

**O‘ZBEKISTON RESPUBLIKASI
OLY VA O‘RTA MAXSUS TA‘LIM VAZIRLIGI
TOSHKENT DAVLAT O‘ZBEK TILI VA ADABIYOTI
UNIVERSITETI**

ABJALOVA MANZURA ABDURASHETOVNA

**TAHRIR VA TAHLIL
DASTURLARINING LINGVISTIK
MODULLARI**

Monografiya

**“Nodirabegim” nashriyoti
Toshkent – 2020**

KBK: 81.320‘z

A12

UO‘K: 811.512.133‘42

ISBN 978-9943-6939-0-6

Abjalova, Manzura

Tahrir va tahlil dasturlarining lingvistik modullari. [Matn] : monografiya / M.A. Abjalova. - Toshkent: Nodirabegim, 2020. – 176 b.

Mazkur monografiyada kompyuter texnologiyalari yordamida matnlarni tahrir va tahlil qiluvchi dastur hamda tizimlarning yaratilish tarixi, imkoniyatlari, lingvistik modul va modullashning ahamiyati asoslangan, o‘zbek tilidagi matnlarni avtomatik tahlil qilish dasturining lingvistik modullari: grafematik tahrir, morfologik tahlil va sintaktik tahlil modullarining lingvistik ta‘minotini yaratishning optimal tamoyillari, morfologik modulda ismlar va fe‘l turkumi uchun morfonologik, orfografik qoidalari algoritmlari, so‘z turkumlarini formallashtirish qoidalari algoritmlari ishlab chiqilgan.

Monografiya filologiya va kompyuter lingvistikasi yo‘nalishida tahsil olayotgan talabalar, magistrantlar, doktorantlar, ilmiy izlanuvchilarga hamda matnni qayta ishlash (NLP) bilan bog‘liq muammolar bilan qiziquvchi keng kitobxonlar ommasiga mo‘ljallangan.

Mas‘ul muharrir:

Dadaboyev H.A.

Filologiya fanlari doktori, professor

Taqrizchilar:

Usmonova Sh.

Filologiya fanlari doktori, professor

Xolmonova Z.

Filologiya fanlari doktori, professor

Monografiya Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti Ilmiy-texnikaviy kengashining 2020-yil 24-oktabrdagi 3-sonli yig‘ilishida nashrga tavsiya etilgan.

© M.A. Abjalova.

© “Nodirabegim” nashriyoti, 2020.

Manzura Abjalova

**LINGUISTIC MODULES OF THE TEXT
EDITING AND ANALYSIS PROGRAM**

Monograph

This monograph describes the history and possibilities of computer-assisted editing and analysis systems, as well as the scientific and theoretical description of the creation of linguistic modules of the Uzbek text analysis program, the degree of study of this problem in the field of natural language processing, automatic editing and analysis systems and technologies are analysed. Theoretical bases of formation of linguistic support of automatic editing and analysis system in Uzbek computer linguistics are created, optimal principles of automatic editing and analytical linguistic support are developed, based on the importance of linguistic module and modulation, Grammatical rules algorithm is developed.

The monograph is intended for students, doctoral students, researchers in the field of philology and computer linguistics, as well as a wide range of readers interested in the problems associated with text processing.

Editor-in-Chief:

Dadaboyev H.A.

doctor of philological sciences, professor

Reviewer:

Usmonova Sh.

doctor of philological sciences, professor

Xolmonova Z.

doctor of philological sciences, professor

QISQARTMA SO‘ZLAR RO‘YXATI

Kompyuter lingvistikasi – KL

Avtomatik tahrir va tahlil qilish – ATT

Lingvistik modullar – LM

Dasturiy ta’minot – DT

Lingvistik ta’minot – LT

Avtomatik tahrir va tahlil sistemasi – ATTS

Grafematik tahrir – GT

Morfologik tahlil – MT

Sintaktik tahlil – ST

Sintaktik tahlil moduli – STM

KIRISH

Jahon tilshunosligida XX asrning 50-yillaridan til va matn muammolari bilan bog‘liq masalalar kompyuter texnologiyalari yordamida hal qilinmoqda. Natijada matnni avtomatik tushunish, mashina tarjimasini, matnni referatlash (gipermatnni asosiy mazmunni o‘zida saqlagan minimatnga keltirish, ya’ni matn hajmini qisqartirish, uni qisqa bayon holiga aylantirish), tasniflash (mavzu, uslub va janr jihatidan o‘zaro yaqin matnlarni guruhlash), matnni tahrir va tahlil qilish, matnni generatsiyalash (bir nechta tabiiy tilga oid hujjatli matnlardan formula, texnik ishlanmalar, dasturiy tizimlarni yig‘ish), matndan ma’lumotni olish, og‘zaki nutqni kodli ma’lumotda berish va aksincha, yozma nutqni og‘zaki nutqqa aylantirish, o‘zga tillar lingvistik korpusini yaratish, ontologik lug‘atlar bazasini shakllantirish kabi dolzarb masalalar kompyuter lingvistikasining mundarijasini egallagan va tabiiy tilni qayta ishlash (Natural Language Processing / NLP) sohasida asosiy vazifa hamda yo‘nalishlariga aylangan. Jumladan, matnni elektron hisoblash mashinalari vositasida tahrir qilish muammosi o‘rganila boshlandi, natijada zamonaviy axborot texnologiyalari dasturlari hamda tabiiy tilning grammatik qonuniyatlariga asoslangan matn tahrir va tahlillarining lingvistik ta’minotini yaratishga qaratilgan salmoqli tadqiqotlar yuzaga keldi. Bu asosda yuz bergan texnologik taraqqiyot dunyodagi yetakchi adabiy tillarda matnni kvantitativ tahrir va tahlil qilish tizimlarini yaratishga olib kelgani ma’lum.

Dunyo tilshunosligida avtomatik tahrirlash va tahlil qilish yo‘nalishida imlo tekshiruvchi (spelling / spell checker), morfologik (morfoanalizator) va sintaktik tahlil qiluvchi (parsing) tizimlarning lingvistik ta’minotlari ishlab chiqilgan. Ular til materialini tahrir va tahlil qiluvchi tezkor hamda iqtisodiy tejamkor tizimlar va dasturlarning yaratilishiga asos bo‘lishi barobarida mashina tarjimasini sifatini oshirishda ham muhim omil sanaladi. Zero, globallashtirish va jadal rivojlanayotgan fan-texnika davrida katta hajmdagi axborot ustida tezlik bilan ishlash va samarali natija olish juda zarur. Shu ma’noda, dunyo tilshunosligi fani oldiga tilni lingvistik modellashtirish, so‘zlarni lemmalash, tahrir va tahlil algoritmini tuzish hamda lingvoanaliz dasturining lingvistik modullarini yaratish vazifasi qo‘yilmoqda.

O‘zbek tilshunoslik sohasida XXI asrning boshlarida kompyuter lingvistikasi (KL) kursining universitetlarda o‘quv jarayoniga joriy qilinishi globallashtirish, optimallashtirish va standartlashuv davrida tilning ijtimoiy ahamiyatiga e’tibor oshganligidan dalolat. Shuningdek, “...ilmiy tadqiqot va innovatsion faoliyatni rivojlantirish, ushbu jarayonda iqtidorli

yoshlar ishtirokini, ijodiy g'oya va ishlanmalarni har tomonlama qo'llab-quvvatlash"¹ga alohida e'tibor qaratilmoqda. O'zbek tilshunosligida matnlarni axborot texnologiyalari dasturlari orqali tanlash va uni qayta ishlash bo'yicha bir qancha ish amalga oshirilgan bo'lsa-da, bu yo'nalishlardagi ilg'or jahon tajribasini tadqiqotlarda samarali qo'llash hamda matn tahrir va tahlillarining lingvistik ta'minotini yaratish ehtiyoji mavjud. Zero, "...endigi vazifamiz – ilmiy tadqiqotlarning natijalarini amaliyotga tatbiq etishga qaratilgan samarali mexanizmlarni yaratish, ilm-fanni, ilmiy-innovatsion faoliyatni barqaror rivojlantirish"dir². O'zbek KL rivojlanishining yangi yo'liga kirgan, jahon KLdagi ilg'or an'analarni tadqiq qilib, o'z yo'nalishiga singdirayotgan davrda KLning muammolarini to'g'ri belgilash va ularni ko'rsata bilish katta muvaffaqiyatlar sari qo'yilgan salmoqli qadamdir. Jumladan, ayni damda o'z yechimini kutayotgan o'zbekcha matnlarni avtomatik tarzda tahrir va tahlil qilish tizimining lingvistik ta'minoti (LT)ni yaratish soha mutaxassisi oldida turgan muhim vazifa hisoblanadi. Hozirgi davrda dunyo miqyosida matn bilan bog'liq texnologiyalarning rivojlanganligi va lingvistik dasturlarning yuzaga kelganligi o'zbek tili uchun yaratiladigan avtomatik tahrir va tahlil (ATT)ning taraqqiyoti hamda bugungi holatini o'rganish, matn tahlili dasturining lingvistik ta'minotini tuzishning umumiy tamoyillarini aniqlash, bu xususda yangicha yondashuvlar asosida tadqiqotlar olib borish muhim ahamiyat kasb etadi.

Xorijda kompyuter lingvistikasi sohasida matnlarni ATT qilish masalasi ko'plab olimlar tomonidan o'rganilgan³. J.Peterson xatolarni tekshiruvchi va ularni tuzatuvchi dasturiy ta'minot (DT)lar borasida so'z yuritgan. Matnning avtomatik tahlil jarayonini lingvistik aspektda o'rganish D.Jurafskiy nomi bilan bog'liq. Olimning matnlarni avtomatik

¹O'zbekiston Respublikasi Prezidenti Shavkat Mirziyoyevning Oliy Majlisga Murojaatnomasi. – Toshkent: O'zbekiston, 2018. – B. 23. / prezident.uz/uz/lists/view/137).

² Mirziyoyev Sh. Milliy taraqqiyot yo'limizni qat'iyat bilan davom ettirib, yangi bosqichga ko'taramiz. I jild.– Toshkent: O'zbekiston, 2017. – B. 168-174.

³Peterson, James. Computer Programs for Detecting and Correcting Spelling Errors. 1980.; Do Prado H.A. Emerging Technologies of Text Mining: Techniques and Applications / Ed. by H. A. Do Prado, E. Ferneda. — Idea Group Reference, 2007. – 358 p.; Earnest, Les. The First Three Spelling Checkers. Stanford University. Archived from the original (PDF) on 22 October 2012.; Earnest, Les. Visible Legacies for Y3K (PDF). Archived from the original.; de Amorim, R.C.; Zampieri, M. Effective Spell Checking Methods Using Clustering Algorithms. Proceedings of Recent Advances in Natural Language Processing. Hissar, Bulgaria. 2013. – P. 172-178.; Zampieri, M.; de Amorim, R.C. Between Sound and Spelling: Combining Phonetics and Clustering Algorithms to Improve Target Word Recovery. Proceedings of the 9th International Conference on Natural Language Processing (PolTAL). Lecture Notes in Computer Science (LNCS). Springer. 2014. – P. 438-449.

qayta ishlash, avtomatik tahlil borasidagi tadqiqotlari mahsuli⁴ keyingi izlanishlarga tamal toshini qo‘ygan.

Avtomatik tahrirlash yo‘nalishida avtomatik tahrir qiluvchi dasturlar ishlab chiqilgan, ular ruscha, inglizcha, nemischa, fransuzcha, italyanacha matnlarni tahrirlashga yo‘naltirilgan. Hatto tatar, boshqird, qozoq, qirg‘iz, ozarbayjon kabi turkiy tillarga mansub matnlar uchun morfologik va sintaktik analizatorlar yaratilgan, MS WORD sistemasiga avto-tahrir paketlari joylashtirilgan⁵. Bunday tahrirlash dasturlari, asosan, imloni tekshirishga mo‘ljallangan. Rus tilini avtomatik tahlil qilish yo‘nalishida R.G.Kotov, V.E.Berzon, V.G.Britvin, I.A.Melchuk, A.B.Kuznetsov, L.I.Belyayeva, V.A.Chijakovskiy, G.G.Belonogov, I.S.Duganova, L.L.Iomdin, E.K.Lavoshnikova, A.V.Sokirko, I.M.Nojoy, E.I.Bolshakova, D.V.Klishinskiy, N.N.Leonteva⁶ va boshqalar samarali tadqiqot olib borishgan.

⁴ Jurafsky D., Martin J. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. – New Jersey: PrenticeHall, 2000. – 927 б.

⁵Орехов Б.В., Слободян Е.А. Проблемы автоматической морфологии агглютинативных языков и парсер башкирского языка [Текст] // Информационные технологии и письменное наследие: материалы международной научной конференции (Уфа, 28–31 октября 2010 г.) / отв. ред. В. А. Баранов. – Уфа. Ижевск: Вагант, 2010. – С. 167-171.; <http://lcpb.bashedu.ru/index.php?go>. Орехов Б., Галлямов А. Башкирский морфологический анализатор (Электронный ресурс); <http://my.ict.nsc.ru/~salerat/kaz/>. Программа генерации словоформ казахского языка.; <http://tamgasoft.kg/morfo/ru/>. Морфология кыргызского языка.; Сулейманов, Д. Ш. Синтаксический анализатор предложений татарского языка / Д. Ш. Сулейманов, А. Р. Гатиатуллин // Труды Математического центра имени Н. И. Лобачевского. – Казань, 1999. – Т. 4. – С. 111-126.

⁶Котов Р.Г. Лингвистические аспекты автоматизированных систем управления. – Москва: Наука, 1977. Он же. Прикладная лингвистика и информационная технология. – Москва: Наука, 1987.; Котов Р.Г., Якунин Б.В. Язык информационных систем. – Москва: 1989.; Он же. Оптимизация речевого воздействия. – Москва: Наука, 1990.; Берзон В.Е. и др. О разработки подсистемы редактирования выходного текста в рамках системы МП // Проблемы внутренней динамики речевых норм. – Минск, 1992, – С. 169-174; Бритвин В.Г. Прикладное моделирование синтагматической семантики научно-технического текста (на примере автоматического индексирования): Дисс. канд. филол.наук. – Москва: МГУ, 1983.; Мельчук И.А. Порядок слов при автоматическом синтезе русского слова (предварительные сообщения) / Научно-техническая информация. 1985, № 12. – С. 12-36.; Беляева Л.И., Чижаковский В.А. Тезаурус в системах автоматической переработки текста. – Кишинев, 1983.; Белоногов Г.Г., Котов Р.Г. Автоматизированные информационно-поисковые системы. – Москва, 1968; Белоногов Г.Г., Дуганова И.С., Кузнецов А.Б. Экспериментальная система автоматизированного обнаружения и исправления ошибок в тексте / НТИ. Серия 2, 1984, № 3. – С. 20-25.; Иомдин Л.Л. Автоматическая обработка текста на естественном языке: модель согласования. – Москва: Наука, 1990. – 180 с.; Лавошникова Э.К. О «подводных камнях» в компьютерных системах проверки правописания // Вестник Московского университета. Серия 9. Филология. 2002, – № 6. – С.151-163.; Сокирко А.В. Семантические словари в автоматической обработке текста: По материалам системы ДИАЛИНГ: Дисс. ... канд.тех.наук. – Москва, 2001. – 120 с.; Ножов И.М. Морфологическая и синтаксическая

Turkiy tillarga mansub matnlarni statistik metodlar asosida lingvistik tahlil qilish sohasida R.G.Piotrovskiy, M.A.Mahmudov, A.A.Piotrovskaya, T.Sadikov ishlari diqqatga sazovor⁷. O‘zbek tili materiallari bo‘yicha KLga oid dastlabki tadqiqotlarni olib borgan olimlar H.Arziqulov, S.Rizayev va S.Muhamedovlardir⁸. Ular, asosan, KLning statistik tahlil yo‘nalishlari bo‘yicha nazariy va amaliy ishlarni amalga oshirgan.

H.Arziqulov o‘zining “Informatika i pererabotka teksta sredstvami vychislitelnoy texniki” (“Информатика и переработка текста средствами вычислительной техники”) nomli monografiyasida kompyuter yordamida katta hajmdagi matnlarga ishlov berishning avtomatik tizimi asoslarini tahlil qilgan. S.Muhamedov va R.Piotrovskiy hammuallifligida yozilgan “Injenernaya lingvistika i opit sistemno-statisticheskogo issledovaniya uzbekskix tekstov” (“Инженерная лингвистика и опыт системно-статистического исследования узбекских текстов”) nomli kitob, shuningdek, R.G.Piotrovskiy, T.Sadikovlarning ilmiy izlanishlarida lingvistik modellar, modellashtirish va uning umumiy tamoyillari haqida fikr yuritilgan, o‘zbekcha matnlarning kvantitativ modellari keltirilgan⁹. S.Muhamedov va R.Piotrovskiy hammuallifligidagi asarda sun‘iy intellekt yaratish va injener lingvistikasi metodlari bilan o‘zbekcha nutqni avtomatik qayta ishlash jarayonida foydalanish uchun zarur bo‘lgan o‘zbekcha matnlarning leksik-morfologik tuzilishi statistik tahlili natijalari keltirilgan. S.Rizayevning “Kibernetika va tilshunoslik” asarida esa tilshunoslikda

обработка текста (модели и программы) сегментации русского предложения: Автореф. дисс. ...канд. филол.наук. – Москва, 2003.; Большакова Е.И., Клышинский Э.С. и др. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб.пособие. – Москва: МИЭМ. 2011. – 272 с.; Леонтьева Н.Н. Автоматическое понимание текста: системы, модели, ресурсы: учеб.пособие. – Москва: Академия. 2006. – 304 с.

⁷Махмудов М.А., Пиотровская А.А., Садыков Т. Система системного анализа и синтеза тюркской словоформы / Переработка текста методами инженерной лингвистики. – Минск, 1982.; Пиотровский Р.Г. и др. Методы автоматического анализа и синтеза текста. – Минск: Высшая школа, 1985.

⁸Арзикулов Х.А., Пиотровская К.Р. Информатика и переработка текста средствами вычислительной техники (учебное пособие). – Самарканд, 1986.; Rizayev S. O‘zbek tilining lingvostatistik tadqiqi: Filol. fan. dok. ...diss. avtoref. – Toshkent, 2008. – 50 b.; Shu muall. O‘zbek tilshunosligida lingvostatistika muammolari (monografiya). – Toshkent: Fan.2005. – 295 b.; Мухамедов С.А. Статистический анализ лексико-морфологической структуры узбекских газетных текстов: Автореф. дис. ...канд.филол.наук. – Тошкент, 1980. – 25 с.

⁹Мухамедов С.А., Пиотровский Р.Г. Инженерная лингвистика и опыт системно-статистического исследования узбекских текстов. – Т.: Фан, 1986.; Пиотровский Р.Г., Садыков Т. Проблемы моделирования тюркской морфологии в инженерной лингвистики. // В кн.: II научные чтения, посвященные памяти академиков И.А.Батманова, К.К.Юдахина, Б.М.Юнусалиева (тезисы докладов сообщений, 18-19 октября 1983 г.). – Фрунзе: Илим, 1983. – С.44-46.; Садыков Т. Проблемы моделирования тюркской морфологии. – Фрунзе, 1987.

aniq metodlarning qo‘llanilishi, bunda kompyuterdan foydalanish, til va nutq hodisalariga statistik yondashishning sabablari, matnlarni avtomatik qayta ishlash, mashina tarjimasini muammolari, shuningdek, harflarning qo‘llanilish chastotasini aniqlashda kompyuterning o‘rni borasidagi ma’lumotlar keltirilgan¹⁰. M.K.Ayimbetov statistik metodlardan foydalangan holda turkiy tillar – o‘zbek, qozoq va qoraqalpoq tillari morfologiyasini modellashtirish orqali ularning qarindosh tillar ekanligini isbotlab, ushbu tillarning kvantitativ tipologiyasini ishlab chiqqan bo‘lsa, X.F.Isxakova turkiy tillarning formal morfologiyasini tadqiq qilish bo‘yicha samarali izlanish olib borgan¹¹.

XXI asr boshiga kelib o‘zbek va ingliz tillarida bir qator izlanishlar qilgan, o‘zbek KL rivojiga salmoqli ulush qo‘shgan A.Po‘latov boshchiligida o‘zbek tilshunosligiga oid masalalarni kompyuter vositasida hal etishga bag‘ishlangan bitiruv malakaviy ishlari, magistrlik dissertatsiyalari himoya qilindi, o‘nlab ilmiy maqolalar chop etildi, o‘quv qo‘llanmalar yaratildi¹² va kitoblar nashrdan chiqarildi¹³. Ayniqsa, fe’l turkumini formallashtirish, ularning modelini ishlab chiqish va fe’llar asosida lingvistik ta’min yaratishga bag‘ishlangan talay ishlar dunyo yuzini ko‘rdi¹⁴. So‘nggi yillarda M.X.Xakimov rahbarligida “Tarjimon L-MX” sistemasi uchun bajarilgan tadqiqot ishlarida o‘zbek va rus tillarining leksik-sintaktik tahlili berilgan¹⁵. Sh.Xamroyeva, N.Abdurahmonova va

¹⁰ Rizayev S. Kibernetika va tilshunoslik. T., 1976. – B.16.

¹¹ 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. – Т., 1997. – 47 с.; Он же. Квантитативная типология тюркского текста (сборник избранных статей). – Нукус.: Илим. 2012.; Исхакова Х.Ф. Исследования в области формальной морфологии тюркских языков (на материале татарского литературного языка в сопоставлении с турецким и узбекским). Дис. канд. филол.наук. – М., 1972.

¹²Muhamedova S. Kompyuter lingvistikasi (metodik qo‘llanma). – Toshkent, 2007.; Po‘latov A., Muhamedova S. Kompyuter lingvistikasi (o‘quv qo‘llanma). – Toshkent, 2008 – 98 b.

¹³Po‘latov A. Kompyuter lingvistikasi. – Toshkent: Akademyashr, 2011. – 520 b.; Rahimov A. Kompyuter lingvistikasi asoslari. – Toshkent: Akademyashr. 2011. – 160b.; Norov A. Kompyuter lingvistikasi asoslari. – Qarshi, 2017. – 136 b.

¹⁴Пўлатов А.Қ., Алиходжаев Б., Джураева Н. Разработка программы компьютерного анализа и синтеза глаголов узбекского языка // O‘zMU xabarlari. – Toshkent, 2002. №2. – С. 17-19.; Po‘latov A.Q., Mo‘minova T., Po‘latova I.O. Dunyoviy o‘zbek tili (O‘zbek tilida fe’lning shakllari va ularning rus, ingliz tillarida berilishi). – Toshkent: Universitet, 2003. – 404 b.; Muhamedova S. O‘zbek tilidagi harakat fe’llari asosida kompyuter dasturlari uchun lingvistik ta’min yaratish. – Toshkent, 2006. – 80 b.

¹⁵Хакимов М.Х. Расширяемый входной язык математического моделирования естественного языка для много-язычной ситуации машинного перевода // ЎзМУ хабарлари, 2009. № 1. – С.75-80.

D.O‘rinboyeva¹⁶lar izlanishlarida matnlarni tahlil qilish bo‘yicha muayyan fikrlar berilgan, bu jihat lingvistik dasturlarni yaratishda lingvistik modullarni ishlab chiqishning muhim ahamiyatga egaligini belgilab bergan.

Yuqorida tilga olingan tadqiqotchilarning ilmiy izlanishlari natijasi o‘laroq matnlarni ATT qilishga doir yetarlicha nazariy bilim hamda muayyan tajribalar to‘plandi. Ularga asoslangan holda mazkur tadqiqotda dastlabki tashabbus – o‘zbek KLda avtomatik tahrirlovchi dasturni yaratish (Ubuntu Linux operatsion tizimida – *Uztextanalysis*) uchun harakat qildik. Ma’lumki, o‘zbek tilshunosligida ham, adabiyotshunoslik sohasida ham o‘zbek tilidagi matnlarning badiiy tahlili va tahriri inson tomonidan amalga oshirib kelinmoqda. Ammo bugungi kunda insonning matn tahriridagi ishtirokini talab qilmaydigan kompyuter dasturlari yaratilmagan, bu borada chuqur ilmiy izlanishlar amalga oshirilmagan. Shu kunga qadar kompyuter dasturchilari tomonidan chiqarilgan tahrir dasturlari (kirill yozuvidagi matnlarni tahrirlovchi) o‘zining ilmiy-nazariy asosiga ega emasligi va nomukammalligi bois foydalanuvchilar e’tiborini torta olmadi. A.Nurmonov shunday deydi: “O‘tgan yarim asrlik vaqt mobaynida kompyuter lingvistikasi sohasida bir qator ilmiy va amaliy natijalarga erishilgan: tabiiy tilda avtomatik tarjima tizimi yaratildi, matndagi ma’lumotlarning avtomatik qidiruv tizimi ishlab chiqildi, og‘zaki nutqning avtomatik analizi va sintezi tizimi yaratildi, bir qator lingvistik ma’lumotlarni hal etuvchi kompyuter dasturlari ishlab chiqildi, inson va mashina (kompyuter) muloqoti optimallashtirildi, tabiiy tilni qayta ishlash (Natural Language Processing) tizimi shakllantirildi”¹⁷. Ma’lum bo‘ladiki, mazkur tadqiqot o‘zbek kompyuter lingvistikasida avtomatik tahrir va tahlil qiluvchi dasturni yaratish borasida yozilayotgan ilk ilmiy ish hisoblanadi. Bajariladigan tadqiqot Respublikada olib borilayotgan ilmiy ishlar bilan hamohang bo‘lib, o‘zbek tilshunosligida kompyuter texnologiyalarining o‘rni masalalariga chuqur e’tibor qaratiladi.

Matnlarni avtomatik tahlil qiluvchi dasturlar lingvistik modullarining yaratilishi mukammal dasturlar ishlab chiqilishiga zamin yaratadi, bu esa o‘zbek adabiy tilidagi matnlarning savodli tayyorlanishiga xizmat qiladi.

¹⁶Xamroyeva Sh. O‘zbek tili mualliflik korpusini tuzishning lingvistik asoslari: Filol.fan.bo‘yicha falsafa doktori (PhD)...dis. avtoref. – Qarshi, 2018. – 53 b.; Abdurahmonova N. Inglizcha matnlarni o‘zbek tiliga tarjima qilish dasturining lingvistik ta’minoti: Filol.fan.bo‘yicha falsafa doktori (PhD)...dis. avtoref. – Toshkent, 2018. – 47 b.; O‘rinboyeva D. Xalq og‘zaki ijodi: janriylisoniy va lingvostatistik tadqiq muammolari: Filol. fan. bo‘yicha dokt. (DSc) ... diss. avtoref. – Samarqand, 2019. – 74 b.

¹⁷ Rahimov A. Kompyuter lingvistikasi asoslari. – Toshkent: Akademnashr. – B.3.

I BOB. KOGNITIV TADQIQOTLAR PARADIGMASIDA MATN TAHRIRI VA TAHLILI

1.1. Hozirgi tilshunoslikda kognitiv paradigma

Ilm-fan, jumladan, tilshunoslikning rivojlanishida ilmiy paradigmalarning bosqichma-bosqich o'zgarishi kuzatiladi. "Ilmiy bilimlar paradigmasi" termini T.Kun tomonidan muomalaga kiritilgan bo'lib, olim paradigmani "barcha odamlar tan olgan ma'lum vaqt davomida ilmiy jamoatchilikka muammolarni hal qilish modeli va ularning yechimlarini taqdim etadigan ilmiy yutuqlar"¹⁸, deb baholaydi. Tilshunoslik tarixida uchta ilmiy paradigma farqlanadi: qiyosiy-tarixiy, sistem-struktur va kognitiv (antropotsentrik)¹⁹.

Qiyosiy-tarixiy paradigma XI-XIX asrlarda tilni tadqiq qilishda qiyosiy-tarixiy metodning hukmronlik qilishi bilan bog'liq. Sistem-struktur paradigma bilish obyektiga, so'zga diqqatni tortishi bilan xarakterlanadi. Kognitiv paradigma antropotsentrik tamoyilga asoslanadi va "til inson tomonidan yaratilgan bo'lib, uning yaratuvchisi hamda foydalanuvchisi bilan bog'liqsiz ravishda uni tushunib ham, tushuntirib ham bo'lmaydi"²⁰, degan tushunchaga asoslanadi.

