii) NP → T + N (NP → NP_{sing} / NP_{pl})
- (iii) VP → V + NP
- (iv) T → the
- (v) N → man, ball
- (vi) V → Aux + V
- (vii) V → hit, took
- (viii) Aux → C (M) (have + en) (be + ing)
- (xix) M → will, can, may, shall, must

Har bir qoida farazga ko'ra X → Y tarzida ifodalansa, "X va Y"ni qayta ifodalashni N.Chomskiy "gaplarning yasalishi" (*derivation of the sentence*) deb nomlaydi.

Uning fikricha, har qanday grammatik struktura tarkib toptirish analiziga asoslanadi. Muayyan grammatika boshlang'ich qator (string)lar to'plami Σ hamda "tuzilgan formulalarning" muayyan F to'plami orqali aniqlanadi. Σ dagi birlik gap, F esa (i)-(vi) kabi qoidalardan iborat. Har bir aniq chekli til terminal belgilar majmui tildir, lekin aniq chekli til bo'lmagan terminal til ham mavjud.

O'zbek va ingliz tillarida N.Chomskiy nazariyasini quyidagi chizmada ifodalash mumkin (5-chizma):



5-chizma

Erkin kontekstli grammatikaga oid qoidalarni daraxtsimon tarzda aks etgani bois matn korpuslari uchun sintaktik tahlil modellari yaratilgan. Sintaktik tarzda annotatsiyalangan korpus Treebank deb nomlanadi.

3.3. Tobelanish nazariyasi. Universal tobelanish (dependency theory universal dependency (UD))

Sintaktik strukturalarni o'rganish tillarning xususiyatidan kelib chiqib, turli nazariyalarni yaratilishiga sabab bo'lgan.

Tobelanish nazariyasi uzoq tarixga ega bo'lib, Panini grammatikasi va arab tilshunoslarining ilmiy qarashlarida o'z ifodasini topgan. Kompyuter lingvistikasi doirasida ushbu nazariya

⁵¹ O'sha joyda. -P.26

1990-yillardan boshlab rivojlana boshlangan⁵². Bu kabi strukturalar franzus lingvisti L.Tenyer (1988) tomonidan qo'llanilgani bois g'arbda bunday sintaktik shakllar "*Tenyer ustunlari*" deb nomlangan. Ayrim adabiyotlarda A.A. Zaliznyak va Y.V. Paducheva⁵³ning tadqiqotlarida dastlab ilgari surilgani, biroq keng darajada ommalashmaganligi haqida fikrlar bildiriladi.

Tobelanish grammatikasining o'ziga xos jihati shundan iboratki, matnda qo'llanilgan so'zlar erkin joylashadi va gap markazi kesim hisoblanadi. Unga ko'ra tobelik nazariyasida to'rt asosiy unsur ustuvorlik qiladi: *tobe-hokim munosabatidagi so'zlarning mantiqiy bog'liqligi, so'zlar qurshovidagi tobelik masofasi, komponentlarning joylashuvi, hokim so'zlar valentligi*. So'zlarning o'zaro bir-biriga bog'lana olish imkoniyati *valentlik* deb nomlanadi.

2011-yil Barselonada bo'lib o'tgan DepLing tobelanish lingvistikasiga bag'ishlangan konferensiyada NLP bo'yicha nazariy va amaliy ahamiyatga ega katta masshtabdagi tadqiqot natijalari muhokama qilinadi. Ilmiy izlanishlarda tobelik nazariyasining tavsifi, matematik ifodasi qanday bo'lishi zarur, degan savollarga Kim Gerdes, Sylvain Kahane, korpus annotatsiyasi uchun sintaktik funksiyalarining to'liq majmuasini rivojlantirish bo'yicha Alicia Burga, Simon Mille, Leo Wanner; sintaktik ma'lumotlar bazasida (tree bank) tobelik xatoliklarini bartaraf etish bo'yicha Katri Haverinen, Filip Ginter, Veronika Laipalla, Samuel Kohonen, Timo Viljanen, Jenna Nyblom va Tabio Salakoski kabi olimlar o'z tadqiqotlari bilan ishtirok etdilar. Bernad Bonet, Leo Wanner va

⁵² Hays, D. G. (1964). Dependency theory: A formalism and some observations. *Language* 40: 511-525.; Tapanainen, P. and Jarvinen, T. (1997). A non-projective dependency parser. *Proceedings of the 5th Conference on Applied Natural Language Processing*, pp. 64-71.; Tesniere, L. (1959). 'El' ements de syntaxe structurale'. Editions Klincksieck.; Schroder, I. (2002). "Natural Language Parsing with Graded Constraints. PhD thesis, Hamburg University; Nivre, J., Hall, J. and Nilsson, J. (2004). Memory-based dependency parsing. In Ng, H. T. and Riloff, E. (eds), *Proceedings of the 8th Conference on Computational Natural Language Learning (CoNLL)*, pp. 49-56. Obrebski, T. (2003). Dependency parsing using dependency graph. In Van Noord, G. (ed.), *Proceedings of the 8th International Workshop on Parsing Technologies (IWPT)*. pp. 217-218.

⁵³ Зализняк, Падучева 1964 – Зализняк А.А., Падучева Е.В. О связи языка лингвистических описаний с родным языком лингвиста // Программа и тезисы докладов в летней школе по вторичным моделирующим системам. Тарту, 1964.

Simon Millening ishlarida esa tobelik nazariyasiga asoslangan semantik strukturalar statistik til generatsiyasi yordamida matn shakliga moslashish mumkinligi haqida fikrlarlarni beradilar. Federiko Gobbo va Marko Benini, Denis Bechet, Alexander Dikovsky ishlarida tobelik grammatikasini formallashtirish masalalari kun tartibiga qo'yiladi⁵⁴.

O'tgan o'n yillikda morfemalar, so'zlarning semantik birliklari o'rtasidagi iyerarxik munosabatlari tobelanish nazariyasida kompyuter lingvistikasining parsing, generatsiya va mashina tarjimasida deyarli standard holatda ifodalanishi kuzatilmoqda⁵⁵. Kompyuterning sintaktik analiz qilish jarayonida 1-usul – to'g'ri usul bo'lib, unda so'zlarning morfologik belgilariga qarab uning gapdagi sintaktik vazifasi aniqlanadi. Masalan, qaratqich kelishi doim aniqlovchi, tushum kelishigi doim to'ldiruvchi bo'lib kelganda. Demak gapdagi fe'lni aniqlab, unga tobelanib keluvchi sintaktik tobelanishlar aniqlanadi. Bunda mantiqiy shart amalidan foydalanib "*agar..., u holda*" bo'lishi mumkin bo'lgan qoidalar jamlanmasidan foydalaniladi.

Sintaktik tahlil universal tobelanish (UT) deb ham ataladi. UT matnning morfosintaktik tasnifidir. Shu bois u ikki grammatik belgisi so'z shakllari (morfologik) hamda tobelanish (sintaktik) munosabatini aks ettiradi. Sintaktik munosabatga kirishganda shakl va funksiya atamalariga izoh kiritib o'tamiz.

Sintaktik shakl – so'z turkumlari va frazada aks etuvchi tarkibiy qismlarga frazalar (so'zdan kattaroq sintaktik birliklar), gapning bosh bo'laklari, sintaktik guruhlar NP, VP, PP kabilar kiradi. Sintaktik vazifalariga ega, kesim, to'ldiruvchi, hol, aniqlovchi kabilar kiradi.

Sintaktik aloqaga kirishgan so'zlarning morfologik holati uch bosqichdan iborat:

- lemma – so'zning semantik mazmuni (so'zlarning lug'atdagi holati);
- so'z turkumi tegi – (POS tag) so'z bilan bog'langan mavhum leksik kategoriya;

⁵⁴ Kim Gerdes, Eva Hajikova, Leo Wanner. *Computational dependency theory*. Netherlands, IOS, 2013, -P.6

⁵⁵ O'sha joyda. -P.3.

- muayyan soʻz shakli bilan bogʻlangan leksik va grammatik xususiyatlar toʻplami (set of features).

Lemma agar lugʻatda topilmasa _ belgisi qoʻyiladi.

Shu boʻyicha tadqiqot olib borgan olimlar⁵⁶ tildagi koʻp tarkibli soʻzlarni obrazli qilib “NLP uchun boʻyindagi ogʻriq” deb baholashadi. Chunki koʻp tarkibli soʻzlarni “soʻzlar kesimidagi idiosinkretik interpretatsiyalar” deb tarʼriflaydilar va muammolar sifatida quyidagilarni sanab oʻtadilar:

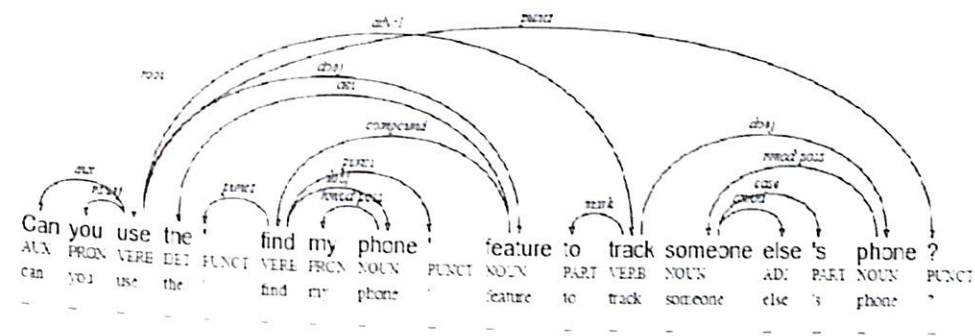
- omonimiya;
- dialektlarda ifodalarning turlicha qoʻllanib, (amerika, britaniya, avstraliya ingliz tili dialektlari misolida) bir nechta generatsiya qilinishi;
- idiomatik muammo (ham tarkibga ajraluvchi, ham sintaktik yaxlit butunlik *kapalagi uchib ketmoq* kabi);
- ajratib yoziluvchi qoʻshma soʻzlar (*sotib olmoq*, *Oʻrta Chirchiq*);
- leksik birikmali soʻzlar (*yomgʻir yogʻib yotibdi*);
- atoqli otlar va qoʻshma otlar.

Sintaktik strukturalar binar assimetrik munosabatlardan iborat leksik birliklardan tashkil topgan. L.Tesnir (1959) oʻz asarida tobelik nazariyasi boʻyicha gap soʻzlardan iborat komponentlardan tuzilishi, har bir soʻz lugʻatdan ajralgan holda muayyan gapga tegishli boʻlishi, soʻz va ular bilan keluvchi yondosh soʻzlar oʻrtasida gap strukturasida umumiy shakllangan umumiy bogʻliqlik mavjudligi, struktur bogʻliqlik soʻzlar oʻrtasidagi tobelik munosabatini yuzaga chiqarishi, shuningdek, har bir bogʻlanishda hokim (governor) va tobe soʻzlardan (subordinate) iborat boʻlishi haqida fikrlarni bildiradi.

Tobelik grammatikasi matnlarning semantik analiziga yaqin, tipologik oʻzgarmas ifodalarni aniqlash usuli oson, kompyuter analizida gaplarni kengaytirish imkoni mavjudligi kabi oʻziga xos afzalliklariga ega. Tenyer gaplardagi anaforik aloqani ham eʼtibordan chetda qoldirmaydi, yaʼni oʻzidan oldingi soʻzga ishora qiluvchi olmoshlar (anoforalar) va oldingi soʻz turkumi otlar (antesedent) kabi guruhlar ajratadi.

⁵⁶ Ivan A. Sag, Timothy Baldwin, Francis Bond, Ann Copestake, and Dan Flickinger. Multiword Expressions: A Pain in the Neck for NLP /http://lingo.stanford.edu

Oʻzbek tilining tobelik grammatikasining sintaktik qolipi quyidagicha: SOV (ega+toʻldiruvchi+kesim), ingliz tiliniki esa SVO (ega+kesim+toʻldiruvchi).



14-rasm

Tobelik grammatikasida soʻzlarning sintaktik oʻrni erkin boʻladi (14-rasm). Ushbu grammatika turkiy tillar uchun mos. Universal tobelik grammatikasida har bir tilning tipologik tabiatidan kelib chiqib maxsus ifodalar (teglar) kiritilgan:

<i>Toʻldiruvchi birikmalarni bogʻlash uchun qoʻllaniladigan teglar</i>
NSUBJ - Nominal ega
DOBJ - bevosita toʻldiruvchi
IOBJ - bilvosita toʻldiruvchi
CCOMP- murakkab toʻldiruvchi
XCOMP- birikmali toʻldiruvchi
<i>Nominal modifikatorli munosabatlar tavsifi uchun qoʻllaniladigan teglar</i>
NMOD - Nominal modifikator
AMOD - Sifat modifikator
NUMMOD-son modifikator
APPOS - joylashtiruvchi modifikator
DET-aniqllovchi
CASE- kelishik/ predlog
CONJ - bogʻlovchi

Universal tobelanish tillararo (70 ta til) sintaktik analiz korpusi asosida yaratilgan loyiha hisoblanib, bunda tillarning tipologik nuqtai nazaridan multilingval parserni rivojlantirish maqsad qilib qo'yilgan. Annotatsiya shakli Stanford tobelanishi, (de Marneffe et al., 2006, 2008, 2014), Google universal so'z turkumlari tegi (Petrov et al., 2012) hamda morfosintaktik tegset uchun interlingval texnologiyaga (Zeman, 2008) asoslangan⁵⁷.

Savol va topshiriqlar

1. Formal til nazariyasida konkatenatsiyaning o'rni qanday?
2. Tildagi ma'lumotlar nima uchun lingvistik analiz qilinishi kerak?
3. Sintaktik shakl barcha gaplar uchun yagona model bo'la oladimi? Nima uchun?
4. N.Chomskiy lingvistik tavsifning tobelik analiziga ko'ra sintaktik darajada shakllanishni qanday aks ettiradi?
5. Grammatikani formallashtirish konsepsiyasining mohiyati nimalardan iborat?
6. Chomskiy formal grammatikasi yana qanday termin bilan qo'llaniladi?
7. Olimlar qanday ishni "NLP uchun bo'yindagi og'riq" deb baholagan?
8. Ko'p tarkibli so'zlar qanday muammolarni keltirib chiqarishi mumkin?
9. Formal grammatikaga o'zingiz namuna keltiring va uni tahlil qiling.
10. O'zbek tilining formal grammatikasini muayyan matn namunasi yordamida tahlil qiling.
11. *The reporter talked the event in Tashkent.* Ushbu gapni ikki xil usulda Chomskiy tahlili bo'yicha gap bo'laklariga ajrating.

⁵⁷ <https://universaldependencies.org/introduction.html>

3-mavzu yuzasidan test

1. Terminal belgilar to'g'ri ko'rsatilgan qatorni aniqlang.
 - a) harflar, qoidalar, raqamlar
 - b) qisqartmalar, raqamlar
 - c) satrlar, simvollar
 - d) harflar, bo'g'inlar
2. Grammatikani formallashtirish konsepsiyasining asoschisi kim?
 - a) Chomskiy
 - b) Kim Gerdes
 - c) Sylvain Kahane
 - d) Simon Mille
3. Sintaktik tahlil yana qanday nom bilan ataladi?
 - a) universal tobelanish
 - b) morfologik tahlil
 - c) kesimni topish
 - d) yadroli tahlil
4. Lemma nima?
 - a) so'zning semantik mazmuni (so'zlarning lug'atdagi holati)
 - b) so'z turkumlari
 - c) so'zning grammatik ma'nosi
 - d) so'zning ko'chma ma'nosi
4. Sintaktik ma'lumotlar bazasida (tree bank) tobelik xatoliklarini bartaraf etish bo'yicha kimlar tadqiqotlar olib brogan?
 - a) Katri Haverinen, Filip Ginter, Veronika Laipalla, Samuel Kohonen, Timo Viljanen, Jenna Nyblom va Tabio Salakosk
 - b) Bernd Bohnet, Leo Wanner va Simon Millen
 - c) Federiko Gobbo va Marko Benini
 - d) Denis Bechet, Alexander Dikovsky
5. Qaysi olimlar tobelik nazariyasiga asoslangan semantik strukturalar statistik til generatsiya yordamida matn shakliga moslashi mumkinligi haqida fikrlarini beradilar?
 - a) Bernd Bohnet, Leo Wanner va Simon Mille
 - b) Federiko Gobbo va Marko Benini
 - c) Denis Bechet, Alexander Dikovsky
 - d) Jenna Nyblom va Tabio Salakosk

6. Tobelanish nazariyasi qachondan boshlab rivojlana boshlangan?

- a) 1990-yillardan
- b) 1980-yillardan
- c) 1975-yillardan
- d) 1985-yillardan

7. Formal grammatika qaysi birliklar asosida ish ko'radi?

- a) word-classes – so'z turkumlari (tub va yasama so'zlar)
- b) phrases - gap bo'laklari
- c) sentences - gap konstruksiyalari sxemalari, modellari
- d) barcha javoblar to'g'ri

8. N.Chomskiy qoidalarning ifodalanish turiga qarab formal grammatikani necha tipga ajratgan?

- a) 4
- b) 5
- c) 3
- d) 2

9. Agar lug'atda lemma topilmasa, o'rniga qanday belgi qo'yiladi?

- a) _
- b) +
- c) ...
- d) ?

4-mavzu. Lingvistik protsessorning grammatik komponenti

4.1. Lingvistik bilimlar bazasi

4.2. Lingvistik bilimlar bazasini loyihalashning fonetik, morfologik, sintaktik va semantik xususiyatlari

4.3. Morfologik ma'lumotlar bazasi

Tayanch so'z va iboralar: *lingvistik protsessor, formal model, lug'at, grammatika, lingvistik jadvallar, qoidalar, morfologik component, morfologik paradigm, omonimiya, omoforma, omograf, omofon, morfologik ma'lumotlar bazasi, turg'un birikmalar, tarjima ekvivalentliklari*

4.1. Lingvistik bilimlar bazasi

Lingvistik protsessor termini⁵⁸ tabiiy tildagi matnlarni analiz va sintez qiluvchi dasturiy majmuadir. Uning uch asosiy komponenti mavjud:

– lingvistik (tabiiy tilning formal modeli, lug'at, grammatika, lingvistik jadvallar, qoidalar);

– matematik-algoritmik (formal til translyatori, matnni qayta ishlovchi algoritmlar);

– dasturiy ta'minot.

Lingvistik bilimlar bazasi mashina uchun matnlarni qayta ishlashda zarur bo'lgan barcha ma'lumotlarni o'z ichiga qamrab oladi. Tabiiy tilining lingvistik sathlari quyidagilardan iborat: leksik, morfologik, sintaktik, semantik. Bularning har biri o'zining struktur ma'lumotlar bazasiga ega. Har bir lingvistik birlik muayyan umumiy belgilariga ko'ra paradigmalariga birlashadi.

Lingvistik ma'lumotlar bazasi (ta'minoti) deyilganda, tilga oid barcha ma'lumotlar tushuniladi. Til tuzilishi hamda matnni analiz va sintez qilishning aniq vazifalariga asosan lingvistik ta'minotida quyidagilarning mavjud bo'lishi talab qilinadi: matn – leksik asos – leksemashakl – so'z birikmasi – sintaktik gap strukturalari. Tarjimadan avvalgi tahlil jarayonida uch asosiy ierarxik bosqich amalga oshiriladi: matn-jumla-leksema shakl. Bunda leksemashakl

⁵⁸ Волкова И.А. Введение в компьютерную лингвистику. Практические аспекты создания лингвистических процессоров, Москва 2006. – С. 6

keng ma'noda matndan olingan istalgan segment birlik sifatida qaraladi. Gap va so'z birikmalarining barcha strukturaviy hamda boshqa xususiyatlari leksema shakl tushunchasida ifodalanadi⁵⁹.

Lingvistik bilimlar bazasiga quyidagilar kiradi:

Leksik resurslar:

- Kompyuter lug'atlari;
- Tezaurus, ontologiya.

Matniy resurslar:

- Matlarning sohaviy majmuasi;
- Matn korpusi.

Aralash resurslar:

FrameNet - so'zlarning ma'lumotlar bazasi.

Internet resurslar:

- WikiPedia, Wiki-lug'at, DBPedia.

Korpus uchun lingvistik ma'lumotlar deyilganda, og'zaki va yozma matnlar tushuniladi. Ular turli me'yorlarga ko'ra umumiy va xususiy belgilari bilan tasniflanadi. Endilikda elektron ko'rinishidagi matnlarning hajmi juda katta bo'lib, ularni saralashda tadqiqot metodologiyasidan kelib chiqiladi. Shuningdek, mavjud manbalardan foydalanish uchun esa mualliflik huquqi ham hisobga olinadi.

Matnlarni korpusga tayyorlashda u yoki bu tilning amaldagi yozuvi ham muhim ahamiyat kasb etadi. Ayniqsa, tillarning yunikodi masalasi matnlarni kompyuter o'qiy oladigan formatga keltirishda muhim omillardan biridir.

O'zbek tilida ikki xil grafika xizmat qiladi: lotin va kiril. Ma'lumot qiqdiruvi uchun har ikki grafikadagi ma'lumotni izlashda korpus bazasiga UTF-8 kodli matnlar kiritiladi.

Korpusning lingvistik ma'lumoti sifatida veb-sahifalar ham muhim o'rin tutadi.

4.2. Lingvistik bilimlar bazasini loyihalash

Fonologiya va amaliy fonetika kompyuter lingvistikasi uchun muhim sanaladi. Negaki fonologik modellar bu matn bilan bog'liq barcha masalalarda dastlabki o'rganish obyekti sanaladi. Umuman olganda, fonetika, leksikologiya, morfologiya, sintaksis va semantika bir-birini hamisha taqozo etadi. Lingvistik analizda har bir sath muayyan darajada ishtirok etadi.

O'zbek tili turkiy tillar oilasiga mansub bo'lib, lotin yozuviga asoslangan alifboda 29 ta harf mavjud:

Aa	Hh	Nn	Tt	O'o'
Bb	Ii	Oo	Uu	G'g'
Dd	Jj	Pp	Vv	Shsh
Ee	Kk	Qq	Xx	Chch
Ff	Ll	Rr	Yy	ng
Gg	Mm	Ss	Zz	'

O'zbek tili lotin alifbosi boshqa ko'plab lotin alifbolaridan farqli xususiyatga ega. Bu alifbodagi 5 harfning, ya'ni *o'*, *g'*, *sh*, *ch*, *ng* harflarning ikkita yunikoddan tashkil topganligi bilan farqlidir. Biroq ushbu belgilar uchun o'zbek tilida Unicode olinmagan. Agar har bir grafema uchun unicode olinsa, texnik xatolik bartaraf etilgan bo'lardi.

Har bir til o'ziga xos morfologik shakllanish belgilariga ega. Morfologik jihatdan fe'l so'z turkumini tahlil qilsak, zamon kategoriyasi tillarda turlicha ifodalanadi. Jumladan, ingliz tilida zamon kategoriyasida ko'makchi fe'llar yoki fe'lning o'zagi bilan bog'liq o'zgarishlar bo'lsa, o'zbek tilida affiksalar bilan shakllantiriladi. Ispan tilida esa har bir fe'lning yasaliş xususiyati o'zgacha. Italyan tilida esa aksariyat so'zlar mustaqil holda qo'llanmasligi mumkin: pomodoro-pomidor (*mj.rod, birlik*), pomodori (*mj.rod, ko'plik*); cipolla -piyoz (*jns.rod, birlik*), cipolle (*jns.rod, ko'plik*). Ushbu holatda pomodor-, cipoll- o'zining asl

⁵⁹ Abduraxmonova N. Mashina tarjimasining lingvistik ta'minoti Toshkent: Akademnashr, 2018. - B. 31.

holatida qo'llanila olmaydi. Fe'l kategoriyasi turli tillarda morfologik jihatdan o'ziga xos shakllanadi:

O'zbek	Ingliz	Ispan		Rus	Nemis	Italyan
		To'g'ri fe'l uchun (hablar-so'zlamog)	noto'g'ri fe'l uchun (tengerbor)			
(Men) gapiraman	I speak	(Yo) hablo	tengo	Я говорю	Ich spriche	Lo parlo
(Sen) gapirasan	You speak	(Tu) hablas	tienes	Ты говориш	Sie sprechen	Tu parli
U gapiradi	He, she speaks	(El, ella) habla	tiene	Он, она говорит	Er, sie spricht	Egli, lei parla
Biz gapiramiz	We speak	Hablam os	tenemos	Мы говорим	Wir sprechen	Noi parliamo
Siz gapirasiz	You speak	Tu hablas	teneis	Вы говорите	Sie sprechen	parli
Ular gapiradilar	They speak	Ellos hablan	tienen	Они говорят	Sie sprechen	Essi parlano

Masalan, xitoy, koreys va yapon tillarida matnni ajratishga xizmat qiluvchi tinish belgilar ishlatilmaydi. Kompyuter orqali segmentlarga ajratish uchun bunday tillar uchun biroz murakkab hisoblanadi. Shu bois har bir matn alohida morfologik analiz qilinishi kerak.

Lingvistik bilimlar bazasini loyihalashning morfologik xususiyati

Matn bilan bog'liq har qanday dasturiy ta'minot lingvistik bazaga tayanadi. Morfologik ma'lumotlar bazasi lingvistik ta'minotning eng muhim sath hisoblanadi. Lingvistik ma'lumotlar bazasini loyihalash va ularni tizimli boshqarishda maxsus tillar bilan ishlash talab qilinadi. Buning uchun kompyuter bazasiga kiritiladigan hujjatlar, masalan, matn korpusini annotatsiyani berishda kompyuter o'qiy oladigan GML, SGML, XML, HTML kabi formatlardan foydalanish belgilab qo'yilgan hamda SgmlQL, SQLlight kabi so'rov tillarida lingvistik ma'lumotlar bazasi

loyihalanadi. SGML tili hujjatlar tavsifini berish uchun xalqaro standart (1990-yilda Goldfarb tomonidan yaratilgan) til hisoblanadi. Uning elementlari quyidagilardan iborat: $\langle n a_1 = v_1 \dots a_2 = V_k \rangle c \langle /n \rangle$, bu yerda $\langle n \dots \rangle$ boshlanish tegi $\langle /n \rangle$ tugallanish tegi; n turning nomi, c uning tarkibi, $a_1 = v_i$ juft qiymatlari atributlari. Har bir elementning tarkibi bo'sh to'plam, satr, element yoki satrlar va elementlarning ketma-ketligi bo'lishi mumkin. Teglar hujjatlarning (bob, qism, bo'lim, paragraf) umumiy tuzilishini belgilashda foydalaniladi. Hujjatlarni teglar orqali yozishda o'zgarmas ifodalar va operatorlar ishlatiladi: yoki -, takrorlash-* yoki +, ixtiyoriy - ? kabi.

Korpus annotatsiyasi uchun SGML tilda yozilgan til strukturasi namuna sifatida keltiramiz⁶⁰:

```

<ScientificBook>
<Book>
<Tit>
<Houn>Bases</NounXPrep>de</PrepXNoun>donnees</Noun>
<Aut>Claude Delobel</Aut>
<Aut>Michel Adiba</Aut>
<Chap>
<Tit>
<Art>Les</ArtXNoun>problemes</Noun>
<Par>
<Sent>
<Prep>dans</Prep>
<Art>un</Art>
<Noun>etat</Noun>
<Adj>coherent</Adj>
</Sent>
</Par>
<Sec>
<TitXNoun>Notes</NounXAdj>bibliographiques</AdjXTit>
<Par>
<Sent>
<Art>Le</ArtXNoun>probleme</NounXPrep>de</Prep>...

```

⁶⁰ John Nerbonne Linguistic Databases USA, 1998. - P. 40.

</Sent>
 </Par>
 </Sec>
 </Chap>
 </Book>
 </ScientificBook>

Ushbu tavsif SQL tilidan ko'ra OSQ (obyektga mo'ljallangan so'rov tili)da qulay sanaladi. SQL tilining *select...from...where* buyrug'i jadvallardagi ma'lumotlarning atributlarini (nom va muallifi)ni belgilashda ishlatiladi. Obyektga mo'ljallangan so'rov tili obyektlar guruhiga asoslanib, *from* buyrug'i barcha obyektlar to'plamini guruhlaydi.

4.3. Morfologik ma'lumotlar bazasi

Morfologik ma'lumotlar bazasi analiz va sintez protsedurasi uchun kerakli bo'ladigan barcha ma'lumotlarni o'z ichiga qamrab oladi.

Agar tanlangan morfologik modelda so'z shakllarning lug'ati qabul qilinsa, u holda muayyan leksemaga tegishli bo'lgan barcha so'z shakllarining grammatik belgisi hisobga olinadi. Agar morfologik modelda so'z asoslari lug'atini qamrab olsa, u holda ma'lumotlar bazasi har bir paradigmatic sinfga tegishli bo'lgan fleksiyaning (*turlanish yoki tuslanish natijasida so'zning oxirgi o'zgaradigan qismi*) ro'yxati lug'atda o'z aksini topadi. Fleksiyaning ma'lumotlari asosida so'zlarning grammatik tavsifi bilan bog'lanadi. Ma'lumotlar bazasida, shuningdek, omonimik leksemalar (to'liq va qisman) ham, albatta, bo'lishi kerak. Morfologik ma'lumotlar bazasida morfologik analiz va sintez protsedura jarayoni uchun zarur bo'lgan barcha ma'lumotlar tarkibiga kiritiladi.

So'z anglatyotgan ma'no va gapda anglashilayotgan mazmun tilning eng muhim tomonidir. So'zning quroli ham u anglatayotgan ma'noda aks etadi. Shu bois semantika tilshunoslikning eng qiziqtirgan va ayrim muammolarga sabab bo'lgan tarmog'i sanaladi. So'zlarning ma'nosiga psixologiya, falsafa va tilshunoslik kabi sohalar nuqtai nazaridan turlicha yondashiladi. Psixologlar uchun kishilarning ma'lumotlardan qay tarzda foydalanish, eslash, tasniflash kabi so'z anglatgan ma'no qiziqtirsa, falsafada ma'no

jihatdan mohiyatan qaysi tushunchalar o'zaro bog'langan yoki zidlangan, shuningdek, ma'no jihatdan nimani qabul qilamiz va anglaymiz kabi savollarga javob izlanadi. Tilshunoslikning barcha sathlari hamda semantika yo'nalishida tildan foydalanish va u bilan bog'liq jarayonlarni tushunish jarayoni birlamchi sanaladi.

Til nutqiy muloqot jarayonida psixologik, lingvistik va pragmativ va kognitiv munosabatlar ta'sirida voqelanadi. Ushbu jarayonda tafakkur va nutq munosabatida psixologik va idrokiy xotira ham ahamiyat kasb etadi. Tilning lisoniy modellari nutqda namoyon bo'lishida inson xotirasida mavjud bilim va tajriba birgalashib murakkab tizimni hosil qiladi.

Olimlar idrokiy ifoda va axborot kommunikatsiyasi kabi ikki asosiy faktorni farqlaydilar. Inson ongidagi idrokiy ifodani psixologlar⁶¹ asosiy kognitiv ifoda qilish tizimi sifatida quyidagicha ifodalaydi:

Kognitiv qatlam	Lingvistik ekvivalent
• doimiy semantik xotira	idrokiy lug'at
• ramziy xotira	muayyan matn (<i>muayyan holat, voqea, shaxs yoki tajribalarimiz aksi</i>)
• ishchi xotira va diqqat	muayyan nutqiy holat

Kompyuter lingvistikasida lingvistik bilimlar bazasini ham inson xotirasi kabi muayyan funksional modellardan tashkil topgan tuzilmadir.

Lingvistik ta'minotda fe'lli frazemalarning turg'un birikma sifatida semantik-mantiqiy imkoniyatlari, algoritmik tavsifi

Til qurilishining asosiy birliklari sifatida lison va nutq tushuniladi. Mana shu ikki asosning o'zaro birligidan tabiiy til elementlari shakllanadi. Til shunday hodisaki, u ichki va tashqi (jumladan, tarixiy, siyosiy, milli y, etnik, geografik, madaniy) omillarning ta'siri natijasida ijtimoiylashadi hamda o'ziga xos struktur tizimni hosil qiladi. Til va nutq birliklari masalasi ko'plab adabiyotlarda munozarali mavzu sifatida o'rganilgan.

⁶¹ Bern Heine, Heiko Narrog. The Oxford handbook of linguistic analysis. UK: Oxford university, 2015 –P. 45

Mashina tarjimasi uchun "tarjima birligi" masalasi muhim ahamiyat kasb etadi. Tarjimada ishtirok etayotgan til juftliklari, masalan, ispan tilidan arab tiliga tarjima qilinayotganda, tarjima birliklari faqat ulargagina tegishli bo'ladi. Yozma tarjimaning birligi morfema, so'z, so'z birikmasi, sintagma, jumla, satr boshi, xattoki, butun bir matnga ham to'g'ri kelishi mumkin. Mashina tarjimasida sintaktik qurilma matnning to'liq mazmunini ko'rsatmaydi. Shu jihatdan mashina tarjimasi tizimi oldidagi asosiy vazifalaridan biri so'z valentliklari hamda turg'un birikmalar korpusini yaratishdir. Bunda ma'lumotlar bazasi ma'lumotlar (informatsiyalar)ni saqlash uchun mo'ljallangan tashkiliy struktura bo'lib xizmat qiladi. Ma'lumotlar bazasi elektron saqlash va qayta ishlash vositalariga bog'liq bo'lishi shart emas. Bu o'zining muayyan strukturaga solingani, qulayligi, vaqt tejankorligi jihatidan ancha samaralidir. Tarjimon dasturi uchun lingvistik ma'lumotlar bazasi (ta'minoti) deyilganda, tilga oid barcha ma'lumotlar tushuniladi. Til tuzilishi hamda matnni analiz va sintez qilishning aniq vazifalariga asosan lingvistik ta'minotida quyidagilarning mavjud bo'lishi talab qilinadi: *matn – leksik asos – leksemashakl – so'z birikmasi – sintaktik gap strukturalari*. Tarjimadan avvalgi tahlil jarayonida uch asosiy iyerarxik bosqich amalga oshiriladi: *matn-jumla-leksema shakl*. Bunda leksemashakl keng ma'noda matndan olingan istalgan segment birlik sifatida qaraladi. Gap va so'z birikmalarining barcha strukturaviy hamda boshqa xususiyatlari leksema shakl tushunchasida ifodalanadi. Y.I.Shemakin "Kompyuter lingvistikasining boshlanishi" asarida avtomatik tarjimaning ikki modelini ko'rsatib o'tgan. Birinchi yondashuv "matn-mazmun-matn" modeli bo'lib, unda semantikaning formal tavsifi muhim o'rin tutadi. Ikkinchi yondashuv "matn-matn" modeli bo'lib, tarjima uning mazmuniy mosligi asosida amalga oshiriladi. Bizningcha, ikkinchi yondashuv formallik jihatdan talabga javob beradi.

O'zbek tili so'z turkumlarini semantik tahlil qilish quyidagi natijalarni voqealantirdi:

Qo'shma otning yasaliş modellari:

1. Ot+ot=>atirgul
2. Ot+fe'l=>dunyoqarash
3. Son+ot=>uchburchak

4. Son+fe'l=>beshotar
5. Ot+sifat=>ustabuzarmon
6. Fe'l+fe'l=>iskabtopar
7. Fe'l+ot=>savacho'p

Affiksatsiya usuli bilan ot yasalganda quyidagi jihatlar kuzatildi: asosi ot turkumidan iborat so'zlardan yasovchi affikslarning 38 turi, fe'l asosli so'zlardan 52 ta affiks, asosi sifatdan iborat bo'lgan 9 ko'rinishdagi ot yasovchi affiks, asosi taqlid so'zdan iborat 5 ko'rinishdagi ot yasovchi affiks turlari aniqlandi; prefiks miqdoran o'ta oz bo'lib, mavjudlari ham o'zlashgan so'zlarda uchraydi.

Qo'shma sifatning quyidagi modellari mavjud:

1. ot+ot=>bodomqovoq
2. sifat+ot=>qimmatbaho
3. ot+fe'l(+ar) =>xalqsevar
4. ravish+ fe'l(+ar) => tezyurar
5. ot+aro(umum, yarim, rang) => xalqaro

Sifat yasovchi affikslarning ja'mi 115 ko'rinishi, old qo'shimchalarning 8 ko'rinishi, ozaytirma daraja qo'shimchalarining 8, orttirma daraja hosil qiluvchi vositalarning 11, ozaytirma daraja hosil qiluvchi vositalarning 4 ko'rinishi borligi aniqlandi⁶².

Qo'shma fe'lning quyidagi modellari aniqlandi:

1. ot+fe'l=>bayon qilmoq
2. sifat+fe'l=>yaxshi ko'rmoq
3. fe'l+fe'l=>sotib olmoq
4. son+fe'l=>bir qilmoq
5. ravish+fe'l=>hozir qilmoq

O'zbek tilida fe'l yasovchi affikslarning 60, modallik ma'noni ifodalovchi 6 turi, fe'l nisbatlarini shakllantiruvchi qo'shimchalarining 26, fe'llarning inkor shaklini ifodalovchi 4, mayl qo'shimchalarining 30, zamon qo'shimchalarining 31, shaxs-son qo'shimchalarining 23 ko'rinishi mavjudligi kuzatildi.

Qo'shma ravishning quyidagi modellari aniqlandi:

1. Olmosh+ravish=>u yerga
2. Olmosh+ot=>har zamon
3. Alla+olmosh=>allaqachon

⁶² Qarang: Abduraxmonova N. Mashina tarjimasining lingvistik ta'minoti (monografiya), Toshkent: Muharrir, 2018. 165 b.

Mashina tarjimasi uchun “tarjima birligi” masalasi muhim ahamiyat kasb etadi. Tarjimada ishtirok etayotgan til juftliklari, masalan, ispan tilidan arab tiliga tarjima qilinayotganda, tarjima birliklari faqat ulargagina tegishli bo‘ladi. Yozma tarjimaning birligi morfema, so‘z, so‘z birikmasi, sintagma, jumla, satr boshi, xattoki, butun bir matnga ham to‘g‘ri kelishi mumkin. Mashina tarjimasida sintaktik qurilma matnning to‘liq mazmunini ko‘rsatmaydi. Shu jihatdan mashina tarjimasi tizimi oldidagi asosiy vazifalaridan biri so‘z valentliklari hamda turg‘un birikmalar korpusini yaratishdir. Bunda ma‘lumotlar bazasi ma‘lumotlar (informatsiyalar)ni saqlash uchun mo‘ljallangan tashkiliy struktura bo‘lib xizmat qiladi. Ma‘lumotlar bazasi elektron saqlash va qayta ishlash vositalariga bog‘liq bo‘lishi shart emas. Bu o‘zining muayyan strukturaga solingani, qulayligi, vaqt tejamlorligi jihatidan ancha samaralidir. Tarjimon dasturi uchun lingvistik ma‘lumotlar bazasi (ta‘minoti) deyilganda, tilga oid barcha ma‘lumotlar tushuniladi. Til tuzilishi hamda matnni analiz va sintez qilishning aniq vazifalariga asosan lingvistik ta‘minotida quyidagilarning mavjud bo‘lishi talab qilinadi: *matn – leksik asos – leksemashakl – so‘z birikmasi – sintaktik gap strukturalari*. Tarjimadan avvalgi tahlil jarayonida uch asosiy iyerarxik bosqich amalga oshiriladi: *matn-jumla-leksema shakl*. Bunda leksemashakl keng ma‘noda matndan olingan istalgan segment birlik sifatida qaraladi. Gap va so‘z birikmalarining barcha strukturaviy hamda boshqa xususiyatlari leksema shakl tushunchasida ifodalanadi. Y.I.Shemakin “Kompyuter lingvistikasining boshlanishi” asarida avtomatik tarjimaning ikki modelini ko‘rsatib o‘tgan. Birinchi yondashuv “matn-mazmun-matn” modeli bo‘lib, unda semantikaning formal tavsifi muhim o‘rin tutadi. Ikkinchi yondashuv “matn-matn” modeli bo‘lib, tarjima uning mazmuniy mosligi asosida amalga oshiriladi. Bizningcha, ikkinchi yondashuv formallik jihatdan talabga javob beradi.

O‘zbek tili so‘z turkumlarini semantik tahlil qilish quyidagi natijalarni voqealantirdi:

Qo‘shma otning yasaliş modellari:

1. Ot+ot=>atirgul
2. Ot+fe‘l=>dunyoqarash
3. Son+ot=>uchburchak

4. Son+fe‘l=>beshotar
5. Ot+sifat=>ustabuzarmon
6. Fe‘l+fe‘l=>iskabtopar
7. Fe‘l+ot=>savacho‘p

Affiksatsiya usuli bilan ot yasalganda quyidagi jihatlar kuzatildi: asosi ot turkumidan iborat so‘zlardan yasovchi affikslarning 38 turi, fe‘l asosli so‘zlardan 52 ta affiks, asosi sifatdan iborat bo‘lgan 9 ko‘rinishdagi ot yasovchi affiks, asosi taqlid so‘zdan iborat 5 ko‘rinishdagi ot yasovchi affiks turlari aniqlandi; prefiks miqdoran o‘ta oz bo‘lib, mavjudlari ham o‘zlashgan so‘zlarda uchraydi.

Qo‘shma sifatning quyidagi modellari mavjud:

1. ot+ot=>bodomqovoq
2. sifat+ot=>qimmatbaho
3. ot+fe‘l(+ar) =>xalqsevar
4. ravish+ fe‘l(+ar) => tezyurar
5. ot+aro(umum, yarim, rang) => xalqaro

Sifat yasovchi affikslarning ja‘mi 115 ko‘rinishi, old qo‘shimchalarning 8 ko‘rinishi, ozaytirma daraja qo‘shimchalarining 8, orttirma daraja hosil qiluvchi vositalarning 11, ozaytirma daraja hosil qiluvchi vositalarning 4 ko‘rinishi borligi aniqlandi⁶².

Qo‘shma fe‘lning quyidagi modellari aniqlandi:

1. ot+fe‘l=>bayon qilmoq
2. sifat+fe‘l=>yaxshi ko‘rmoq
3. fe‘l+fe‘l=>sotib olmoq
4. son+fe‘l=>bir qilmoq
5. ravish+fe‘l=>hozir qilmoq

O‘zbek tilida fe‘l yasovchi affikslarning 60, modallik ma‘noni ifodalovchi 6 turi, fe‘l nisbatlarini shakllantiruvchi qo‘shimchalarining 26, fe‘llarning inkor shaklini ifodalovchi 4, mayl qo‘shimchalarining 30, zamon qo‘shimchalarining 31, shaxs-son qo‘shimchalarining 23 ko‘rinishi mavjudligi kuzatildi.

Qo‘shma ravishning quyidagi modellari aniqlandi:

1. Olmosh+ravish=>u yerga
2. Olmosh+ot=>har zamon
3. Alla+olmosh=>allaqachon

⁶² Qarang: Abduraxmonova N. Mashina tarjimasining lingvistik ta‘minoti (monografiya), Toshkent: Muharrir, 2018. 165 b.

4. Alla-+ot=>allavaqt

Ravish yasovchi affikslarning umumiy 50 ko'rinishi, old qo'shimchalarning 2 turi aniqlandi:

Son so'z turkumi ma'noviy turlarini hosil qiluvchi affikslarning 10 ko'rinishi aniqlandi. O'zbek tilida kishilik, o'zlik, ko'rsatish, so'roq, belgilash, bo'lishsizlik, gumon olmoshlari aniqlandi. Ko'plik qo'shimchasi umumiy holda -lar ko'rinishda bo'ladi. O'zbek tilida 6 ta kelishik mavjud bo'lib, bosh kelishik qo'shimchasiz. Egalik qo'shimchalari birlik va ko'plik shakllarining 10 ta, kichraytirish, erkalash qo'shimchalarining 14, yuklama, ko'makchi va bog'lovchilarning turlari aniqlandi. Barcha yig'ilgan so'z yasovchi affikslar va grammatik ko'rsatkichlar bo'yicha natijalarga erishildi⁶³.

Tilshunoslikda turg'un birikmalar sifatida frazema va idiomalar nazarda tutiladi. Agar ingliz tilidagi frazali fe'llar yoki fe'lli frazemalarni tahlil qilsak, o'zbek tilidan farqli jihatlarini ko'rishimiz mumkin.

Ingliz tilida frazema terminiga har xil yondashuvlar uchraydi. Ayrim adabiyotlarda fe'l yoki ravish bilan keluvchi predloglarning birgalikda muayyan ma'noga tenglashishi⁶⁴, leksikologiyaga bag'ishlangan asarlarda predlog bilan keluvchi ot, sifat, fe'l hamda ravish so'z turkumlarining o'zaro yaxlitlanib yagona ma'noda kelishi tushuniladi⁶⁵. "Zamonaviy ingliz tili leksikologiyasi kursi"⁶⁶da frazeologik birliklar ikki guruhga ajratilgan: frazema va idiomalar. So'z birikmalarida qismlarining biri ikkinchisiga tobe munosabatda bo'ladi: *small hours, three hours, pleasant hours* kabi. Idiomalar tarkib toptiruvchi komponentlarning har biri yaxlitlanib yagona ma'no semasini beradi: *Mare's nest* (alohida tarjima qilinganda *a mare - baytal, nest-in- uya* ma'nosida), lekin bir butun idioma sifatida *mare's nest - aldov yoki to find a mare's-ne to find a mare's-nest* – "noo'rin (bemavrid) gapirib qo'ymoq"ni anglatadi.

⁶³ O'sha joyda.

⁶⁴ Лебедова Л.Д. Введение в курс лексикографии. – Москва: Высшая школа, 2008. – С. 170-176.

⁶⁵ Ginzburg R. S., Khidekel S. S., Knyazeva G. Y., Sankin A. A.. A Course in Modern English Lexicology. – M., 1979. – P. 95-96.

⁶⁶ Yuqoridagi asar. – B. 113.

Tilshunoslarning o'ziga xos yondashuvlari natijasida turg'un birikmalar (TB) turlicha nomlar bilan atalgan: *idioma, frazeologizm, frazema, turg'un birikma, frazeologik birliklar* va h.k. Ushbu terminlarning xususiyatlarini hisobga olib, "turg'un birikma" termini obyektini har tamonlama o'rganishda qo'l keladi, degan fikrni bildiramiz. Ingliz tilida TB lar murakkab tabiatlidir, chunki ikki yoki undan ortiq birliklarning o'zaro semantik, grammatik jihatdan birikuvi natijasida yagona birlik hosil bo'ladi. Ilmiy matnlarda TB larning qo'llanilishi muallifning shaxsiy uslubiga xosdir. Masalan, iqtisodchi M.Dobb va J.M.Keynsning 100 betlik ma'ruzasida 21 ta idioma ishlatilganligi qayd etiladi⁶⁷. Ilmiy matnlar boshqa uslublardan aniqligi, qolipga solinganligi, termin va undagi so'zlarning bir ma'noga egaligi bilan ajralib turadi. Ilmiy matnlarning o'ziga xos xususiyatlari to'g'risida bir qator ilmiy ishlar bajarilgan. Biroq ilmiy va texnikaviy matnlarni tarjima qilish borasidagi ba'zi mulohazalar ko'pincha bahsli munozaralarga sabab bo'ladi. Masalan, Eva Mastnáning "The nature of scientific / technical texts from viewpoint of translation studies"⁶⁸ tadqiqotida tarjimashunoslikda ilmiy matnlarning tabiati tahlil qilingan, shuningdek, ilmiy matnlarni tarjima qilishda mashina tarjimasidan foydalanish samaradorligi, iqtisodiy tejamkorligi bilan kelajakda muhim o'rin egalashini ta'kidlab o'tadi. Mashina tarjimasida muayyan uslubga oid bo'lgan lisoniy qoliplarni, shu jumladan, turg'un birikmalarni ajratib olish va ularning mazmuniy tarjimasini bilan maxsus lug'atlarga kiritish lozim. Quyida lingvistik modellar keltirilgan:

⁶⁷ Вершинина М.И. Идиоматическая фразеология в английском языке делового общения: лексикологический, лексикографический и функциональный аспекты: дисс. ...док. филол. наук. – М., 2000. – С. 24.

⁶⁸ Mastnán E. The nature of scientific/technical texts from viewpoint of translation studies. Bachelor degree diploma. 1984.

Phrases	[We+V V+us our +Noun]	[Verb+suffix] +preposition=> Adj	Adverb	Passive voice	Phrasal verbs
much	we are	to be	sleary	the	look for
less	coming to	attendant on	comple	voltage is	look into
limited	realize	to be	tely	applied	look down
almost	we have	conducive to	conside	the	look up
all of	taken it to be	to be	rably	magnetic	look
which	the tube has	destructive of	essenti	field is	forward
very	shown us	to be	ally	set up	
effecti	we are	incidental to	fairly	the line is	
ve	beginning to	to be	greatly	terminated	
much	see	responsive to	signific	the	
the	we deal with	to be tolerant	antly	switch is	
same	we are now	of		closed	
	speaking				

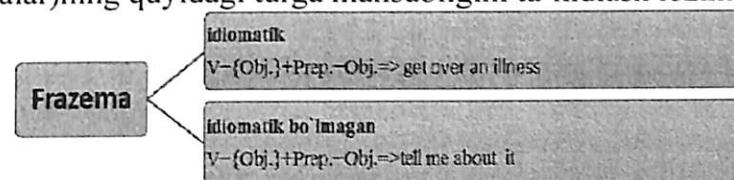
Frazemaga oid soʻz guruhlarining ot (nominal), feʼl (verbal) hamda sifat (adjectival) birikish turlari boʻlib, tuzilish jihatdan ular ikki asosli, uch asosli, toʻrt asosli, besh asosli boʻlishi mumkin: I. Ot fraza: 1) [sifat+ot]=> *foul play*; 2) [feʼl+ot]=> *have a last word*; 3) [ot+ot]=> *body and soul*; II. Ravish fraza: 1) [predlog+ravish]=> *for ever*; 2) [ravish+ravish]=> *now and again*; III. Sifat fraza: 1) [sifat+sifat]=> *alive and kicking*; IV. Feʼl fraza: 1) [feʼl+feʼl]=> *wait and see*; 2) [feʼl+predlog]=> *look through*.

Turgʻun birikma terminiga oid bir necha qarashlar mavjud. Baʼzi adabiyotlarda verbal (feʼllar) iboralar (verbal expression)⁶⁹, boshqalarda esa iborali feʼllar (phrasal verbs), shuningdek, idiomalar (idioms) tarzda talqin etilganlarni kuzatishimiz mumkin. Iborali feʼllar quyidagi turlarga ajratiladi⁷⁰: I. Verb +preposition (transitive) => *Listen to this record Listen!*; II. Verb+particle (transitive) => *Take off your hat. Take your hat off.*; III. Verb+particle (intransitive) => *Sit down!*; IV. Verb+particle +preposition (transitive) => *We've run out of matches. Feʼl+predlog* shaklidagi frazemalarni mazmuniy jihatdan uch guruhga ajratish oʻrinli: 1) asos

⁶⁹ Richard A. Spears Ph.D. Mc. Graw Hill's dictionary of American Idioms and Phrasal verbs. 2005.

⁷⁰ Alexander L.G. Longman English Grammar Practice. (for intermediate students). 1998. - P.123.

feʼl maʼnosi oʻzgaras frazemalar (hear from); 2) asos feʼl maʼnosi oʻzgaruvchi frazemalar (look for, look after); 3) bir asosli feʼllarning predloglar almashinuvi natijasida maʼno nozikligi bilan farqlanuvchi frazemalar (*hear of, hear about*). Miqdoriy tarkibiga koʻra frazemalarning quyidagi turlari borligi aniqlandi: 1) feʼl+predlog yoki ravish=>ikki tarkibli frazema; 2) feʼl+ravish+predlog=>uch tarkibli frazema; 3) feʼl+predlog+otli fraza (Verb + particle + noun phrase); 4) feʼl+ otli fraza + predlog (Verb+ noun phrase+particle); 5) feʼl+ravish+predlog+otli fraza (Verb+adverb+ preposition+noun phrase). Yuqoridagilarni umumlashtirgan holda frazemalar (feʼl frazalar)ning quyidagi turga mansubligini taʼkidlash lozim:



Makoto Nagao konsepsiyasiga qadar professor G.G.Belonogov 1975-yilda frazeologik mashina tarjima konsepsiyasini taklif qilgan. Unga koʻra tarjima qilinayotgan matndagi soʻzlar alohida emas, balki frazeologik birikmalar, ifodalanuvchi tushunchalar va turli vaziyatlar oʻrtasida munosabatlar tarzida tahlil etilishi kerak. Buning uchun frazeologik mashina tarjima tizimida frazemalar, frazeologik birikmalar hamda alohida soʻzlar uchun tarjima ekvivalentliklaridan iborat boʻlgan bilimlar bazasi yaratiladi. Aynan mana shu konsepsiya boʻyicha yana bir qator olimlarning, xususan, A.A. Xoroshilovning “Теоретические основы и методы построения систем фразеологического машинного перевода”, V.A. Chijakovskiyning “Фразеология и машинный перевод (опыт составления работы немецко-русского автоматического словаря для публицистических и научных текстов” mavzusida nomzodlik dissertatsiyasi yoqlandi. Frazeologik birikmalarning toʻgʻri shaklda berilishi matn mazmunini tushinishda aniqligini oshiradi. Ingliz tilidagi *take* soʻzi orqali inglizcha-oʻzbekcha mashina tarjima uchun turgʻun birikmalarning semantik bazasini tuzish boʻyicha ayrim tahlillarga toʻxtalamiz. Lingvistik bazaga quyidagi modellar boʻyicha kiritish mumkin. (Soʻzlarning konnotativ maʼnolari lugʻatda kodlangan boʻladi.)

I. [Take+Noun]=> 1) o'lja; 2) tushum, kirim, yig'im; 3) maosh, oylik; 4) yutuq; 5) ijara (er); 6) ijaraga olingan yer; 7) pora.

II. [Take+Verb]=> 1) olmoq; 2) qabul qilmoq; 3) undirmoq; 4) olib qo'yimoq; 5) ijaraga olmoq; 6) vazifani ado etishga kirishmoq; 7) olmoq; egallamoq (vaqt, joy); 8) qabul qilmoq; 9) yuvintirmoq; 10) to'lamoq

III. [Take+Preposition (away/for/back/in /off /on /out /over /through /up)]=>

IV. take+away=>olib tashlamoq, olib ketmoq

V. [Take+Preposition +Noun]=>take on credit- kredit olmoq

VI. [Take+Preposition +Pronoun]=>take with you- o'zi bilan birga olmoq

VII. [Take+ Adjective+Noun]=>take great care of- ...haqida g'amxo'rlik qilmoq

VIII. [Take+ Adjective(more)+than]=>take more than- ...nafaqat talab qilmoq, balki...

IX. [Take+Noun]=>take account of- hisobga olib chiqmoq, ro'yxatga olmoq, hisoblab chiqmoq

X. [Take+a/an/the +Noun]=>take a photo- rasmga olmoq

XI. [Take +a/an/the+ Noun+Preposition]=>take the place of- o'zgartirmoq; o'zgarib qolmoq; almashtirmoq; yangilamoq; o'miga qo'yimoq

XII. [Take+ Noun+Preposition]=>take care of-qaramoq (economics)

a. Tocomplete care of- g'amxo'rlik qilmoq

XIII. [Take+ to+Noun]=>take to pieces- ajratib tashlamoq (qismlarga)

XIV. [Take+Adverb]=>take down- demontaj qilmoq, qismlarga ajratmoq take outside- ...ga olib chiqmoq

XV. [Take+ Adverb+Preposition]=> take down in the minute- protokolga kiritmoq

XVI. [Take+ Adverb+ a/an/the +Noun]=>take out a lisenca - litsenziya olmoq

XVII. [Take+ Adverb+ Preposition+Noun]=> take out of production- mahsulot sotib olmoq

XVIII. [Take+as +Noun]=>take as a basis - biror narsaning asosi deb bilmoq

XIX. [Take+Verb]=>take leave of-(bilan) - xayrlashmoq

Yuqoridagi namunalardan kelib chiqib shuni aytish mumkinki, *take* fe'lining o'n sakkizta modeli bo'lib, ular turli semalarni ifodalaydi. Biroq keltirilgan modellarda ifodalanayotgan leksema ham denotativ, ham konnotativ ma'no kasb etadi. So'zlarning birikish usullari modellarini berish hamda turg'un birikmalarining lingvistik tabiatini o'rganish mashina tarjima tizimi sifatini yanada oshirishga xizmat qiladi.

Ingliz tilidan o'zbek tiliga tarjima qiluvchi dastur uchun turg'un iboralar, ayniqsa, fe'lli iboralarni chuqurroq tahlil qilish zarur. Unga ko'ra tarkibida fe'l mavjud bo'lgan birikmalar modellari aniqlangandan so'ng ularning semantik bazasi tuzildi. Fe'lli frazemalar ingliz tilidagi eng murakkab tarjima birligi bo'lganligi bois u gapning struktural tuzilishi va semasiga bevosita ta'sir ko'rsatadi.

Manbalarda fe'lli frazemalarga quyidagicha ta'rif keltiriladi: "Fe'lli frazemalar ingliz tilida ko'p qo'llaniladigan muhim kategoriyadir. Avvalo, fe'lli frazemalar kundalik iste'molda tez-tez qo'llaniladi, shuningdek, u boshqa til vakillari uchun juda murakkabdir. Chunki uni grammatik va semantik jihatdan to'g'ri qo'llashda qiyinchiliklarni yuzaga keladi. Ikkinchidan, fe'lli frazemalar tilning boyligi hamdir. Ular orqali kishilarning faoliyati va o'zaro munosabati aks etadi"⁷¹.

Inglizcha matnlarni o'zbek tiliga tarjima qiluvchi dasturiy ta'minot uchun matn birliklarini kompyuter tarjima qilishda ana shunday TBlarning ekvivalentligini to'g'ri tanlash, lingvistik algoritmini tuzish, qolaversa, quyidagi masalalarni hal etish muhim sanaladi: 1) har bir birlikning lingvistik modeli mukammal holda kiritilgan bo'lishi; 2) muayyan leksik birlikning qiymati so'z birikmasi, termin, turg'un birikma, qo'shma so'z tarzida alohida lug'atlarda aks etishi; 3) dasturiy ta'minot uchun o'zbek tilidagi har bir birlikning ekvivalenti tavsifiy xarakterga ega bo'lishi kerak. Mashina tarjimasida bilimlar bazasi hamda lingvistik ta'minot tarjima faoliyatini yengillashtirish bilan birga matnda anglashilgan umumiy konseptni aniqlashtiradi.

⁷¹ Andreea-Rosalia Olteanu. A holistic approach to phrasal verbs, Editura Sfântul Ierarh Nicolae 2012. - P. 16.

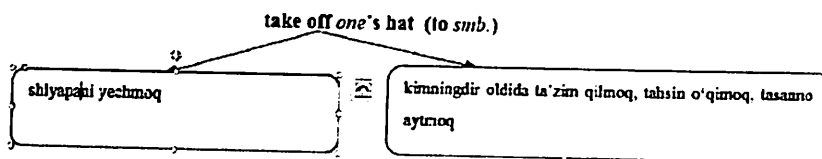
Fe'lli frazemalar (FF) o'timli va o'timsiz (*transitive* va *intransitive verbs*) turlarga bo'linadi⁷²: 1) Verb+preposition (transitive) => *Listen to this record. Listen!*; 2) Verb+particle (transitive) => *Take off your hat. Take your hat off.*; 3) Verb+particle (intransitive) => *Sit down!*; 4) Verb+particle+preposition (transitive) => *We've run out of matches.*

Ingliz tilida predloglarning asosiy vazifasi gap bo'laklarini sintaktik munosabatga kirishiga yordam berishdir. Ular fe'lli frazemaning o'zgarish qismi (particle) deb ataladi. Ba'zi o'rinlarda ular o'zi anglatyotgan ma'noni saqlab qolsa, ba'zan semantik yahlitlanib, bir butun leksemani hosil qiladi. Frazemalarni ma'no ko'chishning o'z va ko'chma ma'no guruhlariga ko'ra ajratish mumkin: a) *drop off* (to 'kilmog- o'z ma'no); b) *drop off* (*mudramog, ko'zi ilinmog* - ko'chma ma'no). Ularni quyidagi namunalarda ham uchratish mumkin: 1. *I think, you have picked up a flu* (*kasal bo'lmoq*); 2. *I picked up Spain when I was in it* (*o'rganib olmoq, bilib olmoq*); 3. *I was picked up for speeding my car* (*so'roqqa tutilmoq*); 4. *I am picked up at 6 p.m.* (*olib ketmoq*); 5. *Fortunately, economy is starting pick up* (*yaxshilanmoq, o'smoq, yuksalmoq*). Fe'lli frazemalar shunday nozik xususiyatga egaki, o'zgaruvchi qismning tartibini o'zgartirish yoki boshqa u bilan turdosh bo'lmagan leksemaning kelishi ma'noning umumiy yahlitligini o'zgartirishi mumkin: *blow off* -1) *uchirib ketmoq, esmoq, esib turmoq*; 2) *g'azablanmoq, jahli chiqmoq*; 3) *vaqtni bekor sarflamoq, blow smb. off* - e'tiborsiz qoldirmoq, e'tiborga olmaslik, *nazar-pisand qilmaslik; blow smth. off* - *kutilayotgan joyga bormaslik yoki qatnashmaslik, biror kimsa bilan uchrashuvni qoldirmoq*. Demak, bu kabi frazemalarning tarkibiy qismlari ajralish yoki ajralmasligi mumkin: 1) ajraladigan (*seperable*) - *shut down* - *shut it down*; 2) ajralmaydigan (*inseperable* - *go in for*). Shu bois ayrim o'rinlarda fe'lli frazeologik birliklar bilan gumon olmoshlarini (GO) (*smth., smb., somewehere*) birgalikda berish maqsadga muvofiq. Bu esa u yoki bu frazemalarni ajralish yoki ajralmasligini ko'rsatishga xizmat qilsa, boshqa bir tomondan, fe'l nisbatlarining ham o'zgarishiga sabab bo'ladi: *wake up* - *uyg'onmoq; wake smb. up* - *uyg'otmoq*.

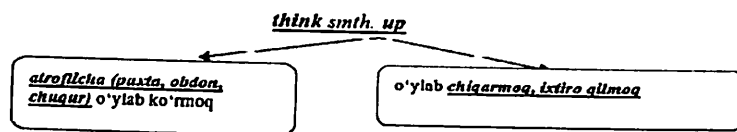
⁷² Alexander L.G. Longman English Grammar Practice. (for intermediate students) 1998. - P. 116-123.

Yana bir jihati gumon olmoshlari (GO) fe'lli frazalar tarkibida bo'lsa, ma'noga ta'sir etmasligi yoki umumiy semani o'zgartirib yuborishi mumkin: *I invited him to come to our party but he turned me down. - Men uni kechamizga taklif etdim, lekin u meni rad etdi; I hope my project will not be turned down. - Umid qilamanki, mening loyiham rad etilmaydi*. Yana bir misolda esa GOlarini o'zgartirish bilan ma'noga ta'sir etganini ko'rish mumkin: *perk smb. up* - *ruhlantirmoq, dalda bermog, ko'nglini ko'tarmog; perk smth. up* - 1) *tozalamoq, yangilamoq, musaffo qilmoq*; 2) (*iqtisodni*) *jonlantirmog, ko'tarilishiga sabab bo'lmoq* (*narxni*); *perk up* -1) (*boshini*) *ko'tarmog, faollashmog*; 2) (*narh-navo haqida*) *ko'tarilmog, o'smog*; 3) *jonlanmog, ruhlanmog, jon kirmog*; 4) (*kasallikdan keyin*) *sog'aymog, tuzalmog, sog'ayib ketmog*. GO lari o'rnida jonli yoki jonsiz ot leksemlarning qo'llanilishi fe'lli frazemalarda anglashilayotgan ma'no umumiy bo'lsa ham, predloglar bilan farqlanishi mumkin: *agree with smb.* - *biror kishi bilan kelishib olmog*; *to agree to* - *biror narsaga kelishmog, qo'shilmog*; *to be angry with smb., to be angry at smth.* - *jahli chiqmog*; *to be sorry for smb., to be sorry about smth.* - *kechirim so'ramog*. Ingliz tilidagi FF o'ziga xos yana bir jihati o'zbek tiliga tarjima qilishda ko'zga tashlanadi. U yoki bu frazemaning ma'nosi bir leksemaga yoki so'z birikmasiga, hatto, ba'zi o'rinlarda iboralarga teng kelishi mumkin. Masalan, *want smb. or smth. out of smth.* - *kimningdir ketishini yoki chiqishini xohlamoq* (so'z birikmasi) yoki *want for smth.* - *ehtiyoj sezmoq, muhtoj* (*hojatmand*) *bo'lmoq, talab qilmoq* (qo'shma fe'l), *want for* - *xohlamoq, istamoq* (sodda fe'l) tarzida.

FFlar faqat fe'l+predlog tarzida baholash to'g'ri bo'lmaydi, chunki fe'ldan keyin keluvchi predlog umuman fe'lga aloqador bo'lmashligi ham mumkin, ya'ni predlog o'zidan keyingi birlikning tarkibi bo'ladi: *I came from Tashkent* - *Men Toshkentdan keldim*. Fe'lli frazemalar fe'l+leksema {predlog} tarzida ham nomoyon bo'ladi. Bunday birlikning murakkab tabiati shunda ko'rinadiki, kontekstdagi holatiga qarab ularni erkin birikma yoki barqaror birikma tarzida ham tarjima qilish mumkin.



Ba'zi FFlar o'zbek tiliga tarjima qilinayotganda ularga qo'shimcha leksemalar qo'shilish holatlari ham kuzatiladi:



Biz ingliz tilida mavjud 12 mingta fe'lli frazemalarni yig'ib, ularning o'zbekcha tarjimasining muqobil variantini ma'lumotlar bazasiga kiritdik. Manbalarda FFlarning lingvistik modellari quyidagicha ifodalangan⁷³: 1) V+ADV=>calm down; 2) V + PREP. => keep on; 3) V+PREP+PREP=> put up with; 4) V+ADV+PREP=> change out of smth. (kiyimlarni yechib almashtirmoq); 5) V+N/PRON+ADV =>call smth. down (to smb.); 6) V+ADV+pro/N =>pull down smth.; 7) V+PRON+ADV=> call smth. down (to smb.); 8) V+ N/PRON + PREP=>give something away; 9) V+a / an+Noun=>make a noise; 10) V+a / an+V=>have a smoke

FFlarning yana bir xarakterli tomoni ayrimlari o'tgan zamon shaklida bo'lib, sifat yoki sifatdoshga aylanadi: bent on doing smth.; bereft of smb. or smth. -1) ayrilgan, judo bo'lgan; 2) bo'sh, quruqdan-quruq, yuksiz. Shu bois bunday tuzilishdagi birliklar ham lug'atda o'rin olgan bo'lishi lozim. Fe'lli frazemalarning ba'zilarida o'zgaruvchi qismlar (fe'llar) hamda o'zgarmas qismlar (predloglar) o'rtasidagi sinonimiya ham kuzatiladi: have at smb.-go at smb. (tashlanmoq, hamla qilmoq, hujum qilmoq); turn off -turn out (o'chirmoq). Shuningdek, fe'lli frazemalardan hosil bo'lgan otlar ham mavjud bo'lib, ular yahlit holga kelib qolgan. Bu kabi otlar ikki tuzilishga ega: 1) predlog+fe'l (offspring) -a) zot zurriyot; b) mahsulot, natija; 2) fe'l+predlog{ravish} (takeover) -a) o'lja; b) davlat to'ntarilishi. FFlarning modellari quyidagilar:

- V + oneself +P+smth. => align oneself with smb. or smth.
- V+ oneself +P=> arch (oneself) over

- V + oneself +P+smb.=> attach oneself to smb
- V + P+ smb. or smth.+P smth. =>arrange with smb. about smth.
- V+ smb. or smth. + P+ smb. or smth. => associate smb. or smth. with smb. or smth.
- V+ smth. +P+ smth.=> balance smth. against smth.
- V+ smth. +P=> bail smth. out
- V+ smb. +P+ smth.=> astound smb. with smth.
- V+smb. +P+ smb. or smth.=> bias smb. against smb. or smth.
- V+smb. +P=> beat smb. up
- V +P+smb.=> bet with smb.
- V +P+smb. or smth.=> attend to smb. or smth.
- V +P+ smth.+P+ smth.=>
- V +P+ smb.+P+ smb.or smth.=> book smb. through (to some place)
- V+P+P=> be in for
- V+P+P+it=> be in for it
- V+P+P+smth.=> be off for smth.
- V+P+P+ smb.or smth.=> bound up with smb. or smth.
- V+P+ P+ smth=>bear up (against smth.)

Kompyuterda tahlil qilinayotgan matndagi frazemalarni leksik analiz qilishning bir nechta texnologiyalari mavjud: qoidaga asoslangan, statistikaga asoslangan va har ikkisinin yig'indisi⁷⁴. Bu borada bizning algoritmimiz gibrid texnologiya hisoblanadi. Leksemani qidirish algoritmiga ko'ra joriy leksema leksemalar jadvaliga (LJ) joylanadi va har bir leksemaning modeli mavjud. LJ da leksemaning tartib raqami, leksema nomi, kodi, turi va tarjimasini beriladi. Joriy leksema (L1) o'qilgandan so'ng, bunday leksemaning mavjudligi yoki mavjud emasligini, agar mavjud bo'lsa, uning qaysi jadvalda joylashgani, kodi qandayligi aniqlanadi. L1 V (verb) jadvalida uchrasa, navbatda leksema (L2) o'qiladi. L2 leksema proN, Prep, Adv jadvallaridan qidiriladi, agar uchrasa, L1+L2 ulanadi, keyingi L3 leksema o'qiladi va keyingi qidiruv ishlari davom ettiriladi. Bu holat L1+...+Ln leksemalar ketma-ketligi

⁷⁴ GULILA ALTENBEK Identification Of The Kazakh basic Phrases Based On The Maximum Entropy Model / Turklang.2013, Astana, 2013. -C. 48

⁷³ Oxford Phrasal Verbs Dictionary for learners of English. Oxford university press. 2002.

turkumlari bilan bog'liq barqaror birikmalarning o'zbek tiliga tarjimasini bo'yicha ishlar amalga oshirilishi nazarda tutilgan.

Adabiyotlar:

1. E.Begmatov, A.Mirvaliyev O'zbek tilining imlo lug'ati Toshkent: Akademiya, 2013.
2. O'zbekiston Respublikasi Vazirlar Mahkamasining "O'zbek tilining asosiy imlo qoidalarini tasdiqlash" haqidagi 1995-yil 24 avgustdagi 339- qarori
3. John Nerbonne Linguistic Databases USA, 1998. 255 p.
4. Abduraxmonova N. Mashina tarjimasining lingvistik ta'minoti (monografiya), Toshkent: Muharrir, 2018. 165 b.

Savol va topshiriqlar:

1. Dunyo tillarining morfologiyasi bo'yicha tasnifiy jadval tuzing va umumiy va farqli jihatlarini matritsa ko'rinishida ifodalang?
2. Kompyuter morfologiyasi nima?
3. Kompyuter morfologiyasi bo'yicha qanday tadqiqotlar amalga oshirilgan?
4. Lingvistik bilimlar bazasi deganda nimani tushunasiz?
5. Olimlar inson ongidagi idrokiy ifodani qanday faktorlarga ajratadi?
6. Muayyan matn kognitiv qatlamning qaysi xotirasida mavjud bo'ladi?
7. Kompyuter lingvistikasida turg'un birikmalar deganda nimalar nazarda tutiladi?
8. Frazologik mashina tarjimasini konsepsiyasiga ko'ra frazemalar qanday tahlil qilinishi kerak?
9. *Take great care of* birikmasini mashina tarjimasini bo'yicha tahlil qiling.
10. Nima uchun fonologik modellar matn bilan bog'liq masalalarda dastlabki qadam hisoblanadi?

4-mavzu yuzasidan test

1. Quyidagilardan qaysi biri lingvistik ma'lumotlar bazasining asosiy yarusiga kirmaydi?
 - a) korpus sathi
 - b) morfologik sath
 - c) fonetik-akustik sath
 - d) semantik-leksik sath
2. Mashina tarjimasini tizimi uchun ma'lumotlar bilan ta'minlashning 5 ta turi mavjud bular qaysilar?
 - a) lingvistik, matematik, dasturiy, informatsion, mantiqiy ta'minot
 - b) morfologik, sintaktik, lingvistik, mantiqiy, dasturiy ta'minot
 - c) lingvistik, mantiqiy, dasturiy, morfologik ta'minot
 - d) mantiqiy, sintaktik, morfologik, algoritmik, dasturiy ta'minot
3. Muayyan dasturiy ta'minot uchun mo'ljallangan, uni saqlash, foydalanish, himoyalash va qayta ishlash ishlash imkoniyatiga ega bo'lgan elektron majmua nima deb ataladi?
 - a) lingvistik ma'lumotlar bazasi
 - b) lingvistik protsedura
 - c) morfologik tahlil
 - d) sintaktik struktura
4. Omograf nima?
 - a) yozilishi o'xshash, biroq urg'u va talaffuziga ko'ra farqlanuvchi so'zlarga aytiladi
 - b) talaffuzi o'xshash, yozilishida farqlanuvchi so'zlarga aytiladi
 - c) ayrim leksemalarning faqat u yoki bu grammatik shakllaridagina teng kelishi
 - d) aytilishi va yozilishi har xil ma'nosi bir xil bo'lgan so'zlar
5. Omofon nima?
 - a) ayrim leksemalarning faqat u yoki bu grammatik shakllaridagina teng kelishi
 - b) aytilishi va yozilishi har xil ma'nosi bir xil bo'lgan so'zlar
 - c) talaffuzi o'xshash, yozilishida farqlanuvchi so'zlar
 - d) yozilishi o'xshash, biroq urg'u va talaffuziga ko'ra farqlanuvchi so'zlar
6. Ingliz tilida predloglarning asosiy vazifasi nima?

a) gap bo'laklarini sintaktik munosabatga kirishiga yordam berishdir

b) otni tobelantirish

c) harakat-nomini shakllantirish

d) qo'shma gapni bir-biriga bog'lash

7. "Теоретические основы и методы построения систем фразеологического машинного перевода" nomzodlik dissertatsiya muallifi kim?

a) A.A. Xoroshilov

b) Chijakovskiy

c) Mitkov

d) Chomskiy

8. "Фразеология и машинный перевод (опыт составления работы немецко-русского автоматического словаря для публицистических и научных текстов" mavzusida nomzodlik dissertatsiyasi muallifi kim?

a) Chijakovskiy

b) Xoroshilov

c) Flin

d) Shennon

9. Qaysi tillarida matnni ajratishga xizmat qiluvchi tinish belgilari ishlatilmaydi?

a) xitoy, koreys va yapon

b) rus, ingliz, ispan

c) ispan, ingliz, italyan

d) hind, arab, ingliz

10. Omoforma nima?

a) ayrim leksemalarning faqat u yoki bu grammatik shakllaridagina teng kelishi

b) talaffuzi o'xshash, yozilishida farqlanuvchi so'zlarga aytiladi

c) yozilishi o'xshash, biroq urg'u va talaffuziga ko'ra farqlanuvchi so'zlarga aytiladi

d) aytilishi va yozilishi har xil ma'nosi bir xil bo'lgan so'zlar

3-MODUL. GRAMMATIK TAHLIL QILISHNING KOMPYUTERLI YONDASHUVI

5-Mavzu. Morfologik tahlilning avtomatik tizimi

5.1. O'zbek tilining morfologik qonuniyatlari

5.2. Tabiiy tilning morfologik modeli

5.3. Morfologik analiz

Tayanch so'z va iboralar: morfologik model, morfotaktik holati, morfotaktik qoidalar, morfologik analiz, morfologik modellashirish, fonologik component, tokenizatsiya, lemmatizatsiya, stemming, tillar tasnifi, inflektiv tillar, agglyutinativ tillar, isolayativ tillar, semantika, giponim, geperonim, FST texnologiyasi

5.1. O'zbek tilining morfologik qonuniyatlari

Morfologiya – grammatikaning bir qismi bo'lib, so'zning grammatik ma'nosi, shakliy (morfologik) va vazifaviy (sintaktik) xususiyatiga ko'ra guruhlanishi – turkumga ajratilishi, grammatik o'zgarishi, grammatik qo'shimcha haqidagi ta'limot.

Sintaksis ham grammatikaning bir qismi bo'lib, so'zning o'zaro erkin bog'lanishi, erkin bog'lanish natijasida hosil bo'lgan so'z birikmasi va gap, uning turi va xususiyatini o'rganadi.

Morfologik komponent—tabiiy tildagi leksemani morfologik analiz va sintezni amalga oshiruvchi dasturiy majmua.

Morfema — so'zning eng kichik ma'noli qismi (o'zak, affiks).

Asos — so'zning yasovchi qo'shimchasiz shakli.

Leksema — so'z, lug'at birligi sanaladi.

So'z shakli — muayyan grammatik shakldagi so'z.

Morfologik paradigma —bir so'zning grammatik kategoriyalar tizimi.

O'zbek tilining umumiy morfologik qonuniyatlariga quyidagilar kiradi:

- So'z turkumlarining mustaqil, yordamchi va alohida guruhni tashkil qiluvchilar kabi guruhlariga bo'linishi hamda har bir guruhning alohida grammatik xoslanishi;

- Morfemalarning morfotaktik qoidalarga bo'ysunishi va istisnosli holatlarning kuzatilishi (fonetik hodisalar, suffikslarning unli yoki undosh fonemalarga ko'ra xoslanganligi);

- Grammatik qo'shimchalar va so'z yasovchi qo'shimchalarning ayrim holatlarda omonimiyani hosil qilishi;

Tillar tasnifi

Inflektiv tillar (hind-yevropa tillari) – biror belgisiga ko'ra farqlanuvchi tillar.

Agglyutinativ tillar (uyg'ur-finland va turkiy tillar) – so'zlar morfemalardan iborat bo'lib, asos va affikslardan iborat bo'ladi.

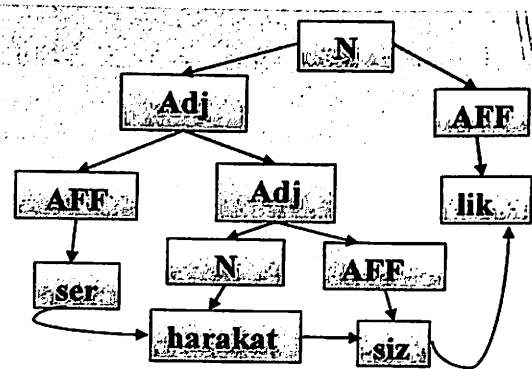
Izolyativ tillar (mandarin xitoy tili)- so'zlar biror chegara va affiksga ega bo'lmaydi.

Asosga qo'shiladigan qo'shimchalar ikki xil bo'ladi:

Infleksional kategoriya –so'zlarning grammatik ma'nosini o'zgartirib, so'zning leksik ma'nosiga ta'sir qilmaydi: *bolaga*->*bola+ga*=> ot+kelishik; *o'qidim*-> *o'qi + di + m* => fe'l+zamon+shaxs-son

Derivativ kategoriya – yangi so'z yasashga xizmat qiladi: *bolalik*->*bola+lik*=> *ot+-lik (ot)*, *bolali*->*bola+li*=>*ot+li (sifat)*

Masalan, *serharakatsizlik* so'zining grafik chizmasi quyidagicha bo'ladi (6-chizma):



6-chizma

Til qurilishining asosini lisoniy modellar va nutqiy birliklar tashkil qiladi. Mana shu ikki asosning o'zaro birligidan tabiiy tilning manzarasi hosil bo'ladi.

Til shunday hodisaki, u ichki va tashqi (jumladan, tarixiy, siyosiy, milliy, etnik, geografik, madaniy) omillarning ta'siri natijasida ijtimoiylashadi hamda o'ziga xos struktur tizimni hosil qiladi. Til va nutq birliklari mavzusi ko'plab adabiyotlarda munozaraga sabab bo'lgan⁷⁵. Mashina tarjimasi uchun "tarjima birligi" masalasi alohida ahamiyat kasb etadi. Tarjimada ishtirok etayotgan til juftliklari, masalan, ispan tilidan arab tiliga tarjima qilinayotganda, tarjima birliklari faqat ulargagina (tarjima jarayonidagi ikki tilga) tegishli bo'ladi. Yozma tarjimaning birligi morfema, so'z, so'z birikmasi, sintagma, jumla, satr boshi, hattoki, butun bir matnga ham to'g'ri kelishi mumkin. Mashina tarjimasida sintaktik qurilma matnning to'liq mazmunini ko'rsatmaydi. Shu jihatdan mashina tarjimasi tizimi oldidagi asosiy vazifalaridan biri so'z valentliklari hamda turg'un birikmalarning lingvistik bazasini yaratishdir.

Ma'lumotlar bazasi elektron ma'lumotlarni saqlash va qayta ishlash vositalariga bog'liq bo'lishi shart emas⁷⁶. Bu o'zining muayyan strukturaga solingani, qulayligi, vaqt tejamlorligi jihatidan ancha samaralidir. Tarjimon dasturi uchun lingvistik ma'lumotlar bazasi (ta'minoti) deyilganda, tilga oid barcha ma'lumotlar tushuniladi. Til tuzilishi hamda matnni analiz va sintez qilishning aniq vazifalariga asosan lingvistik ta'minotida quyidagilarning mavjud bo'lishi talab qilinadi: *matn – leksik asos – leksemashakl – so'z birikmasi – sintaktik gap strukturalari*. Tarjimadan avvalgi tahlil jarayonida uch asosiy iyerarxik bosqich amalga oshiriladi: *matn-jumla-leksema shakl*. Bunda leksemashakl keng ma'noda matndan olingan istalgan segment birlik sifatida qaraladi. Gap va so'z birikmalarining barcha strukturasi hamda boshqa xususiyatlari leksema shakl tushunchasida ifodalanadi. Y.I.Shemakin «Kompyuter lingvistikasining boshlanishi» asarida avtomatik tarjimaning ikki modelini ko'rsatib o'tgan. Birinchi yondashuv "matn-mazmun-matn" modeli bo'lib, unda semantikaning formal tavsifi muhim o'rin

⁷⁵ Qarang: Реформатский А.А. Введение в языковедение. – М., 1967., Смирницкий А.И. Лексикология английского языка. – М., 1956; Кодухов В.И. Введение в языковедение. – М., 1979., Неъматов Х., Бозоров О. Тил ва нутқ. –Т.: Ўқитувчи, 1993., Рахматуллаев Ш. Систем тилшунослик асослари. –Т.: Университет, 2007.

⁷⁶ Сатторов А. Маълумотлар базасини бошқариш системаси. –Т., 2006. – Б. 8.

tutadi. Ikkinchi yondashuv "matn-matn" modeli bo'lib, tarjima uning mazmuniy mosligi asosida amalga oshiriladi⁷⁷. Bizningcha, ikkinchi yondashuv formallik jihatdan talabga javob beradi.

Kompyuter morfologiyasi⁷⁸ so'z va uning shakllarining ham grafematik (yozma) ham fonemik (og'zaki) shaklini tahlil qilish bilan shug'ullanadi. Matnlarni tahrirlovchi dastur yoki nutqiy akti tekshiruvchi ilovalari ham mana shu yo'nalish doirasida o'rganiladi. So'z yasash, shakl va funksiyasiga ko'ra kontekstda qo'llash inson idrokida amalga oshiriladigan psixologik jarayon bo'lgani bois tabiiy til murakkab tizim sanaladi. Dunyodagi barcha tillar o'ziga xos morfologik xususiyatiga ega. Har bir til yuz mingdan ortiq so'zlarni o'z ichiga oladi. Mana shu so'zlar orqali muloqotga kirishganda turli nutqiy hosilalardan so'z shakllari yuzaga chiqadi. "Kompyuter morfologiyasi ana shu nutq aktidagi birliklarni jarayondagi mexanizmlarni aniqlash⁷⁹" va tahlil qilish bilan shug'ullanadi. Morfologiyaning eng kichik qurilish materiali morfemlardir. Shu bois so'zlar yetakchi (leksema) va ko'makchi (affiks) morfemalarga ajratiladi. Morfologiyada so'zlarning eng kichik ma'noli qismigacha bo'lgan birlik aniqlanadi. Turli tillarda bu holat har xil bo'ladi. Masalan, o'zbek tilida muayyan grammatik suffikslarning qo'shilishidan fonetik hodisalar yuz beradi. Aksariyat so'zlarda bu holat bo'lmaydi: *shahar+im=>shahrim, u+ga=>unga, qishloq+im=>qishloq'im, rab+im=>rabbim*. Ingliz tilida ham ba'zi o'zgarishlar kuzatiladi: *leaf+s=>leaves* yoki o'zagida qisman (*choose-chose-chosen*) yoki mutlaq (*catch-caught*) o'zgarish yoki umuman (*run Inf. -run PP.*) o'zgarish sodir bo'lmasligi mumkin.

⁷⁷ Qarang: Шемакин Ю.И. Начала компьютерной лингвистики. –М.: МГОУ, 1992.

⁷⁸ Dastlabki avtomatik morfologik analiz qilish 1950-yillarda mashina tarjimasiga doir Andron (1962), Woyna (1962), Bernard Georges (1962) Brand (1969), Hutchins (2001) kabi olimlarning ishlarida o'z ifodasini topgan. McIlroy ishlarida avtomatik tahrirlovchi dastur ilovasida avtomatik morfologik analiz qilish bo'yicha amaliy natijalarga erishilgan (1982).

⁷⁹ Harald Trost Computational morphology. www.ccl.pku.cn/nlp/Lexical_Analysis/Word_Lemmatization/Introduction/Computational%20Morphology.htm

So'zlarning tuzilishi (morfotaktik holati)

O'zbek tilida grammatik qo'shimchalarning ketma-ketligi aniq belgilab qo'yilgan, biroq qoidadan tashqari holatlar ham uchraydi. Ot va fe'l so'z turkumlarida quyidagicha:

OT	bola+jon+lar+im+dagi+lar+niki+mas+mi
FE'L	yoz+dir+tir+il+ma+yotgan+lar+ga+mi yoz+dir+tir+il+ma+di

Morfotaktik qoidalar morfemalarning o'zaro kombinatsiyalarini boshqaradi. Shu jihatdan matnlarni analiz qilganda so'zlarning asosi va affikslarni aniqlashda morfotaktik tamoyillarga tayaniladi. Matnlarni ovozlashtirish (text-to-speech), mashina tarjimasida kabi dasturlar uchun morfologik tahlil yordam beradi.

Kompyuter morfologiyasining asosiy vazifasi bu string tipidagi elementlar yoki fonemalarni natija sifatida yuzaga chiqarib berishdir. Masalan, biror so'zni morfemalarga ajratish yoki ularni morfotaktik analiz qilish:

Morfologiyada birliklar sifatida quyidagilar tahlilga tortiladi: token, lemma, stem. Token bu matnda ishtirok etayotgan barcha birliklar bo'lib, so'z shakllari ham uning tarkibi sanaladi. Lemma esa muayyan segment birlikning asosiy shakli (lug'at shakli). Stem so'zning negizi, ya'ni asosi (affikslardan holi qismi) hisoblanadi. Masalan, *Men O'zbekistonda yashayman, u mening vatanim* jumlasida oltita token bor, lekin lemma beshta. Chunki *men+ning, men* har ikkisi bitta lemmadan hosil bo'lgan. Jumladagi stemlar soni oltita *men, O'zbekiston, yasha, men, vatan, u*.

So'z turkumlari. O'zbek tilida 12 ta so'z turkumi bo'lib, u uch guruhga bo'linadi:

1. Mustaqil so'z turkumlari: ot, sifat, son, olmosh, fe'l, ravish.
2. Yordamchi so'z turkumlari: ko'makchi, bog'lovchi, yuklama.
3. Undov so'zlar.
4. Taqlid so'zlar.
5. Modal so'zlar.

So'zlarning morfologik tasnifi. Bu tasnifga binoan grammatik qo'shimchalarning qo'shilish imkoniyatiga ko'ra so'zlar ikki guruhga ajraladi: a) o'zgaruvchi so'zlar; b) o'zgarmas so'zlar.

O'zgaruvchi so'zlar lug'aviy shaklni qabul qila oladi. O'zgarmas so'zlar esa bunday xususiyatga ega emas. Mazkur tasnifda so'zning o'rni quyidagicha:

O'zgaruvchi	O'zgarmas
Fe'l	Ko'makchi
Ot	Bog'lovchi
Sifat	Yuklama
Ravish	Taqlid
Son	Undov
Olmosh	Modal

O'zgaruvchi so'zlar uchun lug'aviy shakl mavjud. Masalan, otda birlik, ko'plik, kichraytirish-erkalash shakli, fe'lda nisbat, bo'lishli-bo'lishsizlik, ravishdosh, sifatdosh, harakat nomi qo'shimchasi, ko'makchi fe'li (harakat tarzi), sifat va ravishda daraja. O'zgarmas so'zlar lug'aviy shaklga ega emas.

So'zlarni morfologik jihatdan so'z turkumlarga ajratishda quyidagi lingvistik masalalar masalalarni ham qilish zarur⁸⁰:

So'zni turkumga ajratishning yuqoridagi to'rtta asosini bilish o'zbek tilidagi istalgan so'zni biror turkumga kiritish uchun yetarli emas. Buni murakkablashtiradigan qator holatlar mavjud.

*Birinchi*dan, omonim shaklli so'z birdan ortiq turkumga mansub bo'lishi mumkin. Masalan:

1. *Endi nima qilsam ekan, – deya Otabek o'z-o'zidan so'radi* (o'z-o'zidan – olmosh).

2. *Qizcha o'z-o'zidan raqsga tushib ketdi* (o'z-o'zidan – ravish).

3. *O'z-o'zidan, mamlakat taqdiri hammani qiziqtiradi* (o'z-o'zidan – modal so'z).

Kasal, issiq, sovuq, salqin, nam, ko'k kabi so'zlar ham sifat, ham ot hisoblanadi: sifatlar: *issiq non, sovuq havo, salqin joy, nam iqlim, ko'k choy*; otlar: *issiq ta'sir qilgan, sovuq urgan, salqin tushdi, nam tushdi, ko'k yuzini bulut qopladi*. Ular shakldosh so'zlardir.

*Ikkinchi*dan, so'z mavjud bir ma'nosi yoki ma'nosidan tashqari boshqa turkumga ko'chishi mumkin. Masalan, yo'l so'zi takroriy

shaklda sifatga o'tadi. To'liqsiz fe'l esa zidlov bog'lovchisi guruhiga ko'chgan. Ko'makchi va modal so'zning katta qismi bir ma'nosi bilan shu turkumga ko'chgan so'zdan tashkil topgan.

*Uchinchi*dan, so'z turkumi bir-birining vazifasini bajarib turishi mumkin. Masalan, ot *temir darvoza, cho'yan pechka, yog'och qoshiq, oltin kuz* kabi birikmalarda sifat, *qon yig'lamoq, takror aytmoq, dahshat tayyorlanmoq* kabi birikmalarda ravish vazifasini bajaradi.

Bir soni faqat yuklamasi, *goh-goh bog'lovchisi, juda ravishi* o'rnida qo'llanishi, ya'ni uning vazifasini bajarishi mumkin. **Bilan** ko'makchisi, **-u, -yu, -da** yuklamalari teng bog'lovchi vazifasini bajaradi.

*To'rtinchi*dan, so'zlar nutq jarayonida o'z vazifasini bajara turib vaqtincha yana bir so'z turkumi vazifasini ham zimmasiga olishi mumkin. Otlashish hodisasining 4 muhim belgisi bor:

1. Otlashgan so'zga bog'langan ot tushib qoladi: *Yaxshi (odam) oshini yer, yomon (odam) – boshini*. 2. Otlashgan so'z otning shakl yasovchi qo'shimchasini oladi: *Yaxshidan adashma, yomonga yondashma*. 3. Otning so'rog'iga javob bo'lishi mumkin: *Intilganga (kimga?) tole yor*. 4. Otning sintaktik vazifasini bajaradi: *Shakarning ozi shirin*.

So'z turkumlarining morfologik modelini yaratishda standart morfologik teglardan foydalaniladi. Ushbu teglar orqali tilni tushunish va ularni modellashtirish mumkin.

5.2. Tabiiy tilning morfologik modeli

Morfologiyaning formal modeli so'zning strukturasi, so'zlarning o'zgarish qoidalari, grammatik ifoda shakli va so'z yasash usullarini aks ettiradi. Ma'lumotni mashina yordamida qayta ishlash nuqtai nazaridan morfologik modeli so'zda anglashilgan ma'noni ifodalaydi. So'zda anglashilgan ma'no kontekstda o'z ifodasini topadi.

Tilning morfologik modeli so'zlarning grammatik kategoriyasi va so'z shakllaridan iborat morfologik lug'atda o'z aksini topadi:

⁸⁰ B. Mengliyev, O'. Xoliyorov, N. Abdurahmonova O'zbek tilidan universal qo'llanma (o'quvchi, abituriyent va talabalar uchun) (qayta ishlangan 3-nashri), Toshkent: Akademiya, 2014. –B. 164-165

Tag	properties		Types	definition	questions	Example
N				shaxs, predmet, hodisa, joynomini bildiradigan so'zlarga aytiladi	Kim? Nima? Qayer?	
	Tuzilishi					
SIMP	111	sodda				
CMPL	112	murakkab				
FUSW		1121	qo'shma ot	So'z tarkibidagi ikki yoki undan ortiq asosdan iborat bo'lgan so'zlarga aytiladi.		Qo'rg'ontepa, xontaxta
PAIR		1122	Juft so'z	Ikki so'zning juftlashib yangi ma'no hosil qilishiga aytiladi		Bolachaqa, onabola
CMPN		1123	Murakkab so'z	Ikki yoki undan ortiq leksemalarning birikuvidan hosil bo'lgan so'z		Ishdan nari, tog'yoqalab
ABBR		1124	Qisqartmalar	Bir necha so'zlarning qisqartma holda kelib yangi ma'no hosil qilishi		BMT, ADU
UNDR		1111	Tub so'z	Tarkibida yasovchi qo'shimchalar		Bola, ota, ona,

				bo'lmagan so'zlarga nisbatan aytiladi		aka
DRVT		1112	Yasama so'z	Tarkibida yasovchi qo'shimchalar bo'lgan so'zlarga nisbatan aytiladi		Bolalik sezgirlik
	Ma'nosi					
ANIM	121	jonli		Kim? degan so'roqqa javob berib, shaxsni ifodalab keluvchi otlarga aytiladi.	Kim?	O'qituvchi, bobo
INAM	122	Jonsiz		Kishilardan tashqari nima so'rog'iga javob beradigan so'zlarga aytiladi.	nima?	Non, yostiq, uy.
CMMN	131	Turdosh ot		Bir turdagi shaxs/predmet/hodisa/joyning umumiy nomini bildiruvchi olturdosh ot		Kino, maktab
PRPR	132	Atoqli		Bir turdagi shaxs/predmet/hodisa/joydan biriga atab qo'yilgan nom		Toshkent Mushtariy
	Ma'nosi jihatdan (gram					

	matik ada leksik - seman tik ma'no deb nomla nadi)					
CNCR	141	Aniq otlar		Aniq ot ifodalagan narsani sezgi a'zosi bilan sezish, ko'rish, ushlash mumkin		Gul, gilam
ABST	142	Mavhu m		Mavhum ot ifodalagan narsani sezgi a'zosi bilan sezish mumkin emas		Muhab bat, do'stlik
Cases	Kelish iklar			Kelishik ergash so'z vazifasidagi ismni bosh so'z vazifasidagi ism(qaratqich kelishigi)va fe'l bilan(boshqa kelishik) bog'laydi		
NOM	151	Bosh kelishik			Kim? Nima?	Kitob
GEN	152	Qaratqi ch			Kimnin g? Bolanin g	Kitobni ng

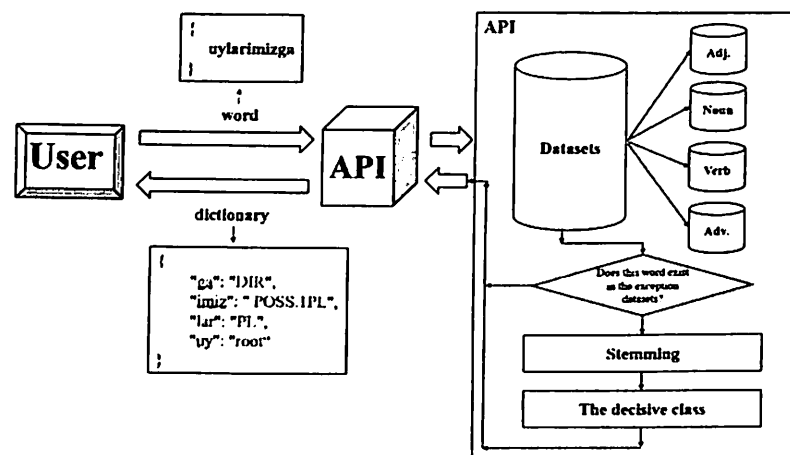
DIR	153	Jo'nalis h			Kimga? nimaga?	kitobga
DIR_LI M						
ACC	154	Tushum			Kimni? Nimani?	kitobni
LOC	155	O'rin payt			Kimda? Nimada ?	kitobda
ABL	156	Chiqish			Kimdan ? Nimada n?	Kitobd an
INST	157	<u>Vosita</u>			Kim bilan? Nima bilan?	Kitob bilan
Plural	Ko'pli k			Otning -lar qo'shimchasi		
SG	161	Birlik				Kitob
PL	162	Ko'plik			kimlar? Nimalar ?	kitoblar
<u>Pers</u>	Shaxs -son qo'shi mchal ari					
P1SG1	171	1 shaxs birlik				Men o'quvc himan
P1SG2						
P2SG1	172	2 shaxs birlik				Sen o'quvc hisan
		2 shaxs ko'plik				Siz o'quvc hisiz
P2SG.P1						
P2SG2						

P2SG.P2						
P3SG	173	3 shaxs birlik				U o'quvchi
P1PL1	175	1 shaxs ko'plik				Biz o'quvchimiz
P2PL1	176	2 shaxs ko'plik				Senlar o'quvchilari
P2PL.P1	178	2 shaxs ko'plik				Sizlar o'quvchilaringiz
P3PL	177	3 shaxs ko'plik				Ular o'quvchilari
POSS	Egalik qo'shimchalari			Egalik shakli predmet-hodisaning qaysi shaxs va songa qarashli ekanligini bildirib, o'zi birikkan bosh so'zni qaratqich kelishidagi ergash so'zga bog'laydi	Kimim? Nimam?	
POSS.1S G	181	1 shaxs birlik			Kimim? nimam?	Mening o'quvchim
POSS.2S G	182	2 shaxs birlik			Kimingi? Nimangi?	Sening o'quvching
POSS.2S G.P	184	2 shaxs birlik (hurmat)			Kimingi? Nimangi?	Sizning o'quvchingiz

POSS.3S G	183	3 shaxs birlik				Kimisi? Nimasi?	Uning o'quvchisi
POSS.1P L	185	1 shaxs ko'plik				Kimingi? Nimangi?	Bizning o'quvchimiz
POSS.2P L	186	2 shaxs ko'plik				Kimilari? Nimalaring?	Senlarning o'quvchilaring
POSS.2P L.P	188	2 shaxs ko'plik				Kimingi? Nimalaringiz?	Sizlarning o'quvchilaringiz
POSS.3P L	187	3 shaxs ko'plik				Kimilari? Nimalari?	Ularning o'quvchilari
Tag	Tarkibi				Shaxs, predmet, hodisa joy nominibildiradigan so'zlarga aytiladi	Kim? Nima? Qayer?	

Morfologik model quyidagi parametrlariga ko'ra farqlanadi. Dastlab morfologik modellar morfologik analizatorga asoslangan ishlarning natijasiga ko'ra alohida xarakterga ega. Morfologik analizator ma'lumotni kiritishda tabiiy tildagi muayyan so'z shakllarni, uzatishda esa barcha so'z bilan bog'liq grammatik belgilari (kelishik, shaxs-soz, zamon va h.k.) kiritiladi. So'ngra morfologik model leksikani to'liq qamrab olishi (ma'lumotlar bazasida barcha leksemalar morfologik darajadagi dasturda qayta ishlanadi) yoki qisman leksik imkoniyatlar hisobga olinishi mumkin (ma'lumotlar bazasiga kiritilmagan leksemlarning yuzaga kelish

imkoniyatlarining hisobga olinishi). Keyingi qadam morfologik modellar soʻz shakllarining ajralishi yoki ifoda etish usullari bilan oʻzaro tafovutga ega boʻladi.



Leksemlarning ifoda etilishi ikki asosiy usul bilan amalga oshiriladi:

1) bazada barcha leksemlarning soʻz shakllari saqlanadi (ehtimol, ularning grammatik xususiyatlari bilan) hamda qandaydir holatda bir soʻz shaklga taalluqli soʻz shakli aniqlanib olinadi. Bunday usulda leksemaning berilishi ayrim flektiv tillar uchun samarali va qulay hisoblanadi, yaʼni bunday tillarda, asosan, fleksiya hodisasi emas, balki muayyan grammatik usulda, masalan, predloglar yordamida soʻz shakllarini hosil qilish muhim sanaladi. Bu kabi tillarga ingliz tilini misol sifatida keltirish oʻrinli.

2) maʼlumotlar bazasida leksemlar va fleksiyaning roʻyhati saqlanadi (grammatik xususiyatlarining maʼnolari bilan), biror soʻz shaklni qabul qilishi uchun asosiga bogʻlangan boʻladi. Bunday usul leksemani berishda turli grammatik kategoriyalar turli oʻzgarishlarga uchraydigan flektiv tillar uchun samaradorligi bilan ahamiyatlidir. Bu kabi tillar sinfiga rus tilini keltirish mumkin. Modellar, leksemlarni uzatish usuli yana ikki guruhga ajraladi: birinchisi sof orfografik asos va fleksiya hisoblansa, boshqasi esa psevdososli (soʻzning bosh oʻzgarmas asosi) hamda psevdofleksiya (soʻzning tugal qismi oʻzgaruvchi) boʻladi. Asoslarning u yoki bu variantlarini aniqlashda, asosan, amalga oshirish samaradorligi hamda morfologik

komponentlarning belgilanganligi inobatga olinadi. Istalgan morfologik modelda har bir leksemaning grammatik xususiyatlari hisobga olinishi kerak: sintaktik guruh (gap boʻlaklari), soʻz oʻzgarish (paradigmatik) sinfi va grammatik kategoriyalar yoki sintaktik tasnifga bogʻliq grammatik turlanish yoki tuslanishi. Sintaktik aloqa ikki xil usulda boʻlishi mumkin: erkin va bogʻlangan grammatik oʻzgarish, masalan, rus tilida ot uchun jonli va rod sifatleri boʻlsa, erkin oʻzgarishda ot uchun son va padej misol boʻla oladi.

Soʻzga bogʻlangan muayyan sintaktik guruhda quyidagilar boʻladi:

- soʻz oʻzgarishlarning umumiy majmuasi;
- erkin soʻz oʻzgarishlarning umumiy majmuasi;
- soʻz oʻzgarishlarning umumiy koʻplik maʼnosi (общее множество значений ГП);
- umumiy sintaktik funktsiya.

Morfologik analizni amalga oshirish uchun undagi qoidalar formal holatda kiritiladi. Morfologiya yoki sintaksisga kompyuter yordamida yondashish bu tabiiy tilni modellashtirish, algoritim, analiz yoki parsing orqali amalga oshiriladi.

Morfologik grammatikani statistik soʻz turkumlari boʻyicha taggerlarga oʻtkazish hamda koʻp modellarni kodlashtirish uchun chekli avtomat va transyuterlardan foydalaniladi. Tabiiyki, qanday qilib avtomat algebrik va nazariy jihatdan fikrlay olishi, chekli avtomatning til va nutqni qayta ishlashdagi rolini aniq bilish kerak. Morfologik analiz faqat u yoki bu kategoriyalarni tahlil qilish bilan cheklanmaydi. Ayniqsa, mashina tarjimasida buning oʻzi kifoya emas. Yuqorida tilga olingan baʼzi muammoli holatlar uchun Rodolfe Delmonte leksikon uchun quyidagi lingvistik kategoriyalarni koʻrsatib oʻtadi: 1) grammatik kategoriyalar; 2) semantik kategoriyalar; 3) diskurs darajasidagi kategoriyalar; 4) sintaktik kategoriyalar; 5) bogʻlangan kategoriyalar; 6) semantik konseptual kategoriyalar; 7) chegaralangan leksik birliklar; 8) grammatik cheklovlar. Bu tasniflash italyan tili uchun xos boʻlib, uning lingvistik tomonlari hisobga olingan. Matnlarni tahlil qiluvchi avtomatik analizatorni yaratishda dastlabki bosqich sifatida oʻzbek tilining elektron morfologik lugʻatini yaratish masalasi qoʻyildi.

Unga ko'ra har bir so'zning qaysi so'z turkumiga tegishli ekanligi statistik tahlilda ko'rsatildi. Har bir leksemaning muayyan morfologik xususiyatlari jadvallarda aks etadi. So'ng har bir leksema so'z turkumlari bo'yicha to'plamlarga tasniflanadi. To'plamlarga ajratishda leksemalarning umumiy semantik, qolaversa, formal jihatlari inobatga olinadi. To'plamlarga ajratilgan har bir so'zga qo'shilishi mumkin bo'lgan qo'shimchalar kombinatsiyasi aniqlanadi. Undan oldin barcha qo'shimchalar funksional belgilariga ko'ra guruhlarga ajratilib, ularning o'zbek tilidagi barcha mumkin bo'lgan mantiqiy bazasi shakllantiriladi.

Avtomatik tahlil qilishning quyidagi turlari mavjud: 1) *so'zlarning lug'atdagi asoslariga ko'ra tahlili*; 2) *so'zlarning lug'atdagi so'zshakllariga ko'ra analizi*; 3) *mantiqiy ko'paytirish metodi orqali tahlili*; 4) *lug'atsiz jadvallar yordamida analiz qilish*. Tilshunoslikning tillar tipologiyasi va qiyosiy tahliliga oid tadqiqotlar generativ tilshunoslik aspektida ko'rib chiqiladi.

Modellashtirish metodi barcha sohalarda muayyan obyektни o'rganishda muhim ahamiyatga ega. Kompyuter bazasi uchun ham tilni modellashtirish eng qulay tadqiqot usulidir. Kompyuterli morfologiya yuqori darajadagi dasturlar uchun zarur hisoblanib, morfologiya doirasida modellashtirish tadqiqotlarda o'rganilib kelinmoqda. Qariyb o'ttiz yildan oshiq vaqt mobaynida turli dasturlar uchun morfoanalizatorni takomillashtirish bo'yicha ishlar amalga oshirilmoqda. Zero, statistik metodga asoslangan tizimlar mashina tarjimasini uchun yaratilgan bo'lsa-da, tabiiy tillarning imkoniyatlari to'liq qamrab ololmaydi. Ushbu yo'nalishda qoidaga asoslangan tizimning grammatik tahlil qilish usuli "tilga oid bo'lgan polisemiya va omonimiya hodisalarini izchil tadqiq etishi hamda turli tipga kiruvchi tillarning grammatik kategoriyalari farqli jihatlari bilan birgalikda o'rganilishi" bilan ham ahamiyatlidir.

Morfologik modellashtirishda lingvistik tavsifning nechog'lik to'g'ri va aniq berilishi muhim sanaladi. Morfologik mezonlarning to'g'riligi quyidagi bosqichlarda tekshiriladi: 1) faqat to'g'ri tuzilgan grammatik so'z shakllarining tahlili; 2) grammatik jihatdan noto'g'ri deb topilgan so'z shakllarining qay tarzda aniqlanishi. So'zshakllarning to'g'riligi faqat to'g'ri deb topilgan modellar yoki

istisno etilgan qoidalardan aniqlashtiriladi, qolgan holatlar esa uning noto'g'riligini tasdiqlaydi.

Ayrim adabiyotlarda morfologik tavsifning komponentlari sifatida quyidagilar tushuniladi: lug'atning maxsus obyektli modeli, lug'at, morfologik qoidalar hamda fonologik va morfofonologik qoidalar. Maxsus lug'atlarda morfem tiplari, barcha turlarining xossalari va tavsifi beriladi. Lug'atdagi har bir morfema faqat grammatik ko'rsatgichga ega bo'lishi kerak. Morfemalarning semantik xususiyatlari ularning strukturasi belgilab beradi. Morfologik qoidalar so'zshakldagi morfemalarning o'zaro bog'lana olish xususiyatlarining chegarasini aniqlay oladi, ya'ni morfemalarning omonimiyaga xosligi shu holatda muayyan turga tegishliligi bilan farqlanadi. Fonologik komponentda morfemalarning allomorfligi natijasi aniqlanib, kontekstda so'zshakllaridagi allomorflar o'zgarishi hamda so'zshakllarining fonotaktik to'g'riligi tekshiriladi.

Tabiiy tillarni qayta ishlashda grammatik tahlil muayyan bosqichlarga ega bo'ladi. Grammatikaning morfologiya va sintaksis bo'limlariga tegishli barcha lingvistik qurilmalar, jumladan, so'z turkumlarining modellarini tuzish, ya'ni kompyuterning o'qish holatiga o'tkazishda tildagi barcha me'yoriy vaziyatlar qamrab olinishi maqsadga muvofiq. E'lon qilingan avvalgi ishlarimizda mashina tarjimasini uchun avtomatik tahlil jarayonlari o'rganilgan. Unga ko'ra leksik tahlil uchun mashina tarjimasining dastlabki *tokenizatsiya->lemmatizatsiya->stemming* bosqichlarida morfologik tahlil amalga oshiriladi. Tokenizator so'zlarni chapdan o'ngga qarab, u yoki bu so'zni qaysi so'z turkumiga tegishliligi hamda asos va qo'shimchalar tekshiriladi.

Tarjimon dastur inglizcha-o'zbekcha yo'nalishda olib borilganligi bois, dastlab ingliz tilidagi otga xos barcha kategoriyalar o'zbek tilining kategoriyalari bilan solishtiriladi. Lingvistik ta'minotda ingliz tilidagi ot so'z turkumining quyidagi paradigmatlari o'rganiladi:

1. *Atoqli va turdosh ot:*

1.1. Atoqli otlar (PropN):

1.2. Turdosh otlar (CN):

1.2.1. Aniq ot:

1.2.1.1. Jonli;

1.2.1.2. Jonsiz;

1.2.2. Mavhum.

1.2.3. Sanalmaydigan ot:

1.2.3.1. Mavhum;

1.2.3.2. Modda va material.

1.2.4. Jamlovchi ot:

1.2.4.1. Kishi nomlari (oila, xalq);

1.2.4.2. Jonsiz otlar (shakar).

2. Otning tuzilishiga ko'ra turlari:

2.1. Sodda ot;

2.2. Qo'shma ot;

2.3. Juft ot;

2.4. Qisqartma ot;

2.5. Takroriy ot.

3. Son:

3.1. Birlik;

3.2. Ko'plik.

4. Kelishik

Tilshunoslikda atoqli otlarning (*Np*) tasniflashi muayyan turlarga ko'ra ajralatiladi: kishi nomlari, geografik nomlar, tashkilot nomlari, mahsulotga qo'yilgan nomlar, lavozim yoki daraja, medallar, taxalluslar v.h. Bosh harfdagi so'zlar bilan boshlangan atoqli otlarni quyidagicha $Np = \{A, B, C \dots Z\}$ ifodalasak, u holda *Np* va *N* bo'ladi. Ayrim vaziyatlarda ingliz tilidagi atoqli otlar o'zbek tilidagi turdosh otlarning qism to'plami ham bo'lishi mumkin, u holda to'plamlar *EngPropN UzbCN* ko'rinishida bo'ladi: a) hafta kunlari (*Monday*); b) oy nomlari (*April*); c) til va millat nomlar (*Uzbek*) – bu turdagi otlar o'zbek tilida turdosh ot sanaladi. Grafematik jihatdan otga tegishli so'zlarni nafaqat tahlil qilish, balki lingvistik ta'minotda ularni morfologik jihatdan loyihalash muhim bosqichlardan biridir. Zero, grammatik modellar semantik tahlillarga ham asoslanishi, qolaversa, atoqli otlarning ayrim farqli tomonlarini tasniflashda alohida o'ringa ega. Atoqli otlarni tarjima qilishda, dastlab har ikki tilda barcha atoqli otlarning inglizcha-o'zbekcha lug'ati elektron baza shaklida tuzilgan bo'lishi kerak. Negaki, tarjimadagi ayrim xatoliklar bevosita shu o'rinda hal etiladi.