Bugungi kunda til aniq maqsadlarga erishmoqchi bo'lgan odamlarning o'zaro muloqot qilish va o'zaro tushunish vositasi bo'lib xizmat qilishda davom etmoqda. Shu sababli zamonaviy til konsepsiyalarini, xususan, kognitiv tilshunoslikni rivojlantirishga alohida e'tibor qaratilmoqda.

Kognitiv ilm XX asrning 60-yillari o'rtalarida Yu.D.Apresyan, A.Vejbitskaya, E.S.Kubryakova, Dj.Lakoff kabi olimlar tomonidan olib borilgan tadqiqotlar natijasida shakllangan²¹. Kognitiv tilshunoslik uzoq

¹⁸ Кун Т. Структура научных революций. – М.: Прогресс, 1977. – С. 11.

¹⁹ Ныгметова Б.Д. Когнитивная парадигма в современной лингвистике. www.rusnauka.com

²⁰ Кравченко А.В. Язык и восприятие: Когнитивные аспекты языковой категоризации. – Иркутск: Изд-во Иркутского ун-та, 1996. – С. 6.

²¹ Апресян, Ю.Д. Исследования по семантике и лексикографии. Т. I: Парадигматика. – Москва: Языки славянских культур, 2009.; Вежбицкая А. Язык. Культура. Познание. Пер. с англ., отв. ред. М. А. Кронгауз, вступ. ст. Е. В. Падучевой. – Москва: Русские словари, 1996. – 416 с.; Она же. Сопоставление культур через посредство лексики и прагматики. – Москва: Языки славянской культуры, 2001. – 272 с.; Кубрякова Е.С. Семантика в когнитивной лингвистике. // Известия РАН. – Сер. лит. и яз. 1999. – Т. 58. – № 5-6. – С. 3-12.; Она же. Начальные этапы становления когнитивизма: лингвистика, психология, когнитивная наука / Е. С. Кубрякова // Вопросы языкознания. 1994. – №4.; Она же. Язык и знание. На пути получения знаний о языке: части речи с когнитивной точки зрения. Роль языка в познании мира. Рос. академия наук. Ин-т языкознания. – Москва: Языки славянской культуры, 2004. – 560 с. – (Язык. Семиотика. Культура); Лакофф Дж. Лингвистические гештальты. В кн.: Новое в зарубежной лингвистике. Москва: Прогресс, 1981. Вып. X. – С. 350-368.

vaqt davomida bir-biri bilan nisbatan aloqasi bo‘lgan yoki umuman aloqasi bo‘lmagan individual tadqiqotlar to‘plamidir²². Ammo ko‘p tadqiqotlar til bo‘yicha amalga oshirilgan ilmlar bog‘liqligi, idrok qilish subyektivi, fikrlash, xatti-harakat va amaliyot bilan bir qatorda tilda tushuncha, tasavvur, tasvir, subyektiv va etnik omillar bilan haqiqiy dunyoni aks ettirishda qarashlar birlashganini ko‘rsatadi. Demak, kognitiv paradigma tilga tubdan yangi yondashuvni talab qiladi. Dunyoning konseptual tasviri (manzarasi)ni shakllantirish mexanizmlari va alohida tushunchalar, shuningdek, ularni tilda obyektivlashtirish kognitiv paradigmaning asosiy tadqiq yo‘nalishlaridan biri hisoblanadi.

Tilni o‘rganishga kognitiv yondashuv bilimlarni (matn yoki nutq shaklida ifoda etilgan) tushunishga erishish zarurati bilan bog‘liq. Bunda tushunish jarayonini murakkablashtiradigan yoki soddalashtiradigan ekstralingvistik omillar hisobga olinadi va tahlil qilinadi.

O‘tgan asrning 70-yillarida kognitologiya doirasida bilimlarni namoyon qiluvchi tillar va ular bilan ishlash tamoyillari (konseptual tushunchalar qoidalari) ishlab chiqildi. Kognitiv bilimlar sohasida tabiiy tilni tushunish uchun kompyuter modellarini yaratish maqsadida ko‘plab tadqiqotlar olib borildi, tilshunoslar bunday tadqiqotlarda kuzatuvchi yoki bevosita ishtirokchi sifatida qatnashdilar. Natijada tilshunoslarning tilga oid materiallarni fikrlash jarayonlari haqidagi ma’lumotlarga mos keltirish borasidagi azaliy ehtiyojlari o‘sha paytda semiotik kategoriyalar bilan fikrlaydigan hisoblash texnikasi (EHM) va informatika mutaxassisleri tomonidan qondirildi.

Shu yillarda sun‘iy intellektda A.Nyuell va G.Saymonlarning “axborotga ramziy ishlov berish”ning (symbolic processing) semiotik paradigmasi ustunlik qildi. Massachusetts texnologiyalar institutining sun‘iy intellekt laboratoriyasi rahbari P.Uinston “hisoblash mashinasini tabiiy tilni tushunishga majburlash – bu aqlni yaratish bilan barobar”ligini ta’kidladi²³. Shundan so‘ng kompyuterli modellashtirish fanga kirib kela boshladi.

Shubhasiz, tabiiy tilni tushunadigan dasturlarning yaratilishi tilning funksional xususiyatlari, xotira nazariyasini ishlab chiqish, bilim tuzilmalarini ko‘rib chiqish, muloqot grammatikasini o‘rganish va ong tabiati falsafiy tadqiqoti haqidagi savollarga javob beradi²⁴.

²² Ныгметова Б.Д. Когнитивная парадигма в современной лингвистике. www.rusnauka.com

²³ Ныгметова Б.Д. Когнитивная парадигма в современной лингвистике. www.rusnauka.com

²⁴ Ныгметова Б.Д. Когнитивная парадигма в современной лингвистике. www.rusnauka.com

R.Shenk sun'iy intellekt, fikrlash va o'zini tushunishga bag'ishlangan maqolasida shunday yozadi: "Sun'iy intellekt tafakkurni tushunishga oid umumiy intilishning bir qismi hisoblanadi. Biz bunday harakatlarga muhim hissa qo'shishga ishonamiz va ular ilm-fanimizning maqsadidir. Biz yozgan dasturlar natija emas, balki tajriba berdi. Bizni qiziqtiradigan narsa intellekt hisoblanadi, artefakt²⁵ (P.S. sun'iy yaratilgan obyekt) emas. Erishilgan natijalardagi yutuqlarimiz, shubhasiz, kundalik hayotning zarur qismiga aylanadigan avtomatik hamkorlarga yo'l ochadi"²⁶. R.Shenk fikriga qo'shimcha ravishda aytish mumkinki, KLda erishilgan natijalardagi yutuqlar, shubhasiz, kelajakda ish-faoliyatni avtomatlashtirishga xizmat qiladi.

Kognitologiya fanining asosiy vazifasi aqliy tasavvur qoidalari va mantiqiy xulosalar qonuniyatlariga tayangan holda tabiiy til tizimini "qayta ishlash"ning talab va haqiqatga mos nazariyasini yaratish hisoblanadi. Shunday g'oyalar zamirida matnlarni qayta ishlashga bo'lgan ehtiyoj oshib bordi va sun'iy intellekt tizimida lingvistik bazalar yuzaga kela boshladi.

O'rganilgan manbalar asosida aytish mumkinki, KL yo'nalishlarining amaliy natijalariga erishish maqsadida yaratilayotgan modellashtirish, formallashtirish va sun'iy intellekt tizimini rivojlantirish jarayonlari kognitiv paradigmalarda ta'sirida voqelanadi.

Informatsion olamning kengayishi va insonlarning qulaylikka hamda tezkorlikka intilishi bois deyarli barcha sohada foydalanuvchiga qulaylik yaratuvchi lingvistik dasturlarga talab yanada oshib bormoqda. Avtomatik tahrir va tahlil kompyuter lingvistikasining asosiy yo'nalishlaridan biri bo'lib, u kompyuterdagi matn muharrirlari (Microsoft Word, Excel, Wordpad, Lexicon) rivojlanishi bilan bog'liqdir. Ayni damda tahrir va tahlil dasturining lingvistik asosi hisoblanmish lingvistik modullarni grafematik, morfologik, sintaktik va semantik bosqichlarda ishlab chiqish kabi masalalar ustuvor ahamiyat kasb etib bormoqda. Kompyuter lingvistikasi fanining asosiy maqsadi ham lingvistik masalalarni hal qilish uchun kompyuter dasturlarini yaratish²⁷ va lingvistik ta'minotni shakllantirishdir.

XX asr oxirida faqat elektron shaklda mavjud bo'lgan matnlar va hujjatlar soni ko'paygan va ma'lumotlarni qayta ishlash bilan bog'liq

²⁵ <https://ru.wikipedia.org/wiki/Артефакт>

²⁶ Schank R.C., Hunter L. The quest to understand thinking // Byte. 1985, apr. – P. 155

²⁷ Новое в зарубежной лингвистике. Вып.24. Компьютерная лингвистика. – Москва: Прогресс, 1989, – С.10.

texnologiyalarni rivojlantirishni talab qilgan. Tabiiy tildagi hujjatlarni qayta ishlash (NLP – “Natural Language Processing”) insonlarga elektron matnlardan foydalanish va hujjatlarni osongina topish hamda qayta ishlashga yordam beradigan yangi texnologiyalar demakdir²⁸. Shunday texnologiyalardan sanalgan matn tahlilining DTini yaratish bo‘yicha rus KLda samarali izlanishlar olib borilgan. Ushbu sohada bir qator nomzodlik va doktorlik dissertatsiyalari yoqlangan. Bu tadqiqotlar, asosan, EHM mutaxassislari tomonidan amalga oshirilgan. Shunday bo‘lsa-da, ular rus tilining nazariy va amaliy jihatlarini tadqiq qilishga munosib hissa qo‘shgan. Xususan, KLda avtomatik tahlilning umumiy masalalari, modellashtirish prinsiplari²⁹, morfologik tahlil³⁰, sintaktik³¹ va semantik tahlil³², tahlil birliklari, tabiiy tildagi matnlarni avtomatik tushunish nazariyasi³³, matnni tahlil qilish bosqichlari, morfologik analiz va sintez, tabiiy tildagi matnni qayta ishlashga bag‘ishlangan bir qator asarlar maydonga keldi. Jumladan, E.I.Bolshakova lingvistik va matematik modellashtirish nazariyasiga to‘xtaladi, avtomatik tahlilning lingvistik obyektlari va birliklarini tadqiq qiladi. N.N.Leonteva matnlarni avtomatik tushunishni amalga oshiruvchi rus va xorijiy tizimlarning yaratilish tajribasini o‘rganadi va asosiy e‘tiborni sifatli ishlanmalarga qaratadi. Ushbu olimlarning xizmati shundaki, ular avtomatik tahrir qiluvchi dasturlar uchun lingvistik ta‘min yaratib berishgan. Buning natijasida kompyuterlardagi ruscha va inglizcha matnlarni tahrirlovchi va lingvistik

²⁸ Баскакова И.Л., Глухов В.П. Практикум по психолингвистике / Учебное пособие. – Москва: Астрель, 2008. – С. 315.

²⁹ Ягунова Е.В. Основы теоретической, вычислительной и экспериментальной лингвистики / Е.В.Ягунова // Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учебное пособие / Е. И. Большакова [и др.]. – Москва: МИЭМ, 2011. – 272 с. – С. 7-89.

³⁰ Клышинский, Э. С. Начальные этапы анализа текста / Э. С. Клышинский // Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учебное пособие / Е. И. Большакова [и др.]. – Москва: МИЭМ, 2011. – 272 с. – С. 106-140.; Марчук, Ю. С. Компьютерная лингвистика / Ю. С. Марчук. – Москва, 2007. – С. 60-70.

³¹ Компьютерный синтаксический анализ: описание моделей и направлений разработок / Г. Д. Карпова [и др.] // Итоги науки и техники. Серия «Вычислительные науки». Т. 6. – Москва, 1991. – 243 с.; Иорданская, Л. Н. Автоматический синтаксический анализ. Т. 2. Межсегментный синтаксический анализ / Л. Н. Иорданская. – Новосибирск, 1967. – 231с.; Мельчук, И. А. Автоматический синтаксический анализ. Том 1. Общие принципы. Внутрисегментный синтаксический анализ. / И. А. Мельчук. – Новосибирск, 1964. – 360 с.

³² Апресян, Ю.Д. Исследования по семантике и лексикографии. Т. I: Парадигматика. – Москва: Языки славянских культур, 2009.; Сокирко А.В. Семантические словари в автоматической обработке текста: По материалам системы ДИАЛИНГ: Дисс. ... канд. тех. наук: – Москва, 2001. – 120 с.

³³ Леонтьева Н.Н. К теории автоматического понимания естественных текстов. – Москва: МГУ, 2000. – 304 с.

tahlil qiluvchi dasturlar ishlab chiqilgan. Bunday dasturlar bugungi kunda barcha kompyuterlarda mavjud.

Bugun o'zbek tilidagi matnlarni tahrir qilish inson mehnati bilan amalga oshirilmoqda, bu esa o'z navbatida matnlarni tahrirlashga ketadigan vaqtning ko'p sarf qilinishi, iqtisodiy harajatlarning oshishi, inson tomonidan yo'l qo'yiladigan xatoliklarning yuzaga chiqishiga olib kelmoqda. Shu bois o'zbekcha matnlarni tahrir va tahlil qiluvchi dasturning lingvistik ta'minotini yaratish muhim ilmiy-amaliy ahamiyat kasb etadi.

1.2. Kompyuter lingvistikasida modul

Ilm-fanning rivojlanishi, yangiliklarning ko'payishi, innovatsion texnologiyalarning taraqqiy etishi natijasida tilga tashqi manba hisobiga yangi terminlar kirib kelmoqda, terminlardan foydalanish ko'lami kun sayin ortmoqda. Xorij kompyuter lingvistikasida faol iste'molda bo'lgan *modul* termini o'zbek kompyuter lingvistikasiga ham kirib, uning terminologiyasini boyitdi.

Dunyoda ko'p foydalanuvchilariga ega "Vikipediya" xalqaro ochiq ensiklopediya saytida "modul" terminining ma'lum sohalar doirasidagi ahamiyati qayd etilgan³⁴. Chunonchi, informatikada: 1) modul – dastur fayli; 2) modul – kodni tashkil qiluvchi obyekt; 3) modul – kompyuter sovutadigan sistema jamlanmasi; 4) MOD musiqiy fayl formati;

matematikada: 1) absolyut balandlik; 2) vektor moduli; 3) avtomorfizm moduli; 4) bir sistemadagi logarifmni boshqa sistemadagi logarifmga o'tkazish koeffitsienti, shuningdek, kattalikning absolyut qiymati va h.k.;

mexanikada: 1) Yung moduli; 2) elastiklik moduli; 3) siljish moduli va h.k.

Boshqa sohalarda: modul – dasturning tugallangan funksional qismi; modulli o'qitish – zamonaviy ta'lim, ya'ni bilim darajasi bo'yicha bosqichma-bosqich o'qitish va h.k.

Kompyuter lingvistikasida *lingvistik modul* termini bugungi kunda muhim ahamiyat kasb etmoqda. Boisi tabiiy tilning kompyuter tiliga o'tkazilishi, ya'ni kompyuter tizimi orqali matnga ishlov berish yo'llarining yaratilishi kuzatilmoqda. Buning uchun chet tillarning lingvistik dasturlari ishlab chiqilgan va bugungi kunda ular takomillashtirilmoqda. Lingvistik modul ana shunday lingvistik

³⁴ <https://ru.wikipedia.org/wiki/Модуль>

dasturlarning *mustaqil tarkibiy qismlari*³⁵, ya'ni dasturiy ta'minotning muayyan lingvistik jarayonini qamragan qismi hisoblanadi. Lingvistik ta'minot, boshqacha aytganda, lingvistik protsessor ushbu qismlar majmuini tashkil etadi. Leksik modulda lug'at qatlami qurshab olinsa, grafematik modulda ramzlar, tinish belgilar, harfiy va boshqa belgilar tahrir qilinadi, orfografik modulda imlo qoidalari, morfologik modulda so'zshakllar analizi (so'zshakldan leksemaga qadar tahlil) va sintezi (leksemaning grammatik shakllanishi tahlili jarayoni), sintaktik modulda supersintaktik birlik – gap yoki so'zlarning o'zaro bog'lanish hodisasi tahlil qilinadi.

LMLar qanchalik aniq bo'lsa, dasturning mukammalligi shunchalik chuqur bo'ladi. Lingvistik dasturlarning LMLarini yaratishda adabiy til me'yorlari asos vazifasini o'taydi.

Kompyuter lingvistikasi taraqqiyoti davomida matnni qayta ishlash dasturlarining algoritmlarida tilda mavjud har bir lingvistik tushunchaga ramz (simvol) berilgan bo'lib, bugungi kunga kelib, bunday ramzlar umumqabul qilingan belgilar sifatida foydalaniladi.

Tahrir va tahlil dasturi bazasida xalqaro iste'molga kirgan lingvistik tushunchalarning maxsus belgilarga ega lingvistik birliklaridan foydalanildi. Quyida mazkur lingvistik birliklarning umumiy simvollari berildi, ish davomida ular xususiylashtiriladi:

1) ot – *N* (noun), ko'plik shaklidagi ot – N_s , turdosh ot – N_{com} , atoqli ot – N_{prop} , otli birikma – NP;

2) sifat – *Adj* (adjective), sifatli birikma – *AdjP*;

3) son so'z turkumi – *Num* (numeral);

4) fe'l – *V* (verb), o'timli fe'l – V_t , o'timsiz fe'l – V_l , fe'lning predikativ (shaxsli shakli – finite form) shakli – V_p , fe'lning nopredikativ shakli (shaxssiz shakli – nonfinite form) – V_{np} , *Vh* – harakat nomi, Vs –sifatdosh, *Vr* – ravishdosh, fe'lli birikma – *VP*, yordamchi fe'llar – *Vaux*;

5) ravish – *Adv* (adverb);

6) olmosh – *Pron* (pronoun);

7) ko'makchi – *RostP* (postposition);

8) bog'lovchi – *Conj* (conjunction);

9) yuklama – *Part* (particle);

10) modal so'zlar – *Mod* (modal);

11) undov so'zlar – *Interj* (interjection);

12) taqlid so'zlar – *Mim*.

³⁵ Демьянков В.З. Когнитивная лингвистика как разновидность интерпретирующего подхода // Вопросы языкознания. – Москва, 1994. – № 4. – С. 22.

Soʻz turkumlarining lingvistik modulini yaratishda amaliy natijaga erishish maqsadida, avvalo, ularga shakl yasovchi affikslar va ularning kombinatsiyasi soʻzlarga biriktirilib, lingvistik bazaga kiritish optimal yoʻl, deb olindi. Zero, soʻz yasovchi affikslar salmoqli miqdorni tashkil etadi. Ishning dastlabki bosqichi tugallangach, kelajakda soʻz yasovchi affikslar va ularning aranjirovkasi ham qamrab olinishi koʻzda tutildi. Ammo bu jihat yasama soʻzlardan voz kechildi, degani emas. *Sanoq* (*son* → *sana* + *q*), *oʻroq* (*oʻr*+*oq*), *oʻquvchi* (*oʻqi*+*v*+*chi*), *anglamoq* (*ong*+*la*) kabi yasama soʻzlar “avtomatik hosila”³⁶ sifatida lingvistik taʼminotga kiritiladi. *Son* ot, *sanoq* sifat, *oʻr* feʼl, *oʻroq* ot, *oʻqi* feʼl, *oʻquvchi* ot, *ong* ot, *angla* feʼl turkumiga oid soʻzlar sifatida olinadi. Bunda soʻzlarga yasovchi affikslar qoʻshilganda oʻzakda roʻy beradigan fonetik hodisalarga alohida eʼtiborning kamayishi va ish hajmi kengayishining oldi olinadi.

Maʼlumki, til oʻzgaruvchan va har bir tilning oʻz lugʼat boyligi mavjud. Ammo hajm qanchalik katta boʻlsa, aniqlik shunchalik mavhumlashadi. Har bir soʻzning qoʻllanish shakllarini model qilib dasturga kiritish mushkul vazifa. Birgina ot turkumiga oid soʻzlarning shakllarini aniqlash tilshunosdan ancha diqqat va mehnatni talab etadi. Maqsadga tezroq erishish uchun har bir turkumning lingvistik moduli yaratildi, yaʼni lingvistik shakllar kategorial tarzda modellashtirildi. Soʻngra oʻsha kategoriyalarga birikish ehtimolidagi affikslar va ularning kombinatsiyalari guruhlandi. Jumladan, ot turkumini modellashtirishda oʻzak soʻzlarga qoʻshiladigan affikslar quyidagicha belgilanadi:

- qarashlilik affiksi: $q_a = -niki$;
- oʻrin-joy oti affiksi: $u_j = -dagi$;
- chegaralash affiksi: $ch_q[3] = \{-gacha, -kacha, -qacha\}$;
- koʻplik affiksi: $Pl_a = -lar$;
- kelishik affikslari (variantlari bilan): $k_a[7] = \{-ning, -ni, -ga, -ka, -qa, -da, -dan\}$;
- egalik affikslari: $e_a[9] = \{-m, -im, -ng, -ing, -lari, -miz, -imiz, -ngiz, -ingiz\}$;
- substantiv shakl yasovchi affiks: $sh_y = -lik$;
- shaxs-son kategoriyasining 1-tur affiksi: $sh_s_1[-man, -san, -miz, -siz; -simiz, -sisiz]$
- affiksli yuklamalar: *-mi, -chi, -gina, -kina, -qina, -dir, -u, -yu, -da, -a, -ya.*

³⁶ “Avtomatik hosila” bu soʻzlarga qoʻshimchalar qoʻshish jarayonida hosila soʻzda roʻy beradigan tovush oʻzgarishlarining lingvistik taʼminotda aynan saqlanishidir.

Keltirilgan affikslarni o‘zakka biriktirish modeliga quyidagi misollarni keltirish mumkin (A=asos, N=hosila so‘z):

1. $N=A \cup q_a$; bolaniki= $bola \cup niki$
2. $N=A \cup u_j$; boladagi= $bola \cup dagi$
3. $N=A \cup ch_a[1]$; bolagacha= $bola \cup gacha$
4. $N=A \cup Pl_a$; bolalar= $bola \cup lar$
5. $N=A \cup k_a[7]$; bolaning= $bola \cup ning$
6. $N=A \cup e_a[6]$; bolam= $bola \cup m$
7. $N=A \cup k_a \cup e_a[6]$; bolalarim= $bola \cup lar \cup im$
8. $N=A \cup k_a[6]$; bolamga= $bola \cup m \cup ga$
9. $N=A \cup Pl_a \cup e_a[6] \cup k_a[6]$; bolalarimga= $bola \cup Pl_a$; $lar \cup e_a[6]$; $m \cup k_a[7]$; ga
10. $N=A \cup e_a[6] \cup u_j$; bolamdagi= $bola \cup m \cup dagi$.

Modellashtirish shu tartibda davom ettiriladi.

Matn tahriri va tahlilini amalga oshiruvchi dasturning lingvistik bazasini yaratishda har bir so‘zning affikslar bilan kelish holati alohida-alohida ko‘rib chiqilsa, katta yutuqlarga erishilgan bo‘lardi. Ammo bunga ko‘p vaqt va jismoniy mehnat talab qilinishi hisobga olinib, ishning optimal varianti o‘ylab chiqildi. Bunda har bir turkum doirasidagi so‘zlar, ya’ni turkum tarkibi umumiy semantikasi bo‘yicha ma’no guruhlariga ajratildi. Shuni nazarda tutib, ot turkumidagi so‘zlar semantikasi va affikslar kombinatsiyasini o‘ziga biriktirishdagi yaqinliklar hisobga olinib, ma’no guruhlariga taqsimlandi. Shuningdek, tadqiqot davomida an’anaviy tilshunoslikda berilgan lingvistik me’yorlar o‘zbek tili so‘z turkumlari modulida leksik-grammatik kodni shakllantirish maqsadida fonologik, morfonologik va orfografik qoidalar algoritmi tuzildi.

1.3. Matn tahriri va tahlilini amalga oshiruvchi dasturiy sistemalar

Matnlarni lingvistik tahrir qilish jarayoniga azaldan asosiy filologik masala sifatida qaralgan bo‘lib, yillar davomida inson tomonidan amalga oshirib kelinmoqda. G‘arbiy Yevropa, rus, xususan, o‘zbek tilshunosligida olib borilgan matn tilshunosligiga doir ko‘pgina ishlarda uning obykti va predmeti, tuzilishi, semantikasi, o‘ziga xos sistemaga ega bo‘lishi kabi matnga aloqador ko‘pgina masalalar ma’lum darajada yoritilgan. O‘zbek tilshunosligida matn nazariyasi haqidagi fikrlarni birinchi bo‘lib G‘.Abdurahmonov ilgari surgan. Mazkur masala haqidagi ba’zi umumiy fikrlar, jumladan, matnning badiiy tahlili borasidagi izlanishlarni A.G‘ulomov, M.Asqarova, A.Mamajonov, B.O‘rinboyev, R.Qo‘ng‘urov,

J.Lapasov, M.Yo‘ldoshev³⁷ning kitoblarida uchratishimiz mumkin. Ushbu tadqiqotlarda matn va uning tiplari, badiiy matn va uning qismlarini bog‘lovchi vositalar, badiiy nutq uslubi va uning ko‘rinishlari, badiiy matnning fonetik-fonologik, leksik-grammatik xususiyatlari hamda badiiy matnni tadqiq etish tamoyillari haqida so‘z yuritilgan. Badiiy asarlardan olingan misollar asosida badiiy asarni lingvopoetik jihatdan tahlil qilish ishlanmalari yaratilgan. O‘zbek tilidagi asarlar va tarjima qilingan badiiy asarlar tilini chuqur tushuna bilish mahorati, asarning badiiy-estetik jihatlari chuqur tadqiq qilingan.

Zamonaviy texnologiya rivojlanib borayotgan ayni vaqtda lingvistik tahrir va tahlil jarayonlari ko‘plab rivojlangan davlatlarda avtomatlashtirilmoqda. Zero, insonning “sun‘iy intellekt” tizimiga oid vositalar bilan muloqoti til tizimi va lisoniy faoliyat tadqiqiga butunlay yangicha yondashuvni talab qiladi³⁸. Maqsadga erishishda nafaqat filologlar, balki informatika mutaxassislari va dasturchilar ham muhim o‘ringa ega. Filolog tabiiy til qonuniyatlarini lingvistik ta‘minotda jamlab beradi, dasturchi ularning algoritmini tuzadi, yagona ishchi dastur holatiga keltiradi. Bir jumla bilan aytganda, ushbu jarayon ko‘p mashaqqat, sinchkovlik va vaqtni talab qiladi.

Kompyuter dasturi ikkita ta‘minot asosida ishlaydi: 1) texnik ta‘minot; 2) lingvistik ta‘minot (LT). Bu taminotlar o‘zaro bir-birini taqazo qilgani uchun kompyuter va tilshunoslik fanlari hamkorligi natijasida KL yuzaga keldi. LT lingvistik daturlarning lisoniy xotirasi, deyish mumkin. Sh.Safarov fikricha, lisoniy xotira nafaqat til birliklari va kategoriyalari haqidagi axborotni saqlovchi psixik tuzilma, balki nutqiy faoliyat uchun zarur bo‘lgan ushbu turdagi axborotni tezda topish, qo‘llash imkonini yaratuvchi manbadir³⁹. Lisoniy xotira negizini “ichki leksikon” ya‘ni xotiradagi lug‘at zaxirasi tashkil etadi. E.S.Kubryakovaning ta‘biricha, “ichki leksikon ma‘lum sharoitda zarur bo‘lgan birlikni izlab topish imkonini ta‘minlovchi majmuaviy “ombor” bo‘lishdan ko‘ra, ko‘proq harakatdagi tizim bo‘lib, bu tizimda har bir birlikning faollashuv imkoniyatlari, ular qo‘llanilishining pragmatik, semantik va formal

³⁷ G‘ulomov A., Asqarova M. Hozirgi o‘zbek adabiy tili (Sintaksis). – Toshkent, 1987.; Mamajonov A. Tekst lingvistikasi (pedagogika institutlarning o‘zbek filologiyasi fakultetlari studentlari uchun maxsus kurs). – Toshkent, TDPI. 1989.; O‘rinboyev B., Qo‘ng‘urov R., J.Lapasov. Badiiy tekstning lingvistik tahlili. – Toshkent: O‘qituvchi, 1990.; Lapasov J. Badiiy matn va lisoniy tahlil. – Toshkent: O‘qituvchi, 1995.; Yo‘ldoshev M. Badiiy matn va uning lingvopoetik tahlili asoslari. – Toshkent: Fan, 2007.

³⁸ Safarov Sh. Kognitiv tilshunoslik. – Jizzax: Sangzor, 2006. – B. 7.

³⁹ Safarov Sh. Kognitiv tilshunoslik. – Jizzax: Sangzor, 2006. – B. 25.

ko'rsatkichlari qayd qilingan"⁴⁰. Shu bois dasturiy ta'minotning lisoniy xotirasi – lingvistik ta'minotni yaratish, lingvistik modullarni mukammal ishlab chiqish muhim vazifa hisoblanadi.

KL an'anaviy tilshunoslikka tayanib ish ko'radi. Uni amaliy jihatdan boyitadi. Ba'zi hollarda an'anaviy tilshunoslikni inkor etadi (ko'p ma'noli, omonim so'zlarni inkor qiladi). An'anaviy tilshunoslik tabiiy til qonuniyatlari asosida ishlaydi. KL esa sun'iy til qonuniyatlariga bo'ysunadi. Sun'iy til programmalash tilidir (*BASIC, Delta, C++* va h.k.). Tilshunoslik tarixida 1887-yili butun dunyo xalqlarini birlashtirishi va hech bir davlat manfaatini ko'zlamagan holda yagona muloqot vositasi sifatida xizmat qilishi maqsadida birinchi bo'lib vrach L.Zamengof tomonidan esperanto tili yaratilgan⁴¹. KLda esperanto tilining ahamiyati shundaki, aniq lingvistik me'yorlarga asoslangan va ixcham grammatikaga ega ushbu til tabiiy tilni optimallashtirish va formallashtirishda qulay lingvistik dastur vazifasini o'taydi.

Matnlarga ishlov berishning avtomatik tahrir yo'nalishi XX asrning 60-yillariga kelib shakllandi. U matn muharrir dasturlari bilan birgalikda yangicha imkoniyatlar asosida rivojlanmoqda. Oddiy muharrirlardan farqi shundaki, unda tahrir avtomatik tarzda qisqa vaqt ichida katta hajmli matnlar tez tekshirilib, xatolarni samarali to'g'rilash imkoniyatlari bo'ladi. Avtomatik tahrirning pragmatik jihati shundaki, u kelajakda taraqqiy etsa, yetarli ma'lumotlar bazasi yaratilsa va maxsus dasturlar ishlab chiqilsa, muharrir kasbi uchun ehtiyoj qolmaydi. Bu esa matn bilan ishlashni osonlashtiradi.⁴²

Ingliz matn muharriri sistemasi tarixi. Bugungi kunda keng tarqalgan ingliz tilidagi matnlarni tahrir va tahlil qilish dasturlarining yaratilish tarixini amerikalik kompyuter texnologiyalari olimi Lester Donald Ernest boshlab bergan. Olim 1959-yildan o'z faoliyatini matnlarni qayta ishlash sohasiga qaratdi. Shundan so'ng u qo'l yozuvini o'qishga mo'ljallangan dastlabki kompyuter tizimi va imloni tekshiruvchi birinchi dasturni ishlab chiqdi⁴³. 1961-yilda tarmoqlanuvchi texnologiyani o'rganishga rahbar bo'lgan Les Ernest 10000 ta maqbul so'zlar ro'yxatiga asoslangan birinchi imlo tekshiruvini tarmoqda amalga oshirishga harakat qildi⁴⁴. 1971-yil

⁴⁰ Кубрякова Е.С. Память и ее роль в исследовании речевой деятельности // Текст в коммуникации. –Москва: Языкознания, 1991. – С.18.

⁴¹Abjalova M. Esperanto tilini bilasizmi? – Toshkent: Mumtoz so'z, 2010. – 64 b. – lug'at: – 8 b.

⁴²Rahimov A. Kompyuter lingvistikasi asoslari. – Toshkent, 2011. – B. 98.

⁴³Earnest, Les. The First Three Spelling Checkers. Stanford University. <http://web.stanford.edu/~learnest/les/>

⁴⁴ https://en.wikipedia.org/wiki/Spell_checker

fevral oyida L.Ernest rahbarligida Stenford universitetining sun'iy intellekt laboratoriyasida aspirant Ralf Gorin (DEC PDP-10 uchun SPELL (imlo tekshirish)) ingliz tilidagi matnlar uchun ilk imlo tekshiruvchi amaliy dasturni yaratadi⁴⁵. Gorin SPELLning tezroq ishlashi uchun dasturni assembler dasturlash tilida yozgan; Bu Gorinning matndagi bir yoki bir nechta qo'shni harflar bilan farqlanuvchi so'zlarni dastur ro'yxatidan ularning to'g'ri variantini qidirish vazifasini bajaruvchi dastlabki orfo muharrir dasturi sanaladi. Gorin SPELL dasturiy ta'minotini SAIL (Stenford sun'iy intellekt laboratoriyasi)ning ko'plab dasturlari kabi ommaviy tarzda taqdim etdi, natijada shaxsiy kompyuterlarning keng tarqalishidan o'n yil oldin yangi ARPAnet sistemasi orqali dastur tezda butun dunyoga tarqaldi⁴⁶. SPELL va uning algoritmlari hamda ma'lumotlar tuzilmasi Unix ispell dasturiga turtki berdi. 1970-yillarga kelib imlo tekshirish mahsulotlari meynfreymlarda keng tarqaldi. Jorjtaun universitetining olti tilshunosdan iborat guruhi tomonidan IBM korporatsiyasi uchun birinchi imlo tekshirish sistemasi yaratildi⁴⁷.

Shaxsiy kompyuterlar uchun dastlabki imlo tekshirgichlari 1980-yilda paydo bo'ldi. Jumladan, 1981-yil yanvarida reklama matnlarini chop etish maqsadida 1980-yilning oxirlarida Commodore sistemasi uchun "WordCheck" (so'z tekshirish) ishlab chiqildi⁴⁸. 1981-yilda Genri Kuchera VAX Digital Equipment Corp texnologiyasi uchun dastur ixtiro qildi⁴⁹. Mariya Mariani va Random House⁵⁰ kabi ishlab chiquvchilar birinchi navbatda shaxsiy kompyuter uchun tez rivojlanayotgan dasturiy ta'minot bozoriga, shuningdek, Apple Macintosh, VAX va Unix uchun OEM paketlarini ishlab chiqishdi. Shaxsiy kompyuterda ushbu imloni tekshirish mahsuloti alohida dastur bo'lib, yetarli xotiraga ega bo'lgan. Biroq yakka tartibdagi paketlar bozori qisqa muddatli bo'lganligi sababli 1980-yillarning o'rtalariga kelib, mutaxassislar WordStar va WordPerfect kabi orfografik tekshirish imkoniyati mavjud bo'lgan matnlarni qayta ishlash paketlarini ishlab chiqardilar va litsenziyaladilar. Avtomatik imlo tekshirish imkoniyati ingliz tilidan Yevropa davlatlari tillariga qadar kengaytirildi. Garchi flektiv tillarni tahrir qiluvchi dasturning LT va

⁴⁵Peterson, James. Computer Programs for Detecting and Correcting Spelling Errors. 1980. https://en.wikipedia.org/wiki/Spell_checker

⁴⁶Earnest, Les. Visible Legacies for Y3K. <http://web.stanford.edu/~learnest/les/>

⁴⁷<https://web.archive.org/web/20090205140452/http://cled.georgetown.edu/faculty/>

⁴⁸Harvey, Charlotte Bruce. Teaching Computers to Spell (obituary for Henry Kučera). Brown Alumni Magazine. 2010. – P. 79.

⁴⁹Advertisement. Micro Computer Industries, Ltd // Compute! Magazine, Issue 8, Vol. 3, № 1. 1981. – P. 119.

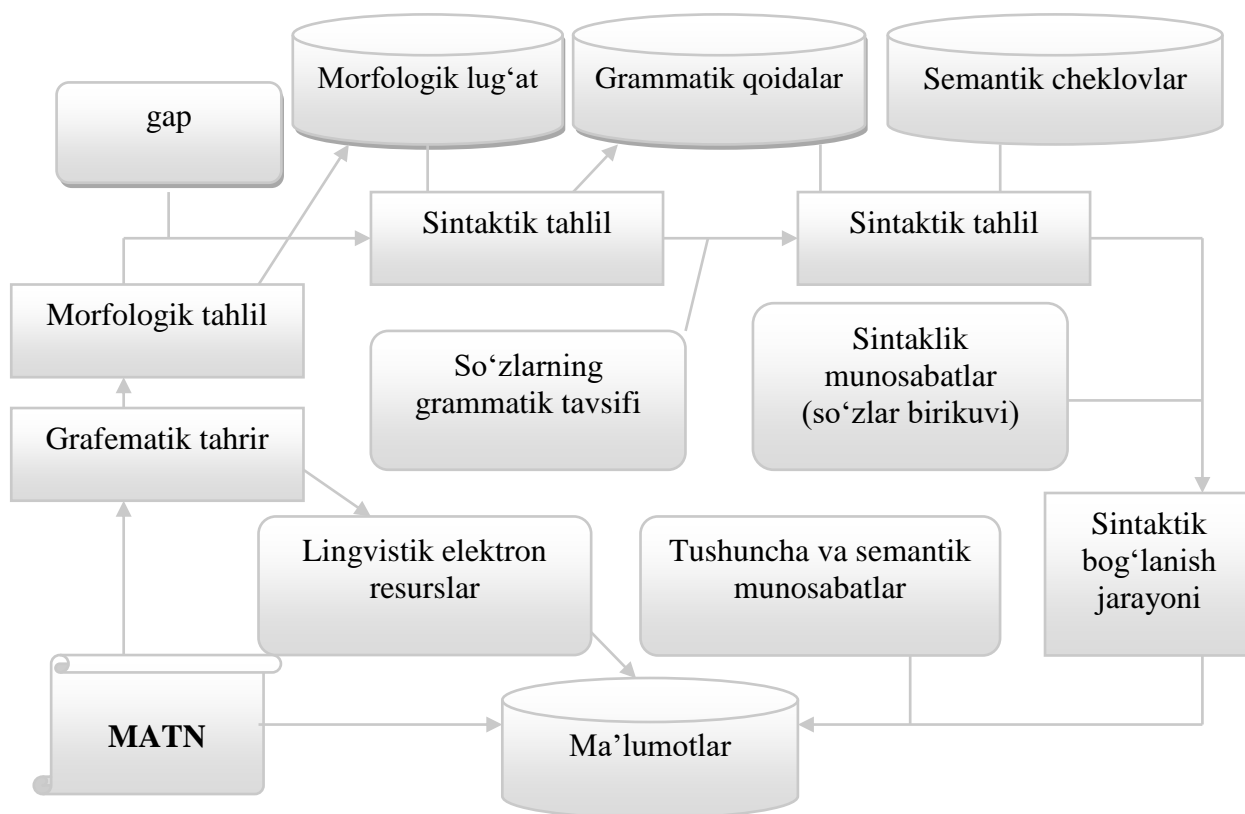
⁵⁰Advertisement. The Spelling Bee Is Over // PC Magazine. 1982. – P. 165.

algoritmi agglyutinativ tabiatli tillar grammtikasiga mos kelmasa-da, keyinchalik bir qancha Osiyo tillari doirasida shunday sistemalar ishlab chiqildi. Jumladan, venger va fin tillaridagi matnlar imlosini tekshirish tizimini yaratish uchun lingvistik ta'minotning morfologiyasini qayta ishlashga to'g'ri kelgan. Bir qancha davlatlar tomonidan milliy tillarga mansub imloni tekshirish sistemasiga sarmoya kiritmagan bo'lsa-da, WordPerfect kabi kompaniyalar o'z dasturlarini global marketing strategiyasi doirasida iloji boricha ko'plab milliy bozorlarga joylashtirishga harakat qilishgan.

Bugungi kunda Firefox 2.0 kabi veb-brauzerlar ko'plab elektron saytlarga, bloglarga va ijtimoiy tarmoqlarga foydalanuvchi tomonidan yoziladigan har qanday hajmdagi matnlarning imlosini tekshirish, Wikitextni tahrirlash imkoniyatiga ega. Shuningdek, Google Chrome, Konqueror i Opera veb-brauzerlari, KMail pochta mijozlari va Pidgin lahzali xabarlashuvlari GNU Aspell va hozirda Hunspell sistemasidan foydalanib, avtomatik imlo tekshirishni taklif qilmoqda. Mac OS X (Makintosh operatsion) sistemasi hozirda boshqa barcha sistemalar uchun avtomatik orfografik tekshirishga ega bo'lib, bu xizmatni yanada takomillashtirish yo'lidan bormoqda. Ba'zi imlo tekshirish mahsulotlari muayyan sohalarga mo'ljallangan⁵¹. Asosiy maqsad soha doirasida yaratilgan LTdagi terminlar, etimologik lug'atlar, qisqartma so'zlar, maxsus belgilar asosida kompyuterga kiritilayotgan matnlarni sifatli tahrir va tahlil qilishdan iborat.

Avtomatik tahrir va tahlil sistemasi (ATTS) matni tahrir va tahlil qilishga mo'ljallangan dastur ustida olib boriladigan ishlar tartibidir.

⁵¹ «Медицинская проверка орфографии для Firefox и Thunderbird».e-MedTools. 2017.; Quathamer, доктор Тобиас «Немецкий медицинский словарь слов» / Доктор Тобиас Куатхамер. 2016.



1-chizma. Avtomatik lingvistik tahlil jarayoni sxemasi.

ATTSni yaratishda qilinadigan ishlarni umumiy tarzda quyidagicha guruhlash mumkin:

1. Tilshunoslik me'yorlari jamlangan lingvistik modullarni yaratish.
2. Modullarda berilgan modellar yordamida dastur algoritmini tuzish.
3. Algoritmdan foydalangan holda dasturni yaratish.

Bugungi kunda matnlarni tahrir qiluvchi yoxud uning tahlilini ham amalga oshiruvchi avtomatik lingvotizim va texnologiyalar yaratilgan bo'lib, ularning ishlash prinsiplari yanada takomillashtirilmoqda. Tubanda shunday tizim va texnologiyalar borasida so'z yuritildi.

Imloni tekshirish tizimi (spell-checker⁵² ingl. spell checker) – kompyuter dasturi bo'lib, kiritilgan matnning orfografik tahririni amalga oshiradi. Aniqlangan imlo xatolari maxsus belgilanadi, ya'ni xato yozilgan leksema tagiga chiziladi. Ko'p hollarda matn teruvchiga imloviy xatolarga ishora qilishdan tashqari dastur maxsus eslatmasi sifatida so'zning to'g'ri yozilish variantlarini ham taklif qiladi. Shuningdek, matnga qanday tuzatish kiritish mumkinligiga izohlar ham beriladi.

⁵² Русский орфографический словарь: около 200 000 слов / РАН. Ин-т рус. яз. им. В. В. Виноградова / Под ред. В. В. Лопатина, О. Е. Ивановой. – Изд. 4-е, испр. и доп. – Москва: АСТ-Пресс Книга, 2012. – С. 709. – (Фундаментал. словари рус. яз.). – ISBN 978-5-462-01272-3.

Matnni avtomatik tahrir qilishga bo‘lgan ehtiyojning o‘sib borishi natijasida imloni avtomatik tekshirish jarayoni shaxsiy EHM dasturiy ta‘minotidan veb-brauzerlargacha ko‘chirilgan bo‘lib, Internet tizimida ishlaydi⁵³. Shuningdek, *orfografik tekshiruv (Speller)* sistemasi ham matnni orfografik jihatdan to‘liq tekshiruvchi modulga ega bo‘lib, uning qulayligi shundaki, keltirilgan ko‘rsatmalar orqali matnni joylashtiruvchi dasturning lingvistik ta‘minotiga yangi so‘zlar, so‘zshakllarni kiritib avtomatik lug‘atni shu zahoti boyitishi mumkin.

Morfologik analizator (Morfoanalizator) – bu so‘zlar va so‘zshakllarini lug‘atdagi shakli (leksikon) bilan taqqoslash, ularning asosi (lemmasi) va so‘zlarning grammatik shakllari (formant / affiks / affiksal morfemalari)ni aniqlash hamda ularga tavsif berish bilan shug‘ullanadigan algoritmlar to‘plamidir, ya’ni dasturiy ta‘minot. Masalan:

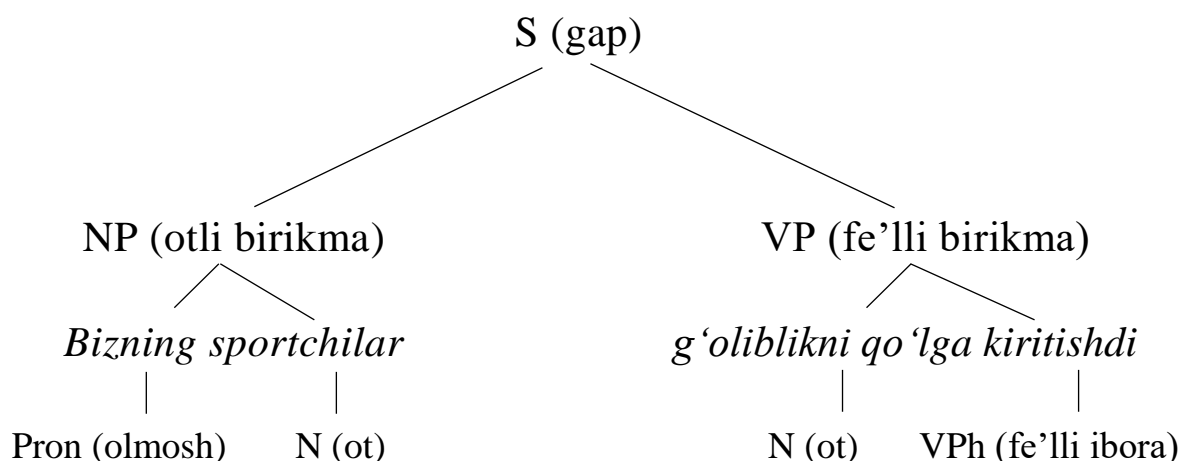
<kitobning> [ot, turdosh ot, ot turi: narsa nomi, birlik, qaratqich kelishigi]

<qizil> [sifat, sifat turi: rang, oddiy daraja, birlik, bosh kelishik]

Sintaktik tahlil tizimi (Parser) (ingl. *parser; parse* – tahlil qilish) yoxud sintaktik analizator – dastur qismi yoki faqat sintaktik tahlilni amalga oshirish uchun yaratilgan maxsus dastur⁵⁴. Kirish ma’lumotlarini (matn) muayyan formatga keltirish orqali tahlil qiladi. Umuman olganda, parser – bu C, C ++, C #, Delphi, PHP, Python kabi tillarda yozilgan va jumlaning tahlil qilish imkoniyatiga ega bo‘lgan dasturiy ta‘minot. Jumla tahlili sol_SyntaxAnalysis funksiyasi yordamida amalga oshiriladi. Tahlilda jumla elementlarining sintaktik munosabatlari aniq belgilab qo‘yilgan sintaktik daraxt, ya’ni tobelik daraxtiga asoslaniladi. Masalan: *Bizning sportchilar g‘oliblikni qo‘lga kiritishdi* jumlasini quyidagicha tobeli qismlarga ajratilib, turkumligi bo‘yicha tahlil qilinadi:

⁵³ Сокирко А. В. Морфологические модули на сайте www.aot.ru (Электронный ресурс) / А. В. Сокирко. Режим доступа: <http://www.aot.ru/docs/sokirko/Dialog2004.htm>.; Russian Link Grammar Parser [Электронный ресурс] / С. Протасов. Режим доступа: <http://sz.ru/parser/>.; Русский морфологический словарь Дмитрия Григорьева [Электронный ресурс]. Режим доступа: <http://www.aot.ru/download/LexGroup.rar>.; Парсер Mystem [Электронный ресурс] / Yandex. <http://company.yandex.ru/teehnology/products/mystem/mystem.xml>!

⁵⁴ <https://ru.wikipedia.org/wiki/Parsing>



Bevosita ishtirokchilar (BI) metodiga ko‘ra tayanch nuqta konstruksiya (qurilma) deb ataladi. Konstruksiya tarkibidan bevosita ishtirokchilar (BeI) va bilvosita ishtirokchilar (BiI) ajratiladi.

Parserda quyidagilar asosiy o‘rinda turadi: tasniflangan leksemalar ketma-ketligi, ierarxik tuzilmalar, ma’lumotlar jadvali, tezauruslar va h.k.

Morfoanalizator va parserlar asosida lemmatizatsiya hamda stemming texnologiyalari mavjud bo‘lib, asosan, so‘zshaklini asos va qo‘shimchalarga ajratib tahlil qilish jarayoni hisoblanadi.

Lemmatizatsiya (lemmatization) – so‘zshakllarni uning lug‘atdagi oddiy shakli – lemmaga keltirish jarayoni⁵⁵, ya’ni so‘zshakl lemma deb atalmish ta’minotdagi so‘zning asosiy yoki lug‘at shakliga qaytariladi. Bu jarayon ayrim rus va xorij manbalarida normallashtirish deyiladi⁵⁶.

O‘zbek kompyuter lingvistikasining lingvistik tahlil yo‘nalishida esa otlar uchun bosh kelishik, birlik son (*kitoblarimni* → *kitob*), fe’llar uchun aniq nisbat, bo‘lishli, fe’lning II shaxs buyruq-istak mayli (*keltirilmagan* → *kel*) va sifatlar uchun uning oddiy darajasi (*oppoq* → *oq*, *pastroq* → *past*) oddiy shakllar hisoblanadi.

Flektiv tillarda lemmatizatsiya qiyin jarayon sanaladi, chunki flektiv tillarda asos va affiks chegarasi yo‘qoladi. Shu bois muayyan so‘zshaklning lemmasi (asosi)ni topish bir muncha qiyinlashadi. Masalan, *vodit* – *voju*, *pit* – *pyu* – *pyosh* – *pey*; *man* – *men*, *tooth* – *teeth* kabi.

Stemming (stemming) – kiritilgan so‘zning asosi (o‘zak)ni topish jarayoni⁵⁷, ya’ni affiksial morfemalarni kesib chiqish texnologiyasi. Bunda topilgan so‘z asosi morfologiyada qabul qilingan so‘z o‘zagiga mos kelishi talab qilinmaydi. Masalan, ingliz tilidagi *flies* (pashshalar) so‘zshaklidagi

⁵⁵ <https://ru.wikipedia.org/wiki/Lemmatization>

⁵⁶ Muller, Thomas; Cotterell, Ryan; Fraser, Alexander; Schütze, Hinrich. Joint Lemmatization and Morphological Tagging with LEMMING (Elektron resurs).

⁵⁷ <https://ru.wikipedia.org/wiki/Лемматизация>

-es ko‘plik qo‘shimchasi kesib tahlil qilinadi, xolos. Uning asosi *fly* ekanligi tahlil qilinmaydi. Shu bois dastur ta‘minotiga “tayyor qolipli so‘zlar” asos sifatida alohida kategoriya qilib kiritilish ehtimoli yuqori bo‘ladi. Masalan, *huquq, tadqiq* kabi leksemalarga egalik affikslari qo‘shilganda leksema so‘ngidagi *q* undoshi o‘zgarishga uchramagani bois bunday istisnoli leksemalar “tayyor qolipli so‘zlar”, ya‘ni “Avtomatik hosila” kategoriyasiga *huquqi, tadqiqi* tarzida kiritiladi va ular stemmingda asos, deb qabul qilinadi.

Kompyuter lingvistikasida so‘zning o‘zagini topish vazifasi uzoq yillar davom etayotgan muammo hisoblanadi.

Stemming, asosan, foydalanuvchining so‘rovi bo‘yicha izlanishni kengaytirish maqsadida qidiruv tizimlari uchun qo‘llaniladi, matnni me‘yorlashtirish jarayonining qismi hisoblanadi. So‘z asosini topishning muayyan usuli stemming algoritmi, uning amalga oshirilishi, ya‘ni dasturning o‘zi stemmer deyiladi.

Ma‘lumki, matnda bir leksemaning turli so‘zshakllari qo‘llanilishi mumkin. Masalan, ingliz tilidagi *organize, organizes, organizing*, o‘zbek tilidagi *kelyapti, kelgan, kelmoqdalar* va h.k. Shuningdek, ifoda va mazmun planiga ko‘ra yaqin so‘zlar oilalari ham mavjud. Masalan, *democracy* (demokratiya) – *democratic* (demokratik) – *democratization* (demokratlashtirish). Lemmatizatsiya va stemming yaqin ifodaga ega so‘zlarga morfemalarning birika olish imkoniyatini ularning asosini topish orqali aniqlaydi. Stemming va lemmatizatsiyaning maqsadi so‘zshakllarini umumiy asos shaklga keltirishdir. Masalan, ingliz tilidagi *cars, car’s, cars’* so‘zshakllarini *car* asos (lemma, stem)ga keltirish kabi. Jarayon bir xil ko‘rinsa-da, stemming lemmatizatsiyadan farq qiladi.

Stemming, odatda, taxminiy jarayon deyiladi. Zero, stemmingda so‘zshakllar oxiridan boshlab lingvistik ta‘minotga kiritilgan asos shaklga qadar kesib kelinadi. Ko‘p hollarda bu holat o‘zini oqlagan.

Lemmatizatsiya so‘zlar leksikonidan foydalanib, ularning morfologik tahlilini amalga oshiruvchi aniq jarayon bo‘lib, bunda faqat fleksiyaga uchragan affikslar o‘chiriladi va lemma deb atalmish ta‘minotdagi so‘zning asosiy yoki lug‘at shakliga qaytariladi.

Yuqorida aytilgan farqni faktik misolda ko‘ramiz: ingliz tilidagi *saw* leksemasi stemmingda *s* harfiga qadar kesib, tekshiriladi, lemmatizatsiya esa uni *see* (fe‘l yoki otga mansubligiga qarab) fe‘lining o‘tgan zamon shakliga qaytaradi. Ko‘rinadiki, stemming bir asosdan yuzaga kelgan so‘zshakllardagi belgilar bilan ishlaydi, lemmatizatsiya esa bir lemmaning

flektiv (affiks qo‘shilishi natijasida o‘zgarishga uchragan) shaklini e’tiborga oladi.

Tokenizatsiya – elektron matndan birliklarni, ya’ni kiritilayotgan belgilar, ramzlar va so‘zlarni ajratib chiqish jarayoni⁵⁸. Tokenizatsiya ingl. *tokenizing* so‘zidan olingan bo‘lib, injiner (kompyuter texnologiyalari mutaxassisi)lar tomonidan qo‘llanishga kiritilgan termin hisoblanadi va tilshunoslikda grafematik tahlil birikmasi bilan tushuntiriladi. Mazkur texnologiya grafematik tahrir bosqichida muhim o‘rin tutadi (qarang: II bob, 2.2-paragraf).

Lingvistik tahlil dasturlash tillarining manba kodi tarjimalari va kompilyator (o‘zgalar tadqiqot natijalarining mustaqil yoki qo‘shimcha ishlovlarsiz berilishi)larida, shuningdek, tabiiy tillarning turli parser – dasturlarning LTlarini yaratishda qo‘llaniladi.