Masalan, *Chomsky* => <*Xomskiy*> yoki <*Chomskiy*>; *Regent Street* => <*Rejent ko'chasi*> yoki <*Rejent Ko'cha / ko'cha*>. Inglizcha atoqli otlarni o'zbek tiliga tarjima qilishda muayyan vaziyatlar o'rganilishi, ya'ni mavjud "qog'oz" lug'atlarda atoqli otlarning transkripsiyasi mavjudligi hamda ularning o'zbek tiliga qay tarzda o'zlashtirilishi aniqlangan holda ma'lumotlar bazasida berilishi maqsadga muvofiq. Tadqiqotda o'zbek tilida atoqli ot hisoblanmaydigan yoki aytilishida farqlanuvchi so'zlarning alohida lug'ati bazaga kiritildi: *English - ingliz tili, Monday - dushanba* kabi.

Otlarning tub yoki yasama so'z ekanligi u darajada muhim emas. Masalan, *dragonfly* so'zining matn ichidan so'z sifatida ajratib olinishida so'zning qanday komponentdan tashkil topganligi e'tiborga olinmaydi. Zotan, aksariyat flektiv tillarda morfemadan keyin ko'makchi morfemalar qo'shilmaydi (-s, -ed va h.k.). Biroq o'zbekcha so'zlarning grammatik tahlilida har bir morfema ahamiyatga ega. Zero, kontekstda *tub so'z - tub so'z, tub so'z - yasama so'z, qo'shimcha - qo'shimcha, tub so'z - grammatik shakllangan so'z omonimiyasi hodisasi ko'p uchraydi: olma - ol + ma; suv + sa - kel + sa, gul + don (ot) - bilim + don (sifat)*.

Ingliz tilida qo'shma otlar qo'shib (*bluebell*) yoki chiziqcha bilan (*mother-in-law*) yoziladi. Aksariyat hollarda bu kabi otlar har ikki tilda alohida tarzda qo'llanishi ham mumkin. Qisqartma so'zlar ham nazardan chetda qolmasligi maqsadga muvofiq. Qisqartmalar ingliz tilida turli so'z turkumiga oid bo'lib ham turdosh otga, ham atoqli otga tegishli bo'lishi mumkin v.s. *yuqoriga qarang, v.v. - aksincha, e.x. - masalan; NATO (North Atlantic Treaty Organization) - NATO, ISBN (International Standard Book Number) - ISBN, biroq the UNO, the UN (the United Nations Organization) - BMT (Birlashgan millatlar tashkiloti), PC (Personal computer), ShK (shaxsiy kompyuter)*. Bundan tashqari Internet muloqotida *onna = oh no, not again "yo'q, qaytarish shart emas"* kabi akronimlar ham ishlatiladi. Qisqartma otlarning o'zgarishsiz qoluvchi shakllari lug'atdan tayyor holatda olinishi, o'zgaruvchi birliklar haqida esa terminografiya va leksikografiyada aniq to'xtamga kelinishi zarur.

So'zning morfologik tahlili gapning umumiy strukturaviy modelini, qolaversa, leksemaning sintaktik vazifasini belgilashga yordam beradi. Otga tegishli deb topilgan so'z tokenizatsiya

(morfemalarga ajratish) jarayoniga o'tkaziladi. Ingliz tilida so'zga qo'shilishi mumkin bo'lgan affikslar turi ko'p emas. Ular sirasiga ko'plik qo'shimchasi (-s, -es) hamda egalik qo'shimchasi ('s, s') kiradi. Ingliz tilidagi otlar son kategoriyasi o'zbek tilidan biroz farq qiladi. Umuman, otlar birlik va ko'plik shaklida grammatik ko'rsatgichga ega bo'ladi, yasaliş xossalariga ko'ra ko'plik kategoriyasi istisnoli holatlarda ham aks etishi mumkin. Shu bois bazada har bir otga tegishli bo'lgan sanaladigan va sanalmaydigan ko'plik turlari an'anaviy lug'atlarda aks etgani kabi indekslanadi. Ko'plik qo'shimchasi <s> yoki <es> bo'lib, ularning qo'shilishi yoki qo'shilmasligi yoxud ularning ishtirokisiz hosil qilinuvchi ko'plik kategoriyasiga mansub otlarni bazada indekslash maqsadida har bir turini shartli ravishda ot turiga ko'ra belgilab olinadi: N1, N2, N3, N4, N5, N6, N7, N8, N9, N10

1-qoida. O'zbek tilida ko'plik qo'shimchasi bevosita qo'shilish imkoniyatiga egaligi bois leksemaning qaysi harf bilan yakunlanishidan qat'iy nazar umumiy yagona model quyidagicha bo'ladi:

N + -lar => kitob+lar; olma+olmalar

2-qoida. Sanaladigan otlar ko'plikda ifodalanishiga ko'ra quyidagi model ko'rinishida bo'ladi:

Noun+(e)s->umumiy ko'rinish:

N1-tipidagi otlar. Unga <s> yoki < es> suffikslari qo'shilganda fonetik hodisaga uchramaydigan leksemalar kiradi hamda ular quyidagicha modelga ega:

N(cons.)+s=> so'z=>Ncount+s=> flowers (gullar), guys (yigitlar);

PropN+KQ=> the Germany+s=>the Germanys (germaniyaliklar);

(N(vow.+y))+s=>day+s=> days (kunlar);

(CN+y)+s=> stand-by+s=>stand-bys (tayyorgarlik holatlari);

(N(+o)count+s)=>photo+s=>photos (rasmlar).

N2-tipidagi otlar. Bu turga kiruvchi otlarga <es> suffiksi qo'shilganda quyidagi modellar hosil bo'ladi:

N(+x|s|sh|asch|z)count+es=>box+es=> boxes (qutilar);

N(+o)count+es=>heroes (qahramonlar), potatoes (pomidorlar).

N3-tipidagi otlar. Bu tipga <s> yoki <es> qo'shilganda leksemalarning oxirgi qismida <f> -> <v> ga o'zgaradi:

N(-f)count+es=>leaf->leaves (yaproqlar);

N(-fe)count+s=> knife->knives (pichoqlar).

N4-tipidagi otlar. Bu turdagi otlar <f> undoshi bilan tugallansa-da, fonetik hodisaga uchramaydi:

N(-f)count+s=>chief+s=>chiefs (rahbarlar).

N5 tipidagi otlar. Bularga <es> qo'shilganda y>i hodisaga uchrovchi ot sinfiga mansub so'zlar kiradi:

N(-y) count+es=>sky+es=>skies (osmonlar);

N(-que) count+es=>soliloque+es>soliloques (monologlar).

N6 tipidagi otlar. Bu sinfga Noun->Noun_i komponentli otlar mansub bo'lib, ko'plik shaklida so'zning o'zagida fonetik o'zgarish bo'luvchi otlar kiradi:

Goose->geese (g'ozlar), tooth->teeth (tishlar);

Child->children (bolalar), ox->oxen (buqalar).

N7 tipidagi otlar. Bularga Noun->Noun ko'rinishli tur tegishli bo'lib, ingliz tilida leksik usulda ko'plik hosil qilinsa-da, hech qanday KQ ifodalovchi suffiks qo'shilmaydi, bu holatda N₂ N₁):

One sheep (bitta qo'y) <=>many sheep (ko'p qo'ylar)

A fish (bitta baliq) <=>a few fish (bir qancha baliqlar)

N8 tipidagi otlar. Bu kabi otlarga boshqa tillardan o'zlashgan so'zlar tegishli bo'lib, o'zagida mutloq yoki qisman o'zgarish sodir bo'ladi:

Corpus->corpora; stimulus->stimuli; criterion->criteria; datum->data

N9 tipidagi otlar. Bunday turga qo'shma ot tipiga mansub so'zlar tegishli bo'lib, unda tarkibidagi ko'plik qo'shimchasi turli holatlarda kelishi mumkin:

N9_both: Tarkibida *man* yoki *woman* leksemasi bo'lgan qo'shma otlar o'zagida o'zgarish bo'luvchi otlarning har ikki qismida KQ beriladi va ular alohida tasniflanadi: *men-servants* (erkak xizmatkorlar), *women-doctors* (ayol shifokorlar);

N9_first: Qo'shma otlarning birinchi qismiga KQ qo'shiladigan otlar (bunday otlar tarkibida, asosan, predloglar ishtirok etadi):

commanders-in chief (bosh sardorlar), *mothers-in low* (qaynonalar);

birinchi qismga KQ qo'shiladigan otlar: *passers-by* (*yo'lovchilar*), *lookers-on* (*tomoshabinlar*);

N₁₀ **last** tipidagi otlarning oxirgi qismiga KQ qo'shiladigan otlar: *forget-me-nots* (*bo'tako'zlar*), *stand-byes* (*tayyorgarliklar*).

N₁₁ **tipidagi otlar**. Bu turdagi otlarning so'nggi qismida <s> qo'shimchasi yaxlitlangan bo'ladi.

Ular birlik ma'nosida kelsa-da, o'zidan keyin keluvchi kesimning ko'plikda kelishini taqozo etadi. Bu qo'shimcha so'zning tarkibida yaxlit holda bo'lgani bois tarjima qilinmaydi: *jeans-djensi*, *scissors-qaychi*, *scales-tarozi* v.h. Ba'zi so'zlar har ikki holatda ham ishlatilishi mumkin, biroq tarjimada ular tafovutga ega: *time* Uncount (vaqt); *time count+s* (*marta*)=>2 times (2 marta). Bu holatda sanaladigan ot sanalmaydigan ot bilan omonimik munosabatga kirishadi. Tarjimada inobatga olinishi kerak bo'lgan yana bir jihati shundaki, otga tegishli so'zlar oldidan keluvchi sanoq sonlardan keyin, albatta, ingliz tilida ko'plik qo'shimchasi qo'shiladi, biroq o'zbek tilida ular tarjima qilinganda birlikda bo'ladi: *2 days ->2 kun*, *5 hours -5 soat*. O'zbek tilida KQ -lar affiksi bo'lib, qo'shilish bilan bog'liq bo'lgan fonetik hodisa uchramaydi. -lar ko'plik ma'nosidan tashqari uslubiy ma'no hosil qiladi: *dadamlar*, *o'zlari*, *kechalari* v.h.

Ingliz tilida artikllar (*a*, *an*, *the*) nafaqat ot so'z turkumidagi leksemalar, balki sifat bilan ham ishlatilishi natijasida otlashish hodisasi yuz beradi: *poor* (*kambag'al*) <-> *the poor* (*kambag'allar*). Otlar tipida bunday so'zlar ham qamrab olinishi maqsadga muvofiq.

Ingliz tilida bosh va qaratqich kelishiklari mavjud. Qaratqich kelishigi qo'shimchasi ikki usulda hosil qilinadi, ya'ni 's va of predlogi orqali.

N+es=>*mother's vase*, *director's name*-> N+ning=>*onaning guldoni*, *direktorning ismi*

N+of+{DET}+N=>the page of the book-> (N+ning)+(N+EQ)=> kitobning varag'i

NPL+s'=>teachers'->NS+lar+ning=>o'qituvchilarning

N1+'s=>children's->N1+lar+ning=>bolalarning

Shu o'rinda so'zlarning <s> shaklida qo'llanilishida chalkashlik seziladi: *She's a doctor*=>*She is a doctor* (U shifokor); *Mother's bag*

(onaning so'mkasi). Bu holatda <s> ishtirokidagi gap bo'laklari modeli quyidagicha bo'ladi:

N|P+'s+NP=> *Nargiza's at home*. (*Nargiza uyda*) yoki *Nargiza's tolerable*. (*Nargiza sabrli*).

N|P+'s+N+VP=> *My uncle's name is Aziz*. (*Mening amakimning ismi Aziz*) yoki *Feruzas's impression was good*. (*Feruzaning taassuroti ijobiy*.)

[N+'s] gap so'nggida kelib o'zidan oldingi so'zga ishora etsa yoki bunday birikmadan keyin ot so'z turkumi qo'llanilmasa, u holda - niki tarzida tarjima qilinadi: NP|VP+(NP+'s)+Ø=>NP+VP+(NP+niki)+Aff {-mas, -dir, -mi, -ya, ...}.

Our house is bigger than yours.=>*Bizning uyimiz siznikidan ko'ra kattaroq(dir)*. *Is our house bigger than yours?*=> *Bizning uyimiz siznikidan ko'ra kattaroqmi?*

The door of the room (*Xonaning eshigi*). Bu holatda egalik qo'shimchasi ingliz tilida ikki xil holatda berilishi mumkin:

N+of+N=>Nning +Neq

The picture of the book => *Kitobning rasmi*

P {PP, both, neither, either, none} + {of+Det.}+N|P=> P

My book=> *Mening kitobim*

Both of you=> *Har ikkingiz*

None of the students => *Talabalarning hech qaysi biri*

O'zbek tilida egalik olmoshi belgisiz holatda ham ishlatilishi mumkin. Ingliz tilida bir necha marta ishlatilgan egalik olmoshi o'zbek tilida takrorlanmaydi, ya'ni u faqat egalik qo'shimchasi sifatida tarjima qilinadi: *He goes his English course* => *U (uning) ingliz tili kursiga boradi*. Bunday holat quyidagicha modellashtiriladi:

N->*book*|*Mushtariy*|*Earth*|UN...

Det->*the*|*a/an*|*my*|*this*|*all*|*both*|*clever*...

V->*go*|*go off*|*have a chat*

Avtomatik tarzda ingliz tilidagi otlarni qo'shimchasiga ko'ra aniqlash yoki uning ega yoki to'ldiruvchiga tegishli ekanligini topish mushkul. Matnni faqat sintaktik strukturasi ko'ra aniqlashda Chomskiyning ilgari surgan daraxtsimon gap tahlili usuli samarali hisoblanadi.

S (NP)+VP (V+NP); bu yerda NP=Det+N // ushbu holatda NP tarkibida ega ishora etuvchi PossP kelsa, u holda sintez bosqichida tushirib qoldirilishi hamda o'zidan keyin keluvchi otga PossPga tegishli bo'lgan egalik qo'shimchasini berishi kerak, shunda tarjima me'yorlashadi: S (NP)+VP (V+NP)=>PP+ning+N+EG+PP+ning+N+EG+V – transformatsion metodning bir ko'rinishi gapda muayyan qismlarning tushirib qoldirilishidir. Yuqoridagi chizmada eng sodda gap qolipi berilgan, biroq gapni, shuningdek, qolgan ikkinchi darajali bo'laklar bilan kengaytirish mumkin, u holda tarmoqlanish yana davom etadi. Natija: *My brother reads his book => Mening akam kitobini o'qiyapti.*

O'zbek tilida egalik va ko'plik qo'shimchalarining qo'shilishiga doir imloviy qoidalar qabul qilingan. Ushbu qoidalarni modellashtirish zarur bo'lgan jihatlarini tahlil qilamiz. Masalan, matn quyidagicha berilgan bo'lsin: *My town is Andijan.*

Har bir komponent morfologik sathda aniqlashtiriladi: PossP+N+is (to be) +PN

Gapning bosh bo'laklari oydinlashtiriladi: S (ega) + P (kesim)

Fe'lgacha bo'lgan qism ega guruhi hisoblanadi va uning modeli aniqlashtiriladi:

Ega guruhining tipi aniqlashtiriladi: ot guruhi yoki nominal birliklar (NP) turkumi tahlil qilinadi.

Ot guruhi (NP) ikkiga bo'linadi: 1) Ph+Noun->otdan oldin keluvchi birikmalar; 2) Noun+Ph-> ot dan keyin keluvchi birikmalar. Har ikki holatda ham birikmalar otga tobelanadi. NP gapda barcha gap bo'laklari bo'lib kela oladi. NP =N+ Prep. + Article+ Adj. +Noun-> *the man in the black coat* (qora kurtkadagi kishi), *the book about my life* (mening hayotim haqidagi kitob). Bu kabi birikmalar aksariyat hollarda ingliz tilidagi predlog bilan bog'lanadi. Ph+N ning strukturasi quyidagicha: Q+Det.+AdjM+NM (noun modifier)+N. Bu kabi birikmalar quyidagi morfologik modelda shakllanadi:

1) {Q (quantifier: one, two...) | M (modifier: noun, adjective, participle, gerund) | D (determiner: Possessive pronoun, article, Pronoun, numeral)} +N=> K (kvantifikator) | M (modifikator: ot yoki sifat) | D (aniqlovchi, PP (kishilish olmoshi) {men, sen+ning(n0-1)})

|| {u, biz, siz, ular+ning} +N+EQ || {u, biz, siz, ular+ning} +N+EQ // tahlil quyidagi algoritm bo'yicha olib boriladi:

egalik qo'shimchasi shaxs-songa qarab tanlanadi: my ->im|-m, your ->ing|-ng, his, her, its ->-i, our->-miz|-imiz, your ->-ngiz|-ingz, their ->-lari

o'zbek tilida egalik qo'shimchasi qo'shilganda fonetik hodisaga uchraydigan ot turkumi turlari aniqlanib, muayyan turga birlashtiriladi: tovush tushish: Ntt {burun, shahar, o'g'il, ko'ngil...} +EQ-> shahar+im=> shahrim;

tovush orttirilishi: Nto {parvo, obro', mavqe, mavzu, avzo...} +y+EQ-> parvo+im=> parvoyim;

tovush almashinish k<g, q<g': Nta {yaxshilik, tilak, bilak, kubok, tayoq...} +EQ=>tilak+im=>tilagim;

Shuningdek, ingliz tilidagi ayrim artikl+ot komponentli birikmalarda semantik xususiyat yaqqol seziladi: *The English – ingliz xalqi; English – ingliz tili.* Ot guruhi tarkibida artikl kelganda muayyan qoidalar kiritilishi kerak: 1) o'zbek tiliga tarjima qilinganda semantik ma'no anglatuvchilar guruhi; 2) tarjimada ma'no anglatmaydigan artiklli otlar guruhi. Noaniq artikl *a, an* quyidagi holatlarda tarjima qilinadi: 1) a|an+Number+N=> a hundred (bir yuz); 2) a|an +Time => an hour; a year| minute| month (bir soat, bir-yil, bir minut, bir oy); 3) a|an+Number+N+N=> *a four floor building* (to'rt qavatli bino).

Sanalmaydigan otlar na birlikka, na ko'plikka mansub. Odatda ularning oldidan a / an artikli ishlatilmaydi. Biroq artikllar hisob so'zlar bilan ham kelishi mumkin: *a pound of butter, a piece of furniture.* Ularning modeli a|an+NW (numerative word)+P (predlog)+UnCN. Ayrim birikmalar quyidagi tarkibda voqealanadi: Num+(N+s)+UncN=>Num+N+N (bu holatda o'zbek tilida KQ qo'shilmaydi) *two spoons of sugar* (ikki qoshiq shakar). So'zmaso'z tarjimaga yo'l qo'ymaslik uchun bu kabi birikmalar ma'lumotlar bazasida berilishi kerak.

Aniq *the* artikl belgilangan qoidalarga ko'ra qo'llaniladi, biroq ayrim hollarda ma'noga ta'sir etgan shakllarni ham inobatga olish zarur. Art+PN+N|Art+(PN+s)=>PN+lar (bu holatda artikl atoqli otga tegishli ismni, unga aloqador boshqa ismlarni ham nazarda tutadi). *The Johnson family / the Johnsons* (Jonson va uning oila

a'zolari). Shuningdek, Art+Adj+V=>Adj+lar (bu holatda sifatning oldidan aniq artikl ishlatilib, uning aniqlanishi bo'lmasa otlashib, ko'plik ma'nosida ishlatiladi). *The poor live in this village (Kambag'allar bu qishloqda yashaydilar).*

Kompyuterli morfologik tahlilda modellashtirish metodi ahamiyatga ega. Ayniqsa, sintaktik tahlilda erkin kontekstli daraxtsimon modellashtirish usuli komponentlarni ajratishda qulay. Model u yoki bu obyektни o'rganishga xizmat qiluvchi vositadir. Tabiiy tilni butunlay bir qolipga solish to'g'ri emas, biroq mashina tili uchun modellashtirish, tilni formal ko'rishga keltirish natijaning aniqligiga xizmat qiladi. So'zlarning semantik maydonlari, valentlik xususiyatlari ham inobatga olinsa, maqsadga muvofiq. Tahlilning to'g'ri olib borilishi dasturning algoritmik bosqichini belgilab beradi.

5.3. Morfologik analiz (morfoanalizator)

O'ttiz yildan oshiqroq davr mobaynida kompyuter morfologiyasi uchun turli texnologiyalarni yaratishga doir tadqiqotlar olib borilmoqda. "Ma'lumotni algoritmik qayta ishlashning lingvistik masalalari" (1983) monografiyasi mualliflarining fikricha: tillar uchun morfologik tahlil qilishning o'nlab algoritmlari ishlab chiqilgan. Algoritmning bir yo'nalishi bu so'z oxiridagi harfiy birikmalarni qamrab oluvchi ma'lumotlar bazasidan foydalanadi. Bunday lug'atda ehtimollik statistik qayta ishlash natijasida ma'lumotlar olinadi. Natijada morfologiyaning universal modeli hosil bo'ladi. Avtomatlashtirilgan morfologik tahlil turlari sifatida quyidagilarni ko'rsatish mumkin:

- Mantiqiy ko'paytirish metodi orqali morfologik analiz;
- Lug'atsiz jadval orqali morfologik analiz;
- So'zlar asosiga asoslangan morfologik analiz (So'zlar asosiga asoslangan morfologik analiz Yevropa tillari uchun moslashtirilgan. Unda qo'shimcha yordamchi jadvallardan foydalaniladi. Bunday analiz 32 blokdan iborat.);
- So'zshakllariga ko'ra morfologik analiz.

So'zshakllariga ko'ra morfologik tahlil qilish ham keng yoyilgan. Unga ko'ra so'z morfemalarga ajratiladi. Bunda quyidagi muammolar kuzatiladi: a) so'z lug'atdan topilmasligi mumkin. U

holda o'sha so'zning qaysi so'z turkumga tegishli ekanligi aniqlanadi. Chunki keyingi grammatik (sintaktik) analizni e'tibordan chetda qoldirmaslik kerak; b) agar har bir so'zshakl alohida leksik birlik bo'lsa, turli so'zshakllarini identifikatsiya qilish uchun boshqa zaruriy atributlar, so'zlarning o'zagi va affiksalar ro'yhati kiritilishi kerak. So'zshakllari orqali morfologik analiz qilish quyidagi operatsiyani amalga oshiradi:

-so'z shakl morfem segmentlarga ko'ra xuddi zanjir bo'lib birikadi;

-so'z shakl morfem segmentlar zanjiri sifatida tartiblanmagan yangi elementlar to'plami – morfemalarga almashtiriladi;

-so'z shakllar morfemalar to'plami sifatida ma'lumotlar yoziladi;

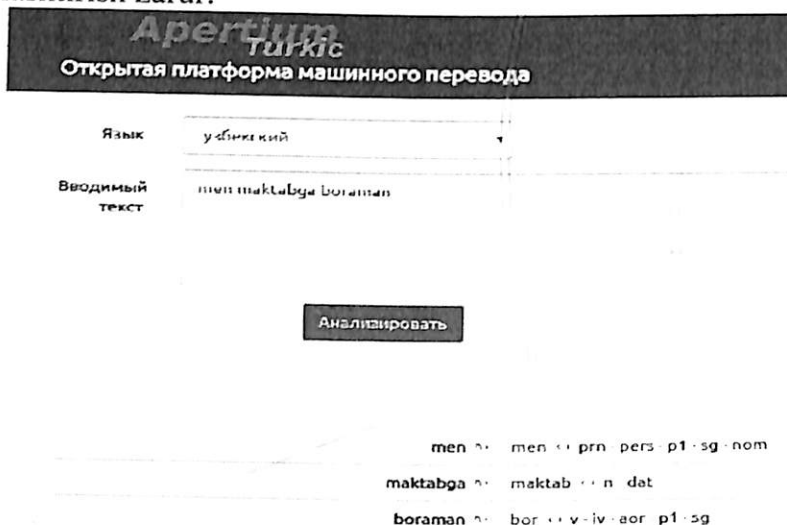
-bu ma'lumot so'z shakllari haqidagi yakuniy xulosa bilan qayta to'ldiriladi.

Mantiqiy ko'paytirish metodi orqali morfologik analiz qilish flektiv tillar uchun foydalaniladi.

Turkiy tillar bo'yicha morfologik analizator yaratish masalasi o'tgan asrning 60-yillaridan boshlab o'rganila boshlangan. O'sha davrdan boshlab ko'plab tadqiqotlar yuzaga chiqdi⁸¹. Xususan, tatar tili morfologik modeli bo'yicha analiz qilish 1990-yillarning o'rtalaridan boshlab amaliyotga tadbiiq etila boshlandi. O'sha davrda

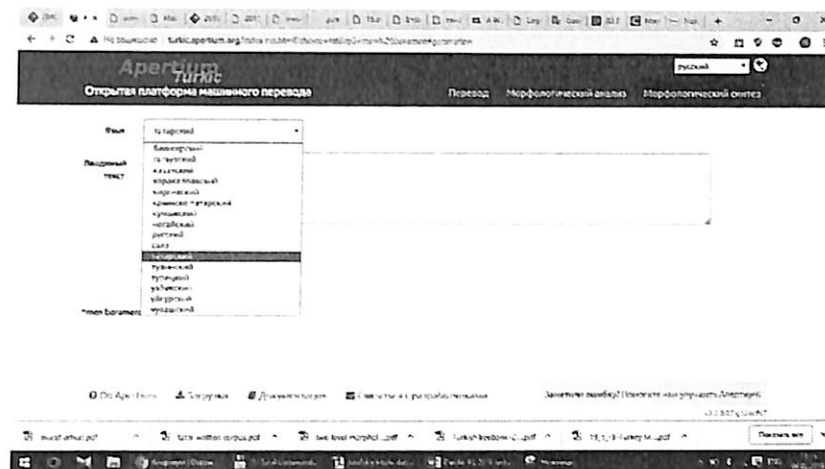
⁸¹ Kemal Ofi azer. Two-level Description of Turkish Morphology. Literary and Linguistic Computing, – Vol. 9, No 2, – 1994.; 2. Altintas K., Cicekli I. A morphological analyzer for Crimean Tatar // Proceedings of the 10th Turkish Symposium on Artificial Intelligence and Neural Networks (TAINN'2001). – 2001. – P. 180–189.; 3. Çağrı Çöltekin (2014) A Set of Open Source Tools for Turkish Natural Language Processing In: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14) Ed. by N. Calzolari et al. 1079–1086.; 4. G. Kessikbayeva and I. Cicekli, "Rule based morphological analyzer of Kazakh language," in Proceedings of the 2014 Joint Meeting of SIGMORPHON and SIGFSM. Baltimore, Maryland: ACL, June 2014, pp. 46–54.; 5. C. Tantug, E. Adali, K. Ofi azer. Computer Analysis of the Turkmen Language Morphology, 5th International Conference on NLP(FinTAL 2006), Turku, Finland, 186–193, 2006.; 6. M. Orhun, C. Tantug, E. Adali. Rule Based Analysis of the Uyghur Nouns, International Journal of Asian Lang. Proc.,19(1), 33–44, 2009.; Дыбо А.В., Шеймович А.В. Автоматический морфологический анализ для корпусов тюркских языков // Филология и культура. 2014. № 2 (36). С. 20–26; 6. Дыбо А.В., Шеймович А.В. Автоматический морфологический анализ для корпусов хакасского и древнетюркского языков // Научное обозрение Саяно-Алтая. Серия: Филология. № 8, 2014. С. 9–30.

tatar tilining morfologik analizatori uch modelga asoslangan: *generativ, paradigmatic va ikki bosqichli model*. Tatar tilining dastlabki morfologik analizatori ikki fayldan iborat bo'lgan. Birinchi fayl –qoidalar fayli (Rules), ikkinchi fayl- leksik birliklardan iborat leksikon (o'zak va affiksial morfemlardan tashkil topgan) hamda ularning morfotaktik qoidalarining izohi ham tavsiflangan. Leksikon o'z ichida ichki leksikondan iborat. Ichki leksikonning tuzilishi graflar bilan bog'liq. Ushbu analiz FST (finite state transducer), ya'ni avtomatik oxirgi bosqich transyutiri texnologiyasi orqali amalga oshirilgan⁸². Endilikda turkiy tillarni o'z ichiga olgan⁸³ morfologik analizator, sintezator (boshqird, gagauz, turk, qozoq, tatar, qoraqolpoq, qirg'iz, qrim-tatar, qo'miq, chuvash, o'zbek, uyg'ur, tuva, saxa, no'g'ay) va mashina tarjimasiga (boshqird, turk, qozoq, tatar, qoraqolpoq, qirg'iz, qrim-tatar, qo'miq) mo'ljallangan platformadan foydalanish mumkin. Ushbu platformada (16-rasm) to'liq morfologik analiz qilish imkoniyati chegaralangan, buni yana takomillashtirish zarur.



⁸² Д.Ш. Сулейманов, Р.А. Гильмуллин, Р.Р. Гатауллин Морфологический анализатор татарского языка на основе двухуровневой модели морфологии / Turklang. 2017, Kazan. -P. 8.

⁸³ <http://turkic.apertium.org>

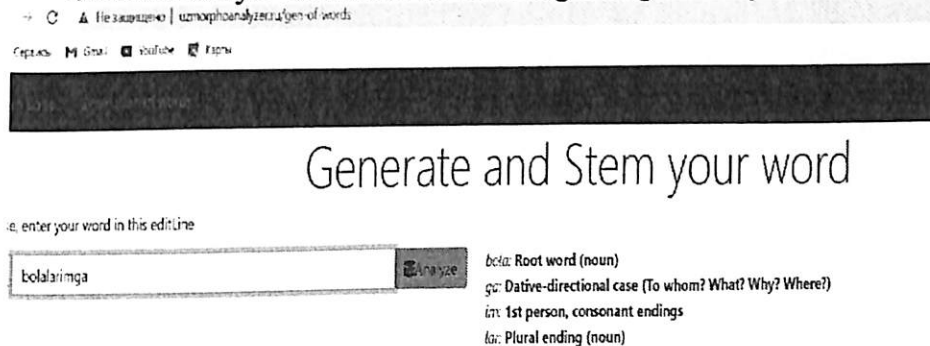


Компьютер орқали матрни (ог'зак ва yozma) анализ qilish quyidagi bosqichlarda olib boriladi:

- fonetik
- grafematik
- morfologik
- sintaktik
- semantik

Matnni normallashtirish jarayoni morfologik analiz bilan bog'liq bo'lib, quyidagi bosqichlarga ega bo'ladi: *tokenizatsiya->lemmatizatsiya->stemmizatsiya*.

- Tokenizatsiya – matndagi so'z shakllarini aniqlab olish;
- Lemmatizatsiya – so'zlarning lug'atdagi ifodasini aniqlab olish;
- Stemmizatsiya– derivativ so'zlarning o'zagini aniqlab olishi.



15-rasmda o'zbekcha so'zlarning morfologik analizi ko'rsatilgan⁸⁴. Ushbu dastur so'zni morfologik analiz qilishda chapdan o'ngga qarab tahlil qilish algoritmiga asoslangan.

Morfologik analiz sintaktik tahlil uchun muhim ahamiyatga ega. Chunki avtomatik tarzda matnni komponentlarga ajratishda u yoki bu so'zning qaysi so'z turkumiga tegishli ekanligini aniqlab olish zarur. Kompyuterlashtirilgan morfologiyaning (computational morphology) asosiy vazifasi "kiruvchi belgilar yoki fonemalardan iborat so'zlarning analiz holatida chiquvchiga yetkazish"⁸⁵.

Ot: bola+jon+lar+im+dagi+lar+niki+mas+mi+kan+a; sodda fe'l: o'qi+t + tir + ma + yot + gan + lig + i + ni; qo'shma fe'l: mashq qil+dir+ish+ayot+gan+lar, ko'makchi fe'lli so'z qo'shilmasi: ber+dir+tir+ib yubor+ma+yot+gan+dan+mi+kan+a va boshqalar. Inson uchun so'zlarning morfologik shakllanishi, ya'ni uni nutqda voqealanishi, so'z asosi va qo'shimchalarga ajratish tabiiy hodisa. Biroq so'zlarning avtomatik tahlilida bu protsedura sifatida qaraladi.

Kompyuter lingvistikasida olib borilgan ishlarda avtomatik tahlil quyidagi bosqichlarni o'z ichiga oladi:

- Frazalarni leksik qayta ishlash;

- Qo'shimchalarni identifikatsiyalash.

An'anaviy tilshunoslikda morfologik analizga "nima tasniflanadi?", degan savol qo'yilsa, kompyuter lingvistikasida matndagi so'zshakllar "qanday tahlil qilinadi?", degan masala ko'ndalang qo'yiladi. Ayrim manbalarda ko'rsatilishicha, morfologik analizning bir qancha yo'nalishlari mavjud. Unga ko'ra, so'z asosi va qo'shimchalarni tekshirish va so'zshakllarining ehtimollik holatiga ko'ra asos va qo'shimchalarni ajratish usuli analizning modellashtirishning klassik sxemasi bo'lib, so'zlardagi oxirgi harfiy birikmalardan iborat ma'lumotlarga asoslanadi. Ushbu ma'lumot lug'atni ehtimolga oid statistik ko'rsatkich bilan belgilaydi.

Flektiv tillarda so'zlarning tartibi gap bo'laklarini aniqlashda yordam beradi, agglyutinativ tillarda kelishik va ko'makchilar

sintaktik vazifani bajarishga xizmat qiladi. Xususan, qaratqich kelishigi qo'shimchasi hamisha qaratqich aniqlovchining grammatik vositasi: *kitobning muqovasi*; shu o'rinda, boshqa kelishik va qo'shimchalar hol va to'ldiruvchining qo'shimchalari bo'lib keladi: *mendan so'ralmoq (to'ldiruvchi) – tongdan (hol) yo'lga chiqmoq va boshq.*

Biroq morfologik analiz bu faqat u yoki bu kategoriyalarni tahlil qilish bilan cheklanmaydi. Yuqorida tilga olingan ba'zi muammoli holatlar uchun Rodolfe Delmonte leksikon uchun quyidagi lingvistik kategoriyalarni ko'rsatib o'tadi⁸⁶:

- grammatik kategoriyalar
- semantik kategoriyalar
- diskurs darajasidagi kategoriyalar
- sintaktik kategoriyalar
- yondashuvli (aspectual) kategoriyalar
- semantik konseptual kategoriyalar
- chegaralangan leksik birliklar
- grammatik cheklovlar

Bu tasniflash italyan tili uchun bo'lib, uning lingvistik tomonlari hisobga olingan.

Matnlarni tahlil qiluvchi avtomatik analizatorni yaratishda dastlabki bosqich sifatida o'zbek tilining elektron morfologik lug'atini yaratish masalasi qo'yiladi. Unga ko'ra, har bir so'zning qaysi so'z turkumining aniq statistikasi ko'rsatiladi. Har bir leksemaning muayyan morfologik xususiyatlari jadvallarda aks etadi. Undan keyin har bir leksemalar so'z turkumlari bo'yicha to'plamlarga birlashadi. To'plamlarga ajratishda ana shu leksemalarning umumiy semantik, qolaversa, formal jihatlari inobatga olinadi. To'plamlarga ajratilgan har bir so'zga qo'shilishi mumkin bo'lgan qo'shimchalar kombinatsiyasi aniqlanadi. Undan oldin barcha qo'shimchalar funksional belgilariga ko'ra guruhlarga ajratilib, ularning o'zbek tilidagi barcha bo'lishi mumkin bo'lgan mantiqiy bazasi shakllantiriladi.

⁸⁴ www.uzmorphoanalyzer.ru

⁸⁵ The Oxford handbook of computational linguistics (edited by Mitkov R.) – Oxford, 2003. – P. 63

⁸⁶R.Delmonte. Computational Linguistic Text Processing: Lexicon, grammar, Parsing and Anaphora Resolution, Nova Science Publishers, Inc. New York, 2008, 4-5

O'zbek tilining morfologik analizi uchun ikki bosqichli chekli avtomat (FST –two level morphology) texnologiyasi mos keladi.

PC-KIMMO kabi ikki bosqichli morfologik analizator instrumentlaridan foydalanuvchilar uchun qo'llash imkoniyati yaratilgan. FST texnologiyasi dastlab fin tili uchun Koskenniemi, ingliz tili morfologiyasi uchun Antvort tomonidan qo'llanilgan⁸⁷. Keyinchalik ikki bosqichli morfologik analizator yapon⁸⁸, koreys⁸⁹, turk⁹⁰, arab⁹¹ va mongol⁹² tillari doirasida foydalanilgan.

Chekli morfologik avtomat uchun quyidagi fayllar yaratilgan: 1) "Qoidalar-Rules" bunda alifbo, fonologik qoidalar va fonetik hodisaga uchraydigan maxsus fonemalar (bunda o'zbek tili uchun har ikki grafema asos qilib olindi: kirill va lotin); 2) Lug'at-lexicon (bunda barcha so'z turkumlarining o'zak va sodda yasama shakli kiritiladi). FST texnologiyadan foydalanib mavjud morfologik leksikon transyuterda, qoidalar ikki bosqichli qoidalar transyuterida (two-level rule transducer) kompilyatsiya qilinadi. So'ng kompilyatsiya qilingan leksikon va qoidalar birlashtiriladi. Buning uchun *compose* buyrug'idan foydalaniladi. Qoidalarga kiritilgan morfologik kategoriyalar uchun maxsus belgilangan teglar (Noun-ot, +Poss-egalik olmoshi, P-shaxs, +SG-birlik, +PL-ko'plik tarzida) *compose* qilingandan so'ng ular birlashtiriladi. Qoidalarni yakunlash uchun ifodalar so'ngida # belgisi qo'yiladi.

O'zbek tilining morfotaktik imkoniyatlari

Morfologik tahlilda morfotaktik holat birlamchi sanaladi (6-chizma). Chunki o'zakdan keyin qaysi qo'shimchalarning ketma-ket

⁸⁷ Antworth, E.L., 1990, PC-KIMMO: A Two-level Processor of Morphological Analysis, Summer Institute of Linguistics, Dallas, TX.

⁸⁸ Alam, Y.S., 1983, Two-level Morphological Analysis of Japanese, Texas Linguistics Forum 22, pp. 229-252

⁸⁹ Kim, D. B., Lee S. J., Choi, K.S., and Kim, G.C., 1994. A two-level morphological analysis of Korean. In Proceedings of the 15th conference on Computational linguistics - Volume 1 (COLING '94), pp. 535-539.

⁹⁰ Oflazer, K. 1994, Two-level description of Turkish morphology, Literary and Linguistic Computing, Literary and Linguistic Computing Volume9, Issue2 pp. 137-148.

⁹¹ Arabic Finite State Morphological Analysis and Generation, In COLING-96, Copenhagen, pp. 89-94.

⁹² Jaimai, P., Zundui, T., Chagnaa, A., and Ock, C.Y., PC-KIMMO-based Description of Mongolian Morphology, International Journal of Information Processing Systems Vol.1, No.1, 2005 pp. 41-48.

qo'llanilishi grammatik qoidaga asoslangan holda shakllantiriladi. O'zbek tilining morfotaktik imkoniyatlari quyidagicha bo'ladi:

Prefiks (1)+o'zak (2)+yasovchi qo'shimcha (3)+lug'aviy shakl yasovchi qo'shimcha (4)+ sintaktik shakl yasovchi qo'shimcha.

Tilning morfologik qoidalarini morfotaktik analiz qilishning ikki xil yondashuvi mavjud: 1) paradigmatic; 2) avtomatik.

FST texnologiyasi bo'yicha o'zbek tilining morfotaktik qoidalarini quyidagicha yozamiz:

LEXICON NumC

+SG: Poss1;

+PL:lar Poss2;

LEXICON NumV

+SG: Poss2 ;

+PL:lar Poss2;

LEXICON Poss1

+PP1+PSG:m Case ;

+PP2+PSG:ng Case ;

+PP3+PSG:si Case ;

+PP1+PPL:miz Case ;

+PP2+PPL:ngiz Case ;

+PP3+PPL:i Case ;

0:0 Case ;

LEXICON Poss2

+PP1+PSG:im Case ;

+PP2+PSG:ing Case ;

+PP3+PSG:i Case ;

+PP1+PPL:imiz Case ;

+PP2+PPL:ingiz Case ;

+PP3+PPL:i Case ;

0:0 Case ;

LEXICON Case

+NOM: PrePrefinal1 ;

+ACC:ning Final ;

+GEN:ni PrePrefinal2 ;
 +DAT:^Ga PrePrefinal1 ;
 +LOC:da PrePrefinal2 ;
 +ABL:dan PrePrefinal1 ;

LEXICON PrePrefinal1

0:0 Final ;
 +PART:mi Final ;
 +PART:ku Final ;
 +PART:^Ya Final ;
 +PART:-da Final ;
 +PART:-chi Final ;

LEXICON PrePrefinal2

0:0 Final ;
 +PART:mi Final ;
 +PART:ku Final ;
 +PART:^Ya Final ;
 +PART:-chi Final #

So'z yasovchi qo'shimchalarning stemmizatsiyasida ba'zi qo'shimchalarni aniqlash murakkab hisoblanadi: *yog'ingarchilik, zargarchilik, paxtachilik*. Har uch so'zning yasovchi qo'shimchasi – *garchilik*, biroq so'z yasovchi qo'shimcha sifatida –*gar, -chi, -lik* affikslari ham mavjud. Shu o'rinda so'zlarning ma'lumotlar bazasiga murojaat etiladi: *yog'in+garchilik, zar+gar+chilik, paxta+chi+lik*.

Ayrim so'zlarga sintaktik shakl yasovchi qo'shimchalar qo'shilganda o'zagida o'zgarish bo'luvchi so'zlar mavjud:

burun+im=>burnim- *tovush tushishi*
 shahar+im=>shahrim- *tovush tushishi*
 u+ga=>unga *tovush ortishi*
 kapalak+im=>kapalagim

Ushbu holatni morfologik analizda hal qilish mumkin. Buning uchun maxsus qoida orqali o'zagida o'zgarish bo'ladigan so'zlarning morfotaktik qoidalari FST tilida quyidagicha yoziladi:

Alphabet (Alifbo)

a b c d e f g h i j k l m n o p q r s t u v x y z ' %- б ч д ф г ғ ж
 қ м л н п р с ш т в х й з ца ээ и о у ё ю я ь

%%#:0
 %^Y:0 %^Y:y %^Y:й
 %^G:g %^G:k %^G:q %^G:ғ %^G:г %G:k
 %^K:k %^K:g %^K:г %^K:к
 %^Q:q %^Q:g %^Q:ғ %Q:к
 %^A:a %^A:0 %^A:a
 %^U:u %^U:0 %^U:y
 %^I:i %^I:0 %^I:и
 %^V:a %^V:0 %^V:a
 %^W:i %^W:0 %^W:и
 %^P:a %P:y %^P:a %^P:й ;

Sets (to'plam)

LatVowel = a i e o u ;
 CyrVowel = а е и о э ё ю я ў ;
 LatCons = b c d f g h j l m n p r s t v x z k q ;
 LatConsMinus = b c d f g h j l m n p r s t v x z ;
 CyrCons = қ б ч д ф г ғ ж м л н п р с ш т в х й з ц к ;
 CyrConsMinus = б ч д ф г ғ ж м л н п р с ш т в х й з ц ;
 LatLetter = a i e o u b c d f g h j l m n p r s t v x z k q ;
 CyrLetter = а е и о э ё ю я ў қ б ч д ф г ғ ж м л н п р с ш
 т в х й з ц к ;

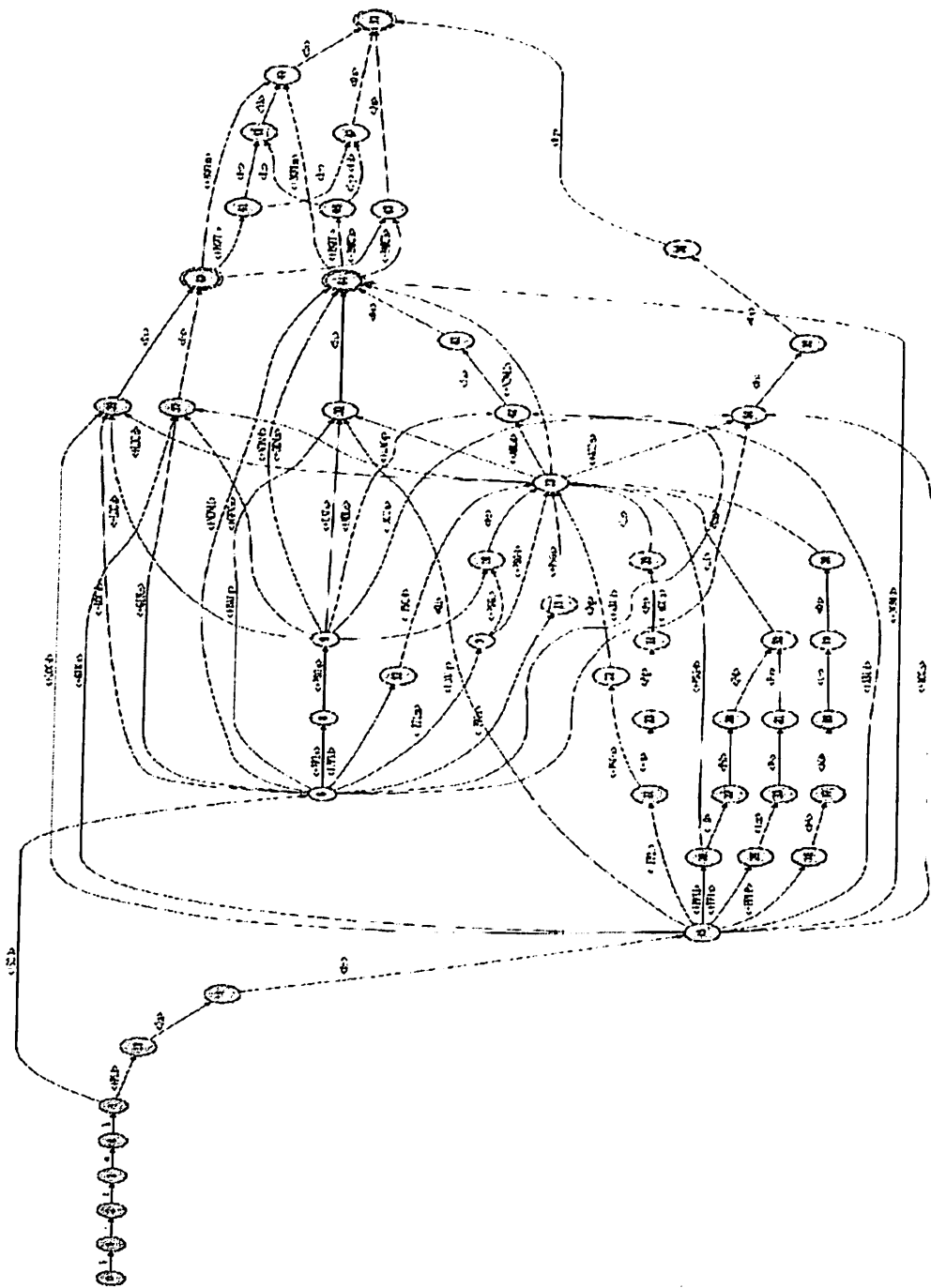
Rules (qoidalar)

"Y is dropped after consonant"
 %^Y:0 <=> [:LatCons | :LatCons] [:']* [:%-]* _ ;
 "Y remains after Latvowels"
 %^Y:y <=> :LatVowel [:']* [:%-]* _ ;
 "Y remains after Cyrvowels"
 %^Y:й <=> :CyrVowel [:']* [:%-]* _ ;

Kitob so'zi misolida uning grafigi quyidagi ko'rinishda bo'ladi (7-chizma):

O'zbek tilidagi matnlar ham kiril, ham lotin grafikasida bo'lgani bois bizning morfotaktik qoidalar uchun yoziladigan qoidalarimizda fonemalar uchun har ikki grafema hisobga olingan. Ushbu morfologik analiz sintaktik parsing uchun ahamiyatlidir.

7-chizma



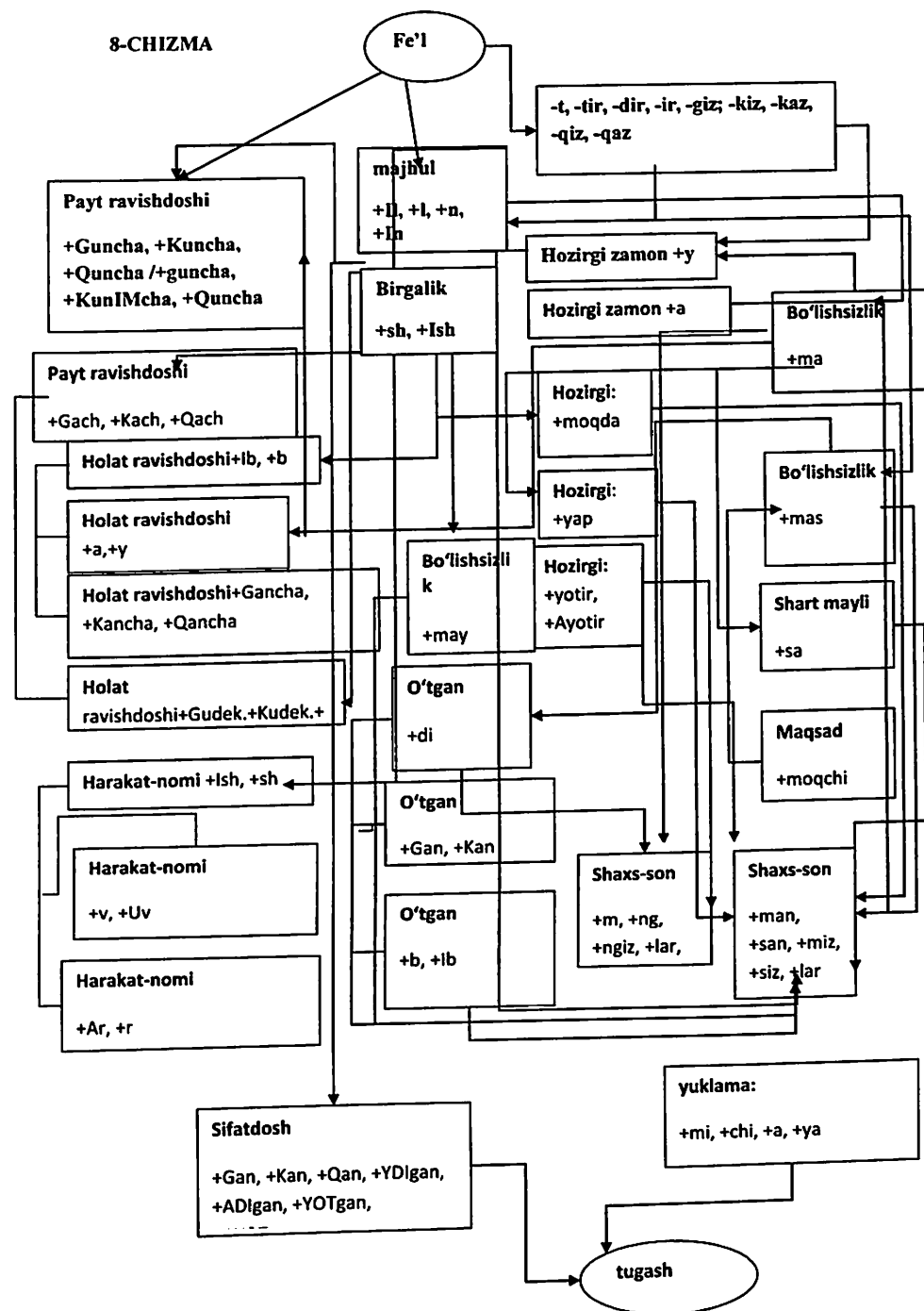
Ushbu holatda matndagi har bir komponent soʻz turkumi sifatida qaraladi. Soʻzlar gapda sintaktik aloqaga kirishar ekan grammatik jihatdan shakllanadi. Buning uchun soʻzlarning lugʻatdagi shaklini aniqlashda lemmatizator kerak boʻladi. Erishilgan natijalarning ayrimlarini namuna sifatida keltiramiz:

- yoz+NEG+PRES+P1+PL+PART:yozmayapmizku
- yoz+NEG+PRES+P1+PL+PART:yozmayapmizmi
- yoz+NEG+PRES+P1+PL+PART:yozmayapmiz-da
- yoz+NEG+PRES+P1+PL+PART:yozmayapmiz-a
- yoz+NEG+PRES+P1+SG:yozmayapman
- yoz+NEG+PRES+P1+SG+PART:yozmayapmanku
- yoz+NEG+PRES+P1+SG+PART:yozmayapmanmi
- yoz+NEG+PRES+P1+SG+PART:yozmayapman-da
- yoz+NEG+PRES+P1+SG+PART:yozmayapman-a
- yoz+NEG+PRES+P3+PL:yozmayapsizlar
- yoz+NEG+PRES+P3+PL+PART:yozmayapsizlarku
- yoz+NEG+PRES+P3+PL+PART:yozmayapsizlarmi
- yoz+NEG+PRES+P3+PL+PART:yozmayapsizlar-da
- yoz+NEG+PRES+P3+PL+PART:yozmayapsizlar-a
- yoz+NEG+PRES+P3+SG:yozmayapti
- yoz+NEG+PRES+P3+SG+PART:yozmayaptiku
- yoz+NEG+PRES+P3+SG+PART:yozmayaptimi
- yoz+NEG+PRES+P3+SG+PART:yozmayapti-da
- yoz+NEG+PRES+P3+SG+PART:yozmayapti-ya
- yoz+NEG+PRES+P2+PL:yozmayapsiz
- yoz+NEG+PRES+P2+PL+PART:yozmayapsizku
- yoz+NEG+PRES+P2+PL+PART:yozmayapsizmi
- yoz+NEG+PRES+P2+PL+PART:yozmayapsiz-da
- yoz+NEG+PRES+P2+PL+PART:yozmayapsiz-a
- yoz+NEG+PRES+P2+SG:yozmayapsan
- yoz+NEG+PRES+P2+SG+PART:yozmayapsanku
- yoz+NEG+PRES+P2+SG+PART:yozmayapsanmi
- yoz+NEG+PRES+P2+SG+PART:yozmayapsan-da
- yoz+NEG+PRES+P2+SG+PART:yozmayapsan-a
- yoz+NEG+PAST+DEF+P2+PL:yozmadingiz
- yoz+NEG+PAST+DEF+P2+PL+PART:yozmadingizku
- yoz+NEG+PAST+DEF+P2+PL+PART:yozmadingizmi

yoz+NEG+PAST+DEF+P2+PL+PART:yozmadingiz-da
 yoz+NEG+PAST+DEF+P2+PL+PART:yozmadingiz-a
 yoz+NEG+PAST+DEF+P2+SG:yozmading
 yoz+NEG+PAST+DEF+P2+SG+PART:yozmadingku
 yoz+NEG+PAST+DEF+P2+SG+PART:yozmadingmi
 yoz+NEG+PAST+DEF+P2+SG+PART:yozmading-da
 yoz+NEG+PAST+DEF+P2+SG+PART:yozmading-a
 yoz+NEG+PAST+DEF+P1+SG:yozmadim
 kitob+SG+NOM:kitob
 kitob+SG+NOM+PART:kitobku
 kitob+SG+NOM+PART:kitobmi
 kitob+SG+NOM+PART:kitob-da
 kitob+SG+NOM+PART:kitob-a
 kitob+SG+NOM+PART:kitob-chi
 kitob+SG+GEN:kitobni
 kitob+SG+GEN+PART:kitobniku
 kitob+SG+GEN+PART:kitobnimi
 kitob+SG+GEN+PART:kitobni-chi
 kitob+SG+GEN+PART:kitobni-ya
 kitob+SG+ACC:kitobning
 kitob+SG+DAT:kitobga
 kitob+SG+DAT+PART:kitobgaku
 kitob+SG+DAT+PART:kitobgami
 kitob+SG+DAT+PART:kitobga-da
 kitob+SG+DAT+PART:kitobga-ya
 kitob+SG+DAT+PART:kitobga-chi
 kitob+SG+ABL:kitobdan
 kitob+SG+ABL+PART:kitobdanku
 kitob+SG+ABL+PART:kitobdanmi
 kitob+SG+ABL+PART:kitobdan-da
 kitob+SG+ABL+PART:kitobdan-a
 kitob+SG+ABL+PART:kitobdan-chi
 kitob+SG+LOC:kitobda

8-chizmada fe'l so'z turkumining grammatik qo'shimchalarini qo'shilishi algoritmi tasvirlangan. Muayyan miqdordagi suffikslar birgalikda qo'shimchalar kombinatsiyasini hosil qiladi.

8-CHIZMA



O'ZBEK TILINING DERIVATIV IMKONIYATLARI

Ma'lumotlar bazasida o'zbek tiliga oid resursning kamligi o'zbek tili morfologiyasida ayrim muammolarni keltirib chiqaradi. Kontekstda ishtirok etayotgan fe'llarni to'g'ri morfemlarga ajratish uchun ularning tasnifi to'g'ri tuzilgan bo'lishi kerak. Grammatik affikslar analiz qilinganda so'zlardagi derivativ imkoniyatlar ham inobatga olinishi kerak. Quyidagi jadvalda fe'l so'z turkumini yasashda ishtirok etuvchi o'zak+affiks larning jadvali keltirilgan:

Ot+	-a =>sana, -an =>kuchan, -i=>ranji, -ik=>ko'zik, -ir=>gapir, -y=> kuchay, -ka=>iska, -la=>gulla, -lan=>faxrlan, -lash=>ommalash, -lashtir =>sahnalashtir, -sit=>aybsit, -sira=>suvsira, -iq =>yo'liq, -g'ar=> jamg'ar, -qar =>boshqar
Sifat+	-a=>qiyina, -i=>tinchi, -ay=>toray, -la =>maydala, -lan=>shodlan, -lash =>osonlash, -lat=> -lashtir=>soxtalashtir, -r=>qisqar, -ar =>oqar, -si =>garangsi, -sin =>yotsin, -sira=>begonasira, -t=>to'lat, -it=>berkit, -iq=>namiq
Son+	-ik=>birik, -lan=>ikkilan, -lash=>birlash
Olmosh+	-la =>sizla, -si =>mensi, -sira=>sensira
Ravish+	-ik=>kechik, -ir=>ko'pir, -ay=>ko'pay, -la=>tezla, -lash=>birgalash, -sit=>kamsit, -chi=>ko'pchi
Taqlid so'z +	-a=>shildira, -illa =>guvilla, -ur=>tupur, -ira=>yaltira, -la=>gumburla, -ra=>ma'ra, -shi=>g'ingshi, qir=>hayqir
Modal so'z+	-la=>yo'qla, -ol =>yo'qol, -ot=>yo'qot
+modal qo'shimchalar+	-imsira=>kulimsiramoq, -inqira=>oqarinqiramoq, -kila=>tepkilamoq, -qila=>chopqilamoq, -gila=>yugurgilamoq, -g'ila=>ezg'ilamoq, -ish=>to'lishmoq, -q=>tutaqmoq, -iq=>toliqmoq, -k=>junjikmoq, -ik=>ko'nikmoq, -la=>savalamoq, -ala=>quvalamoq, -qi=>yulqimoq, -g'i=>to'zg'imoq, -a=>buramoq

So'z turkumlarining 56 turdagi lug'aviy shakl yasovchi qo'shimchalari mavjud. Ma'lumotlar bazasida 80000 ta birlik bo'lib, ularning grammatik kategoriyalari ichki kategoriyalarga ajratilgan.

Ba'zi qo'shimchalar omonim sanaladi. Ular ham ot, sifat va fe'l yasash xususiyatiga ega. Shu bois har bir qo'shimchani morfologik modelini aniqlash orqali bunday muammoni hal qilish mumkin. Shuningdek, ayrim shakldosh leksemalar omonimiyani hosil qilib, ularning hech qanday qo'shimchasiz kelgan holatini aniqlashda sintaktik model yoki morfologik modelga tayaniladi: *och: och rang -qorin och*. Ushbu so'z qo'shma so'zlarning yoki iboralarning ham tarkibi bo'lib kelishi ham mumkin.

Ishlahani och +ib {ber, bo'l, chiq, ket, ko'r, qo'y, tashla} +a {bil, boshla, ol}

Ko'gilni och+ib {ber, ko'r, o'tir, qo'y, tashla, yubor} +a {ol}

FST texnologiyasidagi muammolardan biri ko'p tarkibli so'zlar, chunki ular alohida birlik sifatida analiz qilinadi. Kompyuter tahlili uchun satr (so'z) oxiridagi element so'zning tugallanganligini bildiradi.

Sirni och, yo'l och, fol och, gul och ushbu so'zlar yagona birliklardir, biroq avtomatik tahlilda ular alohida birliklar sifatida analiz qilinadi.

Morfologik analiz borasida olimlar o'zlarining ilmiy qarashlarini bildirganlar: Cerstin Mahlov, Michael Piotrovski⁹³ qo'shimchalar kombinatsiyasi chegarasini to'rt yondashuv asosida belgilaydi: tub so'z, affiks, stem, bevosita yondashuv.

O'tgan asrning 90-yillarida morfologik analiz qilish uchun modelga asoslangan: *generativ model, paradigmatic model, ikki tarkibli morfologik model*⁹⁴. So'z turkumini aniqlovchi morfoanalizator uchun dastlabki algoritmlar ikki tarkibli tuzilishga

⁹³ Krister Lindén, Miikka Silfverberg, and Tommi Pirinen HFST Tools for Morphology – An Efficient Open-Source

⁹⁴ Cerstin Mahlov, Michael Piotrowski (eds.). JSLIM – Computational Morphology in the Framework of the SLIM Theory of Language / State of the Art in Computational Morphology. – Zurich, 2009. –P. 15.

asoslangan⁹⁵. Birinchi bosqich har bir soʻzning turkumini aniqlash boʻlsa, ikkinchi bosqich ajratilgan har bir soʻzning grammatik qoidalariga tuzatishlar kiritish hisoblangan. Oʻzbek tili uchun maʼlumotlar bazasi juda muhim sanaladi. Shuni unutmaslik kerakki, morfoanaliz faqat shakl bilan cheklanib qolmasligi kerak. Kompyuter lingvistikasida chuqur tahlil qilish metodi semantik analiz deb ataladi. Morfoanalizator quyidagi ikki bosqichdan iborat boʻladi:

- Soʻzlarning asosini aniqlash;
- Soʻzlarning soʻz turkumini aniqlash;
- Asosga koʻra har bir affiksni tahlil qilish. Bu jarayon chapdan oʻngga yoki aksincha boʻlishi mumkin;
- Har bir qoʻshimchani turini aniqlash.

Soʻzlarni maʼlumotlar bazasidan oʻqib olish uchun ularning simvollar ketma-ketligiga amal qilinadi. Toki oʻsha soʻz yaxlit birlik sifatida aniqlanmagunicha bu holat takrorlanaveradi. Morfologik analizi uchun komponent tahlil metodidan ham foydalanish mumkin. Unga koʻra soʻzning uzunligi aniqlanadi. Shu tarkibdagi simvollar soniga qarab maʼlumotlar bazasidan qidiriladi. Toki oʻsha soʻz bazadan aniqlanmagunicha tekshirish davom etaveradi.

FST texnologiyasi orqali oʻzagida qoʻshimcha qoʻshilishi natijasida oʻzgarish boʻluvchi soʻzlarning ham modellari va qoidalari kiritildi. Ushbu texnologiya orqali matnlarni tahrirlovchi dastur va kompyuter lingvistikasining morfologik analiz qilishi kerak boʻlgan masalalarida unumli foydalanish mumkin.

Avtomatik tahlil qilishning quyidagi turlari mavjud⁹⁶:

- Soʻzlarning lugʻatdagi asoslarga koʻra analiz qilish
- Soʻzlarning lugʻatdagi soʻzshakllariga koʻra analiz qilish
- Mantiqiy koʻpaytirish metodi orqali analiz qilish
- Lugʻatsiz jadvallar yordamida analiz qilish

Morfologik tahlil qiluvchi avtomatik analizatorning amaliy ahamiyati:

1. Avtomatik tahrirda har bir soʻzning orfografik xatosini tekshirish va soʻzlardan foydalanish;

2. Oʻzbek tilini oʻqitishda oʻrta va Oliy taʼlim boʻgʻinlarida elektron doska va slaydlardan foydalangan holda taʼlim tizimini yanada optimallashtirish;

3. Avtomatik tarjimada u yoki bu tildagi matni oʻzbek tiliga tarjima qilishda gap strukturalarni sintaktik boʻlaklarga ajratishda har bir komponentning qiymati ana shunda berilgan bazada oʻz ifodaini topadi, bu esa tarjima jarayonini yengillashtiradi;

4. Chet ellik talabalar uchun avtomatik oʻqitish tizimini yaratishda muhimligi bilan ajralib turadi.

Amalga oshirilgan ishlarning tahminiy natijasi “B” harfi misolida koʻrsatish mumkin:

<i>Soʻz turkumi</i>	<i>Leksema (namuna)</i>	<i>Statistikasi</i>
<i>ot</i>	Badantarbiya	1460
<i>feʼl</i>	Bastala	279
<i>sifat</i>	Baadab	750
<i>ravish</i>	Baayni	200
<i>son</i>	besh	22
<i>olmosh</i>	Bari	26
<i>bogʻlovchi</i>	balki	10
<i>Yuklama</i>	birgina	4
<i>koʻmakchi</i>	badalida	27
<i>taqlid soʻz</i>	bidir-bidir	20
<i>undov soʻz</i>	borakallo	15
<i>modal soʻz</i>	bilishimcha	4
<i>qoʻshimcha</i>	ba-	8

Soʻzlarni avtomatik morfologik tahlilda biz quyidagi usuldan foydalanamiz. Har bir leksema modellashtirilgan jadvalda aks etadi. Unda soʻzlarning asosi va unga qoʻshimcha jadvallar kiritilgan boʻladi, bu qoʻshimcha jadvallarda soʻzshakllarining istisnoli holatlari aks etadi (EQ-egalik qoʻshimchasi, KQ-kelishik qoʻshimchasi):

-Qoʻshimchalar qoʻshilganda fonetik hodisalar sodir boʻluvchi barcha soʻzlarning roʻyxati: *shahar+EQ=>shahrim, men+KQ=>meni (tovush tushushi); soʻroq+EQ=>soʻrogʻi (tovush almashinishi); u+KQ=>unga tovush orttirilishi;*

-Omonim soʻzlar roʻyxati

-Turgʻun birikmalar va iboralar roʻyxati

-Asosiy qoʻshimchalar va ularning boʻlishi mumkin boʻlgan

⁹⁵ Harris, 1962; Klein and Simmons, 1963; Greene and Rubin, 1971

⁹⁶ Марчук Ю. Компьютерная лингвистика. – М., 2007. – В. 65.

barcha kombinatsiyalari ro'yxati

-So'z birikmasiga omonimlik hosil qiluvchi qo'shma so'zlar va iboralar ro'yxati: *boshiga ko'tarmoq* (o'z ma'nosida), 1) *shovqin solmoq*; 2) *hurmat qilmoq* kabilar kiritiladi.

Morfologik analiz uchun dunyo kompyuter lingvistikasida o'nlab algoritmlar yaratilgan. Bizning texnologiyamizda esa tahlil chapdan o'ngga qarab olib boriladi. O'zak morfema ma'lumotlar bazasidan olinadi. Bu bosqich namunasi quyidagi piton dasturlash tili orqali namunada keltirilgan:

```
k=1
for i in range(0, len(word)):
    if(otlar.__contains__(word[0: i+1])):
        k=i+1
print(word[0: k])
word=word[k:]
k=10
while(len(word)>0):
    if(qoshOtYas.__contains__(word[0:k])):
        print(word[0:k])
        word=word[k:]
        if(len(word)>10):
            k=10
        else:
            k=len(word)
    elif(qoshimchalarOt.__contains__(word[0:k])):
        print(word[0:k])
        word = word[k:]
        if (len(word) > 10):
            k = 10
        else:
            k = len(word)
```

NATIJA: *BOLAJONLARIMGAMI => bola*

jon
lar
im
ga
mi

Foydalanilgan adabiyotlar:

1. Jurafskiy D. Speech and language processing. 2007. – P. 4
2. Mitkov R. The Oxford handbook of computational linguistics. 2010. –P. 62
3. Raül Canals, Anna Esteve, Alicia Garrido et.al. , interNOSTRUM: A SpanishCatalan Machine Translation System, Machine Translation Review, Issue No. 11, December (2000) pp. 21-25
4. Krister Lindén, Miikka Silfverberg, and Tommi Pirinen HFST Tools for Morphology – An Efficient Open-Source
5. Package for Construction of Morphological Analyzers / – Computational Morphology in the Framework of the SLIM Theory of Language / State of the Art in Computational Morphology. – Zurich, 2009. – P. 30
6. Cerstin Mahlow, Michael Piotrowski (eds.). JSLIM – Computational Morphology in the Framework of the SLIM Theory of Language / State of the Art in Computational Morphology. – Zurich, 2009. –P. 15.
7. D. Suleymanov, R. Gilmullin, R. Gataullin Morphological analysis system of the Tatar language based on the two-level morphological model / Turklang 2017. Kazan, 2017. pp. 6-26.
8. Абдурахмонова Н.З. Англо-узбекский словарь фразеологический глаголов. – Toshkent: Turon zamin ziyo. 2017
9. Хомоненко А.Д., Цыганков В.М., Мальцев М.Г. Базы данных: Учебник для высших учебных заведений / Под ред. проф. А.Д. Хомоненко. Издание второе, дополненное и переработанное. СПб.: КОРОНА принт, 2006. 672 с.
10. Финаев В.И, Пушкин А.В.. Информационное обеспечение систем управления. Таганрог: Изд-во ТРТУ, 2001. –С. 91.

barcha kombinatsiyalari ro'yxati

-So'z birikmasiga omonimlik hosil qiluvchi qo'shma so'zlar va iboralar ro'yxati: *boshiga ko'tarmoq* (o'z ma'nosida), 1) *shovqin solmoq*; 2) *hurmat qilmoq* kabilar kiritiladi.

Morfologik analiz uchun dunyo kompyuter lingvistikasida o'nlab algoritmlar yaratilgan. Bizning texnologiyamizda esa tahlil chapdan o'ngga qarab olib boriladi. O'zak morfema ma'lumotlar bazasidan olinadi. Bu bosqich namunasi quyidagi piton dasturlash tili orqali namunada keltirilgan:

```
k=1
for i in range(0, len(word)):
    if(otlar.__contains__(word[0: i+1])):
        k=i+1
print(word[0: k])
word=word[k:]
k=10
while(len(word)>0):
    if(qoshOtYas.__contains__(word[0:k])):
        print(word[0:k])
        word=word[k:]
        if(len(word)>10):
            k=10
        else:
            k=len(word)
    elif(qoshimchalarOt.__contains__(word[0:k])):
        print(word[0:k])
        word = word[k:]
        if (len(word) > 10):
            k = 10
        else:
            k = len(word)
```

NATIJA: *BOLAJONLARIMGAMI => bola*

*jon
lar
im
ga
mi*

Foydalanilgan adabiyotlar:

1. Jurafskiy D. Speech and language processing. 2007. – P. 4
2. Mitkov R. The Oxford handbook of computational linguistics. 2010. –P. 62
3. Raül Canals, Anna Esteve, Alicia Garrido et.al. , interNOSTRUM: A SpanishCatalan Machine Translation System, Machine Translation Review, Issue No. 11, December (2000) pp. 21-25
4. Krister Lindén, Miikka Silfverberg, and Tommi Pirinen HFST Tools for Morphology – An Efficient Open-Source
5. Package for Construction of Morphological Analyzers / – Computational Morphology in the Framework of the SLIM Theory of Language / State of the Art in Computational Morphology. – Zurich, 2009. – P. 30
6. Cerstin Mahlow, Michael Piotrowski (eds.). JSLIM – Computational Morphology in the Framework of the SLIM Theory of Language / State of the Art in Computational Morphology. – Zurich, 2009. –P. 15.
7. D. Suleymanov, R. Gilmullin, R. Gataullin Morphological analysis system of the Tatar language based on the two-level morphological model / Turklang 2017. Kazan, 2017. pp. 6-26.
8. Абдурахмонова Н.З. Англо-узбекский словарь фразеологический глаголов. – Toshkent: Turon zamin ziyo. 2017
9. Хомоненко А.Д., Цыганков В.М., Мальцев М.Г. Базы данных: Учебник для высших учебных заведений / Под ред. проф. А.Д. Хомоненко. Издание второе, дополненное и переработанное. СПб.: КОРОНА принт, 2006. 672 с.
10. Финаев В.И, Пушкин А.В.. Информационное обеспечение систем управления. Таганрог: Изд-во ТРТУ, 2001. - С. 91.

Savol va topshiriqlar:

1. Morfologik analiz grammatik kategoriyalarni tahlil qilish bilan bir qatorda yana qanday vazifalarni bajara oladi?
2. Leksemalarni to'plamga ajratishda ularning qaysi jihatlari inobatga olinadi?
3. Regular ifodalar so'ngida tugallanganlik belgisi qanday ifodalanadi?
4. Quyidagi strukturaga misollar keltiring va yana boshqa holatlarda qo'llang: Verb+NEG+PRES+P1+PL+PART:
5. Avtomatik tahlilning turlarini ayting?
6. So'zlarni avtomatik morfologik tahlilda istisnoli holatlarni keltiring.
7. Tilning uch til tipologiyasi bo'yicha grammatik kategoriyalarni to'ldiring:

Grammatik kategoriya	Inflektiv tillar	Agglyutinativ tillar	Isolayativ tillar
Case (nominative, genitive, dative, ablative, locative, instrumental, accusative, vocative)			
Gender (masculine, feminine, neuter)			
Person (first, second, third)			
Number (singular, dual, plural)			
Mood (Indicative, subjunctive, optative, injunctive, imperative)			
Voice (active, passive)			
Aspect (Tense) Present, past, future			

5-mavzu yuzasidan test

1. Kompyuter lingvistikasida olib borilgan ishlarda avtomatik tahlil qaysi bosqichlarni o'z ichiga oladi?
 - a) frazalarni leksik qayta ishlash, qo'shimchalarni identifikatsiyalash
 - b) sintaktik tahlil
 - c) morfologik tahlil
 - d) etimologik tahlil
2. Tokenizatsiya nima?
 - a) matndagi so'z shakllarini aniqlab olish
 - b) so'zlarning lug'atdagi ifodasini aniqlab olish
 - c) derivative so'zlarni aniqlab olishi
 - d) matndan qo'shma so'zlarni topish
3. Lemmatizatsiya nima?
 - a) so'zlarning lug'atdagi ifodasini aniqlab olish
 - b) matndagi so'z shakllarini aniqlab olish
 - c) matndan qo'shma so'zlarni toppish
 - d) matni sintaktik tahlil qilish
4. Infleksional kategoriya qanday vazifani bajaradi?
 - a) so'zlarning grammatik ma'nosiga ta'sir etadi, so'zning leksik ma'nosiga ta'sir qilmaydi
 - b) yangi so'z yasashga xizmat qiladi
 - c) so'zning grammatik va leksik ma'nosiga ta'sir qiladi
 - d) so'zning ma'nosini kuchaytiradi
5. Izolayativ tillarning (mandarin, xitoy tili) o'ziga xos xususiyati nima?
 - a) biror chegara va affiksga ega bo'lmaydi
 - b) so'zning muayyan shakli affikslar bilan ifodalanib, ularning har biri morfologik paradigamaga tegishli bo'ladi.
 - c) biror belgisiga ko'ra farqlanuvchi tillar
 - d) asosga qo'shimcha qo'lganda o'zgaradi
6. Qaysi tillarda so'zning muayyan shakli affikslar bilan ifodalanib, ularning har biri morfologik paradigamaga tegishli bo'ladi?
 - a) uyg'ur-finland va turkiy tillar
 - b) xitoy, mandarin
 - c) hind-yevropa tillari

d) yapon, xitoy
7. Rodolfe Delmonte leksikon uchun nechta lingvistik kategoriyalarni farqlaydi?

- a) 8
- b) 9
- c) 4
- d) 6

8. Kompyuter morfologiyasi nima bilan shug'ullanadi?

a) so'z va uning shakllarini ham grafematik (yozma) ham fonemik (og'zaki) shaklini tahlil qilish bilan shug'ullanadi

- b) so'zlarni ma'nosini tahlil qilish
- c) so'zlarning tarkibini tahlil qiladi

d) gap tarkibidan yordamchi so'zlarni topadi

9. Tarjimadan avvalgi tahlil jarayonida qaysi asosiy iyerarxik bosqichlar amalga oshiriladi?

- a) matn-jumla-leksema shakl
- b) harf-so'z birikmasi- gap
- c) so'z- morfema- qo'shma so'z
- d) gap-matn-jumla

10. «Kompyuter lingvistikasining boshlanishi» asarining muallifi kim?

- a) Yu.I.Shemakin
- b) Chomskiy
- c) Rodelf Delmonte
- d) Mitkov

6-MAVZU. SINTAKTIK TAHLILNING AVTOMATIK TIZIMI

6.1. Sintaktik strukturalar (modellar)

6.2. Diskurs va matn segmenti

6.3. Sintaktik parsing (tahlil)

Tayanch so'z va iboralar: *Sintaktik strukturalar (modellar), diskurs va matn segmenti, diskurs analizi, so'z birikmalari, tobe bog'lanish, korpus analizi, sintaktik parsing, daraxtsimon qayta qurish, semantik tahlil, CoNLL-U format*

6.1. Sintaktik strukturalar (modellar)

Sintaksis tilning sintaktik strukturasini o'rganadi. Gaplardagi muayyan segmentlar sintaktik qoidalarga ko'ra qismlarga ajraladi. Sintaktik qoidalar formal tilda o'z ifodasini topadi. Sintaksis so'z birikmasi va gap sintaksisiga ajraladi. So'z birikmasi sintaktik birlik sifatida gap qurilishining muhim materiali hisoblanadi. Shuning uchun uning nutq birligi sifatida o'rganilishi muhim ahamiyat kasb etadi. So'z birikmalarini o'rganish masalasi xorijda XVIII asrning boshlarida tilshunoslar tomonidan kun tartibiga qo'yilgan. Xususan, ingliz lingvisti B.Djonson o'zining tadqiqotlarida "so'z birikmalari" terminidan foydalanmagan holda ikkita so'z yig'indisi so'z birikmasini hosil qiladi, deb hisoblagan. XX asrda sintaksis bo'yicha ko'p ishlar amalga oshirilgan⁹⁷. G'arb olimlari ikki yoki undan ortiq

⁹⁷ Шутникова Р.С. К вопросу о сочетаемости прилагательных в современном английском языке. –М., 1966. –С.167., Ханаху, Данна Руслановна Атрибутивные словосочетания в русском и английском языках сопоставительно-типологический анализ): Дис. ... канд. филол. наук: 10.02.01, 10.02.20 Майкоп, 2007. –С. 166.; Кунин, А.В.: Фразеология современного английского языка. –М.: Международные отношения. 1972. – 47 с., Арбекова Т.И. Лексикология англ. яз.(практический курс) учебное пособие для 2-3 курсов ин-ов и фак. Иностр. Уз. –М., «Высшау школа», 1977. –С. 53; Швейцер А. Д. Литературный английский язык в США и Англии. – М.: Высшау школа, 1971. – 200 с.; Шитова Л., Брускина Т. Англо-русский словарь идиом и фразовых глаголов Белуаева Т.М. и др. Английский язык за пределами. – Англии, 1961; Галперин И.Р. Очерки по стилистике английского языка. –М., 1985. – С. 61-104; Бурлакова В.В. Синтаксические структуры современного английского языка. –М.: Просвещение, 1984. –С. 25-41; И.П.Иванова, В.В.Бурлакова, Г.Г.Почепцов Теоретическая грамматика современного английского языка. – М., 1981. – С.138; Гинзбург, С. С. Хидекел, Г. Ю. Кнуазева и А. А. Санкин.

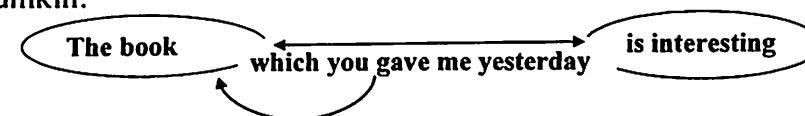
so'zlarning birikuvini *fraza (phrase)* termini bilan ataganlar. Unda turg'un birikmalar va so'z birikmalar bir-biridan ajratilmagan. Bu sohada V.V. Vinogradov so'z birikmasi gapning qurilish materiali sifatida kommunikativ birlik ekanligini ta'kidlagan. Ingliz tilida so'z birikmasi "word collocation", "words that go together", "word combinations" terminlari bilan yuritiladi. Ingliz tili so'z birikmalari lug'atida 14 ming leksema, 75 ming grammatik va leksik so'z birikmalari tartiblangan⁹⁸. O'zbek tili so'z birikmalari lug'ati shu kungacha tuzilmagan. Til birliklari masalasi tilshunoslik maktablarining turli yondashuvlariga ko'ra tadqiq etilgan. Shu o'rinda so'z birikmalariga nisbatan mashina tarjimasida taraqiyotida (1954-yildan 1984-yilgacha) u qadar ijobiy munosabatlar bo'lmaganligini kuzatishimiz mumkin. Negaki, bu davr ichida mashina tarjimasida semantik-sintaktik tahlil konsepsiyasi ilgari surilgan bo'lib, matn so'zma-so'z tarjima qilingan. Keyingi davrlarga kelib tarjimashunoslik fanida matnni sifatli o'girishda "erkin so'z birikmalari"ni o'rganish ehtiyoji borligini amaliy natijalar ko'rsatdi. Chunki bir tildagi matnni ikkinchi bir tilga tarjima qilayotganda tushunchalar mana shu birliklar asosida voqealanadi. Endilikda mashina tarjimasida tizimlarida bu kabi masalalar alohida o'rganilib kelinmoqda. Xususan, RETRANS mashina tarjimasida tizimidagi inglizcha-ruscha politematik frazeologik lug'at 2,6 mln. so'zni o'z ichiga olgan, 250 mingta ruscha sinonimlarning so'z birikmalari shakllari qayd etilgan⁹⁹.

Лексикология английского языка: Учебник для ин-тов и фак. иностр. яз.. – 2-е изд., испр. и доп. – М.: Высш.школа, 1979. –С. 35; Зражевская Т.А., Беляева Л.М. Трудности перевода с английского языка на русский. (на материале газетных статей) – М.: Междунар. отнош., 1972. –С.140., Смирницкий А.И. Синтаксис английского языка. Изд.2. –УРСС, 2007. – С. 67; Слюсарева Н.А. Проблемы функционального синтаксиса современного английского языка. – М.,1981. – С. 206; Плоткин В.Я. Строй английского языка. –М., 1989. – С. 18; Хомский Н. Аспекты теории синтаксиса. – М., 1972. –С. 9-19; Бархударов Л.С. Структура простого предложения современного английского языка. –Издательство: ЛКИ, 2008. – С. 87; Мухин А.М. Структура предложений и их модели. – Л.: Наука, 1968. – С. 52-57; Мешков О.Д. Словообразование современного английского языка. – М., 1976. – С. 187.

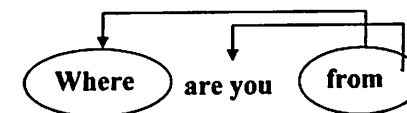
⁹⁸ Benson M., Benson E., Ilson R. The BBI Combinatory Dictionary of English. 1990.

⁹⁹ Хорошилов А.А. Теоретические основы и методы построения систем фразеологического машинного перевода: Дисс. док-ра ...техн. наук. 2006.

Inglizcha-o'zbekcha, o'zbekcha-inglizcha tarjima dasturi uchun matnlardagi so'z birikmalari lug'atini tuzishda tarjimada faqat boshqa ma'no beruvchi turlari hisobga olinadi: *capital letter* – *bosh harf*, *capital punishment* - *oliy jazo* kabi. So'z birikmalari tarkibidagi so'zlar o'zaro leksik-semantik hamda grammatik jihatdan bog'langan bo'ladi. Grammatik vositalarda birikma tartibi muhim ahamiyatga ega. So'z birikmalari tarkibidagi so'zlar nafaqat yonmayon, balki joylashuvi bir-biriga yaqin bo'lmagan so'z birikmasi tarkibidagi so'zlar o'rtasida ham tobe munosabat kuzatilishi mumkin:



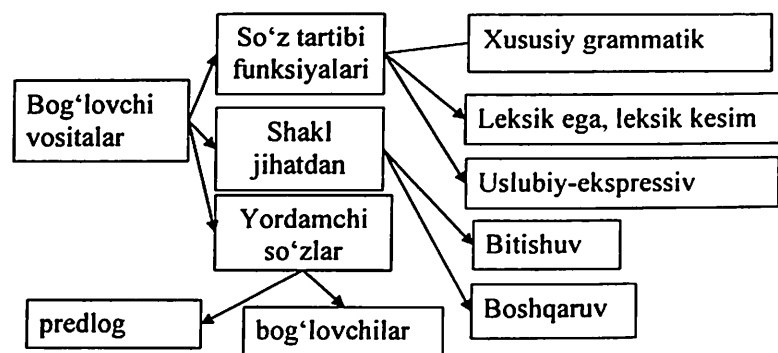
The book which you gave me yesterday is interesting => Sen kecha menga bergan kitob qiziqarli ekan.



Where are you from => Qayerdansen? (Qayerlikсан?)

O'zbek tilida so'z birikmasi bog'lanishining *bitishuv*, *boshqaruv* hamda *moslashuv* usullari mavjud. Ingliz tilida so'z birikmalarining bog'lanish usullari uch turga ajratiladi: *atributiv*, *kompletiv*, *predikativ*. Bularning ichidan atributiv bog'lanish hamisha so'z birikmasini hosil qiladi. Tobelanish munosabatli so'z birikmalarini quyidagicha ajratilgan o'rinlar ham mavjud¹⁰⁰: atributiv (aniqlovchili), obyektiv (to'ldiruvchili), holli. A.I. Smirniskiy inglizcha so'z birikmalaridagi so'zlarni bog'lashga xizmat qiluvchi quyidagi vositalarni ko'rsatadi (8-chizma):

¹⁰⁰ Смирницкий А.И. Синтаксис английского языка. – М., 1957. – С.59-60.



8-chizma

O'zbek tilida so'z birikmalarining bitishuv usulida hokim va tobe munosabatga kirishuvchi so'zlar o'zaro tartib va ohang yordamida bog'lanadi. Boshqaruv usulida so'z birikmasi kelishiklar (bosh va qaratqichdan tashqari), ko'makchilar orqali hosil bo'ladi. Moslashuvda so'z birikmasi qaratqich kelishigi bilan (ba'zan belgisiz) hosil bo'ladi.

—o'zbekcha matnlardagi so'z birikmalarini ingliz tiliga tarjima qilishdagi muammolardan biri so'z birikmalarini qo'shma fe'llardan ajratish sanaladi:

Qo'shma ot [N+V]	So'z birikmasi [N+V]
tashrif buyurmoq- to visit	ovqat buyurmoq —book the meal
qo'l qo'ymoq-to endorse	kitobni qo'ymoq-to put the book

—so'z birikmasi tarkibidagi so'zlarni bog'lovchi vositalarni to'g'ri tarjima qilish:

Uyda bo'lmoq [-da]=>	[at in on] to be at home / to be in the house
Uyga ketmoq [-ga]=>	[to into towards at on]

—so'z birikmasi tarkibidagi hokim va tobe so'zlarni ingliz tiliga tarjima qilishda unga mos sinonimlarni tanlash:

izchil reja	coherent, logical, valid, commonsensical, rational, sensible ...
chiroyli manzara	beautiful, attractive, good-looking, gorgeous, stunning, striking...

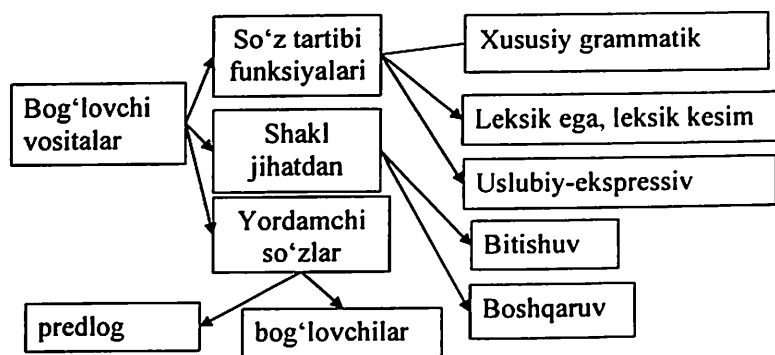
O'zbek tilida so'z birikmalarining ikki xil bog'lanishini uchratamiz: 1) teng bog'lanish; 2) tobe bog'lanish. Teng bog'lanish quyidagi grammatik vositalar yordamida voqealanadi: 1) *biriktiruv*: va, hamda, -u, -yu; 2) *zidlov*: ammo, lekin, biroq; 3) *ayiruv*: goh...goh, dam...dam, ba'zan...ba'zan. Tobe bog'lanishning ham uch turi mavjud: 1) *bitishuv*; 2) *moslashuv*; 3) *boshqaruv*.

Tobe bog'lanishli birikmada hokim so'zning qaysi so'z turkumi bilan ifodalanishiga ko'ra: 1) otli birikma; 2) fe'lli birikmaga ajratiladi.

Ular quyidagicha bo'ladi: 1) otli birikma=tobe so'z+hokim so'z (ot, sifat, son, olmosh, ravish, infinitivli); 2) fe'lli birikma=tobe so'z+hokim so'z (fe'l, ravishdosh)

Bitishuv munosabatli otli birikmalarining tuzilishi:

1. Noun+Noun=> temir uskuna
2. Adj.+Noun=> qulay imkoniyat
3. PNoun+Noun=>hamma ishtirokchilar
4. Num.Noun=> birinchi kun
5. Gerund+Noun=> o'qiyotgan qiz
6. Infinitive+ Noun=>nishonlash kuni
7. Adv.+Noun=> sekin harakat
8. (Noun+dagi)+ Noun=> devordagi rasm
9. (Infinitive+dagi)+ Noun=> ishlashdagi g'ayrat
- 10.(Adv.+dagi)+ Noun=> yuqoridagi qavat
- 11.↓PNoun+↓Adv. +Gerund+Noun=> (kingadir) (sekin) o'qib berayotgan qiz
- 12.Noun|PNoun{ni, ga, da, dan}+Gerund+Noun=> maktabga ketayotgan qiz
- 13.Adj.+Gerund|Past participle+Noun=> yaxshi o'qigan bola
- 14.Adv.+Gerund+Noun=>tez kelgan lahza
15. (Noun+day|dek)+Adj.=> oyday oppoq
- 16.(Noun+dagi)+Adj.=> sinfdagi a'lochi
- 17.Adj.+Num.=> mo'jizaviy yetti
- 18.(Noun+dagi)+Num.=> rasmdagi bir
- 19.Noun+Infinitive=>kitob o'qish (lekin *ism qo'yish, nomushta qilish* bu so'z birikmasi emas, ko'makchi fe'lli so'z qo'shilmasi)
- 20.Adj.+Infinitive=> qulay joylashish
- 21.Adv.+Infinitive=> tez yeyish



8-chizma

O'zbek tilida so'z birikmalarining bitishuv usulida hokim va tobe munosabatga kirishuvchi so'zlar o'zaro tartib va ohang yordamida bog'lanadi. Boshqaruv usulida so'z birikmasi kelishiklar (bosh va qaratqichdan tashqari), ko'makchilar orqali hosil bo'ladi. Moslashuvda so'z birikmasi qaratqich kelishigi bilan (ba'zan belgisiz) hosil bo'ladi.

—o'zbekcha matnlardagi so'z birikmalarini ingliz tiliga tarjima qilishdagi muammolardan biri so'z birikmalarini qo'shma fe'llardan ajratish sanaladi:

Qo'shma ot [N+V]	So'z birikmasi [N+V]
tashrif buyurmoq—to visit	ovqat buyurmoq—book the meal
qo'l qo'ymoq—to endorse	kitobni qo'ymoq—to put the book

—so'z birikmasi tarkibidagi so'zlarni bog'lovchi vositalarni to'g'ri tarjima qilish:

Uyda bo'lmoq [-da]=>	[at in on] to be at home / to be in the house
Uyga ketmoq [-ga]=>	[to into towards at on]

—so'z birikmasi tarkibidagi hokim va tobe so'zlarni ingliz tiliga tarjima qilishda unga mos sinonimlarni tanlash:

izchil reja	coherent, logical, valid, commonsensical, rational, sensible ...
chiroyli manzara	beautiful, attractive, good-looking, gorgeous, stunning, striking...

O'zbek tilida so'z birikmalarining ikki xil bog'lanishini uchratamiz: 1) teng bog'lanish; 2) tobe bog'lanish. Teng bog'lanish quyidagi grammatik vositalar yordamida voqealanadi: 1) *biriktiruv*: va, hamda, -u, -yu; 2) *zidlov*: ammo, lekin, biroq; 3) *ayiruv*: goh...goh, dam...dam, ba'zan...ba'zan. Tobe bog'lanishning ham uch turi mavjud: 1) *bitishuv*; 2) *moslashuv*; 3) *boshqaruv*.

Tobe bog'lanishli birikmada hokim so'zning qaysi so'z turkumi bilan ifodalanishiga ko'ra: 1) otli birikma; 2) fe'lli birikmaga ajratiladi.

Ular quyidagicha bo'ladi: 1) otli birikma=tobe so'z+hokim so'z (ot, sifat, son, olmosh, ravish, infinitivli); 2) fe'lli birikma=tobe so'z+hokim so'z (fe'l, ravishdosh)

Bitishuv munosabatli otli birikmalarning tuzilishi:

1. Noun+Noun=> temir uskuna
2. Adj.+Noun=> qulay imkoniyat
3. PNoun+Noun=> hamma ishtirokchilar
4. Num.Noun=> birinchi kun
5. Gerund+Noun=> o'qiyotgan qiz
6. Infinitive+ Noun=> nishonlash kuni
7. Adv.+Noun=> sekin harakat
8. (Noun+dagi)+ Noun=> devordagi rasm
9. (Infinitive+dagi)+ Noun=> ishlashdagi g'ayrat
10. (Adv.+dagi)+ Noun=> yuqoridagi qavat
11. ↓PNoun+↓Adv. +Gerund+Noun=> (kingadir) (sekin) o'qib berayotgan qiz
12. Noun|PNoun{ni, ga, da, dan}+Gerund+Noun=> maktabga ketayotgan qiz
13. Adj.+Gerund|Past participle+Noun=> yaxshi o'qigan bola
14. Adv.+Gerund+Noun=> tez kelgan lahza
15. (Noun+day|dek)+Adj.=> oyday oppoq
16. (Noun+dagi)+Adj.=> sinfdagi a'lochi
17. Adj.+Num.=> mo'jizaviy yetti
18. (Noun+dagi)+Num.=> rasmdagi bir
19. Noun+Infinitive=> kitob o'qish (lekin *ism qo'yish, nonushta qilish* bu so'z birikmasi emas, ko'makchi fe'lli so'z qo'shilmasi)
20. Adj.+Infinitive=> qulay joylashish
21. Adv.+Infinitive=> tez yeyish

Bitishuv munosabatli fe'lli birikmalarning tarkibi:

1. sifat+fe'l=> yaxshi o'qimoq
2. ravish+fe'l=> astoydil o'qimoq
3. ravishdosh+fe'l=> kulib gapirmoq

Bitishuv munosabatli fe'lli birikmalar:

4. Adj.+Verb=> yaxshi o'qimoq
5. Adv.+Verb=> astoydil o'qimoq
6. Gerund+Verb=> kulib gapirmoq

Boshqaruv munosabatli otli birikma tuzilishi:

Boshqaruv munosabatda kelishiklardan tushum, jo'nalish, o'rin-payt va chiqish kelishigi hamda ko'makchilar orqali bog'lanadi.

Noun+ dan+Noun=> Andijondan kelish

1. Noun+↓dan ham |↓dan ko'ra+Adj.+↓roq=>onadan mehribon
2. Noun+dan+Infinitive=>ustozdan so'rash
3. Gerund+ {Noun} dan +Infinitive => bilgandan so'rash
4. Gerund+dan+Adj.=>ko'rgandan gumon
5. Noun | PNoun + dan + Num.=>hammadan birinchi
6. Noun | PNoun + dan +Adj.=>hammadan ustun
7. Adj. + ↓lar+dan+Num.=> a'lochilardan ikkitasi
8. Adv. + dan+Adv.=> kechagidan erta
9. Adv. + dan+Infinitive=>ko'pdan bilish
- 10.Num.+dan+Num.=>yuztadan bittasi
- 11.Noun+ga+Noun=> vatanga muhabbat
- 12.PNoun+ga+Noun=> hammaga do'st
- 13.Gerund+ga+Noun=>o'qiyotganga omad
- 14.Infinitive+ga+Noun=>o'qishga mehr
- 15.Infinitive|Noun+ga+Infinitive=>o'qishga intilish
16. Noun+ga+Adv.=>bayramga yaqin
17. Noun+da+Noun=>yozuvdv xato
18. Noun+da+Num.=>tartibda birinchi
19. Noun|PNoun+da+Adj.=>menda ko'p
20. Adj.+ni+Infinitive=> qahramonni eslash
21. Noun+ni+Infinitive=>farzandni sog'inish

Boshqaruv munosabatli fe'lli birikmalar

1. Noun+ga+Verb=>maktabga bormoq
2. Noun+ga+Infinitive=>daftarga yozmoq
3. Noun|Pronoun+dan+Verb=>universitetdan qaytmoq

4. Noun|Pronoun +ni+Verb=>hikoyani o'qimoq
5. Noun +ni+ravishdosh=>ishni bajarib
6. Noun+da+Verb=>maktabda o'qimoq
7. Noun+da+Gerund=>osmonda uchib kelayotgan

Ko'makchi –

K₁ (bosh kelishikdagi ot bilan keluvchi ko'makchilar to'plami: bilan (*qalam bilan, sening bilan, sening-la*), uchun, kabi, singari, yanglig', sayin, sari, sababli, orqali, tufayli, chog'li, osha, bo'ylab, uzra, ichra, bo'yi, chamasi, haqda // to'g'rida, haqida // to'g'risida, holda // yo'sinda).

K₂ (jo'nalish kelishikdagi ot bilan keluvchi ko'makchilar to'plami: tomon, qadar, ko'ra, qarshi, qarab, qaraganda, qaramasdan / qaramay, yarasha, doir, asosan, binoan, muvofiq, qarata).

K₃ (qaratqich kelishikdagi ot bilan keluvchi ko'makchilar to'plami: ost, tag, ust, ust, old, orqa, yon, qarshi, bo'yi, ich, o'rt, bosh).

K₄ (chiqish kelishigidagi so'zlar bilan: so'ng, keyin, boshqa, bo'lak, tashqari, o'zga, beri, buyon, nari/nariga, burun, ilgari, oldin, avval, boshlab, tortib).

1. Noun+A₁+Verb=> maktab sari ketmoq
2. Noun+A₂+Verb=>taklifga binoan kelmoq
3. Noun+A₃+Verb=>uyning orqasida turmoq
4. Noun+A₄+Verb=>hammadan burun boshlamoq

Moslashuv munosabatli otli birikmalar

1. {Noun|Pronoun+ning}+{Noun+EK}=> kitobning muqovasi
2. Noun+{Noun+EK}=>talabalar soni

6.2. Diskurs va matn segmenti

XX asrda tilshunoslik til va sotsium bilan bog'liq ravishda yangi yo'nalishlar vujudga keldi¹⁰¹:

- Funksional grammatika
- Kognitiv lingvistika
- Sotsiolingvistika

¹⁰¹ Laura Alba Juez Perspectives on Discourse analysis: theory and practiceUK: Cambridge schoolars, 2009. –P.11

- Pragmalingvistika
- Matn lingvistikasi
- Diskurs analiz

Yuqorida ko'rsatilgan yo'nalishlar o'zaro bir-biri bilan bog'liq. Chunki ularning har biri tilning jamiyatdagi funksiyasi, qo'llanilishi va axborot berishda matn va tilning ishtiroki haqida bahs yuritadi. Kompyuter lingvistikasi bilan bevosita aloqador yo'nalish bu diskurs analizidir. Bu yo'nalishda ko'plab olimlar tomonidan tadqiqotlar olib borilgan: Leech (1983), Schiffrin (1994), Z.Harris (1951, 52), Fairclough (1989), Focault (1980) va boshq.

Kompyuter lingvistikasida matn lingvistikasi va diskurs analizini o'zaro farqlash kerak. Manbalarda qayd qilinishicha¹⁰², matn – bu yozma til, diskurs esa og'zaki nutq bilan hosil bo'ladi. Biroq zamonaviy lingvistikada ma'lumotning istalgan ko'rinishi matn sifatida qaraladi. Schiffrinning nuqtai nazaricha, matn lingvistikasi matnning faqat ichki elementlarini o'rganadi. Diskurs analizi esa matn va kontekst masalasini tadqiq qiladi. Fasold¹⁰³ esa tilning biror aspektida foydalanishini diskurs ilmi deya baholaydi. Aksariyat berilgan korpus analiziga oid izohlarda tilning sotsial holatda ixtiyoriy vaziyatdagi foydalanish usuli, deb baho beriladi. Matnda formal yondashuv, diskursda esa funksional yondashuv o'z aksini topadi. Laura Alba Juez diskurs analizini to'liqroq izohlab berishga harakat qiladi: *Tabiiy holatda mavjud bo'lgan yozma va og'zaki diskursning lingvistik tahlili interaktiv tarzda muayyan sotsial kontekstda kishilar orasidagi dialoglar orqali tildan foydalanish formasi* kabi izohlar beriladi¹⁰⁴.

Diskurs analiz kompyuter lingvistikasining sentiment analiz va korpus lingvistikasida muhim o'ringa ega. Tadqiqot obyekti matn bo'lganligi bois tabiiy tilni qayta ishlash diskurs analizga ham bog'liq jarayon hisoblanadi. Diskurs analizning muhim tomoni ma'lumotlar bazasidan foydalanishdir.

Qanday turdagi diskurslar analiz qilinadi? Zarur ma'lumotlar qanday yig'iladi? Qanday qilib diskurslar tarjima va annotatsiya qilinadi, kabi savollarga javob berilishi zarur.

Kompyuter tizimlarini yaratishda diskurs strukturasi va matni o'rganish tabiiy tilning generatsiyasi va interpretatsiyasi uchun zarur. Diskurs analizi NLP texnologiyasidagi matnlarni qisqartirish, savol-javob tizimlarida qo'llaniladi.

Matn segmenti

Tabiiy tillar uchun matn segmenti masalasi muhim sanaladi. Matn segmenti uchun gaplar muayyan qismlarga ajratiladi. Biroq xitoy, yapon va tailand tillarida so'zlarning chegarasini topish mushkul (ayniqsa, mashina uchun). Bunday tillar uchun *maksimal moslashtirish (MaxMATCH)* metodi mos keladi. So'zlarning umumiy uzunligiga ko'ra so'zlar morfologik jihatdan analiz qilinadi. Matn segmenti uchun asosiy vositalar jumlasiga gap uchun qo'llaniladigan tinish belgilar: nuqta, undov, so'roq, ko'p nuqta kabilar kiradi. Ushbu tinish belgilar nafaqat gap uchun, balki boshqa vazifalarda ham matn tarkibida qo'llaniladi. Masalan, (.) belgisi qisqartmalarda (prof.), raqamlarda (12.10) ishlatilishi mumkin. Bu holatda qisqartmalarning *maxsus* lug'ati zarur bo'ladi.

Dastlab Kleene (1951) chekli avtomat uchun McCulloch-Pitts neuron nazariyasiga asoslanib, o'zgarmas ifodalarni aniqlagan, Ken Tompson (1968) matn qidiruv tizimi uchun o'zgarmas ifodalarning kompilyatsiyasini yaratgan. Stemmerni Lovins (1968) birinchilardan bo'lib yaratgan. Keyinchalik ketma-ket tokenayzerlar yaratila boshlangan.

XX asrning 30-yillarida matnni yaxlit butunlik sifatida o'rganish rus tilshunosligida amalga oshirilgan bo'lsa-da, biroq hanzugacha matn ta'rifi haqida yakuniy xulosaga kelinmagan.

Svetan Todorov matnning uch asosiy kategoriyalari-parametrlarini verbal, sintaktik hamda semantik jihatdan farqlaydi. N.Enkvist matnning uch lingvistik parametri: *mavzu (topic)*, *fokus (focus)* va *bog'lanish (linkage)* borligi haqida fikr bildirgan. Mavzu bu matn mazmuni, fokus belgilangan matn elementlarini ajratish uchun xizmat qiladi (so'z, so'z birikmasi, gap, stilistik usullar), bog'lanish bu turli fikriy munosabatlarni ajratish vositasi

¹⁰²O'sha joyda. – P. 6

¹⁰³Fasold, 1990, -P. 65.

¹⁰⁴ Laura Alba Juez Perspectives on Discourse analysis: theory and practiceUK: Cambridge scholars, 2009. – P.9

hisoblanadi¹⁰⁵. Mashina tarjimasida tanlab olingan obyekt matndir. Matn eng oliy sintaktik birlik sifatida obyektiv olamdagi biror narsa-predmet, voqea- hodisa haqida keng ma'lumot ifodalash uchun xizmat qiluvchi yaxlit tizim hisoblanadi. Matnning asosiy xususiyatlaridan biri tugal fikr anglatishdir. Har qanday matn fikrlar yig'indisi bo'lib, mantiqiy bog'langan mulohazalarni o'z ichiga oladi. Bu esa o'z navbatida nutq jarayonida namoyon bo'ladi.

Mashina tarjimasida matn tushunchasi bir-biriga grammatik va mantiqiy bog'langan gaplar tuzilmasi sifatida olinib, tayyor matnlar asosida ish ko'radi. Unda matnni qanday yaratish emas, balki kiritilgan muayyan tildagi matnni ikkinchi tilda grammatik va mantiqiy shakllangan matn sifatida qayta tiklash masalasi o'rganiladi. Kompyuter matn holatidagi ma'lumotni bazaga kiritadi. Matematik ta'rifga ko'ra matn-maxsus belgilarning ketma-ketligidan iborat to'plamdir. Maxsus belgilar deganda, harflar, raqamlar, shartli belgilar, har xil ishoralar tushuniladi. Maxsus belgilar tartiblansa, chekli tartiblangan to'plam xosil bo'ladi. Har bir belgi tartib bo'yicha raqamlanib, sonlarning ikkilik sanoq sistemasidagi ko'rinishi bilan nisbiy ravishda kodlashtiriladi.

Tarjima jarayonidagi til juftliklarining har bir fragmenti muayyan kodga ega bo'lib, dasturiy ta'minotga kiritilgan buyruq ana shu belgilar asosida amalga oshiriladi.

Matnni shakllantiruvchi birliklar miqdoridan qat'i nazar bir-biri bilan mazmunan bog'langan bo'lishi kerak. Matn mazmuni uning semantikasini tashkil etadi. Inglizcha-o'zbekcha mashina tarjimasida matn quyidagi parametrlarni o'z ichiga olishi kerak: a) matnning axborot uslubi (ilmiy-texnikaviy, rasmiy-idoraviy, ijtimoiy-siyosiy) doirasida bo'lishi; b) matn birliklari (gap modellari va so'z birikmalari)ning lingvistik ta'minotda to'liq aks etishi; c) asl matnning grammatik va mazmun jihatdan tugallangan bo'lishi; d) bog'lovchi vositalarning sintaktik va semantik jihatlarining aniqlanishi; e) terminologik bazani shakllantirish; f) matnlarni tipologik jihatdan ajratish.

Mashina tarjimasida tizimlarining sifatli variantlarini yaratishda matn tipologiyasini o'rganish zarur. Bu borada V.N. Krupnova,

¹⁰⁵ Галперин И.Р. Текст как объект лингвистического исследования. –М., 1981. –С. 19.

L.L.Nelyubin, L.I.Borisova kabi bir qator olimlarning tarjima nazariyasi va modeliga bag'ishlangan tadqiqotlarida u yoki bu tarjima model samaradorligi bilan bog'liq ravishda tarjima qilinayotgan matn tiplariga ilovalar keltirilgan. L.N.Belyayev matnlarni bir nechta turlarga ajratgan: *matnlarning strukturasi, funksional uslubi, uslubiy-janrlari, matn ifodasi* ko'ra va h.k. Uning tadqiqotlarida matnlar bilan bog'liq turli jinsdagi formal belgilarning gaplarga ajratilishi tarjima jarayonida matnni avtomatik tahlil qilish bosqichlari, struktur tahlil qilinishi me'yoriy talablarga ko'ra chegaralash imkonini bergan.

Mashina tarjimasida so'zlardagi polisemiya va omonimiya kabi muammolarni yechishda ham matnlarni tasniflash muhim ahamiyat kasb etadi. Olimlar mana shunday masalalarni hal etishda ikki omilni taklif etadilar. Birinchi omil – kontekstologik lug'at yaratish. Bu tamoyil dastlab AMPAR mashina tarjimasida tizimi ilmiy guruhi tomonidan tuzilgan. Unga ko'ra asos tildagi matn konkordansi va unga munosib tarjima (parallel matnlar) asosiga quriladi. Mazkur lug'at aniqlangan matn bilan bog'liq ravishda undagi har bir ko'p ma'noli yoki omonim so'zlarning tarjima qilinayotgan matndagi lingvistik qurshovi aks etadi. Matn tiplari mashina tarjimasida uchun boshqa lingvistik tarkibiy qismlar: frazeologiya, sintaktik strukturalar, morfologik xususiyatini aniqlashga yordam beradi. Ikkinchi omil tizimini qurish asnosida parallel matnlar konkordansining xususiyatlari bilan tizim imkoniyatlari asos bo'lib hisoblanadi¹⁰⁶.

Samaradorligi jihatdan past bo'lgan mashina tarjimasida tizimlarida, birinchi navbatda, leksik xatolar kuzatiladi. Bu borada L.I.Borisova mutaxassislikka oid ilmiy-texnikaviy matnlarni tarjima qilishda ikki yo'nalish bo'yicha ajratib olishni taklif qiladi. Birinchi yo'nalishda umumiste'moldagi va umumfanga tegishli terminlarni ajratib olish. Ikkinchi yo'nalishda esa xususiy terminlarni alohida tartiblash lozim bo'ladi.

SOKRAT mashina tarjimasida orqali quyidagi matn tarjima qilinganda bu holatni kuzatish mumkin: *Criticism is the analysis and judgment of works of art. It tries to interpret and to evaluate such*

¹⁰⁶ <http://www.nop-dipo.ru/node> Марчук Ю.Н. Типология текстов и машинный перевод.

works and to examine the principles by which they may be understood. (Критика – анализ и решение произведения искусства. Это (он) пытается (пробовать) интерпретировать и чтобы оценивать такие работы (труд) и чтобы изучать (обследовать, проверять) принципы, которыми они могут быть поняты.)

Xususan, ingliz tilida *work* so'zi umumiste'molda *mehnat, ish* tarzida tarjima qilinadi. Biroq ushbu kontekstda termin sifatida *asar* ma'nosida qo'llangan.

O'zbek va ingliz tillar grammatikasining rivojlanishida lingvistik modellashtirish yangilik emas. Ya'ni har ikki tilning tadqiqot tadrijida modellashtirishga oid masalalar nafaqat xorij mutaxassislari, balki o'zbek tilshunoslari va matematik-dasturchilari tomonidan ham izchillik bilan o'rganilgan. Biz gaplarni qayta modellashtirish yoki ulardan yangilik ochish da'vosida emasmiz. Aksincha, asosiy maqsad yo'l qo'yilgan ba'zi bir kamchiliklar va xatolarni hisobga olgan holda tarjimon dasturi uchun lingvistik ta'minotni yaratishda o'z metodlarimizni taklif etish, sinash va rivojlantirishdan iborat. Bunda tilning sof, tabiiy hodisa ekanligini unutmashlik darkor.

Ma'lumki, ingliz tilining sodda gapshakli fe'l kesim [*S+VP*] (*S=Ega VP-fe'lishakl: I come.*) hamda ot-kesimdan [*S+NP*] (*NP-otshakl: I am astudent*) tashkil topadi va bundan ko'rinib turibdiki, har bir kesimning oldida ega, zamon va modal ma'noga ishora qiluvchi grammatik ko'rsatkichlarning bo'lishi talab qilinadi. O'zbek tilida esa sodda gapning eng minimal modeli [*Wp*] (*W – atov birlik, P– predikat, m– marker: Kel-di-m. Talaba-man*) tarzida reallashadi.

Ta'kidlash kerakki, tarjimada gapning umumiy tahlili emas, balki gap bo'laklari tahlili asosida sodda gapshakllar modellashtiriladi. Ayrim adabiyotlarda sodda gapshakllar gapning so'z kengaytiruvchilari va gap kengaytiruvchilari tarzida ajratilgan¹⁰⁷, bu esa gap tarkibiga aniqlik kiritilishida modelning miqdorini shu nuqtai nazardan belgilashda muhimdir. Ya'ni sodda gap shakllarining maksimal gap qoliplarida ega va hol doimiy

¹⁰⁷Qarang: Сайфуллаева Р., Менглиев Б., Бокиев Г., Кубонова М., Юнусова З., Абузалова М. Ҳозирги ўзбек адабий тили. – Т.: Фан ва технология, 2009.

qatnashuvchi sifatida tartiblanadi, garchi gap bo'laklarining miqdori 40 va xatto undan ko'p bo'lsa-da, asosiy me'yor ichki tuzilishda saqlanib qolaveradi. Shuning uchun ingliz tilidagi matnlarni o'zbek tiliga tarjima qilishda ularning har birini model asosida ko'rib chiqish va tarjima jarayonida tekshirish ijobiy samara beradi.

1. Ingliz tili sodda gapshakllari tasnifi va modellari

Ingliz tilidagi sodda gaplar ikki tamoyilga ko'ra tasniflanadi: 1) maqsadiga ko'ra; 2) strukturasi ko'ra. Maqsadiga ko'ra gaplar to'rt turga ajratiladi: darak gaplar: a) bo'lishli gap; b) bo'lishsiz gap. So'roq gaplar: a) umumiy so'roq gap; b) maxsus so'roq gap; v) tasdiq so'roq gap; g) muqobil so'roq gap. Buyruq gaplar: a) talab; b) taklif. Undov gaplar. Gaplar strukturasi ko'ra bir bo'lakli va ikki bo'lakli gaplarga bo'linadi. Ikki bo'lakli gaplar, o'z navbatida, yana ikki guruhga taqsimlanadi: 1) to'liq; 2) noto'liq. Ingliz tilidagi sodda gapshakllari kesimining turiga ko'ra modellarini tuzish mumkin. Bu model kesim fe'l-kesim va ot-kesim tarzida bo'ladi. Fe'l kesim obyektga bo'lgan munosabatiga ko'ra o'timli (tranzitiv) va o'timsiz (intransitiv) hamda har ikki turiga ham mansub bo'lishi mumkin.

1. Kesimi o'timsiz fe'ldan iborat bo'lgan asosiy gap modellari¹⁰⁸:

1. N – V=>The convict escaped.
2. Adj – N – V=>The dangerous convict escaped.
3. Adj – N – Adv – V=>The dangerous convict narrowly escaped.
4. The dangerous convict narrowly escaped through the window.
5. Adj – N – Adv – V-PP free modifiers, compound subject; compound verb=>The dangerous convict, quiet as a mouse, narrowly escaped through the window.
6. The dangerous convict, young and quick, narrowly escaped through the window.
7. The dangerous convict, serving time for murder, narrowly escaped through the window.
8. The dangerous convict, who had beaten up a guard, narrowly escaped through a window.
9. The dangerous convict, who had beaten up a guard, and his hostage narrowly escaped through a window.

¹⁰⁸ http://en.wikiversity.org/wiki/English_sentence_structure/Exercises

10. The dangerous convict, serving time for murder, beat up a guard and narrowly escaped through a window.

II. Kesim yordamchi fe'l (to be: am, is, are, was, were, will be, would be) bilan hosil bo'lgan asosiy gap modellari:

1. N/P – V – Adj=>Sarah was brilliant.

2. N/P – V – N=>Sarah was a student.

3. N/P – V – Adj – N=>Sarah was a brilliant student.

4. N/P – V – N – N=>Sarah was a brilliant student and a kind friend.

5. N/P – N/P – N – V – Adj – N =>Her sister and she were brilliant students.

6. N/P – N/P – V – Adj=>Sarah was a brilliant student and expected to go to college.

7. N/P – V – V – VP free modifiers=>Sarah was a brilliant student but did not expect to go to college.

Yuqoridagi strukturada ega va kesim sodda shaklda kelgan bo'lsa, quyidagi uni kengaytiruvchilari bilan kelganini ko'rish mumkin:

8. A brilliant student, Sarah expected to go to college.

9. Mary, a brilliant student, expected to go to college.

10. Mary was a brilliant student who expected to go to college.

11. Mary and I remained hungry.

12. Mary and I remained hungry and tired.

13. Mary and I remained hungry and tired, but we worked hard.

14. Mary and I remained hungry and tired, but we still worked hard.

III. kesimi o'timli fe'l bo'lgan asosiy gap modellari:

1. N-V-DO=>The girl kicked the dog.

2. Adj-N-V-DO=>The young girl kicked the dog

3. Adj-N-Adv-V-DO=>The young girl viciously kicked the dog.

4. Adj-N-Adv-V-DO-PP=>The young girl viciously kicked the dog down the stairs.

5. Adj-N-V-Conj-Adv-V-PP=>The young girl and her brother viciously kicked the dog down the stairs.

6. The young girl and her brother viciously kicked the dog down the stairs and laughed.

7. The young girl and her brother viciously kicked the dog down the stairs and laughed nervously.

8. The young girl felt threatened and viciously kicked the dog down the stairs.

9. The young girl, feeling threatened, viciously kicked the dog down the stairs.

10. Feeling threatened, the young girl viciously kicked the dog down the stairs.

11. Jhon broke his leg when his car hit the tree.

12. Mary bought her dress after she fought with her mother.

13. Mary brought his mother before he made dinner

Umumiy holda ingliz tilidagi gap modellari quyidagicha modellashtirildi:

Unda darak gaplarning 12 tipi aniqlandi:

1. Noun ⊕ Verb ⊕ ↓ Adverb

2. Article ⊕ ↓ Adjective ⊕ Noun ⊕ ↓ Verb ⊕ ↓ Adjective

3. Article ⊕ ↓ Adjective ⊕ Noun ⊕ ↓ Verb ⊕ ↓ Adverb

4. Pronoun ⊕ Verb ⊕ ↓ Noun ⊕ ↓ Adjective

5. Noun ⊕ Verb ⊕ ↓ Pronoun ⊕ ↓ Noun

6. Noun ⊕ Verb ⊕ ↓ Adjective ⊕ ↓ Noun ⊕ ↓ Adverb

⊕ ↓ Pronoun ⊕ ↓ Verb

7. Pronoun ⊕ Verb ⊕ ↓ Numeral ⊕ ↓ Noun ⊕ ↓ Conjunction

⊕ ↓ Adjective ⊕ Noun

8. ↓ Article ⊕ Noun ⊕ Verb ⊕ ↓ Adjective ⊕ ↓ Conjunction

⊕ ↓ Adjective

9. ↓ Pronoun ⊕ Modal verb ⊕ Verb ⊕ ↓ Adverb

10. Noun ⊕ Modal verb ⊕ Verb ⊕ Adverb

11. modalnoe slovo ⊕ ↓ Pronoun ⊕ Verb ⊕ ↓ Conjunction

⊕ ↓ Pronoun

12. Preposition ⊕ Pronoun ⊕ Noun ⊕ Verb ⊕ ↓ Numeral

⊕ Noun

So'roq gapning 10 turi:

1. Helping verb ⊕ Noun ⊕ ↓ Verb ⊕ ↓ Noun

2. Helping verb ⊕ Noun ⊕ ↓ Pronoun ⊕ ↓ Verb ⊕ ↓ Noun

3. Helping verb ⊕ Noun ⊕ ↓ Noun

4. Helping verb ⊕ Pronoun ⊕ ↓ Noun
5. Modal verb ⊕ ↓ Noun ⊕ Verb ⊕ ↓ Adverb
6. Modal verb ⊕ ↓ Pronoun ⊕ Verb ⊕ ↓ Adverb
7. Helping word ⊕ Helping Verb ⊕ ↓ Noun ⊕ ↓ Pronoun ⊕ Verb
8. Helping verb ⊕ Noun ⊕ ↓ Noun ⊕ ↓ Conjunction ⊕ ↓ Noun
9. Helping verb ⊕ Pronoun ⊕ ↓ Noun ⊕ ↓ Conjunction
⊕ ↓ Noun
10. ↓ Noun ⊕ ↓ Pronoun ⊕ Verb ⊕ ↓ Noun ⊕ Helping Verb
⊕ ↓ Pronoun

Undov gapning 12 turi:

1. Noun ⊕ Helping verb ⊕ Verb
2. Noun ⊕ ↓ Noun ⊕ Helping Verb ⊕ ↓ Verb
3. Helping verb ⊕ ↓ Pronoun ⊕ Conjunction ⊕ Verb
⊕ ↓ Conjunction ⊕ ↓ Article ⊕ ↓ Noun
4. Verb ⊕ ↓ Article ⊕ Noun
5. Verb ⊕ ↓ Article ⊕ ↓ Noun ⊕ Modal word
6. Helping word ⊕ ↓ Adjective ⊕ Noun ⊕ ↓ Pronoun ⊕ Helping
verb ⊕ ↓ Verb
7. Helping word ⊕ ↓ Adjective ⊕ ↓ Article ⊕ ↓ Noun ⊕ Helping
verb ⊕ ↓ Preposition ⊕ ↓ Noun
8. Helping word ⊕ ↓ Adverb ⊕ ↓ Pronoun ⊕ Verb
9. Helping word ⊕ ↓ Adjective ⊕ ↓ Conjunction ⊕ ↓ Adjective
⊕ Helping verb ⊕ ↓ Pronoun ⊕ ↓ Noun
10. Helping word ⊕ ↓ Article ⊕ ↓ Adjective ⊕ Noun
11. Verb ⊕ ↓ Article ⊕ ↓ Noun ⊕ ↓ Noun
12. ↓ Article ⊕ Noun ⊕ ↓ Preposition ⊕ ↓ Article ⊕ Noun
⊕ ↓ Noun

Tarjimon dasturining ta'minotini yaratishda lingvistik modellarning o'rni

Mashina tarjimasi bir tildagi matnni ikkinchi bir tildagi matnga avtomatik tarzda tarjima qilish jarayonidir. Shunga ko'ra, mashina tarjimasini tadqiq etish ikki yo'nalishga bo'linadi: 1) amaliy (axborot uslubidagi (rasmiy, ilmiy-texnikaviy) matnlarning mashina

tarjimasi); 2) nazariy (tadqiqot metodlaridan biri sifatida kishilarning nutqiy faoliyatini modellashtirish, lingvistik tavsif uchun matematik formalarni qayta ishlash hamda tilga oid ob'ektlarni qayta ishlash algoritmlarini qidirish, insoniy tafakkur bilan mashina o'rtasidagi bog'liqlikni tadqiq etish).

Istalgan tildagi matnni avtomatik tarjima qilishda kompyuterda muayyan lingvistik ta'minot bo'lish talab qilinadi. Bu ta'minotning asosini lug'at tashkil etib, unda tartiblangan lingvistik birliklar ro'yhati aks etadi. Har bir lingvistik birlikda muayyan ravishda leksik-grammatik, semantik, semantik-sintaktik masalalar ustidan avtomatik analiz va sintez qilinadi. Semiotik nuqtai nazardan avtomatik lug'atda turli ko'rinishdagi til belgilari modelining tartiblangan ro'yhati aks etadi.

Til tuzilishi hamda matnni analiz va sintez qilishning aniq vazifalariga asosan lingvistik ta'minotida quyidagilarning mavjud bo'lishi talab qilinadi:

- Leksik asos;
- Leksemashakl;
- So'z birikmasi (so'z birikmasi valentliklari);
- Sintaktik gap strukturalari.

O'zbek va ingliz tillari turli til oilalariga mansub bo'lsa-da, ularning lug'at tarkibini tashkil qiluvchi leksemalarning yasash xossalari o'xshash. Tarjimon dasturining lingvistik modulini shakllantirishda bizga birinchi o'rinda morfologik analizator kerak bo'ladi. Chunki matnni tarjima qilayotganda birinchi analiz va undan keyin sintez jarayoni sodir bo'ladi. Xo'sh, u qanday hosil bo'ladi. Dastlab, matnni morfologik holatini muvofiqlashtirish lozimdir. O'zbek tilining izohli lug'ati (A.Madvaliev tahriri ostida. 5 jildlik.-Toshkent, 2006-2008) hamda ingliz tilining ABBY Lingvo 11 lug'atidan foydalangan holda so'z turkumlari lug'ati tuziladi. So'z turkumiga kiruvchi leksemalarning miqdori chegaralangan va uslubiy jihatdan tanlab olingan bo'lishi kerak. Semantik tahlil etilgandan so'ng, har bir so'z turkumining mantiqiy tahlili, ya'ni leksemashakllarning me'yoriy va me'yoridan tashqarida bo'lgan vaziyatlari tadqiq etiladi. Bu o'rinda A.Po'latovning "Kompyuter lingvistikasi" (2011) o'quv darsligining "O'zbek tili axborot

uslubining formal grammatikasi” qismida berilgan tekshirishlar asosida tadqiqotlar olib borilsa maqsadga muvofiq.

Ma'lumki, model bu bevosita kuzatish imkoni bo'lmagan hodisani o'rganish uchun yaratilgan sun'iy mexanizmdir. Tilni matn orqaligina kuzatish mumkin, ammo matn insonning nutqiy faoliyati jarayonida yuzaga chiquvchi tabiiy til hodisalarini to'liq tushunish uchun etarli emasligi adabiyotlarda ko'p bora ta'kidlangan. Shuning uchun tilshunoslikda obyektning tushunishning asosiy vositalaridan biri sifatida modellardan foydalaniladi.

Ingliz tilidagi matnlarni o'zbek tiliga avtomatik tarjima qilishda, avvalo, ingliz tilining xususiyatlaridan kelib chiqqan holda uning modellarini tadqiq etamiz. E'lon qilingan ilmiy tadqiqotlarda ingliz va o'zbek tillarini o'zaro chog'ishtirish barobarida ularning o'xshash va farqli jihatlari ko'rib chiqilgan edi. Bu esa bizning tekshiruvimizdagi modellarga aniqlik kiritishimizga yordam beradi.

Ma'lumki, ingliz tilining sodda gapshakli [S+VP] (S =Ega VP-fe'lishakl: *I come.*) hamda ot-kesimdan iborat [S+NP] (NP-ot shakl: *I am a student*) holida voqealanadi va bundan ko'rinib turibdiki, har bir kesimning oldida uning egasi, zamon va modal ma'noni ishora qiluvchi grammatik ko'rsatkaichlarning bo'lishi talab qilinadi. O'zbek tilida esa sodda gapning eng minimal modeli [WPM] (W-atov birlik, P-predikat, m -marker: *Kel-di-m. Talaba-man.*)

Shuni muhimki, tarjimada gapning umumiy tahlili emas, balki gap bo'laklari tahlili asosida sodda gapshakllar modellashtiriladi. Ayrim adabiyotlarda gapning so'z kengaytiruvchilari va gap kengaytiruvchilari tarzida ajratilgan, bu esa gap tarkibiga aniqlik kiritilishida modelning miqdorini shu nuqtai nazardan belgilashda muhim vazifa bajaradi. Ya'ni sodda gapshakllarining maksimal gap qoliplarida ega va hol doimiy qatnashuvchi sifatida tartiblanadi, garchi gap bo'laklarining miqdori 40 va xatto undan ko'p bo'lsa-da, asosiy me'yor ichki tuzilishda saqlanib qolaveradi. Shuning uchun ham ingliz tilidagi matnlarni o'zbek tiliga tarjima qilishda ularning har birini model asosida ko'rib chiqish va ularni tarjima jarayonida tekshirish samara beradi, degan fikrdan yiroqda emasmiz.

Shu bois dastlab to be yordamchi fe'li qatnashgan gaplarni o'zbek tiliga tarjima qilishning lingvistik ta'minotini tuzishda quyidagi ikkita sababni keltirish o'rinlidir:

- Ingliz tilida eng sodda ot-kesimdan tashkil topgan gaplar to be fe'li orqali yasaladi;

- to be polifunksional vazifa bajaradi va uning ishtirokidagi gaplar turli xarakter kasb etadi.

Dastlab to be (am, is, are) yordamchi fe'l vazifasida kelgan hozirgi zamon aniq nisbatdagi sodda gapshakllarining bo'lishli, bo'lishsiz, so'roq hamda so'roq inkor tipidagi modellarini o'zbek tiliga mos variantlari va tarjimalarini tahlil qilishdan boshlaymiz. Bu tahlilning natijasi xuddi shunday gap modellariga to'g'ri keladigan ingliz tilidagi boshqa zamondagi matnga aylantirishda muhim omil vazifasini o'taydi. Lingvistik modellardan ba'zilarini misol sifatida keltirib o'tamiz: Shartli matematik ifodalar: =>-bir jarayondan ikkinchi jarayonga o'tish< >ma'no o'zgarmaydi, | -yoki , { } - bo'lishi mumkin yoki bo'lishi mumkin emas.

1.1. Noun|P+ to be +Prep.+ Noun=>Ot|O+Ot+KK+ -ShS

I am at office => Men ofisdaman

2.1. Noun1|P1+ to be + P2+ Noun2+s'+Noun3+{-s}=>Ot1|O1+Ot2+K+EK +KK Ot3+{K}+ -ShS

These are our friends'books. => Bular do'stlarimizning kitoblari.

3.1. Noun1|P1 + to be + Prep.(with)+ Noun2| P2+ {Adv.}=> Ot1|O1 + {R}+ Ot2|O2 +KM+ -ShS

I am with him forever => Men doim u bilanman.

- Yuqoridagilardan kelib chiqib, shuni aytish mumkinki, biz tarjima qilayotgan asos til (ingliz tili) tarjima qilinayotgan til (o'zbek tili)dagi matnlarni tarkib toptiruvchi so'zlarning miqdoriga mos tushadi, ya'ni har bir so'z bir qismdan iborat bo'ladi;

- Imkoni boricha so'zlar bir ma'noda qo'llaniladi;

- Tadqiqot ishimizning asosiy maqsadi tarjima dasturining lingvistik ta'minini yaratish masalasi bo'lganligi sababli to be qatnashgan gaplarni chegaralab olamiz, ya'ni ingliz tilining aniq nisbatda ifodalangan sodda gapshakllar tahlil qilinadi;

- Bo'lishi mumkin bo'lgan qisqartma shakllardan foydalanib, ingliz tilidagi sodda gapshakllarni formulalashtiriladi;

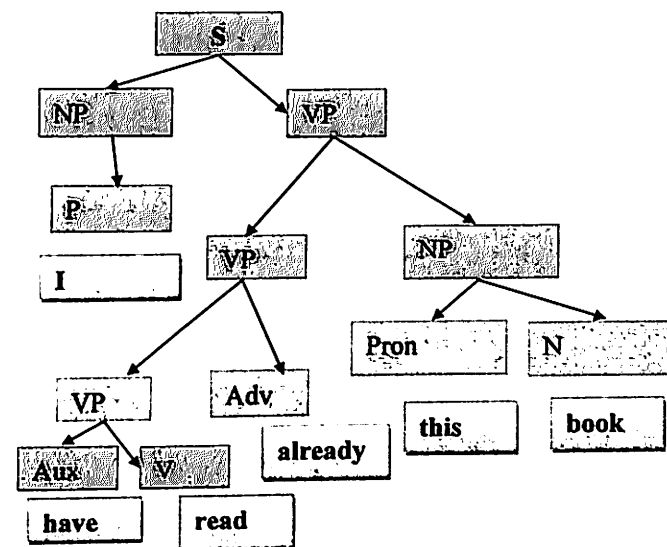
– Dastlab ikki tarkibli, ya'ni ega va kesimdan iborat bo'lgan matnlarni, so'ng so'z hamda gap kengaytiruvchilar ishtirokidagi bir necha tarkibli matnlarning modellari tadqiq etiladi.

6.3. Sintaktik tahlil (parsing)

Juda ko'p masalalarni, ayniqsa, matn yoki ma'lumotlar bazasini qayta ishlashda kompyuter bilan masalalarni hal qilish uchun gap tuzilishining asosiy modellari va unga kiritiladigan ma'lumotlarning to'liqligi talab qilinadi. Sintaktik analiz (parsing) matndagi muayyan sintaktik birlik va strukturalarni avtomatik tarzda analiz qiladi. Tahlilning bu bosqichi tabiiy tilni qayta ishlashda juda muhim bosqich sanaladi. U kompyuter lingvistikasi va NLPning deyarli barcha yo'nalishlari (mashina tarjimai, savol-javob, axborot qidiruv tizimi, sentiment analiz)da foydalaniladi. Chunki gapdagi har bir komponent muayyan vazifani bajarib, u yahlit holatda mazmunni aks ettiradi. Sintaktik parsing uchun sintaktik tavsifni yozish uchun olimlar tomonidan turli metodlar ishlab chiqilgan. Ular: *tobelik grammatikasi, bevosita tarkibiy qismli grammatika, aralashgan nazariya (sintaktik guruh nazariyasi)*.

Xususan, Karloz Gomez erkin-kontekstga asoslangan grammatikani qulay bo'lish bilan birga ayrim tillar uchun mos kelmasligini ta'kidlaydi¹⁰⁹.

Sintaktik parsing metodlaridan biri Noam Chomskiy tomonidan yaratilgan daraxtsimon tahlil qilish usulidir. Unga ko'ra har qanday sintaktik yahlitlik umumiy tugallikka ega bo'lgan ikki qismdan iborat bo'ladi: otli guruh (Noun phrase -NP) va fe'lli guruh (Verb phrase -VP). Gapning boshqa birliklari mana shu gapning asosiy qismlariga tobelanadi. Ushbu tahlil morfologiya bilan bog'langan bo'lib, tahlil qilinayotgan matn so'zlarning sintaktik vazifasiga ko'ra emas, balki uning grammatik shakliga ko'ra ajratiladi. Har bir juftlik tobelanish munosabatida bo'ladi. 9-chizmada ingliz tilida quyidagi jumla keltirilgan: *I have already read this book.*



9-chizma

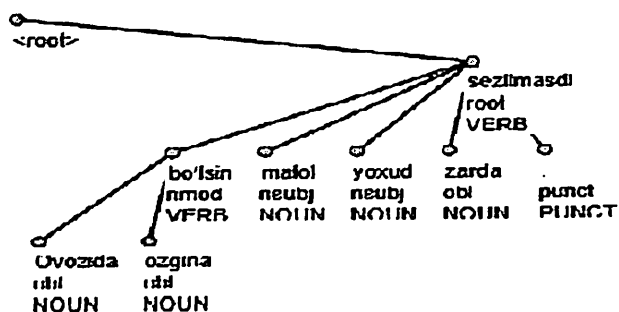
Sintaktik tahlil dasturiy ta'minot uchun muhim bosqich sanaladi. SP tanlovga asoslangan parsing (*constituency parsing*) hamda tobelangan parsingga (*dependency parsing*) bo'linadi.

Nutqiy muloqot jarayonida fikrni ifodalash shakllari turlicha bo'lib, sintaktik shakllangan gap modellari turlicha bo'lishi mumkin. (10-chizma). O'zbek tilining so'zlarning sintaktik vazifasi unga qo'shiladigan Grammatik belgilariga ko'ra belgilanadi. Ingliz tili affikslar alohida shaklan qo'llanilgani bois uning qaysi gap bo'lagi vazifasida kelayotganini o'zidan oldin keluvchi predloglar yoki gap tartibi orqali aniqlash mumkin.

Sintaktik strukturalarning hajmiga ko'ra sintaktik analiz qilish ikkiga ajraladi: a) integral (gap tarkibidagi barcha tashkil etuvchilarni tahlil qiladi); b) lokal (strukturalarning ayrim bo'laklarini tahlil qiladi).

¹⁰⁹ Carloz Gomez Radriguez Parsing schemata for practical text analysis Imperial college press, 2010. -P.7

Qovozida ozgina ho'lsin malol yoxud zarda sezilmasdi



10-chizma

Tobe so'zlarning miqdori qancha bo'lishidan qat'iy nazar hokim so'zning o'rni va mavqei o'zgarishsiz qolaveradi. Tahlildagi asosiy bosqich gapning bosh bo'laklarini aniqlab olishdir. Dastlab gapda ishtirok etayotgan kesimni (ot-kesim yoki fe'l-kesimligini) va unga tobelanib keluvchi bo'lak-eganing o'rmini topish kerak bo'ladi. Ingliz tilidagi matn o'zbek tiliga, odatda, matnning oxirgi qismidan boshlab tarjima qilinadi. Shuning uchun dasturiy ta'minot uchun so'z valentliklari lug'ati ham zarur.

Matn nutqiy jarayonda voqealanar ekan uni sof sintaktik modellarga mos deb aytish qiyin. Negaki, aynan grammatik tamoyillar uchun istisno holatlarni ko'plab uchratamiz. Shuning uchun biz tahlil ostiga olgan inglizcha sodda gapshakllar faqat formallashgan axborot uslubidagi (ilmiy va rasmiy publitsistik uslublarni shu termin bilan ataymiz) tahlil qilinadi.

Matnning sintaktik analizi CoNLLU format orqali tahlil qilinadi. Har bir jumla bir yoki bir necha qatorlardan iborat bo'lib, u quyidagi maydonlardan iborat bo'ladi. Buning uchun so'zlar morfologik jihatdan teglanadi. Morfologik teglash (Part-of-speech tagging) matnga kiritilgan har bir so'zning so'z turkumi bo'yicha markerini belgilash jarayonidir. Teglash algoritmiga kiritiladigan ma'lumot - teglar majmui va so'zlarning ketma-ketligidan iborat bo'ladi.

1. **ID:** so'z indeksi, har bir qator biror jumla boshlanganda yangi qatordan yoziladi.

2. **FORM:** so'z shakli yoki punktuatsion belgi.

3. **LEMMA:** lemma yoki so'zning o'zagi.

4. **UPOS:** universal so'z turkumi tegi (POS tag)

5. **XPOS:** tilning maxsus so'z turkumi tegi agar bo'lmasa, tag osti belgisi _ qo'yiladi.

6. **FEATS:** morfologik kategoriyalar tartibi agar mavjud bo'lsa

7. **HEAD:** so'zning ID raqami (so'zning gapda nechinchii tartibda turgan indeksi) yoki nol (0)

8. **DEPREL:** Universal tobelik munosabatiga ko'ra ichki kategoriyalar.

9. **DEPS:** kengaytirilgan tobelik grafigi (ixtiyoriy), agar yozilmasa _ qo'yiladi.

10. **MISC:** boshqa annotatsiya kodi.

newdoc id = Uzbektext.txt

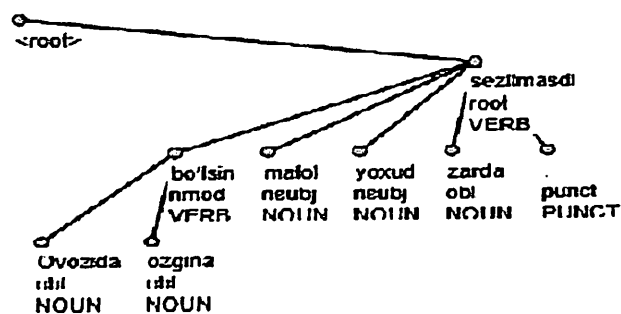
newpar

sent_id = 1

text = *Uchrashuvda innovatsiya sohasida xalqaro hamkorlik, xalqaro loyihalarni ishlab chiqish va gumanitar hamkorlik masalalari muhokama qilindi.*

ID	FORM	LEMM A	UPOS	XPO S	FEATS	H E A D	D E P R E L	D E P S	M I S C
1.	Uchrashuv da	Uchras huv	NOUN	OT	Case=Loc Numbe r= Singular Number [psor]=Singular	5		-	-
2.	innovatsiy a	innovat siya	NOUN	OT	Case=Loc Numbe r= Singular Number [psor]=Singular	6		-	-
3.	sohasida	soha	NOUN	OT	Case=Loc Numbe r= Singular Number [psor]=Singular	6		-	-
4.	xalqaro	xalqaro	ADJ	SF	Degree=Pos	4		-	-
5.	hamkorlik	hamko rlik	NOUN	OT	Case=Nom Numb er= Singular Number [psor]=Singular	3		-	-

Ovozida o'zgina bo'lsin malol yoxud zarda sezilmasdi



10-chizma

Tobe so'zlarning miqdori qancha bo'lishidan qat'iy nazar hokim so'zning o'rni va mavqei o'zgarishsiz qolaveradi. Tahlildagi asosiy bosqich gapning bosh bo'laklarini aniqlab olishdir. Dastlab gapda ishtirok etayotgan kesimni (ot-kesim yoki fe'l-kesimligini) va unga tobelanib keluvchi bo'lak-eganing o'rmini topish kerak bo'ladi. Ingliz tilidagi matn o'zbek tiliga, odatda, matnning oxirgi qismidan boshlab tarjima qilinadi. Shuning uchun dasturiy ta'minot uchun so'z valentliklari lug'ati ham zarur.

Matn nutqiy jarayonda voqealar ekan uni sof sintaktik modellarga mos deb aytish qiyin. Negaki, aynan grammatik tamoyillar uchun istisno holatlarni ko'plab uchratamiz. Shuning uchun biz tahlil ostiga olgan inglizcha sodda gapshakllar faqat formallashgan axborot uslubidagi (ilmiy va rasmiy publitsistik uslublarni shu termin bilan ataymiz) tahlil qilinadi.

Matnning sintaktik analizi CoNLLU format orqali tahlil qilinadi. Har bir jumla bir yoki bir necha qatorlardan iborat bo'lib, u quyidagi maydonlardan iborat bo'ladi. Buning uchun so'zlar morfologik jihatdan teglanadi. Morfologik teglash (Part-of-speech tagging) matnga kiritilgan har bir so'zning so'z turkumi bo'yicha markerini belgilash jarayonidir. Teglash algoritmiga kiritiladigan ma'lumot - teglar majmui va so'zlarning ketma-ketligidan iborat bo'ladi.

1. **ID**: so'z indeksi, har bir qator biror jumla boshlanganda yangi qatordan yoziladi.

2. **FORM**: so'z shakli yoki punktuatsion belgi.

3. **LEMMA**: lemma yoki so'zning o'zagi.

4. **UPOS**: universal so'z turkumi tegi (POS tag)

5. **XPOS**: tilning maxsus so'z turkumi tegi agar bo'lmasa, tag osti belgisi _ qo'yiladi.

6. **FEATS**: morfologik kategoriyalar tartibi agar mavjud bo'lsa

7. **HEAD**: so'zning ID raqami (so'zning gapda nechinchii tartibda turgan indeksi) yoki nol (0)

8. **DEPREL**: Universal tobelik munosabatiga ko'ra ichki kategoriyalar.

9. **DEPS**: kengaytirilgan tobelik grafigi (ixtiyoriy), agar yozilmasa _ qo'yiladi.

10. **MISC**: boshqa annotatsiya kodi.

newdoc id = Uzbektext.txt

newpar

sent_id = 1

text = Uchrashuvda innovatsiya sohasida xalqaro hamkorlik, xalqaro loyihalarni ishlab chiqish va gumanitar hamkorlik masalalari muhokama qilindi.

ID	FORM	LEMM A	UPOS	XPO S	FEATS	H E A D	D E P R E L	D E P S	M I S C
1.	Uchrashuv da	Uchras huv	NOUN	OT	Case=Loc Numbe r= Singular Number [psor]=Singular	5		-	-
2.	innovatsiy a	innovat siya	NOUN	OT	Case=Loc Numbe r= Singular Number [psor]=Singular	6		-	-
3.	sohasida	soha	NOUN	OT	Case=Loc Numbe r= Singular Number [psor]=Singular	6		-	-
4.	xalqaro	xalqaro	ADJ	SF	Degree=Pos	4		-	-
5.	hamkorlik	hamko rlik	NOUN	OT	Case=Nom Numb er= Singular Number [psor]=Singular	3		-	-

6.	,	,	PUNCT	TB	-	-	-	-	-
7.	xalqaro	xalqaro	ADJ	SF	Degree=Pos	4	-	-	-
8.	loyihalarni	loyiha	NOUN	OT	Case=Loc Number= Singular Number [psor]=Singular	3	-	-	-
9.	ishlab	ishla	VERB	F	Fin Conv	3	-	-	-
10.	chiqish	chiq	VERB	F	Fin Vnoun	3	-	-	-
11.	va	va	CONJ	B	-	-	-	-	-
12.	gumanitar	gumani tar	ADJ	SF	Degree=Pos	3	-	-	-
13.	hamkorlik	hamko rlik	NOUN	OT	Case=Nom Number= Singular Number [psor]=Singular	3	-	-	-
14.	masalalari	masala	NOUN	OT	Case=Loc Number= Singular Number [psor]=Plural	2	-	-	-
15.	muhokama	muhok ama	NOUN	OT	Case=Loc Number= Singular Number [psor]=Singular	1	-	-	-
16.	qilindi	qil	VERB	F	Ind= Fin	0	-	-	-
17.	.	.	PUNCT	TB	-	-	-	-	-

Ma'lumotlarga ko'ra morfologik jihatdan boy bo'lgan chex, vengr va turk tillarida ushbu teglash jarayoni ingliz va boshqa flektiv tillarga nisbatan ko'proq bo'ladi. Chunki qo'shimchalar kombinatsiyasini alohida so'z turkumlariga ko'ra teglash turli holatda kelishi mumkin. Masalan, *Noun+A3sg+Pnon+Gen* kabi morfologik tahlil ketma-ketligi katta hajmni egallaydi, zotan ingliz tilida 50-100 ta teg bo'lsa, bu ko'rsatkich turkiy tillarda 4 yoki 10 barobar ko'p bo'lishi mumkin¹¹⁰.

Ko'p tili teglash universal tobelik loyihasining Universal so'z turkumlarining teglarining bazasi turli tillar uchun (yevropa va turkiy tillar uchun) 16 ta teg va unga bog'liq ichki kategoriyalarning markerlari uchun yaratilgan.

Sintaktik birliklarni guruhlash konstruksiyalarni quyidagicha ajratadi¹¹¹.

1. A B / C (B guruhiga kiruvchi a'zolari bilan A guruhidagi a'zolar C guruhiga tegishli bo'ladi.)

2. (A) (B) C / C - (Endosentrik konstruksiya: A va B ixtiyoriy. C guruhdagi tarkib keyingi kattaroq guruhdagi C tarkibdagi bilan bir xil.)

3. A`B / C

(A`: A ning a'zolari birliklar majmuasi bo'lib, konstruksiyadagi biror qo'shimcha birlik sintaktik qurilmadagi birliklarni oshirish uchun konstruksiyalar tuziladi)

4. `A*B / C

(A guruhning faqat muayyan komponentlarigina xususiylashadi. Chunki cheklov faqat sintaktik qurilmalarga tegishli bo'ladi)

5. A B / C`

(C guruhiga kiruvchi a'zolarga nisbatan boshqa tarkib toptiruvchilarda chekli distributsiyalar bo'ladi)

6. A B* / C

(Maxsus qolip uchun birliklarning tartibiga bog'liq bo'ladi)

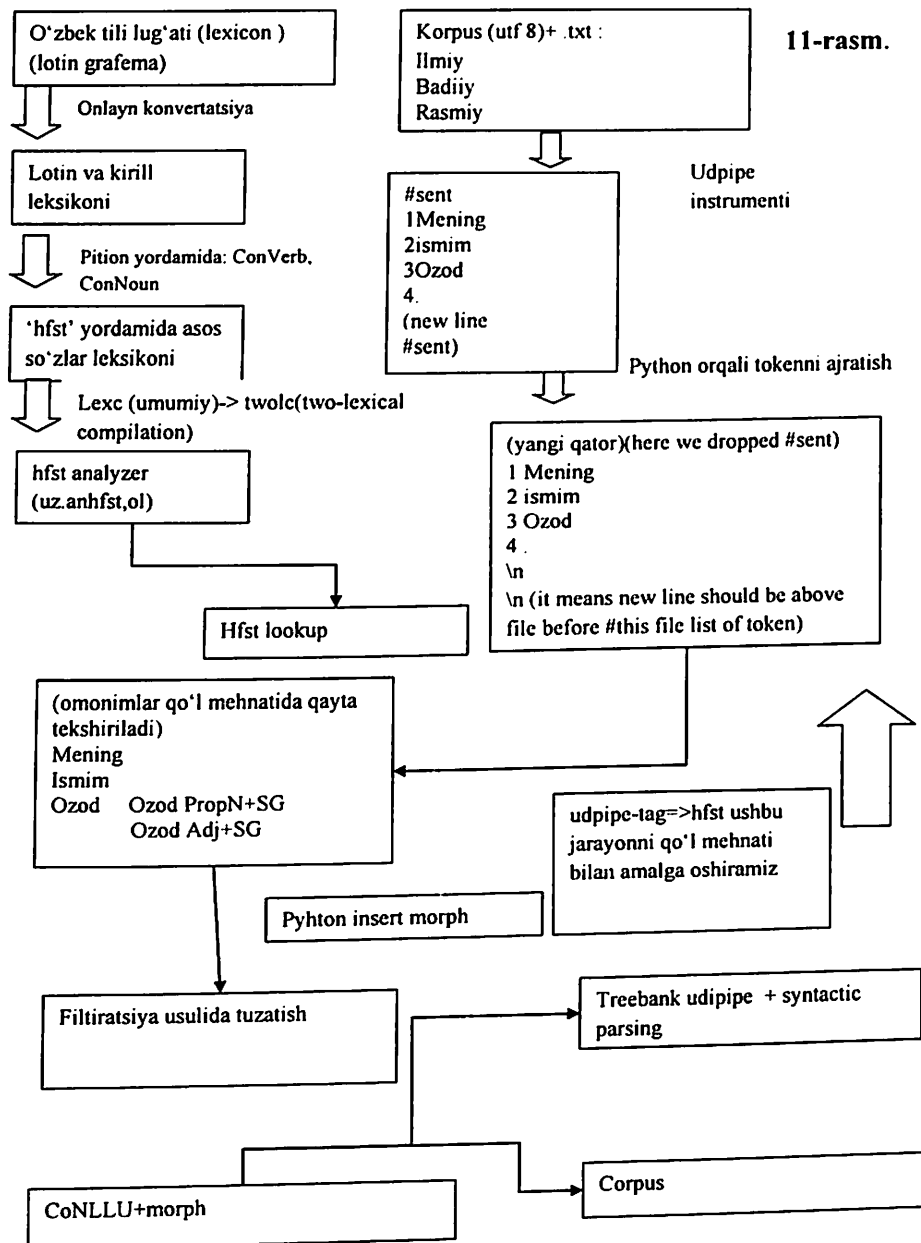
7. A (B:) / A

(B guruhidagi a'zolar umuman takrorlanmasligi yoki ko'p marotaba takrorlanishi mumkin)

11-rasmda o'zbek tili korpusi uchun korpus texnologiyasi orqali amalga oshirish bosqichi tasvirlangan.

¹¹⁰ Jurafskiy D., James H. Martin Speech and language processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 2007 -P. 173

¹¹¹ Sergei Nirenburg, Harold Somers, Yorik Wilks Reading in machine translation / Sydney M.Lamb On the Mechanization of syntactic analysis. Lamb-P. 110



Savol va topshiriqlar

1. O'zbek tilining so'z birikmalar modelini tuzishda qanday jarayonlar amalga oshiriladi?
2. So'z birikmalarining jadvashtirilishi qanday amaliy ahamiyatga ega?

3. O'zbek tilidagi so'z birikmalarini ingliz tiliga tarjima qilishdagi asosiy muammolarni sanab bering.
4. Mashina tarjimasida matn tahlilining qanday turlari mavjud?
5. Gap bo'laklarini aniqlash jarayoni tugallangach sintaktik tahlilning qaysi turidan foydalanish samarali natija beradi?
6. Quyidagi amallarni o'zgaras ifodalar bilan yozing.
 1. alifbo satrning to'plami;
 2. a b bilan tugallangan alifbo tartibidagi satrlarning barcha kichik harflar to'plami;

6-mavzu yuzasidan test

1. Sintaktik strukturalarning hajmiga ko'ra sintaktik analiz qilish qanday turlarga bo'linadi?
 - a) integral, formal
 - b) integral, lokal
 - c) lokal, formal
 - d) lokal, struktiv
2. Diskurs analizi bo'yicha qaysi olimlar tadqiqotlar olib borgan?
 - a) Leech, Schiffirin, Z.Harris, Fairclough, Focault
 - b) Federiko Gobbo va Marko Benini
 - c) Denis Bechet, Alexander Dikovskiy
 - d) Jenna Nyblom va Tabio Salakosk
3. Qaysi analizda matn va kontekst masalasi o'rganiladi?
 - a) diskurs analiz
 - b) morfologik analiz
 - c) sintaktik analiz
 - d) so'z tarkibi bo'yicha analiz
4. Kim tomonidan diskurs analizga quyidagi izoh beriladi: *Tabiiy holatda mavjud bo'lgan yozma va og'zaki diskursning lingvistik tahlili interaktiv holatda muayyan sotsial kontekstda kishilar orasidagi dialoglarda tildan foydalanish kabi tavsiflar beriladi.*
 - a) Laura Alba Juez
 - b) Tabio Salakosk
 - c) Denis Bechet
 - d) Alexander Dikovskiy

5. Sintaktik parsing nimaga asoslanadi?
- Noam Chomskiyning daraxtsimon tahlil qilish metodiga asoslanadi
 - Karlos Gomezning yaratgan nazariyasiga asoslanadi
 - diskurs analiziga asoslanadi
 - gap bo'laklari bo'yicha tahlil
6. Ingliz tilida soda gaplar mazmuniga ko'ra nechaga bo'linadi?
- 4
 - 5
 - 6
 - 3
7. Kimning fikricha, erkin-kontekstga asoslangan grammatika qulay bo'lishi bilan birga ayrim tillar uchun mos kelmaydi?
- Karlos Gomez
 - Chomskiy
 - Shennon
 - Kleene
8. Kim birinchilardan bo'lib Stemmerni yaratgan?
- Lovins
 - Kleene
 - Ken Tompson
 - Karlos Gomez
9. Kleene 1951-yilda qanday tadqiqot olib borgan?
- chekli avtomat uchun McCulloch-Pitts neuron nazariyasiga asoslanib regular ifodalarni aniqlagan
 - matn qidiruv tizimi uchun regular ifodalarning kompilyatsiyasini yaratgan
 - Stemmerni yaratgan
 - tokenizerlar yaratgan
10. XX asrda tilshunoslik til va sotsium bilan bog'liq ravishda qanday yangi yo'nalishlar vujudga keldi?
- funksional grammatika, kognitiv lingvistik
 - sotsiolingvistika, pragmalingvistika
 - matn lingvistikasi, diskurs analiz
 - barcha javoblar to'g'ri

4-MODUL. KOMPYUTER LINGVISTIKASI YO'NALISHLARI VA METODLARI

7-mavzu. Kompyuter leksikografiyasi

7.1. Lug'at tipologiyasi.

7.2. Tezaurus. WordNet va SynSet.

7.3. Ontologik modellar

7.4. Terminologiya va terminografiya

Tayanch so'z va iboralar: *lug'at tipologiyasi, kompyuter leksikografiyasi, ana'anaviy; hisobli lug'atlar, elektron lug'at, ma'lumotlar qidirish, avtomatik referatlash, lingvo lug'ati, tezaurus, ontologiya, WORDNET, EuroWordnet, SynSet, Axborot qidiruvi (IR) tezaurusi, ontologik modellar, Ontologik instrumentlar, terminologiya va terminografiya, CAT, kiberleksikografiya, korpus leksikografiyasi.*

7.1. Lug'at tipologiyasi

O'zbek leksikografiyasi o'zining uzoq tarixiga ega bo'lib, uning ilk namunasi Mahmud Koshg'ariyning "Devoni lug'ati turk" asaridir. Uzoq davriy bosqichlarda qomusiy va lingvistik lug'atlar yaratilib, takomillashtirib borildi. Globalashuv jarayoni nafaqat insonlarning turmush tarzi va madaniyati, balki kishilar muloqotining asosiy vositasi bo'lgan tilga ham o'z ta'sirini ko'rsatdi. Shu bois muloqotning yanada optimal usullarini topish yo'lida turli leksikografik tadqiqotlar yanada rivojlantirilmoqda. Kommunikatsiya uchun yagona universal tilni yaratish prinsiplarini o'rganuvchi interlingvistika fani yuzaga keldi. Shu o'rinda kompyuter texnologiyalarning ham jadal sur'atlarda takomillashib jamiyatning barcha tarmoqlarida o'z o'rnini topayotgani ko'plab sohalarida kompyuter bilan bog'liq yangi yo'nalishlarning paydo bo'lishiga sabab bo'ldi. Ayniqsa, mobil qurilmalar va kompyuter texnologiyalari imkoniyatlarining kengayib borayotganligi til bilan bog'liq ishlarni rivojlantirishda muhim omil sanaladi.

Bu borada kompyuter leksikografiyasi¹¹² (KL) kompyuter lingvistikasining asosiy yo'nalishi sifatida muhim o'ringa ega. Unda kompyuter lug'atlarini yaratish, dasturiy ta'minot uchun lingvistik resurs yaratish vazifalari o'rganiladi. Ayrim manbalarda KL leksikografik manbalarni olish, ular bilan bog'liq lingvistik ma'lumotlarni berilganlar bazasiga yozish, leksik birliklarni kiritish masalalari bilan shug'ullanadi.

L.I.Kolodnyajnyaya lug'atshunoslikni 2 ga ajratadi: an'anaviy hamda raqamli. Tarjima borasida ham ikki yondashuv borligini qayd etiladi: 1) leksikosentrik (kontekstdan kelib chiqib tarjima qilish); 2) slovarsentrik (mashina yordamida analiz qilish).

Kompyuter va lug'atshunoslik o'rtasidagi o'zaro muvofiqlashuv o'tgan asrning 60-yillarida o'z ifodasini topdi. Bu yillarda Brown korpusi, 1978-yilda avtomatik tekshiruvchi tizim hamda tilni qayta ishlash uchun maxsus kodlangan Zamonaviy ingliz tili Longman lug'ati (Longman Dictionary of Contemporary English) yaratildi.

Lug'atlarning yaratilishida uch narsa hisobga olinishi kerak:

- lug'atning yaratilishiga ko'ra maqsad va vazifalari;
- lug'atda tavsiflangan matnlar;
- foydalanuvchilarning kategoriyasi.

Endilikda internet resurslarining tobora kengayib borishi va uni tilga ta'siri natijasida korpus leksikografiyasi yoki kiberleksikografiyasi kabi yangi yo'nalishlar paydo bo'lmoqda.

Korpus leksikografiyasi –ma'lumotlarni saqlash, uzatish va analiz qilish kabi vazifasiga bajarishga yo'naltirilgan kompyuterning muayyan fazasida ko'p qo'llaniladigan matn yoki akustik korpusi asosida tilni o'rganish sohasi hisoblanadi¹¹³. Ushbu soha kompyuter texnologiyalari bilan chambarchas bog'liq. Korpus leksikografiyasining vazifalariga nafaqat tilni analiz qilish, balki kompyuter yordamida matnni tahlil qilishning texnologiya va metodlarini ham o'rganadi.

Kiberleksikografiyada matn bilan bog'liq ishlarda yoki lug'at tuzishda internetdan foydalaniladi.

¹¹² Kompyuter leksikografiyasi elektron leksikografiya yoki raqamli leksikografiya deb ham nomlanadi.

¹¹³ Бархударов С.Г., Новиков Л.А. Каким должен быть учебный словарь? // Русский язык за рубежом. – 2001. – № 3. – С. 46–50.

Shuningdek, katta hajmdagi lingvistik resurslar internet foydalanuvchilar tomonidan boyitilmoqda. Bu soha mutaxassislari tomonidan "ko'ngillilar leksikografiyasi" («волонтерской лексикографии» - user-generated lexicography) deb nomlandi. Ko'plab nashriyotlar internet orqali ma'lumotlarini shu usulda boyitib borishmoqda: *Wikipedia*, *Urban Dictionary*, *Wordnik*, *MacMillan Open Dictionary* kabilar¹¹⁴. I.A.Voronsova elektron ko'rinishdagi lug'atlarni "medialug'at" deb nomlashni tavsiya qiladi¹¹⁵.

An'anaviy va kompyuter leksikografiyasining asosiy vazifalariga lug'atning tuzilishini aniqlash hamda lug'atdagi maqolalarning holati va strukturasi tashkil etish kiradi.

An'anaviy lug'atda u yoki bu so'zlarning tuzilish kompozitsiyasi aks etgan izohlar qay holatda berilishi nazarda tutilsa, elektron (avtomatik, kompyuter) lug'atni yaratish prinsiplariga ko'ra foydalanuvchilar uchun maxsus kompyuter formatida yoki muayyan tarkibiy qism dasturi sifatida berilgan dastur nazarda tutiladi (masalan, mashina tarjimasi). Shu bilan birga foydalanuvchining undan qay tarzda qo'llashiga ko'ra, *elektron lug'at* hamda matnni qayta ishlovchi dastur uchun *avtomatik lug'atdan* farqlanadi. Foydalanuvchi uchun mo'ljallangan avtomatik lug'atlarni erkin tarzda ishlatish mumkin. Ularning sirasiga ayrim lug'atlarni keltirishimiz mumkin:

- Ingliz tilining Oksford lug'ati (www.oed.com)
 - Ingliz tilining avtomatik izohli lug'ati Collins (www.mycobuild.com)
 - Y.D. Apresyan va E.M. Mednikovaning "Yangi katta hajmli inglizcha- ruscha" lug'atining avtomatik varianti (<http://eng-rus.slovaronline.com>)
 - Ojegova onlayn lug'ati (<http://slovarozhegova.ru>)
- Ma'lumot qidirish, matnni avtomatik referatlash va mashina

¹¹⁴ Маник С.А. Новый словарь общественно-политической терминологии и его место в современной лексикографии в эпоху кибберлексикографии // Вестник МГОУ. Серия: Лингвистика 2015 / № 5. –С. 33-34.

¹¹⁵ Воронцова И.А. Медисловарь: принципы организации и перспективы развития // Ярославский педагогический вестник. Том I (Гуманитарные науки). 2014. № 2. С. 176–182.

tarjimasiga mo'ljallangan avtomatik lug'atlar foydalanuvchi uchun yaratilgan avtomatik lug'atlardan maqolalarning strukturasi va interfeysiga ko'ra farqlanadi. Strukturalarning o'ziga xosligi dastur bilan bog'liq bo'lgan mavzularning lug'atga oid materiallari bilan bevosita u bilan bog'liq bo'lgan sohalarni ham qamrab olinganligidadir. Bunday lug'atning tarkibida muayyan maqolaga ilova tarzida bog'langan yuzlab ilovalarni keltiriladi. Leksikografiyaning turli tilshunoslikning morfologik, sintaktik, semantik sathlarni ham nazarda tutadi. Ma'lumki, an'anaviy lug'atlar mavzusiga ko'ra, ensiklopedik hamda lingvistik turlarga ajratiladi. Ularning tuzilishi esa quyidagi qismlardan iborat:

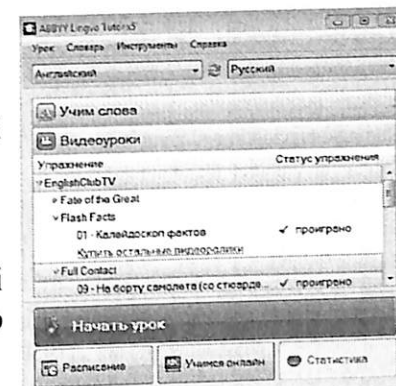
- Kirish, lug'atdagi maqolalar strukturasi haqida ma'lumot hamda lug'atdan foydalanish prinsiplarining berilishi;
- Lug'at birliklaridan iborat bo'lgan so'zlik: morfema, leksema, so'zshakllar yoki so'z birikmasi;
- Har bir birlik lug'atdagi maqolada o'zining izohiga ega;
- Ko'rsatgichlar (indekslar);
- Adabiyotlar ro'yxati;
- Shartli belgilar va alifbo.

Elektron lug'atlarda yuqorida qayd etilgan qismlarning barchasi qamrab olinadi, ammo farqli jihatlari quyidagilarda namoyon bo'ladi: har bir so'zlik mos ravishda muayyan harfning massivida bo'ladi, qolaversa, har bir so'zlik gipperilovalarda taqdim qilinadi. Elektron lug'atlardan foydalanish dasturiy ta'minot sifatida (offline dictionary) hamda Internet tizimida (online dictionary) avtomatik qidiruv tarzida foydalaniladi, ularning har biri foydalanuvchining kuchini tejaydi. Shuningdek, "qog'ozli" lug'atlardan elektron lug'atlar shunisi bilan farq qiladiki, ular multimediali va gippermatnli bo'lishi mumkin. Bunday gipperilovalar lug'atdagi istalgan maqolalarning elementi va lug'atning dasturiy menyusiga biriktirilgan bo'ladi. Bu esa foydalanuvchiga lug'atga oid kerakli ma'lumotlarni tezkorlik bilan qidirish imkoniyatini, bundan tashqari so'zga bog'langan sinonim, antonim, semantik guruhlar, turlanish va tuslanishga oid bo'lgan grammatik paradigmalar haqidagi ma'lumotlarni olish imkoniyatini yaratadi. Lug'atdagi gipperilovalar, shuningdek, turli tipdagi lug'atlar bilan bog'lanish

imkoniyatini ham yaratadi, shuning uchun onlayn yoki offlayn lug'atlar - lug'atlarning majmui yoki portalidir. Foydalanuvchiga so'z ma'nosi haqidagi kerakli ma'lumotni olish, bir ilovaga murojaat etishi bilan boshqa lug'atdagi izohlarni ham kiritish imkoniyatini beradi. Shuningdek, muayyan so'zning maxsus lug'atlardagi (terminologik lug'atda) ma'nosini bilish hamda uning shakli haqida lingvistik ma'lumotga ham ega bo'lish mumkin. Ayrim elektron lug'atlar qo'shimcha imkoniyatlarga ega, masalan, ko'ptillili ABBYY lug'ati Lingvox5 (©2008ABBYY) (ABBYY Lingvo Tutor) o'qitish tizimiga ham egadir. Bugungi kunga kelib bunday lug'atlarning har xillari mavjud:

- 1) tarjimon;
- 2) repetitor tarjimon;
- 3) mashq yoki test bilan bilimni tekshirish mumkin bo'lgan tarjimon;
- 4) videoli;
- 5) rasml;
- 6) yuqoridagilardan ba'zilari yoki barchasini o'zida qamrab oluvchi lug'at va h.k.

Lingvo lug'ati o'zining maqolalarning to'liqliligi bilan boshqa lug'atlardan farq qiladi. Birinchidan, u ko'ptillilik bo'lib yuqorida qayd etganlarimizning deyarli barchasini o'zida aks ettiradi. Qolaversa, so'zlikni uslubiy, adabiy va noadabiy ko'rinishlari hamda ularning so'z birikmasi tarzida ham alohida berilganligi e'tiborga molik. Lug'atdagi tillarning tarkibida qozoq va turk tillari ham o'rin olgan, lekin o'zbek tili hanuz bu ro'yhatda qayd etilmagan. Zotan, o'zbek tarjimashunosligida shunday lug'atlarni yaratish kabi ulkan vazifalar borligini ko'rsatadi. Tarjimonning hamisha mukammal lug'atga ehtiyoji katta. Elektron lug'atlar o'zining tezkorligi, qulayligi, ixchamligi, iqtisodiy jihatdan tejamkorligi bilan ajralib turadi. Uning ko'p funksional imkoniyatlari chet tilini o'rganishda o'ziga xos o'ringa ega. Negaki chet tilini o'rganishda nafaqat so'zlikni yodlash, balki uning talaffuz etilishi va muloqotda hamda boshqa birikmalar bilan birgalikda qo'llanilishini o'rganishda aynan elektron lug'atlar yordamidan foydalanish mumkin.



Lugʻat maqolalaridagi struktura quyidagi soʻz maydonlariga ega boʻladi:

- leksik birlikni kiritish (vocabulary-vocabulary, soʻz hamda lemma, soʻz maydonlari hamda glossa);
- grammatik maʼlumotlar maydoni;
- uslubiy belgilar maydoni;
- semantik maydoni;
- frazeologizmlar maydoni;
- etimologik maydoni;
- adabiyotlardagi namunalar va lugʻatdagi namunalar maydoni

Lugʻatdagi barcha birliklar semantik maydonlarga ajratiladi. Masalan, izohli lugʻatda maʼno maydonlarining berilishi muhim boʻlsa, orfografik lugʻatlarda bunday maydonlarning berilishi shart boʻlmaydi. Agar foydalanuvchiga soʻzlarning faqat talaffuzi kerak boʻlsa, bu kabi lugʻatlarning izohi ham nomoyon boʻladi.

Lingvistik lugʻatlarning quyidagi koʻrinishlari mavjud:

- Izohli
- Tezaurus
- Ikki tilli tarjima lugʻatlar «Inglizcha-oʻzbekcha lugʻat»

V.K.Myuller (1- nashri 1943-yilda nashrdan chiqqan)

- Assosiativ lugʻat «Zamonaviy rus tilining assosiativ lugʻati»
- Tarixiy-etimologik lugʻat
- Til shakllari lugʻati (orfografik, orfoepik, morfem lugʻat)
- Nutqiy qoʻllanishiga koʻra lugʻatlar
- Ontroponimik va toponimik lugʻatlar
- Noanʼanaviy lugʻat. «Ruscha siyosiy metaforalar lugʻati»

(A.N. Baranov va Yu.N. Karayulov)

A.N. Baranov tasnifiga koʻra elektron lugʻatlar yaratishning quyidagi bosqichlari bor:

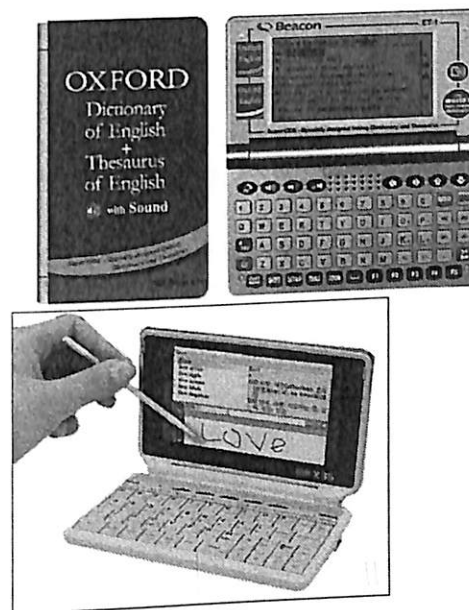
- 1) korpus matnlarini shakllantirish hamda soʻzlikni parallel tarzda yaratish;
- 2) namunalarning korpusini avtomatik shakllantirish;
- 3) maqolalarni yozish;
- 4) maqolalarni maʼlumotlar bazasiga kiritish;
- 5) maʼlumotlar bazasida maqolalarni tekshirib koʻrish;
- 6) maʼlumotlar bazasida matnlarni tekshirib koʻrish;

7) lugʻatdagi matnlarni natijasini olish hamda maketini shakllantirish;

8) lugʻatni chop etish

Soʻzlarni qayta ishlash uchun yaratiladigan lugʻatlar 1940-yillarga borib taqaladi. Negaki mashina tarjimasida olib borilgan ilmiy tadqiqotlarda, avvalo, maʼlumotlar bazasini yaratish masalasiga eʼtibor qaratilgan. Bugungi kunga kelib bunday maqsadlarga qaratilgan lugʻatlar matn ichidagi soʻzshakllari hamda ularning semantik maʼnolarini ifodalovchi kontekstologik konkordansini ham hisobga oladi. Xususan, Prinston kognitiv laboratoriyasi mana shu yoʻnalishda bir qator amaliy izlanishlarni amalga oshirmoqda. Leksikologiya tilshunoslikning barcha sathlari bilan bogʻliq boʻlsa-da, lekin u alohida yoʻnalish sifatida rivojlanmoqda. Negaki, leksikologiyaning tarkibida leksikografiya muhim ahamiyatga ega.

Chet tilini oʻrganishda lugʻatlar hamda zamonaviy informatsion texnologiyalarning oʻrni katta. Fan va texnologiyalarning taʼsirida tilimizda koʻplab yangi terminlar va neologizmlarning kirib kelayotganligi kompyuter leksikografiyasining oldiga bir qator vazifalarni belgilab bermoqda.



Kompyuter lug'atlari quyidagi muhim atributlarga ega bo'lishi kerak¹¹⁶:

-so'z qidiruv tizimiga berilganda bevosita tahrirlovchi dasturiga ega bo'lishi;

-so'zlarning matndagi shakli kiritilganda qidirish imkoniyatining mavjudligi;

-izlanyotgan so'z yoki so'z birikmasi ko'plab kontekstlardan olingan bo'lishi (matn korpusidan foydalangan holda);

-so'rov yuborilganda ilova xotirasiga yozib borilishi;

-tanlab olingan so'z keyingi o'rinlarda ishlatilganda uni tez topish uchun dastlabki harflar birikmasi orqali tavsiya berilishi;

-lug'at maqolalariga izoh shaklida uning bazasi to'ldirib borilishi;

-chet tilidagi so'zlar va so'z birikmalarining talaffuz qilish imkoniyatining mavjudligi.

Internetda lingvistik resurslarning quyidagi turlari bor:

Cambridge Dictionary <http://www.cup.cam.ac.uk>

Collins- COBUILD Dictionaries

<http://www.cobuild.collins.co.uk>

Merriam-Webster's Dictionaries: <http://www.m-w.com>

Oxford Dictionaries <http://www.oup.com>

Random House Dictionaries: <http://www.randomhouse.com>

On-Line Dictionaries: <http://www.facstaff.bucknell.edu/rbeard>

Eng katta lingvistik katoglarni o'zida jamlagan <http://www.yourdictionary.com> platformada 240 tildagi lug'atlardan iborat.

TEZAURUS VA ONTOLOGIYA

Kompyuter leksikografiyasi kompyuter lingvistikasining muhim yo'nalishi hisoblanadi. Fanga doir barcha tadqiqotlar lug'atshunoslik bilan uzviy bog'liq. Tabiiy tilning milliy leksik boyligi, imkoniyatlari lug'atda aks etadi. Amerika, Yevropa va Sharq mamlakatlari (Yaponiya, Xitoy, Koreya) davlatlarida kompyuter lug'atshunosligiga katta e'tibor berilgan. Shu bois lug'atlar mashina o'qiy oladigan formatda bo'lishi, qidiruv tizimiga egaligi,

¹¹⁶ Семенов А.Л. Современные информационные технологии и перевод. –М.: Академия, 2008. –С. 11.

indekslanganligi, kategoriyalashtirilganligi, klasterlashtirilganligi va foydalanuvchiga undan foydalanish imkoniyatining qulayligi va kengligi bilan xarakterlanadi.

Bugungi kunda taksonomiya, tezaurus, ontologiya kabi atamalar keng qo'llanilmoqda. Taksonomiya – maxsus lug'at bo'lib, tarmoqli (shajarasimon) tuzilgan ajdod-avlod (ota-ona – farzand), butun – qism kabi semantik munosabatlar asosiga qurilgan iyerarxik majmuadir.

Tezaurus esa toksonomiyadan kattaroq so'zlardagi ichki munosabatlarni o'z ichiga oluvchi boshqariladigan lug'at turidir. Unda ham iyerarxik, ham ekvivalent munosabatlar aks etadi.

Tezaurus (TS) (thesauros) yunoncha *xazina* ma'nosini anglatishi manbalarda qayd etilsa-da, Duglas Xarper tezaurus yunoncha fe'l so'z turkumi ildizidan kelib chiqqan bo'lib, *joylashtirish*, *joylashish* degan ma'noni bildirishini ta'kidlaydi. Robert Beekesning nuqtai nazaricha, ushbu termin ilk bor XVI-XIX asrlarda "tezaurus" sifatida lug'at yoki ensiklopediya ma'nolarida qo'llanilgan (masalan, "Thesaurus linguatinae"(1532) va "Thesaurus linguagraecae" (1572)).

Ilk tezaurus sifatida Filon Biblskiyning "Sinonimlar lug'ati", termin sifatida "Amar-kosha" (III-IV asr) asarida qo'llanilgan. Ingliz tilidagi dastlabki zamonaviy tezaurus 1805-yilda Piter Mark Roget tomonidan tuzilgan va 1852- yilda "Rogetning Thesaurusi" nomi bilan nashr etilgan.

Murreyning Oksford ingliz tili lug'ati (Oxford English dictionary) 1928-yilda an'anaviy diaxron aspektida tuzilgan va unda leksikologik bilimlarning sinxronik jihatlari inobatga olinmagan. XX asr boshlarida psixolingvist va lingvistlar hamkorligida til resursini lug'at shakliga keltirishda sinxron aspektni ham hisobga olish muhim deb qarala boshlangan. Miller and Johnson-Laird (1976) tilning leksik komponentini tahlil qiluvchi sohani *psixolingvistika* deb nomlaydi. Tezaurusda har bir so'z turkumi o'zining xoslangan kategoriyasiga ko'ra iyerarxik tarzda tasniflanadi. Bu jihatdan WordNet tezaurusga o'xshaydi. Laurens Urdangning tahriri ostida *The Synonym Finder* (1978) va Robert L. Chapman tahriri ostida *Roget's International Thesaurus* (1977) kabi lug'atlar tuzilgan. Biroq tezauruslarning ayrim kamchiligi sifatida bu biror so'zning *Wx*

va *Wy* sinonimlari bo'lsa, ayni sinonimlar alifbo tarzida takror kiritilib, qayta qidirish zaruriyatining borligi e'tirof etiladi. Agar bu vazifa kompyuter ixtiyoriga o'tkazilsa, tez va samarali bo'lishi ta'kidlanadi.

Tezaurus tematik prinsiplariga ko'ra ideografik leksikon hisoblanadi. "Bunday lug'atlarning maqsadi borliq elementlari haqidagi ma'lum tushuncha qanday moddiy vositalar yordami bilan ifodalanishini ochib berishdan iborat bo'ladi. Demak, ideografik lug'atlarda "*borliq+borliqning ongda aks etishi+nom*" tamoyiliga bo'ysuniladi¹¹⁷". V.Morkovkin, Y.Karaulovlarning fikricha, "tushuncha" dan "so'z" ga, "konsept" dan "belgi" ga o'tiladi¹¹⁸. Tezaurusda muayyan soha doirasidagi terminlar o'rtasidagi teskari aloqa mavjud. Ingliz tili doirasida *WordNet*, *EuroWordNet* kabi tezauruslar yaratilgan.

Assotsiativ tezaurus korpuslarni qayta ishlashga asoslangan elektron resurslar turli ilovalarga mo'ljallangan. Kilgarrif, Nikitina tezaurusga turli ta'riflarni keltiradi. Tezaurusda tilning sohaviy tarmoqlariga doir birliklar kiritilib, ularda assotsiativ munosabatlar o'z ifodasini topadi. Shu jihatdan u boshqa tipdagi lug'atlardan ajralib turadi. Tezaurus tarkibidagi birliklar quyidagi *konsept->termin* ko'rinishda bo'ladi. Konsept muayyan tushunchalarning majmuasidir. Masalan, *muhabbat*, *nafrat*, *do'stlik* kabi tushunchalar *hissiyot* konseptiga, u o'z navbatida *mavhum tushunchalar* konseptiga tegishli. "Tezaurusdagi leksik birliklar turli munosabatlarda voqelanadi: assotsiativ, ekvivalent, iyerarxik: tipologik (o'quv quroliga kitob kiradi, biroq kitoblarning barchasi o'quv quroli bo'lmaydi), butun-qism (binar), nomlash munosabatlar kabi. Ekvivalent munosabatlar sinonimlar, kvazisinonimlar (kontekstologik sinonim), antonimlar, omonimlar va polisemantik so'zlar sifatida keladi. Iyerarxik munosabatlardagi terminlar tor va keng doirada bo'lishi mumkin. Keng doiradagi terminlar bir nechta

¹¹⁷ Nurmonov A. Tanlangan asarlar. II jild Toshkent: Akademnashr, 2012. – B. 318.

¹¹⁸ Морковкин В.В. Идеографический словарь. – М., 1970; Караулов Ю.Н. Лингвистическое конструирование и тезаурус литературного языка. – М., 1981. – С. 149.

konseptga tegishli bo'ladi¹¹⁹". (Bu yerda NT- tor doiradagi, BT-keng doiradagi termin).

1960-yillarda fasat analizi iyerarxik munosabatlarni ifodalashda San'at va arxitektura tezaurusini yaratilishi bilan *tezaurufasat* nomini olgan. Fasat analizi bu muayyan o'xshash birliklarni o'zida jamlovchi kenseptlar to'plamidir. La Barrening fikricha, fasat analizi dastlab 1920-30-yillarda Ranganatan ishida o'z ifodasini topgan. U besh asosiy unsur (kategoriya)ga ko'ra leksik birliklarga birlashadi: *shaxsiyat, moddiyat, energiya, fazo va vaqt*. Keyinchalik Brayn Vikery Aristotel kategoriyalariga asoslanib, ushbu ro'yhatni yanada kengaytirgan: *narsa, tur, qism, mulk, material, jarayon, kuzatish, agent, mijoz, mahsulot, ikkilamchi mahsulot, fazo va vaqt*¹²⁰. Ingliz tili uchun yaratilgan Roget tezaurusida fasatlar quyidagicha tasniflangan: I. **Abstrakt munosabatlar**: borliq, munosabat, miqdor, tartib, son, vaqt, o'zgarish, sabab; II. **Fazo**: odatiylik, o'lchov, shakl, motivatsiya; III. **Moddiyat**: odatiylik, organik, organik bo'lmagan; IV. **Tafakkur**: g'oyalar shakli, fikrlar almashinuvi; V. **Kuch**: individual, jamiyat bilan; VI. **Ta'sirlanish**: odatiy, shaxsiy, simpatik, ruhiy, diniy. Ulardagi muayyan turdagi birliklarni jami mingta tipga ajratilgan. Masalan, VI tipga kiruvchi turkumdagi ruhiy tushunchalarni ifodalovchi birliklarga: *majburiyat, kayfiyat* (sentiment), holat kabilar kiradi. *Majburiyat: to'g'rilik, mavhumlik, majburiyat, hurmat*; *kayfiyat: maqtov, tilyog'lamalik, ikkiyuzlamachilik*; holat: *rostgo'ylik, haqiqatparvarlik, mardlik, begunohlik, yaxshi kishi, kechirimlilik va boshqa turdagi birliklar* kiritilgan. Abstrakt munosabatlarga kiruvchi son konseptida quyidagi terminlar kiritilgan:

Number. — N. *number, symbol, numeral, figure, cipher, digit, integer; counter; round number; formula; function; series. sum, difference, complement, subtrahend; product; multipli-cand, -er, -cator; coefficient, multiple; dividend, divisor, factor, quotient, submultiple, fraction; mixed number; numerator, denominator; decimal, circulating decimal, repetend; common measure, aliquot part; reciprocal; prime number; totient. permutation, combination,*

¹¹⁹ Catherine Ryan the construction guidelines: An introduction to thesauri guidelines on their construction, Dublin, 2014. – P. 18, 22.

¹²⁰ O'sha joyda.

variation; election. ratio, proportion; progression; arithmetical -, geometrical -, harmonical progression; percentage. figurate -, pyramidal -, polygonal- numbers; power, root, exponent, index, logarithm, antilogarithm; modulus. differential, integral, fluxion, fluent;

Adj. numeral, complementary, divisible, aliquot, reciprocal, prime, fractional, decimal, figurate, incommensurable; proportional, exponential, logarithmic, logometric, differential, fluxional, integral, totitive; positive, negative; rational, irrational; surd, radical, real, imaginary, impossible.

Kompyuter lug'atshunosligidagi yana bir muhim masala bu metama'lumot va ontologiyani yaratish hisoblanadi. Ontologiya bir nechta tizim elementlari mazmunini tavsiflovchi struktura hisoblanadi. T.Gruber ontologiyaga shunday ta'rif bergan: "Ontologiya bu konseptuallashgan ma'lumotlarning ixtisoslashtirilgan shaklidir". Ontologiya hajm jihatdan katta leksik maydonga ega. Ontologiyalar birlashib yagona ontologiya, web ma'lumotlarga asoslangan web ontologiya kabi turlarni hosil qiladi. Ontologiya nafaqat texnologiya, balki fikrlash san'ati hamdir. Ularni boyitishda quyidagi modellardan foydalaniladi:

OCML (An operational knowledge modelling language) – bilimlar modeli va ontologiyalar uchun ifoda etuvchi til hisoblanadi.

WebOnto - bilimlar modeli va ontologiyalardagi maxsus bilimlar modeli va ontologiyalarni rivojlantiruvchi webga asoslangan vizualizatsiyani ta'minlab beruvchi uskuna.

Lois – bilimlarni qidirish uchun qulay formaga asoslanuvchi interfeys.

Knote – ontologiyani yoyish uchun formaga asoslanuvchi interfeys.

Ontologiya asosan quyidagi komponentlardan tarkib topadi¹²¹:

- sinflar va tushunchalar;
- atributlar;
- munosabatlar;
- aksiomalar;
- namunalar.

¹²¹ Н.В. Лукашевич Тезаурусы в задачах информационного поиска. Москва, 2010, – С. 83

"Rogetning Thesaurusi"dagi maqolalar alifbo tartibida bo'lmay kontseptual mavzularga ko'ra tartiblangan. Roget o'zi tuzgan lug'atiga: "Lug'at sinonimlardan tashkil topsa-da, unda so'zlarning barcha sinonimlari to'liq berilgan, deb qaralmasligi lozim. Boshqa lug'atlardan farqli holda unda so'zlarning izohi berilmagan," – degan fikrlarni bildiradi.

Tezaurusga, shuningdek, LINGVOX5da quyidagi izohlar keltirilgan:

I.

- 1) ideografik lug'at;
- 2) sinonim va antonimlar lug'ati;
- 3) *so'zl. ma'lumotnoma*, ensiklopediya

II. informatikada kalit terminlar orqali qidiruvni amalga oshiruvchi hamda kompyuterda saqlanuvchi muayyan soha bo'yicha ma'lumotlarning tizimlashtirilgan majmuasi.

1. Tillarni avtomatik qayta ishlashda omonim, sinonim, atoqli otlar, iboralarga oid lug'at.

Web-tezaurusi yoki WWW- tezaurusi Internet resurslardan foydalanish hamda axborot qidiruvi uchun WWW global gipermatnli tizimdan iborat. Adam Kilgariff va Kollin Yallop tezaurus haqida "*Tezaurusdagi nima?*" (*What's in a theaurus?*)¹²² maqolasida tezaurusni muayyan guruhga kiruvchi bir-biriga yaqin ma'noli so'zlarning resursi sifatida qaraladi:

1. Roget TS:

- Kitob shaklidagi mualliflar tomonidan yaratilgan *Roget* tezaurus yozuvchilar uchun so'z qo'llashda yordam beradi;
- Wordnet va EuroWordNet, RuWordNet¹²³;
- Axborot qidiruv tizimida tezaurusdan foydalanish;
- Korpuslarini qayta ishlash orqali hosil bo'ladigan avtomatik tezauruslar.

Shuningdek, RuTez va wikilug'atlar ham shular jumlasidandir.

Roget tezaurusida fasadlar quyidagicha tasniflangan: 1) *mavhumlik*; 2) *fazo*; 3) *moddiyat*; 4) *tafakkur*; 5) *kuch*; 6) *ta'sirlanish*.

¹²² www.kilgariff.co.uk

¹²³ <https://wordnet.princeton.edu/>

Rojet tipidagi tezaurusida nafaol soʻzlar, oʻzlashgan soʻzlar, ismlar, eskirgan va fraza boʻlmagan ifodalar hamda iboralar yigʻilgan boʻlib, bu yozuvchiga, qolaversa, ingliz tili chet tili hisoblangan foydalanuvchilarga soʻzni toʻgʻri tanlash va qoʻllashda katta yordam beradi. Undagi leksemalar soʻz turkumlari boʻyicha tuzilgan boʻlib, bogʻlangan kalit soʻzlar orqali katta hajmdagi semantik korpus birlashgan. Xususan, **NATURE (tabiat)** soʻzini soʻz turkumlari boʻyicha quyidagicha tasniflaydi va semasini beradi:

n.	1.	nature, the great outdoors, the wild, tiger country, waste, wilderness area; balance of nature, ecosystem.
	2.	ecology, autecology, bionomics, natural history, nature study...
	3.	naturalist, bionomist, ecologist, physicpher, natre lover, conservationist, greenie
	4.	primitive, child of nature, noble savage, savage
adj.	5.	natural, innate, instinctive, normal, unformed, unschooled
	6.	wild, feral, ladino, tameless, warrigal, wilding (Archaic), wildish; undeveloped, rough, trackless, unimproved...
v.	7.	go back to nature, escape, go bush, go wild, rough it
adv.	8.	naturally, wild; primitively, savagely, wildly; instinctively, by birth

Biroq yuqorida qayd etilgan ushbu soʻzlarning barchasini, yaʼni vergul bilan ajratib koʻrsatilganlarni sinonim deb baholash mumkin emas, jumladan *autecology, bionomics, natural history* oʻzaro maʼnodoshlik kasb etmaydi, ular bir-biriga yaqin sohalar *wild, feral* oʻzaro maʼnosh sanalsa-da, soʻz qoʻllashda maʼno farqlanadi.

Tezaurusdagi leksik birliklar quyidagi munosabatlar asosiga qurilgan:

- 1) nomlash;
- 2) assotsiativ;
- 3) binar;
- 4) iyerarxik;
- 5) ekvivalent

Roje tezaurusida maqolalar kategoriya va podkategoriyalarga, ular esa soʻz turkumlariga ajraladi. Ajratilgan soʻzlar oʻz navbatida

sinonimlar soʻz shakllarga ega. Sinonimlar birlashtiruvchi maʼnoga ega boʻlib, darajalanish, uslubiy xususiyat, ekspressiv boʻyoqdorlik, salbiy va ijobiy maʼno, emotsional maʼno kabi muayyan xususiyatlarga ega boʻladi. Undagi baʼzi sinonimik qatordagi leksemalar davriy jihatdan ham farqlanadi.

Oʻzbek tilida sinonimik qator sinonimlar lugʻatida aks etgan, biroq ular muayyan konsept va giperonimlarga koʻra ajratilmagan:

Avaylamoq, ayamoq, ehtiyotlamoq, ehtiyot qilmoq, ardoqlamoq, eʼzozlamoq, eʼzoz etmoq, papalamoq, poʻpalamoq – bu leksemlarning izohi quyidagicha koʻrsatilgan:

1. *avaylamoq* – tashqi taʼsirdan ehtiyot qilish. Buning maʼnosi 2, 3, 4 da anglashilayotgan maʼnodan kuchliroq.

2. *ayamoq* – yoʻqolish, sarflanishdan ehtiyot qilish

3. *ehtiyotlamoq*

4. *ehtiyot qilmoq* – biror narsa boʻlishidan ehtiyot holda tutmoq

5. *ardoqlamoq* – gʻamxoʻrlik, hurmat maʼnosi

6. *eʼzozlamoq*

7. *eʼzoz etmoq* – (kitobiy uslub) hurmat, gʻamxoʻrlik

8. *papalamoq*

9. *poʻpalamoq* – (soʻzlashuv uslub) erkalash, suyish maʼnosini beradi.

XX asrning 60-yillariga kelib axborot-qidiruvdagi tezauruslar vujudga kela boshlagan. Oʻsha davrning boshlanishida axborot-qidiruvdagi tezauruslarning milliy va xalqaro standartlari yaratildi. Tezaurusda diskriptor va askriptor tushunchalari mavjud. Diskriptor turli tushunchalarning asosiy mazmunini aks ettiruvchi birlik sanaladi.

Axborot-qidiruv tezaurusdagi tiplar oʻrtasidagi munosabatlar quyidagi kategoriyalar boʻyicha tasniflanadi¹²⁴:

- rod – shakl,
- qism – butun,
- sabab – oqibat,
- xom-ashyo – mahsulot,
- tashkiliy – iyerarhiya,
- jarayon – obyekt,

¹²⁴ Wikipedia.org

- funksional – o‘xshashlik,
- jarayon– subyekt,
- antonimiya

Ulardagi munosabat iyerarxik va assotsiativ munosabatlar asosida bo‘ladi. Manbaga ko‘ra iyerarxik munosabat qarindoshlik munosabat, butun-qism munosabati, misol-sinflar munosabatiga ko‘ra tasniflanadi. Leksemalarning munosabatlarini aniqlashda mana shu iyerarxik holatlar hisobga olinadi. Masalan, *bog‘dorchilik* va *bog‘* grammatik jihatdan munosabatni tashkil qiladi, lekin iyerarxik munosabatda emas. *Bog‘dorchilik* va *dehqonchilik* jarayon oti sifatida bir tasnifga kiritiladi. Butun-qism munosabati quyidagi turlarda bo‘lishi mumkin:

1) tizimlar va tana a‘zolari: *Oliy va o‘rta maxsus ta‘lim vazirligi-universitet-fakultet-kafedra*

2) geografik obyektlar: *O‘zbekiston (davlat)- Andijon (viloyat)- Qo‘rg‘ontepa (tuman)-Dehqonchek (qishloq)*

3) ish-faoliyat sohasi: *filologiya- lingvistika-kompyuter lingvistikasi-mashina tarjimasi*

4) harbiy va siyosiy tuzilmalar, jamoa tashkilotlarining bosqichlari: *qonun chiqaruvchi-ijro etuvchi-o‘zini o‘zi boshqarish organlari*

Yuqorida ko‘rsatilganidek, leksemalar o‘rtasidagi birliklarda assotsiativ munosabatlar o‘z ifodasini topadi.

Antik davrda Filo Biblos hozirgi vaqtda “tezaura” deb ataladigan birinchi matnini yozgan.
Sanskrit tilida IV asrda Amarakosha she‘r shaklida tezaurus yaratgan.
Dastlabki zamonaviy tezaura 1805- yilda Piter Mark Roget tomonidan tuzilgan.

WORDNET

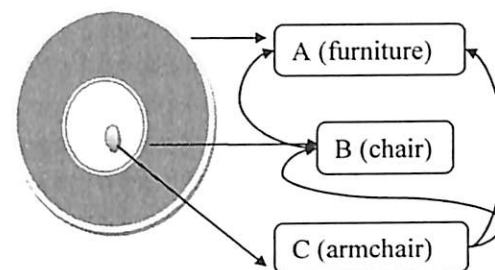
1985-yili Prinston universitetining bir qator lingvist va psixologlari Wordnet nomli leksik ma‘lumotlar bazasini yaratish loyihasi bo‘yicha tadqiqot olib bordi. O‘sha davrda u “leksik ontologiya” deb ham nomlangan. Djorj Miller, psixolingvist insonning semantik xotirasini tushunishga harakat qiladigan sun‘iy intellekt ustida olib borgan tajribalaridan ilhomlangan. Unga ko‘ra WordNet tahminan 95600 so‘zshakllarining (51,500 so‘z va 44,100 so‘z birikmalari) 70,100 ta ma‘nolari, sinonimlar bilan berilgan. Shulardan tahminan 57000si otga tegishli so‘z shakllari va ularning 47000tasi ma‘nolari kiritilgan. Maqolalar miqdori tahminiy aytilishining sababi shundaki, onlayn tizimda bu ko‘rsatkich doimiy ravishda o‘sib, o‘zgarib turadi. WordNet bilan boshqa an‘anaviy lug‘atlarining farqli jihati shundaki, WordNet faqat ot, fe‘l, sifat va ravishdan iborat bo‘ladi. WordNet (Fellbaum, 1998) psixolingvistik tadqiqotlari doir leksik ma‘lumotlar bazasidir. Prinston universitetida tuzilgan bu tizim til muhandisligi tadqiqotlarida foydalaniladi. WordNet ot va fe‘l turkumidagi so‘zlarning sinonimi va giponimlari va sifatlar uchun antonimlar asosida tuzilgan. WordNet fe‘llar uchun troponimlar va otlar uchun giponimlar terminini farqlaydi. Har bir so‘zning ma‘nosi synsetda (sinimlar tizimida) berilib iyerarxik va o‘zaro semantik aloqalari ifodalangan.

1.	Nature, wild, natural state, state of nature-(a wild primitive state untouched by civilization; “he lived in the wild”; “they tried to preserve nature as they found it”) =>state-(the way something is with respect to its main attributes; “the current state of knowledge”; “his state of health”; “in a weak financial state”)
2.	Natural phenomenon, nature-(all non-artificial phenomenon) => phenomenon-(any state or process known through trough the senses rather than by intuition or reasoning)

Endilikda WordNetning turli tillarga doir resursi yaratilgan bo‘lib, tabora xususiy bazadan global bazaga qarab kengayib bormoqda (jumladan, alban tili uchun AlbaNet, daniya tili uchun

DaNet, bolgariya tili uchun BulNet, polsha tili uchun PolNet va h.k.).

WordNet ingliz tilining leksik ma'lumotlar bazasi hisoblanadi. Undagi ot, fe'l, sifat va ravish so'z turkumlari muayyan konseptni ifodalovchi sinonimlar guruhiga (Synset) birlashgan. Synsetlar konseptual-semantik va leksik munosabatlarga ko'ra o'zaro bog'langan. WordNetdan foydalanish va o'rnatib olish qulay. Uning tuzilishi kompyuter lingvistikasi va tabiiy tilni qayta ishlash uchun foydali Wordnet ayrim jihatlari bilan tezaurusga o'xshaydi, biroq ayrim farqli jihatlari ham kuzatiladi. "WordNetning ichki bog'lanishi so'zshakli (harflar qatori) bilan emas, so'zlarning ma'nolari bilan yuzaga kelgan. Natijada so'zlar semantik ko'p ma'noli bo'lgan boshqa so'zlarga yaqin bo'lgan variant bilan topiladi. WordNetda semantik munosabatlaridagi so'zlar tasniflangan bo'lib, tezaurusda muayyan so'zga yaqin bo'lgan so'zlar uchramaydi¹²⁵". WordNetdagi asosiy munosabat sinonimiyaga qurilgan: *katta-ulkan, yig'lamoq-ho'ngramoq* kabi. Bir xil konseptga birlashtirilgan va kontekstdagi o'zaro o'zgaradigan sinonimlar tartiblanmagan guruhlarda umumlashtirilgan. WordNetdagi 117 000 synset kichik hajmli konseptual munosabatlarga ko'ra boshqa synsetlarga bog'langan. Synsetda *qisqa izohlar (glos)* berilgan bo'lib, ko'p hollarda bir yoki undan ortiq synsetdagi birliklarni qo'llanilishi aks etgan. WordNetdagi har bir so'zning ma'noviy shakli bitta bo'ladi. WordNetdagi so'zlar subordanatsion, ya'ni butun qism munosabatga qurilgan. So'zlar iyerarxik ravishda yuqoridan quyi tomon (giperonim, giponim) umumiy {furniture, piece_of_furniture} synsetdan maxsus synsetga qadar {bed} va {bunkbed} birlashadi. Shunday qilib WordNetda *mebel* kategoriyasi *yotog'* va *ikki kishilik krovat* so'zlariga bog'lanadi. Barcha ot so'z turkumiga tegishli so'zlar o'zagiga qarab o'sib boradi. Giponim so'zlar ma'no jihatdan boshqa so'zlarga ham aloqador bo'lishi mumkin: *agar kreslo kursining turi bo'lsa va kursi mebelning turi bo'lsa, u holda kursi ham kreslo ham mebelning turi sanaladi*. Buni quyidagi ifoda bilan berish mumkin (11-chizma):



11-chizma

Muhim jihatlardan biri sinonimiya va polisemiya tarjima bir qator murakkabliklarni yuzaga keltiradi. Masalan, bir so'zning bir nechta ma'nolari va ularning sinonimlari bo'lishi mumkin. Unda "sinonimiya so'zshakllar o'rtasidagi leksik munosabatlar hisoblanib, sinonimiyada so'zlarning o'rtasidagi asosiy farqi uchun {} belgisi, boshqa qo'shimcha leksik aloqalar uchun [] belgilari bilan belgilanadi."

WordNet tizimida ingliz tilidagi barcha so'z turkumlarining semantik ma'nolari ifodalangan va u semantik aloqalar orqali tashkillashtiriladi. Ot so'z turkumi misolida ushbu lug'atning xususiyatlari tahlilga tortilar ekan, uning hajmi 80000 dan ziyod ot so'z turkumiga to'g'ri kelishi va bu ko'rsatgich muayyan leksemani nutqiy birlik sifatida voqealinishida so'z birikmasi holatida qo'llanish imkoniyati qamrab olingan. WordNet kompyuter leksikografiyasida erishilgan innovatsion yutuqlardan biridir. Chunki unga ilova etilgan so'zliklar mashina o'qishi uchun mumkin bo'lgan holatga moslashtirilgan. Odatiy lug'atlarda kiritilgan so'zliklarning talaffuzi, grammatik shakllanishi, yasalishi, etimologiyasi, izohi hamda sinonim, antonim kabi yana bir qator lingvistik xususiyatlari singdirilgan bo'ladi. Bunday belgilar mashina o'qishi mumkin bo'lgan imkoniyati mavjud bo'lmaganligi bois aksariyat jihatlari tushirib qoldiriladi, xususan, WordNet yuqorida qayd etilgan talaffuz, etimologik tavsif va shunga o'xshash ma'lumotlarni o'z ichiga qamrab olmaydi. Ushbu lug'atda o'rin olgan maqolalarning qo'llanishida ifoda ma'nolarining semantik xususiyatlarini ochib berish bilan bog'liq afzalliklarni yaratish, qolaversa, to'g'ri tarjimaga erishishda so'z maydonlarini mantiqiy loyihalash birlamchi mezon bo'lib hisoblanadi. Shu jihatdan WordNet so'zlarning semantik aloqalarini chuqur o'rganishda sinonimiya

¹²⁵ <https://wordnet.princeton.edu>

hodisasi muhim deb qaraladi. WordNetda sinonimlar muayyan bloklarga ajratilib, semantik jihatdan tasniflangan. Sparck Jones (1964, 1986) o'zining semantik tasnif nazariyasiga asos solgan tadqiqotchidir. U o'zining bu boradagi izlanishlarida matndan olingan muayyan so'z shakli boshqa so'z shakllari bilan birga bo'lishi mumkin bo'lgan barcha sinonimlar tizimini yaratadi. Masalan, u ingliz tilidagi qurol-yarog' bilan bog'liq bo'lgan tushuncha *pellet* hamda *injection* (inyeksiya) so'zlarining kontekstdagi ma'nodoshi *shot* leksemasi bilan bog'laydi, bu esa shartli ravishda muayyan holat uchun sinonimik qatorni hosil qiladi. Sinonimlar tizimi (synset) o'zaro semantik bog'lansa-da, har ikki tizimning strukturasi har xil tipga tegishli bo'ladi: {*shot, pellet*} va {*shot, injection*}. Bu ikki to'plam o'zaro hech qanday ma'no jihatdan aloqaga kirishmaydi, ya'ni ularni faqat "*shot*" leksemasigina bog'lab turadi.

Aksariyat Synsetda odatiy lug'atlardagi izohlardan foydalaniladi. Polisemantik so'zlarda bir necha ko'chma ma'nolar ifodalangan glossalar mavjud bo'lsa, Synsetda faqat yagona glossema mavjud bo'ladi. WordNetdagi Synset orqali so'zliklarning leksikallashuv konsepti shakllanadi.

Sinonimlar o'rtasida semantik bog'lanish bo'lsa-da, semantik bog'lanish otlarning leksikallashuv konsepti o'rtasidagi bog'lanishning muhim tarmog'i sanaladi. Bular subordinatsiyalar (guruh yoki muayyan tasniflar)ning bog'lanishi bo'lib, unga kiruvchi elementlar giponimlar deb ataladi. Masalan, *burgut qush* so'zining giponimi bo'ladi, *qush* esa *burgut* so'zining giperonimi bo'ladi. Bunday semantik bog'lanish leksik iyerarxiyani hosil qiladi. Odatiy lug'atlarda ham otlar o'rtasidagi giponimik bog'lanish haqidagi ma'lumotlar beriladi (Amsler, 1980).

Leksik iyerarxiya quyidagicha tarmoqlanadi: {robin, redbreast} @.-> {bird} @-> {animal, animate_being} @-> {organism, life_form, living_thing}.

@.->belgi muayyan so'zni boshqa so'zga qarab xususiylashib borishi, ya'ni generilazatsiyalashuvini bildiradi. Ss @->ushbu belgi ot so'z turkumidan tashkil topgan Synsetni boshqa Sg o'tishini ifodalaydi. Ya'ni u "turi" degan ma'noni ham ifodalishi mumkin. Jumladan, Synset o'rtasida doimiy inversiya hodisasi sodir bo'ladi:

Sg ~-> Ss. boshqacha qilib aytganda, Ss Sgning giperonimi, Sg esa Ssning giponimi hisoblanadi. ~->bu belgi giperonimdan giponimga qarab ma'no kengayishini ifodalaydi. Nafaqat {bird} synset giperonim bilan bog'langan bo'ladi va balki buning tarkibiga barcha qush turlari ham kiritiladi. Bunday bog'lanish haqida ma'lumotlarni berish hamda ularni tuzish bo'yicha qo'yilgan talablar odatiy lug'atlarda ham mavjud, biroq maqolalarni topishni yanada osonlashtirish zarur. Lug'atshunoslar so'zliklarni kiritishda u haqidagi ma'lumotni aylana holatda emas, balki tarmoqli holatda bersa maqsadga muvofiq, deb hisoblaydilar. Tarmoqlanish kompyuter texnologiyalarida ham samarali usullardan biridir. Bu metod ma'lumotlar bazasida umumiy bo'lib, mavjud birliklar haqidagi maxsus joyga ID raqami bilan o'zaro bog'liq ma'lumotlarni biriktirish imkonini yaratadi. Iyerarxiyallashuv katta hajmli ma'lumotlar bazasini shakllantirishda kompyuter dasturchilari tomonidan keng qo'llaniladi (Touretzky, 1986). Chunki tarmoqlanish birliklarga ajratilayotgan joyni tejashda samarali hisoblanadi.

Kompyuter mutaxassisleri bu usulni "voris tizimi" deb nomlashadi, chunki meros keyingi avlodga tarmoqlangan holda uzatiladi. Leksik struktura chiziqli holatda aks etadi: "*Agar oy yerning yo'ldoshi bo'lsa, demak oy sayyora; agar u sayyora bo'lsa, quyosh tizimining tarkibi; agar oy quyosh tizimining tizimi bo'lsa, demak bir xil dinamik holatda harakatlanuvchi...*". Sinonim, antonim va giponimlar o'zaro semantik aloqasi mavjud tushunchalardir. Shuningdek, Word NETda meronimlar (butun-qism munosabatlari) va antonimlar ham o'rin topgan. Masalan, {g'ildirak}so'zi {velosiped} so'zining meronimi hisoblanadi.

Wm #p —> Wh - bu yerda Wm Whning tarkibiy qismi ekanligini bildiradi;

Wm #m —> Wh - Wm Whning a'zosi ekanligini bildiradi;

Wm #s —> Wh - Wm Whning tarkibi ekanligini bildiradi;

#p-WordNetdagi so'zlarning qism to'plam aloqasini bildiradi.

Leksik munosabatlarning muhim sinfini so'zshakllari va morfologik aloqalari tashkil qiladi. Zotan, WordNetga so'zning muayyan grammatik kategoriyasi bilan birgalikda so'z kiritilganda

uning boshlang'ich shaklini topish zarur: *Books=>book, geese=>goose; go-went-gone-going* kabi. Iyerararxik tarzda so'zlarning tasniflanishi muayyan chegaralarga ega. Masalan, meronim butun qism munosabati faqat ot so'z turkumi doirasida bo'lsa, antonim ham shunday holatda bo'ladi. Shu jihatdan so'zlarning uch jihati inobatga olinadi: atributlari, qismlari (meronim), funksiyalari. Shu jihatdan WordNet turli semantik komponentlarni o'z ichiga olgan yigirma beshta fayldan iborat. WordNetdagi asosiy munosabat bir xil so'z turkumi (ST) orasida hosil bo'lgan. WordNet so'z turkumlaridan iborat ichki guruhga va ayrimlari kesishgan ST (POS) bo'lingan. Ushbu kesishuvchi ST munosabatlari o'zakdosh bo'lgan morfosemantik aloqadagi tizimdan iborat: *observe (verb), observent (adjective), observation, observatory (noun)*. Ot-fe'l juftliklari ot bilan fe'lning semantik munosabatida o'zaro juftlashadi: {*sleeper, sleeping_car*} {*sleep*} uchun MAKON va {*painter*} {*paint*} ning AGENTI, shunda {*painting, picture*} NATIJA bo'ladi.

SynSet strukturasi

Semantik munosabat	Ta'rifi	So'z turkumi				Misol
		Ot	Fe'l	Sifat	Ravish	
sinonim	Ma'no jihatdan bir-biriga yaqin so'zlar	+	+	+	+	{ <i>sofa, couch, lounge</i> }=> giperonim { <i>seat</i> }
antonim	Ma'nosi bir-biriga bo'lgan so'zlar	+	+	+	+	{ <i>love</i> } <-> { <i>hate, detest</i> }
giperonim	Tur va tasnif munosabatlarini jamlovchi konsept	+	+			{ <i>feline, felid</i> } giperonim <={ <i>cat, true cat</i> }
giponim	Muayyan turga tegishli so'zlar	+	+			{ <i>wild cat</i> } giponim <={ <i>cat, true cat</i> }
meronim	Mohiyatan biror konseptga tegishli so'zlar	+				{ <i>snowflake, flake</i> } mohiyati <={ <i>snow</i> }

WordNetda fe'l so'z turkumiga kiruvchi troponimlar kiritilgan. Unga ko'ra u yoki bu fe'l boshqa muayyan shakliga ko'ra bog'lanadi. Masalan, *vaysamoq gapirmoq* fa'lining troponimi hisoblanadi. Shuningdek, fe'llar ham antonimik munosabatda bo'lishi mumkin: *turmush qurmoq - ajrashmoq*.

RuWordNetda rus tilidagi ot, sifat va fe'l so'z turkumidagi birliklar kiritilgan. Masalan, ДАВАТЬ (bermoq) fe'lga quyidagi tushunchalar birlashtirilgan:

Sinset: *на руки выдать, дать, дать в руки, давать, отдать, выдать на руки, вручать, вручить, даваться, отдавать, передать, передавать*

Giperonim: *выдать, выдавать преподнести, преподносить торжественно*

Giponim: *вручать торжественно, вручить*

Giponim: *подавать, подавать в руки*

Giponim: *недодать, недодавать*

Giponim: *всунуть, всучить, всовывать, всучивать*

Giponim: *купить, покупать, подкупать, подкупить*

Giponim: *передать безвозмездно, передавать безвозмездно, безвозмездно передать, безвозмездно передавать*

Shuningdek, lug'atda uyadosh (word similarity) va semantik bog'liq so'zlar (word relatedness) ham lingvistik baza sifatida kiritiladi. Masalan, *ananas* va *o'rik* o'zaro sinonim so'zlar emas, lekin ular meva turkumiga kirganligi uchun bitta paradigmaga birlashadi va uyadosh so'zlarni tashkil qiladi. Ularning tarkibida esa ho'l meva va sitrus meva kabi yana ichki turlari mavjud. O'zaro semantik bog'langan so'zlar esa biri ikkinchisini kelishini talab qiladi hamda assotsiativ xotirada ular hamisha yonma-yon keladi. Masalan, *paxta* so'zi bilan *~yig'moq, ~terimi, ~dek yumshoq* kabi so'zlar bilan valentlik hosil qiladi. Bunda *paxta* yuqoridagi leksemalarning hech qaysi biri bilan ma'no o'xshashligi jihatidan bog'liq emas, lekin yonma-yon kelganda bir-birini kelishini talab qilib, muayyan semantik guruhga birlashadi.

Endilikda UzNET deb nomlangan o'zbek tili WordNetining lingvistik bazasi bo'yicha amaliy ishlar bajarilmoqda. Bunda, asosan, ot, sifat va fe'l so'z turkumiga doir so'zlarning semantik maydonlari beriladi:

ishla- leksemasining semantik maydonlari tahlili

1. bosh soʻz: *ishla*
2. semantik uya: *mehnat qilmoq, ter toʻkmoq*
3. leksik-semantik guruh: *boshqarmoq, haydamoq, qurmoq, yasamoq, yaratmoq*
4. semantik toʻda: *zavod, korxonona, ishchi, dehqon, dala, dastgoh*
5. mikramaydon: *mehna, ish*
6. makromaydon: *ish-harakat*
7. sinonim: *mehnat qilmoq, ter toʻkmoq*
8. antonim: *bekor yotmoq, dangasalik qilmoq*
9. geperonim: *mehnat qilish (faoliyat koʻrsatish)*
10. giponim: *ishlamoq*
11. xolonim: *harakat*
12. meronim: *qimirlab qoʻymoq*
13. graduonim: *timirskalanmoq-ishlamoq-ter toʻkmoq*
14. funksionim: *yaratish, vujudga keltirish*
15. sath: *leksik sath*
16. soʻz turkumi: *feʻl*
17. morfologik oʻzgarishi: *oʻtimli&oʻtimsiz, boʻlishli&boʻlishsiz shaklga ega, zamon, mayl, nisbat qoʻshimchalarini oladi (ishladi-ishlatdi; ishladi-ishlamadi; ishladi-ishlayapti-ishlamoqchi-ishla(gin)-ishla(sin), ishlaylik-ishlangiz-ishlasin(lar); ishladi-ishlandi-ishlashdi-ishlattirdi)kabi*
18. tuzilishi: *sodda, yasama (ish+la)*
19. sintaktik aloqa: *zavodda ishlamoq, qattiq ishlamoq*
20. assotsiatsiya: *ishchi, dehqon, xizmatchi, korxonona, tez, sekin, dala, zavod*
21. ibora: *ishlasang-tishlaysan, ishlamasang-tishlaysan*

2. Axborot qidiruvi (IR) tezaurusi.

Oʻzbek tilining tezaurusini yaratish uchun yaratilgan WordNet tipidagi analoglardan namuna sifatida oʻrganish maqsadga muvofiq. Unga koʻra ot, feʻl, sifat va ravish soʻz turkumiga tegishli soʻzlarning yuqorida qayd etib oʻtilgan tasnif boʻyicha semantik maydonlarga ajratib lugʻatini tuzish kerak.

Har qanday lugʻat ikki qismdan iborat boʻladi: 1) indeks; 2) izohi. Tezaurus struktur jihatdan chuqur tahlilga asoslanadi. Chunki

unda soʻzlarning oʻzaro munosabatlari tizimlashtirilgan va strukturaviy loyihalashtiriladi. Biror leksikografik tizimning info leksikografik modeli quyidagicha¹²⁶boʻladi:

$$V(l) = \{A(l), P(l), H\}$$

Bu yerda $V(l)$ - lugʻat birliklari toʻplamidan iborat leksikografik tizim, $A(l)$ - lugʻatning indekslangan qismi, $P(l)$ – lugʻatning ikkinchi izohli tomoni, H -lugʻatning har ikki qismining birlashtirilishi¹²⁷:

$$H: A(l) \rightarrow P(l)$$

7.3. ONTOLOGIK MODELLAR

Ontologiya- soʻzi yunoncha *von* (“mavjud”, “bu”) kabi maʼnolarga ega. Falsafiy ontologiya bundan 2400 yil ilgari yunon falsafasida (Aristotel, Kant, Brentano va boshq.) olamni bilish maʼnosida yuzaga kelgan. Kompyuter ontologiyasi qariyb bundan 30 yil ilgari paydo boʻlgan. Bu termin ontologik yoki ontologiya injeneriyasi deb ham nomlanadi¹²⁸. Kompyuter ontologiyasining ikki xil turi mavjud: 1) ontologiya kategorial analiz sifatida (falsafiy yondashuv); 2) texnologiya sifatida (informatik yonadashuv).

N.V.Lukashevich ontologiya turlarini sanab oʻtar ekan shu kungacha yaratilgan namunalardan misol sifatida foydalanadi. Ontologiyaning soddaroq modeli sinf va sinfosti strukturalardan iborat boʻlishi va taksonomiya, butun-qism munosabatlari asosidagi model – partonomiya deb nomlanishi, rubrikatorlar ontologiyada iyerarxik tarzda tuzilishi, axborot-qidiruv tezauruslari ham ontologik resurs sifatida qaralishi, WordNet tipidagi tezauruslar ham ontologik resursga misol boʻla olishini qayd etadi. Ontologiyani shakllantirishda formal nazariyalar va aksiomalardan foydalaniladi. Formal ontologiyaning formal (mantiqiy) nazariyasi (deksreptiv mantiq, modellashtirilgan mantiq kabi) va DAML+OIL, OWL, CycL, Ontolingua kabi turli tavsiflovchi tillari mavjudligi haqida

¹²⁶ Касилов О. В. Моделирование словаря-тезауруса // Вестник НТУ ХПИ. 2004. №34

¹²⁷ Qarang. Matlatipov S., Aripov M., Abdurakhmonova N. Modeling WordNet type Thesaurus for Uzbek language Semantic Dictionary / International Journal of Engineering Research and Allied Sciences (IJERAS) ISSN : 2455-9660, Volume-01, Issue-09, November 2016

¹²⁸ Roberto Poli Michael Healy Achilles Kameas Theory and Applications of Ontology: Computer Applications Springer Science+Business Media B.V. 2010

ma'lumot keltiriladi. Nerr formal ontologiyani rivojlantirishda muammolardan biri sifatida sohaviy tarmoqlar rivojlanib borayotganligi va yangi tushunchalar paydo bo'layotganligi leksik qatlamning boyitish hamda ular orasidagi munosabatlar ham o'zgarishiga ta'sir etadi. Yangi tushunchalarning mavjud emasligi so'rov bo'yicha qidiruv uchun semantik texnologiyalardan foydalanishga imkon bermaydi.

Kompyuter sohasida 1967-yilda S.H.Mealy tomonidan birinchi marta ontologiya termini qo'llanila boshlangan. Kompyuter sohasida sun'iy intellekt tushunchasi paydo bo'lgandan so'ng sohaviy ontologiyadan foydalanish boshlandi. Ontologiya muayyan soha doirasida muloqot qilishga mo'ljallangan terminlardan iborat bo'ladi.

Ontologiya haqida Jeff Heflin shunday fikrni o'rtaga tashlaydi [Heflen, 2004]: ontologiya bilimlar sohasini ifodalovchi va tasvirlovchi terminlarni aniqlaydi. Gomez-Perez [Gomez-Perez, 2003]ning e'tiroficha, turli ma'lumotlar bazasi va internet, tegishli obyektlardan ma'lumotlar qidiruvining asosi, degan fikrlarni beradi. Ontologiya turli olimlarning qarashlari va fikrlaridan kelib chiqib uning tasniflari keltiriladi. Guarino [Guarino, 1998] ontologiyalarni vazifasiga ko'ra munosabat va maxsus vazifasiga ko'ra tobelanish darajasiga ko'ra to'rt kategoriyaga ajratadi: yuqori daraja, soha, so'rov va ontologiya ilovalari. Lassila [Lassila and McGuinness., 2001] esa ontologiyalarni ularning ta'sirchanlik va berilgan ma'lumotlarni to'liqliligiga ko'ra ajratadi.

O'zbek lug'atshunosligida sinonim, omonim va antonimlar alohida tarzda tuzilgan bo'lib, ularni tezaurus shaklga keltirish matnlarni qayta ishlash va mashina tarjimasiga leksikasi uchun muhim rol o'ynaydi.

Zamonaviy ontologiya quyidagi turlarga ajraladi¹²⁹:

1) diskriptiv ontologiya biror soha doirasida analiz qilinadigan yoki umumiy ma'noda ma'lumotlarning majmui ma'nosida qo'llaniladi;

2) formal ontologiya diskriptiv ontologiyadagi natijalarni

¹²⁹ Poli, R. Descriptive, formal and formalized ontologies. In *Husserl's logical investigations reconsidered*, ed. D. Fissette, and D. Kluwer, Dordrecht: Kluwer, 2003. -P. 193-210.

saralash, filtrlash, kodlash kabi vazifalarni bajarib, uni boshqaradi;

3) formallashgan ontologiya mantiqiy tilda diskriptiv va kategorial konstruksiyalar uchun munosib bo'lgan formal kodlashtirishni nazarda tutadi.

Ontologiya dastlab kompyuter sohasi va sun'iy intellekt sohasida termin sifatida 1980-yillarning oxirida, Gruber 1993-yillarda o'zining ilmiy tadqiqotlarida ontologiyani "konseptuallashgan maxsus soha" sifatida e'tirof etadi.

Ontologiyada modelning o'ziga xos xususiyatlari haqiqiy obyektga (obyekt bilan bog'liq) o'xshash bo'lishi kerak, deb qaraladi. Olimlarning nuqtai nazaricha ontologiya tuzilishining quyidagi to'rt modeli mavjud:

1. Verbal modellar
2. Mantiqqa asoslangan model
3. Struktur model
4. Gibrid model

Ontologiya ontologik modelning interfeys orqali bir nechta ma'lumotlar bazasi iyerarxik tarzda konseptual tuzilmani ta'minlashda MB sifatida tushuniladi. Ontologiya quyidagi tartibdan iborat: 1) sinflar; 2) atributlar; 3) namunalar; 4) aksiomalar.

Hendler va van Harmelen¹³⁰ ontologik tillar sifatida quyidagilarni RDF, RDFS, DAML+OIL, OWL, XML, XHTML va KR qayd qiladilar. Ontologiya yaratishning samarali usuli OWL tilidir.

Ontologiya uchun maxsus dasturiy ilovalar ishlab chiqilgan: Anzo (RDFS va OWL ontologiya muharririni o'z ichiga oladi), CmapTools Ontology Editor, HOZO, Java Ontology Editor (JOE), KAON.

- Javaga asoslangan ontologiya muharriri, ko'plab formatlarni qo'llab-quvvatlaydi).

Turkiy tillar (o'zbek, qozoq, tatar, qirg'iz, turk)ning metatilini yaratish uchun ontologiyadan foydalanish samarador usul

¹³⁰ Hendler, J., and F. van Harmelen. 2007. The semantic web: webizing knowledge representation. In *Handbook of Knowledge Representation, Foundations of Artificial Intelligence*, eds. F. Van Harmelen, V. Lifschitz, and B. Porter, ch. 21, Amsterdam: Elsevier. -P. 821-839.

muammolardan biri sifatida sohaviy tarmoqlar rivojlanib borayotganligi va yangi tushunchalar paydo bo'layotganligi leksik qatlamning boyitish hamda ular orasidagi munosabatlar ham o'zgarishiga ta'sir etadi. Yangi tushunchalarning mavjud emasligi so'rov bo'yicha qidiruv uchun semantik texnologiyalardan foydalanishga imkon bermaydi.

Kompyuter sohasida 1967-yilda S.H.Mealy tomonidan birinchi marta ontologiya termini qo'llanila boshlangan. Kompyuter sohasida sun'iy intellekt tushunchasi paydo bo'lgandan so'ng sohaviy ontologiyadan foydalanish boshlandi. Ontologiya muayyan soha doirasida muloqot qilishga mo'ljallangan terminlardan iborat bo'ladi.