Quyidagi jadvalda avtomatik tahrir va tahlil modullar, ularning vazifalarini amalga oshiruvchi texnologiyalar va dasturiy ta’minotlarning modulga mos ravishdagi nomi yaxlitlandi:

Avtomatik tahrir va tahlil modullari	Texnologiya nomi	Dasturiy ta’minot nomi	
Orfografik tahrir	Spelling	Spell checker	
Grafematik tahrir (tahlil)	Tokenizatsiya/ segmentlash	Tokenizator	
Morfologik tahlil	Stemming	Stemmer	morfoanalizator
	Lemmatizatsiya		
Sintaktik tahlil	parsing	Parser	

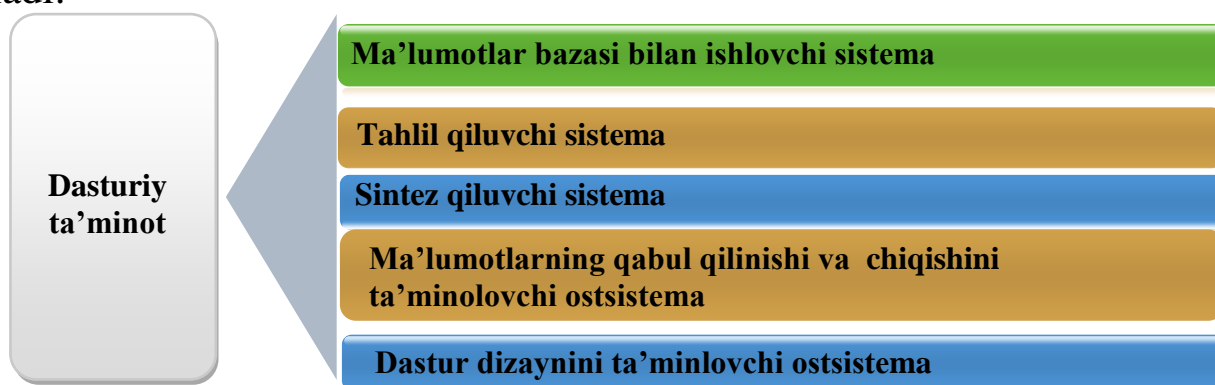
Sistema yoxud texnologiyaning qanday nomlanishidan qat’iy nazar, foydalanuvchi uchun matnni qayta ishlovchi mahsulotning yaratilgani va uning sifat darajasi muhimdir. Chunki matn ustida ishlash, ya’ni rasmiy hujjatlarni yuritish barcha turdagi faoliyatlarning asosi hisoblanadi. Hujjatlar esa o‘z navbatida savodli yozilishi joiz.

O‘zbek tilidagi matnlarni avtomatik tarzda tahrir qilish dasturini yuzaga chiqarishda, albatta, tilshunoslar tomonidan yaratiladigan lingvistik ma’lumotlar bazasi muhim ahamiyatga ega. Bu borada ko‘p bosqichli o‘zbek tili tahrir va tahlil dasturining LT ustida ish olib borilmoqda. Matnlarni avtomatik tahrir qiluvchi dasturlar LMLarining shakllantilishi mukammal dasturlar ishlab chiqilishiga zamin yaratadi, bu esa o‘zbek adabiy tilidagi matnlarning savodli yozilishi va dunyo tillari qatoridan o‘rin olishiga xizmat qiladi. EHM xotirasiga kiritilishi lozim bo‘lgan u yoki bu

⁵⁸ <https://research-journal.org/languages/k-voprosu-o-tokenizacii-teksta/>. Гречачин В.А. К вопросу о токенизации текста [Электронный ресурс]

til (nutq) birliklari, shuningdek, soʻzlararo bogʻlanishning grammatik qoidalari kibernetik apparat uchun tuzilgan dasturda oʻzining qatʼiy matematik tavsifiga ega boʻlgandagina inson tomonidan topshirilgan vazifalar – operatsiyalarning hammasini “aqli” mashinalar aniq bajarishi mumkin⁵⁹.

Dasturning lingvistik taʼminotini yaratuvchi mutaxassisdan DTning mukammal yaratilishi talab qilinadi. Chunki DT tahrir va tahlil dasturining ishlashini taʼminlovchi va uning sifatini namoyon etuvchi bosh sistema hisoblanadi. Bosh sistema, oʻz navbatida, quyidagi sistemalarni oʻz ichiga oladi:



2-chizma. Dasturiy taʼminotni tashkil etuvchi ostsistemalar.

1. Maʼlumotlar bazasi bilan ishlovchi sistema. Bu tizim maʼlumotlar bazasiga kiritilgan maʼlumot bilan ishlashni taʼminlaydi.

2. Tahlil qiluvchi sistema maʼlumotlarni tahlil qiladi.

3. Sintez qiluvchi sistema tahlil qilingan maʼlumotlarning tartiblanishini taʼminlaydi.

4. Maʼlumotlarni qabul qilish va chiqarishni taʼminlovchi sistema dastur uchun kerakli maʼlumotlarning qabul qilinishi va zarur hollarda chiqarilishini boshqaradi.

Maʼlumki, kompyuter kashf etilmasdan ilgari matnlar yozuv mashinkalarida terilgan, bu jarayonda imloviy, punktuatsion, stilistik va grammatik xatolarga yoʻl qoʻyilsa, matnni boshdan oxirigacha qayta yozib chiqishga toʻgʻri kelgan, bu esa foydalanuvchiga noqulaylik tugʻdirgan, vaqtning va ish kuchining ortiqcha sarflanishiga bois boʻlgan. Matnning kompyuter orqali tahlil qilinishi natijasida birmuncha qulayliklar paydo boʻldi. Kompyuterda foydalanuvchi dialogli rejimda ishlaydi va oʻz oʻrnida xatolarni toʻgʻrilash imkoniyatiga ega boʻladi. Hatto matn bilan ishlash jarayonida matnni shakl jihatdan ham tahlil qilish imkoniyati

⁵⁹ Rizayev S. Kibernetika va tilshunoslik. – Toshkent: Oʻzbekiston, 1976. – B.13.

mavjud. Shuni ta'kidlash kerakki, avtomatik tahlillash sistemasi faqat matn muharrirlari dasturlari bazasiga qo'shilgan, masalan, rus, ingliz, fransuz, nemis va boshqa tillarda amal qiladi. Bazada LTi mavjud bo'lmagan tilda kompyuterga matn kiritilsa dastur uni avto-lingvistik tahlil qilolmaydi (aslicha yozilgan baynalmilal so'zlar esa bundan mustasno). Masalan, *chellenj, xesh-teg, karantin, komyuter, internet, klaviatura, modem, telefon, faks* kabi. Avtomatik tahlillash jarayonida Microsoft Word (MS WORD) ta'minotida mavjud bo'lmagan so'z uchrasa, uning ostiga to'liqlik qizil chiziq chiziladi. Mazkur holatda inson – foydalanuvchining ishtiroki zarur, ya'ni foydalanuvchi so'zning imlosi to'g'ri ekanligiga amin bo'lsa, uning avtotahlilda to'g'ri deb tekshirilishi uchun MS WORD lug'atiga "*lug'atga kiritish*" (kontekst menyudagi "*dobavit v slovar*" buyrug'i orqali) mumkin. Matn kiritish davomida so'z yoki so'zlar birikmasi ostiga yashil to'liqlik chiziq chiziladi, bu tinish belgining noto'g'ri qo'llanishi yoxud so'zlarning qusurli grammatik bog'lanishi bilan bog'liq. Bunday holatda foydalanuvchi xatoni o'zi to'g'rilab qo'yishi kerak yoki ostiga chizilgan so'zshaklda sichqonchani o'ng tomoni bosilsa kontekst menyusu ochiladi, unda mazkur so'zshaklning to'g'ri varianti taklif etiladi yoxud to'g'rilash uchun lingvistik ko'rsatma beriladi, so'ng xatoni to'g'rilash kerak bo'ladi. Shuningdek, nuqtadan keyingi kichik harfning avtomatik tarzda bosh harfga o'tkazilishi, bandlarning avtomatik raqamlashtirilishi; xatboshini to'g'rilash kabi mexanik ishlar avtomatik tahlilda erishilgan yutuqlar hisoblanadi.

Ma'lum bo'ladiki, matnlarni tuzatish ikki usulda amalga oshiriladi:

1. Inson tomonidan matnlarning tahrirlanishi (muharrirlik ishi).
2. Kompyuter vositasida tahrirlash (avtomatik tahrir).

Shuni ta'kidlash lozimki, har ikki holatda ham mutaxassisning faoliyati (qobiliyat, temperament, xarakter, irodaviy sifatlar, ijtimoiy holat)⁶⁰ muhim ahamiyatga ega. Muharrir matnlarni bevosita tahrir va tahlil qiladi, avtomatik tahlilda esa tilshunos bilvosita qatnashadi, ya'ni dasturning lingvistik bazasini yaratib beradi.

Dissertatsiyada rus, ingliz, nemis, fransuz kabi tillardagi leksemalar, jumlar, hatto, butun matnni ATT qilishga mo'ljallangan dasturlar yoki global tarmoqda (Internet) qidiruvga berilganda, ularni morfologik, sintaktik, semantik jihatdan tahlil qiluvchi dasturlar borasida ma'lumotlar berildi. Bundan ko'zlangan maqsad jahon tilshunosligida tan olingan va

⁶⁰ Abjalova M., Toshimov R. Matnlarni lingvistik tahrirlashning psixologik asoslari // O'zbekistonning mustaqil taraqqiyot va ijtimoiy-falsafiy tafakkur yangilanishi. Yosh olimlar ilmiy konferensiyasi materiallari, II qism. – Toshkent, 2011. – B. 155-158.

ayni damda foydalanilayotgan asosiy lingvistik dasturlar haqida ilmiy-nazariy faktlar, ularning ishlash darajasi, tahlil yoki tahrir davomidagi imkoniyatlarini o'rganishdir. M.Mahmudiyan ta'kidlaganidek, til, o'z ta'rifiga ko'ra, insoniyat tajribasining barcha sohalarini qoplamog'i lozim⁶¹. Shuni nazarda tutgan holda bugungi kunda ilm-fanning turli sohalarida, kundalik hayotda kompyuterdan foydalanish jarayonida o'zbek tilida matn yozilayotganda matnning to'g'ridan-to'g'ri tahriri va tahlilini ta'minlash masalasi o'zbek tilshunoslarining oldida yechimini kutayotgan masala ekanligini ta'kidlash o'zini oqlaydi.

O'zbek tilidagi matnlarni Microsoft Office Word dasturiga kiritish jarayonida, ularni ATT qilishga erishish, buning uchun dasturning lingvistik ta'minini yaratish maqsadida, bugungi kunda keng qo'llanishda bo'lgan dasturlar chuqur o'rganildi. Ularning amaliy ahamiyati qisqa sharh tarzida berildi.

Lingvistik tahlil dasturlari jahon tilshunosligida quyidagicha guruhlanadi:

1. Lingvistik texnologiyalar va sistemalar⁶².
2. Matnni lingvistik (morfologik, sintaktik) tahlil qilish dasturlari.
3. Matnni statistik tahlil qilish utilitlari⁶¹.

Quyida faqat birinchi guruh tarkibiga kiruvchi dasturlar borasida lingvistik ma'lumotlar berildi.

Nomi, muallif(lar)i	Vazifalari
<p style="text-align: center;">AskNet</p> <p>Intell Servis</p>	<p>Semantik savol-javob tarzidagi qidiruv sistemasi AskNet va ishlov berish instrumentlari rus va ingliz tillaridagi matnlarni lingvistik tahlil qilishni to'liq amalga oshiradi. Lingvistik tahlil qilish moduli o'z zimmasiga morfologiya (lug'atli va lug'atsiz), sintaksis, semantikani oladi. Intell Servisning dasturiy mahsulotlarida to'la versiyalar bilan korporativ (uyushma)li, saytli va shaxsiy (maxsus) qidiruv tizimlari taklif qilinadi. Internet bo'yicha savol-javobli qidiruv metaqidiruvli tizimlar bazasida (www.asknet.ru) amalga oshiriladi. Hozirda analitik qidiruv tizimi AQUA ishlab chiqilmoqda. Bu tizim matnni avtomatik tekshirishda semantik nuqtayi</p>

⁶¹ Махмудиян М. Лингвистика. – Москва, 1985. – С. 173.

⁶² www.krugosvet.ru

⁶¹ Utilita (ingl. utility) operatsion sistema (OS)ning ishi bilan bog'liq ravishda maxsus odatiy vazifalarni bajarishga mo'ljallangan yirik dasturiy ta'minotning bir qismi bo'lgan yordamchi kompyuter dasturi.

	nazardan savollarga javob topadi. Dasturlar savdo-sotiqda keng yoyilmoqda. Lingvistik tahlil darajalari: grafematik, morfologik, sintaktik va semantik. ⁶²
Ontos Ontos	Ushbu dastur matnli hujjatlarni tahlil qilish, annotatsiyaning tuzilishini tekshirish, kiritilgan matnlarga (OntosMiner, LightOntos for Workgroups, Ontos SOA, TAIS Ontos) ishlov berishga mo'ljallangan. Dastur algoritmlari grafematik, morfologik va semantik tahlilga qaratiladi. Sistemalar morfologik lug'atlar va semantik kartrijlardan asosiy tillar (ingliz, nemis, fransuz, rus) uchun foydalanadi. Matnda yo'l qo'yilgan xatolarning to'g'ri javoblari (variantlari) ko'rsatiladi, grafalarning bog'lanishi beriladi. Dasturlar Java.netda amalga oshiriladi va ular Microsoft SQL Serverda ishlatiladi. Lingvistik tahlil darajalari: grafematik, morfologik, semantik – tematik sinonimlar.
Google Mini Google Desktop Google	Dasturli apparat kompleksi Google Mini va Google Desktop dasturlari tarmoq ichi, ya'ni Internetda kompyuter vositasida xabarlarni qidirishga mo'ljallangan. Asosiy tillarda (rus, ingliz va h.k.) kalit so'zlar bo'yicha morfologik jihatdan tekshiruvni amalga oshiradi. Lingvistik tahlil imkoniyatlari: grafematik va morfologik.
Yandex. Server Yandeks	Dastur (NTTR) serveri uchun mo'ljallangan, bir yoki bir necha saytlar bo'yicha qidiruv va raqamlashni amalga oshiradi. Qidiruv rus, ingliz va ukrain tillarining morfologiyasi bo'yicha ishlaydi. Qidiruv natijasi hujjatlar ro'yxatining relevantlik yoki sana asosida tartiblanishi hisoblanadi. Relevantlikda hujjatdagi leksemalarning miqdori, ularning chastotali tavsifi va leksemalararo joylashishi inobatga olinadi. Matnlarning sintaktik va semantik tahlili hisobga olinmagan. Dastur versiyalari Windows va Unix uchun mo'ljallangan. Lingvistik tahlil darajalari: grafematik va morfologik.
Galaktika- ZOOM	Dastur hujjatning ma'noga ega so'z va so'z birikmalarini aniqlaydi, foydalanuvchi tomonidan kiritilgan kalit so'zlar hamda ularning sinonimlari

⁶² www.asknet.ru

Galaktika	yuzasidan tekshiradi, qidiruv ishini olib boradi. Shuningdek, hujjatda uchraydigan soʻzlarning chastotasi boʻyicha statistikani formallashtiradi. Dastur rus tilidagi matnlarning tahlilini taʼminlaydi. Algoritmalar morfologik va statistik tahlildan foydalanishga asoslanadi. Lingvistik tahlil darajalari: grafematik va morfologik.
-----------	---

Demak, lingvistik taʼminoti tarkibining toʻldirib borilishiga koʻra ATT dasturini *turgʻun (statik) va dinamik* (oʻzgaruvchan) turlarga ajratish mumkin. Turgʻun tahlil dasturi LTi faqat mutaxassis tomonidan yaratiladi, uning neologizmlar, yangi lingvistik modellar, til meʼyorlari bilan bogʻliq oʻzgarishlar dasturning keyingi versiyalariga kiritib boriladi. Maʼlum boʻlganidek, LTga foydalanuvchi tomonidan toʻgʻridan-toʻgʻri soʻzshakllarni kiritish imkoniyati boʻlmaydi.

Dinamik muharrir dasturining vazifasi doimiy ravishda leksikografik taʼminotni yangi soʻzlar bilan toʻldirib, boyitib borish hisoblanadi. Koʻp hollarda ushbu vazifa foydalanuvchi tomonidan amalga oshiriladi, yaʼni matnda qoʻllanilayotgan muayyan neologizm, yangi termin maxsus soʻrovlarga binoan taʼminotning zarur moduliga kiritilishi natijasida keyingi oʻrinlarda uning turli shakllarda qoʻllanilganda tahrir va tahlil qilinishiga erishiladi. Bu matn muharrir dasturi rivoji uchun xizmat qiladi, chunki til doimiy harakatda, soʻzlar yangi-yangi maʼnolar kasb etishi, neologizm, arxaizm, istorizmlarning paydo boʻlishini kuzatib borish imkonini beradi. E.I.Bolshakova bunday imkoniyatga morfologik lugʻatni boyitish turlari sifatida qaraydi⁶³ va uni ikki turga ajratadi: 1) lugʻatli boyitish (turgʻun DT); 2) lugʻatsiz boyitish (dinamik DT) imkoniyati.

“Informatik” kompaniyasining lingvistik modullari. Tubanda Microsoft Word matn muharrir sistemasiga joylashtirilgan rus tilidagi turli sintaktik qurilishga ega matnlarni tahrir va tahlil qiluvchi ORFO avtomat tekshiruvchining ishi tahlil qilindi. Ushbu kompaniyaga 1989-yilda asos solingan. Kompaniya lingvistik texnologiyalar doirasidagi masalalarni ishlab chiqadi va ularni amalga oshiradi. 1995-yildan ushbu lingvistik texnologiyalar kompaniyasi Microsoft korporatsiyasi bilan hamkorlikda faoliyat olib borishni yoʻlga qoʻydi. Natijada rus mutaxassislari tomonidan ishlab chiqilgan LMLar dunyoning turli hududlariga yetib bordi. Bugungi kunda kompaniyaning lingvistik mahsulotlaridan dunyoning rivojlangan davlatlarida keng foydalaniladi. A. Mayorova “Doverayay, no proveryay.

⁶³ Большакова Е.И. и др. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика. – Москва: МИЭМ, 2011. – С. 112-122.

Sistemi kontrolya orfografi konseviruyut yazik” (“Доверяй, но проверяй. Системы контроля орфографии консервируют язык” / “Ishon, ammo tekshir. Imloni nazorat qilish sistemalari tilni saqlab qoladi”) maqolasida shunday yozadi: “MS Word uchun matn tekshirish dasturini “Informatik” kompaniyasi ishlab chiqqan. Loyiha rahbari sobiq sovet Fanlar Akademiyasining aspiranti Oleg Grigorev bo‘lgan. Uning hamkorlari Igor Ashmanov va Nina Russova elektron lug‘at, morfologik model va gapni morfologik tahlil qilish modelini ishlab chiqishgan. Dasturning ilk versiyasi DOS tizimi tarkibida ishlagan, 1991-yildan sotila boshlangan. Keyinchalik jamoaning katta qismi O.Grigorev bilan AQShga jo‘nab ketdi. I.Ashmanov sarosimaga tushmasdan tarkibni yangilab, 1992-1993-yillarda Windows sistemasi uchun imloni tekshirish sistemasini yaratadi. Mazkur ishlanma avvalgisidan farq qilmaydi⁶⁴”. ORFO avtomuharriri A.A.Zaliznyak muallifligidagi “Rus tilining grammatik lug‘ati” («Грамматический словарь русского языка»)⁶⁵ asosida yaratilgan. Mazkur lug‘at 1980-yillarda V.M.Andryushenko rahbarligida kompyuter texnologiyalariga o‘tkazilgan.

Kompaniyaning lingvistik modullari matnlar ustida ishlovchi mutaxassislar uchun morfologik kutubxona hisoblanadi. “Informatik” kompaniyasi dasturiy morfologik modullarni matnni morfologik tekshiruvchi dastur yaratuvchilari uchun taklif qiladi. Ushbu dasturiy morfologik modullar ma’lumotli qidiruv sistemasi, analitik sistemalar, elektron hujjatlar katalogi doirasida yirik hajmli matnlarni tekshiruvchi, qidiruv vazifasini bajaruvchi dasturlarning bazasi hisoblanadi.

E.K.Lavoshnikova “О компьютерной проверке синтаксических конструкций в текстах на русском языке” maqolasida rus tilidagi matnlarni uslubiy va grammatik jihatdan tekshirish dasturlari borasida so‘z yuritib, MS WORD matn muharririga joylashtirilgan ORFO 2002 sistemasining afzalliklarini atroflicha ko‘rsatib beradi⁶⁶.

“Informatik” kompaniyasining morfologik modullarida hujjatlarning kontekstli qidiruvida sinonimlarni hisobga oladigan, ularning grammatikasi va orfografiyasini tekshiradigan maxsus texnologiyalar, lingvistik va matematik algoritmlardan foydalanilgan.

⁶⁴ Майорова А. Доверяй, но проверяй. Системы контроля орфографии консервируют язык / «Известия.Ру», izvestia.ru сайтидан

⁶⁵ Зализняк А.А. Грамматический словарь русского языка: Словоизменение: около 100000 слов. – Москва: Русский язык, 1977. – 879 с.

⁶⁶ Лавошникова Э. К. О компьютерной проверке синтаксических конструкций в текстах на русском языке // Информационные процессы. Том5, №3, 2005, – С. 201–212.

Ushbu kompaniya tomonidan yaratilgan morfologik modullar Samsung, Syngenta, Quantum Art, ALP, Yandex kabi yirik tashkilotlar tizimida keng foydalaniladi.

Kompaniya quyidagi modullarni taklif qiladi:

1. Orfografik tekshiruv.
2. Grammatik tekshiruv.
3. Morfologik tahlil. Rus, ukrain, nemis, ingliz, fransuz, ispan, itilyan, portugal, braziliya tillari qamrab olingan.
4. Tezaurus
5. Microsoft uchun qidiruv va ma'lumot tashuvchilarni joylash imkoniyati.

Orfografik tekshiruv (Speller) – matni imlo jihatdan to'liq tekshiruvchi modul. Uning qulayligi shundaki, keltirilgan ko'rsatmalar orqali matni kirituvchi dasturning LTiga yangi so'zlar, so'zshakllarni kiritib, avtomatik lug'atni shu zahoti boyitish mumkin.

Foydalanuvchining avtomatik lug'atiga so'zlarni qo'shish moduli ikki variantda amalga oshiriladi:

1. Ekran interfeysi ichida joylashgan.

Rus tili me'yorlari asosida qo'shilgan so'zlar paradigmasi haqidagi farazlar ro'yxatini generatsiya qiladi. 90% holatlarda berilgan uch farazning birinchisi to'g'ri hisoblanadi. Foydalanuvchi ro'yxatdan to'g'ri farazni tanlab oladi va so'z barcha so'zshakllari bilan lug'atga kiritiladi. Agar so'z murakkab affikslar kombinatsiyasiga ega bo'lsa, foydalanuvchi uning ko'p foydalaniladigan qismini tanlab olishi mumkin. Foydalanuvchi lug'ati "Informatik" kompaniyasining boshqa LMLariga ulanishi mumkin.

2. Ekransiz interfeys.

Rus tili me'yorlari asosida qo'shilgan so'zlar paradigmasi haqidagi farazlar ro'yxatini generatsiya qiladi. Kompyuterning API funksiyasi har bir farazning so'zshakllari ro'yxati va ularning tavsifini olishga yordam beradi. Tanlab olingan farazlar foydalanuvchi lug'atga kiritilishi mumkin. Shundan keyin tekshiruv davomida ulardan foydalanish imkoni bo'ladi. Foydalanuvchi lug'ati "Informatik" kompaniyasining boshqa lingvistik modullariga ulanishi mumkin. Rus, ukrain, ingliz, fransuz, nemis, ispan, italyan, portugal tillariga mo'ljallangan.

Mahsulot Windows OT uchun dinamik lug'at (*.dll) ko'rinishida joylashtiriladi.

Grammatik tekshiruv (Russian Grammar) – rus tilining 40 dan ortiq turli grammatik me'yorlarini, shuningdek, gapdagi tinish belgilarni tekshiradi. Ushbu modul matnning xato yozilgan qismini topadi va o'sha

abzats (gap qismi)ni belgilaydi. Mahsulot Windows OT uchun dinamik lug‘at (*.dll) ko‘rinishida joylashtiriladi.

Morfologik kutubxona – so‘zlarning morfologik tahlilini amalga oshirishga mo‘ljallangan bir qancha modullarni birlashtiradi va quyidagi vazifalarni yechishda yordam beradi:

– So‘zlarni lug‘at shakliga keltirish. Bu bir so‘z shaklini uning boshqa shakli orqali topishni ta‘minlaydi. Masalan: *kitobni* so‘zini *kitobga* so‘zini qidiruvga berish orqali ham topish mumkin bo‘ladi. Ushbu modul tilning imlo lug‘ati yoki morfem lug‘atga muvofiq tayyorlanadi, hajmi cheklanmaydi. Har bir so‘z uchun kerakli ma‘lumot va 4 baytli raqamli xesh-kod (hash-code) xabari beriladi, bu esa matnni indeksirlashga yordam beradi.

– Berilgan so‘zning hamma formalari sintezi.⁶⁷ Modul kiritilgan so‘zning barcha shakllarini taklif qiladi. Morfologik kutubxona ORFO lug‘atdagi mavjud leksemalarning aniq tahlilini amalga oshiradi. Rus tili lug‘at fondini 180 mingga yaqin leksema tashkil qiladi, maxsus lug‘atlarda yana 60 mingga; so‘zshakllari bilan birgalikdagi miqdori 4 mln.dan ortiq, deb hisobga kiritilgan. Shuningdek, ukrain tilining lingvistik ta‘minoti (LT)ga 130 ming, ingliz tilining LTga 115 ming, nemis tilining LTga 100 ming, italyan tilining LTga 80 ming, fransuz tilining LTga 45 mingga yaqin leksema kiritilgan.

DTida mavjud bo‘lmagan leksemalar dastur interfeysi taklif qilgan paradigmalar ko‘rsatkichlari orqali foydalanuvchi tomonidan muayyan yangi leksemaga grammatik xarakteristika berish natijasida DTiga kiritiladi.

Bu jarayon dasturda quyidagi ko‘rinishga ega bo‘lishi mumkin⁶⁸

⁶⁷ *Morfologik sintez* matnga avtomatik ishlov berish jarayonida yoxud uning so‘nggi – natijalarni foydalanuvchi hukmiga havola qilish bosqichida so‘zlarning turli shakllarini chiqarish, analiz natijalarini umumlashtirishdan iborat. Leksema shakllarini sintez qilish jarayoni so‘zshakllarni tahlil (analiz) qilish bilan bir xil ma‘lumotlar bazasi va algoritmlarga asoslanadi. Shubhasiz, har qanday shaklni sintez qilish uchun ta‘minotda leksema va uning grammatik xususiyatlari kiritilishi talab qilinadi, shundagina dastur grammatik parametrlariga ega bo‘lgan har qanday leksemaning shakllarini tahlil va sintez qilish imkoniyatiga ega bo‘ladi.

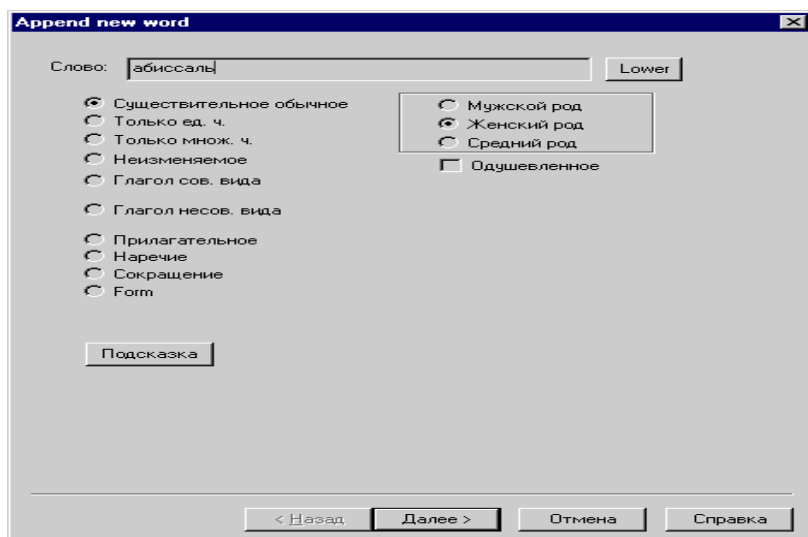
Sintez jarayoni uchun quyidagi parametrlar (xususiyatlar) zarur (<https://www.multitran.ru/morphology.htm>):

1) otlar uchun jins (o‘zga tillarda), son, kelishik;

2) sifat va ravishlarning (mavjud bo‘lsa) oddiy, qiyosiy va ortirma darajalari, ular mansub bo‘lgan jins, soni va turlanishi;

3) fe‘l va sifatdoshning zamoni, tuslanishi, shakli.

⁶⁸ <https://www.multitran.ru/morphology.htm>



3-rasm. Grammatik xarakteristika berish jarayoni.