Ontologiya haqida Jeff Heflin shunday fikrni o'rtaga tashlaydi [Heflen, 2004]: ontologiya bilimlar sohasini ifodalovchi va tasvirlovchi terminlarni aniqlaydi. Gomez-Perez [Gomez-Perez, 2003]ning e'tiroficha, turli ma'lumotlar bazasi va internet, tegishli obyektlardan ma'lumotlar qidiruvining asosi, degan fikrlarni beradi. Ontologiya turli olimlarning qarashlari va fikrlaridan kelib chiqib uning tasniflari keltiriladi. Guarino [Guarino, 1998] ontologiyalarni vazifasiga ko'ra munosabat va maxsus vazifasiga ko'ra tobelanish darajasiga ko'ra to'rt kategoriyaga ajratadi: yuqori daraja, soha, so'rov va ontologiya ilovalari. Lassila [Lassila and McGuinness., 2001] esa ontologiyalarni ularning ta'sirchanlik va berilgan ma'lumotlarni to'liqliligiga ko'ra ajratadi.

O'zbek lug'atshunosligida sinonim, omonim va antonimlar alohida tarzda tuzilgan bo'lib, ularni tezaurus shaklga keltirish matnlarni qayta ishlash va mashina tarjimasini leksikasi uchun muhim rol o'ynaydi.

Zamonaviy ontologiya quyidagi turlarga ajraladi¹²⁹:

1) diskriptiv ontologiya biror soha doirasida analiz qilinadigan yoki umumiy ma'noda ma'lumotlarning majmui ma'nosida qo'llaniladi;

2) formal ontologiya diskriptiv ontologiyadagi natijalarni

¹²⁹ Poli, R. Descriptive, formal and formalized ontologies. In Husserl's logical investigations reconsidered, ed. D. Fiset, and D. Kluwer, Dordrecht: Kluwer, 2003. -P. 193-210.

3) formallashtirilgan ontologiya mantiqiy tilda diskriptiv va kategorial konstruksiyalar uchun munosib bo'lgan formal kodlashtirishni nazarda tutadi.

Ontologiya dastlab kompyuter sohasi va sun'iy intellekt sohasida termin sifatida 1980-yillarning oxirida, Gruber 1993-yillarda o'zining ilmiy tadqiqotlarida ontologiyani "konseptuallashtirilgan maxsus soha" sifatida e'tirof etadi.

Ontologiyada modelning o'ziga xos xususiyatlari haqiqiy obyektga (obyekt bilan bog'liq) o'xshash bo'lishi kerak, deb qaraladi. Olimlarning nuqtai nazaricha ontologiya tuzilishining quyidagi to'rt modeli mavjud:

1. Verbal modellar
2. Mantiqqa asoslangan model
3. Struktur model
4. Gibrid model

Ontologiya ontologik modelning interfeys orqali bir nechta ma'lumotlar bazasi iyerarxik tarzda konseptual tuzilmani ta'minlashda MB sifatida tushuniladi. Ontologiya quyidagi tartibdan iborat: 1) sinflar; 2) atributlar; 3) namunalar; 4) aksiomalar.

Hendler va van Harmelen¹³⁰ ontologik tillar sifatida quyidagilarni RDF, RDFS, DAML+OIL, OWL, XML, XHTML va KR qayd qiladilar. Ontologiya yaratishning samarali usuli OWL tilidir.

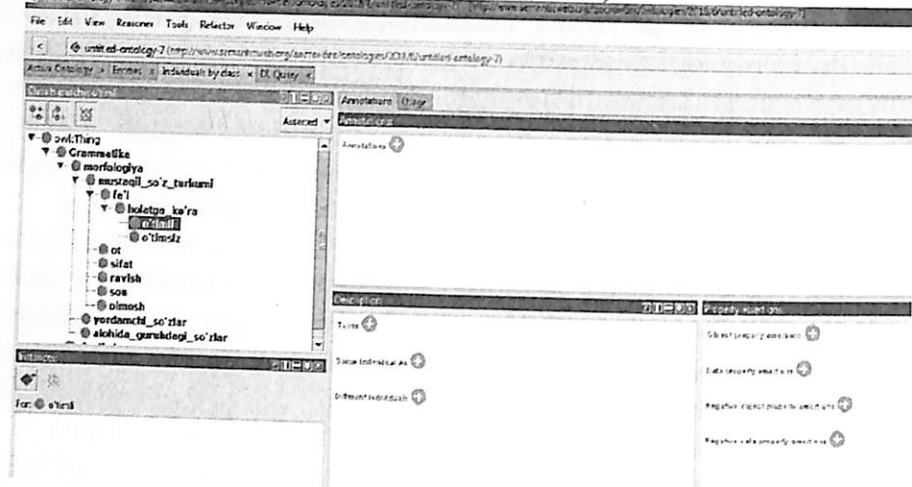
Ontologiya uchun maxsus dasturiy ilovalar ishlab chiqilgan: Anzo (RDFS va OWL ontologiya muharririni o'z ichiga oladi), CmapTools Ontology Editor, HOZO, Java Ontology Editor (JOE), KAON.

- Javaga asoslangan ontologiya muharriri, ko'plab formatlarni qo'llab-quvvatlaydi).

Turkiy tillar (o'zbek, qozoq, tatar, qirg'iz, turk)ning metatilini yaratish uchun ontologiyadan foydalanish samarador usul

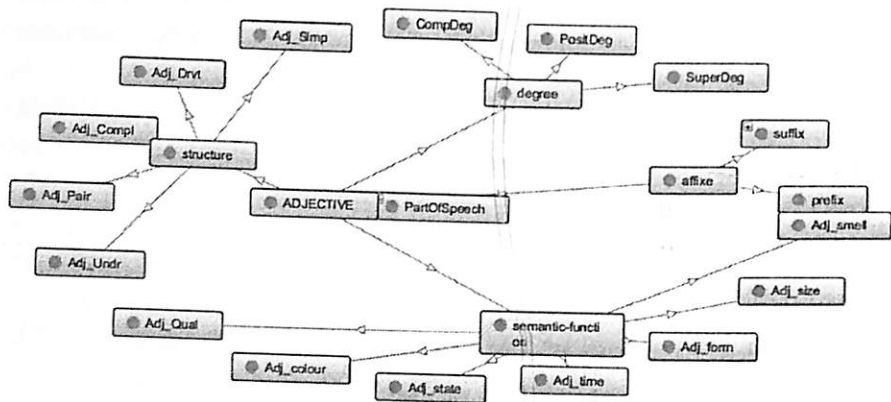
¹³⁰ Hendler, J., and F. van Harmelen. 2007. The semantic web: webizing knowledge representation. In Handbook of Knowledge Representation, Foundations of Artificial Intelligence, eds. F. Van Harmelen, V. Lifschitz, and B. Porter, ch. 21, Amsterdam: Elsevier. -P. 821-839.

hisoblanadi. Quyida o'zbek tilining grammatik qoidalarga oid bilimlar Protégé dasturiga kiritiladi (17-rasm):



17-rasm

Tur va tasnifning grafigi quyidagicha bo'ladi (10-chizma):



Ontologiya rasmiy, maxsus terminlar va sohaviy doirasi aniq tushunchalar o'rtasidagi aloqalarni iyerarxik bog'lash uchun ishlatiladi. Ontologiya qoida asosidagi mashina tarjimasini, axborot qidirish, matn tahlili kabi yo'nalishlarda foydalaniladi. Grammatik ontologiyani yaratishda morfologik teglardan foydalaniladi.

O'zbek va qozoq tillarining ot so'z turkumi doirasida grammatik qoidalarning ontologiyasini solishtirilsa, unda umumiy kategorial

birliklar mavjudligini aniqlaymiz: {SIMP, CMPL, FUSW, PAIR, CMPN, ABBR, UNDR, DRVT, COMP, ANIM, INAM, CMMN, PRPR, CNCR, ABST, Num., SG, COL}, ularning tarkibida esa farqli jihatlarni topish qiyin emas.

Kelishik tegi	O'zbek	Qozoq
NOM	Ø	Ø
GEN	-ning	-ның, -нің, -дың, -дің, -тың, -тің
DIR	-ga, -ka, -qa	-ға, -ге, -қа, -ке, -на, -не, -а, -е
ACC	-ni	Ны
LOC	-da	-да, -де, -та, -те, -нда, -нде
ABL	-dan	-нан, -нен, -дан, -ден, -тан, -тен
INST	Ø	-мен, -бен, -пен, -менен, -бенен, -пенен

Son tegi	O'zbek	Qozoq
Pl	-lar	-лар, -лер, -дар, -дер, -тар, -тер

Protégé dasturida ontologik modellashtirish so'zlarning tur va tasnif, obyekt xususiyatlari, ma'lumotlar tavsifi, birliklar, annotatsiyalash kabi parametrlarga ko'ra amalga oshiriladi.

Ontologik instrumentlar

Ontologiya uchun turli instrumentlardan ma'lumotlarni tasniflash, annotatsiyalash, ma'lumotlarni tavsiflash va solishtirishda foydalanib kelinmoqda. Shulardan ayrimlari haqida ma'lumot beriladi¹³¹:

Text2Onto- matn resurslaridan ontologiyani o'rganish uchun

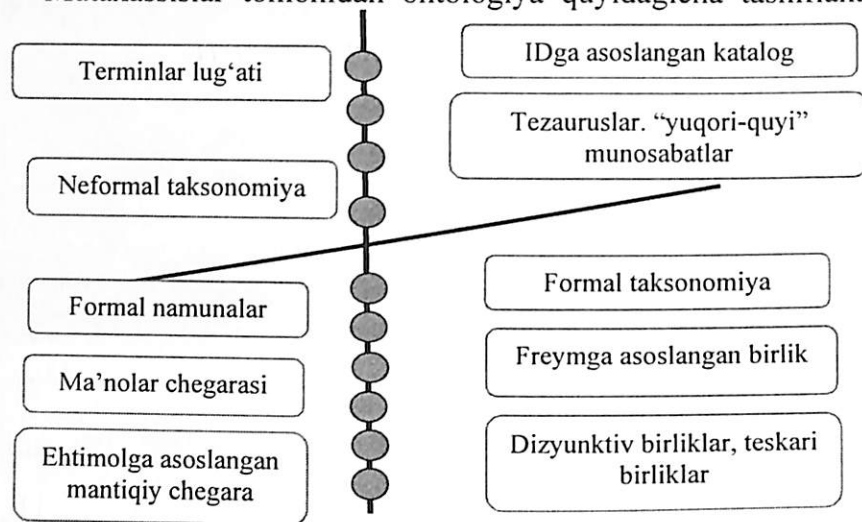
OntoGen- ijtimoiy tarmoqlar orqali ontologiyani yaratish uchun

OntoBuilder - XML yoki HTML tilidagi matnlardan ontologiyani hosil qilish uchun ishlatiladi. Bir nechta ontologiyalardan hosil bo'lgan instrumentlar ham mavjud: *OntoMerge*, *PROMPT*, *FCA-Merge*, *HCONEmerge*, *IF-Map*. Shuningdek, *Match* operatori orqali o'xshash manbalarni solishtirishda ularning ontologiyasini yaratishda quyidagi instrumentlardan foydalaniladi: *Falcon-AO*, *OLA* kabilar. Bir nechta ontologiyalarda aks etgan ma'lumotlar bazasida semantik

¹³¹ Roberto Poli, Michael Healy, Achilles Kameas Theory and Applications of Ontology: Computer Applications Springer Science+Business Media B.V., 2010.

integrallashuv yoki bilimlar almashinuviga erishishda CROSI Mapping System, GLUE, MAFRA, OntoMap R, H-Match kabi instrumentlardan ham foydalaniladi.

Mutaxassislar tomonidan ontologiya quyidagicha tasniflanadi:



7.4. TERMINOLOGIYA VA TERMINOGRAFIYA

Matnlardagi texnikaviy terminlarni o'rganish va ularning muqobil variantlarini avtomatik tarjima qilishda terminologiya sohasining o'rnini katta. Terminlarning turli shakldagi lug'atlarini yaratish terminografiya¹³² (terminologik lug'atshunoslik) sohasida o'rganiladi. "Ko'p tilli terminografiyaning dolzarbligi tez fursatda o'sib borayotgan avtomatlashgan lug'atlar va terminologik ma'lumotlar bankining paydo bo'lgani bilan izohlanadi¹³³". Ma'lumotlarga ko'ra XX asrda nemis tilining barcha ilmiy-texnikaviy terminlar miqdori 1 mln.ni tashkil etgan. XX asr o'rtalariga kelib elektrotexnikaga doir terminlarning o'zi 4 mln.ga yetganligi aniqlangan.

Terminologik lug'atlarni strukturalash va ularni ma'lumotlar bazasida loyihalash muhim masalalardan biri. Mashina tarjimasida terminografiyaning alohida o'rnini bor. Mashina tarjimasida

terminlarga doir sohaviy lug'atlarning yaratilishi tarjimaning sifatini oshiradi. Terminologiya hamisha o'zgarib boradi. Terminografiyada quyidagi masalalar o'rganiladi:

1) u yoki bu terminning ikkinchi bir tilda aynan mavjud bo'lmashligi;

2) terminlarning so'z birikmalari tarzida ifodalanishi;

3) matnning uslubini aniqlash va ilmiy terminlarning ma'nolarini uslubga ko'ra o'zgarishi;

4) muayyan soha doirasida bir leksemaning turli ma'noda qo'llanilishi;

5) muayyan terminning bir necha sohada turli shaklda ifodalanishi.

Masalan, texnika sohasida qo'llaniladigan inglizcha *pin* termini ot so'z turkumi sifatida quyidagi ma'nolarga ega: *barmoq*, *shtift*, *shpilka*, *shplint* va h.k., shuningdek, ushbu leksema boshqa so'z turkumi sifatida ham turli ma'nolar kasb etadi: *hayotini buzmoq*, *so'zidan ilib olmoq* va h.k.

Olimlar fan va texnikaga oid terminlarni standartlashtirish masalasiga alohida munosabatlarni fikrlarni o'rtaga tashlaydilar. Muayyan terminning iste'molga kirib kelishida uning yaratuvchisi yoki ijtimoiy hodisalar ta'sirida ba'zilarining o'rniga boshqa birining qo'llanish holatlarini inkor etib bo'lmaydi. Lingvistik ta'minotda terminlar lug'atining to'liq berilishi tarjima sifatini oshirishga xizmat qiladi.

Ilmiy matnlarning leksik birliklari quyidagi razryadlarga ko'ra bo'linishi mumkin:

- adabiy tilda qabul qilingan so'zlar: *predlog*, *yordamchi so'zlar*, *olmosh*

- umumadabiy tilda qabul qilingan, qoidaga ko'ra ilmiy matnda muayyan ma'noga ega so'zlar: *qism*, *shart*, *qila olmoq*

- ilmiy bo'lmagan matnlarda juda kam uchraydigan va aynan ilmiy matnga xos xarakterni ko'rsatuvchi so'zlar: *tahlil qilmoq*, *tasniflamoq*, *metod*, *neyro*.

- frazeologik birikmalar: *ehtiyojni qondirmoq*, *komponentlarning ta'sirlashuvi...*

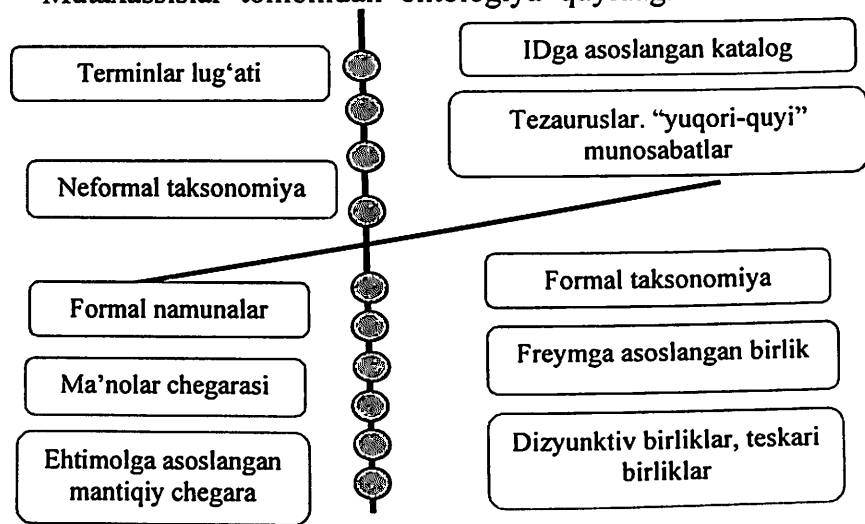
- muayyan fan doirasida maxsus terminologik birliklar: *ehtimollik*, *to'plam*, *formal til*.

¹³² Terminologik lug'atlarni tuzish haqidagi fan.

¹³³ Марчук Ю. Компьютерная лингвистика. – М., 2007. – С. 196.

Mapping System, GLUE, MAFRA, OntoMap R, H-Match kabi instrumentlardan ham foydalaniladi.

Mutaxassislar tomonidan ontologiya quyidagicha tasniflanadi:



7.4. TERMINOLOGIYA VA TERMINOGRAFIYA

Matnlardagi texnikaviy terminlarni o'rganish va ularning muqobil variantlarini avtomatik tarjima qilishda terminologiya sohasining o'rni katta. Terminlarning turli shakldagi lug'atlarini yaratish terminografiya¹³² (terminologik lug'atshunoslik) sohasida o'rganiladi. "Ko'p tilli terminografiyaning dolzarbligi tez fursatda o'sib borayotgan avtomatlashgan lug'atlar va terminologik ma'lumotlar bankining paydo bo'lgani bilan izohlanadi¹³³". Ma'lumotlarga ko'ra XX asrda nemis tilining barcha ilmiy-texnikaviy terminlar miqdori 1 mln.ni tashkil etgan. XX asr o'rtalariga kelib elektrotexnikaga doir terminlarning o'zi 4 mln.ga yetganligi aniqlangan.

Terminologik lug'atlarni strukturalash va ularni ma'lumotlar bazasida loyihalash muhim masalalardan biri. Mashina tarjimasida terminografiyaning alohida o'rni bor. Mashina tarjimasida

¹³² Terminologik lug'atlarni tuzish haqidagi fan.

¹³³ Марчук Ю. Компьютерная лингвистика. – М., 2007. –С. 196.

oshiradi. Terminologiya hamisha o'zgarib boradi. Terminografiyada quyidagi masalalar o'rganiladi:

- 1) u yoki bu terminning ikkinchi bir tilda aynan mavjud bo'lmashligi;
- 2) terminlarning so'z birikmalari tarzida ifodalanishi;
- 3) matnning uslubini aniqlash va ilmiy terminlarning ma'nolarini uslubga ko'ra o'zgarishi;
- 4) muayyan soha doirasida bir leksemaning turli ma'noda qo'llanilishi;
- 5) muayyan terminning bir necha sohada turli shaklda ifodalanishi.

Masalan, texnika sohasida qo'llaniladigan inglizcha *pin* termini ot so'z turkumi sifatida quyidagi ma'nolarga ega: *barmoq*, *shtift*, *shpilka*, *shplint* va h.k., shuningdek, ushbu leksema boshqa so'z turkumi sifatida ham turli ma'nolar kasb etadi: *hayotini buzmoq*, *so'zidan ilib olmoq* va h.k.

Olimlar fan va texnikaga oid terminlarni standartlashtirish masalasiga alohida munosabatlarni fikrlarni o'rtaga tashlaydilar. Muayyan terminning iste'molga kirib kelishida uning yaratuvchisi yoki ijtimoiy hodisalar ta'sirida ba'zilarining o'rniga boshqa birining qo'llanish holatlarini inkor etib bo'lmaydi. Lingvistik ta'minotda terminlar lug'atining to'liq berilishi tarjima sifatini oshirishga xizmat qiladi.

Ilmiy matnlarning leksik birliklari quyidagi razryadlarga ko'ra bo'linishi mumkin:

- adabiy tilda qabul qilingan so'zlar: *predlog*, *yordamchi so'zlar*, *olmosh*

- umumadabiy tilda qabul qilingan, qoidaga ko'ra ilmiy matnda muayyan ma'noga ega so'zlar: *qism*, *shart*, *qila olmoq*

- ilmiy bo'lmagan matnlarda juda kam uchraydigan va aynan ilmiy matnga xos xarakterni ko'rsatuvchi so'zlar: *tahlil qilmoq*, *tasniflamoq*, *metod*, *neyro*.

- frazeologik birlikmalar: *ehtiyojni qondirmoq*, *komponentlarning ta'sirlashuvi...*

- muayyan fan doirasida maxsus terminologik birliklar: *ehtimollik*, *to'plam*, *formal til*.

- ilmiy matnda uchramaydigan, biror muayyan fanning predmeti sifatida umumadabiy tilda qabul qilingan birliklar: *aborigen, absurd*.

Mashina tarjimasini yaratishda zamonaviy tilning qancha leksik boyligini o'z ichiga qamrab olinishi ko'pchilikni qiziqtiradi. Xalqaro tashkilot INFOTERM (Avstriyaning Vena shahridagi Xalqaro terminologik markazi) bergan ma'lumotlarga tayaniladigan bo'lsa, Yevropa tillariga oid terminlar soni 50 million, mahsulot (tovar) nomlarining miqdori 100 millionga yetadi. Bu shuni anglatadiki, jamiyatda yaratiladigan ishlab chiqarilgan har qanday mahsulot yoki biror yangilik muayyan nomga ega bo'ladi va termin sifatida leksikografiyani boyitadi.

Logos (LOGOS), Engspan (ENGSPAN) hamda Sistran (SYSTRAN) mashina tarjimasini tizimlari transfer usulidan foydalanadilar.

1964-yili Koloradodagi tashkil etilgan Sistran kompaniyasining tarjima tizimida 27 til kombinatsiyasi mavjud. So'zlar umumiy hamda texnikaviy ma'noga (termin sifatida) ega bo'lib, so'z kombinatsiyalarining sintaktik va semantik tahlili asosiga yaratilgan. Semantik tasniflangan tizim iyerarxik tarzda 500ta kategoriyaga oid strukturadan iborat. Terminologiyaning muhim vazifalardan biri ko'p ma'noli terminlarni to'g'ri tarjima qilishdir. Sistran tizimida tarjima jarayonining transfer bosqichi uchun maxsus dasturlar ishlab chiqilgan. Unga ko'ra dastlab matn tahlilida so'zning jonli yoki jonsiz ekanligi aniqlanadi. Lug'atdagi biror so'z [+jonli / jonsiz] tarzida kodlangan: (masalan, *seal* (1) *tyulen*; 2) *muhr, tamg'a*), *acquaintance* (1) *tanish*; 2) *tanishish* kabi). Tematik glossariylarda tarjima qilinayotgan so'zning ikkinchi tilda bir nechta ma'nolar bo'lishi mumkin. Jumladan, ingliz tilida *stem* termini bir nechta sohalarda qo'llanilib, ikkinchi tilda turlicha tarjima qilinadi: botanikada *plant stem* - "o'simlik ildizi", anatomiyada *brain stem* - "miya tomiri", tilshunoslikda *word stem* - "so'z negizi" kabi. Ushbu tizimda foydalanuvchi tarjimadan avval ilova etilgan o'nta sohadan birini tanlagandan so'ng, tarjima jarayoni amalga oshiriladi. Bu esa muayyan soha uchun ajratilgan so'zlarni tasniflab berish imkoniyatini beradi. Bunday masalada otli so'z birikmalarni tahlil

qilish uchun korpuslar ahamiyatga ega¹³⁴. Bu borada LOGOS mashina tarjimasini tizimida jiddiy natijalarga erishilgan. Bu tizimda so'zlardagi semantik noaniqliklarni aniqlash uchun turli sohalarga doir besh yuzdan ziyod maxsus ichki lug'atlar kiritilgan. Unda 130 dan ortiq iyerarxik tarzda tuzilgan semantik kategoriyalar mavjud. Shuningdek, VISTA (inglizcha-ruscha, ruscha-inglizcha) mashina tarjimasini tizimi 3 mln. 300 ming leksik birlikni o'z ichiga oladi. Bularning 80% ini uzunligi ikki va o'n yettitagacha bo'lgan so'z birikmalari tashkil etadi. Bunga qo'shimcha tarzda tematik 400 mingdan ortiq leksik birlikdan iborat lug'at kiritilgan¹³⁵. Mashina tarjimasini uchun terminologik lug'atni lingvistik ta'minot sifatida kiritish, ularning umumiy va xususiy ma'nolarini kodlash orqali ingliz tilidagi matnlarni o'zbek tiliga tarjima qilish imkoniyati kengayadi. Buning uchun tanlab olingan sohalarda doirasida ikki tilning terminologik lug'atlarini dasturiy ta'minotga kiritish, ularning semantik tahlilini berish hamda terminlarni standartlashtirish lozim bo'ladi. Chunonchi, ijtimoiy-gumanitar fanlar bo'yicha terminlarning bir xilligini ta'minlash maqsadga muvofiq. Bu masala yuzasidan Xalqaro standartlashtirish tashkiloti (ISO-International Standardization Organization) tomonidan terminlarni standartlashtirish masalalari o'rganilib kelinmoqda. Shu o'rinda bu kabi ajratishning nazariy asoslari yana bir qator asarlarda o'z tasdiqini topganligini e'tirof etamiz¹³⁶.

A.A. Reformatskiyning nuqtai nazaricha, terminologiya og'zaki ifodada mos ravishda mustahkamlangan, muayyan fanning tizimlashtirilgan sohasidir¹³⁷. Mashina tarjimasida terminologiyani alohida soha sifatida o'rganish shu jihatdan zarur hisoblanadi. Terminlarning yasaliishi uch usul bilan amalga oshiriladi: semantik, morfologik, sintaktik. Terminlarning semantik usul bilan yasaliishiga metoforik va metonimik ma'nolarning boshqa terminologiya bilan bog'liq holda xususiylashishi xos. Bunga

¹³⁴ Qarang: <http://www.murieltranslations.com/articles/terminology/term>.

¹³⁵ <http://www.retrans.ru/publikacii/statyi/MT>

¹³⁶ М.Н. Литвинова Практикум по переводу (английский язык): учеб.-метод. пособие / авт.-сост. М.Н. Литвинова, Перм. гос. нац. иссл. ун-т. – Перм, 2012. –С. 74.

¹³⁷ Реформатский А.А. Что такой термин и терминология? / Вопросы терминологии. –М., 1961. –С. 49.

transpozitsiya va transmutatsiya hodisalarini keltirish mumkin. Morfologik usul bilan termin yasashda suffiksatsiya, prefiksatsiya, konversiya va asosni o'zlashtirish hodisalari yetakchilik qiladi. Sintaktik usul bilan termin yasashda transformatsion usulning o'rni sezilarli. Bundan tashqari konkretizatsiya, generilizatsiya, kompensatsiya, transliteratsiya, eksplikatsiya, amplifikatsiya, ma'no differensiyalashuvi va h.k. hodisalarni ham nazardan qochirmaslik kerak¹³⁸. Birgina mashhur inglizcha Webster lug'atidagi 600000 so'zdan 500000 tasining mutaxassislikka oid leksemalardan tashkil topgani terminologiyani chuqur tahlil etishni taqozo etadi¹³⁹. Zero, ilmiy-texnikaviy terminlar jamiyatning barcha sohalarida qo'llaniladi. "London Computer Integrated Translation GmbH" (1987) ko'rsatgan ma'lumotlarga ko'ra, dunyo bo'yicha ilmiy-texnikaviy matnlarni tarjima qilish hajmi yiliga 200 mln. sahifani tashkil etadi, bu esa har yili 15% oshib borayotganligini tasdiqlaydi¹⁴⁰. Shu o'rinda ilmiy matnlarni kompyuter yordamida tarjima qilish mazkur jarayonni optimallashtirishga ko'mak beradi. Ilmiy-texnikaviy matnlarni tarjima qilishda uning oldida turgan mas'uliyatli vazifalar borligini ko'rsatadi.

XX asrning o'rtalarida ilmiy matnlarni tarjima qilish an'anasi shakllandi. Ilmiy matnlarni tarjima qilishning strategik tushunchalari V.Z.Demyankov, R.Karnap, A.N.Kryukov, Y.S.Kubryakov, J.Lakoff, M.V.Nikitin, M.Terner, J.Fokone, R.Shenk kabi atoqli tilshunoslarning g'oyalaridan oziqlandi¹⁴¹.

¹³⁸ Qarang: Онанченко А.В. Способы образования переводческих терминов в русском языке. 2001

¹³⁹ Webster's Third New International Dictionary, Webster 3. 1961.

¹⁴⁰ Гул Хасан Лексико –семантические особенности перевода научно – технического текста с языка урду на русский язык: Дисс. ... док-ра филос.наук. – Исломобод, 2007. –С.18.

¹⁴¹ Qarang: Демуанков В.З. Понятие и концепт в художественной литературе и научном языке // Вопросы филологии. –М., 2001. –№1. – С. 35-47; Крюков А.Н. Перевод как интерпретация // Перевод и интерпретация текста. –М., 1988. – С. 41-65; Крюков А.Н. Понимание как переводческая проблема // Перевод и интерпретация текста. – М., 1988. – С. 65-76. Кубруакова Е.С. Номинативный аспект речевой деуателности. –М.: Наука, 1986. – С. 158; Кубруакова Е.С. Роль словообразования в формировании языковой картины мира // Роль человеческого фактора в языке: Язык и картина мира. – М.: Наука, 1988. – С. 212.; Лакофф Дж. Мышление в зеркале классификаторов // Новое в зарубежной лингвистике. Вып 23. Когнитивные аспекты языка. – М.: Прогресс. 1988. – С.12-52.

Terminografiya vazifalariga: 1) chegaralangan leksemalarning har bir lingvistik sathlar bo'yicha formal va modellashtirilgan tavsifini berish; 2) matn-matn munosabati o'rtasida turgan umumiy va farqli lisoniy paradigmalarni aniqlash; 3) avtomatlashgan terminologik bazani shakllantirish; 4) ilmiy matnlarga oid leksik, morfologik va sintaktik xususiyatlarni ikki til bo'yicha solishtirish va ularni o'xshash hamda farqli jihatlari bo'yicha kategoriyalarga ajratish; 5) ikki tilli ilmiy matnlarda qo'llanadigan leksemalarning miqdori bo'yicha sinonimik, omonimik va polisemantik tizimini tuzish orqali ularning dominantini bazaga kiritish kabilar kiradi.

Terminologiya – ilmiy matnning yadrosi bo'lgani tufayli terminlar ularning kalit so'zlaridir. Bu bo'yicha quyidagilar alohida o'rganilishi kerak: 1) yagona terminologik tizimda qo'llanuvchi leksemalar (*sintaksis, semantik analiz*); 2) bir terminologik tizimda turli ma'no anglatuvchi qo'llanuvchi leksemalar (*phrase –ibora, so'z birikmasi, tasvirlash*); 3) bir terminologik tizimda ma'nolari jihatdan bir-biriga yaqin bo'lgan sinonimlar (*lexeme –leksema, token – leksema: dasturlash tilida sintaktik birlik, lexical item –leksema: leksik (sintaktik) tahlil elementi*); 4) turli terminologik tizimga kiruvchi omonimlar (*stem: word stem – so'z asosi (lingvistika); stem –qabila, jins (iqtisod); stem –qarindoshlik aloqasi (qonun); stem hair–soch ildizi (biologiya); guide stem –boshqaruvchi qism (texnika) v.b.*).

Terminografiyada muhim vazifalardan biri abbreviaturalarning semantik imkoniyatlari o'rganishidir. Ilmiy matnlarda muayyan abbreviaturalar bir nechta sohalarida qo'llanilib, turli ma'no ifodalaydi. Xususan, CAT quyidagi ma'nolarga ega: *Channel-Activity Table* –faol kanallar jadvali (telekommunikatsiya); *Customer Activated Terminal* –aktivlashtirilgan terminal mijozi (marketing); *Corrective Action Team* –muammoli guruh (menejment); *Computerized Axial Tomography* –kompyuterlashtirilgan aksial tomografiya (tibbiyot); *Computer-Assisted (yoki-Aided) Translation* –kompyuterlashtirilgan tarjima (kompyuter lingvistikasi) h.k. Mashina tarjimasida termini tashkil etuvchi komponentlar miqdoriga ko'ra ajratish va ularning tarjimasini to'liq va to'g'ri ma'lumotlar bazasiga kiritish zaruriy masalalardan sanaladi.

Bir tarkibli	Ikki tarkibli	Uch tarkibli	To'rt tarkibli	Besh tarkibli
system	control	aircraft	fly-by-wire	digital fly-by-wire
–	system	control	aircraft control	aircraft control
sistema,	–	system	– system	– system
tizim	boshqarish tizimi	samolyot boshqarish tizimi	samolyotning elektrodistsion boshqarish tizimi	samolyotning raqamli elektrodistsion boshqarish tizimi

Bundan tashqari, kontekstologik qisqartmalar ham mavjud: *n* – noun, *vt* – transitive verb h.k. Ingliz, nemis, fransuz tilidagi ijtimoiy sohalarga oid matnlarda lotin va ba'zan yunon tiliga oid qisqartmalardan foydalanish holatlari ham kuzatiladi: *etc.* – et cetera (va hokazo), *i.e.* – id es (shuningdek, masalan) va h.k.

Ingliz tili Oksford (Oxford) lug'atining 89%ini omonimlar, polisemantik so'zlar, 9.1%ini monosemantik so'zlar tashkil etadi¹⁴². Shu bois ilmiy matnlarni tahlil qilishda tarjima sifatini oshirish uchun olimlar quyidagilarni ko'rsatib o'tadilar¹⁴³: 1. LS (leksik-semantik)-guruhlarini aniqlashtiruvchi ma'lumot: a) kontekstologik identifikator; b) xususiy-belgili identifikator SGr (sifat guruhi), FGr (fe'l guruhi); 2. LS-guruhlarining asos tilning shu til bo'yicha beriladigan ma'lumot: a) leksema nomlari; b) leksik-grammatik tipga kiruvchi guruh; c) leksik-semantik tipga kiruvchi guruh; d) sintaktik aloqalar guruhi; 3. Tarjimadan oldin LSda berilgan tillararo ma'lumot: a) ekvivalent so'zlarning kategorial qismi; b) "Moslashuvchi belgi"; c) "So'zlar tartibi".

Bu bo'yicha sohaviy terminologik baza yaratish va ularni matndagi ma'nolarini ifodalash uchun ilmiy tadqiqotlar amalga oshirilmoqda. Kompyuter lingvistikasi terminlarining inglizcha-o'zbekcha tarjima (izohi bilan) lug'ati yaratildi¹⁴⁴. Ushbu lug'at mashina tarjimasiga uchun kelgusi ishlarda lingvistik baza vazifasini o'taydi. Elektron lug'atlar nafaqat til o'rganish jarayonini tezlashtirish va qulaylikka erishish, balki maxsus maqsadlar uchun

ham zaruriy ehtiyojga aylanmoqda. O'zbek tili bo'yicha ko'plab lug'atlar nashr etilgan. Ushbu lug'atlar kitob shaklida bo'lib, kompyuterga moslashtirilmagan. Shu bois matn tahlili, mashina tarjimasiga oid ko'plab tadqiqotlar kompyuter lug'atshunosligiga bog'liq bo'lganligi bois ayrim muammolar o'z yechimini topganicha yo'q. Shu bois tarjimashunoslikda katta hajmli raqamli lug'atlar yaratishning metodologiyasi ishlab chiqilishi kerak.

Inglizcha-o'zbekcha mashina tarjimasiga uchun kompyuter leksikografiyasi muhim o'ringa ega. Unda o'zbek tilidagi so'zlarning quyidagi semantik maydonlariga ko'ra ajratiladi:

- I. So'zning leksik ma'nolari
- II. So'z turkumi
- III. O'z qatlam yoki o'zlashgan qatlamga oid ekanligi
- IV. So'zning turi va grammatik shakliga ko'ra
- V. So'z birikmasi shaklida kelishi
- VI. Ibora tarkibida kelishi
- VII. So'zning atama bo'lib kelishi
- VIII. So'zning dialektik shakli
- IX. Omonimlar
- X. Sinonimlar
- XI. Antonimlar
- XII. Uslubda qo'llanilishi

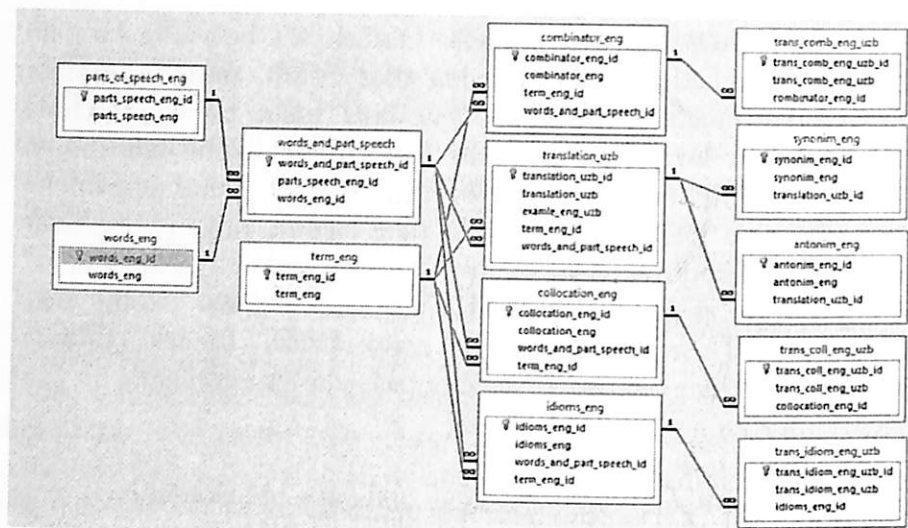
Elektron inglizcha-o'zbekcha tarjimon lug'at ma'lumotlar bazasidagi indekslangan birliklar o'zaro birga - ko'p va ko'pga - ko'pga ko'rinishda bog'lanishi mumkin va hamma jadval inglizcha so'z va uning so'z turkumiga bog'liq holda hosil qilinadi.

Berilganlar bazasi daraxtsimon tuzilishga ega bo'lib, yaratilgan elementdan boshlang'ich elementga oddiy o'tish mumkinligi, ya'ni yaratilgan element uchun faqat bitta boshlang'ich element mavjud va teskari o'tish imkoniyatini (ya'ni bittadan ko'plikka) yaratadi (16-rasm).

¹⁴² http://en.coolreferat.com/Homonyms_in_English_and_their_specific_features

¹⁴³ Л.С.Модина, З.М. Шалуапина Межъязыковые операции в модели японско-русского автоматического перевода ЯРАП^{143[1]}<http://www.dialog-21.ru/Archive/Directions/JPTRANSF.htm>

¹⁴⁴ Ushbu mobil ilova G'aniyeva Sh. hamkorligida tayyorlangan.



16-rasm

Ma'lumotlar bazasidagi maqolalar har ikki tomonlama foydalanish imkoniyati borligi bilan ajralib turadi.

Agar semantik maydon $W=W_1, W_2, W_3, \dots, W_n$ tuzilishda bo'lsa, u holda so'zlarning sinonimi $W=W_1\{A, B, C\}, W_2\{A, B, C\}$ bo'ladi. Shunda boshqa so'zga birlashtirilgan sinonimlar to'plamidagi so'z boshqa so'z bilan ham sinonimik qatorga birlashadi. Bundan tashqari turg'un birikmasi yoki qo'shma so'z holatidagi shakllari ham inobatga olinadi.

Lug'atlarning formati ikki jihatdan farqlanadi: mashina o'qiy oladigan va elektron shakldagi lug'atlar. Biroq elektron shakldagi tayyor lug'atlardan ma'lumotlar bazasi sifatida ham foydalanish mumkin va aksincha. Elektron lug'atlar foydalanuvchiga ilovadan foydalanishning qulay imkoniyat yaratishi, qulay hamda iqtisodiy tejamkorligi jihatidan samarador hisoblanadi. Shu bois elektron lug'atlarning ma'lumotlar bazasi to'g'ri loyihalashtirilishi kerak. Murakkab mantiqiy strukturalar uchun bir nechta ma'lumotlar bazasi mavjud: FileMaker Pro, SQL, Oracle va sodda ma'lumotlar bazasi: Microsoft Excel, Access, Microsoft Word kabilar. Masalan, mobil ilova uchun Android operatsion sistemasini uchun mo'ljallangan "Inglizcha-o'zbekcha va o'zbekcha-inglizcha ilmiy terminlarning izohli lug'ati" (17-rasm) ning funksional imkoniyatlarini tanishtirish

orqali elektron lug'atning ma'lumotlar bazasini loyihalash haqida qisqa ma'lumot beriladi.

Ushbu ilova SQLite ma'lumotlar bazasi boshqaruv tizimida ishlaydi. Ma'lumotlarni loyihalash informatsion tizim doirasining bosqichlaridan biri sanaladi. Yaratilgan ma'lumotlar bazasi uchinchi normal shakl hisoblanadi. Har bir maydon takrorlanmas va bo'linmas bo'lishi lozim. Belgilangan kalit yagona holatda kiritiladigan birliklarni aniqlaydi. Hech qaysi bir kalitga ega bo'lmagan maydon boshqa kalitga ega bo'lmagan maydonga bog'lanmaydi. Ushbu tasvirdagi ma'lumotlar bazasi kompyuter lingvistikasi terminlarining inglizcha-o'zbekcha izohidan iborat. Faol qo'llaniladigan mobil ilovalar uchun Google Android OS uchun quyidagi ilovalar uni takomillashtirishga xizmat qiladi: Android SDK, Eclipse, Java and XML dastulash tillari, SQLite ma'lumotlarni boshqarish tizimi.

Terms	
id	
uzbWord	
uzbExpl	
engWord	
engExpl	

Ilova jarayonning tezligi va xotirani saqlash uchun xizmat qiladi. Bundan tashqari, qidiruv tizimda filter so'zlikning dastlabki harflari kiritilganda faqat o'sha harflar bilan boshlanadigan terminlarning ro'yhatini chiqarib beradi bu esa qidiruv tizimini yanada tezlashtirishga xizmat qiladi. Quyida dasturning kodi berilgan:

```
for (int i=0;i<arl.size();i++)
{
    if(arl.get(i).toLowerCase().startsWith(strName)
arl.get(i).toLowerCase().equals(strName))
    {
        res.add(arl.get(i));
    }
}
```

Boshqa lug'at bilan bog'liq qidiruv tizimi uchun quyidagi so'rov yaratiladi:

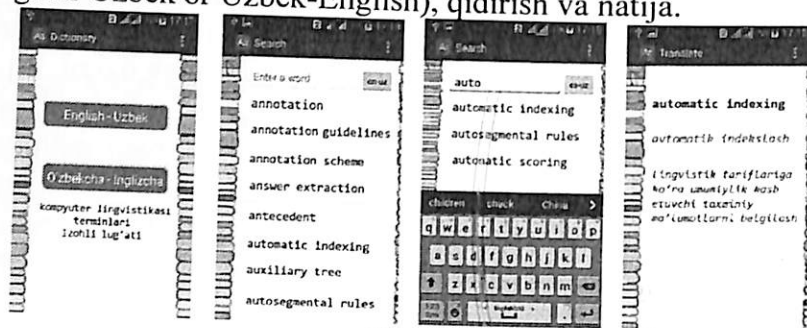
```
final List<Words> wlist = db.getAllUzbWords();
btn.setText("uz-en");
btn.setOnClickListener(new OnClickListener() {
```

```

public void onClick(View v) {
if(position==1)
{
    btn.setText("en-uz");position=2;
    Lang(db.getAllWords());
}
else if (position==2)
{
    btn.setText("uz-en");position=1;
    Lang(db.getAllUzbWords());
}
}
});

```

Ilova uch bosqichni o'z ichiga oladi: lug'atning tilini tanlash (English-Uzbek or Uzbek-English), qidirish va natija.



17-rasm

Adabiyotlar:

1. Collins English Dictionary. 8th Edition first published in 2006 (e-version).
2. John Nerbonne. Linguistic databases. 1996 <http://odur.let.rug.nl/~nerbonne/papers/intro-db.pdf>.
3. Andreea-Rosalina Olteanu. A holistic approach to phrasal verbs, Editura Sfântul Ierarh Nicolae 2012, - P. 16.
4. Хомоненко А.Д., Цыганков В.М., Мальцев М.Г. Базы данных: Учебник для высших учебных заведений / Под ред. проф. А.Д. Хомоненко. 2-изд. СПб.: КОРОНА принт, 2006. 672 с.
5. James Steele, Nelson To. The Android developers CookBook, Addison-Wesley.

6. Atkins S.R. Zampolli. 1994. Computational approaches to lexicon. Oxford press, Oxford 1994. – P.19.

7. Christiane Fellbaum. WORDNET an electronic lexical database the MIT Press Cambridge, Massachusetts London, England 1998 Massachusetts Institute of Technology. – 13 p.

8. https://en.wikipedia.org/wiki/Computational_lexicology

9. Mukanova, B. Yergesh, G. Bekmanova, B. Razakhova, A. Sharipbay. Formal models of nouns in the Kazakh language. Leonardo Electronic Journal of Practices and Technologies. ISSN 1583-1078. Issue 25, July-December, 2014 P. 264-273. ISSN: 15831078. SJR: 0.120

10. B. Yergesh, A. Mukanova, G. Bekmanova, Sharipbay A., B. Razakhova. Semantic Hyper-graph Based Representation of Nouns in the Kazakh Language. Computacion y Sistemas. Vol. 18, No. 3, 2014. pp. 627–635. ISSN 1405-5546. DOI: 10.13053/CyS-18-3-2041. SJR: 0.131

11. Арипов М., Абдурахмонова Н.З., Мухамедова С. О проблемах создания онтологии узбекского языка // Международная конференция “Азербайджанский язык: вчера и сегодня, посвященной 125—летнему юбилею Бекира Чобанзаде, Баку. - С. 120.

12. Константинова Н.С., Митрофанова О.А. Онтологии как системы хранения знаний. –Спб.: Санкт-Петербургский государственный университет, 2009. -С. 67-76.

13. Wikipedia.org

14. WWW. Ontologiya. Ru.

15. Armando Stellato. Dictionary, Thesaurus or Ontology. Disentangling Our Choices in the Semantic Web Jungle // Journal of Integrative Agriculture. 2012, 11(5). –S. 710-719.

16. Лукашевич Н.В. Тезаурусы в задачах информационного поиска. –Москва, 2010. –С.10-19.

17. Christiane Fellbaum. Wordnet. An Electronic Lexical Database. 1998 Massachusetts Institute of Technology, P. 23-44

18. Atkins, B. T. S., and Levin, B. (1991). Admitting impediments. In U. Zernik (Ed.), Lexical acquisition: Exploiting on-line resources to build a lexicon, 233-262.

19. Hillsdale, NJ: Erlbaum. Beckwith, R., Fellbaum, C, Gross,

D., and Miller, G. A. (1991). WordNet: A lexical database organized on psycholinguistic principles. In U. Zernik (Ed.), *Lexical acquisition: Exploiting on-line resources to build a lexicon*, 211-232.

20. Berlin, B., Breedlove, D., and Raven, P. H. (1973). General principles of classification and nomenclature in folk biology. *American Anthropologist*, 75, 214- 42.

21. Matlatipov S., Aripov M., Abdurakhmonova N. Modeling WordNet type Thesaurus for Uzbek language Semantic Dictionary / *International Journal of Engineering Research and Allied Sciences (IJERAS)* ISSN : 2455-9660, Volume-01, Issue-09, November 2016

Savol va topshiriqlar

1. Lug'atlarning tasniflanishi qaysi jihatlariga ko'ra farqlanadi?
2. Avtomatik referatlash va mashina tarjimasiga mo'ljallangan avtomatik lug'atlar foydalanuvchi uchun yaratilgan avtomatik lug'atlardan qanday jihatlari bilan farqlanadi?
3. Elektron lug'atlarning amaliy ahamiyati nimada?
4. Elektron lug'at turlarini sanang va ularning o'zaro farqini ayting.
5. Tezaurus tarkibidagi konseptlarga misollar keltiring.
6. Ontologiya, asosan, qanday komponentlardan tarkib topadi?
7. Tezaurusning boshqa lug'atlardan farqi va afzalliklarini sanang.
8. WordNetdagi leksemalarning semantik munosabatlari nimalarga asoslangan?
9. Terminografiya sohasining asosiy vazifalarini ayting.
10. Elektron lug'at uchun hosil bo'lgan ma'lumotlar bazasi blok sxemasi orqali namuna sifatida kichikroq hajmda lug'at tuzing. Undagi leksik birliklarning bog'lanish usulini tushuntirib bering.

7-mavzu yuzasidan test

1. Toksonomiya deb nimaga aytiladi?
 - a) boshqariladigan lug'at bo'lib, tarmoqli (shajarasimon) tuzilgan ajdod-avlod (ota-ona – farzand), butun – qism kabi boshqa munosabatlar asosida tuzilgan iyerarhik majmua
 - b) so'zlardagi murakkab munosabatlarni o'z ichiga oluvchi boshqariladigan lug'at turi
 - c) yunon tilidan olingan termin bo'lib, xazina ma'nosini beradi
 - d) tillarni avtomatik qayta ishlashda omonim, sinonim, atoqli otlar, iboralarga oid lug'atlar
11. Qaysi lug'at turida iyerarxik va ekvivalent munosabatlar aks etadi.
 - a) tezaurus
 - b) toksonomiya
 - c) chastotali lug'at
 - d) ensiklopedik lug'at
12. WordNetda qaysi so'z turkumiga tegishli so'zlar lug'atda aks etmagan?
 - a) olmosh
 - b) ot
 - c) fe'l
 - d) ravish
13. WordNetda sinonimiyada so'zlarning o'rtasidagi asosiy farqi qanday belgi bilan berilgan?
 - a) {}
 - b) *
 - c) []
 - d) ()
14. Lug'atlarda mavjud birliklar haqidagi maxsus joyga ID raqami bilan o'zaro bog'liq ma'lumotlarni birlashtirish nima deb ataladi?
 - a) Voris tizimi
 - b) To'rli bog'
 - c) koordinatsiya
 - d) Kalit siri
15. "Yevgeniy Onegin" she'riy romanida qo'llanilgan rus tilidagi so'zlarning statistikasi kim tomonidan o'rganilgan ?

D., and Miller, G. A. (1991). WordNet: A lexical database organized on psycholinguistic principles. In U. Zernik (Ed.), Lexical acquisition: Exploiting on-line resources to build a lexicon, 211-232.

20. Berlin, B., Breedlove, D., and Raven, P. H. (1973). General principles of classification and nomenclature in folk biology. American Anthropologist, 75, 214- 42.

21. Matlatipov S., Aripov M., Abdurakhmonova N. Modeling WordNet type Thesaurus for Uzbek language Semantic Dictionary / International Journal of Engineering Research and Allied Sciences (IJERAS) ISSN : 2455-9660, Volume-01, Issue-09, November 2016

Savol va topshiriqlar

1. Lug'atlarning tasniflanishi qaysi jihatlariga ko'ra farqlanadi?
2. Avtomatik referatlash va mashina tarjimasiga mo'ljallangan avtomatik lug'atlar foydalanuvchi uchun yaratilgan avtomatik lug'atlardan qanday jihatlari bilan farqlanadi?
3. Elektron lug'atlarning amaliy ahamiyati nimada?
4. Elektron lug'at turlarini sanang va ularning o'zaro farqini ayting.
5. Tezaurus tarkibidagi konseptlarga misollar keltiring.
6. Ontologiya, asosan, qanday komponentlardan tarkib topadi?
7. Tezaurusning boshqa lug'atlardan farqi va afzalliklarini sanang.
8. WordNetdagi leksemalarning semantik munosabatlari nimalarga asoslangan?
9. Terminografiya sohasining asosiy vazifalarini ayting.
10. Elektron lug'at uchun hosil bo'lgan ma'lumotlar bazasi blok sxemasi orqali namuna sifatida kichikroq hajmda lug'at tuzing. Undagi leksik birliklarning bog'lanish usulini tushuntirib bering.

7-mavzu yuzasidan test

1. Toksonomiya deb nimaga aytiladi?
 - a) boshqariladigan lug'at bo'lib, tarmoqli (shajarasimon) tuzilgan ajdod-avlod (ota-ona – farzand), butun – qism kabi boshqa munosabatlar asosida tuzilgan iyerarhik majmua
 - b) so'zlardagi murakkab munosabatlarni o'z ichiga oluvchi boshqariladigan lug'at turi
 - c) yunon tilidan olingan termin bo'lib, xazina ma'nosini beradi
 - d) tillarni avtomatik qayta ishlashda omonim, sinonim, atoqli otlar, iboralarga oid lug'atlar
11. Qaysi lug'at turida iyerarhik va ekvivalent munosabatlar aks etadi.
 - a) tezaurus
 - b) toksonomiya
 - c) chastotali lug'at
 - d) ensiklopedik lug'at
12. WordNetda qaysi so'z turkumiga tegishli so'zlar lug'atda aks etmagan?
 - a) olmosh
 - b) ot
 - c) fe'l
 - d) ravish
13. WordNetda sinonimiyada so'zlarning o'rtasidagi asosiy farqi qanday belgi bilan berilgan?
 - a) {}
 - b) *
 - c) []
 - d) ()
14. Lug'atlarda mavjud birliklar haqidagi maxsus joyga ID raqami bilan o'zaro bog'liq ma'lumotlarni biriktirish nima deb ataladi?
 - a) Voris tizimi
 - b) To'rli bog'
 - c) koordinatsiya
 - d) Kalit siri
15. "Yevgeniy Onegin" she'riy romanida qo'llanilgan rus tilidagi so'zlarning statistikasi kim tomonidan o'rganilgan ?

- a) Markov
- b) Melchuk
- c) Marchuk
- d) Ayimbetov

16. Quyidagilardan qaysi biri ontologiyaning boyitish texnologiyalariga kirmaydi?

- a) OWL
- b) Lois
- c) KNOTE
- d) Web Onto

17. Elektron lug'atlarning afzalliklari tog'ri berilgan qatorni toping?

- a) talaffuzning mavjudligi, tarjima qilish, ma'lumotni o'tkazish imkoniyati borligi, o'quv dasturlar borligi
- b) o'quv dasturlar borligi, tarjima qilish
- c) ovozli, skaner yordamida tarjima qiladi
- d) kam harajat, yuqori tezlik, talaffuzning mavjudligi

18. 1933-yilda kim tomonidan Esperanto tili asosida tillar o'rtasidagi grammatik boshqaruv usuli bilan taqsimlash metodi va bilingvo lug'atni o'z ichiga oluvchi tizim yaratilgan?

- a) P.P.Smironov Troyanskiy
- b) A.Vaxer
- c) Ch.Bebbidj
- d) J.Astrouni

19. Ontologiya qanday tarkiblardan iborat?

- a) atributlar
- b) sinf va tushunchalar
- c) aksiomalar va namumanlar
- d) teorema va belgilar

8-MAVZU. KOMPYUTER LINGVODIDAKTIKASI

8.1. Kompyuter texnologiyalari yordamida dars berishning didaktik konsepsiyasi

8.2. Til ta'limida AKTdan foydalanish

Tayanch so'z va iboralar: *e-o'rganish (e-learning), MALL (mobile assisted language learning)-mobil yordamida tilga o'qitish, ikkinchi til sifatida o'qitish, simsiz aloqa (wireless communication), test dasturi, MOODLE, MOOC, aralash ta'lim (blending learning), CALL (kompyuter yordamida tilga o'qitish), ta'limiy dastur, elektron nusxa, virtual kutubxona, test dasturi, elektron ishlanmalar konsepsiyasi, elektron multimedia darsligi, korpusga asoslangan ta'lim texnologiyasi*

8.1. Kompyuter texnologiyalari yordamida dars berishning didaktik konsepsiyasi

Jahon kompyuter lingvodidaktikasi kompyuter lingvistikasining tarkibida o'rganilmaydi, biroq til o'qitishda kompyuter texnologiyalardan foydalanish, lug'at yaratish borasida ayrim tahliliy masalalar uchun zarur hisoblanadi. Andragogiya – katta yoshli kishilar ta'limi bo'lib, ushbu ta'lim jarayoni uchun kompyuter lingvodidaktikasi muhim ahamiyatga ega.

Kompyuter lingvodidaktikasi – til bilan bog'liq bo'lgan o'qitish va o'rganish jarayonini avtomatlashtirishga yo'naltirilgan soha. Ushbu sohani rivojlantirishda kompyuter texnologiyalarning imkoniyatlaridan to'liq foydalaniladi.

Bugungi kunda o'qitish jarayonini kompyuterlashtirish konsepsiyasini ishlab chiqishda o'quv faoliyatining samaradorligini oshiradigan, o'rganilayotgan mavzuga nisbatan ijobiy munosabat va qiziqishni shakllantiradigan ta'limiy-elektron ishlanmalarni yaratishning umumiy va xususiy mezonlari yaratilmoqda. Hozirgi paytda Rossiya, Ukraina, Fransiya, Yaponiya, Kanada va boshqa mamlakatlarda ta'limiy-elektron ishlanmalarni yaratishga doir ilmiy-metodik tadqiqotlar olib borilmoqda. Konsepsiyada ishlanmalarning mohiyati, ularni yaratish, tekshirish va qo'llash tartibi belgilangan. 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.

Onlayn kurslarni o'qitish uchun MOODLE platformasi imkoniyatlaridan to'la foydalanish zamonaviy pedagogikaning asosiy shartlaridan biridir. Elektron o'qitish modelining pedagogik va onlayn tizimning samarador usullardan biri o'qituvchi fan imkoniyatlariga tayanib, AKTdan faol foydalanadi. Unda talabaning mustaqil holda ishlashiga ko'proq e'tibor qaratiladi. Onlayn kurslar aralash (Blendid learning) holda bo'lishi mumkin. Unda talaba bilan yuzma-yuz mashg'ulot olib borish bilan birga darslarning ayrim qismlari (amaliy, laboratoriya, loyiha, keys, forum, muhokama, muammoli vaziyatlar) virtual tarzda tashkillashtiriladi. MOODLE (massive open online course) platformasida o'qituvchi va talaba o'rtasida izchil ketma-ketlikka asoslangan tizim yaratiladi.

Onlayn kursni samarali tashkil qilishda quyidagilarni bilish talab qilinadi¹⁴⁵: har bir o'qituvchi e-o'qishning mohiyatini yaxshi anglashi, e-o'qishning turli modellarining farqlarini aniqlay olishi, o'qituvchi kompetenti va e-o'qitishning roli haqida ma'lumotga ega bo'lishi, virtual ta'limda talabalar sifati va rolini aniqlay olishi kerak.

Onlayn o'qitishning ommaviylashyotgani bu AKTga asoslangan (telefon, internet: ijtimoiy tarmoq, e-pochta, sotsial interaktiv aloqalar) o'qitish va o'rganishning davomiylik va axborot yoki bilimga asoslangan jamiyat konsepsiyasi yangi paradigma sifatida kundan kun rivojlanib borayotganligi bilan asoslash mumkin. E-o'qitish an'anaviy o'quv xonalaridan tashqari o'quvchining vaqtiga mos bo'lgan qulay muhit uchun moslashtiriladi. Hamkorlikda ishlash, individual rivojlanish va interaktiv aloqani rivojlantirishga yordam beradi. Shunday bo'lsa-da, izolyativ dars o'tish kishilarda psixologik jihatdan individualizm, ruhiy tushkunlik kayfiyatini uyg'otishi mumkin.

1.1. E-O'QISH MODELLARI

E-o'qish muayyan maqsadlarga ko'ra turli usullar orqali amalga oshiriladi. E-o'qishning quyidagi modellari mavjud:

- Sinxron disrutitiv kurs
- Vebga asoslangan kurs
- Gibrid (Blendid) sinfxona kurslari
- Gibrid (Blendid) onlayn kurs
- Onlayn kurs

Sinxron distributiv kurs – veb-konferensiya yoki boshqa sinxron e-o'qish vositasida o'quv auditoriyasidan tashqarida darsda bo'lmagan hollarda qoldirilgan auditoriyada o'qitilgan dars mashg'ulotlarini o'zlashtirish maqsadida mustaqil ravishda tashkillashtiriladi. Talaba bilan ushbu texnologiya yordamida instruktor ko'magida masofadan mashg'ulot olib boriladi.

Vebga asoslangan kurs – mashg'ulotning umumiy hajmi an'anaviy sinf faoliyati bilan bog'liq bo'lgan internetga asoslanadi.

Gibrid (Blendid) sinf mashg'uloti – bunda 50% dars mashg'uloti sinfda, qolgan 50% darsdan tashqarida ta'lim va interaktiv aloqa uchun auditoriya soatini kamaytirib texnologiyadan foydalangan holda mashg'ulot olib boriladi.

Gibrid (Blendid) onlayn kurs – darsning ko'proq qismi (ma'ruza, muhokama, laboratoriya, loyiha) onlayn tashkillashtiriladi, qolgan qismi esa jadvalga solingan o'quv soati bo'yicha amalga oshiriladi.

Onlayn kurs – bu kurs universitet va talabalar o'rtasida geografik aloqani bog'lashga xizmat qiladi. Bunda uchta asosiy interaktiv munosabat: dars mazmuni, instruktor va talabalardan iborat bo'ladi.

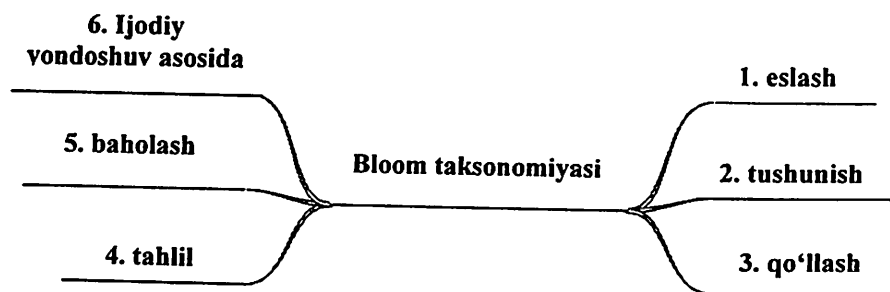
Ushbu modul bo'yicha o'qituvchidan pedagogik ko'nikma, sohaviy bilim, texnologik tajriba kabi kompetentlar talab qilinadi.

E-o'qish jarayonida o'qituvchi ishtirok etmaydi, degan fikrga to'liq qo'shilib bo'lmaydi. O'qituvchi ma'ruzchi bo'lmasligi mumkin, lekin u instruktor, vositachi, repetitor, kuzatuvchi, rahbar vazifasini o'taydi.

Axborot kommunikatsion texnologiyalar yordamida darajali baholash bo'yicha Bloom teksonomiyasidan foydalaniladi.¹⁴⁶

¹⁴⁵ www.moedle.udc.es/pluginfile.php/1102692/mod_resource

¹⁴⁶ www.cmu.edu/teaching/assessment/basics/alignment.html



MOODLE platformasining qulay jihati shundaki, undan foydalanuvchi internet orqali bepul amalga oshiriladigan ochiq resurslardan foydalanishlari mumkin:

- Exelearining – turli logolor yasash
- Genial.ly – animatsiya, taqdimot, interaktiv tasvir, infograflar hosil qilish
- Symbaloo- muayyan maqsad uchun sohaga oid linklarni ilova qiluvchi tool
- Screen-o-matic – video fayllar joylashtirish
- Canva – ma'lumotlarni shakllantirish
- Big Blue Button – veb konferensiya tizimi

Didaktik maqsadga yo'naltirilgan MALL (mobile assisted language learning) (mobil qurilma yordamida til o'rganish (MYTO')) va CALL (computer assisted language learning) (kompyuter yordamida til o'rganish (KYTO')) o'qitish texnologiyalari yaratilgan. Mobil yordamidagi til o'qitish texnologiyasi bo'lib, unda sinfxona va kompyuter oldida uzoq vaqt o'tirish yoki band bo'lish shart bo'lmaydi. Til o'rganish jarayoni joy va sharoitga moslashish xususiyatiga ko'ra har tomonlama pedagogik sohada ijobiy samara bermoqda.

1973-yilda yaratilgan mobil telefonlar hayotimizning ajralmas qismiga aylandi, chunki u nafaqat aloqa vositasi, balki til bilan bog'liq bo'lgan boshqa ko'p vazifalarni bajara oladigan bilimlar bazasi hamdir. Bugungi shiddat bilan rivojlanib borayotgan axborot asrida kishilarning bo'sh vaqt topishi va muayyan maqsadga erishish uchun alohida sharoit yaratishning imkoniyati tobora cheklanib bermoqda.

Horijiy tilni o'rganish va uning lug'ati va grammatikasini o'zlashtirishda tinglash, o'qish, tushunish va gapirish kompetensiyasi muhim jihatlar sanaladi. Ushbu holatda mobil ilovalar har tomonlama qulay hisoblanadi. Unga ko'ra darslarni audio va video holatda eshitish va ko'rish, shuningdek, kitob yoki mashg'ulot uchun yaratilgan ilovalar tushunmagan yoki bilmagan mavzu yuzasidan o'quvchini o'z ustida qayta ishlashi, erishgan natijalarini bilan tanishib, tahlil qilishi, takrorlashi, bilimni mustahkamlash imkoniyatini yaratadi. Biroq mobil ilovalar uchun mo'ljallangan dasturlarning salbiy jihatlarini ham qayd qilib o'tish maqsadga muvofiq. Shunga qaramay kishilarning bo'sh vaqtdan unumli foydalangan holda mobil ilovalardan erkin foydalanishi mumkin. Klopfer va uning hamkasblari mobil qurilmalarning quyidagi xususiyatlarini nazarda tutishadi¹⁴⁷: 1) mobillik (istalgan joyda foydalanish imkoniyati); 2) ijtimoiy interaktivlik (boshqa o'rganuvchilar bilan erkin ma'lumot almashinuvi); 3) sharoitga moslashuvchanlik (joy va vaqt jihatdan ma'lumotlar muayyan joyda saqlanadi); 4) bog'liqlik (mobil telefon boshqa qurilma bilan bog'lana oladi); 5) individuallik (platformadagi imkoniyatdan foydalanuvchi ixtiyoriy ravishda foydalanadi va moslashadi).

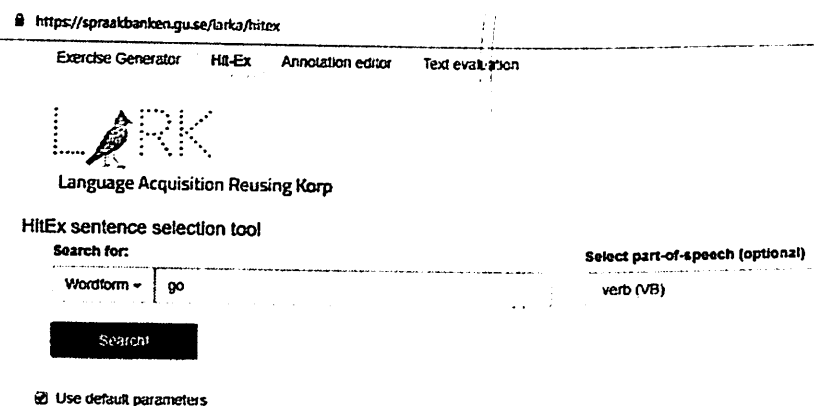
Mobil telefonlarning qulayligi tilni o'rganayotganda lug'at (vocabulary) bilan ishlashga imkoniyat beriladi. Shunda tayyorlangan kichik hajmli lug'atga yo'naltirilgan darslar foydalanuvchiga so'z boyligini oshirish bilan birga bilimni mustahkamlashga xizmat qiladi. Mobil telefon orqali tinglash va grammatikani o'zlashtirish uchun yaxshi imkoniyat yaratiladi.

¹⁴⁷ Klopfer E., Squire K., Jenkins H. Environmental Detectives: PDAs as a window into a virtual simulated world. //in proceedings of IEEE international workshop on wireless and mobile technologies in education. Vaxjo, Sweden: IEEE Computer society. 2002, - P.95-98

8.2. Til ta'limida AKTdan foydalanish

CALL- (kompyuter yordamida til o'rganish – KYTO') tilni xorijiy til sifatida o'qitish va o'rganish imkoniyatini kengaytirishda kompyuter dasturiy vositalaridan foydalanish nazarda tutiladi. KYTO' dasturlari tilni o'qish, tinglash, gapirish va yozish ko'nikmasini shakllantirishga hizmat qiladi. Warschauer¹⁴⁸ KYTO'da kompyuterning quyidagi uch funksiyasini ko'rsatib o'tadi: kompyuter tyutor, kompyuter stimulyator, kompyuter instrument sifatida. Kompyuterning instrument sifatida qo'llash matnlarni avtomatik tahrirlash, elektron lug'atlar va mashina tarjimai kabi tizimlari orqali amalga oshiriladi. Shu bois kompyuter lingvistikasida tilga o'qitish va uni o'rganish lingvistik instrument sifatida muhim vazifalardan sanaladi.

Tilni xorijiy til sifatida va ona tili sifatida o'qitishning turli dasturiy ishlanmalari yaratilgan. *Lärka* intellektual kompyuter yordamidagi o'rganish platformasi bo'lib, u shved tili xorijiy va ikkinchi til bo'lgan talabalar uchun mo'ljallangan. Ushbu platforma



¹⁴⁸ Warschauer M. Computer-assisted language learning: An introduction. In S. photos. Multimedia language teaching / Logos international, 1996.

CLARIN loyihasi doirasida amalga oshirilgan, webga asoslangan ochiq platformadir. Unda leksik resurslar va til texnologiyasi instrumentari uchun mo'ljallangan¹⁴⁹.

Ushbu sohada Yevropa tillari nisbatan ilgarilab ketgan. Masalan, ispan tilini internet tizimi orqali o'rganish uchun o'nlab platformalar va yuzga yaqin dasturiy ilovalar yaratilgan. Ispan tili dunyo bo'yicha o'rganiladigan tillar orasida ikkinchi o'rinda turadi. Ispan tili talaffuzi sodda bo'lganligi, grammatikasi qat'iy formal modelga solinganligi bois o'rganuvchilar soni ortib bormoqda. Ispan tili uchun quyidagi onlayn kurslar yaratilgan:

Proespanol.ru – boshlang'ich bilimlardan boshlab, til haqidagi qiziq ma'lumotlar, turli ko'ngilochar o'yinlar, krassvord va metodik qo'llanmalardan foydalanish mumkin.

Hispanistas.ru — ispan tiliga oid kitoblar, audio, video, forumlar hamda ko'plab kerakli ma'lumotlar berilgan.

Lingvaclub.ru – asosiy leksika va grammatikani boshlang'ich asosini o'qitiladi.

27espanol.com – A1 darajasidagi mashqlar beriladi.

Love-spain.ru – ispan tili jadvallar orqali tushuntirib beriladi.

Spanishgrammarguide.com — sodda tarzda tuzilgan onlayn o'quv darsligi.

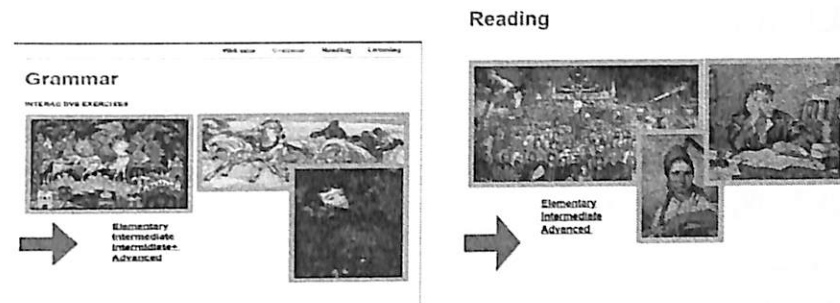
Poliglotty.blogspot.ru – ispan tilidagi 500 ta eng zarur so'zlar kiritilgan; *Hispanista.ru* – leksika, grammatika va onlayn kurslar kiritilgan; *bbc.co.uk* – ushbu sayt orqali BBC ma'lumotlarini ispani tilida olish va boshqa yangiliklarni o'qish mumkin; *Studyspanish.com* – ispan tili maktablari haqida talaffuzi, grammatikasi va lug'ati kiritilgan; *Espato.ru* – ovozlashtirilgan matnlarni tahrirlovchi dasturdan foydalanish va turli mavzulardagi mashqlarni bajarish mumkin; *Donquijote.org* – haftada bir martalik darajani tekshiruvchi dars; *Online-spanish-course.com* – keng mavzudagi lug'at va mashqlar; *Babbel.com* – ushbu sayt ko'ngilochar o'yin va topshriqlar orqali ispan tiliga o'qitiladi; *Diplomas.cervantes.es* – ushbu sayt ispan tili bo'yicha

¹⁴⁹ Qarang: Lärka: From Language Learning Platform to Infrastructure for Research on Language Learning Lärka: From Language Learning Platform to Infrastructure for Research on Language Learning // Selected papers from the CLARIN Annual Conference 2018 Pisa, 8-10 October 2018, -P. 1.

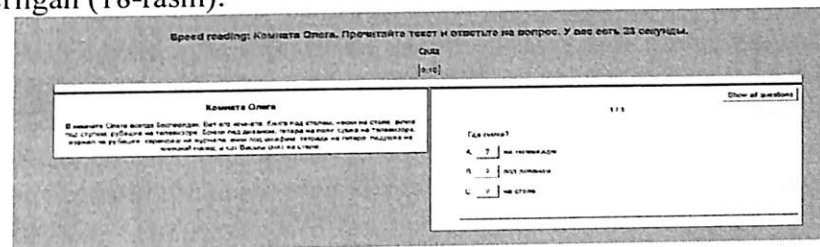
topshiriladigan xalqaro DELE imtihoniga tayyorgarlik ko'rishga mo'ljallangan; *Hispanistas.ru* – kitoblar, audio, video va forumlar joylashgan; *Rae.es* – ispan tilining izohli lug'ati; *Lingus.tv* – turli darajadagi o'rganuvchilar uchun video darslar. Bunda, asosan, diologlar joylashtirilgan; *Recursosdidacticos.es* – ovozlashtirilgan (turli janrdagi pyesa, hikoya, she'riyat va tarix, madaniyat va an'analar haqidagi asarlar) matnlar to'plami; *Spanishgrammarguide.com* – sodda tilda yozilgan onlayn o'quv adabiyoti; *Dirae.es* – qiziqarli lug'at; *Cvc.cervantes.es* – uch darajaga ajratilgan Servantes institutidan olingan matnlar, shuningdek, ijtimoiy tarmoqlar orqali ispan tilini o'rgatishga yo'naltirilgan tarmoqlar mavjud: *Instagram.com* – rasmdagi ispancha so'zlarni o'rganish hamda YouTube, facebook, vk.com, Play.google.com, iTunes.apple.com – so'zlashgich va boshq.

Rus tili bo'yicha ham bepul internet serverlaridan foydalanib turli dasturiy ilovalar yaratilgan: *Learnrussian.rt* – rus tili haqidagi yuzga yaqin darslar, o'quvchining bilimini tekshirish uchun mavzulashirilgan test, to'g'ri talaffuz qilish uchun alifbo tartibidagi materiallar grammatik jadvallar va darslarda o'z ifodasini topgan. *Rus-on-line* - A1 - A2 darajadagi rus tili o'rganuvchilar uchun yaratilgan portal. *Rustest-online* – turli mavzudagi grammatik mashqlar A2-B1 darajada internet qo'llanma. Bunda asosiy e'tibor fe'l so'z turkumiga qaratilgan.

Huddi shunday bir qator internet resurslarini namuna sifatida ko'rsatish mumkin: *Speak-russian.cie/time_new*, *Ru.islcollective*, *Texts.cie* (til trenajori), *Gcserussian*, *Mezhdunami*, *Russian.ucla*, *Vk/russianforeign*, *Russianforeveryone*. Masalan, <http://www.russian.ucla.edu/flagship/russianflagship/Grammar.html> sayti grammatika, o'qish va tinglab tushunish kabi tilni kompetensiyali o'qitishning uch yo'nalishiga asoslangan onlayn test yaratilgan 17-, 18-, 19-rasmlar:

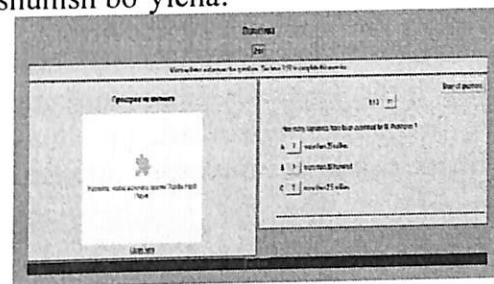


O'qish ko'nikmasini shakllantirish uchun quyidagicha toshiriq berilgan (18-rasm):



18-rasm

Tinglab tushunish bo'yicha:



19-rasm

Listening

1. Наука В Английском языке и другие вопросы по лингвистике	10 Наука Цветы
2. Технологии: Миссия будущего	11 Культура: Возраст
3. Экономика: Кризис в России	12 Культура: Карнавал в Венеции
4. Русское телевидение: Андрей Малахов	13 Путешествие
5. Журналистика: Выступление В.В. Познера в МГУ	14 Лекции по разным дисциплинам
6. Политика: Прохоров на митинге	Участь TV.ru
7. Здравоохранение: Судья обманывает	
8. Спорт: Татьяна Тарасова	
9. Журналистика: Александр Васильев	

20-rasm

Shuningdek, korpusga asoslangan ta'lim texnologiyasi tobora rivojlanib bormoqda. Ayniqsa, korpusdagi lug'at, konkordans hamda turli matndan olingan namunalar til o'qitishda qulay va samarali usul sifatida foydalanib kelinmoqda. Ingliz tilining og'zaki va yozma Longman grammatikasi korpus asosida yaratilgan¹⁵⁰ (1999). Korpusga asoslangan Chemnitz Internet ingliz tili grammatikasidan (<http://www.tu-chemnitz.de/phil/InternetGrammar>) internet tarmog'i orqali foydalanish mumkin. Korpus orqali foydalanuvchilar mustaqil ravishda matndagi biror leksemaning konkordanslari orqali so'zlarning lug'aviy ma'nosi, sinonimlari, grammatik shakllanishi hamda so'z qo'llash uslubini mustaqil ravishda o'rganishlari mumkin.

Adabiyotlar:

1. Anne Rimrott Spell checking in computer-assisted language learning: a study of misspelling by nonnative writers of German. Justus Liebig university, 2002 (M.A. master diploma work), -P.158.

2. Joybrato Mukherjee Corpus linguistics and language pedagogy: The state of the art – and beyond

http://www.unigiessen.de/faculties/f05/engl/ling/staff/professors/jmukherjee/publications/pdfs/Mukherjee2006_pedagogy.pdf

¹⁵⁰ Joybrato Mukherjee Corpus linguistics and language pedagogy: The state of the art – and beyond / http://www.uni-giessen.de/faculties/f05/engl/ling/staff/professors/jmukherjee/publications/pdfs/Mukherjee2006_pedagogy.pdf

3. http://repository.upenn.edu/ircs_reports

Savol va topshiriqlar:

- Elektron ta'limning afzalliklari va kamchiliklarini sanang.
- Onlayn kursni samarali tashkil qilishda qo'yiladigan talablar qaysilar?
- E-o'qishning modellarini aniqlang.
- Blendid (Gibrid) sinf mashg'uloti va Blendid (Gibrid) onlayn kurslari orasidagi farq va o'xshashliklarini toping.
- Sinxron distributiv kurs qanday imkoniyatlarga ega?
- Onlayn kurs moduli bo'yicha o'qituvchiga kerak bo'ladigan komponentlarni izohlang.
- E-o'qishdan foydalanishda o'qituvchining o'rni haqida fikr yuriting.
- Mobil ilovalar uchun mo'ljallangan dasturlarning ijobiy va salbiy jihatlari haqida fikr yuriting.
- E-o'qishning yutug'i va kamchiliklarini jadvalda ifodalagan holda izohlang.
- Ispan tili uchun yaratilgan onlayn kurslar va ularning imkoniyatlari haqida mulohaza yuriting.
- Ushbu jadvalga fikringizni yozing.

E-o'qish	yutug'i	E-o'qitish	yutug'i
	Kamchiligi		kamchiligi

8-mavzu yuzasidan test

1. Katta yoshli kishilarning ta'limi qaysi termin bilan nomlanadi?
 - a. andragagogiya
 - b. akmeologiya
 - c. valeologiya
 - d. eduologiya
2. Mashg'ulotning umumiy hajmi an'anaviy sinf faoliyati bilan bog'liq bo'lgan internetga asoslangan kurs qanday nomlanadi?
 - a. webga asoslangan kurs
 - b. onlayn kurs
 - c. aralash kurs
 - d. an'anaviy kurs
2. Blendid (Gibrid) sinf mashg'ulotining o'ziga xos xarakterini aniqlang
 - a) 50% dars mashg'uloti sinfda, qolgan 50% darsdan tashqarida ta'lim va interaktiv aloqa uchun auditoriya soatini kamaytirib texnologiyadan foydalangan holda olib boriladi.
 - b) darsning ko'proq qismi (ma'ruza, muhokama, laboratoriya, loyiha) onlayn tashkillashtiriladi, ozroq qismi esa jadvalga solingan o'quv soati bo'yicha amalga oshiriladi.
 - c) bu kurs institut va talaba o'rtasida geografik aloqani bog'lashga xizmat qiladi.
 - d) Barchasi to'g'ri
3. Onlayn kurs qanday interaktiv munosabatda bo'ladi. Ularni toping.
 - a) mundarija, instruktor va talabalar
 - b) interfeys va foydalanuvchi
 - c) elektron pochta va foydalanuvchi
 - d) o'qituvchi va talaba
4. O'qitish tizimini avtomatlashtirish qanday nomlanadi?
 - a) kompyuter lingvodidaktikasi
 - b) kompyuter leksikografiyasi
 - c) avtomatik tahrir
 - d) kompyuter lingvistikasi
5. E-o'qishning qanday modellari mavjud:
 - a) sinxron, distributiv kurs, vebga asoslangan kurs

- b) blendid (Gibrid) sinfxona kurslari, blendid (Gibrid) onlayn kurs
 - c) onlayn kurs
 - d) barcha javoblar to'g'ri
6. Mobil telefon yordamidagi til o'qitish texnologiyasi nima deb nomlanadi?
 - a) MALL
 - b) CALL
 - c) diskurs
 - d) blendid kurslari
 7. Klopfer va uning hamkasblari mobil qurilmalarning qaysi xususiyatlarini ko'rsatib o'tishadi?
 - a) mobillik (istalgan joyda foydalanishning mumkinligi)
 - b) ijtimoiy interaktivlik (boshqa o'rganuvchilar bilan ma'lumotlar almsinuvi);
 - c) bog'liqlik (mobil telefon boshqa uskuna bilan bog'lana oladi);
 - d) barcha javoblar to'g'ri
 8. Qaysi darslik ovozashtirilgan (turli janrdagi pyesa, hikoya, she'riyat va tarix, madaniyat va an'analar haqidagi asarlar) matnlar to'plamidan iborat?
 - a) Recursosdidacticos.es
 - b) Spanishgrammarguide.com
 - c) Cvc.cervantes.es
 - d) Hispanistas.ru
 9. Quyidagilardan qaysi biri sodda tilda yozilgan onlayn o'quv adabiyoti hisoblanadi?
 - a) Spanishgrammarguide.com
 - b) Cvc.cervantes.es
 - c) Hispanistas.ru
 - d) Recursosdidacticos.es

9-MAVZU. MASHINA TARJIMASI RIVOJLANISH TENDENSIYASI

9.1. Mashina tarjimasining shakllanishi va taraqqiyot bosqichlari

9.2. Sohada olib borilayotgan tadqiqotlar

9.3. Mashina tarjimasini texnologiyasi va tizimlari

Tayanch soʻz va iboralar: mashina tarjimasini, lingvistik modellashtirish nazariyasi, Ampar, Aspera, PROMT, toʻgʻridan - toʻgʻri tarjima qilish, namunaga asoslangan (example-based), xotiraga asoslangan (memory-based), analogiyaga asoslangan (analogy-based), keysga asoslangan (casebased) mashina tarjimasini, apertium platformasi, elektron lugʻatlar yordamida tarjima, muharrir yordamidagi mashina tarjimasini, dasturiy ilovalar yordamidagi tarjima, qoidaga asoslangan MT, transfer tizim, interlingval tizim, lugʻat asosidagi tizim, statistikaga asoslangan MT, gibrid mashina tarjimasini

9.1. Mashina tarjimasining shakllanish va taraqqiyot bosqichlari

XX asrning oʻrtalarida tarjima jarayonini kompyuter bilan bogʻliq holda oʻrganish gʻoyalari paydo boʻlishi mashina (avtomatik) tarjimasini nomini oldi. Mashina tarjimasining yuzaga kelishiga bir qator omillar taʼsir koʻrsatgan¹⁵¹. XVIII asrda faylasuflar G.V. Leybnis va Deskart tomonidan jumla va soʻzlarning oʻzaro bogʻliqligini kodlash masalasi ilgari surilgan. Mashinaning tarjima qila olish imkoniyati haqidagi gʻoyani Ch.Bebbidj (1791-1871) 1836-1848-yillarda oʻsha davrdan 100-yil keyin paydo boʻlgan elektron raqamli mashinalarning mexanik prototipi - raqamli analitik mashinalar boʻyicha olib borgan loyihasida qayd etgan. XX asrning 30-yillarida "tarjimon mashinalar" amalda qoʻllanilgan boʻlsa, Estoniyada mexanik tarjimini roʻyobga chiqarishda A.Vaxerning nazariy qarashlari "Vaba Maa" nomli roʻznomada eʼlon qilindi (1924). Fransiyada J. Astrouni tomonidan avtomatik bilingval lugʻatlardan foydalanish taklif etilib, u «Mexanik miya»

¹⁵¹ Qarang: Абдурахмонова Н. Машина таржимасининг лингвистик асослари. – Т.: Akademnashr, 2012. –B.7-11.

deb nomlangan mashina tarjimasini ishlanmasini uchun patentga ega boʻldi. 1933-yilda Rossiyada lingvistik arifmometr muallifi P.P.Smironov Troyanskiy tomonidan esperanto tili asosida tillar oʻrtasidagi grammatik boshqaruv usuli bilan taqsimlash metodi va bilingv lugʻatni oʻz ichiga oluvchi tizim yaratildi. Bu tizim uch bosqichni oʻz ichiga olgan: *birinchi bosqichda* muharrir asosiy tildagi soʻzlarni tartibga solib, birliklarning mantiqiy shakllari hamda sintaktik funksiyalarini belgilaydi; *ikkinchi bosqichda* mashina leksik shakllarni tarjima qilinayotgan tilga oʻgiradi; *uchinchi bosqichda* muharrir natijani meʼyorlashtiradi. Biroq olimning bu yondashuvi kompyuterlar ommalashayotgan vaqt – XX asrning 50-yillarigacha nomaʼlumligicha qolaverdi. Mashina tarjimasini oʻtgan asrning 40-yillari oxirida amaliy fan yoʻnalishi sifatida yuzaga keldi. Uning nazariy asosida tilga kodlashtirilgan sistema sifatida qarash yotadi. Mashina tarjimasining dastlabki ilgʻor vakillari matematik va muhandislar tomonidan 40-yillarning oxirida AQShda IBMning yaratilishi uchun kriptografik vazifalarni bajarishga moʻljallangan kodlashtirish nazariyasi asos boʻldi. U.Uiver gʻoyasining zamirida interlingval (Interlingua) konsepsiya mavjud boʻlib, unda ikki bosqichni amalga oshirish koʻzda tutilgan: birinchi bosqichda asosiy matn vositachi tilga oʻgiriladi, ikkinchi bosqichda uning natijasi keyingi tilga tarjima qilinadi. A.But va R.Richans tomonidan mashina tarjimasini boʻyicha dastlabki tajribalar oʻtkazildi (1948) va moliyalashtirildi. AQShda ruscha-inglizcha, Sobiq ittifoqda inglizcha-ruscha matnlarni tarjima qilish diqqat markazda boʻldi. Ingliz matematigi A.Tyuringning ong ustida olib borgan tekshirishlarida mashina fikrlay olishi, yaʼni mashinaning inson kabi muloqotga kirisha olish imkoniyati mavjudligini fanga taqdim qilinishi mashina tarjimasini tarixida katta oʻzgarish yasadi (1950). Mashina tarjimasini boʻyicha ilk ilmiy-amaliy konferensiya Massachutes texnika universitetida oʻtkazilib (1952), Nyu-Yorkdagi Jorjtaun universiteti bilan IBM kompaniyasi hamkorligida birinchi IBM II rus tilidan ingliz tiliga tarjima qilishga moʻljallangan mashina tarjimasini dasturi yaratildi (1954). Bu tadqiqot L.Dorster rahbarligida oʻtkazilib, kimyo sohasi doirasida 250 ta leksik birlik hamda 6 ta grammatik qoidalar bilan chegaralangan 60 ta gap tarjima qilingan. ("Dorster tajribasi" yoki "Jorjtaun sinovi"). Ushbu

9-MAVZU. MASHINA TARJIMASI RIVOJLANISH TENDENSIYASI

9.1. Mashina tarjimasining shakllanishi va taraqqiyot bosqichlari

9.2. Sohada olib borilayotgan tadqiqotlar

9.3. Mashina tarjimasini texnologiyasi va tizimlari

Tayanch soʻz va iboralar: *mashina tarjimasini, lingvistik modellashtirish nazariyasi, Ampar, Aspera, PROMT, toʻgʻridan - toʻgʻri tarjima qilish, namunaga asoslangan (example-based), xotiraga asoslangan (memory-based), analogiyaga asoslangan (analogy-based), keysga asoslangan (casebased) mashina tarjimasini, apertium platformasi, elektron lugʻatlar yordamida tarjima, muharrir yordamidagi mashina tarjimasini, dasturiy ilovalar yordamidagi tarjima, qoidaga asoslangan MT, transfer tizim, interlingval tizim, lugʻat asosidagi tizim, statistikaga asoslangan MT, gibrid mashina tarjimasini*

9.1. Mashina tarjimasining shakllanish va taraqqiyot bosqichlari

XX asrning oʻrtalarida tarjima jarayonini kompyuter bilan bogʻliq holda oʻrganish gʻoyalari paydo boʻlishi mashina (avtomatik) tarjimasini nomini oldi. Mashina tarjimasining yuzaga kelishiga bir qator omillar taʼsir koʻrsatgan¹⁵¹. XVIII asrda faylasuflar G.V. Leybnis va Deskart tomonidan jumla va soʻzlarning oʻzaro bogʻliqligini kodlash masalasi ilgari surilgan. Mashinaning tarjima qila olish imkoniyati haqidagi gʻoyani Ch.Bebbidj (1791-1871) 1836-1848-yillarda oʻsha davrdan 100-yil keyin paydo boʻlgan elektron raqamli mashinalarning mexanik prototipi - raqamli analitik mashinalar boʻyicha olib borgan loyihasida qayd etgan. XX asrning 30-yillarida "tarjimon mashinalar" amalda qoʻllanilgan boʻlsa, Estoniyada mexanik tarjimini roʻyobga chiqarishda A.Vaxerning nazariy qarashlari "Vaba Maa" nomli roʻznomada eʼlon qilindi (1924). Fransiyada J. Astrouni tomonidan avtomatik bilingval lugʻatlardan foydalanish taklif etilib, u «Mexanik miya»

deb nomlangan mashina tarjimasini ishlanmasini uchun patentga ega boʻldi. 1933-yilda Rossiyada lingvistik arifmometr muallifi P.P.Smironov Troyanskiy tomonidan esperanto tili asosida tillar oʻrtasidagi grammatik boshqaruv usuli bilan taqsimlash metodi va bilingval lugʻatni oʻz ichiga oluvchi tizim yaratildi. Bu tizim uch bosqichni oʻz ichiga olgan: *birinchi bosqichda* muharrir asosiy tildagi soʻzlarni tartibga solib, birliklarning mantiqiy shakllari hamda sintaktik funksiyalarini belgilaydi; *ikkinchi bosqichda* mashina leksik shakllarni tarjima qilinyotgan tilga oʻgiradi; *uchinchi bosqichda* muharrir natijani meʼyorlashtiradi. Biroq olimning bu yondashuvi kompyuterlar ommalashayotgan vaqt – XX asrning 50-yillarigacha nomaʼlumligicha qolaverdi. Mashina tarjimasini oʻtgan asrning 40-yillari oxirida amaliy fan yoʻnalishi sifatida yuzaga keldi. Uning nazariy asosida tilga kodlashtirilgan sistema sifatida qarash yotadi. Mashina tarjimasining dastlabki ilgʻor vakillari matematik va muhandislar tomonidan 40-yillarning oxirida AQShda IBMning yaratilishi uchun kriptografik vazifalarni bajarishga moʻljallangan kodlashtirish nazariyasi asos boʻldi. U.Uiver gʻoyasining zamirida interlingval (Interlingua) konsepsiya mavjud boʻlib, unda ikki bosqichni amalga oshirish koʻzda tutilgan: birinchi bosqichda asosiy matn vositachi tilga oʻgiriladi, ikkinchi bosqichda uning natijasi keyingi tilga tarjima qilinadi. A.But va R.Richans tomonidan mashina tarjimasini boʻyicha dastlabki tajribalar oʻtkazildi (1948) va moliyalashtirildi. AQShda ruscha-inglizcha, Sobiq ittifoqda inglizcha-ruscha matnlarni tarjima qilish diqqat markazda boʻldi. Ingliz matematigi A.Tyuringning ong ustida olib borgan tekshirishlarida mashina fikrlay olishi, yaʼni mashinaning inson kabi muloqotga kirisha olish imkoniyati mavjudligini fanga taqdim qilinishi mashina tarjimasini tarixida katta oʻzgarish yasadi (1950). Mashina tarjimasini boʻyicha ilk ilmiy-amaliy konferensiya Massachutes texnika universitetida oʻtkazilib (1952), Nyu-Yorkdagi Jorjtaun universiteti bilan IBM kompaniyasi hamkorligida birinchi IBM II rus tilidan ingliz tiliga tarjima qilishga moʻljallangan mashina tarjimasini dasturi yaratildi (1954). Bu tadqiqot L.Dorster rahbarligida oʻtkazilib, kimyo sohasi doirasida 250 ta leksik birlik hamda 6 ta grammatik qoidalar bilan chegaralangan 60 ta gap tarjima qilingan. ("Dorster tajribasi" yoki "Jorjtaun sinovi"). Ushbu

¹⁵¹ Qarang: Абдурахмонова Н. Машина таржимасининг лингвистик асослари. – Т.: Akademnashr, 2012. –B.7-11.

tajribada uch turdagi lug'at mavjud bo'lgan: 1) so'z o'zamlari lug'ati; 2) so'z oxiridagi morfemalar lug'ati; 3) tarjima lug'at.

Turkiy tillar bo'yicha dastlabki mashina tarjimasi Dorster boshchiligida amalga oshirilgan (1961). Ingliz-turk tili mashina tarjimasi tizimida 700 ta so'zshakldan iborat bo'lib, lug'at turk tili hamda ingliz tilidagi ekvivalentliklarining asos va suffikslari ro'yxatidan iborat bo'lgan¹⁵².

Shunga o'xshash Sobiq ittifoqda mashina tarjimasi bo'yicha ilk bor tajriba I.S. Muxin, L.N. Korolev hamda S.N. Razumovskiy boshchiligida algoritmi I.K.Belskiy tomonidan ishlab chiqilgan inglizcha-ruscha mashina tarjimasi sinovdan o'tkazilgan (1955). Shu yillarda Moskvada A.Lyapunov rahbarligidagi matematik va kibernetiklardan tashkil topgan guruh ko'p aspektli lingvistik ma'lumotlarni o'z ichiga olgan inglizcha-ruscha avtomatik tarjima tizimini (ARAP), Ittifoq Fanlar Akademiyasida Y.A.Motorin rahbarligidagi guruh inglizcha-ruscha mashina tarjimasi dasturini, D.Y.Panov esa uning lingvistik ta'minotini yaratib berdi (1955). 1956-yilda Moskvada Y. A. Motorin rahbarligida mashina tarjimasi laboratoriyasi tashkil etildi. Laboratoriyada faoliyat yuritgan yosh olim Y.N. Marchuk 1958 -yilda siyosiy va ilmiy-texnikaviy sohalarda doirasidagi matnlarni ingliz tilidan rus tiliga tarjima qilib beruvchi iqtisodiy jihatdan eng samarali va qulay bo'lgan birinchi mashina tarjimasi tizimini ishlab chiqdi. Keyinchalik olim "*Mashina tarjimasi muammolari*" nomli monografiyasini e'lon qildi¹⁵³. Uning mazkur tadqiqotida tarjima modelashtirish tamoyillari hamda ilmiy-texnikaviy ma'lumotlarga doir eng dolzarb muammolarini hal etishda qo'llaniladigan tahlil usullari o'rin olgan. Tillararo tarjima moslashuvdagi ikki komponentli tarjima modelida tilning statik tavsifi hamda dinamik tarjima jarayoni aks etishi va uning sifatini oshirishda tahrirlash vazifasini inson resursi orqali foydalanish zaruriyati borligi ilmiy asoslab berilgan. Y.N. Marchuk mashina tarjimasiga tavsiflangan tarjima modellarini tushuntiruvchi kuch hamda til mohiyatini darhol tekshirish imkoniyatini ko'rsatuvchi hamda taklif etilgan ijodiy mulohazalarning samaradorligini

¹⁵² M.G.Mamedova, Z.Yu.Mamedova *Машинный перевод: эволюция и основные аспекты моделирования*. – Баку, 2005. – С. 34.

¹⁵³ Qarang: Марчук Ю.Н. *Проблемы машинного перевода*. – Москва, 1983.

oshiruvchi vosita, deb qaraydi. Shundan so'ng mashina tarjimasi uchun *lingvistik modelashtirish nazariyasi* hamda tabiiy tilning ma'lumotlar bazasini yaratish masalasi yana bir qator asarlarda o'z aksini topgan¹⁵⁴. Boshqa olimlar esa mashina tarjimasi deklarativ bilimlardan iborat, ya'ni lingvist til faktlari tavsifini bazaga kiritganidan so'ng matematik va dasturchi tomonidan tarjima algoritmi tuziladi¹⁵⁵, degan fikrga to'xtaladilar.

O'z davrida mashina imkoniyatlarining chegaralanganligi, dasturlash tilining yuqori darajada ishlanmaganligi, lingvistik muammolarni hal qiluvchi nazariy ta'minotning yo'qligi sabab mashina tarjimasining rivojlanish davri quyi nuqtaga yetdi (1950-yillar oxiri). Dastlabki mashina tarjimasi muayyan sintaktik butunliksiz matnni so'zma-so'z (word-to-word) tarjima qilar edi. Faylasuf Y.Bar-Xillel bunday yo'l bilan mashina tarjimasidagi muammolarni hal etib bo'lmashligini ta'kidladi (1956). Olim ingliz tilidagi omonimiya masalasini *pen – 1) ruchka; 2) manej* leksemasi orqali tushuntirishga harakat qildi. Buning negizida mashina tarjimasini inkor etmagan holda u "*inson-tarjimon*"dan foydalanish zaruriyati borligini qayd etdi. Maxsus Milliy Akademiyada ALPAC (Automatic Language Processing Advisory Committee - Tilni avtomatik qayta ishlash bo'yicha maslahat komiteti) komissiyasi tuzildi (1966). B.Xillelning "Language and machines: Computers in translation in linguistics" (Til va mashinalar: lingvistika va tarjimada kompyuterlar) nomli ma'ruzasidan so'ng texnikaga oid va ilmiy matnlar uchun inson resursidan foydalanish zarur, degan xulosasiga kelindi va AQSh hukumati mashina tarjimasiga oid tadqiqotlarni uning tezligi va sifati past, harajat yuqori bo'lganligi uchun moliyalashtirishni qisqartirib yuboradi.

Mashina tarjimasi nazariyasi kibernetik, tilshunos, faylasuf, matematik-dasturchi va psixolog kabi qator soha vakillarining diqqat markazida bo'lgan. Shu o'rinda mashina tarjimasi uchun asos bo'luvchi metatil, ya'ni matematik modelashtirilgan formal til haqidagi turli munozarali fikrlar o'rtaga tashlangan. Xususan,

¹⁵⁴ Qarang: Пиотровский Р.Г., Билан В.Н., Боркун М.Н., Бобков А.К. *Методы автоматического анализа и синтеза текста*. – Минск, 1985.

¹⁵⁵ Теория машинного перевода <http://rudocs.exdat.com/docs/index-7253.html>. page=2#358615

tajribada uch turdagi lug‘at mavjud bo‘lgan: 1) so‘z o‘zaklari lug‘ati; 2) so‘z oxiridagi morfemalar lug‘ati; 3) tarjima lug‘at.

Turkiy tillar bo‘yicha dastlabki mashina tarjimasi Dorster boshchiligida amalga oshirilgan (1961). Ingliz-turk tili mashina tarjimasi tizimida 700 ta so‘zshakldan iborat bo‘lib, lug‘at turk tili hamda ingliz tilidagi ekvivalentliklarining asos va suffikslari ro‘yxatidan iborat bo‘lgan¹⁵².

Shunga o‘xshash Sobiq ittifoqda mashina tarjimasi bo‘yicha ilk bor tajriba I.S. Muxin, L.N. Korolev hamda S.N. Razumovskiy boshchiligida algoritmi I.K.Belskiy tomonidan ishlab chiqilgan inglizcha-ruscha mashina tarjimasi sinovdan o‘tkazilgan (1955). Shu yillarda Moskvada A.Lyapunov rahbarligidagi matematik va kibernetiklardan tashkil topgan guruh ko‘p aspektli lingvistik ma‘lumotlarni o‘z ichiga olgan inglizcha-ruscha avtomatik tarjima tizimini (ARAP), Ittifoq Fanlar Akademiyasida Y.A.Motorin rahbarligidagi guruh inglizcha-ruscha mashina tarjimasi dasturini, D.Y.Panov esa uning lingvistik ta‘minotini yaratib berdi (1955). 1956-yilda Moskvada Y. A. Motorin rahbarligida mashina tarjimasi laboratoriyasi tashkil etildi. Laboratoriyada faoliyat yuritgan yosh olim Y.N. Marchuk 1958 -yilda siyosiy va ilmiy-texnikaviy sohalar doirasidagi matnlarni ingliz tilidan rus tiliga tarjima qilib beruvchi iqtisodiy jihatdan eng samarali va qulay bo‘lgan birinchi mashina tarjimasi tizimini ishlab chiqdi. Keyinchalik olim “*Mashina tarjimasi muammolari*” nomli monografiyasini e‘lon qildi¹⁵³. Uning mazkur tadqiqotida tarjimani modellashtirish tamoyillari hamda ilmiy-texnikaviy ma‘lumotlarga doir eng dolzarb muammolarini hal etishda qo‘llaniladigan tahlil usullari o‘rin olgan. Tillararo tarjima moslashuvdagi ikki komponentli tarjima modelida tilning statik tavsifi hamda dinamik tarjima jarayoni aks etishi va uning sifatini oshirishda tahrirlash vazifasini inson resursi orqali foydalanish zaruriyati borligi ilmiy asoslab berilgan. Y.N. Marchuk mashina tarjimasiga tavsiflangan tarjima modellarini tushuntiruvchi kuch hamda til mohiyatini darhol tekshirish imkoniyatini ko‘rsatuvchi hamda taklif etilgan ijodiy mulohazalarning samaradorligini

¹⁵² M.G.Мамедова, З.Ю.Мамедова Машинный перевод: эволюция и основной аспект моделирования. – Баку, 2005. – С. 34.

¹⁵³ Qarang: Марчук Ю.Н. Проблемы машинного перевода. – Москва, 1983.

oshiruvchi vosita, deb qaraydi. Shundan so‘ng mashina tarjimasi uchun *lingvistik modellashtirish nazariyasi* hamda tabiiy tilning ma‘lumotlar bazasini yaratish masalasi yana bir qator asarlarda o‘z aksini topgan¹⁵⁴. Boshqa olimlar esa mashina tarjimasi deklarativ bilimlardan iborat, ya‘ni lingvist til faktlari tavsifini bazaga kiritganidan so‘ng matematik va dasturchi tomonidan tarjima algoritmi tuziladi¹⁵⁵, degan fikrga to‘xtaladilar.

O‘z davrida mashina imkoniyatlarining chegaralanganligi, dasturlash tilining yuqori darajada ishlanmaganligi, lingvistik muammolarni hal qiluvchi nazariy ta‘minotning yo‘qligi sabab mashina tarjimasining rivojlanish davri quyi nuqtaga yetdi (1950-yillar oxiri). Dastlabki mashina tarjimasi muayyan sintaktik butunliksiz matnni so‘zma-so‘z (word-to-word) tarjima qilar edi. Faylasuf Y.Bar-Xillel bunday yo‘l bilan mashina tarjimasidagi muammolarni hal etib bo‘lmasligini ta‘kidladi (1956). Olim ingliz tilidagi omonimiya masalasini *pen – 1) ruchka; 2) manej* leksemasi orqali tushuntirishga harakat qildi. Buning negizida mashina tarjimasini inkor etmagan holda u “*inson-tarjimon*”dan foydalanish zaruriyati borligini qayd etdi. Maxsus Milliy Akademiyada ALPAC (Automatic Language Processing Advisory Committy - Tilni avtomatik qayta ishlash bo‘yicha maslahat komiteti) komissiyasi tuzildi (1966). B.Xillelning “*Language and machines: Computers in translation in linguistics*” (Til va mashinalar: lingvistika va tarjimada kompyuterlar) nomli ma‘ruzasidan so‘ng texnikaga oid va ilmiy matnlar uchun inson resursidan foydalanish zarur, degan xulosasiga kelindi va AQSh hukumati mashina tarjimasiga oid tadqiqotlarni uning tezligi va sifati past, harajat yuqori bo‘lganligi uchun moliyalashtirishni qisqartirib yuboradi.

Mashina tarjimasi nazariyasi kibernetik, tilshunos, faylasuf, matematik-dasturchi va psixolog kabi qator soha vakillarining diqqat markazida bo‘lgan. Shu o‘rinda mashina tarjimasi uchun asos bo‘luvchi metatil, ya‘ni matematik modellashtirilgan formal til haqidagi turli munozarali fikrlar o‘rtaga tashlangan. Xususan,

¹⁵⁴ Qarang: Пиотровский Р.Г., Билан В.Н., Боркун М.Н., Бобков А.К. Методы автоматического анализа и синтеза текста. – Минск, 1985.

¹⁵⁵ Теория машинного перевода <http://rudocs.exdat.com/docs/index-7253.html>. page=2#358615

“...metatil bu tilga oid tushunchalarning modeli (I.A. Melchuk), inson psixikasida tushunchalarni yozib oluvchi va saqlovchi gipotetik til hamda semantikaning predmeti (Yu. Apresyan) kabi qarashlar bir necha asarlarda o‘z ifodasini topdi¹⁵⁶. Biroq mashina tarjimasini uchun asos bo‘luvchi tabiiy tilni hech qanday qolipga solib bo‘lmasligi va uni tizim sifatida o‘rganish shartli ekanligi to‘g‘risida N.Z. Kotelova shunday mulohazalarni bildirgan: “Til bu shunday obyekt, uni yaratilgan hech qaysi sun‘iy tizimga solish mumkin emas. Shuning uchun leksik-semantik xususiyatlar aks etgan sun‘iy tilni yaratish masalasi o‘zining utopik va mavhumligini ko‘rsatadi”¹⁵⁷.

Leningraddagi ilmiy tadqiqot institutida matematik hamda mexaniklardan tarkib topgan mashina tarjimasini uchun eksperimental laboratoriya tashkil etildi (1960). Bunda G.G. Belonogov va R.G. Kotov boshchiligidagi ilmiy jamoa faoliyatini alohida ta’kidlash joiz. Mashina tarjimasiga bag‘ishlangan «Mechanical Translation» (1954), «T.A. Informations» (1965; 1960-64-yildan boshlab «Tradition automatic» nomi bilan chiqa boshladi), «Communications of the Association for Computing machinery» (1958), «Научно-техническая информация. Серия 2» (1961) nomli jurnallar, «Машинный перевод и прикладная лингвистика» (1959), «Проблемы кибернетики» (1958) kabi to‘plamlar¹⁵⁸ nashr etila boshlandi. O‘z davrida bildirilgan munosabatlarning turiligi mashina tarjimasining nazariy asoslarini shakllantirdi. N.D. Andreev, I. Bar-Xillel, I.A. Melchuk, I.I. Revzin, V.Y. Rozenveyg va boshqa olimlarning mashina tarjimasining “qog‘oz” algoritmi tamoyili hamda Yu.N. Marchuk, R.G. Piotrovskiy, Y.A. Motorin rahbarligi ostida tuzilgan ilmiy guruhning hamda K.B. Bektaev, A.N. Belyaev, I.K. Belskiy, A.V. Zubov, G.E. Miram, L.L. Nelyubin, V.I. Perebynos, V.A. Chijakovskiy, Y.A.

Shingarev kabi olimlarning monografiyalari fanda muhim o‘zgarishlarga omil bo‘ldi.

Keyingi davrlarda Rossiyada G.G. Belonogov (1975), R.G. Kotov, N.V. Babushkina (1998), Z. Shalyapina, O. Yu. Mansurova (2006), A.S. Panina (2002), A.A. Xoroshilov (2007), V.A. Chijakovskiy kabilarning mashina tarjimasini bo‘yicha bajarilgan doktorlik va nomzodlik ishlari e’tiborga molik. Mashina tarjimasining nazariy ildizlari nafaqat Rossiya, balki ko‘plab xorijiy davlatlarga ham borib taqaladi. AQShda A. But, R. Richans (1948), J. Xatchins, J. Allen, P. Braun; Yaponiyada M. Nagao; Estoniyada A. Vaxer (1924); Fransiyada J. Astrouni; Hindistonda R. Sinha, A. Jain; Germaniyada B. Bleyzer, U. Shvol, A. Storrer kabi olimlarning nazariy qarashlari bugungi kunda mashina tarjimasiga oid turli tizimlarning yuzaga kelishida muhim manba bo‘ldi, desak yanglishmagan bo‘lamiz. Mashina tarjimasini bo‘yicha tadqiqotlarning yangi bosqichi 70-yillarda boshlangan. Bunga muayyan masalalarning kompyuterda modellashtirilganligi sababli “sun‘iy intellekt” nomini oldi. Shuningdek, bu davrda to‘plam nazariyasi bilan ishlovchi *TM* (translation memory–tarjima xotirasi) vujudga keldi. Bunga ko‘ra asos segment (jumla) va matn tarjimasini saqlanib, lingvistik ma’lumotlar bazasini tashkil qiladi. Shu yillarda Sobiq SSSRda ingliz tilidan rus tiliga tarjima qiluvchi AMPAR, ASPERA; nemis tilidan rus tiliga tarjima qiluvchi NERPA; fransuz tilida rus tiliga tarjima qiluvchi FRAP, RANO (ispan-ingliz), SYSTRAN, TAUM (ingliz-fransuz), POLA (xitoy-ingliz), MIAN (fransuz-rus), CULT (xitoy-ingliz) shuningdek, STYLUS, SOCRAT kabi tarjima dasturlari yaratildi. ETAP protsessorining yaratilishida I.A. Melchuk hamda Y.D. Apresyan kabi lingvistlar tomonidan ETAP mashina tarjimasini yaratildi.

Mashina tarjimasining rivojlanishiga ikki omil ta’sir ko‘rsatgan: 1) tarjimaning kompyuterda modellashtirilganligi; 2) mashina tarjimasining ijtimoiy maqsadda tillararo kommunikatsiyani ta’minlashga bo‘lgan imkoniyatning kengligi.

Kanadada Meteo (franzus-ingliz) mashina tarjimasini tizimi yaratiladi va o‘sha davrda u 97% aniqlikda ishlay oladi. Janet va Jim Beyker tomonidan Dragon SYSTEM firmasi (Massachusetts shtatidagi Nyuton shahri, 1982-yilda asos solingan)

¹⁵⁶ Qarang: Апресян Ю.Д. О языке для описания значений слов. – Москва: Известия АН СССР. Серия языка и литературы, 1966. – № 5. – С. 415-429; Мельчук И.А. Опыт теории лингвистических моделей “смысл-текст”. Семантика. Синтаксис. – М.: Наука, 1974. – 314 с.

¹⁵⁷ Котелова Н.Э. Значение олова и его сочетаемость. – Л.: Наука, 1975. – С. 28.

¹⁵⁸ Qarang: <http://bse.sci-lib.com/article074634.html>

“...metatil bu tilga oid tushunchalarning modeli (I.A. Melchuk), inson psixikasida tushunchalarni yozib oluvchi va saqlovchi gipotetik til hamda semantikaning predmeti (Yu.Apresyan) kabi qarashlar bir necha asarlarda o‘z ifodasini topdi¹⁵⁶. Biroq mashina tarjimasini uchun asos bo‘luvchi tabiiy tilni hech qanday qolipga solib bo‘lmasligi va uni tizim sifatida o‘rganish shartli ekanligi to‘g‘risida N.Z.Kotelova shunday mulohazalarni bildirgan: “Til bu shunday obyektiki, uni yaratilgan hech qaysi sun‘iy tizimga solish mumkin emas. Shuning uchun leksik-semantik xususiyatlar aks etgan sun‘iy tilni yaratish masalasi o‘zining utopik va mavhumligini ko‘rsatadi”¹⁵⁷.

Leningraddagi ilmiy tadqiqot institutida matematik hamda mexaniklardan tarkib topgan mashina tarjimasini uchun eksperimental laboratoriya tashkil etildi (1960). Bunda G.G.Belonogov va R.G.Kotov boshchiligidagi ilmiy jamoa faoliyatini alohida ta‘kidlash joiz. Mashina tarjimasiga bag‘ishlangan «Mechanical Translation» (1954), «T.A. Informations» (1965; 1960-64-yildan boshlab «Tradition automatic» nomi bilan chiqa boshladi), «Communications of the Association for Computing machinery» (1958), «Научно-техническая информация. Серия 2» (1961) nomli jurnallar, «Машинный перевод и прикладная лингвистика» (1959), «Проблемы кибернетики» (1958) kabi to‘plamlar¹⁵⁸ nashr etila boshlandi. O‘z davrida bildirilgan munosabatlarning turiligi mashina tarjimasining nazariy asoslarini shakllantirdi. N.D. Andreev, I. Bar-Xillel, I.A. Melchuk, I.I. Revzin, V.Y. Rozenveyg va boshqa olimlarning mashina tarjimasining “qog‘oz” algoritmi tamoyili hamda Yu.N. Marchuk, R.G. Piotrovskiy, Y.A. Motorin rahbarligi ostida tuzilgan ilmiy guruhning hamda K.B. Bektaev, A.N. Belyaev, I.K. Belskiy, A.V. Zubov, G.E. Miram, L.L. Nelyubin, V.I. Perebinyos, V.A. Chijakovskiy, Y.A.

¹⁵⁶ Qarang: Апресян Ю.Д. О языке для описания значений слов. – Москва: Известия АН СССР. Серия языка и литературы, 1966. – № 5. – С. 415-429; Мельчук И.А. Опыт теории лингвистических моделей “смысл-текст”. Семантика. Синтаксис. – М.: Наука, 1974. – 314 с.

¹⁵⁷ Котелова Н.Э. Значение олова и его сочетаемость. –Л.: Наука, 1975. – С. 28.

¹⁵⁸ Qarang: <http://bse.sci-lib.com/article074634.html>

Shingarev kabi olimlarning monografiyalari fanda muhim o‘zgarishlarga omil bo‘ldi.

Keyingi davrlarda Rossiyada G.G.Belonogov (1975), R.G.Kotov, N.V. Babushkina (1998), Z.Shalyapina, O.Yu. Mansurova (2006), A.S. Panina (2002), A.A.Xoroshilov (2007), V.A. Chijakovskiy kabilarning mashina tarjimasini bo‘yicha bajarilgan doktorlik va nomzodlik ishlari e‘tiborga molik. Mashina tarjimasining nazariy ildizlari nafaqat Rossiya, balki ko‘plab xorijiy davlatlarga ham borib taqaladi. AQShda A.But, R. Richans (1948), J. Xatchins, J.Allen, P. Braun; Yaponiyada M. Nagao; Estoniyada A.Vaxer (1924); Fransiyada J. Astrouni; Hindistonda R.Sinha, A.Jain; Germaniyada B.Bleyzer, U.Shvol, A.Storrer kabi olimlarning nazariy qarashlari bugungi kunda mashina tarjimasiga oid turli tizimlarning yuzaga kelishida muhim manba bo‘ldi, desak yangilishmagan bo‘lamiz. Mashina tarjimasini bo‘yicha tadqiqotlarning yangi bosqichi 70-yillarda boshlangan. Bunga muayyan masalalarning kompyuterda modellashtirilganligi sababli “sun‘iy intellekt” nomini oldi. Shuningdek, bu davrda to‘plam nazariyasi bilan ishlovchi *TM* (translation memory–tarjima xotirasi) vujudga keldi. Bunga ko‘ra asos segment (jumla) va matn tarjimasini saqlanib, lingvistik ma‘lumotlar bazasini tashkil qiladi. Shu yillarda Sobiq SSSRda ingliz tilidan rus tiliga tarjima qiluvchi AMPAR, ASPERA; nemis tilidan rus tiliga tarjima qiluvchi NERPA; fransuz tilida rus tiliga tarjima qiluvchi FRAP, RANO (ispan-ingliz), SYSTRAN, TAUM (ingliz-fransuz), POLA (xitoy-ingliz), MIAN (fransuz-rus), CULT (xitoy-ingliz) shuningdek, STYLUS, SOCRAT kabi tarjima dasturlari yaratildi. ETAP protsessorining yaratilishida I.A.Melchuk hamda Y.D.Apresyan kabi lingvistlar tomonidan ETAP mashina tarjimasini yaratildi.

Mashina tarjimasining rivojlanishiga ikki omil ta‘sir ko‘rsatgan: 1) tarjimaning kompyuterda modellashtirilganligi; 2) mashina tarjimasining ijtimoiy maqsadda tillararo kommunikatsiyani ta‘minlashga bo‘lgan imkoniyatning kengligi.

Kanadada Meteo (franzuz-ingliz) mashina tarjimasini tizimi yaratiladi va o‘sha davrda u 97% aniqlikda ishlay oladi. Janet va Jim Beyker tomonidan Dragon SYSTEM firmasi (Massachusetts shtatidagi Nyuton shahri, 1982-yilda asos solingan)

“...metatil bu tilga oid tushunchalarning modeli (I.A. Melchuk), inson psixikasida tushunchalarni yozib oluvchi va saqlovchi gipotetik til hamda semantikaning predmeti (Yu. Apresyan) kabi qarashlar bir necha asarlarda o‘z ifodasini topdi¹⁵⁶. Biroq mashina tarjimasini uchun asos bo‘luvchi tabiiy tilni hech qanday qolipga solib bo‘lmasligi va uni tizim sifatida o‘rganish shartli ekanligi to‘g‘risida N.Z. Kotelova shunday mulohazalarni bildirgan: “Til bu shunday obyekt, uni yaratilgan hech qaysi sun‘iy tizimga solish mumkin emas. Shuning uchun leksik-semantik xususiyatlar aks etgan sun‘iy tilni yaratish masalasi o‘zining utopik va mavhumligini ko‘rsatadi”¹⁵⁷.

Leningraddagi ilmiy tadqiqot institutida matematik hamda mexaniklardan tarkib topgan mashina tarjimasini uchun eksperimental laboratoriya tashkil etildi (1960). Bunda G.G. Belonogov va R.G. Kotov boshchiligidagi ilmiy jamoa faoliyatini alohida ta’kidlash joiz. Mashina tarjimasiga bag‘ishlangan «Mechanical Translation» (1954), «T.A. Informations» (1965; 1960-64-yildan boshlab «Tradition automatic» nomi bilan chiqa boshladi), «Communications of the Association for Computing machinery» (1958), «Научно-техническая информация. Серия 2» (1961) nomli jurnallar, «Машинный перевод и прикладная лингвистика» (1959), «Проблемы кибернетики» (1958) kabi to‘plamlar¹⁵⁸ nashr etila boshlandi. O‘z davrida bildirilgan munosabatlarning turliligi mashina tarjimasining nazariy asoslarini shakllantirdi. N.D. Andreev, I. Bar-Xillel, I.A. Melchuk, I.I. Revzin, V.Y. Rozenveyg va boshqa olimlarning mashina tarjimasining “qog‘oz” algoritmi tamoyili hamda Yu.N. Marchuk, R.G. Piotrovskiy, Y.A. Motorin rahbarligi ostida tuzilgan ilmiy guruhning hamda K.B. Bektaev, A.N. Belyaev, I.K. Belskiy, A.V. Zubov, G.E. Miram, L.L. Nelyubin, V.I. Perebinyos, V.A. Chijakovskiy, Y.A.

¹⁵⁶ Qarang: Апресян Ю.Д. О языке для описания значений слов. – Москва: Известия АН СССР. Серия языка и литературы, 1966. – № 5. – С. 415-429; Мельчук И.А. Опыт теории лингвистических моделей “смысл-текст”. Семантика. Синтаксис. – М.: Наука, 1974. – 314 с.

¹⁵⁷ Котелова Н.Э. Значение олова и его сочетаемость. – Л.: Наука, 1975. – С. 28.

¹⁵⁸ Qarang: <http://bse.sci-lib.com/article074634.html>

Shingarev kabi olimlarning monografiyalari fanda muhim o‘zgarishlarga omil bo‘ldi.

Keyingi davrlarda Rossiyada G.G. Belonogov (1975), R.G. Kotov, N.V. Babushkina (1998), Z. Shalyapina, O. Yu. Mansurova (2006), A.S. Panina (2002), A.A. Xoroshilov (2007), V.A. Chijakovskiy kabilarning mashina tarjimasini bo‘yicha bajarilgan doktorlik va nomzodlik ishlari e’tiborga molik. Mashina tarjimasining nazariy ildizlari nafaqat Rossiya, balki ko‘plab xorijiy davlatlarga ham borib taqaladi. AQShda A. But, R. Richans (1948), J. Xatchins, J. Allen, P. Braun; Yaponiyada M. Nagao; Estoniyada A. Vaxer (1924); Fransiyada J. Astrouni; Hindistonda R. Sinha, A. Jain; Germaniyada B. Bleyzer, U. Shvol, A. Storrer kabi olimlarning nazariy qarashlari bugungi kunda mashina tarjimasiga oid turli tizimlarning yuzaga kelishida muhim manba bo‘ldi, desak yangilishmagan bo‘lamiz. Mashina tarjimasini bo‘yicha tadqiqotlarning yangi bosqichi 70-yillarda boshlangan. Bunga muayyan masalalarning kompyuterda modellashtirilganligi sababli “sun‘iy intellekt” nomini oldi. Shuningdek, bu davrda to‘plam nazariyasi bilan ishlovchi *TM* (translation memory–tarjima xotirasi) vujudga keldi. Bunga ko‘ra asos segment (jumla) va matn tarjimasini saqlanib, lingvistik ma’lumotlar bazasini tashkil qiladi. Shu yillarda Sobiq SSSRda ingliz tilidan rus tiliga tarjima qiluvchi AMPAR, ASPERA; nemis tilidan rus tiliga tarjima qiluvchi NERPA; fransuz tilida rus tiliga tarjima qiluvchi FRAP, RANO (ispan-ingliz), SYSTRAN, TAUM (ingliz-fransuz), POLA (xitoy-ingliz), MIAN (fransuz-rus), CULT (xitoy-ingliz) shuningdek, STYLUS, SOCRAT kabi tarjima dasturlari yaratildi. ETAP protsessorining yaratilishida I.A. Melchuk hamda Y.D. Apresyan kabi lingvistlar tomonidan ETAP mashina tarjimasini yaratildi.

Mashina tarjimasining rivojlanishiga ikki omil ta’sir ko‘rsatgan: 1) tarjimaning kompyuterda modellashtirilganligi; 2) mashina tarjimasining ijtimoiy maqsadda tillararo kommunikatsiyani ta’minlashga bo‘lgan imkoniyatning kengligi.

Kanadada Meteo (franzuz-ingliz) mashina tarjimasini tizimi yaratiladi va o‘sha davrda u 97% aniqlikda ishlay oladi. Janet va Jim Beyker tomonidan Dragon SYSTEM firmasi (Massachusetts shtatidagi Nyuton shahri, 1982-yilda asos solingan)

“...metatil bu tilga oid tushunchalarning modeli (I.A. Melchuk), inson psixikasida tushunchalarni yozib oluvchi va saqlovchi gipotetik til hamda semantikaning predmeti (Yu.Apresyan) kabi qarashlar bir necha asarlarda o‘z ifodasini topdi¹⁵⁶. Biroq mashina tarjimasi uchun asos bo‘luvchi tabiiy tilni hech qanday qolipga solib bo‘lmasligi va uni tizim sifatida o‘rganish shartli ekanligi to‘g‘risida N.Z.Kotelova shunday mulohazalarni bildirgan: “Til bu shunday obyektiki, uni yaratilgan hech qaysi sun‘iy tizimga solish mumkin emas. Shuning uchun leksik-semantik xususiyatlar aks etgan sun‘iy tilni yaratish masalasi o‘zining utopik va mavhumligini ko‘rsatadi”¹⁵⁷.

Leningraddagi ilmiy tadqiqot institutida matematik hamda mexaniklardan tarkib topgan mashina tarjimasi uchun eksperimental laboratoriya tashkil etildi (1960). Bunda G.G.Belonogov va R.G.Kotov boshchiligidagi ilmiy jamoa faoliyatini alohida ta’kidlash joiz. Mashina tarjimasiga bag‘ishlangan «Mechanical Translation» (1954), «T.A. Informations» (1965; 1960-64-yildan boshlab «Tradition automatic» nomi bilan chiqa boshladi), «Communications of the Association for Computing machinery» (1958), «Научно-техническая информация. Серия 2» (1961) nomli jurnallar, «Машинный перевод и прикладная лингвистика» (1959), «Проблемы кибернетики» (1958) kabi to‘plamlar¹⁵⁸ nashr etila boshlandi. O‘z davrida bildirilgan munosabatlarning turliligi mashina tarjimasining nazariy asoslarini shakllantirdi. N.D. Andreev, I. Bar-Xillel, I.A. Melchuk, I.I. Revzin, V.Y. Rozenveyg va boshqa olimlarning mashina tarjimasining “qog‘oz” algoritmi tamoyili hamda Yu.N. Marchuk, R.G. Piotrovskiy, Y.A. Motorin rahbarligi ostida tuzilgan ilmiy guruhning hamda K.B. Bektaev, A.N. Belyaev, I.K. Belskiy, A.V. Zubov, G.E. Miram, L.L. Nelyubin, V.I. Perebiynos, V.A. Chijakovskiy, Y.A.

¹⁵⁶ Qarang: Апресян Ю.Д. О языке для описания значений слов. – Москва: Известия АН СССР. Серия языка и литературы, 1966. – № 5. – С. 415-429; Мельчук И.А. Опыт теории лингвистических моделей "смысл-текст". Семантика. Синтаксис. – М.: Наука, 1974. – 314 с.

¹⁵⁷ Котелова Н.Э. Значение олова и его сочетаемость. –Л.: Наука, 1975. – С. 28.

¹⁵⁸ Qarang: <http://bse.sci-lib.com/article074634.html>

Shingarev kabi olimlarning monografiyalari fanda muhim o‘zgarishlarga omil bo‘ldi.

Keyingi davrlarda Rossiyada G.G.Belonogov (1975), R.G.Kotov, N.V. Babushkina (1998), Z.Shalyapina, O.Yu. Mansurova (2006), A.S. Panina (2002), A.A.Xoroshilov (2007), V.A. Chijakovskiy kabilarning mashina tarjimasi bo‘yicha bajarilgan doktorlik va nomzodlik ishlari e’tiborga molik. Mashina tarjimasining nazariy ildizlari nafaqat Rossiya, balki ko‘plab xorijiy davlatlarga ham borib taqaladi. AQShda A.But, R. Richans (1948), J. Xatchins, J.Allen, P. Braun; Yaponiyada M. Nagao; Estoniyada A.Vaxer (1924); Fransiyada J. Astrouni; Hindistonda R.Sinha, A.Jain; Germaniyada B.Bleyzer, U.Shvol, A.Storrer kabi olimlarning nazariy qarashlari bugungi kunda mashina tarjimasiga oid turli tizimlarning yuzaga kelishida muhim manba bo‘ldi, desak yanglishmagan bo‘lamiz. Mashina tarjimasi bo‘yicha tadqiqotlarning yangi bosqichi 70-yillarda boshlangan. Bunga muayyan masalalarning kompyuterda modellashtirilganligi sababli “sun‘iy intellekt” nomini oldi. Shuningdek, bu davrda to‘plam nazariyasi bilan ishlovchi *TM* (translation memory–tarjima xotirasi) vujudga keldi. Bunga ko‘ra asos segment (jumla) va matn tarjimasi saqlanib, lingvistik ma’lumotlar bazasini tashkil qiladi. Shu yillarda Sobiq SSSRda ingliz tilidan rus tiliga tarjima qiluvchi AMPAR, ASPERA; nemis tilidan rus tiliga tarjima qiluvchi NERPA; fransuz tilida rus tiliga tarjima qiluvchi FRAP, RANO (ispan-ingliz), SYSTRAN, TAUM (ingliz-fransuz), POLA (xitoy-ingliz), MIAN (fransuz-rus), CULT (xitoy-ingliz) shuningdek, STYLUS, SOCRAT kabi tarjima dasturlari yaratildi. ETAP protsessorining yaratilishida I.A.Melchuk hamda Y.D.Apresyan kabi lingvistlar tomonidan ETAP mashina tarjimasi yaratildi.

Mashina tarjimasining rivojlanishiga ikki omil ta’sir ko‘rsatgan: 1) tarjimaning kompyuterda modellashtirilganligi; 2) mashina tarjimasining ijtimoiy maqsadda tillararo kommunikatsiyani ta’minlashga bo‘lgan imkoniyatning kengligi.

Kanadada Meteo (franzus-ingliz) mashina tarjimasi tizimi yaratiladi va o‘sha davrda u 97% aniqlikda ishlay oladi. Janet va Jim Beyker tomonidan Dragon SYSTEM firmasi (Massachusetts shtatidagi Nyuton shahri, 1982-yilda asos solingan)

mikrokompyuterlar uchun nutqni qayta ishlaydigan ALPC deb nomlangan (1983) hamda DARPA nomli nutqni tushunuvchi dasturiy ta'minoti yaratildi (1985). Yaponiyada ham interlingval texnologiyaga asoslangan yuqori tezlikdagi mashina tarjimasini yaratildi¹⁵⁹.

O'tgan asrning 90-yillariga kelib mashina tarjimasida sezilarli o'zgarishlar yuzaga chiqdi. Bunga yangi texnologiyalarning kirib kelishi, statistik metodning qo'llanilishi, konneksionizm konsepsiyasining ilgari surilishi hamda neyron setlardan foydalanish texnologiyasi ishlab chiqilganligini asos qilib ko'rsatish mumkin. Moskvadagi PC Forum shaxsiy kompyuterlarining ko'rgazmasida PROMT (Programmer's machine translation) deb nomlanuvchi o'zining o'ta aniqligi, keng soha va tillar miqdorining ko'pligi bilan tarjimasini bozorining 95 foizini egallagan tijorat mashina tarjimasini dasturi taqdim etildi (1990) va NASA tanlovida muvaffaqiyatga erishdi (1992). Keyinchalik PROMT ingliz, nemis, fransuz, italyan, va ispan tillaridan rus tiliga hamda rus tilidan ingliz tiliga tarjima qilib beruvchi yangi STYLUS nomi ostida chiqqan boshladi, shuningdek, uning bazasi asosida Windows OS uchun mashina tarjimasini tizimini yaratildi (1993). Germaniyalik olimlar Verbmobil nomli loyihasida rasmiy uslubdagi inglizcha matnlarni nemis va yapon tillariga tarjima qilishning asosini ishlab chiqishdi. Windows 3.X/95/NT uchun STYLUS 2.0 versiyasi (1994) va Windows 95/NT uchun STYLUS 3.0 deb nomlangan uchinchi avlodi ruscha-nemischa hamda ruscha-fransuzcha mashina tarjimasini tizimiga yo'naltirilgan (1995-1996-yillar). Shaxsiy kompyuterlar uchun mukammal lug'atlarga asoslangan dastlabki nutqni tanib oladigan Naturally Speaking (Dragon Systems) va Via Voice (IBM) dasturlari yuzaga kelishi 1999-yilda efirga uzatilgan dasturlarni 85% aniqlikda tarjima qildi. IBM kompaniyasi Makintosh uchun *Via Voice* tizimini ishlab chiqdi. Axborot tadqiqotlari institutidan (ISI) Kevin Nayt Janubiy-Kaliforniya universiteti qoshida statistik mashina tarjimasini tizimiga qurilgan EGIPT dasturlarini ishlab chiqish uchun universitetlararo tuzilgan ilmiy guruhga rahbarlik qildi. 2000-yili Yang-Suk Li hamda Klifford Veynsteyn tomonidan Linkoln

Laboratoriyasida (Massachusetts Texnika Universiteti) og'zaki nutqni koreys tilidan ingliz tiliga tarjima qiluvchi tizim prototipi namoyish qilindi¹⁶⁰. Keyingi o'n yillikda multilingval internet tarmoqlari uchun *Quero* nomli fransuz-nemis loyihasi ishlab chiqildi. Bu loyiha nafaqat matnlarni, balki internetdagi video va audio fayllarni ham tarjima qilib beruvchi tezkor qidiruv tizimidir¹⁶¹. 2007-yilda Googlening dastlabki tarjima uskunalari amalda qo'llanila boshlaydi. 2004-yilda Translatica (polyak tili uchun www.translatica.pl) bilingval mashina tarjimasini yaratiladi.

Yuqoridagi tadqiqotlardan kelib chiqib shuni aytish o'rinliki, har qanday fan tarixini davrlashtirishda nazariy qarashlarning burilish bosqichi hisobga olinadi. Jamiyatda sodir bo'layotgan har qanday yangilik insonning tashqi olamga munosabatida o'z ifodasini topadi. Mashina tarjimasiga oid taraqqiyot bosqichlarini ilovada ko'rsatilgan manbaga ko'ra kuzatishimiz mumkin: 1) 1940-yillarning oxiri – 60-yillarning boshi: Mashina tarjimasining to'g'ridan - to'g'ri tarjima qilish (so'zma-so'z) tizimi birinchi avlodining yaratilishi; 2) 1960-yillarning o'rtasi – 70-yillar: sintaktik nazariyalarning intensiv rivojlanishi hamda ular asosida mashina tarjimasini tizimini qayta ishlovchi ikkinchi avlodning paydo bo'lishi; 3) 1970 -yillarning o'rtasi – 80-yillar: mashina tarjimasining ekstensiv taraqqiyoti. Unga ko'ra morfologik va sintaktik tahlilga asoslangan, biroq tarjima semantik strukturalar asosida amalga oshirilmagan, shuningdek, inson vositasida tor sohaviy doirada matnlarning tarjima qilinishi. (Bu davrda uchinchi avlod yaratilmagan.); 4) 80-yillarning ikkinchi yarmi: bilimlar bazasidan foydalangan holda ko'p tilli tizimlarning vujudga kelishi, shuningdek, turli semantik nazariyalarning paydo bo'lishi¹⁶².

Kompyuter dasturiy ta'minotlarning takomillashtirilishi 1990-2000-yillarda mashina tarjimasini tizimlarini ham mukammal variantlarining yaratilishiga omil bo'ldi. Keyingi davrlarda MTning

¹⁶⁰ O'sha asar. – B. 27-33.

¹⁶¹ en.wikipedia.org/wiki

¹⁶² Система машинного перевода «Кросслятор 2.0» и анализ ее функциональности для задачи трансляции знаний http://www.keldysh.ru/papers/2007/prep89/prep2007_89.html

¹⁵⁹ O'sha asar. – B. 25.

korpusga asoslangan, neyroto‘r va sun‘iy intellekt texnologiyasiga asoslangan tizimlari yaratildi.

O‘tgan asrning 90-yillarda bu yo‘nalish turli nomlar bilan atala boshlandi: namunaga asoslangan (example-based), xotiraga asoslangan (memory-based), analogiyaga asoslangan (analogy-based), keysga asoslangan (casebased) mashina tarjimasini bo‘lib, bu oldindan qilingan tarjima namunalar bazasi yoki korpusga asoslanadi¹⁶³. Hozirda jahondagi o‘nlab kompaniyalar tijorat dasturlarini qayta ishlash bilan shug‘ullanmoqdalar: Systren («System of translation» birikmasining qisqartmasi), IBM, L&H (Lernout & Houspie), Language Engineering Corporation, Transparent Language, Nova Incorporated, Trident Software, Atril, TRADOS, Caterpillar Co., Lingvo Ware, Ata Software, Lingvistica, Pragma shular jumlasidandir. Shunday dasturlar sirasida ruscha matnlarni ingliz tiliga tarjima qiluvchi Retrans Vista versiyasini ko‘rsatish o‘rinli. Undagi lug‘atda nafaqat millionlab frazeologik iboralarning ma‘nolari, shuningdek, kundalik so‘zlashuvda qo‘llanuvchi so‘z birikmalari ham berilgan. Bundan tashqari dasturda lug‘atda aks etgan so‘z birikmalarini matndan avtomatik tarzda ajratib oluvchi konseptual tahlil ham mavjud. Retrans Vista tizimining lug‘atini tabiiy fanlar, texnika, iqtisodiyot, biznes, siyosatga oid frazeologik birliklar va terminlar tashkil etadi. Mashina lug‘atida taxminan 3,4 mln. leksema bo‘lib, ularning 1,8 mln. rus-ingliz, 1,6 mln. ingliz-rus leksemalardan tashkil topgan. Lug‘atning 20 % leksemalardan, 80 % o‘rtacha “uzunligi” 2,2 so‘zdan tashkil topgan turg‘un so‘z birikmalaridan iborat. Bu dastur OS Windows 98/NT/2000 boshqaruvi ostida ishlaydi. Endilikda, internet tarmog‘i orqali ham avtomatik tarjima uchun maxsus saytlardan foydalanish mumkin: www.alphaworks.com/hm/aw.nsf/html/mt, www.freetranslation.com, www.t-mail.com, www.translate.ru, www.foreignword.com/Tools/translate.dyn va h.k. Shuningdek, ba‘zi tarjimon dasturlari internet tizimida ishlashga mo‘ljallangan¹⁶⁴:

<u>Alis Technologies Gist-In-Time</u>	ingliz tili, fransuz tili, nemis tili, italyan tili, portugal tili, ispan tili, yapon tili
<u>babelfish.altavista.com</u>	ingliz tili, fransuz tili, nemis tili, italyan tili, portugal tili, ispan tili
<u>Ectaco</u>	alban tili, arab tili, bolgar tili, chex tili, fransuz tili, nemis tili, grek tili, vengir tili, italyan tili, koreys tili, latish tili, fors tili, polyak tili, portugal tili, roman tili, rus tili, serb tili, slovak tili, ispan tili, shved tili, turk tili, yahudiy tili
<u>e-lingo</u>	ingliz tili, fransuz tili, nemis tili, italyan tili, portugal tili, ispan tili
<u>freetranslation.com</u>	ingliz tili, fransuz tili, nemis tili, italyan tili, ispan tili, portugal tili, norveg tili
<u>GPLTrans</u>	golland tili, ingliz tili, fransuz tili, indoneziya tili, nemis tili, italyan tili, lotin tili, portugal tili, ispan tili
<u>IBM alphaWorks and IBM developerWorks</u>	ingliz tili, fransuz tili, nemis tili, italyan tili, ispan tili, yapon tili, xitoy tili
<u>InterTran</u>	bulgor tili, xorvat tili, chex tili, daniya tili, golland tili, fillipin tili, fransuz tili, fin tili, nemis tili, grek tili, venger tili, island tili, italyan tili, yapon tili, norveg tili, polyak tili, portugal tili, roman tili, rus tili, serb tili, slovak tili, slovyan tili, ispan tili, shved tili, uels tili
<u>Language Engineering Corporation LogoVista X</u>	ingliz tili, yapon tili
<u>Lingvistica '98</u>	ingliz tili, nemis tili, rus tili, ukrain tili
<u>Poltran.com (Ectaco)</u>	ingliz tili, polyak tili
<u>PROMT</u>	ingliz tili, fransuz tili, nemis tili, ispan tili, rus tili
<u>Reverso</u>	ingliz tili, fransuz tili, nemis tili, ispan tili, rus tili
<u>Russian Translation USA</u>	ingliz tili, rus tili, ukrain tili
<u>Rustran.com (Ectaco)</u>	ingliz tili, rus tili
<u>Systran</u>	ingliz tili, fransuz tili, nemis tili, italyan tili, ispan tili, portugal tili
<u>Transbot</u>	ingliz tili, fransuz tili, nemis tili, ispan tili
<u>WorldLingo</u>	ingliz tili, fransuz tili, nemis tili, ispan tili, golland tili, rus tili, xitoy tili, yapon tili, koreys tili
<u>Universidad de Oviedo.</u>	ingliz tili, fransuz tili, nemis tili, ispan tili

¹⁶³ The Oxford handbook of computational linguistics (edited by Mitkov R.) – Oxford, 2003. – P. 493.

¹⁶⁴ Qarang: <http://dwheelersite.com/PDFs/ENGL%207702%20Project%20Description.pdf>

1986-yil Hatchins ma'lumotiga ko'ra dastlabki Mashina tarjimasi bo'yicha o'tkazilgan konferensiyalar

- 1.1952-yil, 17-20- iyun Kembridjda MIT konferensiya
- 2.1954-yil, 7-yanvar Nyu-Yorkdagi Jorjtoun IBM tizimi namoyishi.
- 3.1955-yil, avgust Angliyaning Kembridj qirollik Kollejidagi konferensiya
- 4.1956-yil, oktyabr Kembridj, Mass Xalqaro konferensiyasi.
- 5.1959-yil, 15-20 -iyun Fransiyaning Parij shahri, UNESCO konferensiyasi kengashi.
- 6.1960-yil, fevral Los Angeles, UCLA Xalqaro simpoziumi.
- 7.1960-yil, iyul N.J Prinston Vayne davlat universiteti.
- 8.1961-yil, Vashington D.C Jorjtoun universiteti, ikkinchi "Prinston yig'ilishi".
- 9.1961-yil, noyabr Angliya Teddington, Milliy fizika laboratoriyasi.
- 10.1962-yil, N.J. Prinston uchinchi "Prinston yig'ilishi".
- 11.1962-yil, iyun Italiya Venetsiya shahri, NATO Rivojlangan yozgi instituti.
- 12.1965-yil, Las Vegas to'rtinchi "Prinston yig'ilishi".

COLING konferensiyasi

- 1.1965-yil, Qo'shma Shtatlar, Nyu-York.
- 2.1967-yil, Fransiya, Grenobl.
- 3.1969-yil, Shvetsiya, Stokholm.
- 4.1971-yil, Vengriya, Debresen.
- 5.1974-yil, Italiya, Pisa.
- 6.1976-yil, Kanada, Ottava.
- 7.1978-yil, Norvegiya, Bergen.
- 8.1982-yil, Chexislovakiya, Praga.
- 9.1984-yil, Qo'shma shtatlar, Kalif, Stanford.
- 10.1986-yil, G'arbiy Germaniya, Bonn.
- 11.1988-yil, Vengriya, Budapesht.
- 12.1990-yil, Findlaniya, Helsinki.
- 13.1992-yil, Fransiya, Nantes.
- 14.1994-yil, Yaponiya, Kyoto.
- 15.1996-yil, Daniya, Kopengagen.

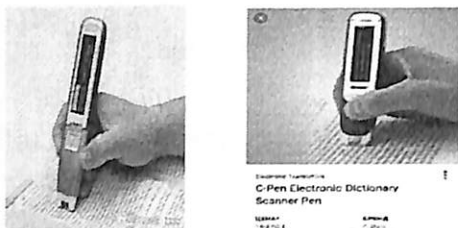
- 16.1998-yil, Kanada, Monreal.
- 17.2000-yil Lyuksemburg
TMI (Tabiiy tillarning mashina tarjimasidagi nazariy va metodologik muammolari)

- 1.1985-yil 14-16-avgust, Qo'shma Shtatlar Nyu-york, Hamilton Colgate universiteti.
- 2.1988-yil, 12-14 iyun, Qo'shma Shtatlar Pittsburgh, Carnegie Mellon universiteti.
- 3.1990-yil, 11-13 iyun, Qo'shma Shtatlar Texas, Austindagi Texas universiteti.
- 4.1992-yil, 15-27 iyun, Kanada Quebec, CCRIT-CWARC/
- 5.1993-yil, 14-16 iyun, Yaponiya Kyoto, Kyoto Xalqaro uyushmasi.
- 6.1995-yil, 5-7-iyun, Belgiya Leuven, Katholieke universiteti.
- 7.1997-yil, 23-25- iyun, Qo'shma Shtatlar N.Mex. Santa Fe. St.John's Kolleji.
- 8.1999-yil, 23-23-avgust, Angliya Chester, universitet kolleji.

Mashina tarjimasi sammiti

- 1.1987-yil, 17-19-sentabr, Yaponiya Hakone.
 - 2.1989-yil, 16-18 -avgust, G'arbiy Germaniya, Munich.
 - 3.1991-yil, 1-4 iyul, Qo'shma Shtatlar Vashington D,C.
 - 4.1993-yil, 20-23-iyul, Yaponiya Kobe.
 - 5.1995-yil, 10-13-iyul Lyuksemburg.
 - 6.1997-yil, 29-oktabrdan 1-noyabrgacha Qo'shma Shtatlar Kalif, San Diego.
 - 7.1999-yil, 13-17- sentabr, Singapur.
AMTA (Amerikaliklarning mashina tarjimasi bilan assotsiya)
 - 1.1994-yil, 5-8 oktabr Qo'shma Shtatlar Kolumbiya.
 - 2.1996-yil, 2-5-oktabr, Kanada Montreal.
 - 3.1998-yil, 28-31 oktabr, Qo'shma Shtatlar Pa., Langhorne.
 - 4.2000-yil, 10-14-oktabr, Meksika, Kuernavasa.
- Quyida mashina tarjimasining funksional imkoniyatlari ko'rsatilgan namunalar bilan tanishingiz mumkin:

Skayner orqali tarjima qiladigan tarjimon ruchka



9.2. Turkologiyada soha doirasidagi tadqiqotlar

Bugungi kunga kelib butun dunyoda qariyb 200 mln.ga yaqin aholi turkiy tilda gaplashadi. Lingvistik ma'lumotlar omborining tobora kengayib borayotgani, qolaversa, har bir tilda muayyan sohaga doir yangi atamalarining kirib kelayotgani aloqa muloqot jarayonida ayrim noqulayliklar yaratmoqda. Shu bois turkiy tillar doirasida mashina tarjimasiga oid izlanishlar rivojlanib bormoqda. Ayniqsa, ijtimoiy tarmoqlardagi muloqot uchun mashina tarjimasiga bo'lgan ehtiyoj ortmoqda.

Dorster rahbarligida ilk bor turkiy tillar bo'yicha mashina tarjimasiga doir tajriba o'tkazilgan (1954). O'sha yillardan buyon turkiy tillarning o'ziga xos tarjima usullari yaratildi. Tarjima texnologiyasi uchun ikki yo'nalish ahamiyatli:

1. Qardosh turkiy tillar o'rtasidagi mashina tarjimasiga.
2. Qardosh bo'lmagan va turkiy tillar o'rtasidagi mashina tarjimasiga.

Shu jihatdan turkologiyadagi bunday tadqiqotlar turli yondashuv va tarjima qilish texnologiyalarni yuzaga keltirdi. Mashina tarjimasida ko'plab izlanishlar olib borilgan bo'lsa-da, turkologiyadagi tadqiqotlarning aksariyati asrimizning keyingi o'n yilligiga borib taqaladi.

Manbalarga ko'ra turkiy tillar doirasida mashina tarjimasiga doir tadqiqotlar o'tgan asrning 90-yillari o'rtalarida boshlandi. Shulardan internetda tarjimon dasturlarning mavjudligi: ruscha-o'zbekcha (www.spells.uz), ruscha-qozoqcha, qozoqcha-ruscha (www.sanasoft.kz), azarbayjoncha-inglizcha, azarbayjoncha-turkcha (www.dilmanc.az), uyg'urcha-xitoycha

(www.jofcis.com/downloadpaper.aspx?), uyg'urcha-yaponcha¹⁶⁵, turkcha-qrim-tatarcha¹⁶⁶, shuningdek, qardosh tillar o'rtasida azarbayjoncha-turkcha va turkcha-qrim-tatarcha mashina tarjimasiga tizimlari qoidaga asoslangan (Rule based MT) mashina tarjimasiga metodiga tayanib qardosh tillar uchun alohida veb-resurs sifatida tarjima tizimi yaratilganligini <http://sem.antat.ru/> saytida ko'rish mumkin¹⁶⁷(21-rasm). Ushbu resurs turkiy tillarning so'z shakllari, gaplar va fayllardan olingan matnlarni tarjima qilishga mo'ljallangan.

ПЕРЕВОД ПРЕДЛОЖЕНИЙ ПЕРЕВОД ТЕКСТА ИЗ ФАЙЛА

ПЕРЕВОД СЛОВОФОРМЫ

Язык перевода: Выберите язык Перевод на язык: Татарский ; Казахский ; Турецкий ; Узбекский ; Русский

Введите слово для перевода:

21-rasm

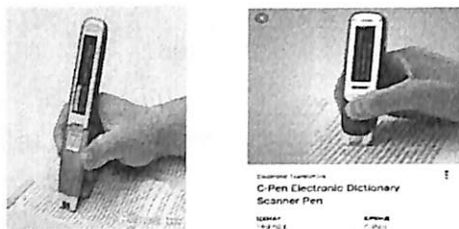
Matni tarjima qilish uchun interfeys orqali quyidagicha foydalanilish mumkin (22-rasm):

¹⁶⁵ Muhtar, M., Casablanca, F., Toyama, K., Inagaki, Y.: Particle-Based Machine Translation for Altaic Languages: the Japanese-Uighur Case. In: Proceedings of the 3rd Pacific Rim International Conference on Artificial Intelligence, Beijing, China, vol. 2, pp. 725-731 (1994).

¹⁶⁶ Kemal Altıntaş Turkish to Crimean Tatar Machine Translation System. MSc Thesis, Bilkent University, Ankara, 2000.

¹⁶⁷ Д.Ш. Сулейманов, А.Р. Гатиатуллин, Р.А. Гильмуллин, М.М. Аюпов К разработке татарско-турецкого машинного переводчика. / Turklang. 2013. Astana, 2013.- P. 271.

Skayner orqali tarjima qiladigan tarjimon ruchka



9.2. Turkologiyada soha doirasidagi tadqiqotlar

Bugungi kunga kelib butun dunyoda qariyb 200 mln.ga yaqin aholi turkiy tilda gaplashadi. Lingvistik ma'lumotlar omborining tobora kengayib borayotgani, qolaversa, har bir tilda muayyan sohaga doir yangi atamalarning kirib kelayotgani aloqa muloqot jarayonida ayrim noqulayliklar yaratmoqda. Shu bois turkiy tillar doirasida mashina tarjimasiga oid izlanishlar rivojlanib bormoqda. Ayniqsa, ijtimoiy tarmoqlardagi muloqot uchun mashina tarjimasiga bo'lgan ehtiyoj ortmoqda.

Dorster rahbarligida ilk bor turkiy tillar bo'yicha mashina tarjimasiga doir tajriba o'tkazilgan (1954). O'sha yillardan buyon turkiy tillarning o'ziga xos tarjima usullari yaratildi. Tarjima texnologiyasi uchun ikki yo'nalish ahamiyatli:

1. Qardosh turkiy tillar o'rtasidagi mashina tarjimasida.
2. Qardosh bo'lmagan va turkiy tillar o'rtasidagi mashina tarjimasida.

Shu jihatdan turkologiyadagi bunday tadqiqotlar turli yondashuv va tarjima qilish texnologiyalarni yuzaga keltirdi. Mashina tarjimasida ko'plab izlanishlar olib borilgan bo'lsa-da, turkologiyadagi tadqiqotlarning aksariyati asrimizning keyingi o'n yilligiga borib taqaladi.

Manbalarga ko'ra turkiy tillar doirasida mashina tarjimasiga doir tadqiqotlar o'tgan asrning 90-yillari o'rtalarida boshlandi. Shulardan internetda tarjimon dasturlarning mavjudligi: ruscha-o'zbekcha (www.spells.uz), ruscha-qozoqcha, qozoqcha-ruscha (www.sanasoft.kz), azarbayjoncha-inglizcha, azarbayjoncha-turkcha (www.dilmanc.az), uyg'urcha-xitoycha

(www.jofcis.com/downloadpaper.aspx?), uyg'urcha-yaponcha¹⁶⁵, turkcha-qrim-tatarcha¹⁶⁶, shuningdek, qardosh tillar o'rtasida azarbayjoncha-turkcha va turkcha-qrim-tatarcha mashina tarjimasida tizimlari qoidaga asoslangan (Rule based MT) mashina tarjimasida metodiga tayanib qardosh tillar uchun alohida veb-resurs sifatida tarjima tizimi yaratilganligini <http://sem.antat.ru/> saytida ko'rish mumkin¹⁶⁷(21-rasm). Ushbu resurs turkiy tillarning so'z shakllari, gaplar va fayllardan olingan matnlarni tarjima qilishga mo'ljallangan.

ПЕРЕВОД ПРЕДЛОЖЕНИЙ ПЕРЕВОД ТЕКСТА ИЗ ФАЙЛА

ПЕРЕВОД СЛОВОФОРМЫ

Язык перевода: Выберите язык Перевод на язык: Татарский ; Казахский ; Турецкий ; Узбекский ; Русский .

Введите слово для перевода:

21-rasm

Matnni tarjima qilish uchun interfeys orqali quyidagicha foydalanilish mumkin (22-rasm):

¹⁶⁵ Muhtar, M., Casablanca, F., Toyama, K., Inagaki, Y.: Particle-Based Machine Translation for Altaic Languages: the Japanese-Uighur Case. In: Proceedings of the 3rd Pacific Rim International Conference on Artificial Intelligence, Beijing, China, vol. 2, pp. 725-731 (1994).

¹⁶⁶ Kemal Altuntas Turkish to Crimean Tatar Machine Translation System. MSc Thesis, Bilkent University, Ankara, 2000.

¹⁶⁷ Д.Ш. Сулейманов, А.Р. Гатиатуллин, Р.А. Гильмуллин, М.М. Аюпов К разработке татарско-турецкого машинного переводчика. / Turklang. 2013. Astana, 2013.- P. 271.

Язык перевода: **Татарский** | Перевод на язык: Татарский | Казахский | Туркменский | Узбекский | Русский

Введите текст для перевода:
 Абай ырымы дуураак кайты.

Введите

Исходный текст	Морфемы исходного текста	Перевод на Казахский
Абай ырымы дуураак кайты.	абай(ысф-0); ыры- (а)Дуураак(ысф-7549) дуу- Дуура- (к)Дуураак(ысф-2056) кайты- Дуураак(ысф-11280).	ага дуураак тагуураак кайты. ага дуураак тагуураак кайты ораам. ага дуураак кайтытагуураак кайты. ага дуураак кайтытагуураак кайты ораам. ага дуураак тагуураак кайты. ага дуураак тагуураак кайты ораам. ага дуураак кайтытагуураак кайты. ага дуураак кайтытагуураак кайты ораам. ага дуураак тагуураак кайты. ага дуураак тагуураак кайты ораам. ага дуураак кайтытагуураак кайты. ага дуураак кайтытагуураак кайты ораам. ага дуураак тагуураак кайты. ага дуураак тагуураак кайты ораам. ага дуураак кайтытагуураак кайты. ага дуураак кайтытагуураак кайты ораам.
		Перевод на Турецкий
		ага кайдыга дуу тагуураак кайты. ага кайдыга дуу тагуураак кайты.

22-рasm

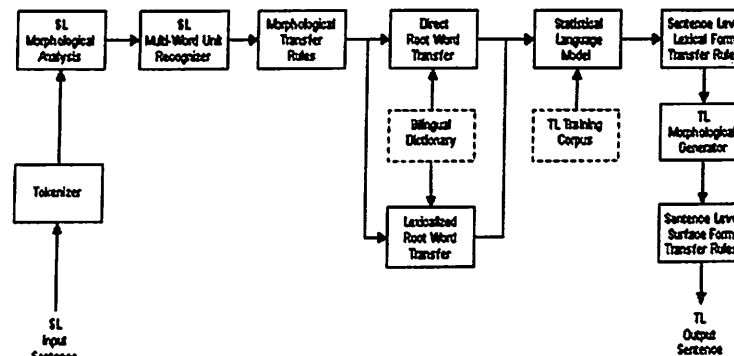
Turkiy tillar o'rtasidagi mashina tarjimasida bo'yicha bir qator izlanishlar olib borilgan. Xususan, turkman tilidan turk tiliga asoslangan mashina tarjimasida bo'yicha Güneyd Tantuğ, Aşref Adalı, Kemal Oflazer¹⁶⁸ tomonidan olib borilgan tadqiqotlarni uchratamiz. Ushbu mashina tarjimasida texnologiyasi tilning statistik modellari mavjud, ko'p ma'noli leksik va morfologik transfer asosidagi qoidaga asoslangan tizim sanaladi. Qardosh tillar o'rtasida MTlarning deyarli barchasida o'xshash modullar amalga oshirilgan: morfologik analizator, gap bo'laklariga ajratish (a part-of-speech tagger), bilingval transfer lug'ati va morfologik generator. Chex->Latish tilida deyarli sintaktik strukturalar ishlatilmaydi, negaki juftlikdagi so'zlarning ketma-ketligi deyarli bir xil. Turkman tilidan turk tiliga tarjima qilish bevosita morfologik transferga asoslangan. Mashina tarjimasida tizimi quyidagi asosiy bloklardan iborat¹⁶⁹ (24-rasm):

1. Asos matnning morfologik analizi

¹⁶⁸ Güneyd Tantuğ, Aşref Adalı, Kemal Oflazer Machine Translation between Turkic Languages / Proceedings of the ACL 2007 Demo and Poster Sessions, pages 189–192, Prague, June 2007. (Association for Computational Linguistics)

¹⁶⁹ Güneyd Tantuğ, Aşref Adalı, Kemal Oflazer Machine Translation between Turkic Languages / Proceedings of the ACL 2007 Demo and Poster Sessions, pages 189–192, Prague, June 2007. (Association for Computational Linguistics)

2. Asos tilning morfologik ko'p ma'noliligi
3. Bir nechta tarkibli so'zli birikmalarni o'qib olish
4. Morfologik transfer
5. So'z asosi transferi
6. Statistik ko'p ma'nolilik va jumlaning qayta tiklash
7. Gap darajasidagi qoidalar
8. Tarjima qilinayotgan tilning morfologik generatori



24-rasm

Shuningdek, Kemal Oflazer¹⁷⁰ ishlarida ingliz tilidan turk tiliga tarjima qilish bo'yicha statistik mashina tarjimasida tizimi yaratilgan. Unda tillarning morfologik va sintaktik qoidalarigagina emas, parallel matnlardagi ma'lumotlarga tayanilgan. Jumladan, 22,500 jumla namunasida so'zlarning kontekstadagi holatiga ko'ra ingliz va turk tillarining grammatik holati aniqlangan (25-rasm).

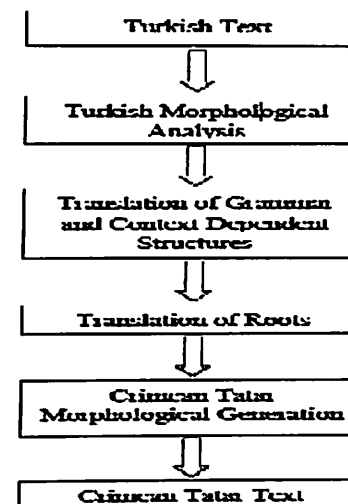
¹⁷⁰ Iknur Durgar El-Kahlout, Kemal Oflazer Initial Explorations in English to Turkish Statistical Machine Translation / Proceedings of the Workshop on Statistical Machine Translation, pages 7–14, New York City, June 2006. (Association for Computational Linguistics)

Wordform	Count	Gloss
faaliyet	3	'activity'
faaliyete	1	'to the activity'
faaliyetinde	1	'in its activity'
faaliyetler	3	'activities'
faaliyetlere	6	'to the activities'
faaliyetleri	7	'their activities'
faaliyetlerin	7	'of the activities'
faaliyetlerinde	1	'in their activities'
faaliyetlerine	5	'to their activities'
faaliyetlerini	1	'their activities (acc.)'
faaliyetlerinin	2	'of their activities'

25-rasm

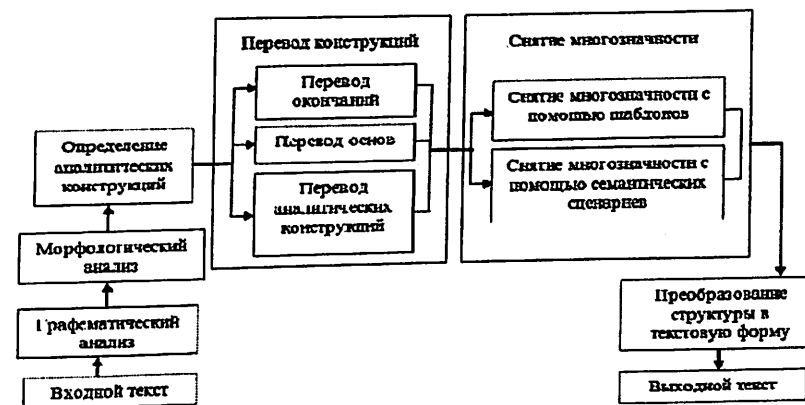
Statistik mashina tarjiması bu jihatdan tezkor va aniqroq ishlashi bilan xarakterlansa, boshqa tomondan resursi kam bo'lgan tillar uchun ushbu texnologiya u qadar o'zini oqlamaydi.

Turkiy tillar bo'yicha tatar tilidan turk tiliga tarjima qilish texnologiyasi Kemal Altintash tomonidan amalga oshirilgan. Unda tatar tiliga turk va rus tillaridan ko'plab so'zlar kirib kelgani bois sintaktik jihatdan turk tiliga o'xshash ekanligi aytiladi. Tarjima qilish tizimi¹⁷¹ 26-rasmda algoritm ko'inishda berilgan:



26-rasm

Shu jihatdan qarindosh tillar o'rtasidagi mashina tarjiması bo'yicha Dj.Suleymanov¹⁷² ishlarida tarjimadagi lingvistik ta'minot, asosan, morfologik va morfo-sintaktik darajada bo'linishini qayd etiladi va quyidagi chizmani pragmatik-mo'ljallangan mashina tarjiması tizimining umumiy arxitekturasi sifatida ifodalaydi (27-rasm).



27-rasm

¹⁷¹ Kemal Altıntaş Turkish to Crimean tatar machine translation system. Turkey, 2001. - P. 43.

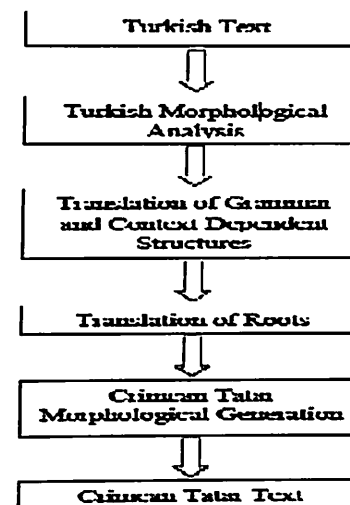
¹⁷² Сулейманов Д.Ш., Гатиатуллин А.Р., Гильмуллин Р.А., Аюпов М.М. К разработке татарско-турецкого машинного переводчика. / Turklang. 2013. - Astana, 2013. - P. 271-276.

Wordform	Count	Gloss
faaliyet	3	'activity'
faaliyete	1	'to the activity'
faaliyetinde	1	'in its activity'
faaliyetler	3	'activities'
faaliyetlere	6	'to the activities'
faaliyetleri	7	'their activities'
faaliyetlerin	7	'of the activities'
faaliyetlerinde	1	'in their activities'
faaliyetlerine	5	'to their activities'
faaliyetlerini	1	'their activities (acc.)'
faaliyetlerinin	2	'of their activities'

25-rasm

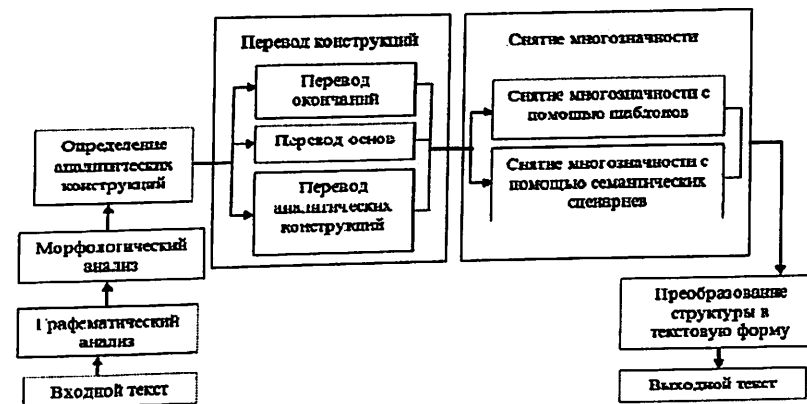
Statistik mashina tarjiması bu jihatdan tezkor va aniqroq ishlashi bilan xarakterlansa, boshqa tomondan resursi kam bo'lgan tillar uchun ushbu texnologiya u qadar o'zini oqlamaydi.

Turkiy tillar bo'yicha tatar tilidan turk tiliga tarjima qilish texnologiyasi Kemal Altıntaş tomonidan amalga oshirilgan. Unda tatar tiliga turk va rus tillaridan ko'plab so'zlar kirib kelgani bois sintaktik jihatdan turk tiliga o'xshash ekanligi aytiladi. Tarjima qilish tizimi¹⁷¹ 26-rasmda algoritm ko'rinishda berilgan:



26-rasm

Shu jihatdan qarindosh tillar o'rtasidagi mashina tarjiması bo'yicha Dj.Suleymanov¹⁷² ishlarida tarjimadagi lingvistik ta'minot, asosan, morfologik va morfo-sintaktik darajada bo'linishini qayd etiladi va quyidagi chizmani pragmatik-mo'ljallangan mashina tarjiması tizimining umumiy arxitekturasi sifatida ifodalaydi (27-rasm).



27-rasm

¹⁷¹ Kemal Altıntaş Turkish to Crimean tatar machine translation system. Turkey, 2001. - P. 43.

¹⁷² Сулейманов Д.Ш., Гатиатуллин А.Р., Гильмуллин Р.А., Аюпов М.М. К разработке татарско-турецкого машинного переводчика. / Turklang. 2013. - Astana, 2013. - P. 271-276.

O'zbek tili bo'yicha mashina tarjimasiga doir deyarli fundamental izlanish olib borilmagan edi. M.Xakimov ishlarida kengaytirilgan matematik modelga asoslanilgan mashina tarjimasi texnologiyasi bo'yicha tadqiqotlar amalga oshirilgan. Biroq o'zbek tilidagi ma'lumotlarning yetarli emasligi dasturiy ta'minotga ta'sir ko'rsatadi.

Ertuğrul Yılmaz, İlknur Durgar El-Kahlout, Burak Aydın, Zişan Sıla Özil, Coşkun Mermer kabi olimlar frazaga va iyerarxik frazaga asoslangan statistik mashina tarjimasi (turkcha-inglizcha) texnologiyasi ko'zga tashlanadi. Gorkem Ozbek¹⁷³ ishlarida esa tarjima jarayoni quyidagicha amalga oshirilgan: birinchi bosqich matnlardan olingan natija tahminiy morfologik ehtimolikka asoslanib, so'z shakllariga ajratiladi va frazalarning asosi aniqlanadi. Bu texnologiya Sabanji universiteti (Turkiya) professori Kemal Oflazer tomonidan ishlab chiqilgan. Ushbu manbada qayd etilishicha, turli janrdagi 22000 dan ziyod gap juftliklari ingliz va turk tillari bo'yicha parallel korpus holda to'plangan. Shuningdek, alifbodagi har xillikni hisobga olib, satr (string) tipidagi ma'lumotlarni yagona formatga keltirish uchun matnni kichik shriftga va milliy yunikod (unicod)dan standart Unicodga o'tkazilgan. Matnni segmentlarga ajratish jarayoni turkcha gaplarni morfologik analizi uchun zarur sanaladi. Shu tarzda matnni stem (asos) va morfemalarga ajratish tarjima modelini hosil qilishda rol o'ynaydi. So'zlarni tahlil qilish quyidagicha:

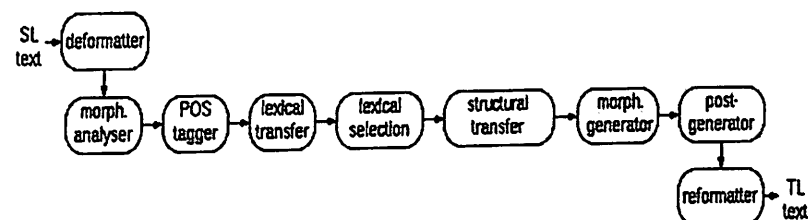
Shakily qismi	To'g'ri tahlil	Tahminiy tahlil	Inglizcha tarjimasi
kelebegin	kelebek #in	kelebe #gin	the butterfly's
yörelerdeki	yöre #ler #de #ki	yöre #ler #de #ki	in the areas

Segmentlash jarayoni quyidagi juft fayllardan iborat: 1-sida so'zlarning asosi, 2-sida asos va affikslardan iborat ma'mumotlar bazasi. Keyingi fayl ushbu fayllardagi ma'lumotlarning qay tarzda

¹⁷³Gorkem Ozbek, Siddharth Jonathan TU R K A L A T O R A Suite of Tools for Augmenting English-to-Turkish Statistical Machine Translation. 2006.

bog'lanishiga javob beradi. So'zlarni ajratishda ushbu holatda ikki bosqichli Giza++¹⁷⁴ instrumentidan foydalanilgan. Ikki bosqichli texnologiyada dastlab bevosita juftlikdagi parallel korpuslar stemmizatsiya qilinadi (so'z shakllari asoslarga ajratiladi). So'ngra ichki moslashish jarayoni olib boriladi. Olimning fikricha "e₁...e_n" va "f₁...f_n" parallel matnlardagi ikki shartdan biri mavjud bo'lsa, hisobga olinadi: agar e_i va f_i lug'atda mavjud bo'lsa, uning birligi sifatida yoki e_i va f_i lar e_j (1 ≤ j ≤ n) yoki f_j (1 ≤ j ≤ m) holati bilan lug'atda mavjud bo'lmaydi. Oxirgi jarayonda Giza++dan ikki fayl olinadi. Shu tarzda ma'lumotlar ustida amal bajariladi.

Ingliz tilidan qozoq tiliga tarjima qilish tizimi bo'yicha ochiq resursli Apertium platformasi orqali struktural transferga asoslangan mashina tarjimasi bo'yicha bir qator olimlar izlanish olib borgan. Manba¹⁷⁵ga ko'ra Apertium platformasi¹⁷⁶ mashina tarjimasining ochiq platformasi bo'lib, unda ingliz tilining morfologik lug'ati tayyor holda kiritilgan. Qozoq tilining morfologik va morfotaktik bazasi asosida har ikki til uchun struktural transfer qoidalari ishlab chiqilgan. Apertium platformasining ishlash tizimi (28-rasm):



28-rasm

¹⁷⁴ Och, Franz J. 2000. Giza++: Training of Statistical Translation Models. Available at <http://www-i6.informatik.rwthachen.de/~och/software/GIZA++.html>.

¹⁷⁵ A. Sundetova, M.L.Forcada, A. Shormakova, A. Aitkulova Structural transfer rules for English-to-Kazakh machine translation in the free/open-source platform Apertium/Turklang. 2013. -Astana, 2013.- P. 322-331.

¹⁷⁶ Apertium platformasi 2005 yili d'Alacant universiteti tomonidan yaratilgan. U qarindosh tillar uchun mo'ljallangan bo'lsa-da, keyinchalik uning boshqa til oilalariga mansub tillar ham kiritila boshlangan.

O'zbek tili bo'yicha mashina tarjimasiga doir deyarli fundamental izlanish olib borilmagan edi. M.Xakimov ishlarida kengaytirilgan matematik modelga asoslanilgan mashina tarjimasi texnologiyasi bo'yicha tadqiqotlar amalga oshirilgan. Biroq o'zbek tilidagi ma'lumotlarning yetarli emasligi dasturiy ta'minotga ta'sir ko'rsatadi.

Ertuğrul Yilmaz, İlknur Durgar El-Kahlout, Burak Aydın, Zişan Sıla Özil, Coş kun Mermer kabi olimlar frazaga va iyerarxik frazaga asoslangan statistik mashina tarjimasi (turkcha-inglizcha) texnologiyasi ko'zga tashlanadi. Gorkem Ozbek¹⁷³ ishlarida esa tarjima jarayoni quyidagicha amalga oshirilgan: birinchi bosqich matnlardan olingan natija tahminiy morfologik ehtimollikka asoslanib, so'z shakllariga ajratiladi va frazalarning asosi aniqlanadi. Bu texnologiya Sabanji universiteti (Turkiya) professori Kemal Oflazer tomonidan ishlab chiqilgan. Ushbu manbada qayd etilishicha, turli janrdagi 22000 dan ziyod gap juftliklari ingliz va turk tillari bo'yicha parallel korpus holda to'plangan. Shuningdek, alifbodagi har xillikni hisobga olib, satr (string) tipidagi ma'lumotlarni yagona formatga keltirish uchun matnni kichik shriftga va milliy yunikod (unicod)dan standart Unicodga o'tkazilgan. Matnni segmentlarga ajratish jarayoni turkcha gaplarni morfologik analizi uchun zarur sanaladi. Shu tarzda matnni stem (asos) va morfemalarga ajratish tarjima modelini hosil qilishda rol o'ynaydi. So'zlarni tahlil qilish quyidagicha:

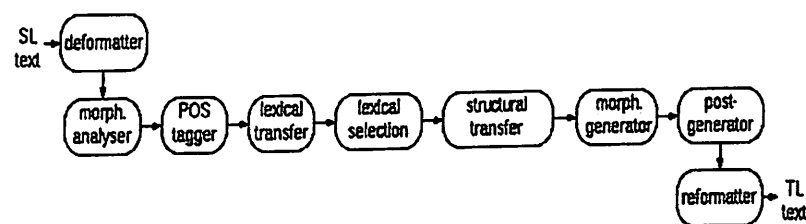
Shakily qismi	To'g'ri tahlil	Tahminiy tahlil	Inglizcha tarjimasi
kelebegin	kelebek #in	kelebe #gin	the butterfly's
yörelerdeki	yöre #ler #de #ki	yöre #ler #de #ki	in the areas

Segmentlash jarayoni quyidagi juft fayllardan iborat: 1-sida so'zlarning asosi, 2-sida asos va affikslardan iborat ma'mumotlar bazasi. Keyingi fayl ushbu fayllardagi ma'lumotlarning qay tarzda

¹⁷³Gorkem Ozbek, Siddharth Jonathan TU R K A L A T O R A Suite of Tools for Augmenting English-to-Turkish Statistical Machine Translation. 2006.

bog'lanishiga javob beradi. So'zlarni ajratishda ushbu holatda ikki bosqichli Giza++¹⁷⁴ instrumentidan foydalanilgan. Ikki bosqichli texnologiyada dastlab bevosita juftlikdagi parallel korpuslar stemmizatsiya qilinadi (so'z shakllari asoslarga ajratiladi). So'ngra ichki moslashish jarayoni olib boriladi. Olimning fikricha "e₁...e_n" va "f₁...f_n" parallel matnlardagi ikki shartdan biri mavjud bo'lsa, hisobga olinadi: agar e_i va f_i lug'atda mavjud bo'lsa, uning birligi sifatida yoki e_i va f_i lar e_j (1 ≤ j ≤ n) yoki f_j (1 ≤ j ≤ m) holati bilan lug'atda mavjud bo'lmaydi. Oxirgi jarayonda Giza++dan ikki fayl olinadi. Shu tarzda ma'lumotlar ustida amal bajariladi.

Ingliz tilidan qozoq tiliga tarjima qilish tizimi bo'yicha ochiq resursli Apertium platformasi orqali struktural transferga asoslangan mashina tarjimasi bo'yicha bir qator olimlar izlanish olib borgan. Manba¹⁷⁵ga ko'ra Apertium platformasi¹⁷⁶ mashina tarjimasining ochiq platformasi bo'lib, unda ingliz tilining morfologik lug'ati tayyor holda kiritilgan. Qozoq tilining morfologik va morfotaktik bazasi asosida har ikki til uchun struktural transfer qoidalari ishlab chiqilgan. Apertium platformasining ishlash tizimi (28-rasm):



28-rasm

¹⁷⁴ Och, Franz J. 2000. Giza++: Training of Statistical Translation Models. Available at <http://www-i6.informatik.rwthachen.de/~och/software/GIZA++.html>.

¹⁷⁵ A. Sundetova, M.L.Forcada, A. Shormakova, A. Aitkulova Structural transfer rules for English-to-Kazakh machine translation in the free/open-source platform Apertium/Turklang. 2013. -Astana, 2013. - P. 322-331.

¹⁷⁶ Apertium platformasi 2005 yili d'Alacant universiteti tomonidan yaratilgan. U qarindosh tillar uchun mo'ljallangan bo'lsa-da, keyinchalik uning boshqa til oilalariga mansub tillar ham kiritila boshlangan.

Qozoq tilidan rus tiliga tarjima qilish bo'yicha bir nechta tadqiqotchilarni sanash mumkin¹⁷⁷. Qozoqcha-ruscha mashina tarjimasi G.Belonogov tomonidan ishlab chiqilgan frazaga asoslangan tarjima qilish texnologiyasi bo'yicha izlanishlar olib borilmoqda¹⁷⁸. Bu tizim sintaktik transformatsiyalar asosida metafrazaga moslashtirilgan va formallashtirilgan sintaktik qurilmalar modeli bo'yicha amalga oshiriladi.

9.3. Mashina tarjimasi texnologiyasi va tizimlari

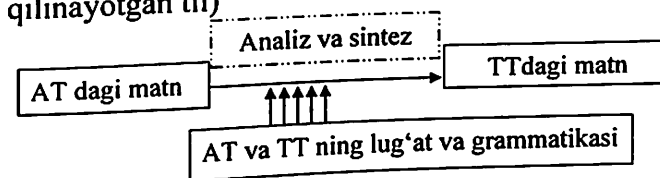
Mashina tarjimasi birikmasidagi *mashina* so'zi *sun'iy intellekt* ma'nosida insonga xos xususiyatlar: o'qish, tekshirish, tarjima va tahrir qilish jarayonlarining bajaruvchisi sifatida qaraladi. Kompyuter termini inglizcha *compute*- *hisoblam*oq so'zidan olingan. *Kompyuter tarjimasi* zamonaviylik kasb etsa-da, mohiyatan to'g'ri emas. Ba'zi o'rinlarda *avtomatik tarjima* birikmasi ham ishlatiladi. Lekin bu birikma mashina so'ziga nisbatan tor doirada, ya'ni faqat dastur orqali bajarish ma'nosida qo'llanadi. Mashina tarjimasi keng tushunchani anglatgani bois hozirda uning to'rt turi mavjud: 1) elektron lug'atlar yordamida tarjima (machine-assisted human translation –MAHT) – inson tomonidan matni tarjima qilishda kompyuter elektron lug'atlar va unga kiritilgan ko'rsatmalaridan foydalanish; 2) dasturiy ilovalar yordamidagi tarjima (computer assisted-translation –CAT) - matni kompyuter vositasida tarjima qilish va unda anglashilgan asosiy tushunchalarni ajratishga xizmat qiladi; 3) muharrir yordamidagi mashina tarjimasi (human-assisted machine translation –HAMT)-kompyuter dasturiy ta'minoti orqali amalga oshiriladigan tarjimaga inson resursi tomonidan tahrir qilish vazifasini yuklash; 4) to'liq avtomatik mashina tarjimasi (fully automatic machine translation –FAMT)-tarjimon dasturi orqali matni ham tarjima, ham tahrir qilish vazifalarini amalga oshirish.

¹⁷⁷ Жуманов Ж.М.1, Амирова Д.Т. Оценка качества машинного перевода в сценарии ассимиляции для англо-казахской и казахско-русской языковых пар / Turklang. 2017. – Kazan, 2017. –P.170-180; Рахимова Д. Исследование моделей и методов семантики машинного перевода с русского языка на казахский язык: Дисс... канд. техн. наук, – Казахстан, 2014. – С. 24-50.

¹⁷⁸ Мейрамбекқызы Ж., Хорошилов А.А. Синтактико-семантические представления для развития казахско-русского фразеологического машинного перевода / Turklang. 2017. – Kazan, 2017. –P. 181-196.

Axborot asrida mashina tarjimasiga oid ko'p tizimlar yaratildi. Ushbu mashina tarjima tizimlari muayyan jihatlarga ko'ra xarakterlanadi¹⁷⁹: 1) tillarning ishtiroki: a) ikki tilli; b) ko'p tilli; 2) tanlangan mavzuning miqdori: a) muayyan uslub va sohaga aloqador; b) bir uslub hamda ko'p sohalarga tegishli; 3) turli uslub va bir necha sohalar bilan bog'liq. Jon Xatchins mashina tarjimasi tizimlarini quyidagicha ajratadi¹⁸⁰: 1. Qoidaga asoslangan MT- MT-Rule-Based MT: 1) transfer tizim (transfer-based machine translation); 2) interlingval tizim (interlingual machine translation); 3) lug'at asosidagi tizim (dictionary-based machine translation); 2. Korpus asosidagi mashina tarjimasi- Corpus based MT: 1) statistikaga asoslangan MT(statistics-based); 2) namuna asosidagi MT(example-based); 3. Gibrid mashina tarjimasi- Hybrid machine translation (HMT). Gibrid mashina tarjimasi qoida hamda statistikaga asoslangan tarjima texnologiyasidan foydalanuvchi tizimdir. Bir nechta mashina tarjimasi kompaniyalari (Asia Online, LinguaSys, Systran, PangeaMT, UPV, Logos) mana shu tizimdan foydalanadi. Jon Xatchinsning yuqorida keltirilgan tasnifini quyidagi chizmalar orqali tushuntirish mumkin¹⁸¹:

1. To'g'ridan-to'g'ri tarjima tizimi (AT-asos til, TT- tarjima qilinayotgan til)



¹⁷⁹ Система машинного перевода «Кросслятор 2.0» и анализ ее функциональности для задачи трансляции знаний

http://www.keldysh.ru/papers/2007/prep89/prep2007_89.html

¹⁸⁰ <http://www.hutchinsweb.me.uk/Valladolid-2003-1-ppt.pdf>

¹⁸¹ Hutchins John. Machine translation: Past, present, future. –London, 1986. – P.118.

Qozoq tilidan rus tiliga tarjima qilish bo'yicha bir nechta tadqiqotchilarni sanash mumkin¹⁷⁷. Qozoqcha-ruscha mashina tarjimasi G.Belonogov tomonidan ishlab chiqilgan frazaga asoslangan tarjima qilish texnologiyasi bo'yicha izlanishlar olib borilmoqda¹⁷⁸. Bu tizim sintaktik transformatsiyalar asosida metafrazaga moslashtirilgan va formallashtirilgan sintaktik qurilmalar modeli bo'yicha amalga oshiriladi.

9.3. Mashina tarjimasi texnologiyasi va tizimlari

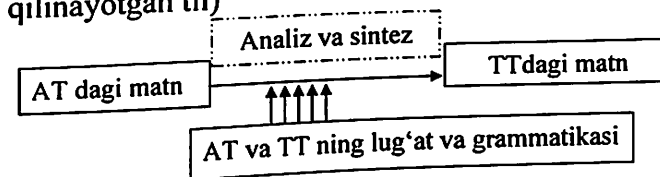
Mashina tarjimasi birikmasidagi *mashina* so'zi *sun'iy intellekt* ma'nosida insonga xos xususiyatlar: o'qish, tekshirish, tarjima va tahrir qilish jarayonlarining bajaruvchisi sifatida qaraladi. Kompyuter termini inglizcha *compute*- *hisoblamog* so'zidan olingan. *Kompyuter tarjimasi* zamonaviylik kasb etsa-da, mohiyatan to'g'ri emas. Ba'zi o'rinlarda *avtomatik tarjima* birikmasi ham ishlatiladi. Lekin bu birikma mashina so'ziga nisbatan tor doirada, ya'ni faqat dastur orqali bajarish ma'nosida qo'llanadi. Mashina tarjimasi keng tushunchani anglatgani bois hozirda uning to'rt turi mavjud: 1) elektron lug'atlar yordamida tarjima (machine-assisted human translation –MAHT) – inson tomonidan matni tarjima qilishda kompyuter elektron lug'atlar va unga kiritilgan ko'rsatmalaridan foydalanish; 2) dasturiy ilovalar yordamidagi tarjima (computer assisted-translation –CAT) - matni kompyuter vositasida tarjima qilish va unda anglashilgan asosiy tushunchalarni ajratishga xizmat qiladi; 3) muharrir yordamidagi mashina tarjimasi (human-assisted machine translation –HAMT)-kompyuter dasturiy ta'minoti orqali amalga oshiriladigan tarjimaga inson resursi tomonidan tahrir qilish vazifasini yuklash; 4) to'liq avtomatik mashina tarjimasi (fully automatic machine translation –FAMT)-tarjimon dasturi orqali matni ham tarjima, ham tahrir qilish vazifalarini amalga oshirish.

¹⁷⁷ Жуманов Ж.М.1, Амирова Д.Т. Оценка качества машинного перевода в сценарии ассимиляции для англо-казахской и казахско-русской языковых пар / *Turklang*. 2017. – Kazan, 2017. –P.170-180; Рахимова Д. Исследование моделей и методов семантики машинного перевода с русского языка на казахский язык: Дисс... канд. техн. наук, – Казахстан, 2014. – С. 24-50.

¹⁷⁸ Мейрамбекқызы Ж., Хорошилов А.А. Синтактико-семантические представления для развития казахско-русского фразеологического машинного перевода / *Turklang*. 2017. – Kazan, 2017. –P. 181-196.

Axborot asrida mashina tarjimasiga oid ko'p tizimlar yaratildi. Ushbu mashina tarjima tizimlari muayyan jihatlarga ko'ra xarakterlanadi¹⁷⁹: 1) tillarning ishtiroki: a) ikki tilli; b) ko'p tilli; 2) tanlangan mavzuning miqdori: a) muayyan uslub va sohaga aloqador; b) bir uslub hamda ko'p sohalarga tegishli; 3) turli uslub va bir necha sohalar bilan bog'liq. Jon Xatchins mashina tarjimasi tizimlarini quyidagicha ajratadi¹⁸⁰: 1. Qoidaga asoslangan MT- MT-Rule-Based MT: 1) transfer tizim (transfer-based machine translation); 2) interlingval tizim (interlingual machine translation); 3) lug'at asosidagi tizim (dictionary-based machine translation); 2. Korpus asosidagi mashina tarjimasi- Corpus based MT: 1) statistikaga asoslangan MT(statistics-based); 2) namuna asosidagi MT(example-based); 3. Gibrid mashina tarjimasi- Hybrid machine translation (HMT). Gibrid mashina tarjimasi qoida hamda statistikaga asoslangan tarjima texnologiyasidan foydalanuvchi tizimdir. Bir nechta mashina tarjimasi kompaniyalari (Asia Online, LinguaSys, Systran, PangeaMT, UPV, Logos) mana shu tizimdan foydalanadi. Jon Xatchinsning yuqorida keltirilgan tasnifini quyidagi chizmalar orqali tushuntirish mumkin¹⁸¹:

1. To'g'ridan-to'g'ri tarjima tizimi (AT-asos til, TT- tarjima qilinayotgan til)



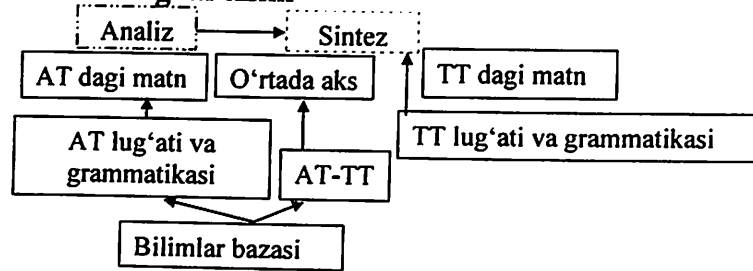
¹⁷⁹ Система машинного перевода «Кросслятор 2.0» и анализ ее функциональности для задачи трансляции знаний

http://www.keldysh.ru/papers/2007/prep89/prep2007_89.html

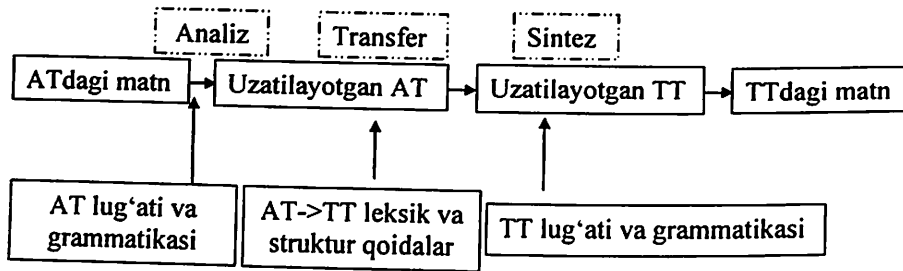
¹⁸⁰ <http://www.hutchinsweb.me.uk/Valladolid-2003-1-ppt.pdf>

¹⁸¹ Hutchins John. Machine translation: Past, present, future. –London, 1986. – P.118.

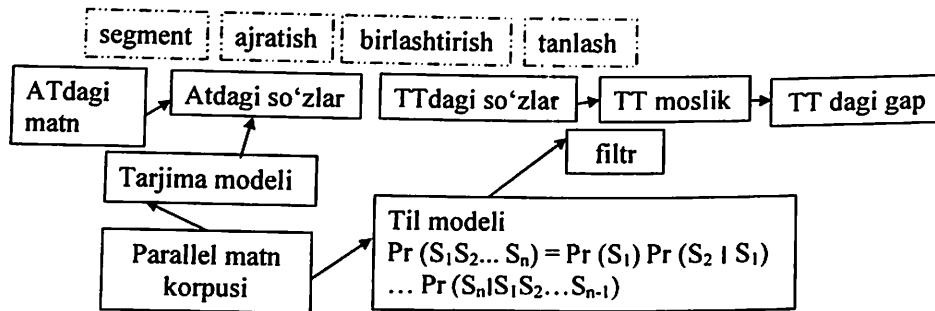
2. Interlingval tizim



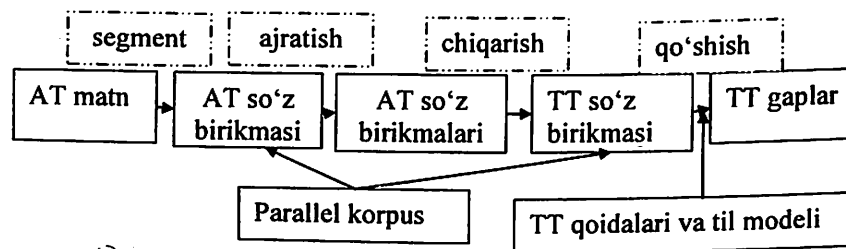
3. Transfer tizim



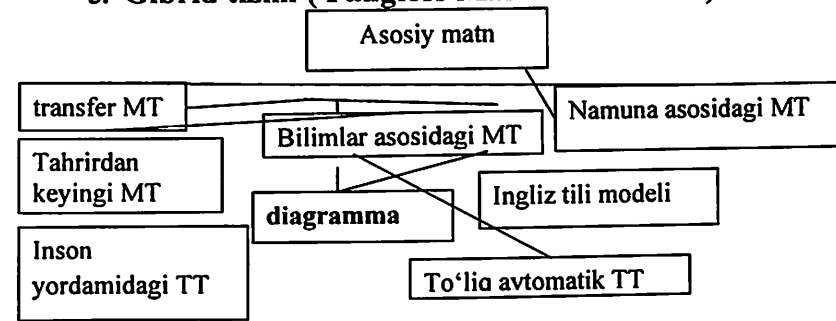
1. Statistika asoslangan tizim



2. Namuna asosidagi tizim



3. Gibril tizim (Pangloss Mark III misolida)

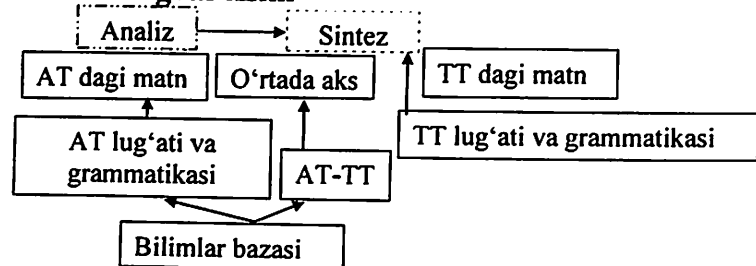


Yuqorida keltirilgan mashina tarjima metodlari bir qator olimlar tomonidan o'rganilgan hamda ularning samaradorlik darajasi o'zaro solishtirilgan. Jumladan, "Qoidaga asoslangan mashina tarjima" (Rule-Based MT) va "Statistik mashina tarjima" (Statistical MT)ning xususiyatlarini ko'rsatib o'tish o'rinli¹⁸². Qoidaga asoslangan mashina tarjima texnologiyasi (Rule-Based MT) har bir til juftliklarining cheksiz ifodalanishi mumkin bo'lgan lingvistik qoidalari hamda millionlab bilingval lug'atlarga tayanadi. Qoidaga asoslangan mashina tarjima, asosan, bir-biriga yaqin bo'lgan til juftliklari uchun qo'llaniladi. Uning ishlash texnologiyasiga ko'ra kompyuter dasturiy ta'minoti matnni tahlil qiladi va uzatilayotgan tavsifni tarjima qilinayotgan tildagi matndan to'playdi. Bu jarayon katta hajmli morfologik, sintaktik, semantik va qoidalardan iborat lug'atga ega bo'lgan ma'lumotlar bazasining bo'lishini talab qiladi. Dastur bu kabi murakkab qoidalardan foydalanib, asosiy tildagi matnni tarjima qilinayotgan ikkinchi tilning grammatik strukturasi o'tkazadi. Tarjima hajman katta lug'at hamda takomillashgan lingvistik qoidalar asosiga quriladi. Terminologiya mashina tarjima samaradorligini oshirishda muhim omil sifatida o'rin egallaydi.

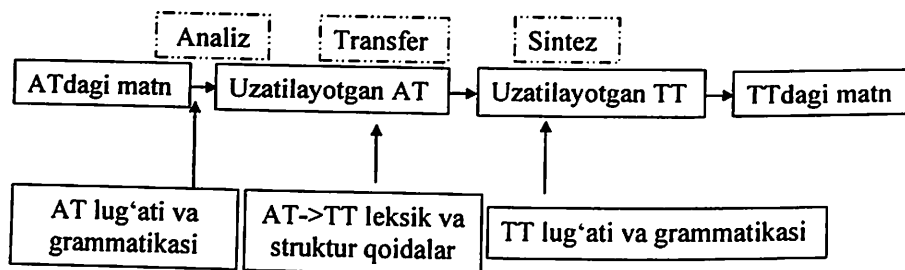
Qoidaga asosidagi mashina tarjima tarjima samaradorligini yaxshilashda tijorat kompaniyalaridan uzoq vaqt va katta mablag'ni talab etadi. Yuqoridagi metodlardan foydalangan holda matnni

¹⁸² Qarang: <http://www.systran.co.uk/systran/corporate-profile/translation-technology/what-is-machine-translation>

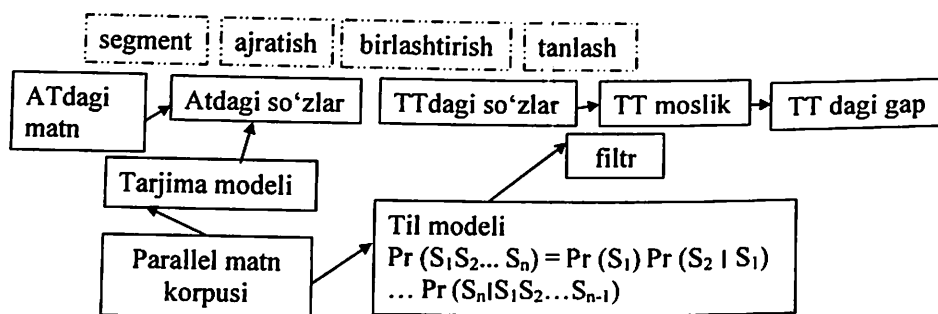
2. Interlingval tizim



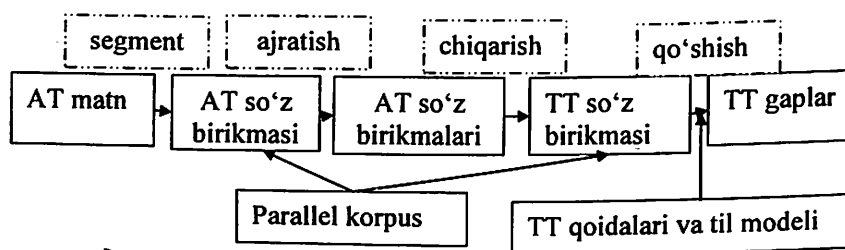
3. Transfer tizim



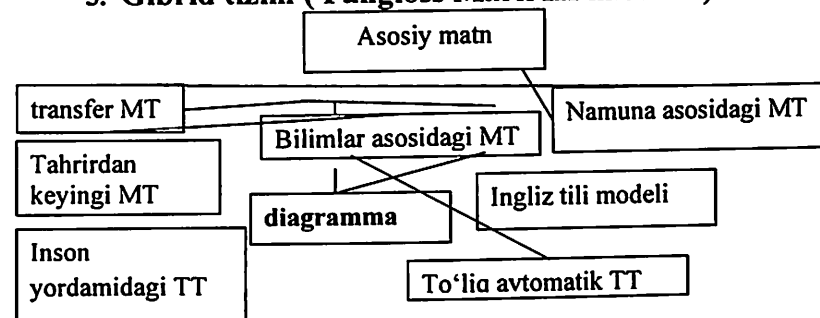
1. Statistika asoslangan tizim



2. Namuna asosidagi tizim



3. Gibril tizim (Pangloss Mark III misolida)



Yuqorida keltirilgan mashina tarjimasi metodlari bir qator olimlar tomonidan o'rganilgan hamda ularning samaradorlik darajasi o'zaro solishtirilgan. Jumladan, "Qoidaga asoslangan mashina tarjimasi" (Rule-Based MT) va "Statistik mashina tarjimasi" (Statistical MT)ning xususiyatlarini ko'rsatib o'tish o'rinli¹⁸². Qoidaga asoslangan mashina tarjimasi texnologiyasi (Rule-Based MT) har bir til juftliklarining cheksiz ifodalanishi mumkin bo'lgan lingvistik qoidalari hamda millionlab bilingval lug'atlarga tayanadi. Qoidaga asoslangan mashina tarjimasi, asosan, bir-biriga yaqin bo'lgan til juftliklari uchun qo'llaniladi. Uning ishlash texnologiyasiga ko'ra kompyuter dasturiy ta'minoti matnni tahlil qiladi va uzatilayotgan tavsifni tarjima qilinayotgan tildagi matndan to'playdi. Bu jarayon katta hajmli morfologik, sintaktik, semantik va qoidalardan iborat lug'atga ega bo'lgan ma'lumotlar bazasining bo'lishini talab qiladi. Dastur bu kabi murakkab qoidalardan foydalanib, asosiy tildagi matnni tarjima qilinayotgan ikkinchi tilning grammatik strukturasi o'tkazadi. Tarjima hajman katta lug'at hamda takomillashgan lingvistik qoidalar asosiga quriladi. Terminologiya mashina tarjimasi samaradorligini oshirishda muhim omil sifatida o'rin egallaydi.

Qoida asosidagi mashina tarjimasi tarjima samaradorligini yaxshilashda tijorat kompaniyalaridan uzoq vaqt va katta mablag'ni talab etadi. Yuqoridagi metodlardan foydalangan holda matnni

¹⁸² Qarang: <http://www.systran.co.uk/systran/corporate-profile/translation-technology/what-is-machine-translation>

avtomatik tarjima qilishda erishilgan natijalarni mutaxassislar quyidagi ko'rsatkichlar bo'yicha belgilaydilar:¹⁸³

Tarjima metodlari	Xatolik [%]
Semantik transfer metod	62
Dialog aktga asoslangan metod	60
Namuna asosidagi metod	52
Statistik metod	29

Statistik mashina tarjimasi texnologiyasi katta hajmli solishtirilgan til juftliklari hamda matn namunalaridan (korpuslar) foydalanuvchi mashina tarjimasi metodi hisoblanadi. Statistik mashina tarjimasi haqidagi dastlabki g'oya 1949-yili Uorren Uiver tomonidan Klauud Shannoning ma'lumot nazariyasi asosida yaratilgan. Keyinchalik ushbu nazariyaga ko'ra 1991 -yilda IBM tadqiqotchisi Tomas J. Vatson tomonidan bu tizim qayta o'rganilib chiqilgan. Statistik tarjima modellari dastlab so'z asosida yaratilgan (Models IBM Hidden dan 1-5 model, Stefan Vogeldan Markov model¹⁸⁴ hamda Frans Jozef Okdan 6 model¹⁸⁵), lekin keyinchalik ibora asosidagi modelar ham paydo bo'ldi¹⁸⁶. Statistik mashina tarjimasi usulidan IBM kompaniyasi foydalanib, IBM Model 1-5 modelini yaratdi. Google kompaniyasi ham bu metoddan foydalana boshladi. Bugungi kunda ayni metod quyidagi tizimlarda unumli qo'llanmoqda: 1) Giza++ ; 2) Moses; 3) Pharaoh; 4) Rewrite; 5) BLEU scoring tool. Statistik mashina tarjimasi monolingval va bilingval korpusning tahlilidan kelib chiquvchi parametrlarning statistik tarjima modellaridan foydalanadi. Statistik tarjima modelini

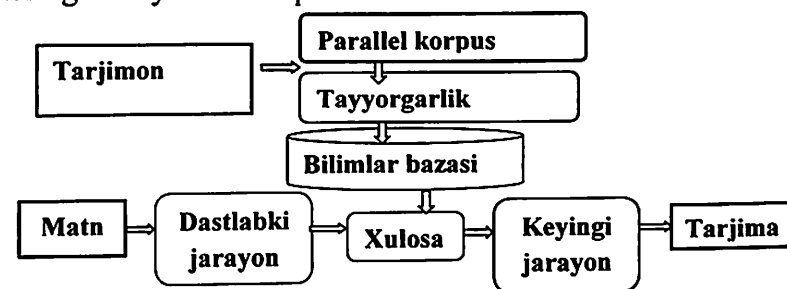
¹⁸³ Qarang: Franz Jozef Och, Hermann Ney. What can machine Translation learn from speech recognition? 2009.

¹⁸⁴ Vogel S., Ney H., Tillmann C. HMM-based Word Alignment in Statistical Translation. In COLING '96: The 16th International Conference on Computational Linguistics, - Copenhagen, 1996.-P. 836-841.

¹⁸⁵ Och F., Ney H., A Systematic Comparison of Various Statistical Alignment Models. Computational Linguistics 29 (1), 2003. -P. 19-51.

¹⁸⁶ Koehn P., Och F.J., Marcu D. Statistical phrase based translation. In Proceedings of the Joint Conference on Human Language Technologies and the Annual Meeting of the North American Chapter of the Association of Computational Linguistics (HLT/NAACL) 2003; Chiang D. A Hierarchical Phrase-Based Model for Statistical Machine Translation. In Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05), 2005.

yaratish qisqa jarayon, biroq u mavjud multilingval korpusga tayanadi. Maxsus soha uchun korpusda kamida ikki million so'z, umumiy til uchun undan ham ko'proq so'z leksikoni bo'lishini talab qiladi. Nazariy jihatdan sifatli tarjimaga erishish mumkin bo'lsa-da, ko'plab kompaniyalarda bunday kerakli tarjima modellarini yaratish uchun katta miqdordagi multilingval korpuslar mavjud emas. Statistik mashina tarjimasining o'rta darajadagi tarjima modellaridan foydalanishda kompyuter qismlarining kuchli hamda katta hajmdagi konfiguratsiyasi talab qilinadi.



(Statistik mashina tarjimasining qurilishi)¹⁸⁷

Mashina tarjimasi usullarini sinovdan o'tkazish natijasida ularning yutuq va kamchiliklari aks etgan ma'lumotga tayanib, quyidagicha ifodalash mumkin¹⁸⁸ (ijobiy tomoni -; salbiy tomoni + bilan belgilangan):

Qoida asosidagi MT (Rule-Based MT)	Statistik MT (Statistical MT)
+ barqaror va kutilgan sifatning mavjudligi	- oldindan bashorat qilinmaydigan natija
+ sohadan tashqari tarjima sifatining yuqoriligi	- sohadan tashqari sifatning pastligi
+ grammatik qoidalarni bilish imkoniyati	- grammatikani bilmaslik holati
+ ish sifatining yuqoriligi va kuchliligi	-katta hajmni egallashga bo'lgan talabning mavjudligi
+variantlar o'rtasidagi barqarorlik	-variantlar o'rtasidagi beqarorlik

¹⁸⁷ O'sha asar.

¹⁸⁸ Qarang: <http://www.systran.co.uk/systran/corporate-profile/translation-technology/what-is-machine-translation>

-ravonlikka ega bo'lmalik	+yuqori ravonlikka egalik
-qoidalardan istisnolarni nazorat qilishning qiyinligi	+ qoidalar istisnolarini topishining osonligi
-tijorat narxi va rivojlanishning yuqoriligi	+tez hamda samarador rivojlanish imkoniyati mavjudligi

<https://ru.smartcat.ai/corp.features>, <https://translate.yandex.ru>, <https://dilmanc.az>, <https://solver.uz>, <https://www.from-to.uz> kabi saytlarda o'zbek tili bilan bog'liq tizimlar yaratilgan. Shular jumlasidan 2014-yil dekabr oyida Google tarjimonga o'zbek tili ham qo'shildi.

Endilikda onlaynda Translate.ru (PROMT), Google Translate, Free Translation

Worldlingo, Babel Fish va Translate Online.ua serverlari mavjud bo'lib yuqori funksional imkoniyatlariga ega. Xususan, Wordlingo 15 ta tilni o'z ichiga oladi va unda tarjima texnologiyasining web sahifalar tarjimasi, matn tarjimasi, elektron pochta uchun tarjima va chat tarjimasi kabi imkoniyatalari mavjud (29-rasm).

WorldLingo

Machine Translation | Professional Translation | About Us | My Account

WorldLingo Translation API Development Tool

Communicate in 15 languages on any application or platform

Today's competitive and political global market has created an urgent need both within corporations and governments for inter-organizational communication and understanding of foreign content. By leveraging technology from WorldLingo and adding the power of advanced translation to your software, intranet, corporate information system or website you can create communications between staff and customers, partners, and suppliers from around the globe.

Simple solution for easy integration

Calling party sends HTTP/HTTPS request and server sends back response

1. Calling party sends HTTP request

2. Server sends back HTTP request

WorldLingo's fully documented Translation API solution is delivered via our trusted API model which means you have direct access to the industry's most sophisticated translation technology capable of translating complex sentences. WorldLingo uses only proven, open Internet protocols and standards (HTTP, HTTPS, SSL, XML), allowing easy integration of the service into software systems on any platform supporting Internet technologies. The API requires no additional

Contact Us

Downloads:

- API Documentation

Customer Stories

- Oddcast Inc.
- Denver Metro Convention & Visitors Bureau
- Education Network of Ontario

Related Products

- Website Translator
- Text Translator
- Email Translator
- Chat Translator

Sender's Info:

Name: _____

Email Address: _____

Language: Arabic

Recipient's Info:

Name: _____

Email Address: _____

Language: Arabic

Subject: _____

Message: _____

Translate & Send Email

Translation powered by WorldLingo. Note: This is a computer translation of text. It is provided for general information only and should not be regarded as complete nor accurate.

29-rasm

Matni avtomatik tarjima qilish jarayoni va unga ta'sir etuvchi lingvistik omillar

Matni sof lingvistik nuqtai nazardan tarjima qilish 2 xil usulda amalga oshiriladi: bevosita va bilvosita. Bevosita tarjima qilishda so'zlarni o'zlashtirish, kalkalash, so'zma so'z tarjima qilish usullarini keltirish mumkin. Bilvosita tarjima qilishga transpozitsiya (gap bo'laklarini almashtirish), modulyatsiya (nuqtai nazarni o'zgartirish), ekvivalensiya (maqolni boshqa shaklda berish), adaptatsiya (tavsiflanayotgan vaziyatni almashtirish) kabilar kiradi. Shu nuqtai nazardan matnda shakl va mazmun munosabatlari turli vaziyatlarda o'z ifodasini topgan holda tarjima jarayonini yuzaga keltirishda omil vazifasini o'taydi. Mashina tarjimasi bu avtomatik tarjimadir. Bu kompyuter dasturiy ta'minoti orqali amalga oshuvchi bir tabiiy tildagi matni (ingliz tilidan) boshqa bir tilga (o'zbek tiliga) tarjima qilish jarayonidir.

Tarjima inson yordamida yoki avtomatik tarzda amalga oshirilsa-da, original (asosiy) tildagi matn mazmuni tarjima qilinayotgan tildagi tarjimada to'liq tiklanishi kerak. Bu jarayon tashqi tomondan oson tuyulsa-da, biroq ancha murakkabligi bilan ajralib turadi. Tarjima faqat so'zma so'z o'ringa qo'yish emas. Tarjimon matndagi barcha elementlarni tahlil va tarjima qilishi hamda muayyan bir so'zni boshqa so'zlarga qanday ta'sir qilishini bilishi kerak. Bunda tarjimondan asos til va tarjima qilinayotgan tilning morfologik, sintaktik (gap strukturasi), semantik (mazmun),

shuningdek, lokal xududi bilan o'xshashligi bo'yicha tajribaga ham ega bo'lish talab qilinadi.

Ikkitadan kam bo'lmagan individual tarjimon yagona til bo'yicha bir xil matnни o'zi xos tarzda tarjima qiladi hamda iste'molchining qoniqishi uchun uni yana bir necha marta tekshirib chiqishni taqozo etadi. Tarjimadan avvalgi tahlil jarayonida uch asosiy iyerarxik bosqich amalga oshiriladi: matn-jumla-so'zshakl. Bunda so'zshakl matndan olingan istalgan segment birlik sifatida qaraladi. Dasturiy ta'minot uchun lingvistik tavsif muhim ahamiyatga egadir. So'zshakl va uning komponentlari to'g'risida ma'lumot berishning uch xil turi beriladi:

- Til va matndagi lingvistik birliklarni tavsiflovchi ma'lumot;
- Ichki til birliklarining tavsifini aniqlovchi ma'lumot;
- Tillararo munosabat va mohiyatini ochib beruvchi ma'lumot.

Ingliz tilidagi sodda gapshakllarini o'zbek tiliga avtomatik tarjima qilish namunasi orqali dasturiy ta'minotga kiritilayotgan bilimlar bazasi lingvistik nuqtai nazardan yondashiladi. Chunki tarjima qilinayotgan tabiiy tillarning xususiyatlaridan kelib chiqib nazariy asoslarini berish orqali sifatli tarjimaga erishish mumkin bo'ladi.

Avtomatik tarjimaning oldida turgan muhim vazifalardan biri birinchi til matnidagi asosiy mazmunni saqlagan holda ikkinchi tildagi matnga tarjima qilishdir. Yuqoridagilardan eng katta mehnatni talab qiluvchi sath bu leksemalarning semantik maydoni va ularning valentliklarini aniqlashdir. Masalan, ingliz tilidagi *capital* so'zi quyidagi semantik maydonga ega:

I. [ot] *kapital; vaziyat, jamg'arma, tejash, sarmoya, daromad*

1)[Fe'l+leksema(capital)]

to borrow capital — qarzga sarmoya olmoq

to establish a capital — mablag' yig'moq

to invest, put up, tie up capital — mablag' sarflamoq

to make capital out of — sarmoya orttirmoq

to raise capital — sarmoya ko'paytirmoq

to withdraw capital — sarmoyani tortib olmoq/olib qo'yimoq

2) [sifat+ leksema(capital)]

borrowed capital — qarzga olingan mablag'

circulating capital — aylanma daromad

foreign capital — horijiy sarmoya

idle capital — iste'moldan chiqqan mablag'

industrial capital — sanoatlashgan jamg'arma

national, state capital — milliy sarmoya

permanent capital — doimiy mablag'

provincial capital — hududiy moliyaviy jamg'arma

world capitals — dunyo sarmoyalari

working capital — muomaladagi pul birligi

fixed capital — o'zgarmas vaziyat

political capital — siyosiy kapitalistlar

II [ot] 1) *poytaxt*

Tashkent is the capital of Uzbekistan. — Toshkent— O'zbekiston

poytaxti.

III. [sifat(capital)+ot] 1) *asosiy, bosh, muhim, katta*

capital letter — katta/bosh harf

Capital letter is used in all proper names. — Barcha atoqli otlar bosh harf bilan yoziladi.

capital city — katta shahar

capital stock — asosiy zahira

IV. (so'zlashuv uslub) [sifat(capital)+ot] 1) *ajoyib, zo'r*

capital boy — zo'r bola

V. (huquq sohasi) [sifat(capital)+ot] 1) *og'ir, katta*

capital punishment-og'ir jazo

VI. (arxitektura) [ot]

Capital -kapitel-ustunning shipga tutashgan qismi

Tanlab olingan va cheklangan leksemalarning lingvistik modulini berishda mana shu vazifalardan kelib chiqish maqsadga muvofiq.

Shuningdek, polisemiya avtomatik tarjimadagi muammoli vaziyatlardan biri leksemalarning polisemantik ma'no kasb etishidir. Ma'lumki, polisemiya bu ko'pma'nolilik bo'lib, semantik jihatdan ular bir-biriga aloqador bo'ladi. Masalan, o'zbek tilidagi bosh leksemasi tananing yuqori qismi, yuz qismi joylashgan a'zo ma'no ottenkalarini ifodalasa, shu ma'noni saqlagan holda, u yana bir qancha so'zlar bilan ham qo'llanaveradi: ko'chaning boshi, ishning boshi, so'z boshi va boshqalar. Ingliz tilida ham polisemantik so'zlarning miqdori lug'at tarkibida katta miqdorni egallaydi.

O'zbek tilida polisemantik so'z faqat bir so'z turkum doirasida kuzatilsa, ingliz tilida ma'noga yaqinlik nuqtai nazaridan u turli so'z turkumiga mansub bo'lishi mumkin. Masalan,

ot	sifat	fe'l	Tarjimasi
act		act	xatti-harakat(qilmoq)
anger		anger	jahl (chiqmoq)
	approximate	approximate	taxminiy (taxmin qilmoq)
attack		attack	hujum (qilmoq)
base		base	asos (solmoq)
	calm	calm	tinch (-imoq)
care		care	diqqat (qaratmoq)
challenge		challenge	qiyinchilik; chaqirmoq
chemical	chemical		kimyoviy modda; kimyoviy
circle		circle	aylana; atrofida aylanmoq
	clean	clean	toza (-lamog)
close	close	close	yakun; yaqin; berkitmoq

Tilshunos Birdetal tadqiqotiga ko'ra (1987), Vebster muallifligidagi yettinchi marta tahrir etilgan ingliz tili izohli lug'atida 60000 ta so'zning 21488 tasi, ya'ni umumiy miqdorning 40% ini ikki va undan ortiq ma'noga ega bo'lgan polisemantik so'zlar tashkil etadi. Masalan, birgina run -"yugurmoq" so'zi 29 ta ma'noga ega bo'lib, 125 ta ichki ma'nolarga ajraladi. Albatta, leksikografiyada lug'at tuzuvchilarning polisemiya bo'lgan qarashlari, nuqtai nazarlari ham turlichadir. Ammo omonimiya hodisasi bilan polisemiya farqlash kerak bo'ladi. Shu jihatdan matni avtomatik tarjima qilayotganda polisemantik so'zlarning semantik maydonlarini alohida tasniflash va o'rganish kerak. Tabiiyki, ingliz tili mutaxassislari tomonidan polisemiya hodisasi o'rganilgan bo'lishi mumkin. Biroq avtomatik tarjima uchun tanlangan va chegaralab olingan barcha so'zlarning to'liq lingvistik ta'minotida bularning barchasi o'rin olishi va o'zbek tilidagi tarjimasi bilan barcha xususiyatlari ochib berilishi kerak.

1960 -yillarda Bar Xillel namuna sifatida pen so'zining "ruchka" va "o'yinchoq qutisi" ma'nolaridan kelib chiqib, "mashina tarjimasini 80 foiz xato", deb munosabat bildirgan edi. Endilikda dunyo bo'yicha o'rganilgan izlanishlar natijasida matni avtomatik tarjima qilishning 80 % muammosi o'rganilgan bo'lib, qolgan qismi

mutaxassislari tomonidan tadqiq etilmoqda. Mazkur tadqiqotlar sifatida Lesk (1986), Gutri (1991), Vilks (1993), Blek (1988), Zernik(1991), Xerst (1991), Gale (1992), Miller, Lekok (1998), Vurxis (1998), Adam Kilgarrif (1997) va boshqa olimlarning ishlarini eslatish o'rinli. Bugungi kunga kelib kontekstlarning korpusi yaratilgan bo'lib, ular o'z navbatida ikkiga bo'linadi: 1) lokal kontekst; 2) mavzuga oid kontekst. Leksemalarning qanday ma'noda qaysi so'z bilan kela olish tartibi va ro'yxati ana shu korpuslarda o'z ifodasini topadi. O'zbek kompyuter lingvistikasi oldidagi ulkan va mas'uliyatli vazifalardan biri o'zbekcha leksemalarning ana shunday strukturasi ega bo'lgan yaxlit korpusini yaratishdir.

Ingliz tilidagi matni o'zbek tiliga tarjima qilishda gap strukturasi joylashish tartibi va shakli yetarli emasligini birgina leksemaning xususiyatlariga ko'ra ham isbotlash mumkin. Biroq ingliz tilida qilingan mavjud ishlardan foydalangan holda berilayotgan ta'minotda ularning kodlangan tizimini ifodalash orqali bu masalani ham hal etish mumkin bo'ladi. Buning uchun tilshunoslarimizning fidoyi mehnati talab qilinadi.

Adabiyotlar:

1. Abdurahmonova N. Mashina tarjimasining lingvistik ta'minoti (monografiya) Toshkent: Muharrir, 2018., 165 b.
2. Марчук Ю.Н. Проблемы машинного перевода. -М.: 1983.
3. Пиотровский Р.Г., Билан В.Н., Боркун М.Н., Бобков А.К. Методы автоматического анализа и синтеза текста. -Минск. 1985.
4. Теория машинного перевода <http://rudocs.exdat.com/docs>

Savol va topshiriqlar

1. Mashina tarjimasining yuzaga kelishiga ta'sir ko'rsatgan omillar qaysilar?
2. Mashina tarjimasidan foydalangan holda multi-lingval internet tarmoqlari uchun qanday fransuz - nemis loyihasi ishlab chiqilgan?
3. Rossiyada lingvistik arifmometr muallifi P.P.Smirnov Troyanskiy tomonidan Esperanto tili asosida tillar o'rtasidagi

O'zbek tilida polisemantik so'z faqat bir so'z turkum doirasida kuzatilsa, ingliz tilida ma'noga yaqinlik nuqtai nazaridan u turli so'z turkumiga mansub bo'lishi mumkin. Masalan,

ot	sifat	fe'l	Tarjimasi
act		act	xatti-harakat(qilmoq)
anger		anger	jahl (chiqmoq)
	approximate	approximate	taxminiy (taxmin qilmoq)
attack		attack	hujum (qilmoq)
base		base	asos (solmoq)
	calm	calm	tinch (-imoq)
care		care	diqqat (qaratmoq)
challenge		challenge	qiyinchilik; chaqirmoq
chemical	chemical		kimyoviy modda; kimyoviy
circle		circle	aylana; atrofida aylanmoq
	clean	clean	toza (-lamog)
close	close	close	yakun; yaqin; berkitmoq

Tilshunos Birdetal tadqiqotiga ko'ra (1987), Vebster muallifligidagi yettinchi marta tahrir etilgan ingliz tili izohli lug'atida 60000 ta so'zning 21488 tasi, ya'ni umumiy miqdorning 40% ini ikki va undan ortiq ma'noga ega bo'lgan polisemantik so'zlar tashkil etadi. Masalan, birgina run -"yugurmoq" so'zi 29 ta ma'noga ega bo'lib, 125 ta ichki ma'nolarga ajraladi. Albatta, leksikografiyada lug'at tuzuvchilarning polisemiya bo'lgan qarashlari, nuqtai nazarlari ham turlichadir. Ammo omonimiya hodisasi bilan polisemiya farqlash kerak bo'ladi. Shu jihatdan matnni avtomatik tarjima qilayotganda polisemantik so'zlarning semantik maydonlarini alohida tasniflash va o'rganish kerak. Tabiiyki, ingliz tili mutaxassislar tomonidan polisemiya hodisasi o'rganilgan bo'lishi mumkin. Biroq avtomatik tarjima uchun tanlangan va chegaralab olingan barcha so'zlarning to'liq lingvistik ta'minotida bularning barchasi o'rin olishi va o'zbek tilidagi tarjimasi bilan barcha xususiyatlari ochib berilishi kerak.

1960 -yillarda Bar Xillel namuna sifatida pen so'zining "ruchka" va "o'yinchoq qutisi" ma'nolaridan kelib chiqib, "mashina tarjimasini 80 foiz xato", deb munosabat bildirgan edi. Endilikda dunyo bo'yicha o'rganilgan izlanishlar natijasida matnni avtomatik tarjima qilishning 80 % muammosi o'rganilgan bo'lib, qolgan qismi

mutaxassislar tomonidan tadqiq etilmoqda. Mazkur tadqiqotlar sifatida Lesk (1986), Gutri (1991), Vilks (1993), Blek (1988), Zernik(1991), Xerst (1991), Gale (1992), Miller, Lekok (1998), Vurxis (1998), Adam Kilgarrif (1997) va boshqa olimlarning ishlarini eslatish o'rinli. Bugungi kunga kelib kontekstlarning korpusi yaratilgan bo'lib, ular o'z navbatida ikkiga bo'linadi: 1) lokal kontekst; 2) mavzuga oid kontekst. Leksemalarning qanday ma'noda qaysi so'z bilan kela olish tartibi va ro'yxati ana shu korpuslarda o'z ifodasini topadi. O'zbek kompyuter lingvistikasi oldidagi ulkan va mas'uliyatli vazifalardan biri o'zbekcha leksemalarning ana shunday strukturasi ega bo'lgan yaxlit korpusini yaratishdir.

Ingliz tilidagi matnni o'zbek tiliga tarjima qilishda gap strukturasi joylashish tartibi va shakli yetarli emasligini birgina leksemaning xususiyatlariga ko'ra ham isbotlash mumkin. Biroq ingliz tilida qilingan mavjud ishlardan foydalangan holda berilayotgan ta'minotda ularning kodlangan tizimini ifodalash orqali bu masalani ham hal etish mumkin bo'ladi. Buning uchun tilshunoslarimizning fidoyi mehnati talab qilinadi.

Adabiyotlar:

1. Abdurahmonova N. Mashina tarjimasining lingvistik ta'minoti (monografiya) Toshkent: Muharrir, 2018., 165 b.
2. Марчук Ю.Н. Проблемы машинного перевода. -М.: 1983.
3. Пиотровский Р.Г., Билан В.Н., Боркун М.Н., Бобков А.К. Методы автоматического анализа и синтеза текста. -Минск. 1985.
4. Теория машинного перевода <http://rudocs.exdat.com/docs>

Savol va topshiriqlar

1. Mashina tarjimasining yuzaga kelishiga ta'sir ko'rsatgan omillar qaysilar?
2. Mashina tarjimasidan foydalangan holda multi-lingval internet tarmoqlari uchun qanday fransuz - nemis loyihasi ishlab chiqilgan?
3. Rossiyada lingvistik arifmometr muallifi P.P.Smironov Troyanskiy tomonidan Esperanto tili asosida tillar o'rtasidagi

grammatik boshqaruv usuli bilan taqsimlash metodi va bilingval lug'atni o'z ichiga oluvchi tizimning bosqichlarini toping.