Modulga qisqa xarakteristika:

- Katta hajmdagi soʻzlar lugʻati qoʻllaniladi.
- Lugʻatni tezkorlik bilan boyitish tizimi: 99 holatlarda kiritilayotgan soʻzlarning oʻzgarishini tizimning oʻzi aniqlaydi.
- Soʻzlarning identifikatorlarini generatsiyalovchi tizim: sistemaga maʼlum soʻzlarning har biriga identifikator qoʻyiladi, bu esa makromatn ichida bir soʻzning boshqa shakli orqali ham topilishini taʼminlaydi.
- Modulning kutubxonasi informatsion qidiruvda sistemaning morfologik tahlil⁶⁹ qilishini taʼminlaydi: soʻzlarga grammatik tavsif beradi, ularning soʻzshakllarini koʻrsatadi.
- Berilgan soʻzlarning barcha shakllari sintezlanadi.

Tezaurus (Russian Thesaurus) – rus tili sinonim, antonim va slavyan tillari oilasiga mansub leksemalar toʻplami boʻlib, semantik tahlilni amalga oshirishga xizmat qiladi.

Rus tilining sinonimlar lugʻatida 70000 dan ortiq soʻz va soʻz birikmalar mavjud⁷⁰. Lugʻat 10 000 ta sinonimlar guruhi (30 000 dan ortiq soʻz va soʻz birikmasidan iborat), 3 500 ta antonim va 14 000 ta slavyan tillar oilasiga mansub leksemalar qatori (20 000 yaqin bir asosli soʻzlar)ni oʻz ichiga oladi.

Tezaurus quyidagi imkoniyatlarni beradi:

- 1) ruscha soʻzlarning har qanday shaklini matndan topa oladi;

⁶⁹ *Morfologik tahlil* soʻzni morfemalarga ajratadi va grammatik maʼlumotlar tahlilini amalga oshiradi.

⁷⁰ Куц Виктор. Система проверки правописания ОРФО 2004. <http://www.comprice.ru/articles/detail.php?ID=43153>

2) har bir soʻzga, uning kiritilgan holatiga sinonim va antonimlarni taklif qiladi.

Microsoft uchun qidiruv. Russian Indexer for Microsoft – rus tilidagi hujjatlar bilan ishlash davomida *Microsoft Indexing Service* va *Microsoft SQL Server* imkoniyatlarini kengaytiradi. Bu dastur morfologik qidiruvni amalga oshiruvchilar, mutaxassislar, sistemali indegatorlar uchun moʻljallangan boʻlib, soʻzning barcha shakllarini topish va morfologik tahlilni amalga oshirish uchun qulay.

Russian Indexer for Microsoft imkoniyatlari: Microsoft SQL Serverda rus tilining morfologik meʼyorlari asosidagi toʻliq matnli indekslarni yaratish (full text search index); qidiruvning aniqligi, toʻliq amalga oshirilishini taʼminlash va tezligini oshirish; saytlar/internet-magazinlar boʻyicha natijali qidiruvdan foydalanish sanaladi.

Russian Indexer for Microsoft:

– Microsoft tajriba maʼlumotlarining tasnifiga mos ravishda ishlab chiqilgan;

– MS Office, XML, html format hujjatlarida ishlay oladi;

– formatlar roʻyxatini kengaytirish imkoniyatiga ega;

– faylli sistemalar katalogi bilan ishlaydi (WEB, arxiv hujjatlar);

– jadvallar boʻyicha ham ish koʻradi;

– lugʻat boyligi 280 ming soʻz (4.5 mln. soʻzshakl)ni tashkil etadi.

Koʻrinadiki, qisqa davr ichida rus matn muharriri katta natijalarga erishgan. Umuman, jahon kompyuter lingvistikasida mazkur soha boʻyicha koʻplab ilmiy va amaliy ishlanmalar amalga oshirilgan, hozirda tinimsiz rivojlanmoqda.

1.4. Matnlarni avtomatik tahrir va tahlil qilish dasturining lingvistik taʼminoti manbalari

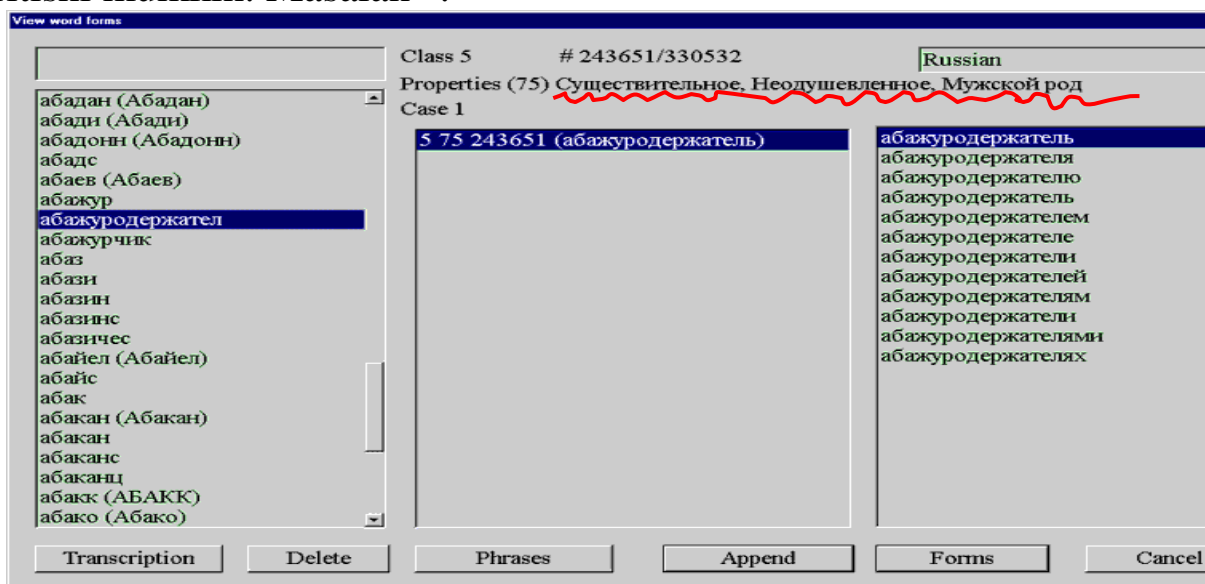
Istalgan tildagi matnlarni avtomatik tahrir va tahlil qilishda kompyuter xotirasida muayyan LT boʻlishi talab qilinadi. Lotin yozuviga asoslangan yangi alifboni isloh qilish masalasi kun tartibida turgani bois mazkur tadqiqotda kirill yozuviga asoslangan imlo qoidalari va leksik manbalariga tayanildi. Kelgusida lotin yozuvidagi matnlarning ATTini amalga oshirish jarayoniga qulaylik yaratish maqsadida ayrim oʻrinlarda lotin imlosiga oid algoritmik qoidalar berildi.

Taʼminot asosini lingvistik qoidalar va meʼyorlar tashkil etadi. Shuningdek, LTdan muayyan tilning lingvistik va filologik lugʻat boyligi ham oʻrin oladi. Shu bois lingvistik protsessorni yaratishda qayta ishlanayotgan tabiiy til haqida yetarli darajada maʼlumotga ega boʻlish

talab qilinadi. Bu o‘z navbatida lingvistik protsessorning mukammal bo‘lishini ta‘minlaydi. Lingvistik ma‘lumotlar tabiiy tilning adabiy me‘yorlaridir.

O‘zbek tilidagi matnlarni avtomatik tahrir va tahlil qilish dasturining lingvistik protsessorini yaratishda tilshunosga o‘zbek adabiy tilining lingvistik lug‘atlari va grammatik qoidalar jamlanmasi zarur bo‘ladi.

Xorij KLda lingvistik lug‘atlar orasida nisbatan keng tarqalgan turi **morfologik lug‘at** hisoblanadi. Bunday lug‘atda leksemaning qaysi so‘z turkumiga oidligi, uning grammatik shakllari ro‘yxati, agar leksema flektiv tilga oid bo‘lsa, o‘zgarish holatlari ham (masalan: *pisat – pishu, vodit – voju; child – children, man – men*) aks etgan bo‘lib, lug‘at matnlarni morfologik tahlil qilishda yordam beradi. Lingvistik protsessorni yaratuvchi o‘zining ish uslubiga ko‘ra lug‘atga grammatik affikslar kiritishi mumkin. Masalan⁷¹:



4-rasm. Lemmani grammatik shakllantirish jarayoni.

Morfologik lug‘at so‘zshaklning, ya‘ni ismlarning turlanishi va fe‘llarning tuslanishini shartli belgilashning maxsus sistemasi orqali aks ettiradigan leksikografik manba hisoblanadi. Lug‘at to‘g‘ri yoki teskari alifbo tartibida tuziladi. Teskari (inversion) tartibda so‘zning so‘nggi harfi e‘tiborga olinadi. Lug‘at avvalida turlanish va tuslanish hodisasi namuna tarzida ko‘rsatilgan “Grammatik ma‘lumotlar” beriladi. Shu bois har bir so‘zda “grammatik ma‘lumot”ga ishora qiluvchi grammatik belgi va indeks bo‘ladi. Bu jihat foydalanuvchiga muayyan so‘zning shakl o‘zgarish hodisasini aniqlashga yordam beradi. Namunani qidirishni tezlashtirish va qulaylik yaratish uchun lug‘atning har bir sahifasi yuqorisi

⁷¹ <https://www.multitran.ru/GIF/ViewWordForms.gif>

yoki kolontitulida sahifada uchragan indekslar hisobi va indekslar ifodalagan “grammatik ma’lumot”lar lug‘atning qaysi sahifasida joylashgani haqida ma’lumot keltiriladi. Morfologik lug‘atni yaratish ko‘lami va uning ahamiyatini yaqqol ko‘rsatish maqsadida quyida taniqli rus leksikografi A.A. Zaliznyakning “Грамматический словарь русского языка” kitobidan fragment berildi:

отретуши́ровать	св 2а	стасова́ть	св 2а
варьи́ровать	нсв 2а	растасова́ть	св 2а ⊗ I
интервьюи́ровать	св-нсв 2а	фасова́ть	нсв 2а
пронтервьюи́ровать	св 2а	расфасова́ть	св 2а ⊗ I
кровать	ж 8а	всова́ть	св 2b ⊗ I (см.)
дива́н-кровать	м, склоняются обе части (косв. формы и определения избегаются)	подсова́ть	св 2b ⊗ I (см.)
крёсло-кровать	с, склоняются обе части (косв. формы и определения избегаются)	колесова́ть	св-нсв 2а
соборо́вать	св-нсв 2а	адресова́ть	св-нсв 2а
особоро́вать	св 2а	переадресова́ть	св 2а ⊗ I
ворова́ть	нсв 2а	пересова́ть	св 2b
наворова́ть	св 2а ⊗ I	интересова́ть	нсв 2а
обворова́ть	св 2а ⊗ I	заинтересова́ть	св 2а ⊗ I
разворова́ть	св 2а ⊗ I	рисова́ть	нсв 2а
поворова́ть	св 2а	зарисова́ть	св 2а ⊗ I
сворова́ть	св 2а	нарисова́ть	св 2а
уворова́ть	св 2а ⊗ I	обрисова́ть	св 2а ⊗ I
озорова́ть	нсв нп 2а	подрисова́ть	св 2а ⊗ I
созорова́ть	св нп 2а	перерисова́ть	св 2а ⊗ I
селитрова́ть	св-нсв 2а	разрисова́ть	св 2а ⊗ I
нитрова́ть	св-нсв 2а	изрисова́ть	св 2а ⊗ I
		пририсова́ть	св 2а ⊗ I
		дорисова́ть	св 2а ⊗ I
		порисова́ть	св 2а
		прорисова́ть	св 2а ⊗ I

O‘zbek tilidagi matnlarni ATT dasturining morfologik tahlil qilish bosqichi modullarini yaratishda ham morfologik lug‘at juda muhimdir. Ammo bunday lug‘atning yo‘qligi bois asosiy manba vazifasini bajaruvchi sifatida morfem lug‘atlardan foydalanildi. Bunday lug‘at birinchi marta A.G‘ulomov, A.N.Tixonov, R.Q.Qo‘ng‘urov tomonidan tuzilib, 1977-yilda “O‘qituvchi” nashriyoti tomonidan chop etilgan⁷².

Imlo lug‘atlaridagi leksemalarning qaysi so‘z turkumiga oidligini aniqlashda “O‘zbek tilining izohli lug‘ati”⁷³ muhim yordamchi manba hisoblanadi. Leksemaning o‘ziga qanday affikslarni biriktira olishi, affikslar kombinatsiyalarini tuzishda “O‘zbek tilining morfem lug‘ati”ga tayanildi.

Ma’lumki, adibning til mahorati, eng avvalo, sinonimik qatorlardan foydalanishda ko‘zga tashlanadi. Sinonimik qatordan mos, kerakli birlikni tanlab olish ekspressivlikni, subyektiv bahoni aniq ifoda etishning eng

⁷² G‘ulomov A. va boshq. O‘zbek tilining morfem lug‘ati. – Toshkent. 1977 – 463 b.; Samad A. “X” va “H”li so‘zlarning imloviy lug‘ati. – Toshkent, 2007. – 346 b.; Mengliyev B., Bahriddinova B. O‘zbek tilining so‘z tarkibi o‘quv lug‘ati. – Toshkent: Yangi asr avlodi, 2007.

⁷³ O‘zbek tilining izohli lug‘ati: 80 000 dan ortiq so‘z va so‘z birikmasi (A. Madvaliyev tahriri ostida). 5 jildli. – Toshkent: O‘zbekiston milliy ensiklopediyasi, 2006.

to'g'ri yo'li bo'lganligi tufayli u matn tuzishda zarur lingvistik vosita sanaladi. Buning ustiga tilda sinonimik variantlar borligi uchun uslubiy me'yorni belgilash imkoniyati mavjud. A.Hojiyevning "O'zbek tili sinonimlarining izohli lug'ati"⁷⁴ ana shu sinonimik variantlarni farqlashda muayyan darajada amaliy yordam beradi.

O'zlashma leksemalar lug'atida o'zbek tiliga boshqa tildan kirib kelgan leksemalar izohlanadi. Bu tildagi lug'at – izohli lug'atning bir ko'rinishi bo'lib, ular qaysi tildan o'zlashgan leksemani izohlashiga ko'ra turlicha nomlanadi⁷⁵. Texnika asridagi yangilanish va o'zgarishlar tufayli terminologiyada ko'plab o'zlashma so'zlar paydo bo'lmoqda, ilmiy va rasmiy terminlar ko'lami oshmoqda. Shu bois o'zlashma so'zlar lug'atining yangi tahririni chiqarish zamon talabidir.

Frazeologizmlar mohiyat e'tibori bilan, asosan, so'zlashuv va badiiy nutq mevasidir. Ulardagi boshqa uslubga xos chegaralanishlar esa ma'lum muddat keyin yuzaga keladi. Chunonchi, birgina *o'lmoq* leksemasining ma'nosini anglatadigan yuzga yaqin evfemizmlar sinonimik qatorining paydo bo'lishi ularning vazifaviy chegaralanish imkoniyatini tug'diradi. Masalan, bu qatorga kiradigan *olamdan o'tmoq, dunyodan o'tmoq, omonatini topshirmoq, qulog'i ostida qolmoq, jon bermoq* so'zlashuv uslubida ishlatilsa, *vafot etmoq, hayotdan ko'z yummoq, dunyodan ko'z yummoq, hayot bilan vidolashmoq* kabilar ommabop va rasmiy uslublarda uchraydi. *Alloh rahmatiga yo'l tutmoq, shahodat sharobini ichmoq, dorilfanodan dorilbaqoga rixlat qilmoq* singlarlar esa badiiy matnga tegishlidir. Bundan ko'rinib turibdiki, frazeologizmlar lug'ati⁷⁶ dasturning mukammalligini ta'minlovchi manbalardan biri hisoblanadi.

Iboralar turg'un so'z birikma hisoblansa-da, komponentlari o'rtasidagi sintaktik derivatsiya dinamik xarakterda bo'ladi. Bu holat iboralarning faqat shakliy tomoniga ta'sir qilib, uning semantikasida ma'no yaxlitligi va majozilik saqlanadi. Birikma komponentlarining sintaktik munosabati o'z o'rnida frazeologik birikmaning jumlada sintaktik munosabatga kirishishida birlamchi hisoblanadi. Masalan: *sabr kosasi to'lmoq* – predikativ belgiga ega ibora, N (N₀+N_{e.a.}=>otli birikma)+V: *Mening sabr kosam to'ldi; Sabr kosang to'lgandir; Uning sabr kosasi ham to'ldi; Sabr*

⁷⁴ Hojiyev A. O'zbek tili sinonimlarining izohli lug'ati. – Toshkent: O'qituvchi, 1974. – 308 b.

⁷⁵ Usmon O., Doniyorov R. Ruscha-internatsional so'zlar izohli lug'ati. – Toshkent, 1965.; Hamrayeva Y. O'zbek tilining o'zlashma so'zlar o'quv izohli lug'ati. – Toshkent: Yangi asr avlodi, 2007.

⁷⁶ Rahmatullayev Sh. O'zbek tilining izohli frazeologik lug'ati. – Toshkent: O'qituvchi, 2001. – 407 b.; Mengliyev B. va boshq. O'zbek tili iboralarining o'quv izohli lug'ati. – Toshkent: Yangi asr avlodi, 2007.

kosamiz to'lib bormoqda. Shu bois LTda 1393 frazema bazasi yaratildi, frazema soʻzga teng birlik [frazema=soʻz] deb olindi va morfologik tahlil moduli uchun ularda grammatik shakllarning oʻzgarish oʻrni belgilandi.

Matni yozish jarayonida nutqda talaffuzi bir-biriga yaqin leksemalarni yanglish almashtirib yoxud aytilishi boʻyicha yozish hollari ham uchramoqda. Masalan, *tanbur* oʻrniga *tambur*, *adil* oʻrniga *adl*, *teri* oʻrniga *tire*, *yonilgʻi* oʻrniga *yoqilgʻi*, *sudxoʻr* oʻrniga *sutxoʻr* qoʻllash kabi. Koʻpchilik AYoQSh abbreviatsiyasining kengaytmasi *avtomobil yonilgʻi quyish shoxobchasi* ekanligiga ahamiyat bermaydi. Shuning uchun *avtomobil yoqilgʻi quyish* (yoxud *qoʻyish*) *shaxobchasi* deb oʻylashadi. Vaholanki, *yonilgʻi* – gaz, kerosin, benzin kabi yonuvchi suyuqliklar majmui; *yoqilgʻi* – oʻtin, koʻmir kabilar majmui⁷⁷. Shuni nazarda tutib, bunday leksemalarning maʼnolarini farqlash maqsadida A.Maʼrufovning “Paronimlar lugʻati”⁷⁸ga murojaat qilamiz. Matni kirituvchi shaxs leksemaning toʻgʻri kiritilganiga shubha qilayotgan paytida dastur protsessoriga joylashtirilgan paronimlar lugʻatiga murojaat qiladi va leksemaning anglatgan maʼnosiga qarab uni oʻz oʻrnida qoʻllaydi.

ATT qilish dasturini yaratishdagi lingvistik muammolardan biri omonimlik hodisasini avtomatik tushunish masalasidir (qarang: II bob, 2.4-paragraf). Ushbu masala birmuncha dolzarb boʻlib, uning echimini topishda, albatta, maxsus modullar muhim ahamiyatga molikdir. Lingvistik modullar dastur bazasidagi mavjud omonim soʻzlar uchun moʻljallangan boʻladi. Shu bois bazaga Sh.Rahmatullayevning “Oʻzbek tili omonimlarining izohli lugʻati”⁷⁹ kitobida keltirilgan, shuningdek, keyinchalik nutqda paydo boʻlgan jami 1378 ta omonim birliklar jamlandi. Mazkur lingvistik taʼminotga jamlangan omonimshakllar statistikasi tubandagicha: otshakl 866 ta, feʼlshakl 227 ta, ravishdosh 2 ta, harakat nomi 3 ta, koʻmakchi feʼlshakl 7 ta, sifatshakl 183 ta, sonshakl 7 ta, ravishshakl 19 ta, olmoshshakl 5 ta, bogʻlovchishakl 5 ta, koʻmakchishakl 5 ta, yuklamashakl 8 ta, modalshakl 7 ta, undovshakl 12 ta, taqlidshakl 16 ta, hisob soʻzshakl 3 ta, predikativ soʻzshakllar 5 ta.

Tezaurus va ontologiyalar lingvistik manbalarning oʻta muhim turlaridan hisoblanadi. Tezaurus (yunoncha thesaurós – *xazina*, *boylik*) lugʻatning oʻziga xos maxsus koʻrinishi boʻlib, unda lugʻaviy birliklar oʻrtasidagi semantik munosabat (sinonimlar, antonimlar, paronimlar,

⁷⁷ Maʼrufov A. Paronimlar lugʻati. – Toshkent: Oʻqituvchi, 1974. – B.20.

⁷⁸ Oʻsha asar.

⁷⁹ Rahmatullayev Sh. Oʻzbek tili omonimlarining izohli lugʻati. – Toshkent: Oʻqituvchi, 1984. – 215 b.

giponimlar, giperonimlar va h.k.) ko'rsatilgan bo'ladi. Shu bois mazkur lug'atlar ATTSning semantik tahlil bosqichida muhim ahamiyatga ega hisoblanadi.

Izohli lug'atdan farqli ravishda tezaurus nafaqat so'zning ma'nosini izohlaydi, balki muayyan so'zning boshqa so'zlar bilan bog'lanishi (valentligi)ni ko'rsatib beradi. Bu esa tezaurus lug'atlardan sun'iy intellekt tizimining bilimlar asosini oshirishda foydalanish mumkinligini ko'rsatadi. Bunday lug'atlarning ommalashuvi informatsion qidiruv vazifasining yechimini beradi⁸⁰.

Tezaurus tushunchasi bilan ontologiya tushunchasi chambarchas bog'liq⁸¹. Ontologiya (yunoncha ontós – *borliq*, logós – *ta'limot*) – borliq haqidagi falsafiy ta'limot. Ontologik lug'atlarning lingvistik hisoblanishining sababi ular ma'lum bir tilning so'z boyligi asosida tuzilishi mumkin.

Ontologiyani til birliklarini nutqiy vaziyat, til egalarining ruhiyati, olam haqidagi bilimlari, intellektual salohiyati, assotsiativ tafakkuri, milliy-madaniy qarashlari kabi ekstralingvistik omillar bilan bog'liq holda to'ldirish muhim ahamiyat kasb etadi. Shundagina, lingvoontologiya mukammal foydalanishga ega bo'ladi. Zero, tildagi mavjud leksemalar inson kognitsiyasi mahsuli sifatida til egalarining olam haqidagi bilimlarini saqlash va uzatish vositasi bo'lib xizmat qiladi.

WordNet⁸² tizimi shunday lingvistik ontologiya hisoblanib, u yirik leksik resursdir. Ushbu tizimda ingliz tilidagi ot, sifat, fe'l, ravish turkumlariga oid leksemalar va ularning semantik bog'lanishi mavjud. Tizimda ko'rsatilgan turkumlarning har biridagi so'zlar antonimlar, giponim va giperonimlar (jins-tur), xolonim va meronimlar (butun-bo'lak) bilan munosabatda bo'lgan sinonimlar guruhi (*sinsetlar*)ga ajratilgan. Tizim resursi 25000 ga yaqin so'zni qamragan. Jins-tur munosabati uchun ierarxik munosabatdagi so'zlar soni esa 15000 tani (7-8) tashkil etadi. Ierarxiyaning eng yuqorigi darajasi olam haqidagi asosiy tushunchalar tizimi – umumiy ontologiyani yuzaga keltiradi. Ingliz WordNet sxemasi bo'yicha boshqa g'arb tillarining leksik resurslari tuzilgan.

Nafaqat matnlarni ATT qiladigan, balki matnlarni qayta ishlovchi barcha lingvistik dasturlarning alohida lingvistik resursi tabiiy tilning

⁸⁰ Лукашевич Н.В., Салий А.Д. Тезаурус для автоматического индексирования и рубрицирования: разработка, структура, ведение // НТИ, Сер. 2, №1, 1996. – С.1-6.

⁸¹ Hirst, G. Ontology and the Lexicon. In.: Handbook on Ontologies in Niformation Systems. Berlin, Springer, 2003.

⁸² Word Net: an Electronic Lexical Database / Edit. by Christiane Fellbaum. Cambridge, MIT Press, 1998.

grammatikasidir. Birinchi o‘rinda grammatika o‘zining me‘yoriy qoidalari bilan muhimdir. Grammatik qoidalar soni morfologiya va sintaksis modeliga bog‘liq.

Xulosa tarzida aytish joizki, lingvistik dasturlarning ta‘minotini yaratishda grammatik lug‘atlar muhim ahamiyatga ega. Grammatik lug‘atlar leksemaning morfologik va sintaktik xususiyatlari haqidagi ma‘lumotlarni o‘z ichiga olgan lug‘atlar bo‘lib, ularda leksemalar to‘g‘ri yoki teskari alifbo tartibida yoziladi. Leksema haqidagi ma‘lumotlar hajmi va leksemalarni tanlash prinsipi har bir grammatik lug‘atni yaratuvchining maqsadiga qarab farqlanadi⁸³. A.Po‘latov hozirgi zamon talablariga ko‘ra samaradorligi katta bo‘lgan lug‘at kompyuter lug‘ati ekanligini ta‘kidlaydi va kompyuterli lug‘at qismlari hamda unga quyiladigan talablarni keltiradi⁸⁴. E.I.Bolshakova o‘z izlanishlarida lingvistik protsessor lingvistik ma‘lumotlar modeliga asoslanishini, ya‘ni DTning lingvistik resurslarini quyidagilar tashkil etishini ta‘kidlaydi: kompyuter lug‘atlari, tabiiy til grammatikasi, so‘z birikmalari bazasi, tezaurus va ontologiyalar, matnlar majmuasi va korpus⁸⁵.

Kompyuterli tezaurus va kompyuterli grammatik lug‘atlarni tuzish dasturning lingvistik modullarini ishlab chiqishdan ko‘ra keng ko‘lamli va sermashaqqat ishdir. Shuning uchun KLning muhim vazifalaridan biri avtomatlashtirilgan lingvistik resurslarni yaratish hisoblanadi⁸⁶.

⁸³ http://gramota.ru/slovari/types/17_5

⁸⁴ Po‘latov A. Kompyuter lingvistikasi. – Toshkent: Akademnashr, 2011. – B. 213.

⁸⁵ Большакова Е.И. и др. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика. Учебное пособие. – Москва: МИЭМ, 2011. – С. 97-99.

⁸⁶ Hearst, M.A. Automated Discovery of WordNet Relations. In: Fellbaum, C. (ed.) WordNet: An Electronic Lexical Database. MIT Press, Cambridge, 1998, – P. 131-151.

Matsumoto Y. Lexical Knowledge Acquisition. In: The Oxford Handbook of Computational Linguistics. Mitkov R. (ed.). Oxford University Press, 2003, – P. 395-413.



5-rasm. ATT manbalari.

Matnlarni ATT qilish dasturini innovatsion taʼlimning bir koʻrinishi sifatida qayd etib, uning quyidagi imkoniyatlarga ega ekanligini koʻrsatish mumkin:

- 1) turli hajmdagi matnlarni tezkor tahrir qiladi;
- 2) matn teruvchining yozma nutq malakasini takomillashtiradi;
- 3) foydalanuvchiga matnlarni mustaqil ravishda toʻgʻri va savodli yozishida koʻmak beradi;
- 4) oʻzbek tilining leksik, grammatik meʼyorlarini oʻrganishda lingvistik manba vazifasini oʻtaydi.

1.5. Oʻzbek tilidagi matnlarni avtomatik tahrir va tahlil qilish sistemasini yaratishda uslub masalasi

Maʼlumki, til – maʼlumot manbai, maʼlumot esa (tilning yozma shaklida) matnda namoyon boʻladi. Matn oʻz navbatida til birliklarining grammatik jihatdan shakllanishidan yuzaga keladi. Demak, til va matn axborot almashuvining yozma shaklida bir-birini taqozo etuvchi vositalar hisoblanadi (“TIL↔MATN”). Shuningdek, til xoh ogʻzaki, xoh yozma shaklda boʻlsin, muloqot manbai sanalanadi. E.V. Yagunova tabiiy tilning har bir foydalanuvchisi yozma tildan foydalanish koʻnikmasini egallasa (ogʻzaki til qatorida) va bir tildan ikkinchi tilga (bir koddan ikkinchi kodga) muloqot talabi bilan oʻta olish qobiliyatiga ega boʻlsa uni bilingv

deb atash haqli, deydi⁸⁷. Ko‘rinadiki, “TIL↔INSON” bog‘liqligi muhim o‘rin tutadi. Yuqoridagi fikr bilan birlashtirganda “TIL↔INSON↔MATN” bog‘lanishi yuzaga keladi, ya‘ni til imkoniyatlari inson orqali yuzaga chiqadi, inson ma‘lumotlarni muloqot vositasida yetkazadi, ma‘lumotlar matnda saqlanadi va matn orqali namoyon bo‘ladi. Bir tabiiy tilning minglab foydalanuvchilari, ularning har birida turli xil ma‘lumotlarni boshqalarga yetkazish maqsadi bo‘ladi. Shu bois bir tilda yaratilgan matnlardagi ma‘lumotlar turi va darajasiga qarab bir-biridan uslub jihatidan farqlanadi.