4. Turkiy tillar bo'yicha dastlabki mashina tarjimasi kim boshchiligida amalga oshirilgan?

5. Y.N. Marchukning mashina tarjimasi haqidagi fikrlariga munosabatingizni bildiring.

6. Tilni tizim sifatida o'rganish shartli ekanligi to'g'risidagi N.Z. Kotelovanning mulohazasiga qanday qaraysiz?

7. Mashina tarjimasining rivojlanishiga ta'sir ko'rsatgan omillar qaysilar?

8. Ruscha matnlarni ingliz tiliga tarjima qiluvchi Retrans Vista versiyasining imkoniyatlari haqida fikr yuriting.

9. M.Hakimov ishlarida kengaytirilgan matematik modelga asoslangan mashina tarjimasi texnologiyasi bo'yicha amalga oshirilgan tadqiqotlar haqida misollar keltiring?

10. Mashina tarjimasining turlarini aniqlang va izohlang.

11. Internetda quyidagi <http://dic.academic.ru>, <http://sobolev-an.narod.ru>, <http://www1.oup.co.uk/oup/elt/software>, <http://www.lingvo.ru> saytlardan kerakli ma'lumotlarni daftaringizga qayd qiling. Ushbu serverdagi mavjud tarjimon dasturlar yordamida tarjima sifati va tarjima qilish imkoniyatini tavsiflang.

9-mavzu yuzasidan test

1. Mashina tarjimasidan foydalangan holda multilingval internet tarmoqlari uchun qanday fransuz – nemis loyihasi ishlab chiqilgan?

- e) Quaero
- f) POLO
- g) ANPAR
- h) Qualcom

2. Mashina tarjimasida metatil tilga oid tushunchalarning modeli kim tomonidan yaratilgan?

- a) I.A. Melchuk
- b) D.Panov
- c) N.Andreyev
- d) K.Bektayev

3. Mashina tarjimasida so'zlarni grammatik shakllarga ajratish

jarayoni qanday nomlanadi?

- a) tokenizatsiya
- b) asper
- c) lemmatizatsiya
- d) stemmizatsiya

4. Qaysi mashina tarjimasida so'zlarning asosi va sintaktik funksiyaga ko'ra leksemalarning rodi, egalik qo'shimchasi va padejlariga ko'ra ajratilgan?

- a) AMPAR
- b) FRAP
- c) Prompt
- d) NERPA

5. Mashina tarjimasining analiz jarayoni necha bosqichda amalga oshiriladi?

- a) 3 bosqich: tokenizatsiya, lemmatizatsiya, stemmizatsiya
- b) 2 bosqich: lemmatizatsiya, stemmizatsiya
- c) 1 bosqich: stemmizatsiya
- d) 4 bosqich: tokenizatsiya, lemmatizatsiya, stemmizatsiya,

integralizatsiya

6. Konteksdan kelib chiqib tarjima qilish-...

- a) leksikotsentrik
- b) slovotsentrik
- c) lokal
- d) integral

7. Sintaksisga asoslangan mashina tarjimasi bo'yicha izlanish olib borgan olimlar to'g'ri ko'rsatilgan qatorni toping.

- a) Vasilevskiy, Marchuk, Narinyani, Shuklin
- b) Vasilevskiy, Chomskiy, Marchuk, Palm
- c) Narinyani, Bekery, Palm
- d) Mialy, Bekery, Vasilevskiy, Ranganatan

8. Avtomatik tarjimada pre-editing qanday vazifani bajaradi?

a) inson matnni mashina-tarjimonga moslaydi, so'ng uni kompyuterga havola etadi.

b) matn (mashina-tarjimon) kompyuter yordamida tarjima qiladi

c) tarjimada murakkablik tug'duruvchi holatlar duch kelganda inson mashina-tarjimon ishiga aralashadi

d) inson-muharrir matnning tarjimasida zarur bo'ladi

grammatik boshqaruv usuli bilan taqsimlash metodi va bilingval lug'atni o'z ichiga oluvchi tizimning bosqichlarini toping.

4. Turkiy tillar bo'yicha dastlabki mashina tarjimasi kim boshchiligida amalga oshirilgan?

5. Y.N. Marchukning mashina tarjimasi haqidagi fikrlariga munosabatingizni bildiring.

6. Tilni tizim sifatida o'rganish shartli ekanligi to'g'risidagi N.Z. Kotelovanning mulohazasiga qanday qaraysiz?

7. Mashina tarjimasining rivojlanishiga ta'sir ko'rsatgan omillar qaysilar?

8. Ruscha matnlarni ingliz tiliga tarjima qiluvchi Retrans Vista versiyasining imkoniyatlari haqida fikr yuriting.

9. M.Hakimov ishlarida kengaytirilgan matematik modelga asoslangan mashina tarjimasi texnologiyasi bo'yicha amalga oshirilgan tadqiqotlar haqida misollar keltiring?

10. Mashina tarjimasining turlarini aniqlang va izohlang.

11. Internetda quyidagi <http://dic.academic.ru>, <http://sobolev-an.narod.ru>, <http://www1.oup.co.uk/oup/elt/software>, <http://www.lingvo.ru> saytlardan kerakli ma'lumotlarni daftaringizga qayd qiling. Ushbu serverdagi mavjud tarjimon dasturlar yordamida tarjima sifati va tarjima qilish imkoniyatini tavsiflang.

9-mavzu yuzasidan test

1. Mashina tarjimasidan foydalangan holda multilingval internet tarmoqlari uchun qanday fransuz – nemis loyihasi ishlab chiqilgan?

- e) Quaero
- f) POLO
- g) ANPAR
- h) Qualcom

2. Mashina tarjimasida metatil tilga oid tushunchalarning modeli kim tomonidan yaratilgan?

- a) I.A. Melchuk
- b) D.Panov
- c) N.Andreyev
- d) K.Bektayev

3. Mashina tarjimasida so'zlarni grammatik shakllarga ajratish

jarayoni qanday nomlanadi?

- a) tokenizatsiya
- b) asper
- c) lemmatizatsiya
- d) stemmizatsiya

4. Qaysi mashina tarjimasida so'zlarning asosi va sintaktik funksiyaga ko'ra leksemalarning rodi, egalik qo'shimchasi va padejlariga ko'ra ajratilgan?

- a) AMPAR
- b) FRAP
- c) Prompt
- d) NERPA

5. Mashina tarjimasining analiz jarayoni necha bosqichda amalga oshiriladi?

- a) 3 bosqich: tokenizatsiya, lemmatizatsiya, stemmizatsiya
- b) 2 bosqich: lemmatizatsiya, stemmizatsiya
- c) 1 bosqich: stemmizatsiya
- d) 4 bosqich: tokenizatsiya, lemmatizatsiya, stemmizatsiya,

integralizatsiya

6. Konteksdan kelib chiqib tarjima qilish...

- a) leksikotsentrik
- b) slovotsentrik
- c) lokal
- d) integral

7. Sintaksisga asoslangan mashina tarjimasi bo'yicha izlanish olib borgan olimlar to'g'ri ko'rsatilgan qatorni toping.

- a) Vasilevskiy, Marchuk, Narinyani, Shuklin
- b) Vasilevskiy, Chomskiy, Marchuk, Palm
- c) Narinyani, Bekery, Palm
- d) Mialy, Bekery, Vasilevskiy, Ranganatan

8. Avtomatik tarjimada pre-editing qanday vazifani bajaradi?

a) inson matnni mashina-tarjimonga moslaydi, so'ng uni kompyuterga havola etadi.

b) matn (mashina-tarjimon) kompyuter yordamida tarjima qiladi

c) tarjimada murakkablik tug'duruvchi holatlar duch kelganda inson mashina-tarjimon ishiga aralashadi

d) inson-muharrir matnning tarjimasida zarur bo'ladi

9. Turkologiyada dastlab kimning rahbarligida mashina tarjimasiga doir tajriba o'tkazilgan?

- a) Dorster
- b) Güneyd Tantuğ,
- c) Aşref Adalı,
- d) Kemal Oflazer

10. Kemal Oflazer mashina tarjimasiga oid qanday tadqiqotlar olib borgan?

a) ingliz tilidan turk tiliga tarjima qilish bo'yicha statistik mashina tarjimasini yaratgan

b) turkman tilidan turk tiliga asoslangan mashina tarjimasini bo'yicha tadqiqotlar olib borilgan

c) so'z birikmasini to'g'ri tarjima qilishga oid tadqiqotlar

d) fransuz tilidan turk tiliga tarjima qilish bo'yicha tadqiqotlar

10. Wordlingvoning qanday imkoniyatlari mavjud?

a) unda tarjima texnologiyasining veb-sahifalar tarjimasini mavjud

b) matn tarjimasini, elektron pochta uchun tarjima va chat tarjima qilish imkoniyati bor

c) 15 ta tilni o'z ichiga oladi

d) barcha javoblar to'g'ri

10-MAVZU. MASHINA TARJIMASINING LINGVISTIK MUAMMOLARI

10.1. Tarjima dasturi uchun lingvistik ta'minot va tahlil tamoyillari

10.2. Omonimiya va polisemiya lingvistik masalalar sifatida

10.3. Mashina tarjimasini pragmatik va semantik muammolar

10.4. Tarjima algoritmi.

Tayanch so'z va iboralar: *terminologiya, matnlarni tahlil etish, tokenizatsiya (matndan so'zshakllarni ajratib olish), lemmatizatsiya (so'zlarni morfem tahlil qilish), stemmizatsiya, omonimiya, polisemiya, algoritm, ma'lumotlar bazasi, matematik modellashtirish*

10.1. Tarjima dasturi uchun lingvistik ta'minot va tahlil tamoyillari

Olimlarning konsept va terminlarga nisbatan yondashuvlari turlicha. "Tilshunoslik maktablarida lingvistik analizning konsept paradigmasi "gap-clause" hisoblanadi. Biroq tobelik nazariyasi (Agel, Fischer va Hudsonlar)da so'z grammatik analizning asosi sanaladi. Funksional diskurs grammatikada esa analizning (Hengeveld, Mackenzie) gapdan-da muhim asosi sifatida *diskurs* akti belgilanadi. Ba'zi olimlarning fikriga ko'ra analizning asosi konstruksiyalardir. Ba'zi til grammatikasini konstruksiyalarning jamlanmasi sifatida talqin qiluvchi yondashuvlarda analizning asosi model deb qaraladi. Korpusga asoslangan tadqiqotlarda (Beber) na grammatik sinflar va na sintaktik strukturalarni yuqori mavqega ega emas, deb hisoblaydilar.

Ayrim yondashuvlarda fokus sifatida subyekt va obyekt olinib, unda argumentlarning munosabatlari tasvirlanadi¹⁸⁹.

Matn biror millatga tegishli xalqning adabiy tiligina emas, balki uning madaniyati va ma'naviyati aks etuvchi lug'at boyligidir. Avtomatlashtirilgan leksikografiyaning o'zi (so'z va so'z birikmasi) matnni tarjima qilib berish uchun yetarli bo'lmaydi. Tillar tipologik

¹⁸⁹ Bern Heine, Heiko Narrog. The Oxford handbook of linguistic analysis. UK: Oxford university, 2015 –P. 17-18

jihatdan farqlanadi. Lingvistik ta'minot tuzishda o'rganilgan va amalda sinalgan tadqiqotlardan foydalanish maqsadga muvofiq.

Mashina tarjimasi, asosan, axborot (ilmiy, rasmiy, publitsistik) uslubi uchun xosdir. Tarjima modellarini tuzishda ana shu uslubdagi me'yoriy talablar hisobga olinadi. Tarjimaning uch asosiy komponenti mavjud, bular: asos matn, unga qo'yilgan masalalarning to'g'riligi hamda terminologiya masalalaridir. Matnni tarjima qilishda quyidagi masalalar o'rganiladi: tahlil jarayoni, so'zlarning matndagi ma'nolarini tushunish jarayoni hamda matnni tiklash jarayoni. Mashina tarjimasida morfologik va sintaktik tahlil muhim ahamiyatga ega. Morfologik analizda kategoriyalarni tasniflashda nimani emas, balki ularni qanday qilib tasniflash nazarda tutiladi. Xususan, turkcha-ruscha mashina tarjimasi ta'minoti quyidagi qismlardan: so'zshakllari so'zligi, lingvistik oborotlardan iborat bo'lgan avtomatik lug'at hamda ot va fe'l so'z turkumiga kiruvchi so'zlarning asosi, qo'shimchalaridan iborat. Turkcha matnlar uchun qo'llanilgan bunday lug'at S. Muhamedovning kvantativ modeli tamoyillari asosida yaratilgan¹⁹⁰.

Ma'lumki, istalgan mustaqil ma'noli leksema iyerarxik semantik maydonlarga ega. Bu semantik maydonlar o'rtasidagi munosabat kontekst orqali voqelanadi. Morfologik tahlilning asosini morfologik va tarjima baza, grammatik qoidalar va istisnoli jadvalga kiritilgan belgilardan tashkil topgan:

- ID->English_Suz
- ID->TranslateUz_Noun
- ID->TranslateUz_Verb
- ID->TranslateUz_Adjective
- ID->TranslateUz_terminology
- ID->TranslateUz_combinatory

O'zbek tili va ingliz tilining morfologik bazasini tuzishda mezonlar aniqlab olinadi. Bunda fe'l so'z turkumining o'zak shakli: *o'qimoq* - <o'qi>, *ochmoq* - <och> kiritiladi. Ushbu holatda ayrim fe'llar boshqa so'z turkumlari bilan omonimlik hosil qilishi mumkin. Shu bois bazadagi ma'lumotlarni dasturiy ta'minotda aniqlay olishi uchun quyidagi belgilar kiritiladi:

¹⁹⁰ Мухамедов С.А., Пиотровский Р.Г. Инженерная лингвистика и опыт системно-статического исследования узбекских текстов. – Ташкент: Фан, 1986. – С. 141-142.

H_W– agar muayyan so'z faqat bir so'z turkum doirasida omonimlik hosil qilsa; H_L– agar muayyan so'z ikkidan ortiq so'z turkumi doirasida omonimlik hosil qilsa; H_Adj. {H_Adv.}– agar ikki so'z turkumi doirasida bo'lsa, ikkinchi so'z turkum ko'rsatiladi.

Ayrim tadqiqotlarda morfologik tahlilni optimallashtirishga oid asosli fikrlar o'rtaga tashlangan¹⁹¹. Matnlarni avtomatik tarzda morfologik tahlil qilish uchun tokenizatsiya, lemmatizatsiya, stemming usullaridan foydalaniladi. Tokenizatsiyalash jarayonida tabiiy tildagi nutq birliklari (token, so'zshakllari) alohida tarzda ularning grammatik ma'nolariga ajratiladi. Lemmatizatsiya bosqichida esa so'zshakllarining dastlabki holati aniqlanadi. Bunda morfologik analizator orqali so'zlarning asos va affikslarga ko'ra morfemalarga ajratiladi. Stemming jarayonida so'zning o'zagi topiladi (Porter algoritmini misol sifatida keltirish mumkin¹⁹²). Quyidagi sxemada ushbu bosqichlar ko'rsatilgan:

¹⁹¹ Qarang: Belonogov G. G. On the use of a similarity method for automatic processing of textual information [Об использовании метода аналогии при автоматическом обработке текстовой информации] Problems of cybernetics [Problemy kibernetiki]. Issue 28. –Moscow: Nauka Publ., 1974; Bolshakov I. A. CrossLexica: A large electronic dictionary of collocations and semantic links between Russian words [КроссЛексика — большой электронный словарь сочетаний и смысловых связей русских слов]. Компьютерная лингвистика и Интеллектуальный Технологии: Труды Международной Конференции “Диалог 2009” [Computational Linguistics and Intelligent Technologies: Proceedings of the International Conference “Dialogue 2009”]. –Moscow, 2009, pp. 45–50.; Gelbukh A. F. Effectively realizable morphologic model of inflective language [Эффективно реализуемая модель морфологии и активного языка]. Научно-техническая Информация [Scientific and Technical Information], series 2, #1, 1992, pp. 24–31.; Sokirko A. V. Bystroslavar': morphological prediction of new Russian words using very large corpora [Быстрословарь: предсказание морфологии русских слов с использованием больших лингвистических ресурсов]. Компьютерная лингвистика и Интеллектуальный Технологии: Труды Международной Конференции “Диалог 2010” [Computational Linguistics and Intelligent Technologies: Proceedings of the International Conference “Dialogue 2010”]. –Moscow, 2010. –P. 450–45; Zaliznjak A. A. Grammatical dictionary of Russian: Инф эсион [Грамматический словарь русского языка: Словоизменение]. –Moscow: Русские Язык Публ., 1977.

jihattan farqlanadi. Lingvistik ta'minot tuzishda o'rganilgan va amalda sinalgan tadqiqotlardan foydalanish maqsadga muvofiq.

Mashina tarjimasi, asosan, axborot (ilmiy, rasmiy, publitsistik) uslubi uchun xosdir. Tarjima modellarini tuzishda ana shu uslubdagi me'yoriy talablar hisobga olinadi. Tarjimaning uch asosiy komponenti mavjud, bular: asos matn, unga qo'yilgan masalalarning to'g'riligi hamda terminologiya masalalaridir. Matnni tarjima qilishda quyidagi masalalar o'rganiladi: tahlil jarayoni, so'zlarning matndagi ma'nolarini tushunish jarayoni hamda matnni tiklash jarayoni. Mashina tarjimasida morfologik va sintaktik tahlil muhim ahamiyatga ega. Morfologik analizda kategoriyalarni tasniflashda nimani emas, balki ularni qanday qilib tasniflash nazarda tutiladi. Xususan, turkcha-ruscha mashina tarjimasi ta'minoti quyidagi qismlardan: so'zshakllari so'zligi, lingvistik oborotlardan iborat bo'lgan avtomatik lug'at hamda ot va fe'l so'z turkumiga kiruvchi so'zlarning asosi, qo'shimchalaridan iborat. Turkcha matnlar uchun qo'llanilgan bunday lug'at S. Muhamedovning kvantativ modeli tamoyillari asosida yaratilgan¹⁹⁰.

Ma'lumki, istalgan mustaqil ma'noli leksema iyerarxik semantik maydonlarga ega. Bu semantik maydonlar o'rtasidagi munosabat kontekst orqali voqelanadi. Morfologik tahlilning asosini morfologik va tarjima baza, grammatik qoidalar va istisnoli jadvalga kiritilgan belgilardan tashkil topgan:

- ID->English_Suz
- ID->TranslateUz_Noun
- ID->TranslateUz_Verb
- ID->TranslateUz_Adjective
- ID->TranslateUz_terminology
- ID->TranslateUz_combinatory

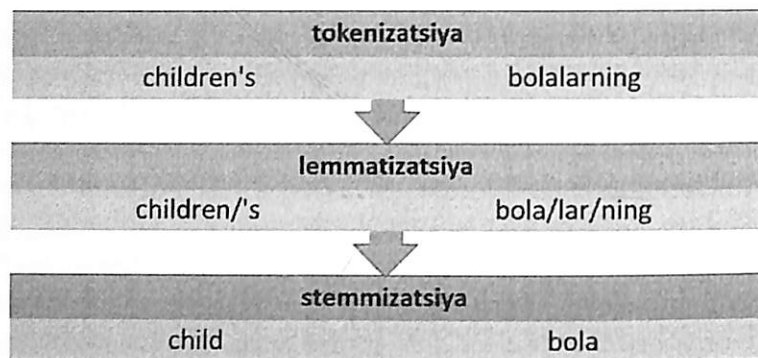
O'zbek tili va ingliz tilining morfologik bazasini tuzishda mezonlar aniqlab olinadi. Bunda fe'l so'z turkumining o'zak shakli: *o'qimoq* - <o'qi>, *ochmoq* - <och> kiritiladi. Ushbu holatda ayrim fe'llar boshqa so'z turkumlari bilan omonimlik hosil qilishi mumkin. Shu bois bazadagi ma'lumotlarni dasturiy ta'minotda aniqlay olishi uchun quyidagi belgilar kiritiladi:

¹⁹⁰ Мухамедов С.А., Пиотровский Р.Г. Инженерная лингвистика и опыт системно-статического исследования узбекских текстов. – Ташкент: Фан, 1986. – С. 141-142.

H_W– agar muayyan so'z faqat bir so'z turkum doirasida omonimlik hosil qilsa; H_L– agar muayyan so'z ikkidan ortiq so'z turkumi doirasida omonimlik hosil qilsa; H_Adj. {H_Adv.}– agar ikki so'z turkumi doirasida bo'lsa, ikkinchi so'z turkum ko'rsatiladi.

Ayrim tadqiqotlarda morfologik tahlilni optimallashtirishga oid asosli fikrlar o'rtaga tashlangan¹⁹¹. Matnlarni avtomatik tarzda morfologik tahlil qilish uchun tokenizatsiya, lemmatizatsiya, stemming usullaridan foydalaniladi. Tokenizatsiyalash jarayonida tabiiy tildagi nutq birliklari (token, so'zshakllari) alohida tarzda ularning grammatik ma'nolariga ajratiladi. Lemmatizatsiya bosqichida esa so'zshakllarining dastlabki holati aniqlanadi. Bunda morfologik analizator orqali so'zlarning asos va affikslarga ko'ra morfemalarga ajratiladi. Stemming jarayonida so'zning o'zagi topiladi (Porter algoritmini misol sifatida keltirish mumkin¹⁹²). Quyidagi sxemada ushbu bosqichlar ko'rsatilgan:

¹⁹¹ Qarang: Belonogov G. G. On the use of a similarity method for automatic processing of textual information [Об использовании метода аналогии при автоматическом обработке текстовой информации] Problems of cybernetics [Problemy kibernetiki]. Issue 28. –Moscow: Nauka Publ., 1974; Bolshakov I. A. CrossLexica: A large electronic dictionary of collocations and semantic links between Russian words [КроссЛексика — большой электронный словарь сочетаний и смысловых связей русских слов]. Компьютерная лингвистика и Интеллектуальный Технологии: Труды Международной Конференции “Диалог 2009” [Computational Linguistics and Intelligent Technologies: Proceedings of the International Conference “Dialogue 2009”]. –Moscow, 2009, pp. 45–50.; Gelbukh A. F. Effectively realizable morphologic model of inflective language [Эффективно реализуемая модель морфологии и активного языка]. Научно-техническая Информация [Scientific and Technical Information], series 2, #1, 1992, pp. 24–31.; Sokirko A. V. Bystroslovar': morphological prediction of new Russian words using very large corpora [Быстрого словаря: предсказание новых русских слов с использованием больших лингвистических ресурсов]. морфологии русских слов с использованием больших лингвистических ресурсов]. Компьютерная лингвистика и Интеллектуальный Технологии: Труды Международной Конференции “Dialog 2010” [Computational Linguistics and Intelligent Technologies: Proceedings of the International Conference “Dialogue 2010”]. –Moscow, 2010. –P. 450–45; Zaliznjak A. A. Grammatical dictionary of Russian: Инф-эстимон [Грамматический словарь русского языка: Словоизменение]. –Moscow: Русские Язык Публ., 1977.



Piton dasturlash tili orqali muayyan matn misolida ular avval lug'at shaklga keltirib, keyin token va lemmalarga ajratish algoritmini ko'ramiz:

```
import io
import GetText
dictit=GetText.getLilst('WordBase.txt')
NoRepeat=list(set(dictit))
for i in NoRepeat:
    print(i)
```

Bunda barcha so'zlar takrorlanmas shaklda lug'at sifatida to'plamga yig'iladi.

```
def getRoot(word):
    for i in range(0, n - 1):
        if (wordBase.__contains__(word[:i])):
            return word[:i]
```

O'zak va qo'shimchalar alohida morfologik jihatdan analiz qilinadi.

Matnlarni tahlil etishning bu kabi usullari ba'zi adabiyotlarda kompyuter ma'lumotlar bazasi –korpus yaratishning maxsus protseduralari hamda dasturlari sifatida tilga olinadi¹⁹³. Inglizcha matnlarni o'zbek tiliga tarjima qilishda, avvalo, morfoklassifikator kerak bo'ladi. Xususan, ingliz tilida sifat va ravish so'z turkumlariga oid so'zlarni morfologik jihatdan tasniflash va ularning o'zbekcha variantlari hisobga olinadi.

¹⁹³ Захаров В.П., Богданова С.Ю. Корпусная лингвистика: учебник для студентов гуманитарных вузов. – Иркутск: ИГЛУ, 2011. – С. 39.

Ingliz va o'zbek tillari boshqa til oilalariga mansub bo'lganligi uchun ham morfologik kategoriyalarning bir-biriga o'xshamasligi, tabiiy. Masalan, ingliz tilidagi ba'zi olmoshlar o'zbek tilida mavjud emas. Ular boshqa kategoriyalar sifatida qaraladi: *few*, *little* olmoshlari o'zbek tilida *kam*, *oz* so'zlariga mos bo'lib, ravish so'z turkumiga tegishli. Bularning yechimi tegishli lug'atlarga kiritilgan so'zlar o'zaro kodlanadi. Tarjimaga qo'yiladigan asosiy talablardan biri matnning shakli va mazmunini saqlab qolishdir. Shu kabi muammolarni hal qilishda transformatsion metod samarador usullardan biridir. Transformatsion usulda to'rt jarayon voqelanishi mumkin: *o'rniga qo'yish*, *almashtirish*, *qo'shish*, *tushurib qoldirish*¹⁹⁴. Shuningdek, transformatsion usulning adyunsiya, substitutsiya turlari ham mavjud. Unga ko'ra, transformatsiya protsessi asosini yadro gaplar deb ataluvchi konstruksiyalar (kernel structures) tashkil qiladi. Bu konstruksiyalar eng kichik yoki sintaksisdagi sodda gaplardan iborat¹⁹⁵. Bunday asosiy bosqichlar analiz va sintez jarayonida o'z ifodasini topadi.

O'rniga qo'yish usuli matndagi so'z va so'z birikmalarini morfologik jihatdan o'rinlashtiradi.

1	2	3	4
I	go	to	school
1	4	3	2
Men	maktab	-ga	boraman

Almashtirish ikki hodisa orqali amalga oshiriladi: 1) *konkretizatsiya*; 2) *generalizatsiya*. Bu ikki hodisa ham morfologik va sintaktik jihatdan yuzaga keladi. Xususan, *and* so'zi biriktiruv bog'lovchi vazifasidan (*va, hamda*) tashqari quyidagi ma'nolarda keladi: *bo'lsa, esa, biroq* – *I shall go and you stay here* – *Men ketaman, sen esa shu yerda qol*. Ayrim fe'lli birikmalar tarkibida kelib, o'zbek tilida fe'ning harakat nomiga mos kelishi mumkin: *try and do it* – buni qilishga harakat qiling, *come and see* – ko'rishga

¹⁹⁴ Бархударов Л. С. Язык и перевод (Вопросы общей и частной теории перевода). – М.: Междунар. отношения, 1975. – С.190-191.

¹⁹⁵ Бўронов Ж.Б. Инглиз ва ўзбек тиллари қиёсий грамматикаси. –Т.: Ўқитувчи, 1973. – Б. 40.

keling, *wait and see* –yashasak ko‘ramiz¹⁹⁶ kabi. *Generalizatsiya* hodisasi ko‘proq sintaktik birliklarda kuzatiladi. *So‘zlarni qo‘shish* usulida ba‘zi morfologik kategoriyalar matnga qo‘shimcha ravishda qo‘shiladi: *I have a book –Menda kitob bor*. Bunda fe‘l va otning o‘zaro munosabati semantik aspektida ko‘rinadi. Inson tomonidan tarjima qilinganda, bu tabiiy holat, biroq mashina tarjimasida bu alohida jarayon. *Tushirib qoldirish* hodisasida tarjima jarayonida u yoki bu so‘z tushirib qoldiriladi: *those–ular, these–bular* ko‘rsatish olmoshlari lug‘atda ko‘plikdagi narasaning predmetlarga nisbatan qo‘llanadi, ammo ular ot bilan kelganda *-lar* tushirib qoldiriladi: *These are mine – Bular meniki ⇔ These children are mine –Bu bolalar meniki*.

Shu jihatdan lingvistik ta‘minotda morfologik tasnif jiddiy masala hisoblanadi. Shu o‘rinda, rus tilidagi *Krossleksika* elektron lug‘atidagi morfologik klassifikatorga 115 guruhga ajratilgan turlanuvchi so‘zlarning tasnifi kiritilgan¹⁹⁷. Morfokvalissifikator lug‘atda mavjud bo‘lmagan so‘zshakllarini ham tuzish imkoniyatiga ega. Lingvistik translyator uch bosqichli tizimdan: analizator va sintezatoridan tarkib topgan *lingvistik protsess*, grammatika va semantika haqidagi bilimlardan tashkil topgan *lingvistik model*, o‘zaro bog‘langan deklarativ va protsedurali qismlarning lingvistik translyatsiyasini aks ettiruvchi *assotsiasiv protsedura*¹⁹⁸ dan iborat. Ko‘plab mashina tarjimasini tizimlarining moduli *analiz->transfer->sintez* bosqichida voqelanishini kuzatamiz. Matnlarni morfologik tahlil qilishda analizator lingvistik jihatdan ta‘minlanadi. Dasturiy ta‘minotda grammatik ta‘minot ishlab chiqilgan: “terminlarning grammatik guruhga kiruvchi so‘zlar (artikl+ot+predlog) hamda terminlarning funksional elementalaridan (ega+kesim) chiziqli tuzilmasi yoki elementlarning ketma-ketligidan tashkil topgan *zanjirli grammatika*. Unda u yoki bu elementlarning o‘zidan oldingi yoki o‘zidan keyingi ketma-ketligi beriladi. Shuning uchun bu kesim

¹⁹⁶ ABBY Lingvo×5

¹⁹⁷ Болшаков И. А., Болшакова Е. И. Автоматический морфоклассификатор русских именных групп. Компьютерная лингвистика и интеллектуальные технологии. По материалам ежегодной Международной конференции «Диалог» (2012) Выпуск 11. – С. 81.

¹⁹⁸ Марчук Ю.Н. Компьютерная лингвистика – М.: Восток, 2006. –С. 272.

oldi sintaktik analiz deb yuritiladi. Keyingisi *tashkil etuvchi grammatika* bo‘lib, unda grammatik element guruhlari haqida ma‘lumot beriladi, masalan, ismlar guruhi (ot, artikl, sifat v.b.dan tashkil topuvchi guruh), predlogli guruh (predlog va boshqa so‘z (ism)lardan tashkil topgan guruh) hamda gap darajalarigacha bo‘lgan elementlar. Shuningdek, *tobelanishga asoslangan grammatika* bo‘lib, unda har bir element o‘zaro tobe munosabatlarga bo‘ysunadi. Tahlil strategiyasi yuqoridan pastga tomon bo‘lib (top-down) gapning markazi fe‘l, ya‘ni kesim aniqlanadi. *Kontekstli erkin grammatikada* transformatsion qoidalarga amal qilinadi. *Unifiksion grammatika* to‘rt komponent (tarkib): unifikatsiya paketi, leksik tavsif va qoidalar uchun interpretator, yo‘naltirilgan graflarni qayta ishlash dasturi hamda graf-sxemalar yordamidagi analizatoridan iborat bo‘ladi. Unifikatsion grammatika semantik valentlik bilan sintaktik valentlik, lug‘atga oid tavsif bilan grammatik qoidalarni aniqlashtiradi¹⁹⁹.

Kompyuter lingvistikasida mashina tarjimasini eng murakkab yo‘nalishlardan biri hisoblanadi. Negaki tabiiy tillar ustida tadqiqot olib borilar ekan, bunda lingvistikaning boshqa psixologik, kognitiv, konseptual hamda lingvokulturologik jihatlari ham inobatga olinishi talab etiladi. Tarjima texnik jarayongina emas, u ijodiy faoliyat bilan bog‘liq bo‘lgan ruhiy va ilmiy asoslangan murakkab yaratish bosqichidir. So‘zlashuv va badiiy uslubdagi matnlar ekstralingvistik omillarning anglashilayotgan ma‘noga va mazmun yaxlitligiga ta‘sir etishi bois, mashina tarjimasining imkoniyatlari yetarli bo‘lmaydi.

Shunga qaramay, o‘tgan asrning 50-yillarida boshlangan mashina tarjimasiga oid tadqiqotlarda barcha uslublarning lingvistik jihatlari qamrab oluvchi sun‘iy intellektning yangi neyro tizimlari, qolaversa, nafaqat yozma, balki og‘zaki matnlarni ham tarjima qiluvchi tizimlarini yaratish borasida ulkan yutuqlarga erishildi. Fan va texnikaning rivojlanishi bugungi globallashuv jarayonida Internet tarmog‘idagi mislsiz axborot oqimini qayta ishlash, ularga tanqidiy yondashish, yangiliklarni o‘rganish va dunyoga yangi axborotni

¹⁹⁹ Мамедова М.Г., Мамедова З.Ю. Машинный перевод: эволюция и основные аспекты моделирования. –Баку: ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ АКАДЕМИИ НАУК АЗЕРБАЙДЖАНА, 2005. – С. 69-72.

uzatish uchun mashina tarjimasidan optimal yechimga ega bo'lgan tizimlarni yaratishni taqozo etmoqda.

Endilikda mashina tarjimasida uchun ochiq platformali lingvistik resurslardan foydalanish mumkin. Masalan, <http://opus.nlpl.eu> saytida quyidagi resurslar, interfeyslar va parallel matnlar bazasi mavjud. *Tanzil* deb nomlangan resursda Qur'on tavsirlarining parallel matnlari kiritilgan bo'lib, bunda o'zbekcha tavsirning parallel matnini ko'rishimiz mumkin.

10.2. Mashina tarjimasida omonimiya va polisemiya lingvistik masalalar sifatida

Omonimiya — shakli va talaffuzi bir xil ma'nosi har xil bo'lgan so'zlar. Omonimlar to'liq yoki qisman bo'lishi mumkin: 1) omoforma; 2) omograf; 3) omofon.

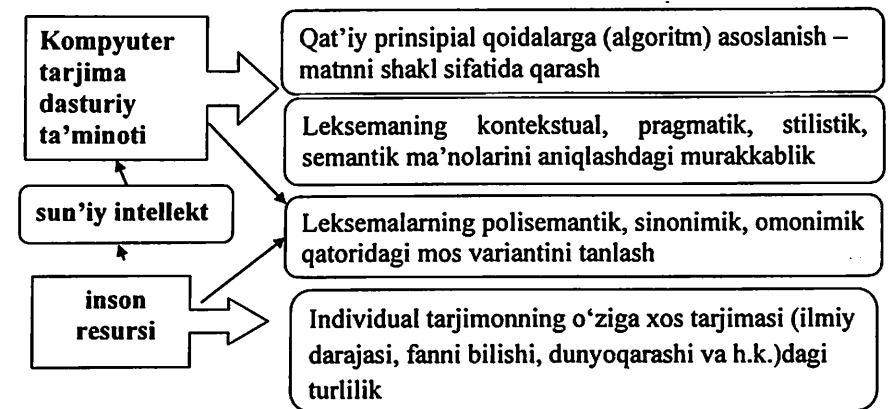
Omoforma – ayrim leksemalarning omonim bo'lib, faqat u yoki bu grammatik shakllaridagina teng kelishi: *saw (see-ko'rmoq)-saw (arra), olma (ot)-olma (fe'l)*. Agar fe'lning bo'lishsizlik shakli olinsa, *olma* so'ziga nisbatan omonimik munosabat yo'qoladi.

Omograf – yozilishi o'xshash, biroq urg'u va talaffuziga ko'ra farqlanuvchi so'zlarga aytiladi: *tom (uyning ustki qismi)-tom (kitobning yirik qismi), akademik (urg'u uchinchi bo'g'inga tushsa, shaxs) – akademik (oxirgi bo'g'inga tushsa, sifat)*.

Omofon – talaffuzi o'xshash, yozilishida farqlanuvchi so'zlarga aytiladi: *ziyrak-zirak, bop- bop kabi*.

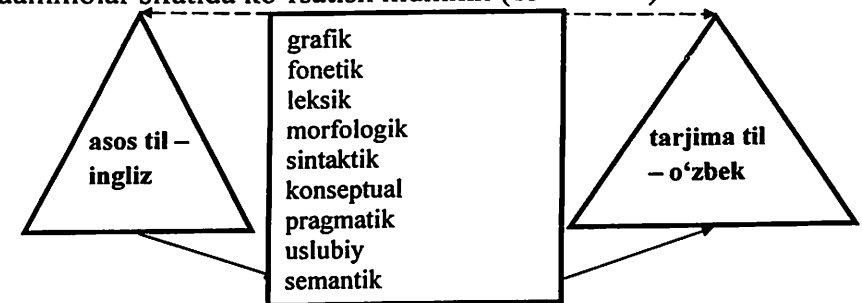
Omonimiya qo'shimchalar, leksemalar, so'z birikmalar va gaplar o'rtasida kuzatiladi. Masalan, sintaksisda buni ajratish uchun so'z turkumini teglashda muammolar yuzaga keladi: *Beshta bolali ayol keldi*. Ushbu gapni tahlil qilganimizda *beshta bolali- ayol keldi* > *beshta-bolali ayol keldi*.

Mashina tarjimasining tadqiqot yo'li yarim asrdan oshgan bo'lsa-da, ba'zi bir muammolarni yechishda hanuz ojizlik qilmoqda. *Til-> vaziyat-> nutq* tabiiy tilni qayta ishlashda ba'zi muammolarni keltirib chiqargan (12-chizma).



12-chizma

Faylasuf Bar Xillel quyidagilarni mashina tarjimasining dastlabki rivojlanish bosqichlaridagi asosiy muammolar sifatida ko'rsatgan²⁰⁰: 1) operatsion sintaksis (Operational Syntax); 2) tabiiy tillarning ichki tarjimasida (Intertranslatability of natural languages); 3) idiomalar (Idioms); 4) universal sintaktik kategoriyalar (Universal syntactic categories). Quyidagilarni mashina tarjimasidagi umumiy muammolar sifatida ko'rsatish mumkin (13-chizma):



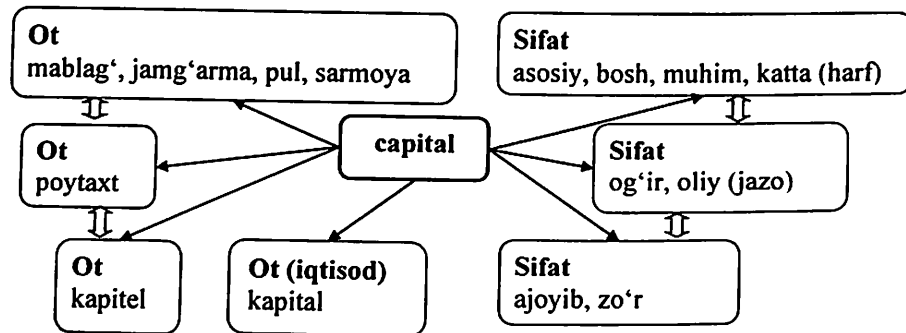
13-chizma

Ma'lumki, ingliz va o'zbek tillari tipologik jihatdan boshqa til oilalarga mansubligi bois, mashina tarjimasida qator muammolarni yuzaga keltiradi: 1) inglizcha matnlardagi leksemalarning polisemik va omonimik xususiyatga egaligi; 2) affikslarning turli funkcionalligi; 3) ba'zi bir frazemalarning lingvokulturologik

²⁰⁰ <http://www.mt-archive.info/Bar-Hillel-1953.pdf> [From: *Philosophy of Science*. Vol.20 (1953). – P.217 – 225

xususiyati; 4) matn tarkibidagi leksemalarning sinonimik qatordan tarkib topishi va ularning uslubiy tomondan bog'liqligi; 5) ilmiy matnlardagi terminlar masalasi va h.k.

Tanlab olingan va cheklangan leksemalarning lingvistik modulni berishda mana shu vazifalardan kelib chiqish maqsadga muvofiq. Masalan, *Capital* so'zi ham boshqa so'z turkumini, ham omonimiyani hosil qilgan (14-chizma):



14-chizma

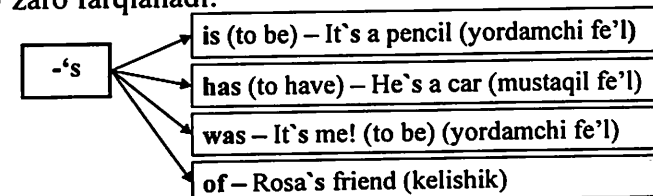
Birdetalning qayd etishicha(1987), Vebster muallifligidagi «Ingliz tilining izohli lug'ati» (7-nashri)dagi 60000 so'zning 21488 tasi, ya'ni umumiy miqdorning 40% ni ikki va undan ortiq ma'noga ega polisemantik so'zlar tashkil etadi. Masalan, birgina *run* – *yugurmoq* so'zi 29 ma'noga ega bo'lib, 125 ta ichki semantik maydonlarga ega²⁰¹. Leksikografiyada lug'at tuzuvchilarning polisemiyaga bo'lgan qarashlari, yondashuvlari turlicha bo'lgani bois omonimiya hodisasi bilan polisemiyani farqlashda ayrim chalkashliklar kuzatiladi. Ingliz tilida polisemiya hodisasi nisbatan yaxshi o'rganilgan. Biroq avtomatik tarjima uchun tanlangan va chegaralab olingan so'zlarning to'liq lingvistik ta'minotida ularning barchasi o'rin olishi hamda o'zbek tilidagi tarjimasida xususiyatlari ochib berilishi kerak.

XX asrning 60-yillarida Bar Xillel namuna sifatida *pen* so'zining «*ruchka*» va «*o'yinchoq qutisi*» ma'nolari ega bo'lishiga tayanib, *mashina tarjimasini 80 % xato*, degan fikrni bildirgan. Endilikda olib borilgan izlanishlar natijasida matnni avtomatik

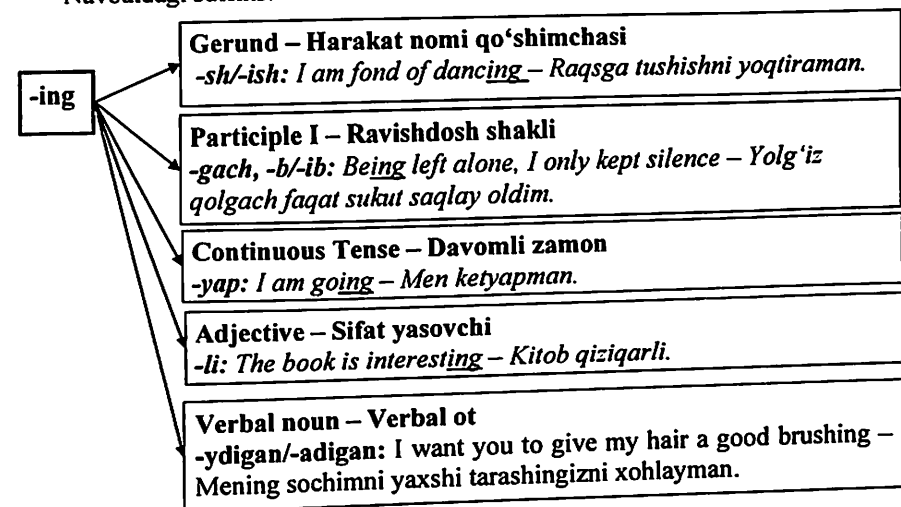
²⁰¹ Yael Ravin, Claudia Leacock Polysemy. Theoretical and computational approaches. Oxford University press. 2000. – P.13.

tarjima qilishning 80% muammosi yechilgan bo'lib, qolgan qismi mutaxassislar tomonidan o'rganilmoqda. Mazkur tadqiqotlar sirasida Lesk (1986), Gutri (1991), Vilks (1993), Blek (1988), Zernik (1991), Xerst (1991), Gale (1992), Miller, Lekok (1998), Vurxis (1998), Adam Kilgarrif (1997) va boshqa olimlarning ishlarini eslatish o'rinli²⁰².

Masalan, ingliz tilidagi -s suffiksining 's shakli ham bo'lib, ular o'zaro farqlanadi:



Navbatdagi suffiks:



So'zlardagi -ing suffiksi sifatdosh vazifasida kelganda, o'zbek tilidagi -(i)sh va -gan qo'shimchalariga mos keladi.

Ingliz tilida sintaktik omonimiya ko'p uchraydi. Asosan, sintaktik omonimiya muayyan kategoriyalar bilan haam bog'liq: gerund, predlod va h.k.

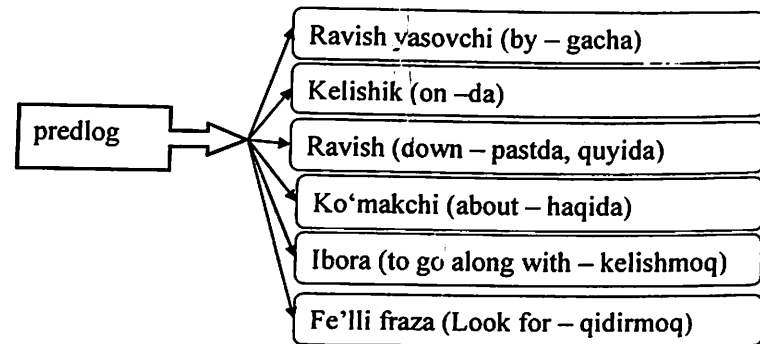
²⁰² O'sha joyda.

Masalan, *I saw the man with binocular.* -1) Men binoklli kishini ko'rdim; 2) Men binokl bilan kishini ko'rdim. Ushbu holatda *with* ham "kishi", ham "binokl" so'zlariga tegishli bo'lishi mumkin.

Mashina tarjimasida sinonimiya hodisasi masalasi alohida o'rganiladi. Ayrim elektron lug'atlarda (ABBY LINGVO) sinonimlar qatori quyidagicha tartiblangan: 1) turkumi; 2) sinonimik qatori; 3) izohi; 4) uslubi (badiiy, so'zlashuv, rasmiy, ilmiy) va unga doir jumlar tarzidagi namunalar; 5) darajasi (kuchsiz, kuchli, neytral); 6) qo'llanilishi (ijobiy, neytral, salbiy); 7) sohasi; 8) valentligi; 9) uning ishtirokida berilgan so'z birikmasi yoki matn namunasi.

Ingliz tili morfologik kategoriyalar turlari va xususiyatlarining o'zbek tili morfologik kategoriyalaridan farqlanishini kuzatishimiz mumkin: *predloglar, modal fe'llar, yordamchi fe'llar, gerundiy, otga xos ba'zi bir xususiyatlar: sanaladigan, sanalmaydigan ot; faqat birlikda, faqat ko'plikda qo'llanadigan otlar; olmoshning ba'zi turlari: much / many (ko'p), little / few (oz) so'zlari va h.k. shular jumlasidandir.*

Ingliz tilida predloglar quyidagi vazifalarga ega²⁰³:



Biror fe'l bilan birikib kelgan predlog so'zning mazmuniga ta'sir qiladi, masalan, *for* predlogi bir qancha ma'no ifodalaydi: 1) -ga, uchun (atalganlik) => *This letter is for you (Bu xat sizga yoki siz uchun);* 2) uchun, tarafdor => *I am for your proposal (Men sizning taklifingiz tarafdoriman);* 3) uchun (sabab) => *He was rewarded for*

²⁰³ Qarang: Blokh M.Y. A course in theoretical English Grammar. - M.: Высшая школа, 2008; Кобрин Н.А., Корнеева Е.А., Оссовская М.И., Гузеева К.А. Грамматика английского языка. Морфология. Синтаксис. - СПб.: Союз, 2008

his bravery (U jasurligi uchun mukofatlandi); 4) davomida (ish-harakatning davom etish muddatini ifodalaydi) => *I have been to Andijon for two weeks (Men Andijonda uch hafta bo'ldim);* 5) *for* predlogi bilan keluvchi turg'un birikmalar => *to ask for (so'ramoq), a contract for (uchun shartnoma);* 6) *iboralar* => *for ever (abadiy, mangu).*

10.3. Mashina tarjimasida pragmatik va semantik muammolar

Til hamisha ma'lumot beribgina qolmay, unda muloqot uchun insonlarning hissiy, ruhiy va madaniy borlig'i ham aks etadi. Shu bois biror matnni ikkinchi tilga tarjima qilishda yuqorida qayd qilingan barcha unsurlar inobatga olinishi kerak. Shu jihatdan tarjimaga pragmatik yondashish zarurligini ko'rsatadi.

Kompyuter lingvistikasida pragmatikaning o'rni haqida Bar Xillel, Wilks, Nirenburglar tomonidan fikrlar o'rta tashlangan²⁰⁴. Pragmatik yondashuv matn tarjimasida uchun kishilarning (muallif, tarjimon va auditoriya) hissiy munosabatlari (nafrat, muhabbat, umid va boshqalar)ga asoslanadi.

Shu jihatdan tarjimaga pragmatik yondashishda quyidagi modellar tizimga kiritilganda quyidagilardan iborat bo'ladi²⁰⁵:

- tarjimonning e'tiqodi (fikri)
- tarjimonning muallif haqidagi fikri
- tarjimonning (original tildagi) adresat fikri uchun bildirgan fikriga munosabati

- tarjimonning (tarjima tildagi) adresat fikriga munosabati

Mashina tarjimasida pragmatik muammolar masalasiga Annely Rothkegel²⁰⁶ o'zining "Mashina tarjimasida pragmatika" deb nomlangan maqolasida asosiy muammolardan biri sifatida tarjima ekvivalentlari o'rtasida moslashish hodisasi ekanligini qayd etadi.

²⁰⁴ Bar-Hillel, Y. (1960). The Present Status of Automatic Translation of Languages. In F.L. Alt (ed.) *Advances in Computers*, 1:91-163. New York: Academic Press.; Wilks, Y. (1975). A Preferential Pattern Seeking Semantics for Natural Language Inference. *Artificial Intelligence*, 6:53-74.; Nirenburg, S., Carbonell, J., Tomita, M., and Goodman, K. (1992). *Machine Translation: A Knowledge-Based Approach*. San Mateo, CA: Morgan Kaufmann.

²⁰⁵ David Farwell, Stephen Helmreich *Pragmatics and Translation* / <http://www.sepln.org/revistaSEPLN/revista/24/24-articulo2.pdf>

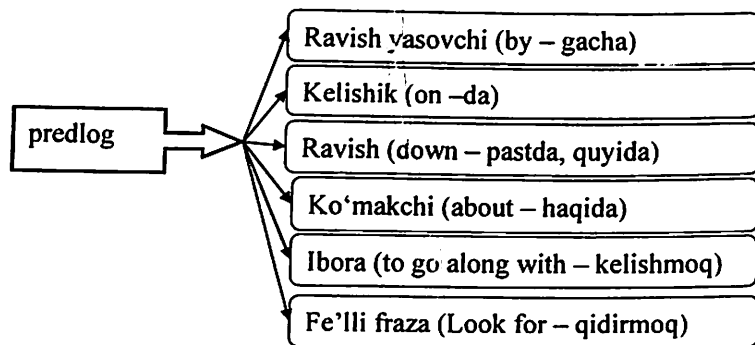
²⁰⁶ <http://www.aclweb.org/anthology/C86-1078>

Masalan, *I saw the man with binocular.* -1) Men binoklli kishini ko'rdim; 2) Men binokl bilan kishini ko'rdim. Ushbu holatda with ham "kishi", ham "binokl" so'zlariga tegishli bo'lishi mumkin.

Mashina tarjimasida sinonimiya hodisasi masalasi alohida o'rganiladi. Ayrim elektron lug'atlarda (ABBY LINGVO) sinonimlar qatori quyidagicha tartiblangan: 1) turkumi; 2) sinonimik qatori; 3) izohi; 4) uslubi (badiiy, so'zlashuv, rasmiy, ilmiy) va unga doir jumlar tarzidagi namunalar; 5) darajasi (kuchsiz, kuchli, neytral); 6) qo'llanilishi (ijobiy, neytral, salbiy); 7) sohasi; 8) valentligi; 9) uning ishtirokida berilgan so'z birikmasi yoki matn namunasi.

Ingliz tili morfologik kategoriyalar turlari va xususiyatlarining o'zbek tili morfologik kategoriyalaridan farqlanishini kuzatishimiz mumkin: *predloglar, modal fe'llar, yordamchi fe'llar, gerundiy, otga xos ba'zi bir xususiyatlar: sanaladigan, sanalmaydigan ot; faqat birlikda, faqat ko'plikda qo'llanadigan otlar; olmoshning ba'zi turlari: much / many (ko'p), little / few (oz) so'zlari va h.k. shular jumlasidandir.*

Ingliz tilida predloglar quyidagi vazifalarga ega²⁰³:



Biror fe'l bilan birikib kelgan predlog so'zning mazmuniga ta'sir qiladi, masalan, *for* predlogi bir qancha ma'no ifodalaydi: 1) - *ga, uchun (atalganlik)* => *This letter is for you (Bu xat sizga yoki siz uchun)*; 2) *uchun, tarafdor* => *I am for your proposal (Men sizning taklifingiz tarafdoriman)*; 3) *uchun (sabab)* => *He was rewarded for*

²⁰³ Qarang: Blokh M.Y. A course in theoretical English Grammar. - M.: Высшая школа, 2008; Кобрин Н.А., Корнеева Е.А., Оссовская М.И., Гузеева К.А. Грамматика английского языка. Морфология. Синтаксис. - СПб.: Союз, 2008

his bravery (U jasurligi uchun mukofatlandi); 4) davomida (ish-harakatning davom etish muddatini ifodalaydi) => *I have been to Andijon for two weeks (Men Andijonda uch hafta bo'ldim)*; 5) *for predlogi bilan keluvchi turg'un birikmalar* => *to ask for (so'ramoq), a contract for (uchun shartnoma)*; 6) *iboralar* => *for ever (abadiy, mangu).*

10.3. Mashina tarjimasida pragmatik va semantik muammolar

Til hamisha ma'lumot beribgina qolmay, unda muloqot uchun insonlarning hissiy, ruhiy va madaniy borlig'i ham aks etadi. Shu bois biror matnni ikkinchi tilga tarjima qilishda yuqorida qayd qilingan barcha unsurlar inobatga olinishi kerak. Shu jihatdan tarjimaga pragmatik yondashish zarurligini ko'rsatadi.

Kompyuter lingvistikasida pragmatikaning o'rni haqida Bar Xillel, Wilks, Nirenburlar tomonidan fikrlar o'rtaga tashlangan²⁰⁴. Pragmatik yondashuv matn tarjimasida uchun kishilarning (muallif, tarjimon va auditoriya) hissiy munosabatlari (nafrat, muhabbat, umid va boshqalar)ga asoslanadi.

Shu jihatdan tarjimaga pragmatik yondashishda quyidagi modellar tizimga kiritilganda quyidagilardan iborat bo'ladi²⁰⁵:

- tarjimonning e'tiqodi (fikri)
- tarjimonning muallif haqidagi fikri
- tarjimonning (original tildagi) adresat fikri uchun bildirgan fikriga munosabati

➤ tarjimonning (tarjima tildagi) adresat fikriga munosabati

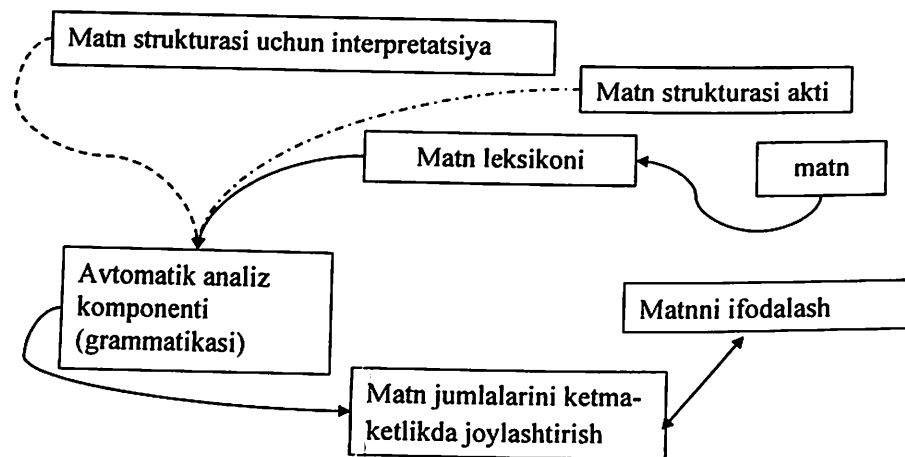
Mashina tarjimasida pragmatik muammolar masalasiga Annely Rothkegel²⁰⁶ o'zining "*Mashina tarjimasida pragmatika*" deb nomlangan maqolasida asosiy muammolardan biri sifatida tarjima ekvivalentlari o'rtasida moslashish hodisasi ekanligini qayd etadi.

²⁰⁴ Bar-Hillel, Y. (1960). The Present Status of Automatic Translation of Languages. In F.L. Alt (ed.) *Advances in Computers*, 1:91-163. New York: Academic Press.; Wilks, Y. (1975). A Preferential Pattern Seeking Semantics for Natural Language Inference. *Artificial Intelligence*, 6:53-74.; Nirenburl, S., Carbonell, J., Tomita, M., and Goodman, K. (1992). *Machine Translation: A Knowledge-Based Approach*. San Mateo, CA: Morgan Kaufmann.

²⁰⁵ David Farwell, Stephen Helmreich *Pragmatics and Translation* / <http://www.sepln.org/revistaSEPLN/revista/24/24-articulo2.pdf>

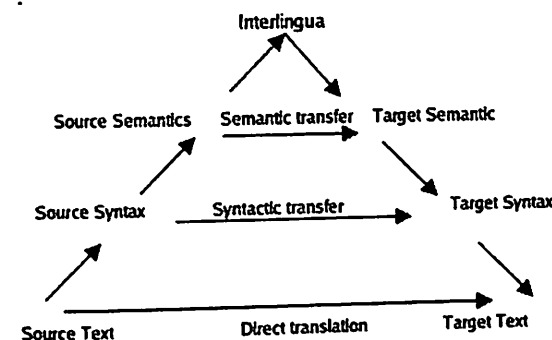
²⁰⁶ <http://www.aclweb.org/anthology/C86-1078>

Unga ko'ra tarjimadagi so'zlar, so'z birikmalar va iboralarning mos variantlarini aniqlash hamda matn turiga qarab tarjima strukturasi belgilash muhim vazifalardan biridir. TEXAN matn analiz qilishning transferga asoslangan tizimi bo'lib, unga ko'ra matnga bog'liq bo'lgan lingvistik ifodalarning modellari yaratiladi. Mashina orqali matn analiz qilinda ma'lumotlar bazasiga kiritilgan lingvistik birliklar lug'at vazifasini o'taydi. Muallif A.Rothkegel ushbu jarayonni quyidagicha ifodalaydi:



Mashina tarjimasida semantik muammolar markazida matndagi lingvistik sathdagi omonimiya hodisasi nazarda tutiladi. Yuqorida mashina tarjima uchun omonimiya masalasi va uning lingvistik muammolariga bag'ishlangan ishlar bilan tanishtirib o'tdik. Ushbu qismda so'zlarning tarjima qilishda semantik ahamiyatga ega bo'lgan muammolari haqida to'xtalib o'tamiz. N. Klyueva o'zining ilmiy izlanishlarida qator ilmiy masalalarga to'xtalib o'tarkan²⁰⁷ mashina tarjima uchburchagini ko'rsatadi (30-rasm):

²⁰⁷ N. Klyueva Semantics in Machine Translation / WDS'07 Proceedings of Contributed Papers, Part I, 141-144, 2007. ISBN 978-80-7378-023-4 © MATFYZPRESS



30-rasm

Y.V.Rojdestvenskiy mashina tarjimasiga sun'iy intellektning eng asosiy muammosi²⁰⁸ sifatida o'rinli baho beradi. Chunki tabiiy tilning imkoniyatlari lisoniy va nolisoniy omillar ta'sirida voqelanadi. Xullas, mashina tarjimasida mukammal ishlangan lingvistik ta'minot dasturning sifatini belgilaydi. Shuningdek, matni tahlil etishda uning grammatik xususiyatlarini to'liqroq o'rganilishi lozim. Inglizcha-o'zbekcha mashina tarjimasini yaratishda ikki tilli lug'atlardagi so'zlarning koefitsentlarini belgilashda ularning kontekstologik xususiyatlari, albatta, inobatga olinishi zarur.

10.4. Tarjima algoritmi

Ilmiy-texnikaviy, diplomatik va madaniy-ma'rifiy millatlararo aloqa almashinuvida tarjimaning o'rni beqiyos. Chunki tarjimaning sifatli amalga oshirilishi kommunikativ muhitni yanada mustahkamlashga xizmat qiladi. Aloqa almashinuvida tarjimaning tezkorligi ham muhim ahamiyatga ega. Negaki bugungi kunning talabi bilan aytadigan bo'lsak, ma'lumotning (xoh verbal, xoh yozma shaklida bo'lsin) yetkazilishi turli oqibatlarni (ijobiy, salbiy) yuzaga chiqarishda omil vazifasini o'taydi. Bu borada kompyuterning kundalik turmushimizda ajralmas qism bo'lib borayotganini hisobga olsak, tarjimon dasturlarini ham mana shunday kommunikativ muhitning bir bo'lagiga aylanib qolayotganligi ham sir emas. Sun'iy intellekt sifatida tarjimon

²⁰⁸ Рождественский Ю.В., Волков А.А., Марчук Ю.Н. Введение в прикладную филологию. – М.: МГУ, 1988. – С. 116.

dasturining ishlash prinsiplari inson tomonidan amalga oshiriladi. Bugungi kunga kelib mashina tarjimasiga oid juda ko'p tizimlarni uchratamiz. Yaratilgan turli mashina tarjima tizimlari quyidagi jihatlarga ko'ra xarakterlanadi:

1. Tillarning ishtirokiga ko'ra:

- ikki tilli;
- ko'p tilli.

2. Tanlangan mavzuning miqdoriga ko'ra:

- bir uslubga tegishli muayyan bir sohaga aloqador;
- bir uslubga tegishli ko'p sohalarga aloqador;
- turli uslublarga tegishli bir necha sohalarga aloqador.

Inglizcha-o'zbekcha mashina tarjimasida tizimini yaratishda tilshunoslar oldida turgan asosiy masalalardan biri sodda gapshakllarini modellashtirish hamda tarjima jarayonidagi nazariy va amaliy tomonlarini tahlil etishdan iborat.

Mashina tarjimasida matn tahlilining quyidagi turlari aniqlangan:

- modullash (Modularity)
- transformatsion tahlil (Transformational parsing)
- filtrlash (Filtering)
- kengaytirilgan usullar yordamida internet tarmog'i orqali bajariladigan sintaktik analiz (Augmented transition network parsing-ATN)
- daraxtsimon qayta qurish (Tree transducers)
- diagramma (Charts)
- semantik tahlil (Semantic analysis)

Ingliz tilida sodda gapshakllar quyidagi turlar bo'yicha farqlanadi ²⁰⁹:

- 1) maqsadiga ko'ra;
- 2) strukturasi ko'ra.

Maqsadiga ko'ra gaplar o'z navbatida to'rt guruhga bo'linadi:

- a) darak gap (declarative);
- b) so'roq gap (interrogative)
- c) undov gap (exclamatory);

d) buyruq gap (imperative).

Mashina tarjimasida, avvalo, gap bo'laklarini aniqlash lozim. Shuning uchun gap bo'laklari modellari bilan tanishib o'tamiz (Shartli belgilar: {}- ixtiyoriy tanlov, [] -model belgisi, | -bo'lishi mumkin.):

Ega (Subject) o'rnida quyidagilar kelishi mumkin: [Noun| Adjective| Pronoun{Personal, demonstrative, defining, indefinite, negative, possessive, interrogative}| Adjective | Participle | Numeral| Infinitive | Gerund| It]

Kesim (sodda) (predicate) o'rnida quyidagilar kelishi mumkin: [Verb| phraseological unit| Word combinations Noun| Adjective| Pronoun {Personal, demonstrative, defining, indefinite, negative, possessive, interrogative, reflexive}| Adjective | | Numeral| Infinitive | Participle II | Participle| Adverb]; Kesim (murakkab) o'rnida quyidagilar kelishi mumkin: [{can| may|must|should|would|ought| dare|need+Verb}| phrasal verbs|word compound verbs]

To'ldiruvchi (the object) o'rnida quyidagilar kelishi mumkin: [noun (in the common case) | pronoun (personal in the objective case, possessive, defining, reflexive, demonstrative, indefinite) | Adjective | Participle|Infinitive| infinitive phrase| Infinitive construction| Gerund| Gerundial phrase| Gerundial construction]; to'ldiruvchi quyidagilarga to'belanishi mumkin: [Object-> Verb | Noun | Adjective].

Aniqlovchi (the attribute) o'rnida quyidagilar kelishi mumkin: [Adjective|Pronoun{ demonstrative, defining, indefinite, possessive, interrogative}| Numeral {cardinal | ordinal} Noun| Prepositional phrase Participles I and II or a participial phrase | Infinitive| Infinitive phrase | Infinitive construction]; aniqlovchi quyidagilarga to'belanishi mumkin: [Attribute -> Noun | Pronoun].

Hol (the adverbial modifier) o'rnida quyidagilar kelishi mumkin:

[Adverb | Prepositional phrase | than+ {Noun| Pronoun| Adjective, Infinitive| Participle|Prepositional phrase} | Participle |Participial phrase | Infinitive | Participial phrases | Prepositional phrases]

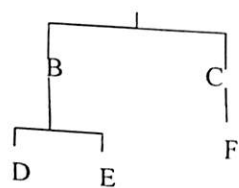
Inglizcha sodda gapshakllarini o'zbek tiliga tarjima qilishda ega va kesim moslashuvi o'zbek tilida aynan muvofiqlashmasligi

²⁰⁹ Каушанская В.Л., Ковнер Р.Л., Кожевникова О.Н., Прокофьева.В., Райнес З.М., Сквирская С.Е., Цырлина Ф.Я. Грамматика английского языка. Ленинград .- 1963.С-3-25.

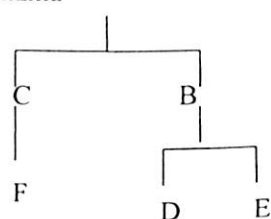
mumkin. Chunki soʻz turkumlari boʻyicha sintaktik tahlil qilinayotganda grammatik tafovut yaqqol koʻzga tashlanadi. Ingliz tilidagi ega va kesim moslashuvi va uni oʻzbek tiliga tarjima qilish masalasini keyingi tadqiqotlarda koʻrib chiqamiz.

Gap boʻlaklari aniqlab olingandan soʻng, matn morfologik va sintaktik tahlil qilinadi. Buning uchun biz yuqorida koʻrsatilgan tahlil usullaridan daraxtsimon qayta qurish turini namuna sifatida tekshiriladi. Bu tahlil usuli dastlab TAUM mashina tarjimasi loyihasining Q-tizimida vujudga kelgan. Mazkur jarayon sintaktik tahlil hisoblanib, u quyidan yuqoriga (bottom-up) tahlil metodiga oʻxshash. Uning strukturasi quyidagicha:

15-chizma



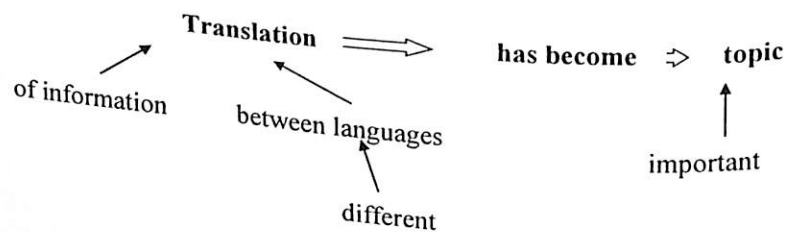
16-chizma A



Agar 15-chizmaning formulasi $A(B(D,E),C(F))$ tarzda boʻlsa, yaʼni gap tashkil etuvchilarining soni qancha va qaysi oʻrinda turishidan qatʼiy nazar qiymati bir xil boʻladi: $(A(B(*), C(*))) > A(C(*), B(*))$ (16-chizma). Koʻrsatilgan ushbu chizmani matn orqali tahlil etamiz:

Translation of information between different languages has become an important topic. Sodda gapshakl modeli quyidagicha boʻladi:

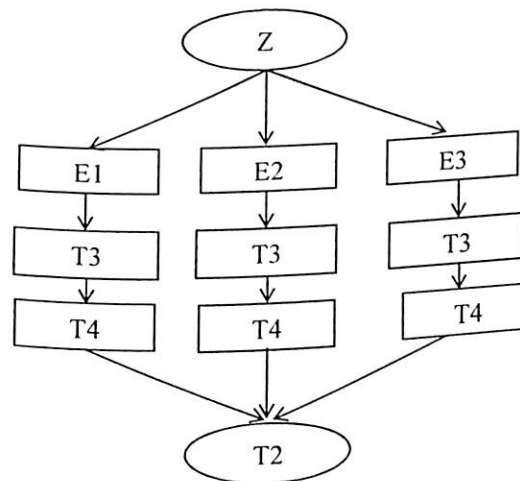
Noun(1) + {Prep.} + Noun(2) + prep. + Adjective + Noun(3) + Verb + Adjective + Noun =>



Dastlab matnlarni tarjima qilishda matnlarni morfologik tahlil qilish bosqichlari amalga oshiriladi. Maʼlumki, maʼlumotlar bazasi, deganda tizimlashtirilgan va strukturalashtirilgan zaruriy holatda ulardan foydalanish hamda saqlash imkoniyatiga ega boʻlgan maʼlumotlar yigʻindisi tushuniladi. Ikki tilli tarjimon muhitida algoritmlar va dasturiy taʼminotni yaratish uchun maʼlumotlar bazasiga quyidagi belgilarni kiritish oʻrinli.

Bazalarning nomi	Funksiyasi
R _i	Predmet soʻhalar boʻyicha terminlar va iboralar bazasi
Q1	Tildagi barcha oʻzak soʻzlar bazasi
K1	Tildagi barcha yasama soʻzlar bazasi
V2	Gap boʻlaklari bazasi
V3	Soʻz turkumlari bazasi

Har bir til uchun yuqoridagi jadvallar yaratiladi. Tarjimon muhitining ushbu naqli ilmiy matnlarni tarjima qilishga xizmat qiladi. Matnni tashkil qiluvchi har bir gapning tuzilishini, muayyan soʻz turkumi qanday gap boʻlagi vazifasida kelishini aniqlashda, albatta, har bir tilning grammatikasiga murojaat qilanadi. Tarjimon muhitining ushbu naqli oʻzbek-ingliz, ingliz-oʻzbek yoʻnalishlardagi tarjimoni amalga oshiradi: Kiritilgan matnni (Z) gap boʻlaklariga ajratib, har bir soʻzni boshqa tildagi tarjimasini terminlar bazasidan olib, shu tilning grammatikasi asosida gap boʻlaklari joylashtiriladi. Tarjima algoritmining funksional chizmasini keltiramiz:



Tabiiy tillarni matematik modellashtirishdagi kengayuvchi kirish tiliga²¹⁰ quyidagi belgilashlar kiritiladi:

- T_{3i}-soʻzning boshqa tildagi tarjimasi va uning gapdagi vazifasidan iborat boʻlgan toʻplam, $1 \leq i \leq m$;
- T_{4j}- soʻzning boshqa tildagi tarjimasi, $1 \leq j \leq m$;
- T₂-tarjima qilingan matn;
- E₄-ega; G₂ -kesim; E₅-aniqlovchi; E₆-toʻldiruvchi; E₇- hol.

Muayyan tildagi gap tuzilishining boshqa tildagi mos gap tuzilishlari modelini yaratamiz. Quyida oʻzbek tilidagi gap tuzilishlariga mos ingliz tilidagi gap modellarini koʻrib chiqamiz.

a) oʻzbek tilidagi darak gap tuzilishining matematik modellari quyidagicha boʻlsin:

- I. 1. $\langle E_4 \rangle \downarrow \oplus \langle E_5 \rangle \downarrow \oplus \langle E_6 \rangle \downarrow \oplus \langle E_7 \rangle \oplus \langle G_2 \rangle$.
2. $\downarrow \langle E_5 \rangle \oplus \langle E_4 \rangle \downarrow \oplus \langle E_6 \rangle \downarrow \oplus \langle E_7 \rangle \oplus \langle G_2 \rangle$.
3. $\downarrow \langle E_5 \rangle \downarrow \oplus \langle E_5 \rangle \oplus \langle E_4 \rangle \downarrow \oplus \langle E_6 \rangle \downarrow \oplus \langle E_7 \rangle \oplus \langle G_2 \rangle$.
4. $\langle E_4 \rangle \downarrow \oplus \langle E_5 \rangle \downarrow \oplus \langle E_6 \rangle \oplus \langle G_2 \rangle$.
5. $\langle E_4 \rangle \oplus \langle G_2 \rangle$.
6. $\langle E_4 \rangle \downarrow \oplus \langle E_7 \rangle \oplus \langle G_2 \rangle$.
7. $\langle E_4 \rangle \downarrow \oplus \langle E_6 \rangle \oplus \langle G_2 \rangle$.

[1,3,4] dan keltirilgan matematik modellarda gap tarkibi soʻz turkumlaridan tashkil topgan, shu matematik modellarga biroz qoʻshimchalar kiritamiz. Yaʼni modellardagi soʻz turkumlarini qaysi gap boʻlagiga oidligi aniqlanib, keyin uning boshqa tildagi tarjimasi olinadi va gap tuzilishiga joylashtiriladi. Modellardan eng optimal va maʼnoli tarjimani xosil qilish algoritmining vazifasiga yuklatiladi. Yuqorida keltirilgan oʻzbek tilidagi darak gap tuzilishlariga mos ingliz tilidagi darak gaplarning matematik modellari qoʻyidagicha shakllantiriladi:

- II. a) 1. $\langle E_4 \rangle \downarrow \oplus \langle E_5 \rangle \oplus \langle G_2 \rangle \downarrow \oplus \langle E_6 \rangle \downarrow \oplus \langle E_7 \rangle$.
2. $\downarrow \langle E_5 \rangle \oplus \langle E_4 \rangle \oplus \langle G_2 \rangle \downarrow \oplus \langle E_6 \rangle \downarrow \oplus \langle E_7 \rangle$.
3. $\downarrow \langle E_5 \rangle \downarrow \oplus \langle E_5 \rangle \oplus \langle E_4 \rangle \oplus \langle G_2 \rangle \downarrow \oplus \langle E_6 \rangle$.
4. $\downarrow \langle E_5 \rangle \oplus \langle E_4 \rangle \oplus \langle G_2 \rangle \downarrow \oplus \langle E_6 \rangle$.

²¹⁰ Хакимов М.Х. Расширяемый входной язык математического моделирования естественного языка для многоязычной ситуации машинного перевода. // О'ЗМУ хабарлари, 2009. -№ 1 -С.75-80.

5. $\langle E_4 \rangle \oplus \langle G_2 \rangle$.
6. $\langle E_4 \rangle \oplus \langle G_2 \rangle \downarrow \oplus \langle E_7 \rangle$.
7. $\langle E_4 \rangle \oplus \langle G_2 \rangle \downarrow \oplus \langle E_6 \rangle$.

b) oʻzbek tilidagi sodda soʻroq gapning matematik modellaridan birini olaylik:

1. $\langle M_4 \rangle \downarrow \oplus \langle E_5 \rangle \downarrow \oplus \langle E_5 \rangle \downarrow \oplus \langle E_6 \rangle \oplus \langle G_2 \rangle$
2. $\langle M_4 \rangle \downarrow \oplus \langle E_6 \rangle \downarrow \oplus \langle E_5 \rangle \oplus \langle G_2 \rangle$
3. $\downarrow \langle E_6 \rangle \oplus \langle M_4 \rangle \oplus \langle G_2 \rangle$
4. $\langle M_4 \rangle \downarrow \oplus \langle E_5 \rangle \downarrow \oplus \langle E_6 \rangle \oplus \langle G_2 \rangle$
5. $\downarrow \langle E_5 \rangle \downarrow \oplus \langle E_6 \rangle \oplus \langle M_4 \rangle \oplus \langle G_2 \rangle$

Ushbu soʻroq gaplarga mos ingliz tilidagi gap tuzilishlari quyidagicha boʻladi:

1. $\langle M_4 \rangle \oplus \langle G_2 \rangle \downarrow \oplus \langle E_7 \rangle \downarrow \oplus \langle E_6 \rangle$
2. $\langle M_4 \rangle \oplus \langle G_2 \rangle \downarrow \oplus \langle E_6 \rangle$
3. $\langle M_4 \rangle \oplus \langle G_2 \rangle \downarrow \oplus \langle E_5 \rangle \downarrow \oplus \langle E_7 \rangle$
4. $\langle M_4 \rangle \oplus \langle G_2 \rangle \downarrow \oplus \langle E_6 \rangle \downarrow \oplus \langle E_7 \rangle$

Yuqorida keltirilgan gap tuzilishlari va terminlar baʼzasidan foydalangan holda tarjima algoritmini keltiramiz.

Quyidagi belgilashlardan foydalanamiz:

Q1_uz="SELECT * FROM `Q1_uz`"-oʻzbek tilidagi barcha oʻzak soʻzlar;

K1_uz="SELECT * FROM `K1_uz`"-oʻzbek tilidagi barcha yasama soʻzlar;

Q1_eng="SELECT * FROM `Q1_eng`"-ingliz tilidagi barcha oʻzak soʻzlar;

K1_eng="SELECT * FROM `K1_eng`"-ingliz tilidagi barcha yasama soʻzlar;

E_i- Z matndan ajratib olingan gaplardan, $1 \leq i \leq n$;

L_{1,j}- E_i-dan ajratib olingan soʻzlar, $1 \leq j \leq n$;

[2] da keltirilgan algoritmi bajarilishidan soʻng, quyidagi "soʻz izlash" algoritmi Z gaplarga, gaplar esa soʻzlarga ajratilib olingach, har bir soʻz dastlab oʻzak soʻzlar baʼzasidan izlanadi, agar u bazada yoʻq boʻlsa, yasama soʻzlar bazasidan qidiriladi. Soʻz topilgach, uning boshqa tildagi tarjimasi olinadi. Masalan, oʻzbek-ingliz yoʻnalishi uchun I-tarjima algoritmi quyidagicha boʻladi.

1. L_{1j} dagi har bir soʻz $Q1_{uz}$ dan izlansin. Topilsa, 2- qadamga aks holda 4-qadamga oʻtilsin;
2. $Q1_{uz}$ dan shu oʻzak soʻzning ingliz tilidagi tartibi (ID) olinsin;
3. $Q1_{eng}$ dan oʻzak soʻzning tarjimasini olinsin va 7- qadamga oʻtilsin;
4. L_{1j} dagi har bir soʻzni $K1_{uz}$ dan izlansin;
5. $K1_{uz}$ dan shu yasama soʻzning $K1_{eng}$ dagi tartibi (ID) olinsin;
6. $K1_{eng}$ dan yasama soʻzning tarjimasini olinsin;
7. Soʻzning gapdagi vazifasi ham aniqlansin va $T3_{i1}$ toʻplamga joylashtirilsin;
8. $T3_{i1}$ toʻldirilgan toʻplamni UzbekIngliz ($T3_{i1}$) funksiyasiga uzatilsin;

UzbekIngliz ($T3_{i1}$) funksiyasining natijalari $T2$ ga joylashtirilsin; Bu yerda UzbekIngliz ($T3_{i1}$) [2] da keltirilgan oʻzbek-ingliz yoʻnalishining tarjima algoritmi asosida yozilgan funksiya. UzbekIngliz($T3_{i1}$) funksiyasi quyidagi koʻrinishda yoziladi: Funksiyani yozishda quyidagi belgilashlardan foydalanamiz: $ET3_{k1}$ - oʻzbek tilidagi gap tuzilishlariga mos boʻlgan ingliz tilidagi gap tuzilishlari $1 \leq k1 \leq m2$;

1. $T3_{i1}$ toʻplamga yuklatilgan soʻzlarning gapdagi vazifalari $E8_k$ ga yuklansin;
 2. $E8_k$ ga mos gap tuzilishi $ET3_{k1}$ dan izlab topilsin;
 3. $ET3_{k1}$ dan izlab topilgan gap boʻlaklari vazifasida kelgan soʻzlar olinsin va $T2$ ga yuklatilsin;
- Bu funksiyaning dasturlash tilidagi koʻrinishini quyidagicha keltiramiz:

```
private String UzbEng(String suz) throws
ObjectNotFoundException {
    int engId=0;
    String engSuz="";
    int gapBulagiId=0;
    UzakSuzlar
    us=uzakSuzUzbekDao.getUzakUzbekByWord(suz);
    if(us.getUzakSuzlar().equals(suz)){
        engId=us.getUzakEnglishId();
```

```
List<UzakEnglish>
ueList=uzakSuzEnglishDao.getuzakSuzlarListByRid(engId);
for (UzakEnglish ue : ueList) {
    engSuz=ue.getUzakEnglish();
}
} else {
    YasamaSuzlar
    ys=yasamaSuzUzbekDao.getYasamaUzbekBySuz(suz);
    if(suz.equals(ys.getYasamaSuzlar())){
        engId=ys.getYasamaEnglishId();
        YasamaEnglish ye=(YasamaEnglish)
        yasamaSuzEnglishDao.getYasamaEnglishListByRid(engId);
        engSuz=ye.getYasamaEnglish();
    } else {
        engSuz=suz;
    }
}
return engSuz;
}
```

Ingliz- oʻzbek yoʻnalishi uchun 2-tarjima algoritmi quyidagicha boʻladi.

1. L_{1j} dagi har bir soʻz $Q1_{eng}$ dan izlansin. Topilsa, 2- qadamga aks holda 4-qadamga oʻtilsin;
 2. $Q1_{eng}$ dan shu oʻzak soʻzning ingliz tilidagi tartibi (ID) olinsin;
 3. $Q1_{uz}$ dan oʻzak soʻzning tarjimasini olinsin va 7- qadamga oʻtilsin;
 4. L_{1j} dagi har bir soʻzni $K1_{eng}$ dan izlansin;
 5. $K1_{eng}$ dan shu yasama soʻzning $K1_{uz}$ dagi tartibi (ID) olinsin;
 6. $K1_{uz}$ dan yasama soʻzning tarjimasini olinsin;
 7. Soʻzning gapdagi vazifasini ham aniqlansin va $T3_{i1}$ toʻplamga joylashtirilsin;
 8. $T3_{i1}$ toʻldirilgan toʻplamni InglizUzbek ($T3_{i1}$) funksiyasiga uzatilsin;
- InglizUzbek ($T3_{i1}$) funksiyasining natijalarini $T2$ ga joylashtirilsin;

Bu yerda InglizUzbek ($T3_{il}$) [2] da keltirilgan ingliz-o'zbek yo'nalishining tarjima algoritmi asosida yozilgan funksiya. InglizUzbek ($T3_{il}$) funksiyasi quyidagi ko'rinishda yoziladi: Funksiyani yozishda quyidagi belgilashlardan foydalanamiz: $ET4_{k1}$ -ingliz tilidagi gap tuzilishlariga mos bo'lgan o'zbek tilidagi gap tuzilishlari $1 \leq k1 \leq m2$;

1. $T3_{il}$ to'plamga yuklatilgan so'zlarning gapdagi vazifalari $E8_k$ ga yuklansin;
2. $E8_k$ ga mos gap tuzilishini $ET4_{k1}$ dan izlab topilsin;
3. $ET4_{k1}$ dan izlab topilgan gap bo'laklari vazifasida kelgan so'zlarni olinsin va $T2$ ga yuklatilsin;

Ushbu funksiyaning Java dasturlash tilidagi tegini quyidagicha keltiramiz:

```
private String EngUzb(String suz) throws
ObjectNotFoundException {
    int Id=0;
    String uzbSuz ="";
    int gapBulagiId=0;
    UzakEnglish
ue=uzakSuzEnglishDao.getUzakEnglishByword(suz);
    if(ue.getUzakEnglish().equals(suz)){
        uzakId=ue.getUzakSuzlarId();
        List<UzakSuzlar>
usList=uzakSuzUzbekDao.getuzakSuzlarListByRId(uzakId)
        for (UzakSuzlar us : usList) {
            uzbSuz=us.getUzakSuzlar();
        }
    }else{
        YasamaEnglish
ye=yasamaSuzEnglishDao.getYasamaEnglishByWord(suz)
        if(suz.equals(ye.getYasamaEnglish())){
            uzakId=ye.getYasamaSuzlarId();
            yasamaSuzlar
            yasamaSuzUzbekDao.getYasamaSuzlarListByRId(uzakId);
            uzbSuz=yu.getYasamaSuzlar();
        }else {
            uzbSuz=suz;
```

```
}
}
return uzbSuz;
}
```

Xulosa qilib aytganda, fan olamida o'zbek tili bilan bog'liq tarjima dasturlarining yaratilishini birlamchi vazifalardan, deb baholash mumkin. Bizning tadqiqotimiz, avvalo, qoidaga asoslanuvchi (Rule-based translation) ikki tilli tarjima muhitida yaratiladi. O'rganilgan va erishilgan ijobiy natijalardan foydalanib, ko'p tilli tarjimon muhitini yaratish istiqboldagi rejalardan sanaladi.

ADABIYOTLAR:

1. Abdurakhmonova N. Z. Grammatical analyze in machine translation 1-я Международная конференция / Компьютерная обработка тюркских языков. Латинизация письменности. Казахстан, Астана, 2013
2. R.Delmonte. Computational Linguistic Text Processing: Lexicon, grammar, Parsing and Anaphora Resolution, Nova Science Publishers, Inc. New York, 2008, 4-5 –P.
3. Абдурахмонова Н., Ҳақимов М.Х. Логико-лингвистические модели слов и предложений английского языка для многоязычных ситуации компьютерного перевода. Компьютерная обработка тюркских языков. Латинизация письменности 1-я Международная конференция Казахстан, Астана, 2013, – С. 302-306
4. Abdurakhmonova N. Automatic morphological analyze for English-Uzbek system // Известия Кыргызский государственный технический университет им. И.Раззакова. Теоретический и прикладной научно-технический журнал № 2 (38) 2016, –С. 12-18
5. Абдурахмонова Н. Ҳақимов М.Х. Семантические базы английского языка для многоязычной ситуации компьютерного перевода. / Труды научной конференции Проблемы современной математики. 22-23 апреля 2011 Карши, –С. 311-314
6. Ахмедова Х.И. Моделлашган компьютер таржимаси технологиясининг алгоритмлари / Амалий математика ва информатсион технологияларнинг долзарб муаммолари Ал-Хоразмий 2014. – Самарқанд 15-17-сентябр 2014
7. Ҳақимов М.Х. Математические модели узбекского языка. ЎзМУ хабарлари, № 3, 2010, –С.185-188
8. Ҳақимов М.Х. Семантические базы и математические модели русского языка для многоязычных ситуаций компьютерного перевода // ЎзМУ хабарлари, № 2, 2011, –С.57-64.
9. Шаляпина З.М.Текст как объект автоматического перевода. – В кн.: Текст и перевод. – М.: Наука, 1988, –С. 113-129.
10. Марчук Ю.Н. Компьютерная лингвистика (учеб. пособ.) Москва, Восток Запад, 2007, –С. 616.

Savol va topshiriqlar

1. Tarjimaning asosiy komponentlarini toping va izohlang.
2. Matnni tarjima qilishda o'rganiladigan asosiy lingvistik masalalar qaysilar?
3. O'zbek tili va ingliz tilining morfologik bazasini tuzishda aniqlab olingan mezonlar nimalardan iborat?
4. Transformatsion metod haqidagi fikringiz qanday?
5. Morfokvalissifikatorning qanday imkoniyatlari mavjud?
6. Lingvistik translyatorning bosqichlarini sanang?
7. Faylasuf Bar Xillel qaysi kategoriyalarni mashina tarjimasining dastlabki rivojlanish bosqichlaridagi asosiy muammolar sifatida ko'rsatgan?
8. Tanlab olingan va cheklangan leksemalarning lingvistik modulini berishda qanday vazifalardan foydalanish maqsadga muvofiq?
9. Ingliz tili morfologik kategoriyalarning turlari va xususiyatlarini o'zbek tili morfologik kategoriyalari bilan solishtirgan holda teglar to'plamini yarating.
10. Mashina tarjimasida sinonimiya hodisasini internet tizimidan foydalangan holda ikki til matni asosida lingvistik tahlil qiling.

10-mavzu yuzasidan test

1. Ko'plab mashina tarjimai tizimlarining modulini qanday bosqichda voqealanadi?
 - a) analiz->transfer->sintez
 - b) transfer->analiz->sintez
 - c) analiz->sintez
 - d) sintez-> transfer-> analiz
2. Lingvistik translyatorning qaysi bosqichida o'zaro bog'langan deklarativ va protsedurali qismlarning lingvistik translyatsiyasi aks etadi?
 - a) assotsiativ protsedura
 - b) lingvistik protsessor
 - c) lingvistik model
 - d) lingvistik struktura
3. Unifikatsion grammatikaning vazifasi to'g'ri ko'rsatilgan qatorni toping.
 - a) semantik valentlik bilan sintaktik valentlik, lug'atga oid tavsif bilan grammatik qoidalarni aniqlashtiradi
 - b) tobelanishga asoslangan grammatika bo'lib, unda har bir element o'zaro tobe munosabatlarga bo'ysunadi
 - c) transformatsion qoidalarga amal qilinadi
 - d) unda grammatik element guruhlarida haqida ma'lumot beradi
4. Faylasuf Bar Xillel qaysi muammolarni mashina tarjimasining dastlabki rivojlanish bosqichlaridagi asosiy muammolar sifatida ko'rsatgan?
 - a) operatsion sintaksis (Operational Syntax)
 - b) tabiiy tillarning ichki tarjimai (Intertranslatability of natural languages), idiomalar (Idioms)
 - c) universal sintaktik kategoriyalar (Universal syntactic categories)
 - d) barcha javob to'g'ri
5. Kim mashina tarjimasiga sun'iy intellektning eng asosiy muammosi sifatida qaraydi?
 - a) Y.V.Rojdestvenskiy
 - b) Ashraf Adali
 - c) Delmonte
 - d) Chomskiy

6. Ingliz va o'zbek tillari tipologik jihatdan boshqa oilalarga mansubligi bois, qanday muammolarni yuzaga keladi?
 - a) inglizcha matnlardagi leksemalarning polisemik va omonimik xususiyatga egaligi; affikslarning turli funktsionalligi;
 - b) ba'zi bir frazemalarning lingvokulturologik kasb etishi;
 - c) matn tarkibidagi leksemalarning sinonimik qatordan tarkib topishi va ularning uslubiy tomondan bog'liqligi;
 - d) barcha javoblar to'g'ri
7. Terminlarning grammatik guruhga kiruvchi so'zlar (artikl+ot+predlog) hamda terminlarning funktsional elementlaridan (ega+kesim) chiziqli tuzilma yoki elementlarning ketma-ketligidan tashkil topgan grammatika nima deb nomlanadi?
 - a) zanjirli grammatika
 - b) unifikatsion grammatika
 - c) kontekstli erkin grammatika
 - d) tashkil etuvchi grammatika
8. O'zbek tili va ingliz tilining morfologik bazasini tuzishda mezonlar aniqlanib, bunda fe'l so'z turkumining qanday shakli kiritiladi:
 - a) o'zak shakli
 - b) nisbat shakli
 - c) o'tgan zamondagi shakli
 - d) shart maylidagi shakli
9. Tokenizatsiyalash jarayoni qanday jarayon?
 - a) tabiiy tildagi nutq birliklari (token, so'zshakllari) alohida tarzda xususiy ma'nolariga ajratiladi
 - b) so'zshakllarining dastlabki holati aniqlanadi
 - c) so'zning o'zagi topiladi
 - d) so'zning gap tarkibidagi vazifasi aniqlanadi
10. Rus tilidagi Krossleksika elektron lug'atidagi morfologik klassifikatorga nimalar kiritilgan?
 - a) 115 guruhga ajratilgan turlanuvchi so'zlar tasnif
 - b) ot so'z turkumiga kiruvchi so'zlar
 - c) yasama so'zlar
 - d) qo'shma so'zlar

11-MAVZU. KORPUS LINGVISTIKASI

11.1. Korpus lingvistikasining nazariy konsepti

11.2 Korpus turlari

11.3. Korpus annotatsiyasi

Tayanch so'z va iboralar: *Korpus lingvistikasi, diaxron korpus til, Brown korpusi, Sketch Engine tizimidagi korpuslar, parallel korpus, korpus annotatsiyasi, lemma, token, konkordans, matnlarni teglash, korpusga asoslangan tahlil usuli*

11.1. Korpus lingvistikasining nazariy konsepti

Korpus lingvistikasi tadqiqot yo'nalishi sifatida bir necha yillar davomida tildan foydalanish va uning variativligini empirik tarzda tadqiq etish usuli sifatida rivojlanib kelmoqda. Korpus matnlar jamlanmasi bo'lib, u mashina o'qiy oladigan formatda matnlarni lingvistik analiz qilish, tabiiy tilni qayta ishlashga mo'ljallangan bo'ladi.

Korpus lingvistikasi haqida olimlarning nuqtai nazarlari turlicha. Qayd etilishicha, u til modeli emas, uni muayyan darajada metodologik yondashuv sifatida qarash to'g'ri bo'ladi. Douglas Biber uning quyidagi xususiyatlarini sanab o'tadi²¹¹:

- u tabiiy matndagi zarur birliklarni empirik tahlil qiladi
- analiz uchun "korpus" sifatida tabiiy matnlarning katta va tizimlashtirilgan jamlanmalarini birlashtiradi
- analiz uchun kompyuterning ham avtomatik va interaktiv texnologiyalaridan foydalanish imkonini beradi
- u analitik texnologiyaning miqdor (stattistik) va sifat xususiyatlarini o'z ichiga oladi.

Korpus lingvistikasi fan yo'nalishi sifatida ba'zi adabiyotlarda amaliy tilshunoslikning, ba'zilarida esa Kompyuter lingvistikasining yo'nalishi sifatida, ba'zilarida esa alohida fan sifatida o'rganiladi. Kaliforniya universiteti mutaxassisi Stefan Gries korpus lingvistikasi uchun berilgan ta'riflardagi metod, nazariya yoki model kabi

²¹¹ Bern Heine, Heiko Narrog. The Oxford handbook of linguistic analysis. / Douglas Biber Corpus-based and Corpus-driven analysis of language variation and use UK: Oxford university, 2015 -P. 193

munosabatlarga nisbatan uni metod(ologiya) deb baholaydi. Til nazariyasining generativ nuqtai nazaridan tilga aloqasi yo'q, degan fikrlarni o'rta tashlaydi. Chunki u deskriptiv va amaliy jihatdan ayrim metodlardan foydalangan bo'lsin ayrim misollar bilan o'z fikrini dalillaydi. Unga ko'ra agar tilshunos biror leksemani korpus ichida grammatik birlik sifatida o'rganmoqchi bo'lsa, demak u, grammatik nazariyaga, agar u yoki bu millatning ikkinchi til sifatida murakkab konstruksiyalardan qay darajada qo'llash imkoniyatini baholamoqchi bo'lsa, ikkinchi tilni o'rganish nazariyasiga asoslanadi, ya'ni har bir izlanuvchi o'ziga xos metodni tanlaydi, degan mulohazalarni bildiradi. Shuningdek, korpus lingvistikasi hamisha chastotaga asoslangan analiz olib boradi. Korpusdagi ma'lumotlar veb sahifalaridagi ma'lumotlardan farqi matnlardan tashqari nutqiy jarayonda yaratilgan turli uslubdagi yozma va og'zaki materiallar (gazeta va jurnal materillari) hamda audio ko'rinishidagi ma'lumotlar ham o'rin oladi. Shu jihatdan korpus umumiy va maxsus turlarga ajraladi. Maxsus turdagi korpuslar janr, uslub, davrlarga ko'ra farqlansa, har ikki tipdagi korpus o'z navbatida diaxron va sinxron shaklida ham bo'ladi. Diaxron korpus tilning davrlar osha qay darajada o'zgarganligini glottoxronologik va statistik tahlil qilsa, sinxron esa ayni vaqtda qo'llanilayotgan nutqiy birliklarning qo'llanilishini o'rganadi. Bundan tashqari korpus monolingval va parallel matnlarning majmuasidan iborat bo'lib, turli sohalar uchun o'rganish obyekt bo'la oladi.

Korpus lingvistikasining vazifalaridan biri lingvistik resurslarning kompyuterli analizi va bazasini yaratishdan iborat.

Korpus lingvistikasi, asosan, XX asrning oxirida rivojlantirilib, XXI asrda ularning turli tillarga oid milliy korpuslari yaratilgan.

Korpus tarixi

Korpus lingvistikasini o'rganish XVIII asrga borib taqaladi. Korpus lingvistikasi dastlab turli maqsadlar uchun foydalanilgan, masalan, Kading (1897) 11mln.dan iborat nemischa so'zlardagi harflar ketma-ketligi va qo'llanish chastotasini o'rgangan, (1876-1926 -yillarda) bolalar tilini o'rganish uchun, xorijiy til pedagogikasi

bo'yicha Fries va Traver (1940), Bongers (1947), Kennedy (1992) va h.k.²¹²ilmiy izlanish olib borgan.

1968-yili Kvirik (Quirk) (Survey of English Usage) lug'atida 1 mln.ta so'z qo'llash bo'yicha bazani 25-yil ichida qo'lyozma nusxadagi 17 satrdan iborat bo'lgan. 1989-yilga kelib uning raqamli shakli yaratilgan. Ushbu korpusga Londondagi Yuniversiti kollejining maxsus ruxsati orqali kiriladi. Elektron korpuslar quyidagi bosqichlarni bosib o'tgan:

1.1960-yillar: Brown korpusi (AQSh), 1 mln.so'z, 500ga yaqin namunalar kiritilgan.

2.1970-yillar: LOB (Lancaster-Oslo/Bergan) korpusi (Buyuk Britaniya, Norvegiya (Lankaster)), 1 mln. so'z

3.1980-yillar: rus tilining mashina fondi

4.Upsala rus tilining korpusi (Shvetsiya), 1 mln. so'z

5.1990-yillar: British National Corpus milliy korpusi 100 mln. so'z (vengr, italyan, xorvat, chesh, yapon tili)

6.1990-yillar: Penn korpusi (Penn Treebank (PTB) - sintaktik razmetkaga asoslangan korpus) – 4.5 mln. so'zni o'z ichiga olib, unga Wall street jurnalidagi aksariyat matnlar kiritilgan va ovozlashtirilgan.

7. Ingliz tili banki Birmingham (Collins Cobuild), 600 mln. so'z

8. 2000-yillar: Amerika milliy korpusi, 100 mln. so'z

9. Zamonaviy Amerika ingliz tili korpusi, 400 mln. so'z

10.Rus tilining milliy korpusi, 140 mln so'z

11.Gigaword korpusi: ingliz, arab, xitoy tili, 2 mlrd. so'z

12.Oxford English korpusi, 2 mlrd. so'z

13.Portugaliya va Galitsiya tillari uchun CRPC, NILC, CORGA

14.Ispan tili uchun CREA

15.Qozoq tili uchun Kazakh-KTB

Braun (Brown) korpusi 1963-64-yillarda Brown universitetida yaratilgan. Unda dastlab 500 ta yozma matndan iborat bo'lgan 1 mln.ta so'z bo'lgan. Switchboard korpusida 1990-yillarning boshida kishilar o'rtasidagi o'zaro 2430 ta muloqot (har bir muloqot

²¹² www2.sal.tohoku.ac.jp/ling/corpus1/1early.htm

o'rtacha 6 daqiqali bo'lib, umumiy 240 soat, 3 mln. so'zdan iborat)²¹³dan iborat.

Rus tilida yaratilgan birinchi korpus 1980-yilda Shvetsiyaning Upsala universitetida tuzilgan. Ungacha L.N.Zasorin rahbarligida 1960-1970-yillarda rus tilining chastotali lug'ati yaratilgan. Ushbu lug'at to'rt janrdagi: dramaturgiya, turli sohadagi siyosiy va badiiy matnlardan iborat 1 mln. so'z qo'llash bo'yicha matnli fayllardan tashkil topgan. Bunday fayllarni yaratishda V.Zaxarov quyidagi bosqichlarni ko'rsatib o'tadi:

-reprezentatsiya

-grafematik analiz

-normallashtirish

-lemmatizatsiya

Korpus menejeri maxsus korpusdagi ma'lumotlarni qidirish uchun dasturiy ta'minotga bog'langan statistik ma'lumotlarni foydalanuvchi uchun qulay shaklda taqdim etuchi qidiruv tizimi hisoblanadi.

Korpus atamasiga quyidagicha ta'rif berish mumkin: Korpus bu - kompyuterning ma'lumotlar bazasida saqlanuvchi og'zaki va yozma matnlar majmui. Korpus juda kichik (50000 so'z) yoki katta hajmda (milliondan ortiq so'z) bo'lishi mumkin. Masalan, Kembridj xalqaro korpusidagi matnlarda 700 mln. so'z mavjud bo'lib, uning hajmi yanada kengaymoqda.

Rus matematigi A.A.Markov tomonidan 1913-yillarda Pushkinning "Yevgeniy Onegin" she'riy romanidagi ruscha so'zlarning statistikasi ishlab chiqilgan. Bu hodisa o'z davrida tilshunoslik va pushkinshunoslik uchun muhim ahamiyatga ega bo'lib, buni olimlar tomonidan korpus yaratishning dastlabki bosqichi deb e'tirof etildi. Hisoblash metodining keyingi bosqichini matnlar korpusi tashkil etadi.

Bugungi kunga qadar Britaniya milliy korpusi (www.natcorp.ox.ac.uk), Amerika milliy korpusi (www.americannationalcorpus.org), ingliz tilining bank sohasi uchun korpus (www.collins.co.uk/Corpus/Corpus Search.aspx), rus

²¹³Jurafskiy D., James H. Martin Speech and language processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 2007 - P. 46

tili milliy korpusi (www.ruscorpora.ru), rus adabiy tili milliy korpusi (www.narusco.ru), XX asr rus tili gazetalari matnining kompyuter korpusi (www.philol.msu.ru/~lex/corpus), A.S.Griboyedev tilining lugʻat-korpusi²¹⁴ yaratilgan.

11.2. Korpus turlari

Matnlar korpusini yaratishda turli meʼyorlarga amal qilinadi. Bular quyidagilar:

- 1) berilganlar bazasining turlariga koʻra: a) ogʻzaki; b) yozma; v) aralash;
- 2) matnlarning tiliga koʻra: a) ingliz; b) rus; s) nemis...
- 3) matn tarjimalarining paralleligiga koʻra: a) ikki tillik; b) uch tillik;
- 4) uslubiga koʻra: a) soʻzlashuv; b) publitsistik; s) badiiy; d) rasmiy; e) ilmiy;
- 5) bazadan foydalanish imkoniyatiga koʻra: a) ochiq; b) yopiq;
- 6) geografik holatiga koʻra: faqat bir davlatga tegishli va h.k.

Yuqorida qayd etilgan ogʻzaki nutqqa xos boʻlgan matnlar korpusi xususiyatiga koʻra soʻzlarning talaffuzi (transkripsiyasi) va turli dialoqlarda qoʻllanilgan nutqiy vaziyatlar keltiriladi. Bu kabi korpuslar taʼlim jarayonida, yaʼni u yoki bu tilni chet tili sifatida oʻqitishda muhim ahamiyat kasb etadi.

V.Zaxarov korpusni quyidagicha tasniflaydi²¹⁵ :

BELGISI	KORPUS TURI
Maqsadi	koʻp maqsadli; muayyan maqsadga yoʻnaltirilgan
Tilga oid maʼlumotlarning turi	ogʻzaki; yozma; aralash
Adabiy tilga oidligi	dialektal; soʻzlashuv; terminologik; aralash
Janr	badiiy; folklor; dramaturg; publisistik
Koʻrinishi	tadqiqotga oid; illyustrativ
Oʻzgaruvchanlik	dinamik; barqaror
Lingvistik annotasiya (razmetka)	razmetkalanmagan; razmetkalanmagan

²¹⁴ Michael McCarthy Touchstone from corpus to course book Cambridge University Press, 2004.63

²¹⁵ Захаров В. П., Богданова С. Ю. Корпусная лингвистика/Учебник для студентов гуманитарных вузов. — Иркутск: ИГЛУ, 2011. — С. 16

Annotasiya (razmetka) turi	morfologik; sintaktik; semantik; anaforik; prosodik
Foydalanish imkoniyati	barcha uchun ochiq; yopiq; tijorat maqsadli
Matn hajmi	toʻliq matnli; matndan berilgan parcha (matn fragmenti)

Korpusning yozma shakli turli manbalardan olinadi: veb sahifalar, yozma materiallar, gazeta va jurnallar, diniy manbalar va h.k. Xullas, yozma matn koʻrinishidagi barcha elektron va qogʻoz holdagi resurslardan korpus sifatida foydalanish mumkin. Ogʻzaki matn korpusi M.Z.Kurdi nuqtai nazariga koʻra quyidagicha boʻladi²¹⁶:

-**Ogʻzaki yozib olish.** Bunda maxsus yigʻilgan matnlar turli kategoriyaga kiruvchi (yosh, jins, kasb, nutqiy imkoniyat, sheva va h.k.) kishilar tomonidan oʻqitib olinib, ulardagi fonetik oʻziga xoslik yozib olinadi;

-**Gapiruvchi buyruq operatorlari.** Bunda ogʻzaki matnlar televizor yoki robot orqali yozib olinadi. Ushbu holatda tabiiy muhit oʻzgaradi, chunki unda matnning beixtiyoriy buzish, soʻzlarni tez-tez takrorlash yoki tushurib qoldirish kabi holatlar kuzatilmaydi.

-**Inson – mashina diologi.** Bunda mashinaning imkoniyati ancha cheklangan boʻlib, inson mashinaga moslashish uchun sodda birliklardan foydalangani bois, bunda lingvistik fenomenning xilma-xiligi yoʻqoladi;

-**Mashina yordamidagi inson – inson diologi.** Ushbu holatda ogʻzaki yoki yozma matn ikki kishi tomonidan yaratilsada, mashina undagi yozishmalarni ketma-ketlikda yozib oladi. Ayniqsa, tarjima tizimida bu jarayondan foydalaniladi.

-**Koʻpmodelli dialog.** Dialog inson bilan inson yoki inson bilan mashina oʻrtasida boʻlsa-da mashina ular oʻrtasidagi muloqotni taʼminlashga vositachi boʻladi. Bunday diologlarda soʻzlar va ishoralar aralash tarzda ishlatiladi.

²¹⁶ Mohamed Zakaria Kurdi. Natural Language Processing and Computational Linguistics: Speech, Morphology and Syntax. —Great Britain, USA: Wiley-ISTE 2016. — P. 4.— 300 p.

Korpuslarning tiliga ko'ra monolingval, bilingval va ko'p tili bo'lishi tadqiqotchi tomonidan o'rganilayotgan ob'ektning xususiyatlaridan kelib chiqadi. Dunyo bo'yicha juda ko'p tillarning korpuslari yaratilgan: Brown korpusi, ingliz tilining yozma va og'zaki matnlar korpusi Switchboard korpusi, fransuz tili uchun Frantext korpus, fransuzcha og'zaki va yozma matnlar uchun OTG korpusini alohida qayd etish o'rinli bo'ladi. Parallel korpus bir necha tildagi matnlar majmuasidan iborat bo'lib, ular ma'lumotlar bazasida o'zaro bog'lanadi. Masalan, rasmiy saytlardagi ma'lumot yoki yangiliklar ikki yoki undan ortiq tillarda beriladi. Bu o'z navbatida matnning tarjima muqobillari tarzida ma'lumotlar bazasi (MB) sifatida yig'ilib mashina tarjimasini uchun resurs bo'lib xizmat qiladi. Multilingval (ko'p tili) korpuslarning to'plami parallel korpuslardan iborat bo'lmay, berilgan matn boshqa tilda aynan o'xshash bo'lmaydi. Namuna sifatida uchta dialekt va o'n ikki tildagi telefon suhbatlaridan tashkil topgan CALLFRIEND korpusi va olti tildagi dialoglardan tarkib topgan CALLHOME korpusini keltirish mumkin²¹⁷.

M.Z.Kurdi korpuslarning tematik ko'lamini matnlarning sohaviy tasnifiga ko'ra uch turga ajratadi:

-muvozanatlashgan korpus – bunda olingan tematik mavzu teng miqdorda taqsimlanadi;

- piramidali korpus– bunda mavzuning dolzarbligiga ko'ra markaziy va kichik matnlar majmuasidan iborat bo'ladi;

- imkoniyat darajasidagi korpus –lingvistik resursi yetarli bo'lmagan holatda berilgan til yoki ilova uchun yaratilgan bo'ladi.

Istalgan turdagi korpus yaratilishida quyidagi savollarga javob izlanishi kerak:

-korpus qanday qilib yaratiladi;

-qaysi turdagi matnlardan tuziladi;

-nima maqsadda foydalaniladi;

-imkoniyatlari nimalardan iborat bo'ladi.

So'ng yig'ilgan matnlar muayyan holatda annotasiyalanib, muayyan maqsadga yo'naltiriladi. Demak dastlab korpus yaratish

²¹⁷ Mohamed Zakaria Kurdi. Natural Language Processing and Computational Linguistics: Speech, Morphology and Syntax. —Great Britain, USA: Wiley-ISTE 2016. —P. 7.

uchun matnning shakli (og'zaki yoki yozma), hajmi, uslubi va tematikasi aniqlab olinishi kerak.

Bir tildagi matn korpuslari o'zining hajmiga ko'ra farqlanadi: nemis –DeReKo, 5,4 mlrd. so'z, [14], ingliz tili – BNC, 100 mln. so'z [15], ingliz tilining amerikacha variant – 450 mln. so'z [16], xitoy – LIVAC Synchronous Corpus, 1 mlr. so'z [17], vengr tili – 100 mln. so'z [18], ispan tili – 100 mln. so'z [19], italyan tili – 100 mln. so'z [20], chexiya tili 200 mln. so'z [21], rus tili НКРЯ, 500 ming. so'z [22] berilgan²¹⁸.

Korpus uchun turli manbalardagi matnlar skaynerlashtirilishi yoki internetdan yuklab olinishi mumkin. An'anaviy usulda yaratiladigan korpuslarning kompilyatsiyasi juda katta mehnatni talab etib, bu jarayon asta-sekinlik bilan amalga oshiriladi. Bugungi kunga kelib ingliz, nemis, fransuz va rus tilidagi katta hajmdagi WEBga asoslangan korpuslarni yaratish borasida loyihalar amalga oshirilmoqda. Ana shunday tadqiqotlar sirasiga Google (Google Books N-gram Viewer) orqali milliy kutubxonalardagi kitoblarni massiv holatda raqamlashtirish, "General Internet Corpus of Russian" (GICR, 19.7 billion tokens) veb-sahifalar (Veb-jurnal), ijtimoiy axborot va yangiliklarni qamrab oluvchi integral veb majmualarga asoslangan korpuslar misol bo'la oladi²¹⁹.

Korpus lingvistik tahlil uchun tilshunoslikning sotsiolingvistika, psixolingvistika, matn tilshunosligi kabu yo'nalishlarda unumli foydalaniladi. Otto Jespersenning ko'p jildli Tarixiy tamoyillarga ko'ra zamonaviy ingliz tili grammatikasi (*A Modern English Grammar on Historical Principles* (1909–49) kitobida lingvistik strukturalarni ifodalashda ko'plab ingliz amerika yozuvchilarning asarlaridan namunalar ketirilgan (Shekspir, Jespersen, Swift va h.k.), lingvistik tahlil uchun olingan namnalar kompyuter o'qiy oladigan shakliga keltirilgan. Birinchi marta yaratilgan Braun korpusi

²¹⁸ З.А. Сиразитдинов, Б.З.Сиразитдинов Корпусные проекты в башкирском языкознании. / Turklang 2013 international confrence - С.59 Turklang 2013 international confrence

²¹⁹ Benko V., Zakharov V.P. Very large Russian corpora: new opportunities and new challenges. // Компьютерная лингвистика и интеллектуальные технологии. По материалам ежегодной международной конференции «Диалог» (2016). Выпуск 15. – С. 86.

elektron holatda bo'lib, kompyuter platformasiga moslashtirilgan (Macintosh, DOS/Windows, and Unixga asoslangan kompyuterlar).

Endilikda korpuslar turli vazifa va maqsadlariga ko'ra tasniflanadi. Braun korpusi "me'yorlashtirilgan" korpus bo'lib, 2000 so'zli ingliz tilidagi yozma matnlarning turli janrlarni o'z ichiga oladi: matbuot reportajlari, nashriyot, hukumat qarorlari va hujjatlari, texnik va badiiy matnlar. Ushbu korpus turli janrga oid matnlarni solishtirish va ingliz tilidagi alohida olingan yozma asarlarni tizimli ravishda o'rganishga mo'ljallangan. Penn daraxtsimon korpusga turli janrdagi matnlar kiritilgan. Brouan korpusi ixtisoslashtirilgan bo'lib, muayyan lingvistik tadqiqot va tahlilga mo'ljallangan. Collins korpusida ingliz tilidagi *must* modal fe'lining (majburiyat va zaruriyat uchun ishlatiladigan semasi) to'rt janr, ya'ni matbuot xabarlari, suhbatlar, ilmiy proza, parlament debatlarida qo'llanilishi tahlil qilingan. Ushbu korpus kompyuterga oid tabiiy tilni qayta ishlash bilan bog'liq bo'lgan tadqiqotlarga mo'ljallanib, tegger va parserlar ustida samarali izlanishlar olib borilgan. Bunda matndagi leksik birliklar so'z turkumlarga (ot, sifat, fe'l...) hamda sintaktik strukturalarga (frazali strukturalar, qo'shma gapning qismlari, gap turlari) avtomatik tarzda ajratiladi.

Korpus lingvistikasini kompyuter lingvistikasining asosiy yo'nalishi sifatida e'tirof etish mumkin, chunki matnni qayta ishlovchi dasturlar korpusga asoslanadi. Turli tipdagi korpuslar lingvistik tahlillarga yo'naltiriladi. Masalan, CHILDES korpusiga bolalar tiliga xos og'zaki matnlarning transkripsiyasi kiritilgan. Ushbu korpus psixolingvistlar uchun tadqiqot obyekti sifatida bolalarning til o'rganish jarayonini ilmiy tekshirishga amaliy yordam beradi. Helsinki korpusi esa Ingliz tilining rivojlanish evolyutsiyasini tarixiy aspektda tahlil qilishda muhim manba hisoblanadi. COLT korpusi London o'smir yoshlarining nutqini o'zida aks ettirgan turli ijtimoiy yoshdagi kishilarning sotsiolingvistik aspektda tahlil qilishga yo'naltirilgan.

Korpus tilning funksional imkoniyatlarini ochib berishga katta yordam beradi. Generativ tilshunoslik tildagi grammatik belgilarning qay tarzda ishlatilishini nazarda tutsa, funkcionl lingvistika o'sha grammatik qolip (marker)larning nutqiy vaziyatda qay tarzda qo'llanilishini o'rganadi. Masalan, *maktab tomon ketmoq* yoki

maktabga ketmoq birikmasida ko'makchi va jo'nalish kelishigi qo'shimchasi o'zaro ma'nodosh hisoblanadi, biroq aniqlik va maqsadga yo'nalganlik mazmuniga ko'ra farqlanadi. Shunga ko'ra korpus lingvistikasi tildagi lisoniy va nutqiy vaziyatlarni aniq baholash, solishtirish va ularni kontekst ichida o'rganishda muhim resurs sanaladi.

Korpus tilni tizimli ravishda o'rganishda ahamiyatga ega. Masalan, Amerika xalqaro korpusida ingliz tilidagi elleptik koordinatsiyalar (EK) obyekt sifatida tanlab olinib, ularning asosi nutqiy vaziyatga qaraganda yozma shaklda ko'proq ishlatilishi, qariyb 100 ming so'zdan iborat turli janrga tegishli matnlar tahlil qilinganda hukumat hujjatlarida majhul nisbat (agentsiz matn), diologlarda esa *men*, *biz* kishilik olmoshlari ko'p qo'llanilishi korpus orqali o'z isbotini topgan.

*Rus tilining zamonaviy korpuslari*²²⁰

Rus tilining milliy korpusi ([http:// ruscorpora.ru/](http://ruscorpora.ru/)). 2013-yilgi mart oyidagi ma'lumotga ko'ra, ushbu korpus 500 mln.ga yaqin so'zdan iborat. Matnlar ilmiy, rasmiy, ommabop, badiiy hamda so'zlashuv uslubi, elektron muloqot matnlari hamda multimedia ma'lumotlaridan iborat (<http://ruscorpora.ru/search-murco.html>) (<http://ruscorpora.ru/corpora-sem.html>)da. Multimediali rus korpusi 1930–2000-yillarda yaratilgan kinofilmlar fragmentidan tuzilgan.

Xelsinki annotatsiyalashgan korpusi ([http:// www.ling.helsinki.fi/projects/hanco/](http://www.ling.helsinki.fi/projects/hanco/)). 2000-yil boshida Xelsinki universitetida ushbu korpus "Rus tilining funksional sintaksisi" loyihasining bir qismi sifatida yaratilgan. Muntazam rivojlantirilib boriladigan ushbu korpusda «Итоги» jurnalida 2001-yilgacha e'lon qilingan maqolalar asosida 100 ming so'zshaklni o'z ichiga oladi. Korpusda morfologik va sintaktik qidiruv tizimi mavjud. Korpusning o'ziga xosligi ma'lumotlarning lingvistik tavsifi izchil ishlangan format hamda avtomatik razmetkalar natijalarni to'liq vizual tekshirish imkoniyati va 2000 ta ko'p tarkibli barqaror birikmalar kiritilgan.

(Buyuk Britaniya) Lids shahridagi universitet korpusi ([http:// corpus.leeds.ac.uk/](http://corpus.leeds.ac.uk/)). 2000-yilda tarjimonlik tadqiqotlari markazida

²²⁰ V.P.Zaxarov tomonidan berilgan rus tili uchun yaratilgan korpuslar tasnifi

turli tillar uchun (ingliz, arab, xitoy, fransuz, nemis, italyan, yapon, ispan, polsha tili) katta hajmdagi korpus yaratilgan. Rus tili (<http://corpus.leeds.ac.uk/ruscorpora.html>) uchun O.N. Lyashevskaya, S.A. Sharov muallifidagi rus tili chastotali lug'atida 116 mln. so'z qo'llanilgan. Bundan tashqari bu saytda (2001–2004-yillardagi) rus tilidagi gazetalar (76 mln. so'z qo'llanilishi) hamda internetdan olingan rus tilidagi matnlarning korpusi ham ilova etilgan (160 mln. so'z qo'llanilishi). IMS Corpus Workbench korpus menedjeriga asoslangan Leeds CQP qidiruv interfeysi o'ziga xos imkoniyatiga ega. U leksik-grammatik qidiruv tizimiga ega bo'lib, maxsus til so'rovlaridan foydalanish imkoniyatiga ega. (<http://corpus.leeds.ac.uk/tools/>).

Moshkov kutubxona korpusi (<http://aot.ru/search1.html>). Bu korpus A.Sokirko tomonidan 680 mln. so'zdan iborat Moshkov kutubxonasidagi matnlardan tuzilgan.

Sketch Engine tizimidagi korpuslar. Ingliz tilidagi "Lexical Computing Ltd." (A. Kilgariff) platformasi qirqqa yaqin turli tildagi korpusni o'z ichiga olgan. Wacky texnologiyasiga asoslanib, 20 mlrd. so'zdan iborat internetdan olingan matnlar majmuasidir. Sketch Engine (<http://sketchengine.co.uk/>) korpus menedjeneri yordamida tuzilgan. U chastotali lug'at, leksik-semantik maydondagi leksik birliklarni guruhlaydi, alohida sintaktik model hamda so'z birikmali konkordanslar uchun standart qidiruv tizimiga ega.

Og'zaki matnlarning korpusi ham yaratilgan. Bunga misol sifatida 10 mln. so'z qo'llanilgan og'zaki nutq korpusining (<http://ruscorpora.ru/sbornik2008/07.pdf>) multimediali, dialektal va aksenteologik korpusni keltirish mumkin.

Rus tilining milliy korpusi²²¹ 18-asr o'rtalaridan 21-asr boshigacha bo'lgan davriy manbalarni qamrab oladi. Ushbu korpus o'tgan va hozirga davr o'rtasidagi sotsiolingvistik variantlar: adabiy, dialektal, so'zlashuv tiliga xos bo'lgan turli adabiy (drama, proza, poeziya) turga oid adabiyotlar, shuningdek, ocherk, esse, jurnalistik nashr, ilmiy va ommabop ilmiy adabiyot, kommunikativ nutq, yozishma va kundalik hujjatlardan iborat.

Rus milliy korpusi quyidagi kichik korpuslardan iborat:

1. Annotatsiyalashgan morfologik va sintaktik tahlilga asoslangan korpus
2. Ingliz, nemis, ukrain, belorus va multilingval tillarga mo'ljallangan parallel korpus
3. Dialektal morfologik variantlardan iborat korpus
4. Turli janrlarni qamrab oluvchi poezik xarakterdagi korpus
5. Rus maktablarida til o'qitish uchun grammatik omonimlarning matnda qo'llanilishini ko'rsatib beruvchi ta'limiy korpus
6. Rus kinolari va turli nutq vaziyatlardan yozib olingan audio korpus (1930-2007).

Tatar tilining "Tugan Tel" korpusi²²² 180 mln. so'zni o'z ichiga oladi. Unda turli janrdagi matnlar o'rin olgan. PC-KIMMO dasturiy instrumenti yordamida tatar tilining ikki bosqichli morfologik analizi yaratilgan. D.Suleymanov, O.A. Nevzorova, R.A.Gilmullin, A.R.Gatiatullin, A.M.Galiyeva, M.M.Ayyupov, va boshqalar ushbu korpusni yaratishda ishtirok etganlar.

Parallel korpus mashina tarjimasi uchun muhim manba hisoblanadi. Shuningdek, ushbu sohada erishilgan natijalardan terminologik ma'lumotlar bazasini boyitishda asosiy manba sanaladi. Korpuslar ikki tilli, uch tilli va ko'p tilli parallel korpuslardan iborat bo'lishi mumkin.

Ushbu rasmda o'zbekcha-inglizcha parallel korpus namunasi (31-rasm) berilgan. Ushbu subkorpus o'zbek tili korpusining (www.uzbekcorpus.uz) negizida yaratilgan.

²²¹ www.ruscorpora.ru/en/corpora-intro.html

²²² <http://tugantel.tatar/?lang=ru>

jamiyat		
UZLASH	TUZLASH	
Uzbek	English	
Fuqarolar o'z huquq va erkinliklarini amalga oshirishda boshqa shaxslarning, davlat va jamiyatning umumiy manfaatlari, huquqlari va erkinliklariga putur yetkazmasliklari shart	The exercise of rights and freedoms by a citizen shall not encroach on the lawful interests, rights and freedoms of other citizens, the state or society	
Davlat organi va mansabdor shaxslar jamiyat va fuqarolar oldida mas'uldir	State bodies and officials shall be accountable to the society and the citizens	
Davlat jamiyatning madaniy, ilmiy va texnikaviy rivojlanishiga qatnashuvchi qilmish	The state shall promote the cultural, scientific and technical development of society and shall promote the cultural, scientific and technical development of society	
O'zbekiston Respublikasida qonunda belgilangan tartibda royxatdan o'tkazilgan kasaba uyushmalari, siyosiy partiyalar, oilalarning jamiyatlari, zotn-qizlar, faxriy va yoshlar tashkilotlari, ijodiy uyushmalar, ommaviy harakatlar va fuqarolarning boshqa uyushmalari jamoat birlashmalari sifatida e'tirof etiladi	Trade unions, political parties, and scientific societies, as well as women's, veterans' and youth leagues, professional associations, mass movements and other organizations registered in accordance with the procedure prescribed by law, shall have the status of public associations in the Republic of Uzbekistan. Scientific societies, professional associations, mass movements and other organizations registered in accordance with the procedure prescribed by law, shall have the status of public associations in the Republic of Uzbekistan.	

31-rasm

Kelajakda ushbu yo'nalishdagi ishlarni yanada rivojlantirish hamda o'zbek tilidan xorijiy tilga va aksincha tarjima qilingan asarlarning parallel korpuslarini yaratish, qiyosiy tarjimashunoslik, leksikografiya, mashina tarjimasi va tilga o'qitish kabi sohalar uchun ijobiy natijalar uchun xizmat qiladi.

11.3. Korpus annotatsiyasi

Leech²²³ qoidasiga asoslangan annotatsiyalash tamoyillari²²⁴:

1. Sof matnga qaytish uchun korpusdan olingan annotatsiya olingan bo'lish kerak. Buning uchun morfologik teglash usulidan foydalaniladi. Morfologik teglash turli tamoyillarga asoslanadi: "Mushtariy_NP1kitob_NN2o'qiydi_VVZ" => *Mushtariy kitob o'qiydi*.

2. Matnning o'zidan annotatsiyalar olinishi mumkin.

3. Annotatsiya strukturasi oxirgi foydalanuvchiga yo'naltirilgan bo'lishi mumkin. Ko'p korpuslarda annotatsiyaning to'liq shakli keltirilgan yo'riqnoma bo'ladi. Bu foydalanuvchiga annotatsiya haqida to'liq ma'lumot berishga xizmat qiladi.

4. Annotatsiya kim tomonidan va qanday yaratilganligi ko'rsatiladi. Korpus annotatsiyasi xoh inson yoki avtomatik tarzda amalga oshirilsa ham inson tomonidan tuzatilishi kerak.

5. Annotatsiya strukturasi umum tomonidan qabul qilingan neytral prinsiplarga asoslanadi.

6. Hech bir annotatsiya o'zidan oldingi annotatsiyalarni standart, deb olishi shart emas. Standart me'yorlar tajriba davomida yuzaga chiqadi.

Matnlarni kompyuterda qayta ishlash jarayoni avtomat, yarim avtomat va qo'l mehnati yordamida amalga oshirilgan. Korpus strukturasi shakllantirish uchun muvozanatlashgan korpus tamoyillariga asoslanildi. Ya'ni olingan matnlarning janri, yaratilgan davri, mavzusi va uslubi proporsional tarzda teng taqsimlandi. Lingvistik resurs sifatida olingan manbalarning original holatini saqlagan holda o'zbekcha matnlarning ham kiril, ham lotin yozuvidagi shakllari bazaga kiritildi. Matnni kompyuterda qayta ishlash uchun unga tegishli kontekstga doir ma'lumot muayyan darajada annotatsiyalanishi kerak. Ishning lingvistik annotatsiyalash (razmetka) bosqichi sifatli natijaga erishishda muhim omil hisoblanadi. Razmetka quyidagi jihatlariga ko'ra tasniflanadi²²⁵:

-Ekstralingvistik (metarazmetka) – matn (muallifi, nashr yili, nashriyot nomi, yili, janri, mavzusi) va uning muallifi haqida ma'lumot;

-Matn tuzilishi – sarlavha, satrboshi, gap, so'z shakli;

-Lingvistik sathiga ko'ra:

1. Morfologik (POS-tagging) razmetka;

2. Sintaktik razmetka – bunda so'zlar o'rtasidagi sintaktik aloqalar va gap turlariga ko'ra ma'lumotlar keltiriladi;

3. Semantik razmetka – matnda ifodalangan tushunchalarning predmet va predmet bo'lmagan nomlari, faoliyat turlari hamda semantik munosabatlari tavsiflanadi.

-Anaforik – matnning biror elementi boshqa matnda anglashilgan mazmun bilan aloqadorlikda aniqlanadi va lingvistik jihatdan tahlil qilinadi;

²²³ Geoffrey Leech matn annotatsiya bo'yicha qoidalarni rivojlantirgan ilk korpus tilshunosi.

²²⁴ www.lancaster.ac.uk/fss/course/ling/corpus/corpus2/2MAXIMS>HTM

²²⁵ К. К. Боярский Введение в компьютерную лингвистику. Санкт-Петербург 2013, - С. 28.

-Prosodik – urg‘u, ritm, mantiqiy urg‘u va boshqlariga ko‘ra matn lingvistik jihatdan ifodalanadi.

Dunyo tajribasida lingvistik annotatsiyaning turlari xilma-xil. Masalan, Leech²²⁶ qoidasiga asoslanlangan annotatsiyalash tamoyillariga e‘tibor berilsa, quyidagilarga guvoh bo‘lamiz:

7. Asos matnga qidirish uchun korpusdan olingan annotatsiya o‘sha matndan aynan olingan bo‘lish kerak. Buning uchun morfologik teglash usulidan foydalaniladi. Morfologik teglash turli tamoyillarga asoslanadi: “Mushtariy_NP1kitob_NN2o‘qiydi_VVZ” => *Mushtariy kitob o‘qiydi*. Ushbu gapda har bir so‘z so‘z turkumiga ko‘ra tavsiflanadi: NP (nounphrase) – ot shakli, NN (noun) – 2-ot shakl, VV (Verb) – fe‘l shakli.

8. Shuningdek, matn qolipiga mos bo‘lgan boshqa matndan annotatsiyalar olinishi mumkin.

9. Annotatsiya strukturasi oxirgi foydalanuvchiga yo‘naltirilgan bo‘lishi mumkin. Aksariyat korpuslarda annotatsiyaning to‘liq shakli keltirilgan yo‘riqnoma bo‘ladi. Bu foydalanuvchiga annotatsiya haqida batafsil ma‘lumot berishga xizmat qiladi.

10. Annotatsiya kim tomonidan va qanday yaratilganligi ko‘rsatiladi. Korpus annotatsiyasi xoh inson yoki avtomatik tarzda amalga oshirilsa ham inson tomonidan tuzatilishi kerak.

11. Annotatsiya strukturasi umum tomonidan qabul qilingan me‘yorlarga asoslanadi.

12. Hech bir annotatsiya o‘zidan oldingi annotatsiyalarni standart, deb olishi shart emas. Standart me‘yorlar tajriba davomida yuzaga chiqadi.

Korpusning lingvistik annotatsiyalash bosqichlari

Korpus annotatsiyasi GML (Generalized Markup Language), SGML (->HTML-Standard Generalized Markup Language, standart umumiy lashgan kodli til), XML (Extensible Markup Language-kengaytirilgan razmetkalar tili), RDF (Resource Description Framework) kabi kodlovchi tillar orqali amalga oshiriladi.

Korpus matnining formati ishlash yo‘nalishiga yordam beradi. Masalan, SGML-XML format korpus lingvistikasiga doir tadqiqotlar

uchun, CoNLLU formatidagi matnlar NLP bo‘yicha tadqiqotlarga xizmat qiladi.

To‘liqroq tasavvurga ega bo‘lish uchun XML tilining orqali manbadan keltirilgan namunasini ilova qilamiz²²⁷:

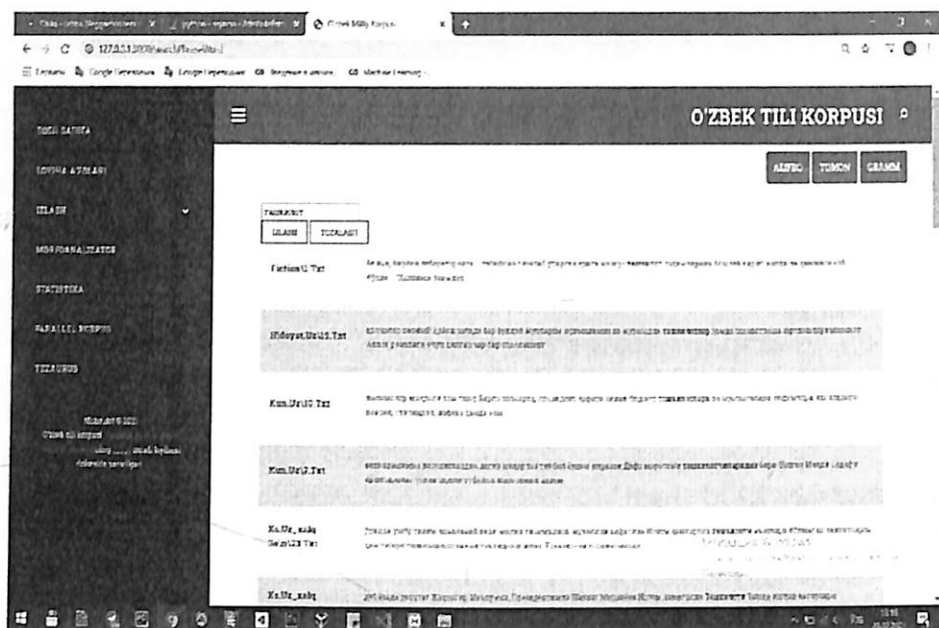
```
<anthology>
<poem>
<heading>The SICK ROSE</heading>
<stanza>
<line>O Rose thou art sick.</line>
<line>The invisible worm,</line>
<line>That flies in the night</line>
<line>In the howling storm:</line>
</stanza>
<stanza>
<line>Has found out thy bed</line>
<line>Of crimson joy:</line>
<line>And his dark secret love</line>
<line>Does thy life destroy.</line>
</stanza>
</poem>
<!-- more poems go here -->
</anthology>
```

Shuningdek, korpus annotatsiyasi quyidagi ma‘lumotlardan iborat bo‘lishi mumkin:

- metadata (metama‘lumot), sarlavha, paragraf
- tokenlar
- lemmalar
- so‘z turkumi tegi (part of speech-POS)
- morfologik belgilar
- sintaktik tahlil

²²⁶ Geoffrey Leech matn annotatsiya bo‘yicha qoidalarni rivojlantirgan ilk korpus tilshunosi.

²²⁷ <https://tei-c.org/release/doc/tei-p5-doc/en/html/SG.html>



32-rasm

32-rasmda *tashkilot* soʻzining token boʻyicha korpus mendjeringining qidiruv natijasi aks etgan. N-gram, masalan, bigram boʻyicha qidirilganda (odatda konkordanslarning oʻzidan oldingi va keyingi soʻzlar qurshovi) matndan olingan unga mos keluvchi namunalar (fragment)ni olish mumkin.

Ayrim manbalarda annotatsiyaning ikki xil shakli sifatida berilib, quyidagicha guruhlanadi: soʻz turkumlarini teglash (POS-parts of speech tagging) va razmetka (razmetka). Birinchi shaklda korpusdagi yoki segmetlardagi har bir leksik birlik grammatik belgilari (ot, sifat, feʼl) teg bilan belgilanadi. Birinchi shaklda avtomatik ravishda har bir segment birlik grammatik jihatdan belgilanib olinadi, soʻng mutaxassis tomonidan qayta tahrirlanadi. Chunki tabiiy tilning istisnoli vaziyatlari ayrim meʼyoriy qoidalarga mos tushavermasligi mumkin va ular tahlilda, shubhasiz, hisobga olinish kerak.

Korpusni teglash matnni lingvistik analiz qilish uchun foydalaniladi. Matnni razmetka qilish turli jihatlarni qamrab olishi mumkin. Masalan, yozma matnlar razmetkasida matnning struktur birliklari, sarlavhasi hisobga olinsa, ogʻzaki matnlar uchun tekshirish obyekti sifatida olingan soʻzlovchilarning yoshi, jinsi, kasbi kabi

tomonlarini ham ifodalashi mumkin. Kiritilgan maʼlumotlarni mashina oʻqiy oladigan formatda berish uchun matnni kodlash yoʻriqnomasi (Text Encoding Initiative (T.E.I.)) orqali standart tilda beriladi. Agar shu kungacha yaratilgan korpuslarga eʼtibor berilsa, Brown korpusi soʻz turkumlari jihatidan teglanganligini, *Penn Treebanki* gaplarning sintaktik analiz qilish boʻyicha annotatsiyalanganligi, *SemCor* korpusida esa soʻzlarning maʼnolariga koʻra semantik jihatdan farqlangan annotatsiyalanganligini, Bangla, Hindi, Marathi, Telugu tillarini oʻz ichiga olgan hind korpusida soʻz turkumlariga koʻra lingvistik annotatsiyalanganligini kuzatish mumkin.

Endilikda korpus lingvistikasi tilshunoslikning ajralmas qismiga aylandi. Negaki kompyuter lingvistikasi, sotsiolingvistika, pedagogika, tarjimashunoslik, diskurs analiz kabi sohalarida korpuslardan unumli foydalanilib, ijobiy natijalarga erishib kelinmoqda.

Oʻzbek tilining elektron korpusini yaratishda dunyo miqyosidagi korpuslar taksonomiyasidan foydalaniladi, shu bois ishimizda ularning ilmiy tavsifiy tomonlariga eʼtiborimizni qaratmoqchimiz.

Makkeneri tadqiqot uchun korpusning quyidagi uch maʼlumotdan iborat boʻlishini taʼkidlaydi²²⁸: metamaʼlumot, matnga oid razmetka va lingvistik annotatsiya. Metamaʼlumot matn haqida maʼlumot beradi. Agar korpus yozma materiallardan iborat boʻlsa, u holda uning metamaʼlumoti manbaning muallifi, qachon nashr qilingani va qaysi tilda yozilganligi haqida maʼlumot beradi. Masalan, Amerika milliy korpusi (<http://www.anc.org/data/masc/corpus/>) (Open American National Corpus (OANC)) 1990-yildan buyon amalda qoʻllanilib kelinmoqda. Uning hajmi 15 mln. soʻzni tashkil qiladi. Korpus quyidagi birliklardan tashkil topgan: DOCUMENT NAME asosiy maʼlumot fayli uchun berilgan nom va barcha annotatsiya qilingan bogʻlangan maʼlumotlar. WORD COUNT hujjatlardagi mavjud soʻzlarning miqdori, TYPE matnning yozma yoki ogʻzaki shakli, GENRE matnni tasniflash uchun korpusning lugʻatidan olingan maʼlumotlarning janri, SUB-CATEGORY mavjud maʼlumotlarning

²²⁸ Charlez M. English corpus linguistics: An introduction. Cambridge, 2004. 12., 29

qo'shimcha tasnifi, ANNOTATIONS mantiqiy tuzilishi (matnning asosiy qismi, paragrafi, sarlavhasi)dan asosiy to'plamdan iborat hujjatlar, lists annotations for the document beyond the basic set, which includes logical structure (body, paragraph, headings, etc.); jumalarning chegaralari, token va so'z turkumlari, ot va fe'l so'z turkumini hamda nomlar (shaxs, hudud, sana va tashkilot)ni aniqlovchi annotasiyalar mavjud²²⁹.

Semantik razmetkada boshqa razmetkalarda bo'lganidek, yagona standart shakl bo'lmasa ham, harf va raqam yoki faqat raqamdan iborat kodlardan foydalaniladi. Birinchi harf yoki raqam umumiy semantik ma'noni, keyingi belgi esa so'z ma'nosini yanada maxsuslashtiruvchi kichik semantik guruhni ifodalaydi. Semantik razmetka nafaqat so'z, balki ko'plab birikmalarni ham semantik guruhlarga birlashtiradi, bunday paytda turli birikuvdagi bir ma'noni bildiruvchi birikmalar bitta belgi bilan kodlanadi. Idiomatik ibora tarkibidagi so'zlar miqdorini bildiruvchi axborot ham razmetkadan joy oladi. Semantik razmetka korpusdagi so'z ma'nosining ixtisoslashuvi, omonimlik va sinonimlik, ma'noviy guruhga ajratish kabi muammoni hal qiladi. V.P. Zaxarov, S.Yu. Bogdanovlar rus tili milliy korpusini tuzishda semantik razmetkalashning o'z variantini taklif qiladi²³⁰. Bu korpusda har bir so'zga uch xil – so'z razryadi, leksik-semantik tavsif, derivatsion izoh beriladi.

Anaforik razmetkalash jarayonida matnga ishlov berishda ayrim holatlarda qiyinchilik tug'diruvchi olmosh so'z turkumi va matn birliklariga ishora qiluvchi nutq birliklariga e'tibor qaratiladi. Ishora so'zning matndagi ma'nosini ajratib olish uchun razmetkaning alohida ko'rinishiga ehtiyoj tug'iladi. Anaforik razmetka shu xildagi ma'lumotni qo'shish uchun zarur. Ma'nosi olmosh bilan ifodalanayotgan so'z alohida kodlanib, keyingi o'rinda shu so'zga ishora qilayotgan olmosh yoniga shu kod biriktiriladi. Natijada olmoshning matndagi ma'nosi aniqlanadi hamda kerakli tadqiqotlarda korpusga asoslanish imkoniyati paydo bo'ladi.

Prosodik razmetkalash ovoz transkripsiya qilingan korpusda urg'u va ohangni ifodalovchi izoh kiritiladi. Razmetkaning diskurs

²²⁹ Абдурахмонова Н. Замоновий корпусларнинг компьютер моделлари Ўзбекистонда хорижий тиллар. –2020. –№ 1(30). – Б. 50–58. [https://doi.org/ 10.36078/](https://doi.org/10.36078/)

²³⁰ Захаров В.П., Богданова С.Ю. Корпусная лингвистика. - Иркутск: ИГЛУ, 2011.

deb ham ataluvchi ushbu turi sharh, izoh, eslatma, takrorlardagi to'xtamlarni bildirish uchun ishlatiladi. Korpusni razmetkalash (annotatsiyalash) dasturlash usuli bilan amalga oshiriladi. Bunda, avvalo, vaqtni tejash va mehnatni kamaytirish nazarda tutilsa, ikkinchidan, matnga avtomatik ishlov berish muammosiga yechim topiladi. Hozircha anaforik va prosodik razmetka qiyinligicha qolib ketmoqda hamda ularning aksariyati qo'l mehnati bilan amalga oshirilmogda.

Morfologik va sintaktik razmetka esa tegger va parsing yordamida amalga oshirilsa ham, bu dasturlarning ham aksariyati avtomatik razmetkadan keyingi tuzatishni talab qiladi. O'zbek tili uchun aynan istisnoli vaziyatlar ko'p bo'lganligi sababli qayta tahrirlash zaruriyatini keltirib chiqaradi. Chunonchi, morfologik omonimiya (ko'proq flektiv tillarga xos) va sintaktik ko'p ma'nolilik holatida dastur natijasining bir nechta shakllarini taklif qiladi, mutaxassi bularning orasidan eng munosibini tanlaydi.

Yangi avlod korpuslari hajmining haddan ziyod kengayib borayotgani mutaxassislar oldiga razmetkaning to'liq avtomatlashtirilgan turiga o'tish, yangi tegger va parsinglar yaratish vazifasini qo'ymoqda. Avtomatik morfologik tahlil (tegger) yordamida har bir leksik birlikka (so'z turkumi, lemma, grammema guruhi) alohida grammatik kategoriyalari (shaxs-son, kelishik va boshq.) beriladi. Masalan, Braun korpusida so'zning chastotasini aniqlash oson. Faqat bu so'z shaklning (korpus tilida token) chastotasi bo'ladi. Leksemaning chastotasini aniqlash uchun esa har bir so'zga uning lemmasi biriktirilgan bo'lishi kerak. V.P.Zaxarov fikricha, lingvistik razmetkaning morfologik, sintaktik, semantik, anaforik, prosodik kabi turlari quyidagi tamoyillar asosida amalga oshiriladi²³¹:

- 1) razmetka sxemasini tavsiflash (asoslash);
- 2) umumiy lingvistik tushunchalar tizimini aniqlash;
- 3) foydalanuvchi uchun ma'lum bo'lgan tahlil sxemasini shakllantirish;
- 4) razmetka sxemasining nazariy an'anaviyligiga erishish;
- 5) xalqaro andozalarga amal qilish.

²³¹ Захаров В.П., Богданова С.Ю. Корпусная лингвистика. - Иркутск: ИГЛУ, 2011.

Bundan tashqari, ayrim adabiyotlarda soʻz turkumlarini annotatsiyalashning boshqa bir qator turlarini keltiriladi²³²: bordiyu muayyan soʻzning soʻz turkumlari sinfi bilan maʼlumot keltirilishi zarur boʻlsa, ichkiqator (inline) annotatsiya yaratiladi: *The inhabitants of Egypt were called mummies*. Ushbu gap quyidagicha teglanadi: *The/Determiner inhabitants/Noun of/Preposition Egypt/Name were/Verb called/Participle mummies/Noun /Punct*. Yaʼni gapda ishtirok etgan har bir soʻz turkumi nominal yoki verbal guruhdan qaysi biriga tegishli ekanligiga koʻra ajratiladi va / belgisi qoʻyiladi. Olim qoʻshimcha sifatida yana bir annotatsiyalash turi sifatida *yonma-yon (stand-off) annotatsiyani* keltirib oʻtadi. *w1:Det w2:Noun w3:Preposition w4:NE w5:Verb w6:Participle w7:Noun w8:Punct*. Ushbu annotatsiyalash bosqichida soʻzni qayta aslyatdagi matndan topib, u haqidagi maʼlumotni chiqarishda qoʻl keladi.

Annotatsiya jarayonida muayyan xulosani chiqarish uchun annotatsiya modellaridan (annotation scheme) foydalaniladi. Ular annotatsiya qoʻllanmalarida qayd etiladi. Masalan, ot son kategoriyasi boʻyicha annotatsiya qilinadi, degan xulosaga kelish uchun qoʻllanmada keltirilgan tavsialardan foydalanish oʻrinli. Chunki ayrim matn tahlili bilan bogʻliq murakkab vaziyatalarda u yoki bu soʻz haqida lingvistik maʼlumot olish uchun undagi berilgan maʼlumot juda muhim. Masalan, oʻzbek tilida ayrim atoqli otlar turdosh otga (Nyuton) yoki turdosh otlar atoqli otga aylangan hollari koʻp kuzatiladi. Ularni bosh harf bilan kelish kelmasligi mavhum, chunki ular gapning boshida ham kelishi mumkin. Ushbu vaziyatda annotatsiyadan foydalaniladi.

Annotatsiyalashning keyingi talabalaridan biri bu matn haqidagi maʼlumotni qay tarzda ifodalashdir. Odatda matnni annotatsiyalash qoʻl mehnati bilan, avtomatik va yarim avtomat tarzida amalga oshiriladi. Qoʻl mehnati bilan amalga oshirishda inson idroki va uning shaxsiy tajribasiga tayaniladi. Avtomatik usulda esa kompyuter lingvistikasidagi metodlardan foydalangan holda natijaga erishiladi. Yarim avtomat yordamida annotatsiyalashda avtomatik

²³² Sandra Kübler and Heike Zinsmeister CORPUS LINGUISTICS AND LINGUISTICALLY ANNOTATED CORPORA Bloomsbury Academic, New York, 2015, - P. 23

dasturiy metod yordamida bajariladi, erishilgan natijalari inson tomonidan tekshirib chiqiladi.

Shuningdek, *temporal (vaqt bilan bogʻliq) annotatsiya* turi ham mavjud²³³. Ushbu annotatsiyada voqealar oʻrtasidagi munosabatlar va tildagi vaqt va voqeani ifodalovchi ifodalar aniqlashga eʼtibor beriladi. Bu kabi maʼlumotlarning berilishi voqealar izchilligini topishda yordam beradi. Temporal annotatsiya toʻrt turdagi maʼlumotlarni beradi: voqealar, vaqtni bildiruvchi soʻzlar, kalit (siganl) soʻzlar hamda havolalar. Ingliz tilida soʻzdan oldingi qism aniq vaqni ifodalaymaydi, balki qolgan yoki davom etayotgan ikki hodisa oʻrtasidagi vaqtinchalik munosabatni bildiradi. Vaqtni ifodalashda signal soʻzlar havola berilganda qoʻllaniladi, havolalar esa hodisalar orasidagi bogʻlanish boʻlib hizmat qiladi. Oʻzbek tilida quyidagi misolni namuna sifatida keltirish mumkin: *Covid 19ning dunyo boʻylab tarqalishi davlat va jamiyatning turli sohalariga jiddiy taʼsir koʻrsatib, rivojlanish koʻrsatkichlarining pasayishiga olib keldi*. [_{B1}Covid 19]ning [_{B2}dunyo boʻylab tarqalishi] [_{B3}davlat va jamiyatning turli sohalariga] [_{B4}jiddiy taʼsir koʻrsatib], [_{B5}rivojlanish koʻrsatkichlarining pasayishiga olib keldi]. Bunda [_{B1}Covid 19]-> [_{B5}rivojlanish koʻrsatkichlarining pasayishiga olib keldi] komponent biror voqeaning sodir boʻlishidan avvalgi holar va shu ikki tarkibning oʻrtasidagi signallar [_{B2}dunyo boʻylab tarqalishi], [_{B4}jiddiy taʼsir koʻrsatib]dan iborat.

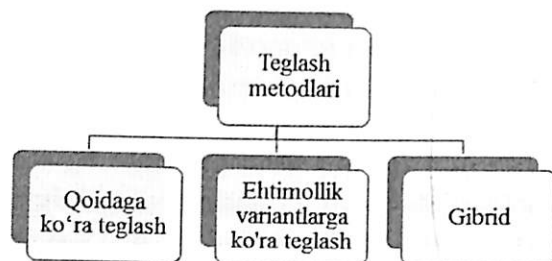
Gaplarning maʼnolarini ifodalashning yana bir usuli bu semantik freymlardan foydalanishdir. Semantik freymlarda ensiklopedik bilimlar bilan bogʻlanadi²³⁴. Masalan, ot yoki feʼl soʻz turkumi doirasida u yoki bu soʻz ifodalagan maʼno munosabatlari (sinonimlar, uyadosh soʻzlar, giponim va giperonimlar kabi) orqali uning semantik maydonlarini hosil qiladi. Shu kabi birliklar kontekstning ichidan semantik annotatsiya uchun model vazifasini bajaradi. Semantik freymlar bilan annotatsiyalangan korpus

²³³ Sandra Kübler and Heike Zinsmeister CORPUS LINGUISTICS AND LINGUISTICALLY ANNOTATED CORPORA Bloomsbury Academic, New York, 2015, - P. 24

²³⁴ Sandra Kübler and Heike Zinsmeister CORPUS LINGUISTICS AND LINGUISTICALLY ANNOTATED CORPORA Bloomsbury Academic, New York, 2015., -P. 29

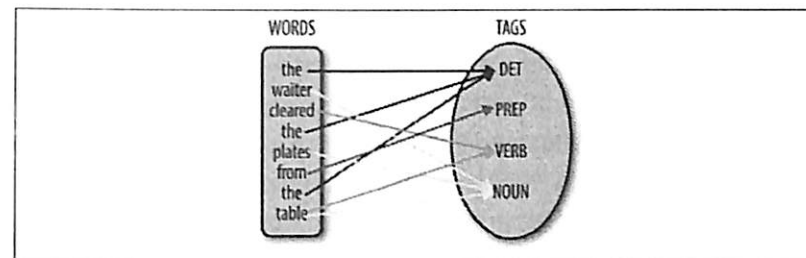
FreymNet deb yuritiladi. Masalan, o'qimoq_harakat-holat, kitob, kitobxon, kutubxona_vosita kabi birliklar freymlar hisoblanib, o'qimoq, mutolaa qilmoq, ko'z yugurtirmoq, ko'zdan kechirib chiqmoq kabi fe'llar ma'no jihatidan farq qilsada, bitta semantik maydonni tashkil qiladi. Yani, ushbu fe'llar freymlar o'rtasidagi munosabatni yuzaga chiqaruvchi vositalari bo'lib, FreymNetning leksik birliklari hisoblanadi. Demak, freymlar semantikaga asoslanadi. Gaplarning sintaktik strukturalari darajasida lingvistik tavsifini berish diskurs annotatsiyada ifodalanadi. Ushbu holatda annotatsiya mantiqliy tamoyiliga asoslanadi. Ya'ni mantiqiylik deyilganda matn tarkibida so'zlovchilarning og'zaki va yozma nutqiy jarayonida ajratilgan yoki o'zaro bog'lanmagan so'z, fraza va gaplarning yig'indisi tushuniladi. Qisqa paragraflar turli darajadagi mantqiylikni ko'rsatadi.

MATNLARNI TEGlash (TAGGING)



1. Qoidaga asoslangan teglash leksikonga asoslanadi, biroq o'zi uchun noma'lum bo'lgan konstruksiya uchun teglarni aniqlay olmaydi.
2. Ehtimolga asoslangan teglash – ehtimollik nazariyasi metrikasi so'z noma'lum grammatik sinfga tegishli bo'lsa, qo'shimcha ma'lumot bazasidan topishga harakat qiladi.
3. Gibrid teggerga (CLAWS) kiradi.
POS (Part of speech tagging), ya'ni so'z turkumi teglariga quyidagilar kiradi:²³⁵

²³⁵James Pustejovsky, Amber Stubbs Natural Language Annotation for Machine Learning. 2013 USA, - P. 16.



Penn tree-bank tegseti quyidagicha ifodalanadi:

Number	Tag	Description
1.	CC	Coordinating conjunction
2.	CD	Cardinal number
3.	DT	Determiner
4.	EX	Existential there
5.	FW	Foreign word
6.	IN	Preposition or subordinating conjunction
7.	JJ	Adjective
8.	JJR	Adjective, comparative
9.	JJS	Adjective, superlative
10.	LS	List item marker
11.	MD	Modal
12.	NN	Noun, singular or mass
13.	NNS	Noun, plural
14.	NNP	Proper noun, singular
15.	NNPS	Proper noun, plural
16.	PDT	Predeterminer
17.	POS	Possessive ending
18.	PRP	Personal pronoun
19.	PRP\$	Possessive pronoun
20.	RB	Adverb
21.	RBR	Adverb, comparative
22.	RBS	Adverb, superlative
23.	RP	Particle
24.	SYM	Symbol
25.	TO	to
26.	UH	Interjection
27.	VB	Verb, base form
28.	VBD	Verb, past tense
29.	VBG	Verb, gerund or present participle
30.	VBN	Verb, past participle
31.	VBP	Verb, non-3rd person singular present
32.	VBZ	Verb, 3rd person singular present
33.	WDT	Wh-determiner
34.	WP	Wh-pronoun
35.	WP\$	Possessive wh-pronoun
36.	WRB	Wh-adverb

Korpus texnologiyasi quyidagi yo'nalishlarda maqsadli foydalaniladi:

1. Matndagi muallif yoki xarakterning psixologik nuqtai nazarini aniqlash. Bunda annotatsiyalangan korpus yaratiladi. Tegger moduli

orqali statistik jihatdan diskurs o'rganiladi. So'ng diskurs teggeri korpusda munosabat bildiruvchi birliklarni ifodalarni belgilab chiqadi. Ushbu yo'nalish murakkab jarayon bo'lib, hech qanday formal mezonlarga asoslanmaydi.

2. Mashina tarjimasi. Bunda parallel matnlar orqali tarjima amalga oshiriladi. Tayyor matnlardagi birliklar tarjima uchun me'yor vazifasini bajaradi. Statistik tahlil asosida matnda eng ko'p qo'llanilgan tarjima tanlanib, tarjima qilinayotgan til strukturasi joylashtiriladi.

3. Nutq sintezatori. Korpusdagi matnlar orqali qo'shimchalar kombinatsiyasi yoki so'z shakllariga tayanib, ehtimollik variantlariga ko'ra leksik birliklar avtomatik ovozlashtiriladi.

4. Avtomatik tahrirlovchi dastur. Tokenlardagi grammatik shakllar orqali matn korpusidan olingan variantlarga ko'ra analiz qilinadi. Bunda so'zlarning fonetik, morfologik sintaktik va semantik analiz qilish imkoniyatini yaratadi.

Adabiyotlar:

1. Рыков В.В. Курс лекций по корпусной лингвистике. URL: <http://rykovcl.narod.ru/c.html>
2. narod.ru/c.html
3. Kennedy, Graeme. An Introduction to Corpus Linguistics / Graeme Kennedy. - London: Longman, 1998. - 315 p
4. Tony McEnery, Andrew Wilson. Corpus Linguistics. - Edinburgh University Press, 2001. URL: http://books.google.com/books?id=nwmgdvN_akAC
5. Developing linguistic corpora: a guide to good practice. Edited by Martin Wynne. URL: <http://www.ahds.ac.uk/creating/guides/linguistic-corpora/index.htm>
6. D. Biber, S. Conrad, R. Reppen. Corpus Linguistics: Investigating Language Structure and Use 3. - Cambridge University Press, 1998. URL: <http://books.google.com/books?id=2h5F7TXa6psC>
7. Соснина Е. П. Введение в прикладную лингвисти. Учебное пособие по курсу «Основы теоретической и

прикладной лингвистики» для студентов направления «Лингвистика» Ульяновск УлГТУ 2012 С. 74

8. Сайхунов М. Р., Ибрагимов Т. И., Хусаинов Р. Р. Письменный корпус татарского языка [Электронный ресурс]. Казань, 2012. Режим доступа: <http://corpus.tatar>, свободный.

9. Aksan, Y., Aksan, M. Building a national corpus of Turkish: Design and implementation. Working Papers in Corpus-based Linguistics and Language Education no. 3, pp.299-310. Tokyo: TUFU, 2009.

Savol va topshiriqlar

1. Korpus lingvistikasiga metod, nazariya, metodologiya yoki model sifatida qarashlarga shaxsiy munosabatingizni bildiring?
2. Korpusdagi ma'lumotlar veb sahifalarida ilova etilgan ma'lumotlardan farqi nima?
3. Korpus nima?
4. Elektron korpuslar qanday bosqichlarni bosib o'tgan?
5. So'z qo'llash bo'yicha matnli tashkil topgan fayllarni yaratishda V. Zaharov qanday bosqichlarni ko'rsatib o'tgan?
6. Pushkinning "Yevgeniy Onegin" she'riy romanida qo'llanilgan rus tilidagi so'zlarning statistikasini kim tayyorlagan?
7. Matnlar korpusini yaratishda qanday aspektlarga asoslaniladi?
8. CHILDES korpusini izohlang.
9. London o'smir yoshlarining nutqini o'zida aks ettirgan, turli ijtimoiy yoshdagilarning sotsiolingvistik tadqiq qilish mumkin bo'lgan korpus qaysi?
10. Leech qoidasiga asoslanlangan annotatsiyalash tamoyillari qaysilar?

10-mavzu yuzasidan test

1. 100mln. soʻz (vengr, italyan, xorvat, chesh, yapon tili)ni oʻz ichiga olgan korpus berilgan qatorni toping.
- British National Corpus milliy korpusi
 - Upsala rus tilining korpusi
 - Brown korpusi
 - Penn korpusi
2. CRPC, NILC, CORGA nomli korpuslar qaysi tillar uchun yaratilgan?
- Portugaliya va Galitsiya tillari uchun
 - Ispan tili uchun
 - Qozoq tili uchun
 - Rus tili uchun
3. Rus tilida yaratilgan birinchi korpus qayerda yaratilgan?
- Shvetsiyaning Upsala universitetida
 - Turkiyaning Istanbul universitetida
 - Amerikaning Brown universitetida
 - Rossiyaning Moskva davlat universitetida
4. Pushkinning “Evgeniy Onegin” sheʼriy romanida qoʻllanilgan rus tilidagi soʻzlarning statistikasi kim tomonida tayyorlangan?
- A.A.Markov
 - A.S.Griboedev
 - Chomskiy
 - Mitkov
5. CHILDES korpusi qanday korpus hisoblanadi?
- bolalar tili transkripsiyasi kiritilgan turli nutqiy vaziyatlar tahlil qilingan korpus
 - qoʻshma feʼllar kiritilgan korpus
 - ot soʻz turkumiga oid soʻzlar kiritilgan
 - sifat soʻz turkumiga oid soʻzlar kiritilgan
6. London oʻsmir yoshlarining nutqini oʻzida aks ettirgan korpus berilgan qatorni toping.
- COLT
 - CHILDES
 - CRPC
 - NILC
7. Multimediali rus korpusi qanday matnlardan tuzilgan?

- 1930–2000-yillardagi kinofilmlar fragmentidan tuzilgan
 - 1930 – 2000- yillardagi yaratilgan romanlar matnidan foydalanilgan
 - 2000-yillardagi filmlar matnidan tuzilgan
 - 2001- yillardagi qoʻshiqlar matnidan foydalanilgan
8. Moshkov kutubxona korpusi berilgan qatorni toping.
- <http://aot.ru/search1.html>
 - <http://corpus.leeds.ac.uk/ruscorpora.html>
 - <http://corpus.leeds.ac.uk/tools>
 - [http:// www.ling.helsinki.fi/projects/hanco/](http://www.ling.helsinki.fi/projects/hanco/)
9. Rus tili milliy korpusi qanday manbalarni oʻz ichiga oladi?
- 18-asr oʻrtalaridan 21-asr boshigacha boʻlgan davriy manbalarni qamrab oladi
 - 19-asr oʻrtalaridan 21-asr boshigacha boʻlgan davriy manbalarni qamrab oladi
 - 20-asr oʻrtalaridan 21-asr boshigacha boʻlgan davriy manbalarni qamrab oladi
 - 18-asr oʻrtalaridan 20-asr boshigacha boʻlgan davriy manbalarni qamrab oladi
10. Rus tilining milliy korpusi ([http:// ruscorpora.ru/](http://ruscorpora.ru/)) ning oʻziga xos xususiyatlari berilgan qatorni toping.
- multimediali korpus (<http://ruscorpora.ru/search-murco.html>) hamda semantik format (<http://ruscorpora.ru/corpora-sem.html>)
 - matnlar turli janrdagi matnlar: ilmiy, rasmiy, ommabop, badiiy hamda soʻzlashuv va elektron muloqot matnlarini qamrab olgan
 - (2013-yilgi mart oyidagi maʼlumotga koʻra) soʻz qoʻllashning 500 mln. hajmga ega
 - barcha javoblar toʻgʻri

12-MAVZU. KOMPYUTER LINGVISTIKASIDA METODLAR

12.1. Statistik metod

12.2. Modellashtirish metodi

12.3. Korpusga asoslangan tahlil metodi

Tayanch soʻz va iboralar: *n-gram algoritmi, statistik metod, modellashtirish metodi, ehtimolga asoslangan metod, korpusga asoslangan metod, original modellar, funksional modellar, struktur modellar, generativ grammatika, transformatsion grammatika*

12.1. Statistik metod

“Statistika” termini lotin tilidan olingan boʻlib, *status –holat* maʼnosini beradi. Oʻrta asrda ushbu termin davlatning *siyosiy holati* maʼnosida qoʻllanilgan. XVIII asrda nemis olimi Gorfrid Axenval tomonidan fanga termin sifatida kiritilgan.

Kompyuter lingvistikasi metodlari haqida bir qator tadqiqotlar mavjud²³⁶. Kompyuter lingvistikasi uchun samarador metodlardan biri statistik metod hisoblanadi. Kompyuter lingvistikasining mashina tarjimasini, korpus lingvistikasi, sentiment analiz, matn klassifikatsiyasi, matn tahlili kabi yoʻnalishlarining rivojlanishida statistik tahlil metodining ahamiyati katta.

Markov tomonidan 1913-yilda Pushkinning “Yevgeniy Onegin” asaridagi undosh yoki unli harflarning ketma-ket kelishini aniqlash uchun yaratilgan *n-gram* algoritmi fanga *markov zanjiri (bigram, trigram)* termini bilan kiritilgan²³⁷. Soʻzlarning ehtimoliy holatda ketma-ketlikda kelishi til modellari deb ataladi²³⁸. Soʻzlarning ketma-ketlikda kelishi *n-gram* deb nomlanadi. *n-gram*dagi *n* soʻzning ketma-ketligidagi sonni bildiradi. 2-gram soʻzlarning ikki

tarkibli ketma-ketligi (*men bozorga*), 3-gram soʻzlarning uch tarkibli ketma-ketligi (*men bozorga ertalab*), 4-gram soʻzlarning toʻrt tarkibli ketma-ketligi (*men bozorga ertalab bordim*). Muayyan soʻzning oʻzidan oldingi va keyingi soʻzlar bilan ketma-ket kelish ehtimolligini aniqlash mashina tarjimasini, nutqiy sintezator kabi tizimlar uchun muhim sanalib, ushbu holatda korpus lingvistikasining oʻrni juda katta. Chunki til va nutqiy vaziyat shu darajada betakror, tuzilgan gaplar uchun mukammal modellarni yaratishning imkoni yoʻq. Biroq unda ishtirok etayotgan birikmalarning muayyan darajada bazasini yaratish mumkin. Aytaylik, soʻzning oʻzidan oldingi ehtimollik darajasini aniqlash uchun $P(w_n|w_1^{n-1})$ bigram (ikki til juftligining ketma-ket kelishi) modelidan foydalaniladi²³⁹. Shuningdek, oʻzidan oldin yoki keyin ikki soʻzning kelish ehtimolligini aniqlash trigram, oʻzidan uchta oldin yoki keyin kelishi mumkin boʻlgan soʻzlarning ehtimolligini aniqlash 4-gram va h.k. kabi modellar uchun quyidagi algoritmlar tuzildi (30-rasm):

```
import io
myFile=io.open('Words.txt',mode='r', encoding='utf-8')
sentences=[]
for i in myFile:
    sentences.extend(i.split())
n=int(input("n grammdagi birliklar sonini kiriting"))
for i in range(0, len(sentences)-n):
    print(sentences[i:i+n])
```

²³⁶ Чеповский А. М. Методы компьютерной лингвистики и анализ текстов на естественных языках. Учебное пособие. М.: Национальный открытый университет «ИНТУИТ», 2015.

²³⁷ Jurafskiy D., James H. Martin Speech and language processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 2007 -P. 60.

²³⁸ Jurafskiy D., James H. Martin Speech and language processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 2007 -P. 38

²³⁹ Jurafskiy D., James H. Martin Speech and language processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 2007 -P. 39



33-rasm

Ehtimollik nazariyasida kombinatorik tahlil

Kombinatorika (kombinatorik tahlil) — bu diskret matematikaning diskret to‘plam elementlarini berilgan qoidalar asosida tanlash va joylashtirish bilan bog‘liq bo‘lgan masalalarni yechish usullarini o‘rganuvchi bo‘limidir²⁴⁰.

Matematikaning muayyan yo‘nalishi sifatida B.Paskal ishlarida o‘z ifodasini topgan. G.V.Leybnits, Y.Bernulli, L.Eyler kabi olimlar tomonidan kombinatorik tahlil rivojiga ulkan hissa qo‘shilgan.

Muayyan belgilardan iborat guruhlariga kombinatsiyalar deb aytiladi. Ana shu kombinatsiyalarni tashkil etgan predmetlarga elementlar deyiladi. Uch xil turdagi kombinatsiyalar mavjud: o‘rin almashtirish, o‘rinlashtirish va mosliklar. 1 dan n gacha bo‘lgan natural sonlar ko‘paytmasi «n faktorial» deb ataladi va qisqacha $n!$ kabi yoziladi: $n! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot (n - 1) \cdot n$. ($0! = 1$).

O‘rin almashtirishlar n ta elementli o‘rin almashtirishlar deb bir-biridan faqat elementlarining tartibi bilan farq qiladigan n ta elementli birikmalarga aytiladi. Masalan, uchta A, B, C elementdan oltita o‘rin almashtirish bajarish mumkin: ABC, ACB, BAC, CBA, BCA, CAB. n ta elementli o‘rin almashtirishlar soni P_n bilan

²⁴⁰ A.S.Rasulov, G.M.Raimova, X.K.Sarimsakova Ehtimollar nazariyasi va matematik statistika.Toshkent, 2006. –B. 4.

belgilanadi va quyidagi formula bilan hisoblanadi: $P_n = 1 \cdot 2 \cdot \dots \cdot (n-1) \cdot n = n!$

O‘rin almashtirishlar n ta elementdan m tagacha bo‘lgani o‘rinlashtirishdir. Har birida berilgan n ta elementdan t tasi olingan shunday birikmalarning har biri hech bo‘lmaganda bitta elementi bilan yoki faqat ularning joylashish tartibi bilan farq qiladi²⁴¹. Masalan, uch element A, B, C dan ikkita elementli oltita o‘rinlashtirish mavjud: AB, AC, BC, BA, CA, CB. n ta elementning m tadan turli o‘rinlashtirishlar soni A_n^m bilan belgilanadi va quyidagi formula bilan hisoblanadi:

$$A_n^m = \frac{n!}{(n-m)!} = n(n-1)(n-2) \dots (n-m+1). \quad (0 \leq m \leq n).$$

$$A_n^1 = n \text{ va } A_n^0 = 1.$$

12.2. Modellashtirish metodi

Ma’lumki, model bevosita kuzatish imkoni bo‘lmagan hodisani o‘rganish uchun yaratilgan sun’iy mexanizmdir. «Model (qolip) (fran. modele<ital. modello – namuna, qolip) – biron til hodisasi tarkibi yoki qismlarining joylashish holatini ko‘rsatuvchi sxema yoki namuna»²⁴².

Yer kurrasini teleskop bilan, mikroorganizmni mikroskop bilan kuzatganimizdek tilni ham modellashtirish orqali kuzatish mumkin²⁴³.

Modellashtirishda quyidagi jihatlar e’tiborga olinadi:

I. Aniqlashtirish: obyekt butun va qism sifatida o‘rganiladi:

- 1) umumiy g‘oya ayrim vositalari orqali tekshirish;
- 2) obyektни qayta nomlash;
- 3) u yoki bu muhitdan obyektни material holatda ajratish mumkin.

II. Strukturani soddalashtirish: obyekt mantiqiy ichki tuzilma sifatida tekshiriladi, bunda turli qismlardagi farqlar o‘rganiladi

III. O‘zaro mustaqillik, funksiallik

²⁴¹ O‘sha joyda. –B. 5.

²⁴² Хожиев А. Тилшунослик терминларининг изоҳли луғати. – Тошкент, 2002. – Б.65.

²⁴³ Computational approaches to morphology and Syntax. Brian Roark, Richard Sproat. Oxford university press. 2007.

IV. Originallik

V. Barqarorlik

Modellashtirish metodida tadqiqotchi obyektning o'zini emas, balki uning modelini o'rganadi. Original bilan model o'rtasida o'zaro bog'liqlik va mutanosiblik mavjud bo'ladi.

Tilshunoslikda modelning 3 ta turi bor:

1. Original modellar.
2. Funktsional modellar.
3. Struktur modellar

Original modellar – obyektning tuzilishini, funktsional modellar – original modellarning qanday ishlashini, struktur modellar – har ikkalasi haqida ma'lumot beradi.

Original modellarga quyidagi talablar qo'yiladi:

- 1) model tabiiy obyektning aynan nusxasi bo'lishi;
- 2) original model o'zida elementlarning murakkab tuzilishini namoyon eta olishi;
- 3) tabiiy obyektning hamma xususiyatlari original modelga to'g'ri kelishi;
- 4) model evristik funksiyaga ega bo'lishi, ya'ni u yangi g'oyalar bera olishi va uni amaliyotda sinab ko'rish imkoniyati bo'lishi zarur.

Ma'lumki, zamonaviy kompyuter lingvistikasi «lingvistik model» tushunchasini struktur lingvistikadan meros qilib olgan. Umuman, kompyuter lingvistikasining dunyoga kelishi struktur lingvistikada lingvistik modelning shakllanishi bilan bog'liq. Lingvistik model tushunchasi ilmiy muomalaga struktur tilshunoslikning E.Sepir, L.Blumfild, R.Yakobson, X.Chomskiy, Z.Xarris, Ch.Hokket kabi namoyandalari asarlarida o'z ifodasini topgan. Uning taraqqiyoti XX asrning 60 – 70-yillari (matematik va kibernetik lingvistika rivojlana boshlagan davr) ga to'g'ri keladi.

Demak, struktur lingvistika va kompyuter lingvistikasini bog'lab turuvchi mustahkam ko'prik - bu lingvistik model tushunchasidir. Model bevosita kuzatish imkoni bo'lmagan hodisani o'rganish uchun yaratilgan sun'iy mexanizm. Ammo matn inson nutq faoliyati asosida yotuvchi til mexanizmini to'liq tushunish uchun yetarli emas. Shuning uchun tilshunoslikda ob'ektni tushunishning asosiy vositasi sifatida modellardan foydalaniladi.

Lingvistik modelning asosiy xususiyatlari:

1. Faqat obyektning funktsional struktur tomonini modellashtirish mumkin. Uning semantik tabiati bilan bog'liq tomonlari modellashtirilmaydi. Boshqacha qilib aytganda, model obyekt kabi harakat qilishi talab qilinadi, xolos. Obyektning materiali (moddiy qobig'i) doimo modeldan farq qiladi. Masalan, tildagi kelishiklar tizimi obyekt sifatida miyaning asab kataklari holatida kodlashtirilgan, ammo uning modeli qalam bilan qog'ozga yozilgan bo'lishi mumkin. Bunda model bo'yicha hosil qilingan kelishik qoidalari inson miyasi faoliyati bilan bir xil ishlashi, bir xil natija berishi model uchun ahamiyatli xolos.

2. Modelda deyarli doimo obyekt ideallashtiriladi. Hayotiy hodisalar juda murakkab. Bu murakkabliklarni o'rganish uchun, avvalo, eng sodda va umumiy hodisalarni bilish lozim.

Tilshunoslikda bu metod ilgari mavjud bo'lib, struktur tilshunoslikda o'tgan asrning 60-70-yillarida faol qo'llanila boshlangan. Masalan, to'liqsiz gaplarni tahlil qilish uchun ular «to'liqlashtiriladi». Ideallashtirish tabiiy hodisani qaysidir ma'noda qo'pollashtirishga olib keladi. Ammo sxemalashtirmasdan ilmiy yondashuv yaratish mumkin emas. L.Yelmsev aytganidek, «Ilmiy konsepsiya kartina emas, balki diagrammadir».²⁴⁴ Hodisani barcha jihatlarini baravariga tekshirmoqchi bo'lgan olim hech qanday ilmiy muammoni hal qila olmaydi.

3. Model, odatda, obyekt haqidagi aniq tushunchalar bilan emas, balki, konstrukt bilan ish ko'radi, ya'ni model bilan ishlashda tajribaga asoslangan ma'lumotlardan bevosita hosil bo'lgan tushunchalardan emas, balki kuzatishlar yig'indisi va kuzatuvchi intuitsiyasiga asoslangan ayrim umumiy gipotezalarga tayaniladi. Misol uchun «Nol ko'rsatkichli leksema» deb ishlatamiz. Aslida bu leksemada «nol» yo'q, ammo ko'rsatkichning yo'qligi shu konstrukt dan foydalanishga asos bo'lgan.

4. Har qanday lingvistik model formal bo'lishi shart. Modelning formal bo'lishi deganda, modelning « mantiqiy davomiylilik + bir ma'nolilik + aniqlik »ka ega bo'lishi tushuniladi. Bu xususiyat nazariyaning asosiy shartlaridandir.

²⁴⁴ Ельмслев Л. Метод структурного анализа в лингвистике // Звегинцев В.А. История языкознания в очерках и извлечениях. Ч.2 –М., 1965. -С.103.

5. Har qanday lingvistik model eksplanatorlik xususiyatiga, ya'ni tushuntirish kuchiga ega bo'lishi kerak. Agar model biror xususiyatga ega bo'lsa, u:

1) eski nazariya tushuntirib bera olmagan muammoni hal qiladi;

2) obyektning ilgari kuzatilmagan, ammo kelajakda amalga oshishi mumkin bo'lgan tomonini kashf qiladi.

N.Chomskiy ta'kidlaganidek, model faqat so'zlovchi nutqining amaliyoti davrida uchragan til obyektlarini qurish bilan cheklanishi mumkin emas, balki model hali so'zlovchi nutqi amaliyotida uchramagan, ammo uchrashi mumkin bo'lgan til hodisalarini ham qamrab olishi kerak²⁴⁵. Faqat shunday modellargina so'zlovchining har qanday yangi gaplarni tuzish (noto'g'ri gaplardan tashqari) va tinglovchining har qanday yangi gaplarni tushunish qobiliyatini qamrab oladi. Bunday modellar bola tomonidan tilning o'zlashtirilishini ham tushuntirib berishi mumkin.

Lingvistik strukturalar model bo'lishi mumkin. Biroq sintaktik jihatdan qaralganda modellashtirish lingvistik o'ziga xos jihatlarni, ya'ni sintaktik omonimiya masalasini qamrab olmaydi. Mark Jonson²⁴⁶ ushbu hodisaga daraxtsimon model orqali izoh beradi. *I saw the man with the telescope*. Ushbu gapni sintaktik jihatdan ikki xil modellashtirish mumkin: agar ikkinchi darajali bo'lakni tobelanuvchisi fe'lga qarashli bo'lsa, tarjima *Men teleskop bilan u kishini ko'rdim deb* (1), agar ikkinchi darajali bo'lakka qolgan qism tobelansa, *Men teleskop ushlab olgan kishini ko'rdim deb* (2) tarjima qilinadi.

Tilni matn orqali kuzatish mumkin, ammo matnning inson nutqiy faoliyati jarayonida yuzaga chiquvchi tabiiy til hodisalarini to'liq tushunish uchun yetarli emasligi adabiyotlarda ko'p bora ta'kidlangan. Shuning uchun tilshunoslikda obyektning tushunishning asosiy vositalaridan biri sifatida modellardan foydalaniladi. Matn tilshunosliklikning barcha sathlari uchun asos vazifasini o'taydi²⁴⁷.

²⁴⁵ Хомский Н. Синтактические структуры // Новое в лингвистике.-М., 1962.

²⁴⁶ <http://web.science.mq.edu.au>

²⁴⁷ Qarag'ang: 3. Tuxsanov M. Микротекст и система средств выражения его когерентности в узбекской художественной речи (замена и повтор): Автореф. дис. канд. филол. – Ташкент: Наук, 1987.; Ўринбоев Б., Қўнгулов Р., Лапасов Ж. Бадиий текстнинг лингвистик тахлили. –Тошкент: Ўқитувчи, 1990; Ҳакимов М.Х. Ўзбек илмий матнининг синтагматик ва прагматик хусусиатлари:

Negaki, matn nutqiy vaziyatning yozma (shuningdek, ba'zi adabiyotlarda og'zaki) shaklidir. Uslubiy jihatdan matn birliklari haqida ham turli munozarali fikrlar bildirilgan. Ya'ni gap va supersintaktik birliklar (abzas) matnni tashkil etuvchilari sifatida e'tirof etiladi.

Model lingvistik obyektning formal tavsifini berishda qulay vosita sanaladi. Uning quyidagi turlari mavjud²⁴⁸:

Model - chizmalar tilga oid obyektlarining mazmuni va strukturasini ifodalashda qo'llaniladi;

Model – tur (language pattern) muayyan barqaror birliklar uchun (so'z birikmalari, sintaktik strukturalar);

Model – nazariya formallashtirilgan ilmiy nazariya, elementlar strukturasini va metatilga moslashtirilgan qoidalar asosidagi tavsif.

Kompyuter texnologiyalarga asoslangan inson bilimi, tili va tafakkurini modellashtirishda sui'iy intellekt texnologiyasi shug'ullanadi.

12.3. Korpusga asoslangan tahlil metodi

Tilshunoslik va til ta'limi uchun korpusga asoslangan yondashuv so'nggi qirq yillikda, xususan, 1980-yil o'rtalaridan boshlab mashhur bo'ldi. 90-yillarning boshlarida korpusga oid tadqiqotlarning kashfiyotlarini til pedagogikasiga qo'llash ortdi. Lankaster, Oksford, Graz, Bertinoro, Granada, Parij va Lissabonda Ta'lim va Til korporasi bo'yicha o'tkazilgan 8ta xalqaro anjuman qiziqishlar ortib borayotganidan dalolat beradi²⁴⁹.

Филол. фан. номз....автореф. –Тошкент. 1993.; Тожиев Ё. Матн кўринишлари ва матн таърифи. Тилшуносликнинг долзарб масалалари. (Илмий мақолалар тўплами). –Тошкент, 2010. –Б.164-171; Накимов М.Х.Ўзбек тилида матннинг прагматик талқини: Докторлик дисс. автореф. –Тошкент, 2001; Каменская О.Л. Текст и коммуникация. – М., 1990. –С. 8; Чинникулов Н.Пешлавха коммуникатив бирлик сифатида // Ўзбек тили ва адабиёти. 2006. 2-сон. –Б.90-93; Шодиев И. Матннинг коммуникатив турлари // Ўзбек тили ва адабиёти. –Т., 2006. – № 2, – Б. 86-88; Алронова Ш. Матннинг қабул қилинишига доир // Ўзбек тили ва адабиёти. –Т., 2006. – № 2. – Б.73-76.

²⁴⁸ Соснина Е. П. Введение в прикладную лингвистику Учебное пособие Ульяновск УлГТУ 2012, - С. 36.

²⁴⁹ Tony McEnergy and Richard Xiao. "What Corpora Offer in Language Teaching and Learning". 2010. – 364-373.

Yevropa olimlari tomonidan til o'rganishda korpusga asoslangan yondashuv to'g'risida bir qancha ilmiy tadqiqotlar amalga oshirilgan. Tonny McEnery va Richard Xiao tilni o'qitish va o'rganishda korpusning afzalliklari borasida ilmiy izlanishlar olib brogan va buni isbotlab berishgan. Eric D. Brill, Samina Dazdarevic, Amela-Lukac Zoranic, Fahreta Fijuljaninlar ham tilni o'qitish va o'rganishda korpusga asoslangan yondashuvning foydaliligi haqida izlanishlar olib borishgan.

Word	Count	Category
1	1	1
2	1	1
3	1	1
4	1	1
5	1	1
6	1	1
7	1	1
8	1	1
9	1	1
10	1	1
11	1	1
12	1	1
13	1	1
14	1	1
15	1	1
16	1	1
17	1	1
18	1	1
19	1	1

34-rasm

Korpusga asoslangan metod kompyuter lingvistikasining mashina tarjimai, morfologik analiz, sintaktik tahlil (34-rasm), Treebank (sintaktik ma'lumotlar ombori) kabi yo'nalishlarda keng foydalanilmoqda. Ushbu metod orqali ma'lumotlar analiz qilinganda lisoniy va nutqiy vaziyatlari korpusda o'z ifodasini topadi.

Bundan tashqari korpus bilan bog'liq lingvistikaning boshqa psixolingvistika, sotsiolingvistika, tarjima nazariyasi, dialektologiya kabi yo'nalishlarida ham korpusga asoslangan metoddan unumli foydalanib kelinmoqda.

Dialektologiya muayyan hududdagi so'zlarning so'zlovchilar nutqida qo'llanilishi va uning boshqa dialektlardan farqini yaxshiroq o'rganishda muhim o'rni bor. Turli sotsial qatlamning uslubi, leksikasi va muloqot tilini tahlil qilishda undagi faktlar katta

ahamiyatga ega. Korpus lingvistikasi tilning barcha aspektlarini tahlil qila olgani uchun lingvistikasining barcha yo'nalishlarida faol ravishda qo'llanilmoqda (35-rasm).

Word	Count	Category
1	1	1
2	1	1
3	1	1
4	1	1
5	1	1
6	1	1
7	1	1
8	1	1
9	1	1
10	1	1
11	1	1
12	1	1
13	1	1
14	1	1
15	1	1
16	1	1
17	1	1
18	1	1
19	1	1
20	1	1

35-rasm

Tilni har jihatdan o'rganish imkoniyati mavjudligi bois korpusga asoslangan ta'lim texnologiyasi bugungi kunda eng faol qo'llanilayotgan ta'lim usulidir.

Adabiyotlar:

1. Abdurakhmonova N., Urdishev K. Corpus Based Teaching Uzbek As A Foreign Language Journal of Foreign Language Teaching and Applied Linguistics (J-FLTAL) Volume 6 - Number. 1 - 2019, - P. 131-137
2. А.С. Восковская, Т. А. Карпова Английский язык для вузов (Учебное пособие) Ростов-на-Дону «Феникс» 2008.
3. Винокуров А.А., Чуканов В.О. Новый метод оценки машинного перевода. // Информационные технологии и системы. Hardware Software Security. Тенденции и перспективы - Сборник статей: М., Международная академия информатизации, 1997.
4. Шаляпина З.М. Текст как объект автоматического перевода. — В кн.: Текст и перевод. — М.: Наука, 1988, с. 113-129.
5. Хакимов М.Х. Расширяемый входной язык математического моделирования естественного языка для

многоязычной ситуации машинного перевода. ЎзМУ хабарлари, № 1, 2009, с.75-80.

6. Хакимов М.Х. Формальные системы машинного перевода в многоязычной ситуации. Материалы республиканской научной конференции «Современные проблемы математики, механики и информационных технологий», НУУз, Институт Математики и ИТ АН РУз, Т, 2008, с.297-301

Savol va topshiriqlar:

1. Tilshunoslikda model turlari va ularni izohlang.
2. Original modellarga qanday talablar qo'yiladi?
3. Lingvistik modelning asosiy xususiyatlari qaysilar?
4. Modellashtirish metodidan foydalanib til yaruslari bo'yicha amaliy mashg'ulot olib boring.
5. Chomskiy N. ta'kidlaganidek, model faqat so'zlovchi nutqida uchragan til obyektlarini qurish bilan cheklanishi mumkin emas, balki model hali so'zlovchi nutqi amaliyotida uchramagan, ammo uchrashi mumkin bo'lgan til hodisalarini ham qamrab olishi kerak. Fikrni izohlang.
6. N.Enkvist matnning qanday lingvistik parametrlari haqida fikr yuritgan?
7. Inglizcha-o'zbekcha mashina tarjimasida matn qanday parametrlarni o'z ichiga olishi kerak?
8. L.N.Belyayev matnlarni qanday turlarga ajratgan?
9. L.I.Borisova mutaxassislikka oid ilmiy-texnikaviy matnlarni tarjima qilishda qanday fikrni taklif qiladi?
10. Tarjimon dasturining ta'minotini yaratishda lingvistik modellarni ingliz va o'zbek tillari misolida yarating.

12-mavzu yuzasidan test

1. Tilshunoslikda modelning qanday turlari bor?
 - a) original, funksional, struktur modellar
 - b) struktur, nazariy, qiyosiy
 - c) qiyosiy, nazariy
 - d) lingvistik, texnologik
3. Funksional modellar nima?
 - a) original modellarning qanday ishlashini o'rganadi
 - b) obyektning tuzilishini o'rganadi
 - c) elementlarning murakkab tuzilishini o'rganadi.
 - d) barchasi to'g'ri
4. Original modellar...
 - a) funksional va struktur modellarni o'rganadi
 - b) funksional modellarning qanday tuzilishini o'rganadi
 - c) modellarning qanday ishlashini o'rganadi
 - d) obyektning tuzilishini o'rganadi.
5. Tilshunoslikka oid bo'lgan grammatik belgilarni, ko'rsatkichlarni va strukturalarni qolipga solish-... deyiladi.
 - a) modellashtirish
 - b) formallashtirish
 - c) identifikatsiyalash
 - d) korpuslashtirish
6. Mashina tarjimasida tizimlarining sifatli variantlarini yaratishda matn tipologiyasini o'rganish bo'yicha qaysi olimlar tadqiqot olib borgan?
 - a) V.N. Krupnova, L.L.Nelyubin, L.I.Borisova
 - b) L.N.Belyayev, L.L.Nelyubin, L.I.Borisova
 - c) L.I.Borisova, Svetan Todorov
 - d) N.Enkvist, Svetan Todorov
7. Kim matnning uch lingvistik parametri: mavzu (topic), fokus (focus) va bog'lanish (linkage) borligi haqida fikr bildirgan?
 - a) N.Enkvist
 - b) Svetan Todorov
 - c) L.N.Belyayev
 - d) L.L.Nelyubin
8. Lingvistik model tushunchasi ilmiy muomalaga struktur tilshunoslikning kimlarning asarlari orqali kirib kelgan?

- a) E.Sepir, L.Blumfild
- b) R.Yakobson, X.Chomskiy
- c) Z.Xarris, Ch.Hokket

d) barcha javob to'g'ri

9. Svetan Todorov matnning uch asosiy kategoriyasi sifatida qaysi kategoriyalarni sanab o'tadi?

- a) verbal, sintaktik hamda semantik
- b) mavzu (topic), fokus (focus) va bog'lanish (linkage)
- c) so'z, so'z birikmasi, gap, stilistik usullar
- d) harf, so'z, so'z birikmasi, gap

10. «Ilmiy konsepsiya kartina emas, balki diagrammadir» Ushbu fikr muallifi kim?

- a) L.Yelmsev
- b) Z.Xarris
- c) Ch.Hokket
- d) Svetan Todorov

13-MAVZU. KOMPYUTERDA MATNLARNI QAYTA ISHLASH TIZIMLARI (NLP)

- 13.1. Sun'iy intellekt texnologiyasi
- 13.2. Matn interfeysi va generatsiyasi
- 13.3. Qidiruv va ekspert tizimlari
- 13.4. Savol-javob tizimi
- 13.5. Nutq sintezatori
- 13.6. Tahrirlovchi dastur (Spell-checker) texnologiyasi
- 13.7. Sentiment analiz

Tayanch so'z va iboralar: *Sun'iy intellekt, axborot qidiruv tizimi, matn interfeysi va generatsiyasi, qidiruv va ekspert tizimlari, diologning kompyuterga oid modellari, savol-javob tizimi, nutq sintezatori, spell-checker (tahrirlovchi dastur) texnologiyasi, Djaro Vinkler algoritmi, sentiment analiz*

13.1. Sun'iy intellekt (SI) texnologiyasi

Sun'iy intellekt tushunchasiga manbalarda kompyuter sohasida muayyan vazifalarni (artificial intelligence - ba'zida mashina intellekti deb ham nomlanadi), mashina orqali boshqariladigan algoritmlar tushuniladi. Kaplan va Haenllein SIga quyidagicha ta'rif beradi: *sistemaning shunday qobiliyatiki, tashqi ma'lumotlarni to'g'ri interpretatsiya qiladi, ushbu ma'lumotlarni o'rganib moslashuvchan adaptatsiya orqali maxsus maqsadlar va vazifalarga erishish uchun ma'lumotlardan foydalanadi*²⁵⁰. SI katta soha hisoblanib, endilikda u turli fan tarmoqlarida qo'llanilmoqda: inson nutqini tushunish, yuqori darajadagi o'yinlarni boshqarish (shaxmat kabi), avtomatik boshqariladigan mashinalar va boshq. Kaplan va Haenllein SI tizimini uch turga ajratadi: analitik, kishini ruhlantiruvchi (human-inspired-HISI), insoniylashtirilgan SI (humanized AI-HSI). Analitik SI da faqat kognitiv intellekt bo'lib, avvalgi bilimlarni keyingi vaziyatlar uchun qo'llaydi. HISI da emotsional intellekt, tushunish, kognitiv belgilar bo'lib qaror chiqarish imkoniyatiga ega. HISI da insonga xos deyarli barcha

²⁵⁰ en.m.wikipedia.org

kompetensiyalar (kognitiv, emotsional va sotsial intellekt) mavjud bo'lib, barcha o'zaro faoliyatga nisbatan xabardor bo'lish va hal qilish imkoniyatiga ega.

Sun'iy intellekt tushunchasiga 1956-yilda asos solingan. Endilikda u kompyuter sohasi, informatsion muhandislik, matematika, psixologiya, kompyuter lingvistikasi, falsafa va boshqa ko'plab sohalarda qo'llanilmoqda.

Axborot qidiruv tizimi uzatish, saqlash, loyihalash va ma'lumotlarni kiritish kabi vazifalarni bajaradi. Bunday ma'lumotlar birligiga aniq hujjatlar, paragraflar, veb sahifalar, nutqiy matnlar, tasvirlar, rasmlar, musiqa, video va boshqalar tegishli bo'lishi mumkin (Baeza-Yates and Ribeiro-Neto, 1999). Axborot qidiruv tizimi noaniq bo'lmaganda va foydalanuvchi Veb qidiruvga to'g'ridan to'g'ri so'rovlar yuborishi mumkin bo'lganda zarur ma'lumotlar tavsifi juda qisqa bo'lar edi. So'rovlar bilan bog'liq hujjatlarni qayta ishlashda ayrim muammoli vaziyatlar sirasiga nutqdagi turli xatolar, matndagi omonim va sinonimlar kiradi. Tillarda so'zlarning deyarli barchasi polisemantik xarakterga ega. Ba'zi ma'nolar bir-biriga bog'liq bo'ladi, ba'zilari esa tubdan farq qiladi. Sinonimiya hodisasi ham ma'lumot qidiruvda muhim hisoblanadi. Chunki muayyan tushunchani bir necha so'zlar bilan ifodalanishi mumkin. Samarador qidiruv tizimi hujjatlar (yoki so'rovlar) bilan bevosita ishlamaydi. Hujjatlar yoki so'rovlarning asosiy semantik aspektini ifodalash uchun turli texnologiya va strategiyalardan foydalaniladi. Bu jarayon indekslash (indexing) deb nomlanadi. Indekslish ikki komponentdan iborat²⁵¹: 1) indeks hajmi; 2) indekslash jarayoni. Hujjatlarning semantik mundarijasini ifodalashda tashqi belgilari nashr etilgan sana, muallif nomi, sahifa raqami, kitob narhi, nashri, tili kabi matnga bog'liq ma'lumotlar kiritiladi.

Indekslish jarayoni hujjatlar uchun so'zlar, musiqa uchun musiqiy belgilar, rasmlar uchun ranglar indekslash birligi hisoblanadi. Yozma hujjatlar faqat alohida so'zlar bilan emas, balki qo'shma so'zlar yoki tezaurus guruhlaridan iborat bo'lishi mumkin. Bundan tashqari jumlar yoki so'zlar n xarakterdagi (n -gramma

²⁵¹ Handbook of Natural Language Processing (second edition). – Cambridge UK, 2010

indekslash) (McNamee and Mayfield, 2004)) segmentlarga ajratiladi. $n=4$ deb belgilansa, "white house" birikmasi to'rt-gramma tartibi bo'yicha quyidagicha bo'linadi: "whit," "hite," "ite h," "te ho," ..., va "ouse." Bunday ajratish morfologik analizatori bo'lmagan tillar uchun samarali usul hisoblanadi (Savoy, 2005). Ma'lumotlarni qidirishda uning yanada to'liqroq foydalanuvchiga taqdim qilinishida berilayotgan termin yoki so'z o'zidan oldingi yoki keyingi birikmasi bilan kelishi mumkin: ichimlik (drink) – alkogolsiz ichimlik (soft drink). Shuningdek, axborot qidiruvda so'zlarning variativligi ham inobatga olinadi: *ko'payish*, *ko'paytirish*. Orfografik variantlar ham nazardan chetda qolmaydi: *qism* – *qisim*.

Daraxtsimon sintaktik tahlil qilish gaplarni lingvistik strukturalarga ajratishda standart ifoda etish usuli sanaladi.

Matn interfeysi va matn generatsiyasiga bag'ishlangan aksariyat ishlarda jumlar o'rtasidagi o'zaro bog'lanish kalit so'zlar bilan amalga oshiriladi²⁵². Matn generatsiyasi ikki jumla o'rtasidagi yoki gap tarkibidagi qismlarni o'zaro bir-birga bog'lash jarayonidir. Bu holatda daraxtsimon sintaktik tahlil qilish usuli samarador usul bo'lib hizmat qiladi. OpenNLP [1OpenNLP 2013. <http://incubator.apache.org/opennlp/documentation/manual/opennlp.htm>] hamda Stanford NLP [2Stanford NLP. <http://nlp.stanford.edu/>] tobelangan daraxtsimon modelga asoslanadi. Generizatsiya jarayonida matnni tahlil qiluvchi dasturlar tasniflash, kategoriyalash, mundarijasini hosil qilish kabilar amalga oshiriladi²⁵³. Matn qismlarining o'zaro o'xshashliklarni aniqlashda generizatsiya jarayoni tahlil qilinadi. Matn o'rtasidagi diskursiv aloqadorlik hamda o'xshash tomonlarni aniqlash uchun jumlar o'rtasidagi sintaktik munosabatlar to'g'ri aniqlanishi kerak. Matnning daraxtsimon modelini tahlil qilishda B. Galitsky²⁵⁴ quyidagicha izoh beradi: Agar T_1 matnda ikki P_1 va P_2 sintaktik strukturalar bo'lsa, r

²⁵³ Galitsky, B. Machine Learning of Syntactic Parse Trees for Search and Classification of Text. Engineering Application of AI, <http://dx.doi.org/10.1016/j.engappai.2012.09.017>, (2012b)

²⁵⁴ Galitsky B., Ivovsky D., Kuznetsov S. Parse thicket representations of text paragraphs, «Диалог» (2013) Выпуск 12. 241 p.

o'zaro aloqa uchun: P_1 va P_2 o'rtasidagi $P_{1j} \rightarrow P_{2j}$ bo'lsa, u holda T_1 ning $P_{1, i-2}, P_{1, i-1}, P_{1, i}, P_{2, j}, \dots$ qismlari T_2 dagi bir nechta jumladagi elementlarni muayyan jumlagi moslashtirishimiz mumkin.

Ikki matn o'rtasidagi o'xshashliklarni aniqlashda quyidagi usullardan foydalaniladi:

- 1) (bag-of-words) So'zga asoslangan yondashuv umumiy kalit so'zlar/ n-grams va ularning chastotasini hisoblash;
- 2) juftlashtirib moslashtirish: bir juft gaplarni generalizatsiya qilish hamda umumiy natijasini qisqartirish;
- 3) paragraf bilan paragrafni moslashtirish.

13.2. Matn interfeysi va generatsiyasi

Matn generatsiyasi deganda, uni alohida gaplarga ajratish jarayoni tushuniladi.

Ma'lumotlarni to'plash va ularni qayta ishlashda ma'lumotlarni qayta ishlash tizimi keng qo'llanilmoqda. Ma'lumotlar bazasidan axborot qidirish uchun foydalanuvchidan so'rov tilini yaratish talab qilinadi. Tabiiy til interfeysi bo'yicha Wang (1999), ELF Software Co., (2000), Torgersson, Falkman (2002), Hallett (2005), Bertomeu (2006) kabi mutaxassislar tomonidan izlanishlar olib borilgan²⁵⁵.

Tabiiy til interfeysi ikki asosiy masalani hal qilishi kerak:

1. Tabiiy tilni so'rov tiliga o'tkazish
2. So'rov natijasida jadvaldan olingan axborotdan foydalanib natijalarni generatsiya qilinadi.

Tabiiy til interfeysining asosiy vazifasi tabiiy til so'rovini yaratish hamda qisqa javob shaklida foydalanuvchilarga uzatishdan iborat. Matn generatori tizimlari uchun quyidagi ma'lumotlar bazasi kerak bo'ladi: SQL so'rov tili, so'zlar lug'ati, taurus, sintaktik va morfologik qoidalar. Agar berilayotgan savolda quyidagi *solishtirish, ro'yxatlash va baholash* (Compare, List, and Evaluate²⁵⁶) kalit so'zlari topilsa, berilgan savollarga javob berish jarayoni amalga oshiriladi. Generatsiya qilinadigan matnlarning freymlarni ajratish uchun savol turlarni aniqlash va tasniflash zarur.

Har bir freym savol turiga qarab nomlanish tavsiya qilinadi: *set frame, list frame* kabi²⁵⁷. Mutaxassislar matn generatsiya uchun erkin kontekstli grammatikadan foydalanishni tavsiya qiladilar. Shunda ajratilgan freymlar terminal va noterminal belgilarga ko'ra tasniflanadi. Huong Thanh Le ushbu freymlarga qo'shimcha ravishda tasvir freymidan ham foydalanishi haqida to'xtalib o'tadi. Bunda matn generatsiyasi uchun olingan so'zlarning turli atributlari o'rtasidagi munosabatlar va har bir freymdagi so'zlarning barcha atributlarini qamrab olinishi kerak. Masalan, tadqiqotchi "*talaba - tbl_student*" so'ziga birlashtirilgan munosabatlarni quyidagicha ifodalaydi:

(1) $[frame_student] \text{ \AE } [studname] ([sex]) [vp_studId]. [studname] [vp_DOB], [vp_cityborn].$

$[studname] [vp_classId].$

(2) $[frame_student] \text{ \AE } [studname] ([sex]) [vp_studId]. [studname] [vp_classId]. [studname]$

$[vp_DOB], [vp_cityborn].$

(3) $[vp_studId] \text{ \AE } has\ the\ student\ code\ [studId]$

(4) $[vp_DOB] \text{ \AE } was\ born\ on\ [DOB]$

(5) $[vp_cityborn] \text{ \AE } in\ [cityborn]$

(6) $[vp_classId] \text{ \AE } is\ a\ student\ of\ the\ class\ [classId]$

(7) $[vp_classId] \text{ \AE } studies\ in\ the\ class\ [classId]$

[] qavsda berilgan qoidalar to'plami, satrlar non-terminal belgilar, qolganlari esa terminal belgilar hisoblanadi.

Matn generatsiya uchun Joyce tizimi Ulusses dasturiy ta'minotining muhitining interfeysi sifatida yaratilgan. Ushbu generator matn morfologiyasi jarayonidan olingan natijalarga asoslanadi. Joyce tizimi bir-biridan alohida tuzilgan quyidagi modullardan iborat²⁵⁸:

- matnni loyihalash

- jummalarni loyihalash

- lingvistik realizator – til leksikonida aks etgan morfologik va sintaktik bilimlarni analiz qilish.

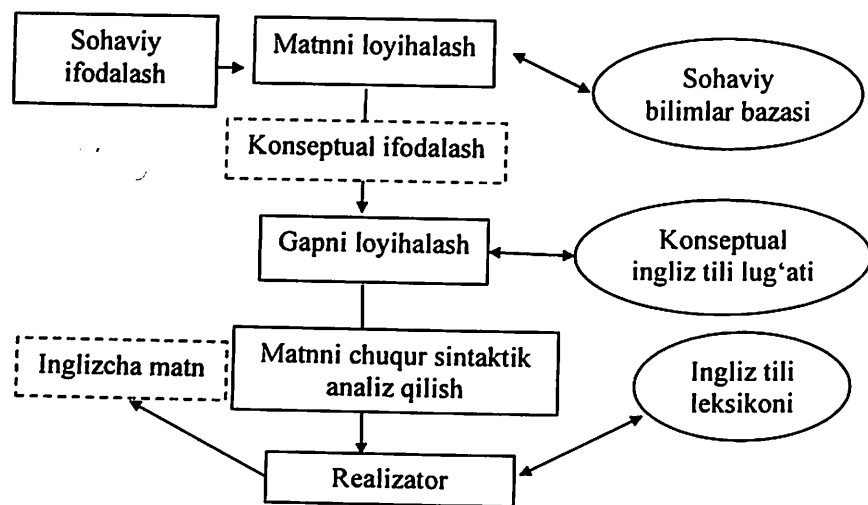
²⁵⁵ Huong Thanh Le A Frame-based Approach to Text Generation / <http://www.aclweb.org/anthology/Y07-1019>

²⁵⁶ Huong Thanh Le A Frame-based Approach to Text Generation / <http://www.aclweb.org/anthology/Y07-1019>

²⁵⁷ O'sha joyda.

²⁵⁸ Owen Rambow, Tanya Korelsky Applied Text Generation / <http://aclweb.org/anthology/A92-1006>

Ushbu komponentni olimlar quyidagicha tasvirlaydi (17-chizma)²⁵⁹:



17-chizma

13.3. Qidiruv va ekspert tizimlari

Qidiruv tizimi dastlab Archie deb nomlangan bo'lib, Arxiv (archive so'zidagi vee tushib qolgan) demakdir. U 1990-yil Kanadaning Montreal shahrida joylashgan McGill universiteti talabasi Alan Emtage tomonidan yaratilgan. Ammo u indekslanmagani, qolaversa, ma'lumot hajmi cheklangan bo'lgani uchun uncha samarali bo'lmagan.

Qidiruv tizimida 1994-yil Carnegia Mellon universiteti tomonidan yaratilgan WebCrawler dasturining yaratilishi tizimning keng yoyilishiga sabab bo'ldi. U avtomatik ravishda kiritilgan matndagi barcha so'zni indeks sifatida qabul qilib, kerakli vaziyatda saralab bera olgan.

Izlash tizimini cho'qqiga ko'targan loyiha 2000-yil Google kompaniyasining so'nggi asoschilari Sergey Brin va Larry Pagening ilmiy ishlariga asoslangan "Qidirish tizimi anatomiyasi" (Anatomy of a Search engine) nomli PageRank dasturi bo'ldi. U boshqa saytlardagi ma'lumotlarni skayner qilib oladi va tartiblab foydalanuvchiga uzatadi. Googlening mashhur bo'lib ketishiga

uning raqobatchilardan farqli ravishda *Mystery seeker* (sirli ravishda qidirish) tizimi ham sabab bo'ldi.

Qidiruv tizimi – veb-interfeysli dasturiy apparat majmuasi hisoblanadi. U orqali internet tizimidan muayyan kalit so'zlar orqali ma'lumot qidiriladi. Qidiruv tizimi deyilganda, odatda, ma'lum bir sayt nazarda tutiladi. Axborot qidiruv tizimlari internet resurslari ichidan so'ralgan ixtiyoriy ma'lumotni qidirib topish imkoniyatiga ega. Qidiruv tizimi informatsion ma'lumot berish tizimidan (Information retrieval system) kelib chiqqan. Bu sistema ma'lumotlar bazasi uchun kalit so'zlar indeksini yaratadi, natijada ushbu kalit so'z qatnashgan qidiruv so'rovnomasiga javoban natijalar ro'yxatini ko'rsatish imkoni paydo bo'ladi. Axborot qidiruv tizimi axborot turi, manbai va ular olish usuliga ko'ra farqlanadi.

Qidiruv mashinasi 5 qismdan iborat: 1) veb-sahifalarni server dasturiga yuklash; 2) yuklangan veb-sahifalarni saralash; 3) saralangan ma'lumotlarni tartiblash; 4) to'plangan ma'lumotlarni server ma'lumotlar bazasiga joylashtirish; 5) foydalanuvchi va server bilan qidiruv uchun aloqa vositasida sifatida ishtirok etish. Qidiruv tizimlariga Google, Altavista, Yandex, Rambler, Aport, Milliy kabilar kiradi.

Axborot qidiruvi quyidagi xoslikka ega:

- berilgan ma'lumotning to'liqligi
 - ma'lumotning aniqligi
 - ma'lumotning g'oyib bo'lishi
 - information shovqin
- borot turi

Ekspert tizimi biror soha doirasida biror masalani yechish yoki tavsiya berish uchun muayyan bilimga asoslangan kompyuter dasturi hisoblanadi. U sun'iy intellektning o'rganish sohasidir. NLPning ushbu yo'nalishi *bilimlar injeneriyasi* yoki *bilimga asoslangan tizim* deb ham ataladi.

Ekspert tizimlarining quyidagi vazifalari mavjud²⁶⁰:

1. Ma'lumotlarni (geofizik o'zgarishlar) interpretatsiya qilish;
2. Ko'p funksional vazifalarni diagnoz qilish (biror asbobdagi xatolik yoki insonlardagi kasallik);

²⁵⁹ O'sha joyda. –P. 42

²⁶⁰ <http://www.cs.bham.ac.uk/~jxb/IAI/w9.pdf>

3. Murakkab obyektlarning konfiguratsiyasi yoki struktural analizi (kimyoviy tarkib yoki kompyuter tizimlari);

4. Harakat ketma-ketligini rejalashtirish (robotning xatti-harakati);

5. Kelajakni bashorat qilish (ob-havo, narh-navo, iqlim va h.k)

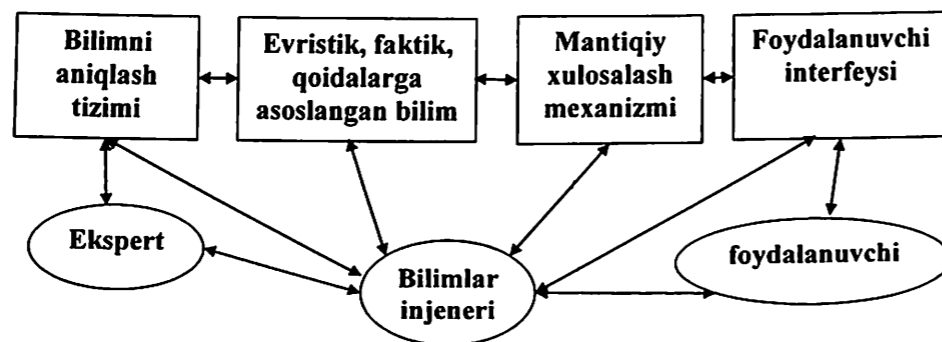
Lisp, Prolog va Miranda kabilar ushbu ekspert tizimini yaratishga qulay dasturlash tillari sanaladi. Negaki qoidalarning sintaktik strukturalari Prologning mantiqiy xulosalash modeliga muvofiq keladi.

Ekspert tizimi ikki bazaga asoslanadi:

-ma'lumotlarga asoslangan yondashuv (bunda ma'lumotlardagi kalit so'zlar orqali) javob qidiriladi;

-bilimlarga asoslangan yondashuv.

Ekspert tizimining arxitekturasi quyidagicha bo'ladi²⁶¹(18-chizma):



18-chizma

13.4. Savol-javob tizimi

Globalashuv davrida ma'lumotlarni tez qidirish, uzatish va ularni tahlil qilish juda muhim jarayon hisoblanadi. Qachondir afsona bo'lib tuyulgan tasavvurlar bugun katta ilmiy tekshirish institutlarda ro'yobga chiqmoqda. Ular hayotimizga shu qadar tez kirib boryaptiki, yangi yaratilgan kashfiyot ertaga yangilik bo'lmay qolayotgani tez sezilmoqda. Ertaklarda o'zi uchar gilamlar bugun turli uchish transportlariga, o'zi suhbatlashadigan ko'zgulalar bugun

²⁶¹ O'sha joyda.

chatbotlarga aylanib ulgurdi. Chatbotlarning ishlash texnologiyasi savol-javob va diolog tizimiga asoslangan.

Savol-javob tizimi (SJT) bo'yicha dunyo bo'yicha ko'p tadqiqotlar olib borilgan. Dastlabki Eliza nomli SJT 1964-yili Jozef Weizenbaum tomonidan yaratilgan²⁶². Manbalarda SJT umumiy ishlash texnologiyasi ikki xil usul bilan amalga oshiriladi: 1) axborot qidiruviga asoslangan SJT; 2) bilimga asoslangan SJT.

Savol-javob tizimi quyidagi bosqichlardan iborat²⁶³: 1) sahifalarni yuklab olish; 2) sahifalardagi linklarni aniqlash va ularni yuklab olish; 3) sahifalardagi kalit so'zlarni aniqlash; 4) har bir kalit so'zning indeks raqamini qo'yish; 5) har bir so'z va sahifani indekslash. Olimning fikricha, bunday axborot qidiruvga asoslangan texnologiya hind-yevropa tillariga nisbatan turkiy tillar uchun biroz murakkab sanaladi. Qidirilayotgan so'zlarga qo'shiladigan bir nechta affiks va prefikslar qo'shilishi natijasida ularning asosiy o'zak qismi aniqlanishi lozim bo'ladi. Shu bois ushbu tizim uchun morfologik analiz juda muhim hisoblanadi.

SJTda savollarning strukturasi va mavzu ko'lami turlicha bo'ladi: qisqa yoki uzun, tematik yoki ixtiyoriy, mantiqiy yoki faktik. Gaplarning turi ham turlicha bo'lishi mumkin: ritorik, maxsus so'roq gap, umumiy so'roq gap. Agar berilayotgan savolning mavzu ko'lami kengroq bo'lsa, qidiruv tizimidagi kalit so'zlarning ehtimollik holati ham kengroq bo'ladi. E.Adali javobga asoslanuvchi quyidagi lingvistik faktorlarni keltirib o'tadi:

- Kishi nomlari
- Qisqartmalar
- Institutlar
- Joy-makon (davlat, shahar)
- Narsa nomlari (uy, oziq-ovqat)
- Vaqt (asr, yil, oy, davr)
- Sonli qiymatlar

²⁶² J. Weizenbaum, ELIZA A Computer Program For the Study of Natural Language Communication Between Man And Machine MIT Cambridge, Communication of the ACM Volume 9 / Number 1 / January, 1966

²⁶³ Adali E. Question and answering system for turkic languages / Turklang.2018. Tashkent, 2018. -P. 6.

Bilinga asoslangan SJT esa o'zining ma'lumotlar bazasiga ega bo'ladi²⁶⁴.

- **Ob-havo ma'lumotlari**

- **Diyin natijalari**

- Mehmonxonada xonaga buyurtma berish
- Chiptaga buyurtma berish
- Elektron tijorat ilovalari

Savol-javob tizimining asosiy vazifasi kalit so'zlar yordamida axborot qidiruviga berilgan so'rovni kengaytirishdan iborat. Ayrim tizimlarda beriladigan javoblar quyidagi ma'lumotlarga ko'ra kengaytiriladi:

- javob turi– javobdagi birlik turi (shaxs, makon, vaqt va bshq.);
- markaz – savoldagi so'zlar satri biror topilgan javobdagi satr

orqali o'rniga qo'yiladi;

- savol turi– beriladigan savollarni aniqlash.

Bilinga asoslangan savol-javob tizimi semantik analiz orqali amalga oshiriladi.

Ko'p diologlar monolog matnga oid diskursga o'xshaydi. Monolog matnlardan biroz farqlansa-da, diologlar ham anafora va diskursning strukturaviyligi va manqiyiligini ham ko'rsatadi²⁶⁵. Diologda matn navbatlanib keladi. Masalan, kutubxonachi (A) va talaba (B) o'rtasidagi diologni kuzatamiz – inson-inson diologi:

A1: Xo'sh, keling. Nima xizmat?

B1: Menga ayrim kitoblar kerak edi...ilmiy kitoblar.

A2: Yaxshi. Qaysi fan bo'yicha?

B2: Tilshunoslikka oid.

A3: Katalogdan kerakli kitoblarni tanlang va raqamlarini yozing.

Keyin buyurtma qilishingiz mumkin.

B3: Xo'p, rahmat.

Diologning murakkab tomoni shundaki, unda inversiya hodisasi yoki savolga nisbatan so'zlovchi va tinglovchi o'rtasida noverbal vositalarning ishtiroki, qolaversa, pragmatik unsurlarning bevosita ishtirokini kuzatish mumkin. Inson-inson o'rtasidagi muloqotni

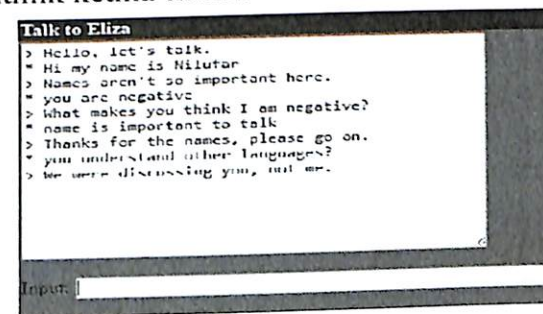
²⁶⁴ O'sha joyda. –P. 10.

²⁶⁵ Daniel Jurafsky and James H. Martin An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition 1999 . -P. 716

tushunish qiyin emas. Negaki nazarda tutilayotgan mavzu har ikki tomonga aloqador bo'ladi.

Diologda qatnashayotgan ishtirokchilar kimdan so'ng yoki qachon gapirishlarini biladi. Bu holat muloqot tahlili muloqot analizida (Conversation Analysis – CA) o'rganiladi²⁶⁶. Biroq inson va robot o'rtasidagi muloqotni ta'minlashda kompyuter lingvistikasi oldida bir qator masalar hal qilinishi kerak.

Quyidagi saytlar orqali <http://www.manifestation.com/neurotoys/eliza.php3>, <https://www.cyberpsych.org/eliza/#.WZ19toiGNEY>, <http://psych.fullerton.edu/mbirnbaum/psych101/Eliza.htm> Eliza bilan suhbat qurish mumkin. Eliza ilovasi bu mantiqiy fikrlar asosiga qurilgan algoritmik ketma-ketlik.



13.5. Nutq sintezatori

Insonning artikulyatsion apparati nutq tovushlarini hosil qiladi. Nutq apparati orqali biz fikrimizni ifodalashimiz uchun zarur bo'lgan nutq tovushlarini talaffuz qilamiz. Har bir tovush akustik xususiyatga ega: nutqiy oqim (tezligi, baland-pastligi). Nutq tovushlari ma'no farqlash, ma'no ajratish belgilariga ega. Shu boisdan ham ular biologik va fizik tovushlardan farqlanadi. Muayyan tovushlarning tartibli ketma-ketligi muayyan ma'noga va funksional mazmunga ega so'zni hosil qiladi. Ushbu texnologiya nutqni paydo bo'lish jarayoniga ko'ra fizika bilan, uni maqsadli yo'naltirish jarayoni psixologiyaga bog'liq. Endilikda nutq sintezatoridan foydalanilayotgan Google qidiruv tizimi yoki muloqot roboti (chatbox) akustik signallar asosida ishlamoqda. Tillarning

²⁶⁶ Yuqoridagi asar - P.718.

Bilimga asoslangan SJT esa o'zining ma'lumotlar bazasiga ega bo'ladi²⁶⁴.

- Ob-havo ma'lumotlari
- O'yin natijalari
- Mehmonxonada xonaga buyurtma berish
- Chiptaga buyurtma berish
- Elektron tijorat ilovalari

Savol-javob tizimining asosiy vazifasi kalit so'zlar yordamida axborot qidiruviga berilgan so'rovni kengaytirishdan iborat. Ayrim tizimlarda beriladigan javoblar quyidagi ma'lumotlarga ko'ra kengaytiriladi:

- javob turi– javobdagi birlik turi (shaxs, makon, vaqt va bshq.);
- markaz – savoldagi so'zlar satri biror topilgan javobdagi satr orqali o'rniga qo'yiladi;
- savol turi– beriladigan savollarni aniqlash.

Bilimga asoslangan savol-javob tizimi semantik analiz orqali amalga oshiriladi.

Ko'p diologlar monolog matnga oid diskursga o'xshaydi. Monolog matnlardan biroz farqlansa-da, diologlar ham anafora va diskursning strukturaviyligi va manqiyiligini ham ko'rsatadi²⁶⁵. Diologda matn navbatlanib keladi. Masalan, kutubxonachi (A) va talaba (B) o'rtasidagi diologni kuzatamiz – inson-inson diologi:

A1: Xo'sh, keling. Nima xizmat?

B1: Menga ayrim kitoblar kerak edi...ilmiy kitoblar.

A2: Yaxshi. Qaysi fan bo'yicha?

B2: Tilshunoslikka oid.

A3: Katalogdan kerakli kitoblarni tanlang va raqamlarini yozing.

Keyin buyurtma qilishingiz mumkin.

B3: Xo'p, rahmat.

Diologning murakkab tomoni shundaki, unda inversiya hodisasi yoki savolga nisbatan so'zlovchi va tinglovchi o'rtasida noverbal vositalarning ishtiroki, qolaversa, pragmatik unsurlarning bevosita ishtirokini kuzatish mumkin. Inson-inson o'rtasidagi muloqotni

²⁶⁴ O'sha joyda. –P. 10.

²⁶⁵ Daniel Jurafsky and James H. Martin An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition 1999 . -P. 716

tushunish qiyin emas. Negaki nazarda tutilayotgan mavzu har ikki tomonga aloqador bo'ladi.

Diologda qatnashayotgan ishtirokchilar kimdan so'ng yoki qachon gapirishlarini biladi. Bu holat muloqot tahlili muloqot analizida (Conversation Analysis – CA) o'rganiladi²⁶⁶. Biroq inson va robot o'rtasidagi muloqotni ta'minlashda kompyuter lingvistikasi oldida bir qator masalar hal qilinishi kerak.

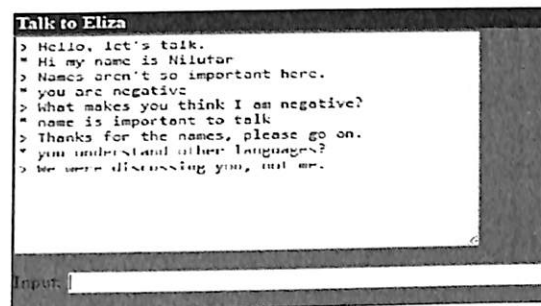
Quyidagi saytlar orqali

<http://www.manifestation.com/neurotoys/eliza.php3>,

<https://www.cyberpsych.org/eliza/#.WZ19toiGNEY>,

<http://psych.fullerton.edu/mbirnbaum/psych101/Eliza.htm> Eliza

bilan suhbat qurish mumkin. Eliza ilovasi bu mantiqiy fikrlar asosiga qurilgan algoritmik ketma-ketlik.



13.5. Nutq sintezatori

Insonning artikulyatsion apparati nutq tovushlarini hosil qiladi. Nutq apparati orqali biz fikrimizni ifodalashimiz uchun zarur bo'lgan nutq tovushlarini talaffuz qilamiz. Har bir tovush akustik xususiyatga ega: nutqiy oqim (tezligi, baland-pastligi). Nutq tovushlari ma'no farqlash, ma'no ajratish belgilariga ega. Shu boisdan ham ular biologik va fizik tovushlardan farqlanadi. Muayyan tovushlarning tartibli ketma-ketligi muayyan ma'noga va funksional mazmunga ega so'zni hosil qiladi. Ushbu texnologiya nutqni paydo bo'lish jarayoniga ko'ra fizika bilan, uni maqsadli yo'naltirish jarayoni psixologiyaga bog'liq. Endilikda nutq sintezatoridan foydalanilayotgan Google qidiruv tizimi yoki muloqot roboti (chatbox) akustik signallar asosida ishlamoqda. Tillarning

²⁶⁶ Yuqoridagi asar - P.718.

turli-tuman xarakterga egaligi nutq sintezatorini o'rganish usullarini o'zgartiradi.

G.Fant, G.Xall, R.Yakobson kabi tilshunoslar tomonidan artikulyatsion belgilariga ko'ra universal tovushlar tasnifi yaratilgan.

Nutqni tanuvchi dastur ilk bor 1952-yilda paydo bo'lgan. U inson tomonidan talaffuz qilingan raqamlarni tanish imkoniyatiga ega bo'lgan. Endilikda nutqni tanuvchi dastur uchun bir nechta metodlar mavjud. Undan ham oldinroq Pushkinning "Yevgeniy Onegin"da qatnashgan so'zlarning unli va undosh harflarining birin-ketin kelish ehtimolini 1913-yilda Markov tadqiq qilgan. Bu yashirin markov modeli deb nomlanadi.

Dunyoda turli tillarga mo'ljallangan "text-to-speech" va "speech to text" texnologiyasini sun'iy intellekt texnologiyalari va RNN algoritmi, NLP texnologiyalari yaratilgan hamda turli texnologiyalarga asoslangan platformalar ham mavjud (<https://github.com/CSTR-Edinburgh/merlin>, <https://github.com/google/voice-builder>).

O'zbekcha nutq-generatorini yaratish o'zbek tili texnologiyasi uchun muhim masalalardan biridir. Jahon til texnologiyasi va NLP sohasida bu kabi chuqur o'rganilgan tadqiqotlar amalga oshirilgan bo'lib, ular ovozli matn korpusi sifatida matnlarni qayta ishlash texnologiyasi uchun ham foydalaniladi:

1. Text To Speech with Tacotron-2 and FastSpeech using ESPnet
2. Natural TTS Synthesis by Conditioning WaveNet on Mel Spectrogram Predictions
3. VoiceLoop: Voice Fitting and Synthesis via A Phonological Loop
4. Deep Voice 3: Scaling Text-to-speech With Convolutional Sequence Learning
5. Parallel WaveNet: Fast High-Fidelity Speech Synthesis

O'zbekcha matnlarni ovozlashtirish va ovozli ma'lumotlarni matnga o'tkazish o'ziga xos izchil tadqiqotlar olib borishni, kompyuter texnologiyalari, sun'iy intellekt texnologiyasi hamda lingvistik resurslarning mashina fondini yaratishni taqozo etadi.

Nutqni tushunuvchi texnologiyasini yaratish o'zbek tili uchun ilmiy va texnik muhim masalalardan sanaladi. Nutq sintezatori uch yo'nalishda ishlash funksiyalariga ega: *nutqdan-> matnga*,

matndan-> nutqqa, *nutqdan-> nutqqa*. Ko'rish orqali ko'p ma'lumotni qabul qilamiz. Natijada vaqtdan unumli foydalanish hamda ko'zni toliqtirmaslik maqsadida eshitish orqali zarur bo'lgan ma'lumotni olishga harakat qilamiz. Shu bois nutq sintezatori muhim ahamiyatga ega. Ushbu texnologiya nafaqat matnga bog'liq, balki informatikaning boshqa amaliy yo'nalishlarida ham qo'llanilmoqda: robot texnologiyasi, axborot qidiruvi tizimi va boshqalar.