O‘zbek tilidagi matnlarni ATT qilish dasturi lingvistik bazaning yaratilishida milliy tilning funksional uslublari e‘tiborga olinishi darkor. Chunki til uslublariaro xilma-xillik, bir-biridan farqlanuvchi xususiyatlar LTning to‘liq va mukammal shaklda qisqa vaqt ichida dastur protsessoriga yetkazilishiga to‘sqinlik qiladi. Bu, asosan, ilmiy, rasmiy va publitsistik uslublarning so‘zlashuv hamda badiiy uslublardan vazifa va xususiyatlari jihatidan faqlanishida ko‘rinadi.

O‘zbek tili uslublari vazifaviy jihatdan tasnif qilish ikki omilga – til va tildan tashqarida bo‘lgan omillarga tayanadi⁸⁸. Uslublarning A.Shomaqsudov, I.Rasulov, R.Qo‘ng‘urov, E.Begmatovlar tomonidan tavsiya etilgan variantlari mavjud bo‘lib, ularda asosan beshta uslub – so‘zlashuv, ommabop, ilmiy, rasmiy va badiiy uslublari e‘tirof etiladi⁸⁹

So‘zlashuv uslubi maishiy muloqotdan iborat bo‘lib, unda sheva, jargon, argo, varvarizm, vulgarizm kabilarni o‘z ichiga olgan til qatlami yetakchilik qiladi. Shu bois so‘zlashuv uslubi ATT dasturining lingvistik ta‘minotida o‘z o‘rniga ega emas.

Badiiy uslub rang-barang xarakterga ega. Unda milliy tilning barcha tasvir vositalari va hodisalari namoyon bo‘ladi. Hatto she‘riy matnlarni hisobga olmaganda ham, ko‘pgina badiiy uslubdagi matnlarni bir xil xususiyatlar bo‘yicha tuzish qiyin. Badiiy uslubdagi matnlarda ma‘lumot tashish vazifasi yo‘q bo‘lib, asosan, estetik vazifani bajaradi. Rasmiy, ilmiy va publitsistik uslublardagi matnlar asosida ma‘lumot yotadi. Misol sifatida qonun, shartnoma, farmon, farmoyish matnlari (yuridik kuchga ega bo‘lgan va to‘g‘ri ma‘noda tuzilgan), gazeta, jurnal, ilmiy nashrlarni

⁸⁷ Автоматическая обработка текстов на естественном языке и компьютерная лингвистика. Учебное пособие / Большакова Е.И., Клышинский Э.С., Ландэ Д.В., Носков А.А., Пескова О.В., Ягунова Е.В. – Москва: МИЭМ, 2011. – 272 с.

⁸⁸ Xudoyqulov M. Badiiy publitsistika janrlari. – Toshkent: Universitet, 2010. – B. 26.

⁸⁹ Shomaqsudov A., Rasulov I., Qo‘ng‘urov R., Rustamov H. O‘zbek tili stilistikasi. – Toshkent: O‘qituvchi, 1983, – 246 b.; Qo‘ng‘urov R., Begmatov E., Tojiyev Yo. Nutq madaniyati va uslubiyat asoslari. – Toshkent: O‘qituvchi, 1992.

sanab o'tishimiz mumkin. Ushbu uch uslubning funksional xususiyatlari deyarli bir xil, ya'ni bunday matnlarda aniqlilik, qisqalik, mantiqiy izchillik yetakchilik qiladi. Ammo ta'kidlash o'rinliki, rasmiy uslub boshqa uslublarga nisbatan qat'iy lingvistik talablarga bo'ysunadi va amal qiladi. Quyida o'zbek tilidagi mantlarni ATT dasturining qaysi uslubdagi matnlarni tahlil qilishga mo'ljallanganligining sabablari yoritildi.

Rasmiy uslubning yozma shaklida qonun matnlari, buyruq, farmoyish, farmon kabi rasmiy ish qog'ozlari, diplomatik shartnomalar va boshqa huquqiy hujjatlar yoziladi. Demak, ushbu hujjatlardagi bayon qilingan fikr ish qog'ozining nutqini tashkil qiladi.

Rasmiy uslubdagi matnlar noaniqlikdan yiroq bo'lib, fikrlar sodda, aniq mazmunda va tushunarli qilib yoziladi: *Davlat xalq irodasini ifoda etib, uning manfaatlariga xizmat qiladi. Davlat organlari va mansabdor shaxslar jamiyat va fuqarolar oldida mas'uldirlar* (O'zR Konstitutsiyasi, 2-modda).

Rasmiy uslubda yozilgan hujjatlardagi muhim holat shuki, ularda tayyor, formulaga kirib qolgan, andozalangan jumlar ishlatiladi, o'ziga xos ifodalariga ega: *buyuraman, yuklatilsin, ma'muriy javobgarlik, fuqarolik holati, fuqarolik javobgarligi, aybdor, gumondor, jabrlanuvchi, belgilanadi, asoslanadi, hisoblanadi, himoya qilinadi, amalga oshiradi, tan olinadi, ish ko'radi, ado etadi, kafolatlanadi, taqiqlanadi, tavsiya etiladi, jamoat kafilligi, surishtiruv, jabrlanuvchi bilan yarashuv* singari leksik birliklar, birikmalar qo'llanadi.

Shuningdek, hujjatlarda yangi leksemalar qo'llanilmaydi, agar leksemalar terminologik ma'noda bo'lsa ular umumiste'moldagi sodda leksemalar bilan almashtiriladi. Qisqartmalardan foydalanilganda ehtiyotkorlik talab qilinadi.

Ushbu uslub uchun jargonlar, oddiy so'zlashuvga xos leksemalar, emotsional-ekspressiv bo'yoqqa ega bo'lgan leksemalarning ishlatilishi me'yor sanalmaydi va shu jihati bilan boshqa uslublardan keskin farq qiladi.

Rasmiy uslubning grammatik me'yorlaridagi muhim jihat shuki, so'zshakllar morfemik qoida asosida (asos+so'z yasovchi affiks+lug'aviy shakl yasovchi affiks+sintaktik shakl o'zgartiruvchi affiks: *buyur+uq+lar+dagi=buyruqlardagi*) yoziladi. Bu esa dasturning lingvistik modullarini yaratishda tilshunosni *uylarimdagilarim, egnilarimda, bolalikdagilarimmi* kabi nome'yoriy so'zshakllar ustida vaqt sarflab, bosh qotirishdan saqlaydi.

Rasmiy uslubning sintaktik alomatlari ham matnda darhol ko‘zga tashlanadi. Unda darak gaplar, ayniqsa, qo‘shma gap shakli ko‘p ishlatiladi. Yoyiq va murakkab so‘z birikmalari mahsuldor hisoblanadi, murakkab tipdagi nomlar keng qo‘llaniladi. Gap tuzilishida o‘zbek tilining me‘yoriga amal qilinadi.

O‘zbekiston Respublikasi Adliya vazirligining 2014-yil 28-fevraldagi “Idoraviy normativ-huquqiy hujjatlarni tayyorlash va qabul qilish qoidalarini tasdiqlash to‘g‘risida”gi 53-mh-sonli buyrug‘ida idoraviy normativ-huquqiy hujjatlarni tayyorlash va qabul qilish qoidalari berilgan bo‘lib, ushbu qoidalar matnlarni avtomatik tarzda tahrir qiluvchi dastur ta‘minotini shakllantirishda rasmiy uslubdagi matnlar uchun tayanch manba bo‘lib xizmat qiladi. Ushbu buyruqda *idoraviy normativ-huquqiy hujjat* yoyiq so‘z birikmasiga vazirlik, davlat qo‘mitasi va idora tomonidan belgilangan tarzda qabul qilingan, umum majburiy davlat ko‘rsatmalari sifatida huquqiy normalarni belgilash, o‘zgartirish yoki bekor qilishga qaratilgan rasmiy hujjat⁹⁰, deb izoh berilgan.

Buyruqning 1-ilovasi, 2-§, 25-26-27-bandlarida quyidagi qoidalar kiritilgan bo‘lib, ular idoraviy normativ-huquqiy hujjatning ham shakliy, ham grafematik tuzilishini belgilab beradi:

“25. Huquqiy normalar arab raqamlarida belgilanib, nuqta qo‘yilgan bandlar ko‘rinishida bayon qilinadi va sarlavhaga ega bo‘lmaydi. Idoraviy normativ-huquqiy hujjat bandlari izchil ketma-ketlikda raqamlanishi kerak. Bandlar doirasimon qavslar bilan berkitiladigan kichik harflar yoki raqamlar ko‘rinishidagi kichik bandlarga hamda xatboshilarga bo‘linishi mumkin.

26. Idoraviy normativ-huquqiy hujjatda quyidagilarga:

– ikkitali raqamlarni ishlatishga (masalan, «1.1», «5.1.1», «1.a», «a.a» va h. k.);

– bir vaqtning o‘zida kichik bandlarni va mustaqil (kichik bandlar bilan bog‘lanmagan) xatboshilarni bayon etishga;

– xatboshilar boshlanishida defis yoki boshqa belgilar ishlatishga yo‘l qo‘yilmaydi.

27. Hajmiga ko‘ra idoraviy normativ-huquqiy hujjat kamayib boruvchi quyidagicha tuzilmaviy birliklarga bo‘linadi: bo‘limlar, boblar, paragraflar, bandlar, kichik bandlar va xatboshilar. Ikkinchi va keyingi xatboshilarni bo‘lishga yo‘l qo‘yilmaydi.

⁹⁰“Idoraviy normativ-huquqiy hujjatlarni tayyorlash va qabul qilish qoidalarini tasdiqlash to‘g‘risida”gi O‘zbekiston Respublikasi Adliya vazirligining 2014-yil 28-fevraldagi 53-mh-sonli buyrug‘i, 1-§, 1-bo‘lim, 2-band.

Bo‘limlar rim, boblar esa arab raqamlari bilan raqamlanadi va sarlavhalarga ega bo‘ladi. Paragraflar «§» belgisi bilan belgilanadi, arab raqamlari bilan raqamlanadi va sarlavhalarga ega bo‘ladi⁹¹.

Ilmiy uslub ilmiy asarlar uslubidir. Mazkur uslubda dalil va asoslarga tayanilishi, fikriy izchillikning yetakchilik qilishi jihatdan rasmiy uslubga yaqin turadi.

Ilmiy uslubda so‘zlar bir ma’noda qo‘llanadi, sinonimik qatordagi variantlardan masalaning mohiyatini birmuncha aniq ifoda etadigan varianti tanlanadi yoki neytral varianti (dominanta)ni qo‘llash taqozo etiladi. Terminlarni ishlatish bu uslubning asosiy xususiyati sanalgani bois tahrir va tahlil dasturining lingvistik ta’minotini yaratishda terminlarga ham jiddiy e’tibor berish talab etiladi.

Mantiqiy izchillik ilmiy bayon uslubining o‘ziga xos xususiyatini tashkil etadi. Ilmiy uslub matndagi so‘zlar, gaplar, abzatslarning o‘zaro mantiqiy bog‘lanishini talab qiladi. Shuning uchun ham ilmiy matnlar bu izchillikni yuzaga keltiruvchi *ma’lumki, ma’lum bo‘ladiki, ta’kidlash lozimki, ta’kidlash o‘rinli, darhaqiqat, binobarin, shunday ekan, birinchidan, ko‘rinadiki, ta’kidlash joizki, xulosa qilib aytganda, xulosa qilganimizda, birgina misol* kabi ko‘plab bog‘lovchi vazifasini bajaradigan leksema va birikmalar qo‘llaniladi. Lingvistik ta’minotda bu kabi ifodalar “tayyor ifodalar” tarzida bir guruhga kiritiladi.

Bu uslubda xuddi rasmiy uslubda bo‘lgani kabi, leksemalar ko‘chma ma’nolarda qo‘llanmaydi, tasviriy vositalardan unumli foydalanilmaydi. Subyektiv emotsionallik, obrazlilik, og‘zaki nutq elementlari, umuman, g‘ayri adabiy unsurlar kam ishlatiladi. Mumkin qadar bayon bir xil me’yorda ifoda etiladi. Ilmiy uslubning morfologik xususiyatlari to‘g‘risida M.Mukarramovning “Hozirgi o‘zbek adabiy tilining ilmiy stili” monografiyasida quyidagi fikrlar bayon qilingan:

Ilmiy uslubda faqat adabiy tilda me’yor sifatida e’tirof etilgan grammatik ko‘rsatkichlardan foydalaniladi. So‘z turkumlaridan ot faol, ko‘pincha takror qo‘llanadi. Subyektiv baho shakllari xos emas. *-lar* ko‘plikdan boshqa semantik-uslubiy ma’nolarni ifodalamasada, termin hosil qilishda ishtirok etadi. Umumiy egalik ustun turadi, I-II shaxs qo‘shimchalari ishlatilmaydi, III shaxs shakli faol. Undov va taqlid so‘zlar

⁹¹“Idoraviy normativ-huquqiy hujjatlarni tayyorlash va qabul qilish qoidalarini tasdiqlash to‘g‘risida”gi O‘zbekiston Respublikasi Adliya vazirligining 2014-yil 28-fevraldagi 53-mh-sonli buyrug‘i, 1-bob, 2-§, 25-26-27-bandlar.

ham bu uslubga xos emas. Yuklamalarning esa imkoniyati chegaralangan⁹².

Ilmiy nutq uchun *deyiladi, aytiladi, yoritiladi, bayon qilinadi, aniqlanadi, tekshiriladi, foydalaniladi, ta'riflanadi* singari xoslangan fe'llar mavjud. Fe'lning majhul daraja shakli ham bu uslub uchun xos.

Rasmiy va ilmiy uslublarning o'ziga xos xususiyatlari quyidagilardan iborat:

1. So'zlar o'z ma'nosi (denotativ)da va har bir sohaning o'ziga xos terminlaridan foydalaniladi.

2. Darak gaplar, qaror, buyruq ko'rsatmalardan foydalaniladi.

3. Gap bo'laklari odatdagi tartibda bo'ladi.

4. Rasmiy uslubdagi hujjatlar qisqa, aniq, lo'nda barcha uchun tushunarli qilib tuziladi.

5. Tasviriylik va majozilikka yo'l qo'yilmaydi.

6. Buyurish va undash ohangi kuchli bo'ladi.

7. Rasmiy hujjatlarning qolip matnlari – tayyor andozalari ishlab chiqilgan.

8. Yozma nutq shaklida amal qilinadi. Ifoda aniq, ixcham, izchil va tushunarli bo'ladi.

Til taraqqiyotida muhim omil sanaladigan publitsistikada tilning imkoniyatlaridan keng foydalanadi. Publitsistik janrda yozilgan asarlarning mohiyatan hozirjavobligi, axborot hamda targ'ibot-tashviqot xarakterda bo'lishi va ommani dunyo yoki mamlakat miqyosida sodir bo'layotgan voqea-hodisalardan zudlik bilan xabardor qilish natijasida yuzaga keladigan novatorlik, ularda tabiiy ravishda yangi ijtimoiy-siyosiy terminologiyaning qo'llanilishi va tilda me'yorlashishiga sabab bo'ladi. Boshqacha aytganda, ommabop uslub tilning yangi so'z va iboralar hisobiga boyib borishiga ko'maklashadi.

Ommabop uslubda shakllangan matnlar obrazlilik, ta'sirchanligi, tasviriy vositalarning faol qo'llanilishi bilan badiiy uslubga yaqinlashsa, dialektizmlar, istorizmlar, argo va jargonlar qo'llanilmasligi nuqtayi nazardan undan uzoqlashadi. Ifodaning aniqligi, publitsistik janrlarga xoslangan hamda ijtimoiy-siyosiy terminologiyaning qo'llanilishi jihatdan ilmiy uslubga o'xshaydi. Shu bilan birga, ifodadagi qisqalik, lo'ndalik, ixchamlik, yorqinlik kabi jihatlar uni ilmiy uslub bilan yonma-yon qo'yadi.

⁹² Mukarramov M. Hozirgi o'zbek adabiy tilining ilmiy stili. – Toshkent: Fan, 1984. – B.13.

Ijtimoiy-siyosiy terminlar asosan *-chi, -parast, -parvar, -xona, -noma, -dor, -kor* kabi affikslar yordamida yasaladi: *terrorchi, qo'poruvchi, aqidaparast, millatparvar, devonxona, bojxona, hissador, tadbirkor* kabi.

Kezi kelganda mazkur uslubda so'z yasaliş imkoniyatining boshqa vazifaviy uslublarga nisbatan mahsuldor ekanligini ta'kidlash o'rinli. Shuningdek, *MDH, BMT, NATO, NASA, YUNESKO, O'zMU, TDO'TAU* kabi qisqartmalardan keng foydalaniladi. Shu o'rinda tahrir dasturining grafematik tahrirlash bosqichining muhim ahamiyat kasb etishi yaqqol ko'rinadi.

Badiiy uslub o'zbek tili vazifaviy uslublari orasida o'ziga xos mavqega ega bo'lib, ayni paytda alohida me'yorlarga ham ega. Til materialini qamrab olish imkoniyatining kengligi, umumxalq tilida mavjud barcha lingvistik birliklar, shu bilan birga, boshqa vazifaviy uslublarga xos elementlarning ham ishlatilaverishi, ularning muhim bir vazifa – badiiy-estetik vazifani bajarishga xizmat qilishi badiiy uslubning asosiy xususiyati hisoblanadi.

Badiiy adabiyot tilidagi fonetik o'zgarishlarning mavjudligi (*So'rma mendan, kim diloru / Do'stmi yo janona deb. E.Vohidov; Ne balolig' kun edikim. Oshino bo'ldim sango. Alisher Navoiy; Quyosh orqasidan – behisob chirog' / Nuriga kiradi mamlakat shu chog'.* Zulfiya), shuningdek, orfografik me'yor talablariga muvofiq kelmaydigan *qaro, yamon, yaro, oshno, talosh* singari so'zlarning ishlatilishi lingvistik ta'minotni yaratishda badiiy uslubni cheklab turishga sabab bo'ladi. Bundan tashqari, hozirgi o'zbek adabiy tilida so'z yasovchi, ko'plik, egalik, kelishik, daraja, shaxs-son, zamon, mayl, nisbat kabi ma'nolarni ifodalovchi affikslarning nutqdagi ishtirokida, yordamchi so'zlarning turlicha ko'rinishlari va variantlarida namoyon bo'lishi badiiy uslubga oid matnlarda ko'p uchraydi. Bu esa o'z navbatida:

birinchidan, affikslarning har bir varianti va ular orqali hosil qilinadigan affiklar kombinatsiyasi ustida ishlashda ko'p vaqtni talab qiladi. Natijada ishni yakunga yetkazish sezilarli darajada orqaga suriladi;

ikkinchidan, affikslar kombinatsiyasi ulkan hisobni tashkil etishi bilan birga dastur protsessoridan ham katta hajmni talab etadi;

uchinchidan, hosil qilingan minglab kombinatsiyalarning har biri qaysi leksemaga birikishi yoxud har bir leksemaning o'sha kombinatsiyalar bilan bog'lana olish yoki bog'lanmasligini tekshirishda tilshunosdan o'ta jiddiylik va diqqat e'tiborni talab qildi. Chunki o'zbek tili leksikasi boy til hisoblanadi, ularning har birini hosil qilingan minglab

affikslarning kombinatsiyalari bilan birika olish xususiyatini tekshirib ko‘rish imkonsizdir.

Bundan tashqari, rasmiy, ilmiy va ommabop uslublarda grammatik shakllarning adabiy-me‘yoriy variantlari ishlatilsa, badiiy uslubda bu grammatik shakllarning barcha ko‘rinishlari – lahja va tarixiy variantlari ham asarning mavzu talabi bilan qo‘llanilaveradi.

So‘zlashuv uslubi kishilarning o‘zaro erkin muomalalarini yuzaga keltirishda xizmat qiladi. Bugungi kunda so‘zlashuv uslubi nafaqat og‘zaki, balki yozma shaklda ham amalda. Buni, asosan, global tarmoqda yoshlar o‘rtasidagi yozishmalarda ko‘rishimiz mumkin. Vaqtni tejash maqsadida *w-sh*, *6-o‘*, *4-ch*, *u-o‘* harflari almashib *yaxwi* (yaxshi), *ki4ik* (kichik), *yow* (yosh), *yu* yoki, *yo* (yo‘q), *nimidi* (nima edi), *k64a* (ko‘cha) kabi leksemalarni yozishda qo‘llanilishi so‘zlashuv uslubida yozilgan matnlarni avtomatik tahrirlashda butunlay chetlab o‘tishni taqozo etadi. Bu uslubning o‘ziga xos xususiyati shundaki, u xoh og‘zaki, xoh yozma shaklda bo‘lsin, nutq jarayonida til va tildan tashqari omillar faol qo‘llaniladi, ayniqsa, sheva elementlari yetakchilik qiladi, varvarizmlarga keng o‘rin ajratiladi. Ayni paytda, bu jihat so‘zlashuv uslubidagi me‘yoriy holatlarni belgilashni qiyinlashtiradi.

E‘tiborli tomoni shundaki, nutqda tovushlarning uyg‘unlashuvi (*ketti*, *ottan*, *yigichcha*), bir tovush o‘rnida ikkinchisining talaffuz qilinishi (*traktir*, *zaril*, *direktr*, *kampitr*), tovushlarning o‘rin almashishi (*bunni*, *turpoq*, *aynalmoq*), tovushlarning orttirilishi (*o‘ramol*, *fikir*, *banka*, *ustol*), tushirib qoldirilishi (*gazet*) kabi fonetik hodisalar avvalo so‘zlashuv uslubida namoyon bo‘ladi.

Ko‘rib o‘tganimizdek, matn uslubiy mohiyatiga ko‘ra ilmiy matn (tezis, maqola, ma‘ruza, annotatsiya, taqrizlar), badiiy matn (nasriy va nazmiy asarlar), rasmiy matn (ma‘lumotnoma, qaror, buyruq, farmon, farmoyish, tavsifnoma, tavsiyanomalar, rasmiy nutq matni), publitsistik matn (kundalik nashrlardagi maqolalar), so‘zlashuv matni (internet tarmoqlaridagi yozma muloqot) kabi tarmoqlarga bo‘linadi.

Zamonaviy kompyuter texnologiyalaridan unumli foydalanish ko‘lami barcha sohada oshmoqda. Hatto texnologiyasiz ish faoliyatini yuritish ham qiyin kechadi. O‘zbek tilidagi matnlarga avtomatik ishlov beruvchi DTning mukammal bo‘lishini ta‘minlash, uning aniqligi va sifat ko‘rsatkichini oshirish maqsadida badiiy va publitsistik uslublarga xos matnlarda arxaizm, istorizm va dialektizmlarning qo‘llanishi me‘riy hol hisoblanishi bois, ilk qadamda, ilmiy va rasmiy matnlar tadqiqot obyekt sifatida olindi. Matnlarni tahrir va tahlil qilish dasturining yaratilishi matn

teruvchi yozma nutq malakasining takomillashuvi, matnlarni mustaqil ravishda to‘g‘ri va savodli yozishida ko‘mak beradi. Shu bois lingvistik ta‘minotni mukammallashtirish jarayonida keyingi bosqichlarda badiiy uslub qamrovi ham e‘tiborga olingan.

1.6. Matnlarni avtomatik tahrir qilishdagi muammolar

ATT dasturining lingvistik bazasini yaratishda o‘zbek tili doirasida muammolar ko‘zga tashlanadi. O‘zbek tili affikslarining ko‘p variantliligi tahrirlash dasturining lingvistik ma‘lumotlar ta‘minotini yaratish, lisoniy-axborot bazasini kengaytirishda birmuncha noqulayliklarni keltirib chiqaradi. Bu noqulayliklar quyidagilarda ko‘rinadi:

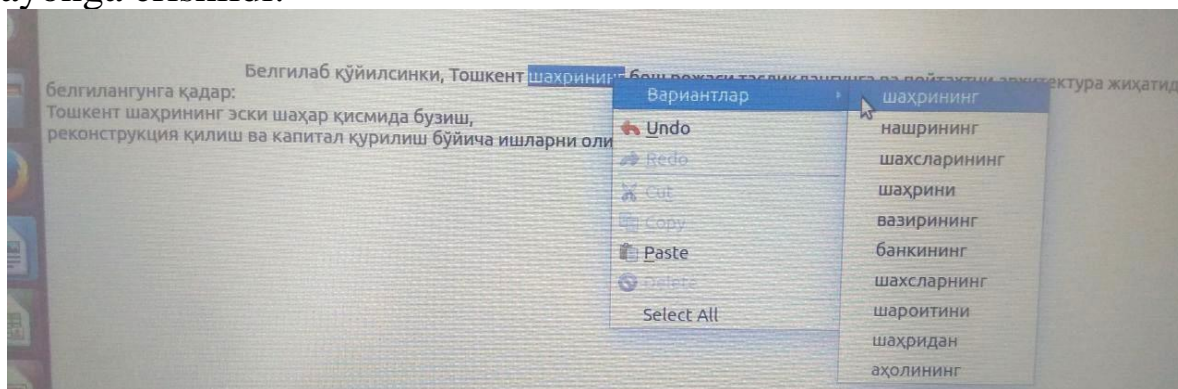
1. Bir so‘zning uslublar doirasida o‘z qo‘llanish variantiga egaligi va ushbu uslublarning me‘yorlarini hisobga olib aytish mumkinki, so‘zning hamma uslubdagi ko‘rinishi hamda affikslar bilan kelish shaklini yaratish murakkab vazifa. Shu sababli ATT dasturining LTini yaratishda ilmiy va rasmiy uslublarda qo‘llaniladigan so‘z va affikslarning olinishi ishning samaradorligini oshirishga xizmat qiladi. Masalan, *samo*, *osmon*, *ko‘k*, *falak* sinonimik qatoridagi *samo* va *osmon* so‘zlari ot turkumi, *ko‘k* so‘zi esa faqat rang bildiruvchi so‘z sifatida olinib affikslar biriktirilsa, lisoniy bazaning kengayishi va ish hajmi ortishining oldi olingan bo‘ladi. *Falak* so‘zi esa badiiy uslubga mansub bo‘lganligi uchun alohida guruh tarkibiga kiritiladi (lug‘at boyligi sifatida) va keyingi bosqichlarda dasturni mukammallashtirish maqsadida bunday guruhlar ham ishlab chiqiladi.

2. Affikslarning ko‘p variantliligi, omonimlik, bir tomondan, axborot bazasini kengaytiradi, ikkinchi tomondan, dasturlash jarayonida murakkab nostandart dasturlar ishlab chiqilishiga olib keladi va har bir uslub, har bir yangi matnga alohida dasturlar yaratishni talab qiladi. Bu holat mutaxassislardan ko‘p mehnat, katta shijoatni taqozo etadi. Masalan, *-gina* “iliq” munosabatni ifodalovchi affiks va ayiruv yuklamasi vazifasida keladi (*onaginam* – *onamgina*). *-la* affiksi ayni damda so‘z yasovchi va shakl yasovchi affiks hisoblanadi: *ishla* – *quvla*. Dastur o‘z ta‘minotiga kiritilgan ma‘lumotga binoan buni foydalanuvchiga ko‘rsata oladi. Ammo affikslarning nomi va vazifasini ta‘minotga kiritib qo‘yishning o‘zi yetarli emas. Asosiy ish affikslar kombinatsiyalarini har bir so‘zga qo‘shilib kelish holatida modellashtirish, ularning aranjirovkasini berish hisoblanadi.