Nutqni tushunuvchi dasturlar uch asosiy tamoyilga bo'ysunadi: *avtomat faqat belgilangan kishilarning nutqini tanish xususiyatiga ega (bank raqamlari faqat uning egasi tovushi bilan ochiladi); avtomat faqat cheklangan so'zlarni taniydi; barcha tovushlarni tanish xususiyatiga ega.*

Bu yo'nalishda turli darajada ilmiy-amaliy izlanishlar olib borilmoqda. Nutq sintezatori uchun uning ma'lumotlar bazasini yaratish uchun quyidagi yondashuvlarga tayaniladi: 1. So'z va uning shakllariga asoslangan texnologiya. Bunda qo'shimchalari kam yoki so'zlarga umuman qo'shilmaydiga tillar uchun mos keladi (flektiv tillar); 2. Bo'g'inlarga asoslangan texnologiya. Bunda fonotaktik qoidalar va mantiqiy modellarga asoslanib, affiksial morfemalar ko'p qo'llaniladigan tillar uchun mos keladi (aglyutinativ tillar).

O'zbek tili aglyutinativ tillar sirasiga kirgani uchun ikkinchi texnologiya mos kelishi mumkin. Bunda so'zlarni hosil qiluvchi fonetik tarkibli modellar tuziladi. Nutqning boshqa bo'laklarga bo'linmaydigan eng kichik birligi tovush hisoblanadi. Bir havo zarb bilan aytiladigan tovush yoki tovushlar yig'indisidan bo'g'in hosil bo'ladi. Ular so'zni hosil qiladi. Nutq tovushlarining muayyan ketma-ketlikdagi ifoda plani fonetik so'z hisoblanadi. Ikki kichik pauza orasiga joylashgan va yagona urg'u bilan aytiladigan bir yoki bir nechta bo'g'inning yig'indisi takt deyiladi. Har ikki tomondan chuqur pauza bilan ajratiladigan fonetik birlikka fraza (jumla) deyiladi. Fraza fonetik birlik hisoblanadi. So'zda nechta unli bo'lsa, bo'g'in miqdori ham shuncha bo'ladi. *Mak-tab* (2 ta unli 2 bo'g'in), *ma-da-ni-yat* (4 ta unli 4 bo'g'in). Bo'g'in tuzilishi jihatidan quyidagi tiplarga ajratiladi:

1. *Bir unlidan iborat: o-ta, o-na, a-ka.*
2. *Bir unli va bir undoshdan iborat:*

a) unli + undosh – ot, ish, ol

b) undosh + unli – lo-la, do-na, to-za

3. Bir unli va ikki undoshdan iborat:

a) undosh + unli + undosh: bosh-liq, dur

b) unli + ikki undosh: aql, ilm, erk

v) ikki undosh + unli: fra-za, sta-kan

4. Bir unli va uch undoshdan iborat:

a) undosh + unli + ikki undosh: baxt, qirq, dard

b) ikki undosh + unli + undosh: stol, trak-tor, xlor

5. Bir unli va 4 undoshdan iborat:

a) ikki undosh + unli + ikki undosh: sport, shrift

b) undosh + unli + uch undosh: punkt, tekst

v) uch undosh + unli + undosh: skver, shtraf, skrip-ka

Bo'g'inning CCV, CCVC, CCVCC, CVCCC, CCCVC kabi turlari, asosan, rus tilidan o'zlashgan so'zlarda uchraydi. Turkiy tilga xos sof o'zbekcha so'zlarda bir tovushdan 4 tovushgacha bo'lgan bo'g'in tiplari mavjud. Bo'g'in unli yoki undosh fonema bilan tugashiga ko'ra 2 xil bo'ladi: unli bilan tugasa, ochiq bo'g'in: bo-la, lo-la; undosh bilan tugasa, yopiq bo'g'in: maktab, dehqon.

Matnni ovozlashtirishda bir nechta muammolar ko'rinadi:

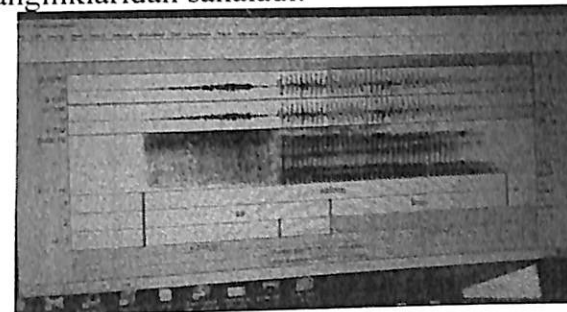
- ovoz aniqligi
- tovushlarning talaffuzi
- ovoz tezligi va tembri
- orfografik holati
- talaffuzida biroz farqlanuvchi so'zlar (omofonlar)
- omograflar

Matnni ovozlashtirish uchun ritm, urg'u, to'xtam ahamiyatga ega. Grafemalarni akustik holatga o'tkazishda statistik metod samarali usuldir. Lingvistik annotatsiyalangan ma'lumot xalqaro fonetik alifbo bo'yicha transliteratsiya qilinadi:

No	Harflar	IPA	namuna
1.	B	b	bahor
2.	D	d	do'st
3.	F	f	Fil
4.	g	g	go'sht
5.	H	h	hukumat
6.	J	j	jasur
7.	K	k	kerak
8.	L	l	mamlakat
9.	L	ʃ	quloq

10.	M	m	maskan
11.	N	n	Non
12.	Ng	ŋ	Tong
13.	P	p	posbon
14.	Q	q	quloq
15.	R	r	ravon
16.	g'	ɣ	g'ildirak
17.	S	s	Sabr
18.	Sh	ʃ	shamol
19.	T	t	to'sqin
20.	Ch	tʃ	chol
21.	V	v	voris
22.	X	x	xurlik
23.	Z	z	zina
24.	Ts	ts	animatsiya

Endilikda nutq sintezatori kompyuter lingvistikasining boshqa yo'nalishlari: mashina tarjimasi, axborot qidiruvi tizimlarida ham qo'llanilmoqda. Mashina tarjimasining ovozi dasturi yoki ovoz orqali ma'lumotlarni qidirish kompyuter lingvistikasining innovatsion yangiliklaridan sanaladi.



36-rasm

Praat dasturi yordamida audiofayllarni matnga o'tkazish va korpus uchun annotatsiyalash imkonini beradi (36-rasm). Bunda har bir talaffuz qilingan so'zlarning intonatsiyasi va bo'g'inlarga ko'ra segmentlarga ajratadi. Umid qilamizki, akustik fonetika ham kompyuter lingvistikasining istiqbolli masalalardan biri sifatida rivojlanadi.

13.6. Tahrirlovchi dastur (Spell-checker) texnologiyasi

Matnlarni kompyuterda analiz qilishda kompyuter texnologiyalarning rivojlanishi va ularning kun sari takomillashtirilyotganligi insonlar uchun har tamonlanma qulayliklar yaratmoqda. Matnlarni orfografik tahrirlovchi dastur bugungi kunda

nafaqat Offis dasturlari uchun, balki matn bilan bog'liq boshqa tizimlarda ham qo'llanilmoqda: mashina tarjimasini, axborot qidiruvi, veb-brauzerlar, veb ilovalari. Sun'iy intellekt texnologiyasining yaratilishi bir qator amaliy masalalarning yechimini topayotgan davrda hanuz matn bilan bog'liq ayrim masalalar ochiqlikligicha qolmoqda. Endilikda matnlarni tahrirlovchi dasturlar quyidagi funksional imkoniyatlarga ega (37-rasm):

- matn tarkibida tushirib qoldirilgan yoki ortiqcha qo'llanilgan so'z yoki iboralarni aniqlash;

- imloviy xatolarni aniqlash

- uslubiy xato qo'llanilgan so'zlarni aniqlash

- matnning sintaktik strukturasi xatolarni aniqlash

- grammatik jihatdan xato qo'llanilgan shakllarni aniqlash

- so'zlarning sinonimlarini tavsiya qilish orqali semantik imkoniyatini oshirish.

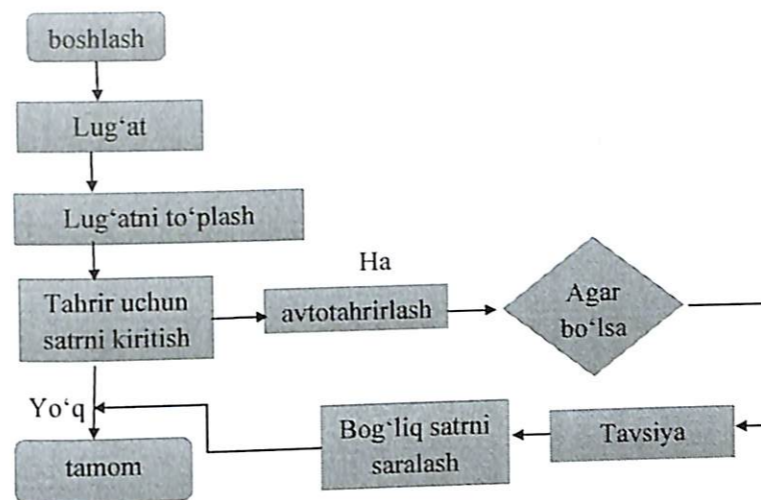
Ixtiyoriy tahrirlovchi dasturning algoritmi quyidagi bosqichlardan iborat bo'ladi:

1. Matn tarkibidagi so'zlarni kompyuter o'qib oladi;

2. Har bir so'z lug'atdagi imlosi bilan solishtiriladi. Leksik va grammatik atributlariga ko'ra so'zlarning lug'ati tuziladi;

3. Ajratilgan segment birliklar morfologik va sintaktik lingvistik me'yorlarga ko'ra tekshiriladi.

Tahrirlovchi dastur algoritmining blok-sxemasini quyidagicha ifodalash mumkin²⁶⁷(19-chizma):



19-chizma

Matnlarni tahrirlovchi dastur turk, tatar, qozoq, qirg'iz tillari uchun yaratilgan. O'zbek tili uchun tahrirlovchi dasturning lingvistik ta'minotini yaratish bo'yicha ham bir nechta tadqiqotlar olib borilmoqda. Tahrirlovchi dasturning ta'minotini yaratish bo'yicha quyidagi texnologiyalar mavjud:

-Djaro Vinkler algoritmi orqali ehtimollik nazariyaga asoslangan texnologiya;

-sof lingvistik bazaga asoslangan texnologiya;

-HFST (finite state transducer) algoritmiga asoslangan texnologiya.

-Levenshteynning oraliq masofa nazariyasi

Djaro Vinkler algoritmi

Ushbu metrika ikki o'xshash satrni o'zaro solishtirish uchun qo'llaniladi, mazkur algoritmda qo'shish, tushirib qoldirish va o'rniga qo'yish amallari bajariladi.

Ikkita satr quyidagicha hisoblangan:

$$d_j = \begin{cases} 0 & m \text{ bo'lganda} = 0 \\ \frac{1}{3} \left(\frac{m}{|s_1|} + \frac{m}{|s_2|} + \frac{m-t}{m} \right) & \text{qolgan holatlarda} \end{cases}$$

qayerda:

$|s_i|$ - s_i satr uzunligi;

m - o'xshash simvollar miqdori;

²⁶⁷ Vibhakti V. Bhaire, Ashiki A. Jadhav, Pradnya A. Pashte, Mr. Magdum P.G SPELL CHECKER / International Journal of Scientific and Research Publications, Volume 5, Issue 4, April 2015 1 ISSN 2250-3153

t – transpozitsiyalar miqdorining yarmi.

Agar o'xshash bo'lsa va $\lceil \frac{\max(|s_1|, |s_2|)}{2} \rceil - 1$ dan oshib ketmasa, s_1 va s_2 satrlardan ikkita simvol olinadi. Satrdagi har bir c_1 dagi simvol c_2 dagi simvollar bilan solishtiriladi. Turli holatdagi o'zaro o'xshash raqamlar 2 ga bo'linadi, bu transpozitsiya raqami hisoblanadi. Masalan, **ODAMLAR** va **OMADLAR** so'zini solishtiradigan bo'lsak, har ikki satrning uzunligi $m=7$ ga teng. Biroq undagi M va D simvollarining o'zni turlicha holatda kelgan. Bundan O va A bir xil holatda joylashgan. Chunki D va M ning holati 1 dan kattaroq. Shuning uchun bularni o'xshash simvollar, deb ayta olmaymiz.

Vinkler masofasi (Distance of Djaro-Winkler)

Djaro Vinkler masofasida p o'lchov koeffitsienti bo'lib, u satr boshidan prefix deb nomlangan l uzunligacha bo'lgan bir-biriga o'xshash satrlar uchun aniqlangan o'lchov birligi hisoblanadi. Djaro Vinklerda kamida ikkita satr solishtiriladi: $d_w = d_j + (lp(1 - d_j))$

$d_j - s_1$ va s_2 satrlar uchun Djaro masofasi

l – satr boshidan 4-simvolgacha bo'lgan umumiy prefiks

p – o'lchov koeffitsienti. U 0,25 dan katta bo'lmashligi kerak. Unga teskari holatlarda orasidagi masofa 1 dan katta bo'lishi mumkin. Lekin u 0 va 1 orasida bo'lishi kerak. Agar masofasi 0 ga teng kelsa, ikki satr o'rtasida umuman o'xshashlik yo'qligini, agar 1 ga teng kelsa, ikki satr bir xil ekanligini bildiradi. Vinklarning ishlarida 0 va 1 koeffitsient birligida hisoblanadi. Biz ham ushbu qiymatni qabul qilamiz:

TAROQ va **TAYOQ** so'zlarni solishtirsak, quyidagi jadval hosil bo'ladi:

	T	A	R	O	Q
T	1	0	0	0	0
A	0	1	0	0	0
Y	0	0	0	0	0
O	0	0	0	1	0
Q	0	0	0	0	1

O'xshash simvollar uchun 1, qolgan holatlar uchun 0 qo'yiladi. Birinchi so'zning har bir simvoli ikkinchi so'zning barcha simvollar bilan solishtiriladi. 1-so'zning qiymati $m=4$. Bir-biriga mos tushmagan simvollarga $t=0$ yoziladi.

$$d_j = \frac{1}{3} \left(\frac{4}{5} + \frac{4}{5} + \frac{4}{4} \right) = 0.8(6)$$

Ikki so'z va Djaro Vinkler masofasi uchun $l=2$:

$$d_w = 0.8(6) + (2 \cdot 0.1 \cdot (1 - 0.8(6))) = 0.89(3)$$

Agar Djaro Vinkler masofasi va Djaro prefiksi o'xshash bo'lmagan ikki so'zni solishtirsak ular bir xilligini ko'rishimiz mumkin.

	T	A	R	A	M	O	Q
Q	0	0	0	0	0	0	0
A	0	1	0	0	0	0	0
R	0	0	1	0	0	0	0
A	0	0	0	1	0	0	0
M	0	0	0	0	1	0	0
O	0	0	0	0	0	1	0
Q	0	0	0	0	0	0	1

Yuqorida aks etgan TARAMOQ va QARAMOQ solishtirsak: $m=6$, $t=0$, $|s_1|=|s_2|=7$:

$$d_j = \frac{1}{3} \left(\frac{6}{7} + \frac{6}{7} + \frac{6}{6} \right) = 0.9047$$

$l=0$, $p=0.1$:

$$d_w = 0.9047 + (2 \cdot 0.1 \cdot (1 - 0.9047)) = 0.92376$$

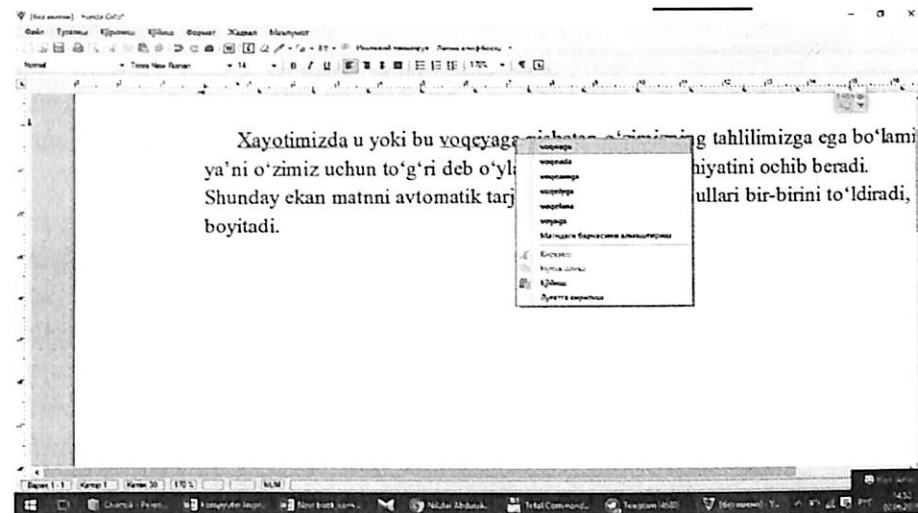
Piton dasturlash tilida ushbu natijani sinab ko'rilsa, undagi natijasi quyidagicha bo'ladi:

BAHOR	0.8666666666666667	0.9066666666666667
NAHOR	0.7333333333333334	0.7333333333333334
HABAR	0.7	0.7
KITOB	0.8666666666666667	0.9066666666666667
KILOB	0.7333333333333334	0.7866666666666667

Ushbu algortim orqali muayyan so'zdagi harflar tushirilib qoldirilishi, o'rniga qo'yishi, almashtirilishi mumkin. Shunga ko'ra u so'zlardagi harflar lingvistik bilimlar bazasi orqali tekshiriladi.

Tahrirlovchi dasturlarning onlayn tizimlari ham yaratilgan bo'lib, ularning imkoniyatlari bir-biridan farqli ravishda matnni har jihatdan tekshirishga yo'naltirilgan²⁶⁸:

1. Grammarly
2. Reverso
3. Ginger Software
4. WhiteSmoke
5. LanguageTool
6. PaperRater
7. Hemingway Editor
8. Pro Writing Aid
9. Online Correction.com
10. Spell Check Plus
11. Grammar Slammer
12. Virtual Writing Tutor
13. Microsoft Word
14. Google Docs
15. Slick Write
16. GrammarCheck
17. WordPerfect Office X8
18. SentenceChecker.org
19. After the Deadline
20. Textly.ai



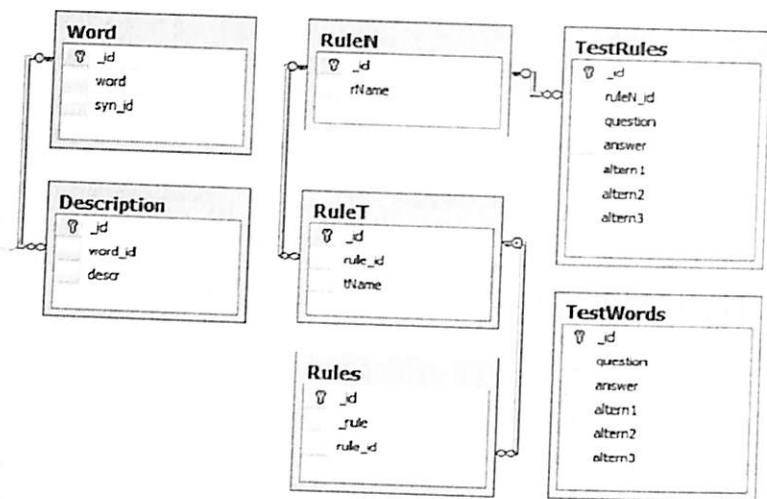
38-rasm

Masalan, imloni tekshiruvchi mobil ilovaning berilganlar bazasini yaratishda quyidagi komponentlardan iborat bo'ladi:

- o'zbek tilining imlo lug'ati;
- o'zbek tilining imlo qoidalari;
- o'zbek tilidagi paronim so'zlarining kontekstda berilishi bilan bog'liq o'rinlar (ham-xam, rux-ruh) va ularning izohi;
- imlo va qoidalarni tekshiruvchi testlar.

Ushbu mobil ilovani yaratish jarayonining asosiy bosqichlaridan biri, imloni tekshirishda foydalaniladigan so'zlar va imlo qoidalarini saqlash uchun berilganlar bazasini loyihalashtirishdir. Yaratilgan berilganlar bazasi 3-normal shaklga keltirilgan bo'lib, munosabatlar ustida amallar bajarish jarayonida hosil bo'ladigan anomalialar yo'qotilgan. Quyida berilganlar bazasi sxemasi keltirilgan:

²⁶⁸ <https://financesonline.com/top-20-grammar-checker-software-solutions/>



Ilovaning qidiruv tizimi mobil qurilmadagi lokal berilganlar bazasi bilan ishlaydi.

Qoidalarni topish uchun SQL so'rov quyidagi qadamlardan iborat bo'ladi:

```

SELECT RulesN.rname, Rules.rule
FROM RulesN INNER JOIN Rules ON RulesN.[_id] =
Rules.rules_id

```

WHERE (((RulesN.rname)="rname
Sintaktik qoidalar"));

So'zni topish quyidagi so'rov orqali amalga oshiriladi:

```

SELECT Words.word, Description.descrip
FROM Words INNER JOIN Description ON Words.[_id] =
Description.word_id

```

WHERE (((Words.word)="shox"))

13.7. Sentiment analiz

Dunyoda sodir bo'layotgan turli voqealar, jamiyatda bo'layotgan o'zgarishlar, qolaversa, sotsiumda bildiralayotgan turli munosabatlarni ommaviy axborot vositalari (gazeta, internet, televidenie, radio, sotsial tarmoqlar) orqali bilishga intilamiz. Bu esa bizning dunyoqarashimiz, u yoki bu sohadagi bilimlarmizga o'z ta'sirini ko'rsatmay qolmaydi. Matn analizi orqali kishilar tomonidan bildirilgan kayfiyat va munosabatlarni tahlil qilish, monitoring qilish yo'nalishi sentiment analiz deb yuritiladi.

Matnning tonalligi (тональность текста) – bu matnda aks etgan emotsional munosabatni bildiradi. O'quvchida muayyan tasavvur paydo qilishda matn muallifi obyektga nisbatan o'zining muayyan emotsional bo'yoqdorlikka ega bo'lgan leksikasidan foydalanadi.

Sentiment termini ilk bor 2001-yilda Das va Chen tomonidan bozordagi mahsulotlarni baholash va sentiment leksikonini qo'llash haqidagi maqolada ilk bor qayd etilgan²⁶⁹.

Sentiment analiz (*ingl. sentiment analysis*) – kompyuter lingvistikasi sohasining bir yo'nalishi bo'lib, matnda ifodalangan munosabat yoki emotsiyani avtomatik tahlil qilishga yo'naltiriladi. Bunda obyekt va subyekt o'rtasidagi salbiy, ijobiy yoki neytral munosabatlar tahlil qilinadi. Sentiment analiz jamiyatda sodir bo'layotgan o'zgarishlarni baholashga yo'naltirilgani bois u amaliy jihatdan, marketing, sotsiologiya, siyosatshunoslik, psixologiya kabi sohalarda o'zining amaliy ahamiyatiga egaligi bilan ajralib turadi. Endilikda matn bo'yoqdorligini analiz qilib beruvchi bir nechta metodlar mavjud²⁷⁰. Ular sirasiga qoidaga asoslangan, lug'atga asoslangan, inson ishtirokisiz mashina tahlili yoki gibril metodlarni keltirish mumkin.

Sentiment analizda – 1.5 – 1.5 shkala orasidagi munosabatlar kalit so'zlar orqali tahlil qilinadi. Agar minus ko'rsatkichni ifodalasa, salbiy; musbatni ko'rsatsa, ijobiy munosabatlarni bildiradi. Sentiment analizda matn korpusi va uning fragmentidan foydalaniladi. Bunda vektor semantikasidan foydalaniladi. Ya'ni muayyan tushunchaga yoki umumiy assotsiativ ma'noni anglatuvchi so'zlar baza sifatida to'planadi. Insonlar tomonidan qabul qilingan ijobiy va salbiy ma'noga ega bo'lgan birliklar dominant so'z atrofiga birlashadi. Har bir so'zlar toksonomik, semantik bog'liqlik, sinonimik va antonimik, uyadoshlik kabi munosabatlarga ko'ra matn tahlil qilinadi. Bu nafaqat sentiment analiz, balki axborot qidiruv tizimi, matnlarni qisqartirish, savol-javob tizimi, diolog tizimi,

²⁶⁹ Jurafskiy D., James H. Martin Speech and language processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 2007 -P. 397

²⁷⁰ Зверева П.П. Сентимент-анализ текста (на материале печатных текстов газеты «The New York Times» о россии и россиянах) / Вестник МГОУ. Серия: Лингвистика, 2014 / № 5. –С. 33.

mashina tarjimasi, matn tasnifi kabi boshqa yo'nalish uchun ham muhim manba bo'lib hizmat qiladi. Sentiment analizda vektor maydoni sifatida korpus matnlari qo'llaniladi. Vektorning vertikal holatida so'z, gorizantal holatida hujjatlar ushbu so'zning qo'llanilish chastotasini ko'rsatadi. Shuningdek, semantik vektorda vertikal holatida so'z, gorizantal holatida ham so'z qo'llanilib, kesishish nuqtasi bo'yicha ham analiz qilish mumkin²⁷¹.

Matn mazmunida aks etgan fikr ta'sirchanligini kompyuter orqali qanday aniqlash mumkin? Bunda kishilarning quyidagi xususiyatlari aks etadi²⁷²:

-emotsiya (his-hayajon) –kishilarning muayyan ichki va tashqi ta'sirga nisbatan javobi: g'urur, jahl, xursandchilik, qo'rquv, hafalik...;

-kayfiyat – muayyan ta'sir natijasida kishilarning voqega nisbatan his-tuyg'usi: tushkunlik, xursandchilik...;

-sotsial munosabat– muayyan vaziyatda kishilar orasidagi sotsial aloqalar: do'stlik, iliqlik, sovuqlik...;

-hissiy munosabat – muayyan davrgacha davom etuvchi kishilar orasidagi aloqa: sevish, yomon ko'rish, xohish, nafratlanish...;

-shaxsiy xulq-atvor – biror kishi uchun insonlarning o'ziga xos xulq-atvori: rashk qilish, asabiylashish, janjallashish...

Sentiment analizda mana shu turga kiruvchi so'zlarning lingvistikk bazasi yaratiladi va ular ijobiy va salbiy turiga qarab matnlar avtomatik analiz qilinadi.

Sentiment analiz jamiyatning turli sohalarida qo'llanilmoqda:

-sport sohasida (sportchilarning reytingini aniqlash uchun)

-tijorat sohasida (mahsulotlarni baholash uchun)

-ta'lim sohasida (universitet reytinglarini aniqlash uchun)

-turizm sohasida (mehmonxonalar ko'rsatkichini baholash uchun) va boshqalar.

Ingliz, italyan va nemis tillari uchun Google cloud, Microsoft Azure Text Analytics API, Microsoft Azure Emotion API, Social

²⁷¹ Qarang: Jurafskiy D., James H. Martin Speech and language processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 2007 -P. 110

²⁷² Scherer, K. R. (2000). Psychological models of emotion. In Borod, J. C. (Ed.), The neuropsychology of emotion, Oxford, pp.137–162.

Mention, Sentiment140, «SentiStrength»; Semantria, SentiFinder kabi matn analizi bilan shug'ullanuvchi dasturlar yaratilgan.

Matnlarni sentiment analiz qilishning quyidagi bosqichlari mavjud²⁷³:

1) tilning leksik tonalligi aks etgan lug'atni yaratish;

2) gaplarning sentimentini aniqlovchi qoidalarni yaratish va formallashtirish;

3) lug'at va qoidaga asoslanib, matnni sentiment analiz qilish algoritmini ishlab chiqish.

Matnning sentiment analizida tezaurus juda katta rol o'ynaydi. Ingliz tili uchun WordNet-Affect, SentiWordNet, SenticNet, MPQA Opinion Corpus; rus tili uchun РусСентилекс kabi so'zlarning sentiment ma'nolarini ifodalagan ochiq resurslar yaratilgan. Biroq o'zbek tili uchun na sentiment analiz, na tazaurus bo'yicha fundamental tadqiqot olib borilgan.

Qozoq tilida esa bu yo'nalish bo'yicha muayyan darajada jiddiy tadqiqotlar olib borilmoqda. Sentiment analiz bo'yicha amalga oshirilgan ishlarda so'zlarning sentiment analizi uchun so'zlarni quyidagi sinflarga ajratadi²⁷⁴:

• sifatga oid so'z va so'z birikmalar (baholash: [-2; 2]);

• otlar (baholash: [-1;1]);

• fe'lga oid so'z va so'z birikmalar (baholash: [-1;1]);

• ravish (ortirma daraja).

Bundan ko'rinib turibdiki, ushbu baholash shkalasi quyidagi ottenkaga ega:

-2	-1	0	1	2
Juda yomon	Yomon	Neytral	Yaxshi	Juda yaxshi
bashara	Aft	yuz	chehra	oraz

²⁷³Ергеш Бану Определение тональности текстов на казахском языке на основе словаря эмоциональной / Turklang.2017 international conference –Kazan, 2017 -C.63-64.

²⁷⁴ O'sha joyda. -C.65.

Ushbu baholash belgisiga ko'ra so'zlar aniqlangandan so'ng, qaysi birliklar ko'p ishlatilishi yoki kam qo'llanilishi statistik metod orqali tahlil qilinadi.

Adabiyotlar:

1. X. Li. D. Roth, *Learning Question Classifiers: The Role of Semantic Information*, Natural Language Engineering 1 (1): 000-000. Cambridge University Press. 2004
2. Bryan S. Todd An introduction to expert systems. Oxford university (monograph), 1992. 101 -p.
3. Jurafskiy D., James H. Martin Speech and language processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 2007 -P. 110
4. Scherer, K. R. (2000). Psychological models of emotion. In Borod, J. C. (Ed.), *The neuropsychology of emotion*, Oxford. pp. 137-162.
5. Зверева П.П. Сентимент-анализ текста (на материале печатных текстов газеты «The New York Times» о россии и россиянах) / Вестник МГОУ. Серия: Лингвистика, 2014 / № 5. - С. 33.
6. Vibhakti V. Bhaire, Ashiki A. Jadhav, Pradnya A. Pashte, Mr. Magdum P.G SPELL CHECKER / International Journal of Scientific and Research Publications, Volume 5, Issue 4, April 2015 1 ISSN 2250-3153
7. Galitsky, B. Machine Learning of Syntactic Parse Trees for Search and Classification of Text. Engineering Application of AI, <http://dx.doi.org/10.1016/j.engappai.2012.09.017>, (2012b)
8. Galitsky B., Ivovsky D., Kuznetsov S. Parse thicket representations of text paragraphs, «Диалог» (2013) Выпуск 12, p 241
9. Abdurakhmonova N. Tuliyeu U. Spell checking analysis of Uzbek texts using Djaro Winkler algorithm / Turklang. 2018 international conference. - Tashkent, 2018 -P. 311-314
10. Ергеш Бану Определение тональности текстов на казахском языке на основе словаря эмоциональной / Turklang. 2017 international conference. - Kazan, 2017. - С. 63-64.

Savol va topshiriqlar

1. Sun'iy intellekt tushunchasini izohlang.
2. Indeksplash jarayoni haqida fikr yuriting.
3. Matn generatsiyasi deganda nimani tushunasiz?
4. Savol-javob tizimi (SJT) bo'yicha dunyo bo'yicha qanday tadqiqotlar olib borilgan.
5. Conversation Analysis (CA) sohasi haqida internet ma'lumotlariga tayanib amaliy mashg'ulot darsida muhokama qiling.
6. Bugungi kunda matnlarni tahrirlovchi dasturlar qanday funksional imkoniyatlarga ega?
7. Matnlarni avtomatik tahrirlash algoritmlari qaysilar?
8. SJTning umumiy ishlash texnologiyasi qanday usul bilan amalga oshiriladi?
9. Diologning kompyuterga oid modellari, deganda nimani tushunasiz?
10. Bugungi kunda sun'iy intellekt qaysi sohalarda qo'llanilmoqda?

Ushbu unilarning fonemataktikasini daftaringizga yozing va misollar keltiring. O'zbek tili bo'inlar modellariga boshqa namunalarni toping.

i	
u	
e	
o	
o'	

13-mavzu yuzasidan test

1. Hujjatlarning yoki so'rovlarning asosiy semantik aspektini ifodalashda turli texnologiyalardan foydalaniladi. Bu jarayon nima deb nomlanadi?
 - a) indeksplash (indexing)
 - b) search (qidirish)
 - c) analise (qidirish)
 - d) checking (tekshiruv)

2. Dastlabki savol-javob tizimi (SJT) kim tomonidan rivojlantirilgan?
- Weizenbaum
 - Eshref Adali
 - Ch.Hokket
 - Troyanskiy
3. Monolog va diolog o'rtasidagi farq nima?
- diologda matn navbatlanib keladi
 - monologda so'z miqdori ko'p bo'ladi
 - dialog ilmiy matnlardan iborat bo'ladi
 - monologda qidiruv tizimi bo'lmaydi
4. Tahrirlovchi dasturning ta'minotini yaratish bo'yicha qanday texnologiyalar mavjud?
- Djaro Vinkler algoritmi orqali ehtimollik nazariyaga asoslangan texnologiya
 - sof lingvistik bazaga asoslangan texnologiya
 - HFST (finite state transducer) algoritmiga asoslangan texnologiya
 - barcha javoblar to'g'ri

FOYDALANILGAN ADABIYOTLAR:

Dissertatsiya va avtoreferatlar:

- Abduraxmonova N. Inglizcha matnlarni o'zbek tiliga tarjima qilish dasturining lingvistik ta'minoti. Fil.f. falsafa doktori dissertatsiyasi. – Toshkent, 2018. 165 b.
- Kemal Altuntaş Turkish to Crimean Tatar machine translation system (a thesis submitted to the department of computer engineering and the institute of engineering and science of Bilkent university in partial fulfillment of the requirements for the degree of master of science) 2001. – P. 30
- Бондарева Н.А. Прикладные аспекты моделирования терминологии (на материале терминосистем уголовного права в русском и английских языках): Дисс...канд. филол. Наук. – Санкт-Петербург, 2003.
- Казымбетова М.М. Семантический анализ русского языка для системы «Таржимон Л-МХ»: Магист-я дисс., Научный руководитель: Хакимов М.Х. НУУз, 2006.
- Каримов О. «Таржимон Л-МХ» тизими учун ўзбек тилини лексик ва синтактик таҳлил қилишнинг математик модел ва алгоритмларини тузиш: Магист-к дис. – Тошкент, 2005.
- Луконина Е.К. Идиоматическая фразеология в языке и речи (на материале современного английского языка): Дисс. ... канд. филол. наук. – М., 1986.
- Мадраҳимов И. Ўзбек тилида сўзнинг серкирралиги ва уни таснифлаш асослари: Филол. фан. номз... дис. афтореф. – Тошкент, 1994. – Б. 25.
- Муратов У.М. Построение математической модели и алгоритмов лексического и синтаксического анализа русского языка для системы «Таржимон Л-МХ». Магист-я дисс., Научный руководитель: Хакимов М.Х. НУУз, 2005.
- Тухсанов М. Микротекст и система средств выражения его когерентности в узбекской художественной речи (замена и повтор): Автореф. дис. ...д-ра.филол.наук. – Ташкент, 1987. – 24 б.
- Ханаху Д. Р. Атрибутивные словосочетания в русском и английском языках. (сопоставительно-типологический анализ): Дисс... канд. филол. наук. – Майкоп, 2007. –166. с.

11. Хорошилов А.А. Теоретические основы и методы построения систем фразеологического машинного перевода: Дисс... канд. техн. наук. – Москва, 2006. – 120. с.
12. Хакимов М.Х. Ўзбек тилида матнинг прагматик талқини: Филол. фан.д-ра....дис. автореф. – Тошкент, 2001. – 24. б.
13. Шелманов А. Исследование методов автоматического анализа текстов и разработка интегрированной системы семантико-синтаксического анализа: Дисс... канд. техн. наук. – Москва, 2015. –210 с.
14. Рахимова Д. Исследование моделей и методов семантики машинного перевода с русского языка на казахский язык: Дисс... канд. техн. наук, – Казахстан, 2014. – С. 24-50.
15. Махмудов М. Разработка системы формального морфологического анализа тюркской словоформы (на материале азарбейджанского языка): Дисс... канд. филол. наук. – Баку, 1982. –С. 20.
16. Ахмеджанов Т. Таржимон дастурлар ёрдамида техника терминлари ўқитиш: Магистрлик диссертацияси, – Тошкент, 2005.

Jurnal va to'plamlar:

17. Abdurakhmonova N. The bases of automatic morphological analysis for machine translation // Известия Кыргызский государственный технический университет им. И.Раззокова теоритической и прикладной научно-технический журнал.– Bishkek, 2016. –№ 2 (38). –С. 12-17.
18. Abdurakhmonova N.Z. Grammatical analyses in machine translation between English and Uzbek // Proceedings of the I International conference on Computer processing of Turkic Languages (TurkLang.2013), – Astana, 2013. –P. 294-297.
19. Belonogov G.G. On the use of a similarity method for automatic processing of textual information [Об использовании метода аналогии при автоматическом обработке текстовой информации] Problems of cybernetics [Проблем кибернетики].– Moscow: Наука. Issue 28. 1974.
20. Gelbukh A. F. Effectively realizable morphologic model of inflective language [Эффективно реализуемая модель

морфологии флективного языка] // Научно-техническая Информация [Scientific and technical information], 1992. series 2, #1, –PP. 24–31.

21. S. Vogel, Ney H. and Tillmann C. 1996. HMM-based Word Alignment in Statistical Translation. / In COLING '96: The 16th International Conference on Computational Linguistics, Copenhagen, Denmark. –PP. 836-841.
22. Zangeneid R., Sonnenhauser B. Russian Verbal Aspect and Machine Translation / Computational Linguistics and Intellectual Technologies Papers from the Annual International Conference “Dialogue” 2014. –№ 13, PP. 744-745.
23. Абдурахмонова Н. Хакимов М.Х. Семантические базы английского языка для многоязычной ситуации компьютерного перевода. / Труды научной конференции «Проблемы современной математики» –Карши, 22-23 апреля 2011. –С. 311-314.
24. Абдурахмонова Н., Хакимов М.Х. Логико-лингвистические модели слов и предложений английского языка для многоязычных ситуаций компьютерного перевода. // 1-я Международная конференция “Компьютерная обработка тюркских языков. Латинизация письменности” Казахстан, – Астана, 2013. –С. 302-306.
25. Абдурахмонова Н.З. Компьютерда илмий матнларни таржима қилишнинг асосий масалалари // Илм сарчашмалари журнали. –Хоразм, 2013. –№ 5. –Б. 40-45.
26. Алиқулов З. Истиклол даврида ўзбек терминологияси тўғрисида / Тилшуносликнинг долзарб масалалари. –Тошкент, 2008. –Б. 146.
27. Алпонова Ш. Матнинг қабул қилинишига доир // Ўзбек тили ва адабиёти. 2006. 2-сон. –Б. 73-76.
28. Андреев Н.Д. Машинный перевод и проблема языка-посредника // ВЯ. 1957. №5. – С. 117-121.
29. Апресян Ю.Д. О языке для описания значений слов. // Известия АН СССР. Серия языка и литературы № 5, 1966. –С. 415-429.
30. Болшаков И. А. CrossLexica: A large electronic dictionary of collocations and semantic links between Russian words [КроссЛексика — большой электронный словарь сочетаний и

смысловых связей русских слов] // Компьютерная Лингвистика и Интеллектуальные Технологии: Труды Международной Конференции "Диалог 2009" [Computational Linguistics and Intelligent Technologies: Proceedings of the International Conference "Dialogue 2009"]. – Moscow, 2009. – P. 45–50.

31. Винокуров А.А., Чуканов В.О. Новый метод оценки машинного перевода // Информационные технологии и системы. Hardware Software Security. Тенденции и перспективы – Сборник статей. – Москва, Международная академия информатизации, 1997.

32. Винокуров А.А., Чуканов В.О. Новый метод оценки машинного перевода // Информационные технологии и системы. Hardware Software Security. Тенденции и перспективы: Сборник статей. – М.: Международная академия информатизации, 1997.

33. Говорун М. Машинный перевод в XX веке : как это было, как этого нет / М. Говорун, А. Сергеев // Мир Интернет. 2001. – №2. – С.16-26.

34. Дрейзин Ф.А. Об алгоритмизации составления алгоритма анализа языка. (на примере морфологии агглютинативного узбекского языка) // Научные труды Ташкентского университета, 1961, вып. 189. Матем. Науки, кн. –С.121.

35. Дрейзин Ф.А. Об алгоритмизации составления алгоритма анализа языка. (на примере морфологии агглютинативного узбекского языка) // Научные труды Ташкентского университета, 1961, вып. 189. Матем. Науки, кн. –С.121.

36. Дрисслер В. Синтаксис текста // Новое в зарубежной лингвистике. – М.: Прогресс, 1978. конференции «Диалог» (2012) Выпуск 11, –С.81.

37. Лютикова Е. А. Формальное Моделирование падежного варирования: параметрический подход. // Компьютерная лингвистика и интеллектуальные технологии. По материалам ежегодной международной конференции «Диалог» (2016) Выпуск 15, –С. 461.

38. Мелчук И.А. Опыт теории лингвистических моделей "смысл-текст". Семантика. Синтаксис. –Москва: Наука, 1974. – С.314.

39. П.Беляева Л. Н. Практические системы машинного перевода – теория, технология, перспективы // Структурная и прикладная лингвистика: межвуз. сб. – СПб., 1998. – С.172-183.

40. Расулов И. Мураккаб синтактик бутунлик // Ўзбек тили ва адабиёти. –Тошкент, 1983. – №1. – Б. 22.

41. Реформатский А.А. Что такой термин и терминология? / Реформатский А.А. // Вопросы терминологии, 1961. –С. 49.

42. Сайфуллаева Р., Абузалова М. Гапнинг энг кичик курилиш колиплари хақида // Ўзбек тили ва адабиёти. – Тошкент, 1991. – №5. – Б. 42-47.

43. Тожиев Ё. Матн кўринишлари ва матн таърифи / Тилшуносликнинг долзарб масалалари. (Илмий мақолалар тўплами). – Т., 2010. – В.164-171.

44. Хакимов М.Х. Расширяемый входной язык математического моделирования естественного языка для многоязычной ситуации машинного перевода // ЎЗМУ хабарлари. –Тошкент, 2009. –№1. – С.75-80.

45. Хакимов М.Х. Формальные системы машинного перевода в многоязычной ситуации // Современные проблемы математики, механики и информационных технологий. Материалы республиканской научно-технической конференции. – Ташкент, 2008. – С. 297 – 300.

46. Хакимов М.Х. Математические модели узбекского языка // ЎЗМУ хабарлари. – Тошкент, 2010. – № 3. –С. 185-188.

47. Хакимов М.Х. Семантические базы и математические модели русского языка для многоязычных ситуаций компьютерного перевода // ЎЗМУ хабарлари. – Тошкент, 2011. – №2 – С. 57-64.

48. Шаляпина З.М. Грамматика и ее соотношение со словарем при словоцентрическом подходе к языку (на опыте формализованного лингвистического описания) // Вопросы языкознания. 1991. – №5. – С.42-54.

49. Шаляпина З.М. К проблеме моделирования языковой компетенции и языковой деятельности человека в рамках общей

- модели языковой действительности // Референция и проблемы текста образования. – М., 1988. – С. 214-234.
50. Шаляпина З.М. Текст как объект автоматического перевода // Текст и перевод. – М.: Наука, 1988. – С.113-129.
51. Шемакин Ю.И. Начало компьютерной лингвистики. – М.: МГОУ, 1992.
52. Шодиев И. Матнинг коммуникатив турлари // Ўзбек тили ва адабиёти. 2006. 2-сон. –Б. 86-88.
53. Ю.Беляева Л. Н. Автоматический (машинный) перевод / Л.Н.Беляева, М. И. Откупщикова // Там же. – СПб.: Изд-во С.-Петербург. ун-та, 1996. –С. 360-388.
54. Modeling analytic forms of verb in Uzbek as stage of morphological analysis in machine translation // Iranian Journal of Social Sciences and Humanities Research. UCT. J. Soc. Human. Resear. (UJSSHR).Takestan, Iran, 2017. Volume 5, Issue 4. P. (№5 Global Impact factor, Impact factor-1,78)
55. Компьютер лингвистикаси асослари / “Филология фанларининг долзарб масалалари” мавзусидаги республика илмий-назарий конференция материаллари. Бухоро: Дурдона, 2017. – Б. 16-19.
56. Personal names spell-checking – a study related to Uzbek // Iranian Journal of Social Sciences and Humanities Research. UCT. J. Soc. Human. Resear. (UJSSHR).Takestan, Iran, 2018. Volume 6, Issue 1. P. (№5 Global Impact factor, Impact factor-1,87)
57. Morphological analysis by finite state transducer for Uzbek-English machine translation // Horijiy filologiya 3-son Samarkand journal, 2018. –P. 64-71
58. New keyboard or Unicode for Uzbek Latin alphabet in process of the text via computational technology // Ozbek tili va adabiyoti ta’limi 6-son, 2018 yil
59. Inglizcha-o‘zbekcha tarjima lug‘ati uchun semantik ma’lumotlar bazasi / Translation, information, communication – political and social bridge” nomli xalqaro konferensiya. Samarqand, 2018 yil. 4-may. – B. 147-148.
60. O‘zbek tilining ontologiyasini yaratish masalalari / Til va adabiyot ta’limida zamonaviy axborot va pedagogik texnologiyalar Respublika konferensiya / Toshkent, 2018. – B.172-174
61. Tezaurus yaratishda leksik munosabatlar tavsifi / Til va adabiyot ta’limida zamonaviy axborot va pedagogik texnologiyalar Respublika konferensiya / Toshkent, 2018. – B.177-180
62. Modeling WorNet type thesaurus for uzbek language semantic dictionary // International Journal of Systems Engineering 2018; 2(1): 26-28 <http://www.sciencepublishinggroup.com/j/ijse>doi:10.11648/j.ijse.20180201.16
63. Database of mobile application of specific terms on computational linguistics / PAPERS PRESENTED TO THE 3th FORUM OF SOCIAL SCIENCES “THE GREAT STEPPE” / Astana, 2018. –P.220-224
64. Mashina tarjimai lingvistik ta’minoti uchun o‘zbek tilidagi ko‘makchi fe’llarning semantik maydonlarini berish masalasi / Oliy ta’limda darsdan tashqari mashg‘ulotlar – talabalarning kasbiy kompetentligini shakllantirish omili” mavzusidagi respublika ilmiy-amaliy konferensiya. Navoiy viloyati, 2018. – B. 67-68.
65. Spell checking analysis of Uzbek text using Djaro Winkler algorithm / Turklang.2018 international conference. Tashkent.
66. Ontology of grammar rules as example of noun of Uzbek and Kazakh languages Turknag.2018 international conference / 13-15 September 2018. P. 37-38
67. Ontological model of Uzbek language (as example morphology) / Computational Models in Language and Speech» TEL-2018, -Kazan, 5-10 pp.
68. Necessity Uzbek corpus for machine translation Modern linguistics, translatology, linguodidactics international conference, Tashkent, 4-8 October 2018. –P. 120-122
69. Til texnologiyasini rivojlantirishda o‘zbek tili milliy korpusining innovatsion ahamiyati / “Yoshlar Innovatsion faolligini oshirishning dolzarb vazifalari” mavzusidagi Respublika ilmiy-amaliy konferensiya , Toshkent, 2019- B. 159-162
70. Corpus Based Teaching Uzbek As A Foreign Language // Journal of Foreign Language Teaching and Applied Linguistics, Volume 6 - Number. 1 – 2019 ISSN: 2303-5528– P.131-137 (hammuallif: K.Urdishev)

71. Корпус ёрдамида тезаурус яратишнинг концептуал аҳамияти / Translation, information, communication – political and social bridge Самарқанд, 2019 йил 7 июнь
72. Proceedings of the international scientific conference, - P. 36-39 (hammuallif: A.Собиров)
73. A two-level morphological analysis of the Uzbek corpus / МАТЕРИАЛЫ IV международного научного конгресса иностранная филология. социальная и национальная вариативность языка и литературы / 2019/4/1 Симферополь, -P. 425-431
74. Syntactic Structures for Ontological Models (As Example of Uzbek) / Turklang , 2020
75. Семантическое поле связанное со словом маънавият и его лингвистические свойства на основе корпусного анализа / PAPERS PRESENTED TO THE IVst FORUM OF SOCIAL SCIENCES "THE GREAT STEPPE" by INTERNATIONAL TURKIC ACADEMY, Nur-Sultan, 2019 С. 719-727 (hammuallif: A.Sobirov)
76. Composing modification of the Uzbek Phonetic alphabet based on international Phonetic alphabet // International Journal of Advanced Research in Science, Engineering and Technology, Vol. 6, Issue 9, September 2019 Hindiston (hammuallif: Norov A., M.Aripov)
77. O'zbek tilida WordNet yaratish masalalariga doir O'zbekistonda xorijiy tillar (ilmiy metodik elektron jurnal) № 4 (27) 2019, - B. 19-28
78. Syntactic modelling machine translation between Russian and Uzbek Turklang. 2019 (hammuallif: H.Ahmedova)
79. The use of morphological ontological models for Turkish-Uzbek machine translation // O'zbekiston Milliy universiteti xabarлари, 2019, [1/3] ISSN 2181-7324, 128-132 B.
80. Разработка онтологической модели синтаксических правил узбекского языка / Abstracts on the international scientific conference "Actual problems of applied mathematics and information technologies", Tashkent, 2019 (14-noyabr), - С. 281-283

81. Syntactic parsing based on Uzbek corpus "Language technology for all" International conference by UNESCO France, Paris 4-6 December
82. Использование корпуса для обучения синонимии на узбекском языке / Электронная письменность народов Российской Федерации: опыт, проблемы и перспективы Материалы II Международной научной конференции(Уфа, 11-12 декабря 2019 г.), Уфа – 2019, С. 99-104 (hammuallif: Хайруллаева Г., Урдишев К.)
83. ERASMUS+ PROJECT EXPERIENCES ON THE COMPUTATIONAL LINGUISTICS MASTER DEGREE IN THE CENTRAL ASIA UNIVERSITIES / 14th annual International Technology, Education and Development Conference March 2nd-4th, 2020 Valencia (Spain) Indexing Web Of Science Pages: 4803-4812 Publication year: 2020 ISBN: 978-84-09-17939-8 ISSN: 2340-1079 doi: 10.21125/inted.2020.1321 Conference name: 14th International Technology, Education and Development Conference
84. Dates: 2-4 March, 2020 Location: Valencia, Spain (hammualliflar: M. Aripov, A.Duarte, K. Georgouli, C. Gomez-Rodriguez, Y.Kandalina, B.Maia, A. Sharipbay, U. Tukeyev, G.Urazboev, Z.Vetulani, O.Yusupov)
85. Замонавий корпусларнинг компьютер моделлари // Ўзбекистонда хорижий тиллар. —2020. — № 1(30). — Б. 50–58. <https://doi.org/10.36078/>
86. Ot so'z turkumini o'qitishda korpusga asoslangan ta'lim texnologiyasi // Ilm sarchashmalari, 4-Son, 2019, - B. 112-116 (hammuallif: G.Hayrullayeva)
87. Brief Outlook On Creation Of Uzbek Thesaurus By Sketch Engine Technology / "Filologiyaning dolzarb masalalari" mavzusidagi Respublika ilmiy-uslubiy anjumani, Qo'qon 2020-yil 25-aprel, - B. 145-147 (hammuallif: S.Kholmatova)
88. Korpus lingvistikasida matnlarni lingvistik annotatsiyalash tamoyillari / "Ilm-fan ta'limining rivojlanish istiqbollari" mavzusidagi ilmiy konferensiya, Toshkent, 2020-yil 27-aprel, - B. 332-336 (hammuallif: M.Sadikova)
89. Sifat so'z turkumini o'qitishda korpusga asoslangan ta'lim texnologiyasi / Міжнародної науково-практичної інтернет-

конференції «ТЕНДЕНЦІЇ ТА ПЕРСПЕКТИВИ РОЗВИТКУ НАУКИ І ОСВІТИ В УМОВАХ ГЛОБАЛІЗАЦІЇ» 29 травня 2020 року Вип. 59, Переяслав – 2020 (hammuallif: G.Hayrullayeva)

90. Ochiq manbali terminologik ma'lumotlar bazasini yaratishning lingvistik aspekti / "Davlat tili o'qitishning dolzarb masalalari: muammo va yechimlar" xalqaro ilmiy-amaliy anjuman materiallari to'plami, Farg'ona, 2020. – B. 97-100 (hammuallif: U.Babajanov)

91. Creating multilingual thesaurus as example of the concept of "Empire" in Russian, English and Uzbek based on corpus analysis "Davlat tili o'qitishning dolzarb masalalari: muammo va yechimlar" / xalqaro ilmiy-amaliy anjuman materiallari to'plami, Farg'ona, 2020. – B. 139-142 (hammuallif: V.Zakharov)

92. O'zbek tili terminlar bazasi lingvistik portalining funksional imkoniyatlari xususida / "Davlat tili - ijtimoiy taraqqiyot va milliy yuksalish mezon" (O'zbek tiliga Davlat tili maqomi berilganligining 31 yilligiga bag'ishlangan respublika ilmiy-amaliy konferensiya materiallari), Buxoro, 2020. – B. 252-255. (hammuallif: U.Babajanov)

93. Linguistic Issues Of Creating Parallel Corpora For Uzbek Multilingual Machine Translation System // BuxDU ilmiy axborotnomasi № 6, 2020, - P. 60-68.

94. O'zbek tilining lingvistik ma'lumotlar bazasini shakllantirishda "Turkiy morfema" portali instrument sifatida / O'zbek tilini dunyo miqyosida keng targ'ib qilish bo'yicha hamkorlik istiqbollari xalqaro ilmiy-amaliy anjuman, 2020 20-oktabr (hammuallif: A.Gatiatullin)

95. Turkic Morpheme Web Portal as a Platform for Turkology Research International Conference on Information Science and Communications Technologies ICISCT 2020 (Indexing Scopus) Applications, Trends and Opportunities 4th, 5th and 6th of November 2020, Tashkent Uzbekistan (hammualliflar: D. Suleymanov, A. Gatiatullin, N. Prokopyev)

96. UZWORDNET: A Lexical-Semantic Database for the Uzbek Language / 11th International Global Wordnet Conference (GWC2021) will be held from 18 to 21 January 2021

(Xalqaro konferensiya (Indexing Scopus))(hammualliflar: Alessandro Agostini, Timur Usmanov, Ulugbek Khamdamov, Nilufar Abdurakhmonova and Mukhammadsaid Mamasaidov)

97. Recent Results of "TurkLang-7" Project / XVI международная конференция по компьютерной и когнитивной лингвистике TEL'2020 (12-13 ноября 2020 г., Казань, Россия) (Indexing Scopus), Vol.2780, 2020, -P. 90-101 <http://ceur-ws.org/Vol-2780/> (hammuallif: Aidar Khusainov, Dzhavdet Suleymanov, Rinat Gilmullin, Nilufar Abdurakhmonova, Lenara Kubedinova)

98. Creation Lexical Alignments for Uzbek-Russian Parallel Corpora / XVI международная конференция по компьютерной и когнитивной лингвистике TEL'2020 (12-13 ноября 2020 г., Казань, Россия) (Indexing РИНЦ)

99. Синтаксические структуры для онтологических моделей (на примере узбекского языка)/ VIII Международной конференции по компьютерной обработке тюркских языков TurkLang'2020 18-20 октября 2020 г. Уфа 2020 (hammualliflar: M.Aripov, A.Norov)

O'quv adabiyotlar:

100. Abdurahmonova N. Mashina tarjimasining lingvistik ta'minoti (monografiya). –Toshkent: Muharrir, 2018. 165 b.

101. Anke Lüdeling, Merja Kytö Corpus Linguistics An International Handbook (Volume 1), Walter de Gruyter. Berlin, New York, 2008.

102. Anke Lüdeling, Merja Kytö Corpus Linguistics An International Handbook, Vol. 1, Berlin, New York: Walter de Gruyter. 2008. 81(2), —P. 246–247 DOI: 10.1080/00393270903392342

103. Antoinette Renouf, Andrew Khoe Corpus linguistics: Refinements and reassessments New York, 2009. 471 p.

104. Bern Heine, Heiko Narrog The Oxford handbook of linguistic analysis. / Douglas Biber Corpus-based and Corpus-driven analysis of language variation and use UK: Oxford university, 2015 –P. 193.

105. Bern Heine, Heiko Narrog. The Oxford Handbook of

- Linguistic Analysis. / Douglas Biber Corpus-based and Corpus-driven analysis of language variation and use UK: Oxford university, 2015. 193 p.
106. Bolshakov A., Gelbuk A. Computational Linguistics Models, Resources, Applications. 2004.
107. Brian R. Richard S. Computational Approaches to Morphology and Syntax, 2007, OXFORD. – P.63, 100.
108. Cerstin M., Michael P. State of the art in computational morphology Workshop on Systems and Frameworks for Computational Morphology, SFCM 2009 Zurich, Switzerland, September 4, 2009. Proceedings, –P. 39
109. Cerstin Mahlow, Michael Piotrowski (eds.). JSLIM – Computational Morphology in the Framework of the SLIM Theory of Language / State of the Art in Computational Morphology. – Zurich, 2009. –P. 15.
110. Charlez Meyer English corpus linguistics: An introduction. Cambridge, 2004. – P. 12.
111. Charlez Meyer English corpus linguistics: An introduction. Cambridge University Press, 2004. —168 p.
112. Cyril Goutte, Nicola Cancedda, Marc Dymetman, and George Foster Learning Machine Translation Cambridge, Massachusetts, London, England, 2009. - P.6
113. D. Suleymanov, R. Gilmullin, R. Gataullin Morphological analysis system of the Tatar language based on the two-level morphological model / Turklang 2017. Kazan, 2017. pp. 6-26.
114. Daniel J., James H. Martin. Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition. 2006.
115. Delmonte R. Computational Linguistic Text Processing: Lexicon, grammar, Parsing and Anaphora Resolution, Nova Science Publishers, Inc. New York, 2008, –P. 4-5.
116. Douglas A. Machine Translation: An Introductory Guide. London, 1994.
117. Geoffrey Sampson, Diana McCarthy Corpus linguistics: Readings in a Widening Discipline New York, 2005. 541 p.
118. Hans Lindquist Corpus linguistics and the description of English Edinburg, 2009. 241 p.
119. Hutchins W. Machine Translation: Past, Present, Future. – N.-Y., 1986. –P.118
120. Jean Veronis. Parallel text processing: alignment and use of translation corpora Springer Science+Business Media Dordrecht, 2000
121. Jeffrey Heinz, Colin de la Higuera, Menno van Zaanen Grammatical inference for computational linguistics, Morgan & Claypool Publishers, 2015.
122. John Nerbonne. Linguistic databases. 1996 [http:// odur.let.rug.nl/~nerbonne/papersintro-db.pdf](http://odur.let.rug.nl/~nerbonne/papersintro-db.pdf).
123. Joybrato Mukherjee, Magnus Huber Corpus linguistics and variation in English New York, 2012. 256 p.
124. Jurafskiy D., James H. Martin Speech and language processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 2007 -P. 12-13– P. 140.
125. Karin Aijmer, Bengt Altenberg Advances corpus-based contrastive linguistics. USA: John Benjamins, 2013.
126. Koehn P., Och F.J., Marcu D. Statistical phrase based translation. In Proceedings of the Joint Conference on Human Language Technologies and the Annual Meeting of the North American Chapter of the Association of Computational Linguistics (HLT/NAACL). 2003.
127. Krister Lindén, Miikka Silfverberg, and Tommi Pirinen HFST Tools for Morphology – An Efficient Open-Source
128. Marianne Hundt, Nadja Nesselhauf, Carolin Biewer Corpus linguistics and Web. New York, 2007. 312 p.
129. Mitkov R. The Oxford handbook of Computational linguistics. Oxford university press, 2003.
130. Mohamed Zakaria Kurdi. Natural Language Processing and Computational Linguistics: Speech, Morphology and Syntax. — Great Britain, USA: Wiley-ISTE 2016. — 300 p.
131. Och F., Ney H. A Systematic Comparison of Various Statistical Alignment Models. Computational Linguistics, 29 (1) (2003) 19-51.
132. Och F., Ney H. What can machine Translation learn from speech recognition? 2009

133. Package for Construction of Morphological Analyzers / – Computational Morphology in the Framework of the SLIM Theory of Language / State of the Art in Computational Morphology. – Zurich, 2009 P. 30
134. Raül Canals, Anna Esteve, Alicia Garrido et.al. , interNOSTRUM: A SpanishCatalan Machine Translation System, Machine Translation Review, Issue No. 11, December (2000) –PP. 21-25
135. Sirgei N., Somers H.L. Reading in machine translation. – MIT press, 2003.
136. State of the Art in Computational Morphology Workshop on Systems and Frameworks for Computational Morphology, SFCM 2009 Zurich, Switzerland, September 4, 2009, Proceedings, –P. 31, 88.
137. Vogel S., Ney H. and Tillmann C. 1996. HMM-based Word Alignment in Statistical Translation. // In COLING '96: The 16th International Conference on Computational Linguistics. Copenhagen, –Denmark.– P. 836-841.
138. Yael Ravin, Claudia Leacock Polysemy. Theoretical and computational approaches. Oxford University press. 2000. – P.13.
139. Yael R., Claudia L. Polysemy. Theoretical and computational approaches. Oxford University press 2000. – P.13.
140. Yorick W. Machine translation. Its scope and limits. Spring science+Business Media LLC. 2009 UK. –P. 139.
141. Абдурахмонова Н.З. Машина таржимасининг лингвистик асослари. (Ўқув қўлланма)–Тошкент: Академнашр, 2012. 160 б.
142. Анисимов, А. Компьютерная лингвистика для всех: Мифы. Алгоритмы. Язык. – Киев: Наук. думка, 1991.
143. Апресян, Ю.Д. Идеи и методы современной структурной лингвистики. – М.: Просвещение, 1966 г. 34 с.
144. Апресян, Ю.Д. Лингвистический процессор для сложных информационных систем / Ю.Д. Апресян [и др.]. – М.: Наука, 1992. – 256 с.
145. Арбекова Т.И. Лексикология англ. яз.(практический курс) учебное пособие для 2-3 курсов ин-ов и фак. Иностр. Яз. – М.: Высшая школа, 1977. –С. 53.

146. Баранов, А.Н. Введение в прикладную лингвистику: учебное пособие. – М.: Эдиториал УРСС, 2001. – 360 с.
147. Бархударов Л.С. Структура простого предложения современного английского языка. Издательство: ЛКИ. 2008. –С. 87.
148. Бархударов Л.С. Язык и перевод (Вопросы общей и частной теории перевода). –М.: Междунар. отношения, 1975. – С. 190-191.
149. Белоногов, Г.Г. Компьютерная лингвистика и перспективные информационные технологии. – М.: Русский мир, 2004. – 248 с.
150. Беляева Л. Н. Автоматизация в лексикографии / Л. Н. Беляева, А. С. Герд, И. И. Убин // Прикладное языкознание: Учебник. – СПб., 1996. – С. 318 – 333.
151. Большакова Е.И. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие – М.: МИЭМ, 2011. – 272 с.
152. Боярский, К.К. Введение в компьютерную лингвистику: учебное пособие. – СПб: НИУ ИТМО, 2013. – 72 с.
153. Бурлакова В.В. Синтаксические структуры современного английского языка. – М.: Просвещение, 1984. –С. 25-41.
154. Волкова И.А. Введение в компьютерную лингвистику. Практические аспекты создания лингвистических процессоров: учебное пособие. – М., 2006. – 43 с.
155. Галперин И.Р. Текст как объект лингвистического исследования. – М.: Наука, 1981. – С.19
156. Гинзбург, Хидекел С. С., Князева Г. Ю., Санкин А. А. Лексикология английского языка: Учебник для ин-тов и фак. иностр. яз. –2-е изд., испр. и доп. – М.: Высш. школа, 1979. –С. 35.
157. Зализняк А.А. Из заметок о любительской лингвистике. – М.: Русский Мирь, 2008. – 240 с.
158. Захаров В. П., Богданова С. Ю. Корпусная лингвистика/Учебник для студентов гуманитарных вузов. — Иркутск: ИГЛУ, 2011. — 161 с.

159. Зубов, А.В., Зубова, И.И. Основы искусственного интеллекта для лингвистов: учебное пособие. – М.: Университетская книга; Логос, 2007. – 320 с.
160. Кобрин Н.А., Корнеева Е.А., Оссовская М.И., Гузеева К.А. Грамматика английского языка. Морфология. Синтаксис. – СПб.: Союз, 2008.
161. Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 26–30 мая 2010 г.). Вып. 9 (16). – М.: Изд-во РГГУ, 2010.
162. Компьютерная лингвистика: научное направление и учебная дисциплина: сборник научных статей. Вып. 1 / В. И. Коваль (ответств. ред.) [и др.]; М-во образования РБ, ГГУ им. Ф. Скорины. – Гомель: ГГУ им. Ф. Скорины, 2010. – 236 с.
163. Котелова Н.Э. Значение олова и его сочетаемость. – Л.: Наука, 1975. – С. 28.
164. Кунин, А.В. Фразеология современного английского языка. – М.: Международные отношения, 1972. – С. 47.
165. Курбонова М. Ҳозирги замон ўзбек тили (Содда гап синтаксиси учун материаллар). – Тошкент, 2002.
166. Лебедова Л.Д. Введение в курс лексикографии. – Москва: Высшая школа, 2008. – С. 170-176.
167. Лингвистика для всех. Летние лингвистические школы 2005 и 2006 / Ред.-сост. Е.В. Муравенко, О.Ю. Шеманаева. – М.: МЦНМО, 2008. – 440 с.
168. Мамажонов А. Текст лингвистикаси. – Тошкент, 1989.
169. Мамедова М.Г., Мамедова З.Ю. Машинный перевод: эволюция и основные аспекты моделирования. – Баку, 2005. – С. 34, 69-72.
170. Марчук Ю.Н. Компьютерная лингвистика (учеб. пособ.) – Москва: Восток Запад, 2007. – С. 61, 65, 272.
171. Марчук Ю.Н. Модели перевода. – Москва: Академия, 2010.
172. Маскальская О.И. Грамматика текста. – Москва: Высшая школа, 1981.

173. Мелчук И.А., Равич Р.Д. Автоматический перевод. 1949 – 1963: Аналитико-библиографический справочник. – Москва, 1967.
174. Мельчук, И.А. Опыт теории лингвистических моделей «Смысл Текст». – М.: Школа «Языки русской культуры», 1999. – 346 с.
175. Мешков О.Д. Словообразование современного английского языка. – М., 1976. – С. 187.
176. Миртожиев М.М., Мухамедова С.Х. Ўзбек тилида сўз валентликлари. – Тошкент, 2011.
177. Мухин А.М. Структура предложений и их модели. – Л.: Наука, 1968. – С. 52-57.
178. Пентус, А.Е., Пентус М.Р. Теория формальных языков: учебное пособие. – М.: МГУ, 2004. – 80 с.
179. Переводная машина П.П.Троянского / Отв. ред. Д.Ю.Панов. – Москва, 1959.
180. Пиотровский Р.Г. Текст, машина, человек. – Л.: Наука, 1975.
181. Пўлатов А. Инглиз тили. – Тошкент: Академнашр, 2011.
182. Пўлатов А. Компьютер лингвистикаси. – Тошкент: Академнашр, 2011.
183. Пўлатов А.К., Муминова Т., Пулатова И. Дунёвий ўзбек тили. Т.1. (Ўзбек тилида феълнинг шакллари ва уларнинг рус, инглиз тилларидаги берилишлари). – Тошкент, 2003.
184. Пўлатов А.К., Мухамедова С. Компьютер лингвистикаси. – Тошкент, 2007.
185. Раҳимов А. Компьютер лингвистикаси асослари. – Тошкент: Академнашр, 2011. – Б. 75-77.
186. Раҳматуллаев Ш. Ҳозирги адабий ўзбек тили. – Тошкент, 2010.
187. Рецкер Я.И. Теория перевода и переводческая практика. – М., 1974.
188. Рождественский Ю.В., Волков А.А., Марчук Ю.Н. Введение в прикладную филологию. – М., 1988. – С. 87.
189. Розентал Д.Э., Теленкова М.А. Справочник лингвистических терминов: Пособие для учителя. – М., 1972.

190. Смирницкий А.И. Лексикология английского языка. – М.: Московский государственный университет. Филологический факультет, 1998.
191. Смирницкий А.И. Синтаксис английского языка. Изд.2. УРСС. 2007. –С. 67.
192. Тожиев Ё. Ўзбек тилида аффиксал синонимия. – Тошкент: Университет, 1991.
193. Шаляпина З.М. Грамматика и ее соотношение со словарем при словотсентрическом подходе к языку (на опыте формализованного лингвистического описания). – Вопросы языкознания, 1991. –С. 42-54.
194. Шаляпина З.М. Текст как объект автоматического перевода. – В кн.: Текст и перевод. –М.: Наука, 1988. –С. 113-129.
195. Шемакин Ю.И. Начало компьютерной лингвистики. – М.: МГОУ, 1992.
196. Шемякин, Ю.И. Начала компьютерной лингвистики: учеб. пособие. – М.: Иэд-во МГОУ, А/О «Росвузнаука», 1992. – 53 с.
197. Юлдашев А.А. Аналитические формы глагола в тюркских языках. –Москва: Наука, 1965.

Интернет ресурслари:

198. http://www.promt.com/company/technology/hybrid_technology
199. <http://www.dialog-21.ru/Archive/Directions/JPTRANSF.htm>
200. <https://universaldependencies.org/>
201. <https://www.clarin.eu/resource-families/parallel-corpora>
202. <https://www.sketchengine.eu/quick-start-guide/parallel-corpus-lesson-7/>
203. <http://opus.nlpl.eu/>
204. <https://www.llas.ac.uk/resources/gpg/1444>
205. <http://www.systran.co.uk/systran/corporate-profile/translation-technology/what-is-machine-translation>
206. <http://www.systran.co.uk/systran/corporate-profile/translation-technology/what-is-machine-translation>

207. http://teneta.rinet.ru/rus/pe/parshin-and_theoria-i-praktika-perevoda.htm

208. http://comp.potrebitel.ru/action=model_list&num_id=71&at_id=669.

209. <http://cl.www.essex.ac.uk/group/> The Computational Linguistics (CL) and Machine Translation (MT) Group (University of Essex)

210. Kay Martin. *History of machine Translation*. URL: <http://www.lsadc.org/Kay.html> (9 May 2002).

211. Hutchins, W. John. Computer-based Translation Systems and Tools. URL: http://www.eamt.org/archive/hutchins_intro.html (8 July 2002).

212. Hutchins, W. John. *The State of Machine Translation in Yeuropa and Future Prospects*. URL: <http://www.hltcentral.org/page-917.0.shtml> (10 March 2002).

213. Hutchins, W. John. *Translation. Technology and the Translator*. URL: <http://www.bcs.org.uk/siggroup/nalatran/nalamtr7/nalamt76.htm> (9 May 2002).

GLOSSARIY

ANNOTATSIYA	ANNOTATION	Maxsus lingvistik belgilar, munosabatlar yoki matndagi strukturalarning tuzilishi va semantik jihatdan qisqartirilgan shakli
ANNOTATSIYA INSTRUMENTI	ANNOTATION TOOL	Annotatsiya jarayonini tezlashtirish va qulayligini ta'minlash uchun maxsus annotatsiya tuzilishini shakllantiruvchi instrument
IKKI TILLI PARALLEL VOSITALAR	BILINGUAL ALIGNMENT TOOLS	Bir matndan ikkinchi tilga o'girilgan parallel korpus asosida bilingval lug'at yaratishga mo'ljallangan instrument.
BILINGUAL CONCORDANS	BILINGUAL CONCORDANCE	Ikki tilli parallel korpuslardan olingan leksik birliklar: termin, so'z va iboralarning ikkinchi tildagi mos tarjimalari ro'yxati.
IKKI TILLI KORPUS	BILINGUAL CORPUS	Ikki tilda tarjima qilingan parallel matnlarning elektron to'plami.
IKKI TILLI LUG'AT	BILINGUAL DICTIONARY	Bir tildan ikkinchi bir tilga tarjima qilish uchun mo'ljallangan lug'at.
CHOMSKIY IYEXRIYASI	CHOMSKY HIERARCHY	Kontekstga bog'liq bo'lmay ichki guruhlarining muayyan tobelanish asosida yaratilgan paradigmatic munosabatlar yig'indisi.
QIYOSIY KORPUS	COMPARABLE CORPORA	Bir nechta tildagi matnlardan olingan namunalarni solishtirish asosida yaratilgan korpus.
KOMPYUTER LEKSIKOGIRAFIYA-SI	COMPUTATIONAL LEXICOGRAPHY	Lug'atlarni kompilyatsiya qilishga yordam berish uchun kompyuter texnologiyalardan foydalanish yoki mavjud

		lug'atlar asosida tabiiy tilni qayta ishlashga yo'naltirilgan soha. Ushbu yo'nalish elektron yoki raqamli leksikografiya deb ham nomlanadi.
KOMPYUTER MORFOLOGIYASI	COMPUTATIONAL MORPHOLOGY	Mavjud metodlar orqali avtomatlashgan morfologik analiz va generatsiyasini amalga oshiruvchi tizim
KOMPYUTER FONOLOGIYASI	COMPUTATIONAL PHONOLOGY	Fonologik ma'lumotni qayta ishlash va ifodalash uchun formal kompyuter modellarga asoslangan tizim
KOMPYUTER PRAGMATIKASI	COMPUTATIONAL PRAGMATICS	Asosan, ovozli dialog tizimlari, kompyuterning turli tizimlarini rivojlantirish uchun qo'llaniladigan pragmatik metodlar va nazariyalar bilan shug'ullanuvchi soha
KOMPYUTER SEMANTIKASI	COMPUTATIONAL SEMANTICS	Tabiiy tildagi lingvistik birliklarning semasini kompyuter orqali tahlil qilishga mo'ljallangan texnologiya
KOMPYUTER TERMINOLOGIYASI	COMPUTATIONAL TERMINOLOGY	Kompyuter yordamida terminologik ma'lumotlarni to'plash, uzatish, saqlash yoki foydalanishga asoslangan tizim
KOMPYUTER YORDAMIDA TIL O'RGANISH	COMPUTER-ASSISTED LANGUAGE LEARNING (CALL)	Til o'rganishda kompyuter texnologiyalardan foydalanish va ular bilan bog'li yangi metod va usullar bilan shug'ullanuvchi soha
KOMPYUTERLA SHTIRILGAN LINGVISTIK USKUNA	COMPUTATIONAL LINGUISTIC TOOL	Lingvistik ma'lumotlar bilan bog'liq jarayonlarni amalga oshiruvchi dastur bo'lib, u analiz qilish, qayta yaratish yoki shunga o'xshash lingvistik vazifalarni

		bajarishda boshqa foydalanuvchilar uchun ham tilga oid ma'lumotlarni qo'shish yo'ji o'zgartirish imkonini beradi.
KONKORDANS	CONCORDANCE	Korpusda aniqlangan barcha so'z yoki frazalarning ularga bog'liq ravishda birikma hosil qilingan ro'yhati
KONTEKST	CONTEXT	1) korpus yoki matn lingvistikasida muayyan so'z yoki frazalarning nutqiy aktda ifodalangan matnning muayyan qismi; 2) diskurs analizi yoki pragmatikada so'zlarning shakl va ma'no jihatlariga ta'sir qiluvchi nutqiy vaziyat
ERKIN KONTEKSTLI GRAMMATIKA	CONTEXT-FREE GRAMMAR	So'z yoki kategoriyalarning iyerarxik tasniflangan muayyan yagona kategoriyalarga birlashtirilgan qoidalar tizimi
KONTEKSTGA BOG'LIQ GRAMMATIKA	CONTEXT-SENSITIVE GRAMMAR	Kontekstga bog'liq grammatika bo'lib, muayyan simvollar zanjiri kontekst orqali aniqlanadi. Bu grammatika tabiiy til elementlarining generatsiya bosqichida qo'llaniladi.
KORPUS	CORPUS	Matnlarning mashina o'qiy oladigan formatda shakllantirilgan va lingvistik tahlil qila oladigan to'plami
KORPUS ANNOTATSIYA SI	CORPUS ANNOTATION	Korpusning lingvistik xususiyatlari va unga bog'liq boshqa ma'lumotlari berilgan izohli qismi
KORPUSGA	CORPUS-BASED	Leksikograflar tomonidan

ASOSLANGAN LUG'AT	DICTIONARY	konkordanslar va boshqa interaktiv instrumentlardan foydalanib mashina o'qiy oladigan formatda matnlar asosida yaratiladigan lug'at
KORPUSGA ASOSLANGAN LEKSIKOGRAFIYA	CORPUS-BASED LEXICOGRAPHY	Katta hajmli korpus asosida so'zlarning ma'nolarini analiz qiluvchi soha
KORPUS LINGVISTIKASI	CORPUS LINGUISTICS	Kompyuterga asoslangan metodologiya bo'lib, lingvistika uchun og'zaki va yozma nutqqa tegishli masalalarni empirik analiz qiladi
MA'LUMOTLAR BAZASI	DATA COLLECTION	Keng maqsadlar uchun mo'ljallangan elektron ma'lumotlar bazasi
MA'LUMOT QIDIRISH	DATA MINING	Muayyan maqsadga yo'naltirilgan ma'lumotlar bazasining avtomatik qidiruv tizimi
LUG'AT	DICTIONARY	Leksema yoki frazemalarning ma'lumotlar majmui. An'anaviy lug'atlarda ushbu ma'lumotlarga so'z shakllari, talaffuzi, so'z turkumi, etimologiyasi, ma'nosi va qo'llanilish doirasi kiradi.
DISKURS	DISCOURSE	Ma'lumot almashinish yoki uzatish uchun bir yoki bir necha kishilar tomonidan ketma-ketlikda tuzilgan gaplarning kengaytirilgan shakli.
DISKURS MODEL I	DISCOURSE MODEL	Tabiiy til bilan interaktiv tizim uchun muloqot asnosida yaratilgan voqealar va ularning bir-biriga aloqadorligi aks

		etgan strukturaviy sxema
DISKURS SEGMENTI	DISCOURSE SEGMENT	Diskursning muayyan alohida tarkibiy qismi. Diskurs segmentining o'ziga xosligi butun diskursni qamrab olgan mavzuning ma'lum bir qismini o'z ichiga oladi.
EUROWORDNET	EUROWORDNET	Ingliz, ispan, italyan, fransuz, german, chex, daniya, va estoniya tillari uchun yaratilgan ko'p tilli elektron ma'lumotlar bazasi
CHEKLI TIL	FINITE-STATE LANGUAGE	Cheklangan holatlar uchun foydalaniladigan til
CHEKLI AVTOMAT MORFOLOGIYASI	FINITE-STATE MORPHOLOGY	O'zgarmas ifodalarga asoslanib morfologik birliklarni morfologik komponentlar asosida chekli avtomat transyuteri yordamida so'zlarning morfotaktikasini o'rganuvchi soha. FST yondashuvi ikki bosqichli morfologiyaga tayanadi.
CHEKLI AVTOMAT TEXNOLOGIYASI	FINITE-STATE TECHNOLOGY	Chekli avtomat transyuterini yaratish va uni qo'llashda turli instrumentlar va metodlardan foydalanish
CHEKLI AVTOMAT TRANSYUTERI	FINITE-STATE TRANSDUCER	Regulyar munosabatlarni aks ettiruvchi FST avtomati
FORMAL GRAMMATIKA	FORMAL GRAMMAR	Tilning matematik strukturasi $G = (N, T, S, P)$. Bu yerda N, T aniq (to'g'ri) alifbo elementlari, S aksioma va P aniq belgilangan qoidalar.
FORMAL TIL	FORMAL LANGUAGE	Alifbodan tuzilgan satrlar jamlanmasi. Tabiiy tilni kompyuterga moslashtirishda (matematika, mantiq,

		kompyuter dasturi uchun) maxsus yaratilgan til. Ushbu tilda sintaktik va semantik munosabatlar uchun mos ravishda maxsus simvol va formulalar yaratiladi.
GENERATSIYA	GENERATION	Tabiiy til matnlarini avtomatik tarzda muayyan semantik, kommunikativ yoki sintaktik ma'lumotlarga ajratish jarayoni
GRAMMATIKA	GRAMMAR	1) (keng ma'noda) yaxlit tizim va til strukturasi; (tor ma'noda) sintaksis va morfologiyaning tizimli majmui; 2) (formal grammatika uchun grammatika) cheklangan noterminal alifbo, cheklangan alifbo, aksioma va chekli to'plamlardan iborat generativ vosita.
IYERARXIYA	HIERARCHY	Muayyan konseptlarning boshqa umumiy konseptlari bilan o'zaro ontologik tasniflangan mantiqiy ketma-ketligi
OMOGRAF	HOMOGRAPH	Yozilishi o'xshash, biroq urg'u va talaffuziga ko'ra farqlanuvchi so'zlar
OMOFON	HOMOPHONE	Talaffuzi o'xshash, yozilishida farqlanuvchi so'zlar
INSON-KOMPYUTER MULOQOTI	HUMAN-COMPUTER INTERACTION	Kompyuter tizimi bilan inson-foydalanuvchi o'rtasidagi aloqani ta'minlovchi interfeys
INSON-INSON MULOQOTI	HUMAN-HUMAN DIALOGUE	Kompyuter vositachi bo'lishi yoki bo'lmasligidan qat'iy nazar ikki yoki undan ortiq kishilar o'rtasidagi muloqot
INSON-MASHINA	HUMAN-MACHINE	Kompyuter tizimi bilan bir yoki undan ortiq kishilar

MULOQOTI	DIALOGUE	o'rtasidagi diolog
GIPONIM	HYPONYM	Umumlashtiruvchi so'zdan farqli ravishda xususiy ma'no kasb etuvchi bir necha so'zlar to'plami. Masalan, <i>chumchuq</i> va <i>kanareyka qush</i> so'zining giponimidir.
MA'LUMOTNI KENGAYTIRISH	INFORMATION EXTRACTION	Matnda tanlangan birliklarni lingvistik bog'liqlik asosida ma'lumot turlarini avtomatik aniqlash jarayoni
BILIMLAR BAZASI	KNOWLEDGE BASE	Ma'lumotlar tizimi, tabiiy tilni qayta ishlash (NLP) dasturi yoki ilovalarda qo'llaniladigan bilimlardan iborat ma'lumotlar bazasi
BILIMGA ASOSLANGAN MASHINA TARJIMASI	KNOWLEDGE-BASED MACHINE TRANSLATION	Sifatli tarjimaga erishish maqsadida muayyan soha haqidagi ontologik bilimlar bilan boyitilgan lingvistik ma'lumotlarga ega mashina tarjimasi metodi
TIL	LANGUAGE	Muayyan jamiyat yoki umumiy kishilar tomonidan foydalaniladigan kommunikatsion tizim. Til tabiiy til (o'zbek tili), formal (kompyuter dasturlash tili yoki mantiqiy tizim) bo'lishi mumkin.
TIL MODELI	LANGUAGE MODEL	Statistik matnlarni qayta ishlash jarayonida so'zlarning ketma-ketlik ehtimolligini belgilovchi model
LEMMA	LEMMA	Asos so'zning morfologik qoidalarga ko'ra olingan barcha paradigmatic shakllari

LEMMATIZATSIYA	LEMMATIZATION	Grammatik shakllangan so'zlarni morfologik jihatdan yetakchi va ko'makchi morfemalarga ajratilishi: <i>bolabolajon, bolalar, bolam, bolagina</i>
LEKSIK BIRLIK	LEXICAL ENTRY	So'z turkumlari, kategoriyalarning ichki turlari, ma'nosi, talaffuzi, atamalarga bog'liq bog'lovchilar yoki leksikondagi turli xil boshqa ma'lumotlari aks etgan so'z yoki frazalar
LEKSIK NOMINAL FRAZALI ANAFORA	LEXICAL NOUN PHRASE ANAPHORA	Anaforalarda aks etgan muayyan otli frazalar
LEKSIK SEMANTIK TARMOQ	LEXICAL SEMANTIC NETWORK	Semantik tizimda (so'zlarning gapdagi qo'llanilishi) leksik birliklarni yonida qo'shni so'zlar bilan qo'llanilishi, shuningdek, bir-biriga o'zaro bog'langan terminlar bazasi. Ingliz tilidagi WordNet bunga misol bo'la oladi.
LEKSIK SEMANTIK MUNOSABAT	LEXICAL SEMANTIC RELATION	Tildagi leksikallashuv jarayoni o'rtasidagi semantik munosabatlar
LEKSIK TRANSYUTER	LEXICAL TRANSDUCER	Morfologik analiz va generatsiyasi uchun mo'ljallangan chekli avtomat. U turlangan yoki tuslangan shakllar (lemmalar)ni leksik kategoriyalarga ko'ra muvofiqlashtiradi.
LEKSIK O'ZGARUVCHA NLIK	LEXICAL VARIABILITY	Bir xil ma'nodagi muayyan so'zning ko'plab turli kontekstlarda yoki vaziyatlarda ko'ra o'zgarishi

LEKSIKALLASH UV JARAYONI	LEXICALIZATION	Avtomatik matn generatsiyasi jarayonida berilgan semantik elementlarning tipik fazasi uchun leksik birliklarning generatsiyalashuv jarayoni
LEKSIKOGRAFIYA	LEXICOGRAPHY	Har bir leksik birlikning orfografiyasi, talaffuzi, so'zshakllari, so'z turkumlari, ma'nosi yoki tarjimasi, qo'llanilishi, frazeologik holati, tarixi va etimologiyasini o'rganuvchi soha
MASHINA O'QIY OLADIGAN LUG'AT	MACHINE-READABLE DICTIONARY	Elektron shaklidagi matndan olingan (odatda inson foydalanishi uchun tuzilgan) lug'at. Mashina o'qiy oladigan lug'at qog'oz shaklidagi lug'atda ifodalanmagan ma'lumotlar bilan ta'minlangan bo'lib, tabiiy tilni qayta ishlash uchun qo'shimcha ma'lumotlardan iborat bo'ladi.
MASHINA TARJIMASI	MACHINE TRANSLATION (MT)	Tabiiy tildagi matnlar o'rtasida inson ishtirokisiz (ba'zan inson ishtiroki yordamida) amalga oshiriladigan avtomatik tarjima jarayoni
BIR TILLI KORPUS	MONOLINGUAL CORPUS	Bir tilga tegishli matnlarning elektron majmuasi
BIR TILLI LUG'AT	MONOLINGUAL DICTIONARY	Yaratilgan lug'atda maqolalarning faqat bir tildagi izohi beriladi.
MORF	MORPH	So'zlarning kichik ma'noli qismi
MORFEMA	MORPHEME	Morfologiyaning kichik ma'noli qismi bo'lib, so'z qurilish vositasidir. U grammatik tahlilning minimal

		birligi hisoblanadi.
MORFOLOGIYA	MORPHOLOGY	So'zning ichki strukturasi, so'z shakllarini o'rganish sohasi
MORFOFONOLOGIYA	MORPHOPHONOLOGY	Morfemalarning voqeanishida fonologiyaning ta'siriga ega bo'lgan lingvistikaning bir qismi
MORFOTAKTIKA	MORPHOTACTICS	So'z shaklining morfem tuzilishini, struktur chegarasi aniqlovchi soha
MT	MT	Mashina tarjimasining qisqartmasi
NLP	NLP	Tabiiy tilni qayta ishlash jarayoni
ONTOLOGIYA	ONTOLOGY	Iyerarxik tarzda tuzilgan bir-biriga bog'liq tushunchalar, obyekt yoki jarayonlarning xususiy xarakterini muayyan soha doirasida tasnifiy berilishi
SO'Z TURKUMI	PART-OF-SPEECH	So'zlarning asosiy grammatik sinflaridan biri: <i>ot, sifat, fe'l</i>
SO'Z TURKUMI TEGI	PART-OF-SPEECH TAG	So'z turkumlarining maxsus shartli belgisi
SO'Z TURKUMI TEGGERI	PART-OF-SPEECH TAGGER	So'zlarning grammatik sinflari uchun kiritilgan belgilarni aniqlovchi kompyuter dasturi
SO'Z TURKUMI TEGINI YOZISH	PART-OF-SPEECH TAGGING	Kompyuter dasturi orqali so'zlarning grammatik sinflarni simvollar bilan belgilash
FONETIKA	PHONETICS	Tilning tovush tizimi, tuzilishi, nutq organi, bo'g'in va uning turi, so'zda tovush o'zgarishi, shuningdek, urg'u va uning o'rnini kabilarni o'rganuvchi soha.
OVOZLASHTIRISH	PHONETIZATION	Fonemalarning ketma-ketligini kompyuterlashtirish orqali so'z, fraza yoki gaplar talaffuz

		qilinadi. Harfni ovozashtirish yoki grafemani fonemaga o'tkazish kabi terminlar alternativ ravishda ishlatiladi.
FONOLOGIYA	PHONOLOGY	Tilda qo'llaniladigan tovushlar, so'z va frazalardagi bo'g'in qismlarini o'rganadigan lingvistika bo'limi.
FRAZA	PHRASE	Matnda yakka birlik sifatida olinuvchi so'zlarning ketma-ketligi
FRAZA STRUKTURASI	PHRASE STRUCTURE	Fraza bo'lib keluvchi so'z turkumlariga tegishli daraxtsimon kesishmalardagi frazalarning iyerarxik joylashgan formal ifodasi.
EHTIMOLLIK MODELLARI	PROBABILISTIC MODELS	Ma'lumot qidiruv tizimida ehtimollik nazariyasiga asoslangan metod turlari. Ehtimollik metodlar orqali elementlar o'rtasidagi o'xshashlik aniqlanadi.
SAVOL-JAVOB TIZIMI	QUESTION ANSWERING (QA)	NLP texnologiyalaridan foydalanib, onlayn hujjatlarning katta hajmidagi to'plamlaridan foydalaniladigan savol-javob tizimi
SXEMA	SCHEMA (PLURAL SCHEMATA)	Matn strukturasi bo'lib, sxemalar strukturalarning ichki tuzilishi va lingvistik voqelanishi, matn mazmuniga doir muayyan ma'lumot beruvchi komponentlarni o'zida aks ettiradi.
SEGMENT BIRLIK	SEGMENT	Segment birliklar – so'z yoki morfema tarkibida gorizontaal chiziq bo'ylab birin-ketin

		keladigan birliklardir. Ularga nutq tovushi, bo'g'in, fonetik so'z, takt va fraza kiradi.
SEGMENTLASH	SEGMENTATION	Yozma matnni so'z, gap va paragraf kabi ma'noli qismlarga ajratish jarayoni.
UMUMIY BILIMLARNING YUZAKI TAHLILI	SHALLOW PARSING SHARED KNOWLEDGE	Fe'l va to'ldiruvchi yoki predlogli fraza kabi faqat tanlab olingan sintaktik strukturalarga ko'ra matnni sintaktik analiz qilish jarayoni.
NUTQNI TUSHUNISH	SPEECH RECOGNITION	So'zlardagi ketma-ket kelgan nutq tovushlarining transkripsiyasini o'qish jarayoni
OVOZLI KORPUS	SPOKEN CORPUS	Og'zaki tabiiy nutqda sharoitida qo'llanilgan so'zlarni qidirish va tahlil qilishga mo'ljallangan korpus. Maxsus yozib olingan og'zaki matnlarning transkripsiyasi beriladi.
OG'ZAKI MASHINA TARJIMASI	SPOKEN LANGUAGE MT	Og'zaki nutqni avtomatik tarjima qiluvchi tarjima texnologiyasi
SPONTAN NUTQ	SPONTANEOUS SPEECH	Oldindan tayyorgarlichsiz yoki yozma ishora birliklarsiz erkin shakllantirilgan nutq shakli (masalan, dialog shaklidagi yuzma-yuz yoki telefondagi suhbat)
STATISTIK MASHINA TARJIMASI	STATISTICAL MACHINE TRANSLATION	Parallel korpusdan olingan statistik asosga ko'ra bo'lishi mumkin bo'lgan tarjimani amalga oshiruvchi mashina tarjimasi metodi
QISQARTIRISH	SUMMARIZATION	Foydalanuvchi yoki dasturiy ta'minot uchun ma'lumotning

		asosiy mazmunini olish uchun matnni qisqartirish texnologiyasi
LINGVISTIK STRUKTURALAR	LINGUISTIC STRUCTURES	Shakllangan gaplardagi barcha so'zlarning o'zaro sintakmatik munosabatlari nazarda tutiladi.
SINTAKSIS	SYNTAX	Predikat-argument munosabatlarini aks ettiruvchi so'z yoki frazalarning tashqi tuzilishi va funksiyalarni o'rganuvchi bo'lim
TEG	TAG	So'zlarning kontekstda semantik va boshqa holatlarga ko'ra so'z turkumi sifatida beriluvchi grammatik belgisi
TEG TO'PLAMI	TAG SET	1) kompyuterda tilni qayta ishlash uchun matnni teglashda foydalaniladigan belgilar to'plami; 2) lug'atga birliklarni tizimli tarzda tuzishga yordam beruvchi XML yoki boshqa belgilar to'plami
TEGGING	TAGGING	Matndagi so'zlar va ifodalar uchun teglarning berilishi
TARJIMA QILINGAN TIL	TARGET LANGUAGE	Mashina tarjimasi tizimida tarjima qilingan matn tili
TARJIMA QILINGAN SO'Z	TARGET WORD	Lug'at yoki unga bog'liq holda matn tarkibida tarjima qilingan so'z
TAKSONOMIYA	TAXONOMY	Obyektning iyerarxik tasnifi yoki tasnifning asosiy tamoyillari aks etgan munosabatlar
TERMIN CHASTOTASI	TERM FREQUENCY	Muayyan hujjatdagi so'z yoki termin chastotasining o'lchov birligi. Termin chastotasi hujjat tarkibida qanchalik ishlatilganligini bildiradi.

TERMINOLOGI K BOSHQARUV	TERMINOLOGY MANAGEMENT	Terminologiyaning ko'ptilli ma'lumotlar bazasini qidirish va yaratishning kompyuterli tizimi
MATNNI TASNIFLASH	TEXT CATEGORIZATION	Moliya, sport, yangilik va boshqa sohalarga doir matnlarni kategoriya yoki turlarga ajratish
MATNNI OVOZLASHTIRISH	TEXT-TO-SPEECH SYNTHESIS	Elektron formatdagi matndan tabiiy nutq tovushlarini avtomatik talaffuz qilish texnologiyasi
MATNNI NUTQQA O'TKAZUVCHI SINTEZATOR	TEXT-TO-SPEECH SYNTHESIZER	Matnni ovozlashtiruvchi dasturiy ta'minot
TOKENIZATSIYA	TOKENIZATION	Matnni so'z, tinish belgi, raqamlar kabi grammatik birliklarning segmentlash jarayoni
TOKENAYZER	TOKENIZER	Matndagi alohida tokenlar (so'z, raqam, tinish belgilari) uchun matnni tokenizatsiya qiluvchi dasturiy ta'minot
TRANSFER TIZIMI	TRANSFER SYSTEM	Tarjima qilinadigan tilning dastur generatsiyasi bilan asos til uchun mashina tarjimasi dasturining analiz, transfer va generatsiya kabi bosqichlaridan iborat texnologiya
TRANSFORMATSIYA	TRANSFORMATION	Muayyan tilning tashqi strukturasi kelib chiqib, transformatsion grammatikadagi biror qoidalardan foydalangan holda gapning ichki strukturasi aylantirilishi
TARJIMA	TRANSLATION	Matnlarning tarjima qilingan

XOTIRASI	MEMORY	namunalari bilan avvalgi tarjima qilingan ma'lumotlar bazasidan foydalanuvchi instrument hisoblanadi. Tarjimon keyingi matni tarjima qilishda ushbu namunalardan foydalanishi mumkin.
SINTAKTIK MA'LUMOTLAR BAZASI / TREEBANK	TREEBANK	Matn korpusining sintaktik analizi
MAXSUS BELGILI STRUKTUR GRAMMATIKA	TYPED FEATURE STRUCTURE GRAMMAR	Ushbu grammatikada graflar maxsus belgilar (yoki belgilar to'plami) sifatida kelganda frazaning sintaktik, semantik, fonologik, (ayrim holatlarda) pragmatik / kontekstual xususiyatlari aks etadi.
UNIVERSAL TIL	UNIVERSAL LANGUAGE	Formal til nazariyasi va avtomatda bo'sh satrdan va muayyan uzunlikdagi barcha satrlarni jamlovchi til.
UNIVERSALIYALAR	UNIVERSALS	Mashina tarjimasi transfer yoki interlingval tizimiga kiritilgan barcha tabiiy tillardagi umumiy tomonlari olingan til elementlariga aytiladi.
FLEKSIYA	FLECTION	Turlanish yoki tuslanish natijasida so'zning oxirgi o'zgaradigan qismi
KORPUS LEKSIKOGRAFIYASI	CORPUS LEXICOGRAPHY	Ma'lumotlarni saqlash, uzatish va analiz kabi kompyuterning muayyan fazasida ko'p qo'llaniladigan matn yoki akustik korpusi asosida tilni o'rganish sohasi hisoblanadi.
EKSPERT TIZIMI	EXPERT SYSTEM	Ekspert tizimi muayyan soha doirasida biror masalani

		yechishi yoki tavsiya berish uchun muayyan bilimga asoslangan kompyuter dasturi hisoblanadi.
KOMPYUTER LINGVODIDAKTIKASI	COMPUTATIONAL LINGVODIDACTICUS	Til bilan bog'liq bo'lgan o'qitish va o'rganish jarayonini avtomatlashtirishga yo'naltirilgan soha.
MOODLE	MOODLE – massive open online course	Platformada o'qituvchi va talaba o'rtasida interaktiv ta'limni amalga oshiruvchi muhit
MOBIL QURILMA YORDAMIDA TIL O'RGANISH	MALL – mobile assisted language learning	Mobil yordamidagi til o'qitish texnologiyasi
KOMPYUTER YORDAMIDA TIL O'RGANISH	CALL-COMPUTER ASSISTED LANGUAGE LEARNING	Tilni horijiy til sifatida o'qitish va o'rganish imkoniyatini kengaytirishda kompyuter dasturiy vositalaridan foydalanish nazarda tutiladi.
ELEKTRON LUG'ATLAR YORDAMIDA TARJIMA	MAHT-MACHINE-ASSISTED HUMAN TRANSLATION	Inson tomonidan matni tarjima qilishda kompyuterning elektron lug'atlar va unga kiritilgan ko'rsatmalaridan foydalanish
KOMPYUTER YORDAMIDAGI TARJIMA	CAT – COMPUTER ASSISTED-TRANSLATION	Kompyuter texnologiyalari yordamida inson tomonidan amalga oshirilgan tarjimani qulaylashtirish va tezlik va sifatini oshirishga xizmat qiluvchi texnologiya
MUHARRIR YORDAMIDAGI MASHINA TARJIMASI	HAMT (HUMAN-ASSISTED MACHINE TRANSLATION)	Kompyuter dasturiy ta'minoti orqali amalga oshiriladigan tarjimaga inson resursi tomonidan tahrir qilish vazifasini yuklaydi.
TO'LIQ AVTOMATIK	FAMT- FULLY AUTOMATIC	Tarjimon dasturi orqali matni ham tarjima, ham tahrir qilish

MASHINA TARJIMASI	MACHINE TRANSLATION	vazifalarini amalga oshiradi.
MASHINA TARJIMASI	MACHINE TRANSLATION	Bir tildagi matnni ikkinchi tilga avtomatik tarzda og'zaki yoki yozma tarzda tarjima qilish jarayoni
KORPUS ASOSIDAGI MASHINA TARJIMASI	CORPUS BASED MT	Parallel matnlar asosida tarjima qilish texnologiyasi
GIBRID MASHINA TARJIMASI	HYBRID MACHINE TRANSLATION	Ham statistik ham qoidaga asoslangan tarjima qilish texnologiyasi
QOIDAGA ASOSLANGAN MASHINA TARJIMASI	RULE BASED MT	Lug'at va muayyan qoidalarga asoslangan mashina tarjimasi texnologiyasi
INTERLINGVAL TIZIM	INTERLANGUA L MT	Bir tildan ikkinchi tilga tarjima qilishda o'rtada vositachi til orqali tarjima amalga oshirish mexanizmi
ALPAC	AUTOMATIC LANGUAGE PROCECCING ADVICORY COMMITTY	Tilni avtomatik qayta ishlash bo'yicha maslahat komiteti
METATIL	METALANGUA GE	matematik modellashtirilgan formal til
TARJIMA XOTIRASI	TRANSLATION MEMORY	asos segment (jumla) va matn tarjimasi saqlanib, lingvistik ma'lumotlar baza
PROMT	PROGRAMMER'S MACHINE TRANSLATION	Avtomatik tarjima dasturiy ilovasi
NAMUNAGA ASOSLANGAN MT	EXAMPLE BASED MT	Oldindan qilingan tarjima namunalar bazasi yoki korpusga asoslanadigan mashina tarjimasi texnologiyasi

UNIVERSAL SINTAKTIK KATEGORIYAL AR	UNIVERSAL SYNTACTIC CATEGORIES	Tilni avtomatik qayta ishlash bo'yicha maslahat komiteti
OPERATSION SINTAKSIS	OPERATIONAL SYNTAX	matematik modellashtirilgan formal til
TARJIMA XOTIRASI	TRANSLATION MEMORY	asos segment (jumla) va matn tarjimasini saqlovchi lingvistik ma'lumotlar baza
OMONIMIYA	HOMONYMY	Shakli o'xshash ma'nosi tubdan farqlanuvchi til va nutq birliklari
POLISEMIYA	POLISEMY	So'zning ko'chma ma'noda qo'llanilishi
PARALLEL MATN	PARALLEL TEXT	Ikki tilda aynan bir mazmunda tarjima qilingan matnlar majmuasi
KO'P TILLI PARALLEL KORPUSLAR	MULTILINGUAL PARALLEL CORPORA	Bir necha tillarning tarjima qilingan electron matnlarning majmuasi
MUQOBILLIK	ALIGNMENT	Bir tildagi til birligi ikkinchi tilga tarjima qilinganda unga mos keluvchi lingvistik jihatdan mos kelishi
NEYRO MASHINA TARJIMASI	NEURO MACHINE TRANSLATION	Statistik texnologiya asosida amalga oshiriladigan mashina tarjimasi
KORPUS	CORPUS	Muayyan turga (uslub va janr...) tegishli matnlarning elektron majmui
KORPUS LEKSIKOGRAFIYASI	CORPUS LEXICOGRAPH Y	Ma'lumotlarni saqlash, uzatish va analiz qilish kabi vazifasiga bajarishga yo'naltirilgan kompyuterning muayyan fazasida ko'p qo'llaniladigan matn yoki akustik korpusi asosida tilni o'rganish sohasi

MATN FRAGMENTI	CHUNK	Parallel (tarjima) korpusda biror tilga tegishli matnning muayyan qismi.
FE'LLI FRAZEMA	PHRASAL VERBS	Ikki yoki undan ortiq analitik shaklda hosil bo'lgan fe'lli birikmalar
TURG'UN BIRIKMA	IDIOMA	Ikki yoki undan leksik birliklarni birikuvidan hosil bo'lgan til birligi
LEKSIKOLOGIY A	LEXICOLOGY	So'zlarning shakl va ma'nosini o'rganuvchi bo'lim.
LINGVISTIK MODEL	LINGUISTIC MODEL	Til birligi sifatida qabul qilingan lisoniy qurilma
KONNOTATIV MA'NO	CONNOTATIVE MEANING	So'zlarning ko'chma ma'nosi
DENOTATIV MA'NO	DENOTATIVE MEANING	So'zlarning kontekstdan tashqari ma'nosi

Shartli qisqartmalar:

| (mantiqiy "yoki" belgisi)
Art- artikl (article)
CN- qo'shma ot (compound noun)
EQ- egalik qo'shimchalari
F-fe'l
FST – finite state transducer
KL – kompyuter lingvistikasi
KQ- ko'plik qo'shimchasi
MT – mashina tarjimasi (machine translation)
N (vow)- unli bilan tugovchi ot
N(cons.)- undosh bilan tugovchi ot
n0-1– birinchi elementdan ilk elementni olish
N1 – irregular plural noun
NLP – natural language processing
N- ot(noun)
NPL- ko'plikdagi ot (noun plural)
NSG- birlikdagi ot (noun singular)
Nta– so'z o'zagining oxirgi qismidagi undosh muayyan harfga o'zgaruvchi otlar
Nto– so'z o'zagining oxirgi qismiga 'y' undoshi qo'shiluvchi otlar
Ntt – o'zagida ikkinchi bo'g'in unlisi tushib qoluvchi otlar
O- olmosh
OtBK- bosh kelishikdagi ot

OtChK- chiqish kelishikdagi ot
OteQ- egalik qo'shimchasi bilan kelgan ot
OtJK- jo'nalish kelishikdagi ot
OtoK- o'rin-payt kelishikdagi ot
OtoK- qaratqich kelishikdagi ot
OttK- tushum kelishikdagi ot
POS – Part of speech
PP- possessive pronoun
P- pronoun
Prt.- Participle
R- ravish
Rsh- ravishdosh
Ssh- sifatdosh
S- sifat
UT – universal tobelik
P- Pronoun
Num- Numeral
V- verb
G- gerund
PP- participle
Adv.- adverb
Adj.- adjective
ADP- adposition

Shartli belgilar:

-ixtiyoriy qatnashish
 {}-bo'lishi mumkin
 =>ekvivalentlik hosil qiladi
 -tegishli
 Ø-bo'sh to'plam
 // izoh belgisi
 |-yoki

MUNDARIJA

KIRISH	3
1-MODUL KOMPYUTER LINGVISTIKASINING RIVOJLANISH TENDENSIYASI	
1-MAVZU. "KOMPYUTER LINGVISTIKASI" FANIGA KIRISH	
1.1. Fanning maqsadi va vazifasi.....	6
1.2. Kompyuter lingvistikasi fan sifatida rivojlanishida nazariy va amaliy tadqiqotlar.....	12
1.3. Kompyuter lingvistikasining boshqa fanlar bilan aloqasi (matematika, psixologiya, informatika, dasturlash asoslari)...	18
2-MAVZU. KOMPYUTER LINGVISTIKASINING ASOSIY YO'NALISHLARI	
2.1. Kompyuter lingvistikasi yo'nalishlaring nazariy mazmuni	25
2.2. NLP texnologiyasi va lingvistik instrumentlar.....	28
2-MODUL. KOMPYUTER LINGVISTIKASIDA LINGVISTIK MASALALAR	
3-MAVZU. GRAMMATIK NAZARIYALAR	
3.1. Formal til nazariyasi.....	46
3.2. Chomskiy iyerarxiyasi.....	49
3.3. Tobelanish nazariyasi.....	55
4-MAVZU. LINGVISTIK PROTSESSORNING GRAMMATIK KOMPONENTI	
4.1. Lingvistik bilimlar bazasi.....	63
4.2. Lingvistik bilimlar bazasini loyihalash.....	65
4.3. Morfologik ma'lumotlar bazasi.....	68
3-MODUL. GRAMMATIK TAHLIL QILISHNING KOMPYUTERLI YONDASHUVI	
5-MAVZU. MORFOLOGIK TAHLILNING AVTOMATIK TIZIMI	
5.1. O'zbek tilining morfologik qonuniyatlari.....	87
5.2. Tabiiy tilning morfologik modeli.....	93
5.3. Morfologik analiz.....	112
6-MAVZU. SINTAKTIK TAHLILNING AVTOMATIK TIZIMI	
6.1. Sintaktik strukturalar (modellar).....	135

6.2. Diskurs va matn segmenti.....	141
6.3. Sintaktik tahlil (parsing).....	154
4-MODUL. KOMPYUTER LINGVISTIKASI YO'NALISHLARI VA METODLARI	
7-MAVZU. KOMPYUTER LEKSIKOGRAFIYASI	
7.1. Lug'at tipologiyasi.....	163
7.2. Tezaurus. WordNet va terminologik lug'atlar.....	170
7.3. Ontologik modellar.....	187
7.4. Terminologiya va terminografiya.....	192
8-MAVZU. KOMPYUTER LINGVODIDAKTIKASI	
8.1. Kompyuter texnologiyalari yordamida dars berishning didaktik konsepsiyasi.....	207
8.2. Til ta'limida AKTdan foydalanish.....	212
9-MAVZU. MASHINA TARJIMASI RIVOJLANISH TENDENSIYASI	
9.1. Mashina tarjimasining shakllanishi va taraqqiyot bosqichlari.....	220
9.2. Sohada olib borilayotgan tadqiqotlar.....	232
9.3. Mashina tarjimasining texnologiyasi va tizimlari.....	240
10-MAVZU. MASHINA TARJIMASINING LINGVISTIK MUAMMOLARI	
10.1. Tarjima dasturi uchun lingvistik ta'minot va tahlil tamoyillari.....	255
10.2. Omonimiya va polisemiya lingvistik masalalar sifatida..	262
10.3. Mashina tarjimasining pragmatik va semantik muammolari.....	267
10.4. Tarjima algoritmi.....	269
11-MAVZU. KORPUS LINGVISTIKASI	
11.1. Korpus lingvistikasining nazariy konsepti.....	284
11.2. Korpus turlari	288
11.3. Korpus annotatsiyasi	296
12-MAVZU. KOMPYUTER LINGVISTIKASIDA METODLAR	
12.1. Statistik metod.....	312
12.2. Modellashtirish metodi.....	315
12.3. Korpusga asoslangan metod.....	319

**13-MAVZU. KOMPYUTERDA MATNLARNI QAYTA
ISHLASH TIZIMLARI (NLP)**

13.1. Sun'iy intellekt texnologiyasi.....	325
13.2. Matn interfeysi va generatsiyasi.....	328
13.3. Qidiruv va ekspert tizimlari.....	330
13.4. Savol-jabob tizimi.....	332
13.5. Nutq sintezatori.....	335
13.6. Spell-checker (tahrirlovchi dastur) texnologiyasi.....	339
13.7. Sentiment analiz.....	346
Adabiyotlar.....	353
Glossariy.....	372

Abduraxmonova Nilufar Zaynobiddin qizi

KOMPYUTER LINGVISTIKASI

darslik

“Nodirabegim” nashriyoti

Nashriyot litsenziyasi AI № 313. 24.11.2017 y.

Bosishga ruxsat etildi: 12.03.2021.

“Times New Roman” garniturasini. Qog'oz bichimi: 60x84 1/16

Nashriyot bosma tabog'i 24,6. Adadi 500 nusxa.

100129, Toshkent shahri, Shayxontohur tumani,

Navoiy kochasi, 30-uy.

OOO “AKTIV PRINT” bosmaxonasida chop etildi.

Toshkent, Chilonzor 25, Lutfiy 1A.



**Abduraxmonova Nilufar
Zaynobiddin qizi**

Filologiya fanlari bo'yicha falsafa doktori (PhD), dotsent. Kompyuter lingvistikasi, mashina tarjimasi, korpus lingvistikasi bo'yicha mutaxassis.

“Biz kutgan bahor” she'riy to'plam (2003), “Mashina tarjimasining lingvistik asoslari” o'quv qo'llanma (2012), “O'zbek tilidan universal qo'llanma” (hammualliflikda, 2013), “English-Uzbek dictionary of phrasal verbs” lug'at (2017), “Mashina tarjimasining lingvistik ta'minoti” monografiya (2018),

“Kompyuter lingvistikasi terminlarining izohli lug'ati” elektron lug'at (2018), shuningdek, yuzdan ortiq ilmiy maqola, tezis hamda o'n beshga yaqin baza va dasturiy ta'minot uchun patent va ilmiy guvohnomalar muallifi.

ISBN 978-9943-6940-4-0



9 789943 694040