3. O‘zbek adabiy tilida so‘zga affiks qo‘shilishi natijasida tovush tushishi (*shahar* – *shahri*), tovush almashinishi (*quloq* – *qulog‘i*), tovush orttirilishi (*u* – *unga*) hodisalarini kuzatishimiz mumkin. Shu bois

algoritmlar ishlab chiqishda o‘zakda ro‘y bergan o‘zgarishlar hisobga olinib, hosil qilinadigan so‘zlarni “avtomatik hosila” sifatida dastur ta’minotiga kiritish lozim, deb topildi. “Avtomatik hosila” bu so‘zlarga affikslar qo‘shish jarayonida hosila so‘zda ro‘y beradigan tovush o‘zgarishlarining lingvistik ta’minotda aynan saqlanishidir. Masalan, *singil*, *ko‘ngil*, *qorin* so‘zlariga egalik kategoriyasining I shaxs-birlik affiksi qo‘shilishi natijasida quyidagi o‘zgarishlar ro‘y beradi: {*singil*, *ko‘ngil*, *qorin* } + -im = {*singlim*, *ko‘nglim*, *qornim* }. *Quloq*, *qiroq*, *bo‘yoq*, so‘zlariga egalik kategoriyasining II shaxs-birlik affiksi qo‘shilganda so‘z oxiridagi *q* tovushi *g*‘ tovushiga almashadi: {*Quloq*, *qiroq*, *bo‘yoq*, + -ing = {*qulog‘ing*, *qirog‘ing*, *bo‘yog‘ing*,}. Avtomatik tahrir dasturining lingvistik ta’minotiga hosil qilingan *singlim*, *ko‘nglim*, *qornim*, *qulog‘ing*, *qirog‘ing*, *bo‘yog‘ing* so‘zlari avtomatik hosila (tayyor o‘zak) sifatida kiritiladi, so‘ngra ularga affikslar kombinatsiyasining birikish algoritmi modullar orqali tuziladi. Dasturni yaratishda o‘zbek tilining mana shunday orfografik me’yorlariga e’tibor qaratiladi va shunga binoan dastur algoritmi tuziladi.

Biz tomondan yaratilgan dastur Ubuntu 16.4 (Linux) operatsion sistemasi Python 3 dasturlash tilida (Uztextanalysis) bo‘lib, unda quyidagi jarayonga erishildi:



6-rasm. Mazkur tasvarda tadqiqotimiz natijasida *shaxrining* so‘zshakli dastur tomonidan tahlil qilinib, shu so‘z bilan bog‘liq to‘g‘ri variantlarni taklif qilishiga erishilgani ko‘rinadi.

O‘zbek tilida yozilayotgan matnlar, turli sohalarga oid hujjatlarni avtomatik ravishda tahrir qilishga erishish maqsadida yuqorida ilk qadamda e’tiborga olinishi kerak bo‘lgan muammolar morfologik tahlil doirasida ko‘rib chiqildi va ATT dasturining lingvistik ta’minotini yaratish borasida muhim xulosalar chiqarildi.

BIRINCHI BOB BO‘YICHA XULOSA

O‘zbek tilidagi matnlarni Microsoft Office Word dasturiga kiritish jarayonida matnni avtomatik tahrir qilishga erishish, buning uchun dasturning lingvistik ta‘minini yaratish maqsadida bugungi kunda keng qo‘llanishda bo‘lgan dasturlar chuqur o‘rganildi. Ularning amaliy ahamiyati qisqa sharh tarzida berildi. Vazifalari, ish faoliyati chuqur tadqiq etildi va matnlarga avtomatik ishlov beruvchi dastur ta‘minotini yaratishda quyidagi lingvistik imkoniyatlarning muhim ekanligi aniqlandi:

1. Matnlarning leksik, morfologik va sintaktik sathlarda tahrir va tahlilini amalga oshiruvchi spell-checker, orfografik tekshiruv (Speller), morfoanalizator, parser kabi sistemalarga o‘zbek tilidagi matnlarni tahrir va tahlil qilish dasturini yaratishda asosiy dasturiy vosita sifatida tayanish ish samaradorligini oshiradi.

2. Tokenizatsiya, lemmatizatsiya (lemmatization), stemming (stemming) texnologiyalari matnlarni tahrir va tahlil qilish dasturida eng sermahsul jarayon hisoblanadi.

4. Har qanday dasturning lingvistik bazasini yaratishda kerakli manba – “O‘zbek tilining morfologik lug‘ati”dir. Morfologik lug‘at tilshunosning vaqtini tejaydi va belgilangan natijaga erishishni tezlashtiradi. Zero, bunday lug‘atda bir leksemaning qaysi turkumga oidligi, uning so‘zshakllari keltiriladi. Bu esa dasturning mukammal chiqishini ta‘minlaydi. Shu bois o‘zbek tilidagi matnlarga ishlov beruvchi dasturlarning lingvistik ta‘minotini yaratishda qulaylik va natijaga tezkor erishish maqsadida ushbu lug‘atning ishlab chiqilishi maqsadga muvofiqdir.

5. O‘zbek tilining tipik agglyutinativ til ekanligi va binobarin, tilimizda affikslarning ko‘p variantlilik, omonimlik, bir tomondan, axborot bazasini kengaytiradi, ikkinchi tomondan, dasturlash jarayonida murakkab nostandart dasturlar ishlab chiqilishiga olib keladi va har bir uslub, har bir yangi matnga alohida dasturlar yaratishni taqozo etadi. Bu holat mutaxassislardan ko‘p mehnat, katta shijoatni talab qiladi. Masalan, -gina affiksi “iliq” munosabatni ifodalovchi affiks va ayiruv yuklamasi vazifasida keladi (onaginam-onamgina). -la affiksi ayni damda so‘z yasovchi va shakl yasovchi affiks hisoblanadi: *ishla-quvla*. Dastur o‘z ta‘minotiga kiritilgan ma‘lumotga binoan buni foydalanuvchiga ko‘rsata oladi. Ammo affikslarning nomi va vazifasini ta‘minotga kiritib qo‘yishning o‘zi yetarli emas. Asosiy ish affikslar kombinatsiyalarini har bir so‘zga qo‘shilib kelish holatida modellashtirish hisoblanadi.

3. O‘zbek adabiy tilining grammatik qoidalarini o‘zida jamlagan “O‘zbek tilining qisqartirilgan grammatikasi”ni yaratish darkor. Bunday

grammatika lingvistik bazani tuzishda muhim manba bo‘lib xizmat qiladi. Chunki tayyor manbaning mavjud bo‘lishi tilshunos yoki dasturchini o‘nlab grammatika ustida vaqt sarflashidan saqlaydi.

4. O‘zbek adabiy tiliga mansub leksemalarni o‘zida jamlagan “O‘zbek tilining tezaurus lug‘ati”ni yaratish lozim. Shunda avtomatik tahlil dasturining sintaktik va semantik tahlil bosqichlari muammolarining yechimi hal etiladi.

5. O‘zbek tilining milliy korpusini yaratish talab qilinadi. Bu tarkibiy sohalar bo‘yicha guruhlangan matnlar jamlanmasi bo‘lib, lingvistik modullarni yaratish borasida tadqiqotlar olib borishda qo‘l keladi.

6. Dasturning ilk bora joriylanayotgani e‘tiborga olinib, ishning avvalida rasmiy-idoraviy va ilmiy uslublardagi matnlar tahriri va tahliliga mo‘ljallangan lingvistik ta‘minot yaratish maqsadga muvofiq sanaladi. Bir so‘zning uslublar doirasida o‘z qo‘llanish variantiga egaligi va ushbu uslublarning me‘yorlarini hisobga olib aytish mumkinki, so‘zning hamma uslubdagi ko‘rinishi hamda affikslar bilan kelish shaklini yaratish murakkab vazifa. Shu bois ATT dasturining LTini yaratishda ilmiy va rasmiy uslublarda qo‘llaniladigan so‘z va affikslarning olinishi ishning samaradorligini oshirishga xizmat qiladi. Masalan, *samo*, *osmon*, *ko‘k*, *falak* sinonimik qatoridagi *samo* va *osmon* so‘zlari ot turkumi, *ko‘k* so‘zi esa faqat rang bildiruvchi so‘z sifatida olinib affikslar birlashtirilsa, lisoniy bazaning kengayishi va ish hajmi ortishining oldi olingan bo‘ladi. Falak so‘zi esa badiiy uslubga mansub bo‘lganligi uchun alohida guruh tarkibiga kiritiladi (lug‘at boyligi sifatida) va keyingi bosqichlarda dasturni mukammallashtirish maqsadida bunday guruhlar ham ishlab chiqiladi.

7. O‘zbek adabiy tilida so‘zga affiks qo‘shilishi natijasida tovush tushishi (*shahar* – *shahri*), tovush almashinishi (*quloq* – *qulog‘i*), tovush orttirilishi (*u* – *unga*) hodisalarini kuzatishimiz mumkin. Shu bois algoritmlar ishlab chiqishda o‘zakda ro‘y bergan o‘zgarishlar hisobga olinib, hosil qilinadigan so‘zlarni “avtomatik hosila” sifatida dastur ta‘minotiga kiritish lozim, deb topildi. “Avtomatik hosila” bu so‘zlarga affikslarni qo‘shish jarayonida hosila so‘zda ro‘y beradigan tovush o‘zgarishlarining lingvistik ta‘minotda aynan saqlanishidir. Tegishli morfemik o‘zgarish avtomatik hosila (tayyor o‘zak) sifatida kiritiladi, so‘ngra ularga affikslar kombinatsiyasining birikish algoritmi modellar orqali tuziladi. Dasturni yaratishda o‘zbek tilining mana shunday orfografik me‘yorlariga e‘tibor qaratiladi va shunga binoan dastur algoritmi tuziladi.

II BOB. MATNNI AVTOMATIK TAHRIR VA TAHLIL QILISH SISTEMASI MODULLARI

2.1. Matnlarni avtomatik tahlil qilish borasida amalga oshirilayotgan tadqiqotlar yo‘nalishi va sistema bosqichlari

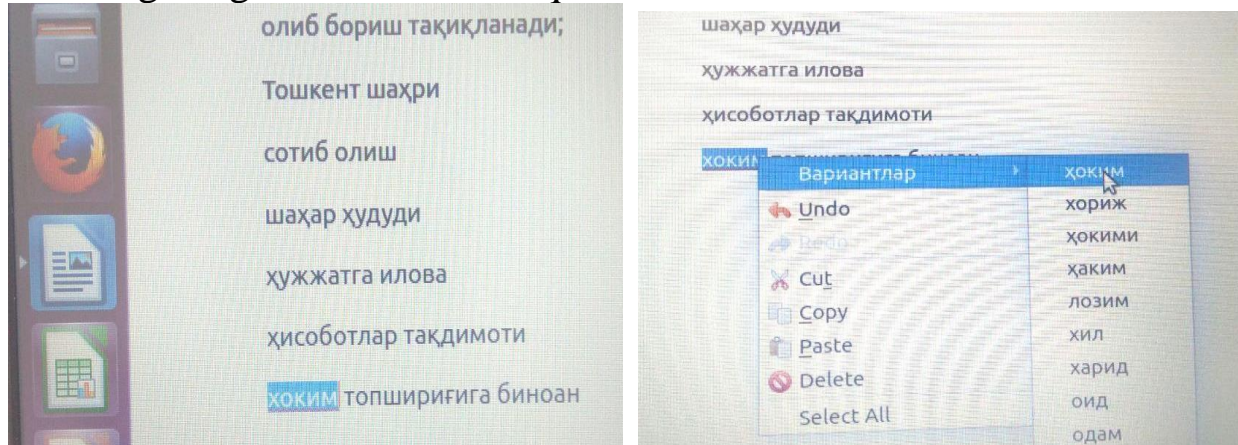
Kundalik turmushning ajralmas qismiga aylangan kompyuter texnologiyalarida matn asosiy o‘rinni egallaydi. Shu bois amaliy tilshunoslik vakillari dastavval matnni avtomatik ravishda bir tildan boshqa bir yoki bir necha tabiiy tilga tarjima qiluvchi, matnni morfologik, sintaktik va semantik jihatdan avtomatik tahrir qiluvchi, matndagi so‘zlarning chastotasini ko‘rsatuvchi va boshqa shu kabi dasturlarni yaratish ustida ish olib borishdi. Tabiiy tillardagi matnlarning kompyuter tahriri borasida xorijdagi tilshunoslar tomonidan keng ko‘lamli ishlar amalga oshirilmoqda, katta hajmli hujjatlar ham matematik metodlar asosida tahrir qilinmoqda. Shu bilan bir qatorda, qidiruv tarmog‘ida kiritmalarni to‘g‘ridan-to‘g‘ri tahlil va tahrir qilinishiga erishilmoqda. Bugungi kunda matnni texnik tahlil qilish borasida yaratilayotgan ishlar asosan quyidagi yo‘nalishda olib borilmoqda:

- 1. Kompyuterli grammatika⁹³
 - Kompyuterli morfologiya
 - Kompyuterli sintaksis
- 2. Kompyuterli semantika

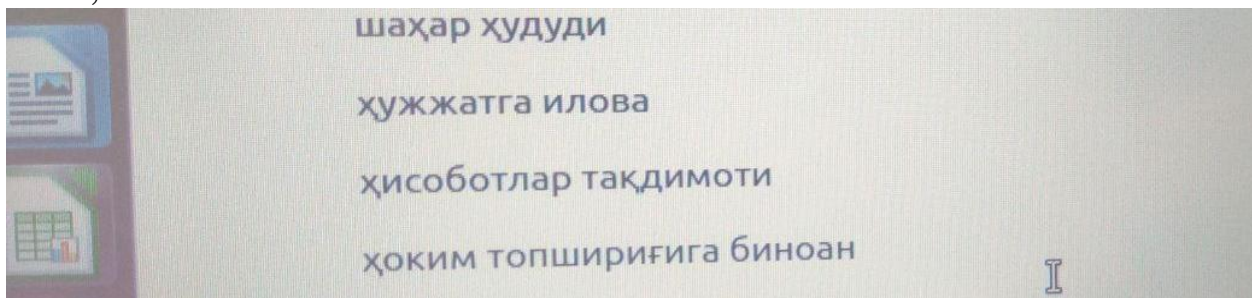
Matnlarga ishlov berish yo‘nalishida ATT muhim o‘rin tutadi. Kompyuter vositasida amalga oshiriladigan bunday tahlil usuli inson tomonidan qilinadigan lingvistik tahrir va tahlildan vaqt hamda moddiy mablag‘ning tejalishi bilan ajralib turadi. Shuningdek, ATT ish jarayonida, ya‘ni matnni kompyuterga kiritish paytida foydalanuvchining savodini oshirishga ham xizmat qiladi. Matnni kirituvchi tomonidan yo‘l qo‘yilgan orfografik, sintaktik yoxud semantik xatoni tahlil qilish dasturi o‘sha onda aniqlaydi va to‘g‘ri yozish variantlarini havola qiladi. Deylik, *hokim* so‘zi kiritilishi kerak, ammo dastur foydalanuvchisi so‘zni *xokim* tarzida yozdi,

⁹³ X.S.Yasulova “Математические модели синтаксиса” (2011). <https://cyberleninka.ru/article/v/matematicheskie-modeli-sintaksisa>) maqolasida KL an’anaviy tilshunoslikdagi terminlarning kompyuter bilan aloqador holda boyitilganini e’tirof etadi va hatto KLdagi barcha terminlar oldidan “kompyuterli/kompyuterbop” so‘zini qo‘yish joiz, deydi. Boisi KL fanining asosiy maqsadi lingvistik masalalarni hal qilishning kompyuter dasturlarini yaratishdir. Mazkur tadqiqotda “компьютерная морфология” birikmali termini (BT)ni o‘zbek tiliga “kompyuterli morfologiya”, “компьютерный синтаксис” BTni “kompyuterli sistaksis”, “компьютерная семантика” BTni “kompyuterli semantika” tarzida kalkalash maqsadga muvofiq hisoblandi.

shunda dastur o‘ziga joylangan lingvistik modullarda so‘zni tahlil qilib, uning xato yozilganini aniqlaydi va foydalanuvchiga *hokim*, *hakim* variantlarini taqdim etadi. Natijada foydalanuvchi o‘zbek imlo lug‘atida *xokim* tub so‘zi mavjud emasligini anglaydi. Fikrimizga Uztextanalysis dasturidagi fragmentlar bilan aniqlik kiritamiz:



- 1) “xokim” tarzida yozilgan so‘z tagiga qizil to‘lqinli chiziq chizildi;
- 2) tagiga chizilgan so‘z belgilandi, sichqonchanning o‘ng tomoni bosilib xato so‘zga taklif qilinayotgan so‘zshakllardan kerakli so‘z tanlab olindi;



- 3) xato bartaraf etildi.

E’tiborli jihati shundaki, dastur tomonidan foydalanuvchiga taklif qilinayotgan so‘zlar qatorida *xokim* leksemasi bilan bog‘liq so‘zshakllar uchramadi. Bu jihat foydalanuvchiga mazkur so‘zning imlo lug‘atida mavjud emasligi va keyingi holatlarda shu so‘z bilan bog‘liq xatoga yo‘l qo‘ymaslik kerakligi borasida xulosani beradi. Bu dasturning bevosita ta’limiy ahamiyatini namoyon etadi.

Tahrir dasturi nafaqat so‘zlarning to‘g‘ri yozilayotganini tekshiradi, balki belgilarning o‘z o‘rnida qo‘llanilayotganini ham nazoratga oladi. Deylik, «Siyosiy fanlar» fani tarmog‘i ixtisosliklari bo‘yicha tayanch doktoranturada 4 kishi, va doktoranturada 2 kishi miqdorida belgilash to‘g‘risidagi takliflari ma‘qullansin⁹⁴. jumlasini grammatik jihatdan to‘g‘ri

⁹⁴ O‘zbekiston Respublikasi Prezidentining 2019-yil 29-yanvardagi «Siyosiy fanlar sohasida kadrlarni tayyorlash, fundamental va amaliy tadqiqotlar samaradorligini oshirish chora-tadbirlari to‘g‘risida»gi PQ-4139-son Qarori. <http://www.lex.uz/docs/4182353>.

tuzilgan, lekin vergul belgisi *kishi* soʻzidan soʻng bir boʻsh joy – probel tashlab yozilgan (grafematik xato), vergul hamda *va* bogʻlovchisi bir oʻrinda notoʻgʻri qoʻllanilgan. Lingvistik tahrir dasturi ushbu kichik xatolarni inobatga olib tuziladi.

Lingvistik protsessor koʻp bosqichli boʻlib, har bir bosqichga chuqur ishlov berish talab qilinadi. Xususan, ularga alohida modullar ishlab chiqilishi kerak. Matn tahlili jarayonida modullar quyidagilarni amalga oshiradi:

1) leksik-morfologik – soʻzshakldan uning leksik-grammatik xususiyatlariga oʻtish;

2) sintaktik jummalarni ifodalovchi leksik va grammatik xususiyatlar zanjiridan uning sintaktik tuzilishiga oʻtish;

3) semantik – sintaktik tahlil qilingan lingvistik birlikdan semantik ishlovga oʻtish.

Umuman, matnlarni tahrir va tahlil qiluvchi dasturning mukammal boʻlishi uchun quyidagi lingvistik modullar ishlab chiqilishi lozim⁹⁵:

1. Grafematik tahrir modullari.

2. Morfologik tahlil modullari.

3. Sintaktik tahlil modullari.

4. Semantik tahlil modullari.

Grafematik tahrir – matnda qoʻllangan turli ramzlar, belgilar va alohida qoʻllangan harflarni tahrir qilish bosqichi⁹⁶.

Morfologik tahlil modullari – soʻzshaklidan uning lemma (leksemaning lugʻatdagi shakli) yoki asosi (soʻzning oʻzak, yadro qismi)ga qadar tahlil qilinadi.

Sintaktik tahlil – matndagi jumlaning grammatik tarkibini aniqlash.

Semantik tahlil – leksemani semasi boʻyicha matnda qoʻllanilish oʻrnini tahlil qilish.

2.2. Grafematik tahrirlash moduli va uning texnologiyasi

Grafematik tahlil (GT) matnni lingvistik jihatdan avtomatik tekshirishning ilk bosqichi hisoblanadi, shu bois vazifa nuqtayi nazaridan ishda mazkur modul tahrir bosqichi, deb hisoblandi. Ushbu bosqichda matndagi abzats, soʻz, gap, raqam, punktuatsion va boshqa maxsus ramziy belgilar aniqlanadi. *Grafema* – yozma matn birligi (harf, tinish belgilar),

⁹⁵ Большакова Е.И. и др. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика. – Москва, 2011. – С.106.; Леонтьева Н.Н. Автоматическое понимание текстов: системы, модели, ресурсы. – Москва, 2006. – С.50.

⁹⁶ Леонтьева Н.Н. Автоматическое понимание текстов: системы, модели, ресурсы. – Москва, 2006. – С. 50.

ingliz tilida *token* deyiladi, *tokenization* soʻzi shundan kelib chiqqan. GT bosqichining maqsadi matndagi eng kichik birliklarni aniqlash va tasniflashdir. Bunday birliklarga soʻz, abzats, tinish belgilari, sanalar, pul birliklari ramzlari, soʻz-raqamli birikmalar, sonlar, IP-manzillar va fayl nomlari, telefon raqamlari oid. I.M.Nojov oʻz tadqiqotida GT birligi har ikki tomondan boʻshliq (probel) bilan ajralib turadigan belgilar majmuasi, deb taʼkidlaydi⁹⁷.

GT jarayonida quyidagi vazifalar bajariladi:

1) matn oqimini soʻzlarga segmentlab, ularni tasniflash. Bu birliklarga jarayon davomida keyingi tahrir uchun izoh beriladi. Masalan, AA – soʻz, faqat bosh harflardan iborat. Aa – soʻz, bosh harf bilan boshlanadi kabi. Buning uchun alifbodagi bosh va kichik harflarning barchasi dastur moduliga kiritiladi (ilova: 1-jadval);

2) matndan olingan soʻzlarni chastota boʻyicha tartibga solish;

3) soʻzlarning ayrim kombinatsiyalarini yirikroq birliklar – “barqaror iboralar”ga birlashtirish (ibora boshi va oxiriga belgi qoʻyadi: BI1... BI2);

4) soʻzlar (yashirin, noaniq)ga maxsus belgi berish;

5) abzatsni aniqlash va raqamlash;

6) qisqartma soʻzlar (abbreviaturalar)ni aniqlash.

Ushbu vazifalarni amalga oshirish uchun tilshunoslik metodlariga tayangan holda formal yondashuvni ishlab chiqish talab qilinadi.

Grafematik tahrirlash bosqichida matnda qoʻllanilgan elementlar oqimini soʻzshakl va turli ramzlarga ajratish, matnni morfologik tahlil jarayoniga tayyorlash, matnni korpusga kiritish maqsadida *tokenizatsiya* texnologiyasidan foydalanish maqbul hisoblanadi. Boisi tokenizatsiya elektron matndagi barcha birliklarni ajratish jarayonidir. Buni quyidagi *Oʻzbekiston Respublikasi Prezidentining 2005-yil 15-iyundagi «Sharq taronalari» xalqaro musiqa festivaliga tayyorgarlik koʻrish va uni oʻtkazish toʻgʻrisida»gi PQ-103-sonli qarori oʻz kuchini yoʻqotgan deb hisoblansin* jumlasini segmentlash, yaʼni tokenlarga ajratish bilan koʻrsatish mumkin:

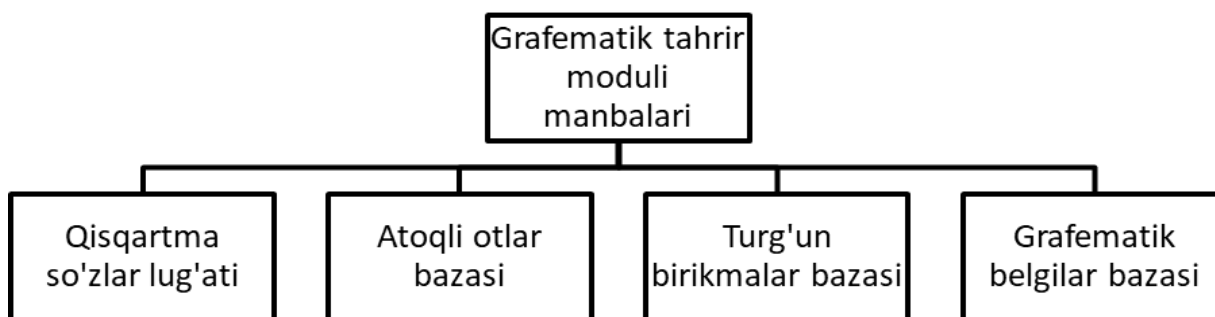
[‘Oʻzbekiston’, ‘Respublikasi’, ‘Prezidentining’, ‘2005’, ‘-’, ‘yil’, ‘15’, ‘-’, ‘iyundagi’, ‘«’, ‘Sharq’, ‘taronalari’, ‘»’, ‘xalqaro’, ‘musiqa’, ‘festivaliga’, ‘tayyorgarlik’, ‘koʻrish’, ‘va’, ‘uni’, ‘oʻtkazish’, ‘toʻgʻrisida’, ‘»’, ‘gi’, ‘PQ’, ‘-’, ‘103’, ‘-’, ‘sonli’, ‘qarori’, ‘oʻz’, ‘kuchini’, ‘yoʻqotgan’, ‘deb’, ‘hisoblansin’, ‘.’]

⁹⁷ Ножов И.М. Морфологическая и синтаксическая обработка текста (модели и программы): Дис. канд. филол.наук. – Москва, 2003. – С. 59.

Bu yerda kvadrat qavs matnning boshlanish va yakunlanish joyini ko'rsatadi. Birtirnoq elementni ajratuvchi belgi, vergul ajratilgan belgilarni bir-biridan alohidalashga xizmat qiladi. Mazkur misolda tokenizatsiya jarayonida 'gi' bo'g'ini (tokenizatsiyada "qism" deyiladi) ma'noga ega bo'lmagani bois kesib tashlanadi. Mazkur jarayondan so'ng elementlarni teglashtirish, ularni hisoblash, guruhlashtirish, ortiqcha qo'llangan belgi yoki ma'noga ega bo'lmagan so'zlarni o'chirish mumkin.

Boshqird davlat universiteti aspiranti V.A.Grechachin matnni tokenizatsiyalash jarayonining algoritmini Python tilida tuzadi⁹⁸. Tokenizatsiya texnologiyasi Daniel Jurafskiy hamda Jeyms Martinlarning «Speech and Language Processing»⁹⁹ kitobida amaliy dasturlar misolida yoritib berilgan.

Grafematik tahrir manbalari sifatida quyidagilar xizmat qiladi:



7-chizma. Grafematik tahrir manbalari.

Demak, GT modullari asosini quyidagi belgilar va birikmalar qatori tashkil etadi:

1. Lotin, kirill va ingliz alifbolaridagi bosh va kichik harflar (ilovada 1-jadval).

2. 0 dan 9 gacha bo'lgan sonlar va rim raqamlari (2-jadval). Qolgan murakkab sonlarning ushbu o'nta son orqali yuzaga keltirilishi dasturga algoritm qilib kiritiladi.

3. Tinish belgilari: nuqta, vergul, qo'shtirnoq, qavs, tire, ikki nuqta, nuqtali vergul, ko'p nuqta, so'roq va undov (3-jadval).

4. Matematik ifodalar: qo'shish (+), ayirish (-), bo'lish (\div ; :), ko'paytirish (\times ; ·), qavs va uning ko'rinishlari (), [], cheksizlik belgisi (∞), darajaga ko'tarish (x^2), ildiz osti ($\sqrt{\quad}$), tenglik (=) va teng emaslik (\neq) belgilari, katta (>) va kichik (<) ishoralari kabilar. Algoritmga ularning faqat sonlar orasida kelish ma'lumoti beriladi (3-jadval).

⁹⁸ Гречачин В. А. К вопросу о токенизации текста. <https://research-journal.org/languages/k-voprosu-o-tokenizacii-teksta/>

⁹⁹ Jurafsky, D. Speech and Language Processing / D. Jurafsky, J. H. Martin. – 2nd – New Jersey: Prentice Hall, 2008. – 1024 p.

5. Tire (–) va chiziqcha (-)ning alohida belgilar ekanligi yoziladi.

6. Tutuq belgisining vazifalari ko‘rsatiladi.

7. Diakritik belgilar. Matn lotin yozuvida yozilayotganda dastur algoritmi o‘ va g‘ harflari ustidagi «‘» diakritik belgisi «’» tutuq belgisiga almashtirib qo‘yilganda, bunday holatni xato, deb hisoblaydi va matn kirituvchini ushbu xato haqida ogohlantiradi. Mazkur jihatlar ATT dasturi lotin yozuvida kiritilayotgan matnlar tahririni amalga oshirishga mo‘ljallangan vaziyatlarda zarur hisoblanadi.

8. Havola belgilari: raqamli (2,3…) va yulduzchali (*) qo‘rinishlari.

9. Maxsus belgilar: ^, v, %, &, \, { } _ , № kabilar.

10. Atoqli o‘rin-joy otlari sifatida qit‘a, o‘lka, davlat, viloyat, shahar, tuman, qishloq, mahalla, posyolka, daha, ko‘cha nomlari (toponimlar);

11. Kishi ismlari, laqablar, taxallus (antroponim)lar, yulduz va sayyoralar nomlari (astroponimlar), hayvonlarga berilgan maxsus nom (zoonim)lar kabi atoqli otlarni kiritish darkor.

12. Qisqartma so‘zlar: *NDKI, O‘zMU, YUNESKO, SamDU, TDO‘TAU* kabi.

Qoida: qisqartma so‘zlarning yozilishi:

1) faqat bosh harf bilan yoziladi, harflardan so‘ng tinish belgi (nuqta), probel qo‘yilmaydi: *TDO‘TAU, BMT, AQSH, MDH, NATO* va h.k.;

2) birinchi so‘zning dastlabki bo‘g‘ini, keyingi so‘zlarning bosh harfi olinadi: *O‘zMU, SamDU, BuxDU, O‘zFA* va h.k.;

3) birinchi so‘zning birinchi, ikkinchi bo‘g‘ini va keyingi so‘zning to‘liq shakli olinadi: *santexnik, agitpoyezd*;

4) so‘zlarning dastlabki bo‘g‘inlari olinadi: *filfak, ijroqo‘m*.

Matndagi belgilarning asosiy qismini tabiiy til alifbosidagi harflar tashkil etadi. So‘zlarni tashkil qiluvchi harflar ma’lum bir yozuv (kirill, lotin va h.k.)da kiritilayotganda so‘z tarkibida o‘zga yozuv elementi yozilib ketishi mumkin. Masalan, *maktab, makr, g‘arb, aksiya* kabi. GT bunday xatolarni aniqlaydi va o‘zidan keyingi (morfologik) bosqichga o‘tkazadi. Ayrim hollarda matnlarda maxsus belgilarni uchratish mumkin. Masalan, haroratni ko‘rsatuvchi belgi, ©, *, &, №, &, <, > kabi belgilar. Bunday belgilar tahrir dasturining lingvistik ma’lumotlar omboriga “maxsus belgilar – MaxB” nomi ostida har birining vazifasi yozilgan holda kiritiladi. Ular qatoriga bo‘sh joy ajratuvchi probel va xatboshi, satr oxirini ko‘rsatuvchi belgilar ham kiradi. Masalan, (“leksema”), – tarzida ketma-ket kelgan belgilar tizimi quyidagicha izohlanadi:

- (– ochuvchi qavs;
- “ – ochuvchi qo‘shirnoq;
- leksema – lug‘aviy birlik;
- ” – yopuvchi qo‘shirnoq;
-) – yopuvchi qavs;
- , – vergul;
- – tire.

Bunday izohlar sintaktik yoki semantik tahlilda belgilarning o‘rni hamda ahamiyati uchun muhimdir. Bundan tashqari, yozuvda shunday holatlar yuzaga chiqadiki, ularni e‘tibordan chetda qoldirib bo‘lmaydi. Masalan, so‘zlardagi harflar bir bo‘shliq – probel qo‘yib yozilgan bo‘lishi mumkin (*A R I Z A*). Agar bunday holatlar yuzasidan grafematik tahrir modullariga ma‘lumot kiritilmasa, dastur *A R I Z A* so‘zidagi harflarni bir so‘zni tashkil etuvchilar deb emas, balki alohida qo‘llangan harflar ketma-ketligi deb hisoblaydi. Shuningdek, kasr sonlarni ifodalashda qo‘llaniladigan nuqta yoki vergul raqamlarni ajratuvchi belgilar sifatida emas, balki raqamlarning bir butunlikda qabul qilinishini bildiradi. Masalan, *1.25* va *5,5* raqamlari bir butun sonlar.

Lotin grafikasida bitilgan o‘zbekcha matnlarni kompyuter vositasida yozish jarayonida diakritik belgilarning turli holatda qo‘yilganiga guvoh bo‘lamiz.

Noto‘g‘ri	To‘g‘ri
O`o` G`g`	O‘ o‘
O'o' G'g'	G‘ g‘
O'o' G'g'	
fe`l	fe’l
a`lo	a’lo

Bunday xatoga ommaviy axborot vositalari va reklama matnlarida yo‘l qo‘yilayotgani achinarli hol. Chunki har bir belgi o‘z o‘rni hamda vazifasiga ega. Lotin alifbosidagi o‘ va g‘ harflarini ifodalashga xizmat qiladigan diakritik belgini to‘g‘ri qo‘yish qoidasi (tutuq belgisi (‘)ning yuqoriga qaragan teskari holati(‘)) “O‘zbek tilining asosiy imlo qoidalari”da yozib qo‘yilgan¹⁰⁰. Tutuq belgisining vazifalaridan (tutuq belgisi unli tovushdan keyin qo‘yilsa, uning cho‘zib o‘qilishini ta‘minlaydi: *she’r*, *ma’no*; undosh tovushdan so‘ng qo‘yilsa, o‘zidan keyin

¹⁰⁰ O‘zbek tilining asosiy imlo qoidalari. 1995-yil 24-avgust, 339-son / <http://lex.uz/docs/1625271>

kelgan unidan ajratib o‘qilishga xizmat qiladi: *sur’at, qit’a*) anglash mumkinki, diakritik belgi bilan tutuq belgisining almashtirib qo‘llanilishi aslo mumkin emas.

Lingvistik tahrir dasturining GT bosqichi modullarini yaratishda yuqoridagi holatlarning har biri matn elementlari tiplari tarzida yuzaga chiqariladi¹⁰¹. Quyida tiplarning qanday birliklarni o‘z tarkibiga olishi keltirildi:

LekB –leksik birlik, bir yozuvga mansub alifbodagi harflardan yuzaga kelgan leksema (*valida, buyruq, yurak, hayot, oldin* kabi);

ChetL – chet tili leksemasi (*printer, byuro*);

RB – raqamli butunlik (*1986, 18/04/2012, 5.05, 19,25*)

HRB – harf-raqamli butunlik (*Boing-767, SU-27, “Kelajak ovozi – 2017”, “Yil ayoli – 2018”*)

ABBR – abbreviatura (*BMT, MDH, O‘zR, DXX, O‘zMU*)

QB – qisqartirilgan birikmalar (*va h.k., va boshq., m-n:*)

Lingvistik tahrir va tahlil dasturi ma’lumotlar bazasiga GT bosqichida quyidagi belgilar kiritildi:

№	Tushuncha yoxud termin nomi	ramzi
1	Bosh harf	Â
2	Kichik harf	ă
3	Probel (bo‘shliq)	□#
4	abzats	¶
5	Tinish belgilar ramzlari (ilova: 4-jadval)	

Bir qarashda ahamiyatsiz tuyulgan belgilar tahriri matnlarning savodli yozilishida o‘z o‘rniga ega. Bir juz’iy lingvistik xato ham matnning sifatiga putur yetkazishi mumkin. Shu bois o‘zbek tilidagi matnlarni avtomatik tahrirlovchi dasturning lingvistik ta’minotini yaratishda, asosan, kirill va lotin yozuvidagi qisqartma so‘zlar, grafemalar tahririni amalga oshiruvchi grafematik tahrirlash modullarini tashkil etuvchi lingvistik belgilarga alohida urg‘u berildi. Grafematik tahrir ko‘p bosqichli tahrirlash dasturining ilk bosqichi hisoblanib, u keyingi bosqichlarning samarali va to‘laqonli bo‘lishida xizmat qiladi.

¹⁰¹ Леонтьева Н.Н. Автоматическое понимание текстов: системы, модели, ресурсы: учебное пособие. –Москва: Академия, 2006. – С.58.

2.3. Morfologik tahlil qilish modulini yaratish prinsiplari va uning algoritmi

Morfologik tahlil (MT) tarixi qadimgi hind tilshunosi Panini bilan bog'liq. U Aṣṭādhyāyī matnidan foydalanib sanskrit tili morfologiyasining 3959 ta qoidasini yaratadi¹⁰². Yunon-rim grammatikasi an'anasi ham MT bilan shug'ullanish bo'lgan¹⁰³. 1859-yilda A.Shleyxer tilshunoslikdagi "morfologiya" terminini o'ylab topadi¹⁰⁴. O'tgan asrning 60-70-yillariga kelib, mashinali morfologiya qamrovidagi barcha tadqiqotlar mashina lug'atini yaratish bilan boshlangan¹⁰⁵.

MTda, asosan, matn imlosini tekshirish, so'zlarning grammatik shakllanishi va ularning tahlili ko'zda tutiladi. Ushbu maqsadga erishish uchun quyidagi ishlar amalga oshirilishi kerak:

1) tilning lug'at tarkibini komyuterga kiritish, ya'ni elektron kompyuter lug'atini tuzish;

2) lug'atdagi adabiy tilga mansub so'zlarni ajratib olish;

3) ajratib olingan so'zlarni uslubiy (ilmiy, badiiy, publitsistik, rasmiy-idoraviy va so'zlashuv) jihatdan guruhlash;

4) adabiy so'zlarni turkumlarga ajratish;

5) so'zlarni turkum doirasida semantik guruhlash (masalan, ot turkumidagi so'zlarni shaxs otlari, o'simlik nom (fitonim)lari, joy nom (toponim)lari kabi guruhlarga ajratish);

6) guruhlangan so'zlarning affikslar kombinatsiyasini tuzish; bunda affikslar kombinatsiyasining amalda adabiy til doirasida foydalaniladigan hajmi qamrab olinadi (ilovada 5-jadval). Kombinatsiyalar qatorini tuzishda affikslar ketma-ketligiga e'tibor qaratiladi. M-n, ot turkumida *ko'plik+egalik+kelishik* kategoriyasi va h.k.

Grammatik vositaning so'z tarkibidagi tartibi ma'lum qonuniyatga ega bo'lib, grammatik vositaning joylashishidagi tartib va izchillik uning ma'no va grammatik xususiyati bilan bog'liq: *yangi lug'aviy ma'no hosil qiluvchi vosita* birinchi (1), *lug'aviy ma'noga ta'sir qiluvchi vosita* ikkinchi (2) va *lug'aviy ma'noga ta'sir qilmaydigan, lekin so'zni bog'laydigan vosita* uchinchi (3) bo'lib qo'shiladi: S+1+2+3.

Affikslarning joylashuvidagi me'yoriy holat ba'zan buziladi: *opa-*

¹⁰² Leonard Bloomfield (1927). On some rules of Pāṇini. Journal of the American Oriental Society. 47. American Oriental Society. – P. 61-70.

¹⁰³ [https://en.wikipedia.org/wiki/Morphology_\(linguistics\)](https://en.wikipedia.org/wiki/Morphology_(linguistics))

¹⁰⁴ Schleicher, August. Zur Morphologie der Sprache. Mémoires de l'Académie Impériale des Sciences de St.-Petersbourg. VII. I, N.7. St. Petersburg. 1859. – P. 35.

¹⁰⁵ Ножов И.М. Морфологическая и синтаксическая обработка текста (модели и программы): Дис. канд. филол.наук. – Москва, 2003. – С. 54.

lar-in, opa-m-lar, ayt-di-ng-lar, ayt-di-lar-ing kabi. Bu soʻzlashuv va badiiy uslubdagi matnlarda kam uchraydigan hodisa. Quyida faqat shakl hosil qiluvchi affikslar doirasida fikr yuritildi.

Lugʻaviy shakl hosil qiluvchining odatdagi tartibi:

– otda: *kichraytirish-erkalash + son*. Kichraytirish-erkalash shaklining kam qoʻllanadigani koʻp qoʻllanadiganidan (ular bazada koʻrsatiladi) avval keladi: *kitobcha, xatcha*;

– sifatda *-roq* qoʻshimchasi *-ish, -imtir* affiksidan keyin, shuningdek, ravishdosh va sifatdoshga qoʻshilganda ham ularni hosil qiluvchi affiksdan keyin qoʻshiladi: *oqishroq, koʻkimtirroq, sovinqiraganroq, tortinibroq* va h.k.;

– sonda: taxmin bildiradigan *-cha* affiksi «dona» maʼnosini ifodalovchi *-ta* (urgʻusiz) unsuridan keyin keladi: *oʻntacha* kabi;

– olmoshda: gumon bildiruvchi *-dir* (urgʻusiz) elementi, umuman, yuklama tabiatida boʻlib, hamma turdagi affikslardan keyin qoʻshiladi: *nima-lar-ni-dir*;

– feʼlda: oʻzlikdan boshqa nisbat koʻrsatkichi orttirma nisbat affiksidan keyin qoʻshiladi: *tarqa-til-di, yugur-tir-ish-di, tik-tir-ish-di*; feʼlning tarz affiksi va kuchaytiruvchisi (*chayqa, bura, toʻzgʻi*); boʻlishsizlik affiksi (*urintirma, chayqatma*); zamon affiksi; shart mayli affiksi va buyruq maylining kuchaytiruvchisi (*bordi, boradi, borsa, boray, borgin/borgil*), feʼlning xoslangan shaklini hosil qiluvchi affiks (*bormoq, borgan, borib* va h.k.) oxirida qoʻshiladi.

Sintaktik shaklning tartibi:

– ismlarda: egalik affiksi+kelishik affiksi: *martabasini, oʻqiganimni*;

– feʼllarda: tasdiq-inkor + zamon / mayl + shaxs + son: *oʻtmayapsan*.

Affiks-yuklama hamma turdagi affiksdan keyin qoʻshiladi: *davlatimiznimi, loyihangami, oʻqimayapsanmi*.

7) Tuzilgan affikslar kombinatsiyasi soʻzlarga biriktiriladi.

8) Tilshunoslar tomonidan amalga oshirilgan yuqoridagi ishlar dasturchi tomonidan taʼminotga kiritiladi.

Morfologik tahlil texnologiyalari yaratilgan boʻlib, bugungi kunda samarador texnologiyalarga aylangan stemming va lemmatizatsiya jarayonini qayd etish mumkin. Dissertatsiyaning 1.3-paragrafida misollar bilan tushuntirilganidek, stemming bir asosdan yuzaga kelgan soʻzshakllardagi belgilar bilan ishlaydi, yaʼni soʻzshakldagi affikslarni asosdan ajratib tahlil qiladi (tovush oʻzgarishi hodisalari tahlil qilinmaydi):

подснежник → *под/снеж/ник*, *водный* → *вод/ный*; *happiness* → *happi/ness*, *babies* → *babi/es*, lemmatizatsiya esa bir lemmaning flektiv (affiks qo‘shilishi natijasida o‘zgarishga uchragan) shaklini e‘tiborga oladi, asosning asliy holatini tiklaydi: *подснежник* → *под/снег/ник*, *водный* → *водa+ный*; *happiness* → *happy+ness*, *babies* → *baby+es*. Ma‘lum bo‘ladiki, so‘zning eng kichik birligi – bitta tovush (harf / grafema)ga qadar tahlil qilish (stemming) va tovush o‘zgarishlari o‘rnini aniqlab, asosni asliy holatiga binoan tahlil qilish (lemmatizatsiya) texnologiyalari o‘z vazifa imkoniyati jihatidan ahamiyatli sanaladi.

Ma‘lumki, egalik va kelishik affikslari bilan o‘zgarish xususiyatiga ega so‘zlar ismlar atamasi ostida birlashtiriladi. Ular tarkibiga ot, sifat, son, olmosh, taqlid so‘z, fe‘lning sifatdosh va harakat nomi kiradi. Shuni nazarda tutib, tilshunoslikdagi turlanish va boshqa grammatik kategoriyalarning so‘zlarga birikib kelish hodisasi ismlarga mansub turkumlarda kuzatilgani bois dasturning LTni yaratishda morfologik sintez ismlar doirasida amalga oshirildi. Buning uchun lug‘aviy shakl yasovchi va sintaktik shakl hosil qiluvchi affikslar bazasi Accessda yaratildi (gr_form). Ko‘plik shakliga ega yoxud semantik ko‘plikka ega so‘zlar (asosan, sanalmaydigan otlar, mavhum otlar) sol_CorrNounNumber guruhiga birlashtirildi.

O‘zbek tilidagi ismlarning grammatik shakllanishi uchun maxsus dasturiy qism (sol_GenIsmForm) yaratildi. Bunda 1) grammatik shakllar nomi jadvalga maxsus belgi berib kiritildi (id_entry,iform); 2) id_dims maydoniga ega affikslar ID_Wordga bog‘lanadi, natijada List of derivate maydonida so‘zshaklni yuzaga keltiradi.

Ai orqali affikslar quyidagi holatda ifodalanadi:

Ai	Belgi izohi	ID	Lemma turkumi
A1	ko‘plik affiksi	k_a	Ism asosli shakllar
A2	egalik affiksi	e_a	
A3	kelishik affiksi	ke_a	
A4	o‘rin-joy ot affiksi	u_j	
A5	qarashlilik affiksi	q_a	
A6	chegaralash affiksi	ch_a	
A7	sifat darajasi affikslari	Adj_a	
A8	affiksli yuklamalar	aff_part	
A9	inkor shaklni hosil qiluvchi aff.	ink_a	
A10	ajratish (-gina)	aj_a	
A11	tegishlilik (-ligi)	teg_a	
A12	o‘xshatish, solishtirish (-chalik, -day, -dek)	us_a	

```

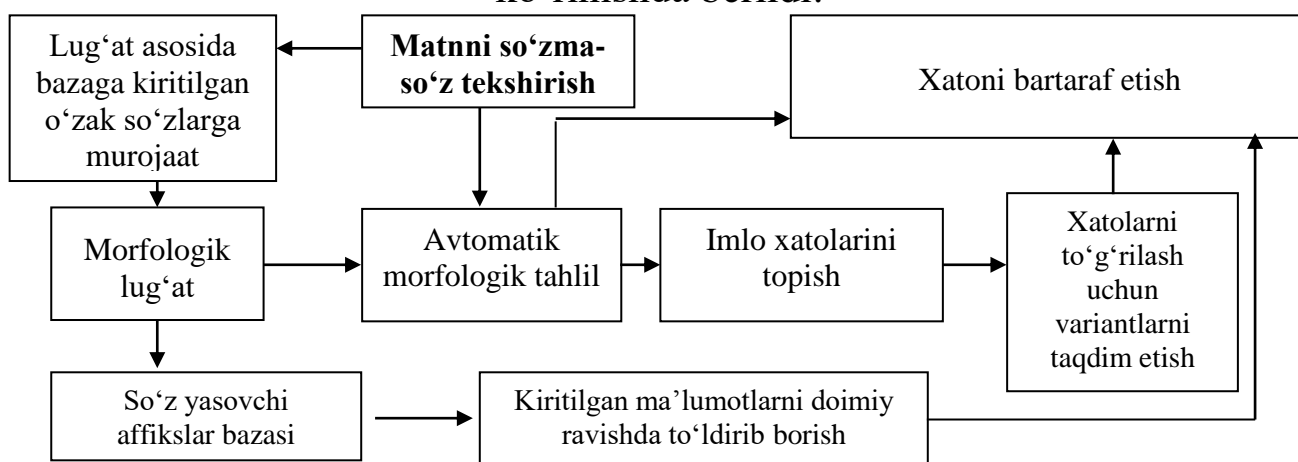
SELECT sg_form.name
FROM sg_entry, sg_form, coord_pairs
WHERE sg_entry.name='davlat'
      AND sg_form.id_entry=sg_entry.id
      AND coord_pairs.id=sg_form.id_dims
      AND coord_pairs.str_pairs LIKE '%egalik+kelishik%' – egalik
va kelishikdagi so'zshakl.

```

List of derivate: davlatimni / davlatingga / davlatidan / davlatingizning / davlatlarida / davlatimizni / davlatingizda / davlatning ...

Dasturning LT wordform_set_coord maydoniga esa affikslar tartibi kiritildi. Jumladan, wordform_set_coord(lemma+*x*s_ya +sh_ya + sin_ya). Bu yerda lemma – asos, s_ya – so‘z yasovchi affiks, sh_ya – shakl hosil qiluvchi affiks, sin_ya – sintaktik shakl hosil qiluvchi affiks. “*x*” belgisi affiksning birikish tartibi qat’iy talab qilinmaydigan holatni anglatadi. Shundan so‘ng so‘zshaklning matnda uchraydigan holati namoyon bo‘ladi: wordform_refresh(wrd). Aynan morfologik tahlilning shu jarayoni mashina tarjimasida muhim o‘ringa ega.

Kompyuter yordamida matni MT qilishning texnik zanjiri quyidagi ko‘rinishda berildi:



8-chizma. Morfologik texnik tahlilning zanjirli ko‘rinishi.

Morfologik tahlildagi asosiy birliklar:

1. Ifoda plani birliklari (grafemalar).
2. Ma’no tashuvchi kichik birliklar (morfemalar).
3. Ma’noli qismlar (so‘zlar).
4. Ma’lumotning kommunikativ tashuvchilari (gaplar).

O‘zbek tilida yozilayotgan matnlar, turli sohalarga oid hujjatlarni ATT qilishga erishish maqsadida yuqorida ilk qadamda e’tiborga olinishi kerak

bo'lgan jihatlar MT doirasida ko'rib chiqildi, tahrirlovchi dasturning lingvistik ta'minotini yaratish borasida muayyan xulosalar chiqarildi.

Matnni avtomatik tahrir va tahlil qilish algoritmi. Ma'lumki, har qanday dastur algoritmlar asosida ishlaydi, algoritmlar esa ma'lumotlar manbaiga tayanadi. Matnlarni qayta ishlovchi lingvistik dasturlar ham milliy tilning lingvistik qoidalariga tayanib tuzilgan algoritmlar asosida o'z vazifasini bajaradi. MT algoritmidagi quyidagi belgilardan foydalaniladi:

U_s – lingvistik ta'minotdagi so'zlar bazasi, $U_s = \text{"SELECT * FROM 'U_s'"};$

Y_s – o'zbek tilidagi barcha yasovchi affikslar bazasi, $Y_s = \text{"SELECT * FROM 'Y_s'"};$

S_q – o'zbek tilidagi grammatik kategoriyalar bazasi, $S_q = \text{"SELECT * FROM 'S_q'"};$

S_i – W matndan ajratib olingan so'zshakllar, $1 \leq i \leq n$, n – W matndagi so'zshakllar soni;

S_{qj} – S_q bazadagi affikslar, $1 \leq j \leq m$, m – S_q bazadagi affikslar soni.

T_z – so'zning qusurli yozilganligini vizual ko'rsatuvchi va yozilgan xato so'zga maqbul so'zshakl variantlarini beruvchi maxsus funksiya.

Quyida gaplarni tokenlarga ajratgan holda, lemma bo'yicha o'zbek tilidagi so'zlar bazasidan izlanadi, topilmasa o'zbek tilidagi barcha yasama so'zlar bazasidan izlanadi. Asos yoki yasalma bazadan topilgach, unga birikish ehtimolidagi affikslar ID_{S_i} bo'yicha o'zbek tilidagi barcha affikslar bazasidan olinadi. Demak, tahlil algoritmi quyidagicha bo'ladi:

1. S_i dagi har bir so'z, U_s bazasidan izlansin. Topilsa, keyingi qadamga, aks holda 5-qadamga o'tsin.

2. S_i so'zning U_s bazadagi ID (tartib raqami) olinsin.

3. S_i so'zning ID raqamiga to'g'ri keladigan affiks S_{qj} – S_q bazadan izlansin.

4. $S_i + S_{qj}$ to'g'ri bo'lsa 10-qadamga o'tilsin, aks holda keyinga qadamga o'tilsin.

5. S_i dagi har bir so'z, Y_s bazasidan izlansin. Topilsa, keyingi qadamga, aks holda 3-qadamga o'tilsin.

6. S_i so'zning Y_s bazadagi ID (tartib raqami) olinsin.

7. S_i so'zning ID raqamiga to'g'ri keladigan affiks S_{qj} – S_q bazadan izlansin.

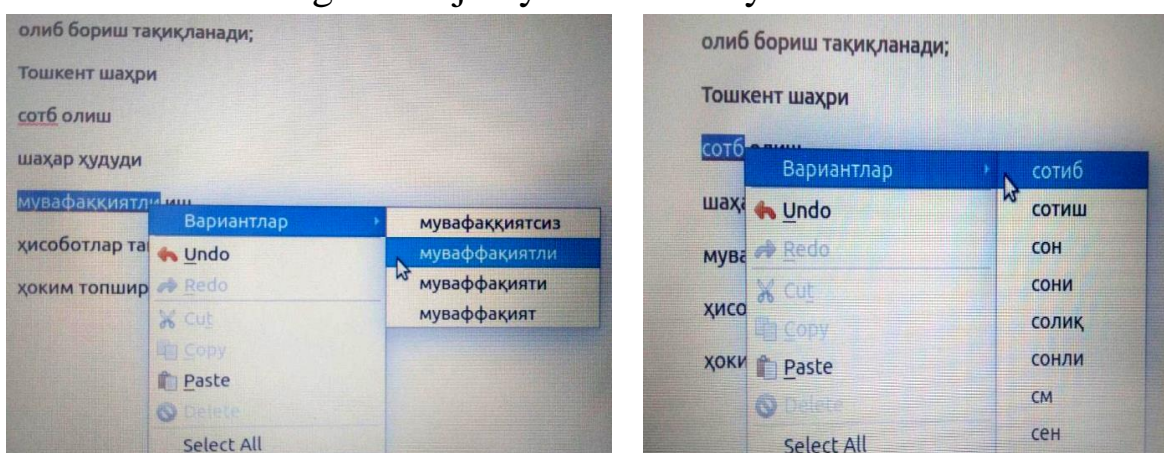
8. $S_i + S_{qj}$ to'g'ri bo'lsa, 10-qadamga o'tilsin, aks holda keyingi qadamga o'tilsin.

9. T_z fuksiya amalga oshirilsin va 10-qadamga o'tilsin.

10. Soʻz toʻgʻri yozilganligi tasdiqlansin va keyingi S_{i+1} soʻzga oʻtilsin.

Algoritm ish tartibini misol orqali tahlil qilamiz: *isteʼmolchilarning* soʻzshakli stemming qilinadi, yaʼni *wordform_set_coord* maydonida berilgan affikslar tartibi boʻyicha kesib chiqiladi: *isteʼmol/chi/lar/ning*. Shundan soʻng *isteʼmol* lemma imlosi List of Words bazasidan tekshiriladi, *-chi*, *-lar*, *-ning* affikslari S_{q_j} – S_q bazasidan qidiriladi, bunda *-lar*, *-ning* shakl yasovchi va sintaktik munosabat shakli ekanligi aniqlanadi, *-chi* yasovchi affiksni topish uchun Y_s bazasiga murojaat qilinadi. Lemma va affikslarning LTda mavjudligi topilgach, soʻzning toʻgʻri yozilganligi tasdiqlanadi va keyingi soʻzga oʻtiladi.

Misolni dasturdagi tahlil jarayoni bilan boyitamiz:



9-rasm. Xato yozilgan yasama soʻz va ravishdosh bilan shakllangan xato soʻzning dastur tomonidan tahlil qilinishi natijasida ularga toʻgʻri variantlarni taklif etish jarayoni.

2.4. Omonimlarni tahlil qilish texnologiyasi

Maʼlumki, omonimlar yozilishi bir xil fonemalardan iborat shakllar hisoblanadi. Shu bois omonimiyada ifoda plani yetakchi¹⁰⁶. Omonim hodisada yasama leksema va tub leksema, shuningdek, ularga grammatik maʼno ifodalovchi affiks qoʻshilganda voqe boʻladigan shakliy teng kelish ham qamrab olinadi. Masalan, *terim* soʻzshakli *ter* – feʼl leksemasiga *-im* leksema yasovchisining birikishi natijasida hosil qilingan leksemaning birlik, bosh kelishik formasiga teng.

Aniqroq boʻlish uchun, bir necha turdagi omonimlarni ajratib koʻrsatish mumkin:

1) nutqning bir qismi yoki lugʻatdagi shakli (lemma) farqlanadi;

¹⁰⁶ Rahmatullayev Sh. Oʻzbek tili omonimlarining izohli lugʻati. – Toshkent: Oʻqituvchi, 1984. – B.5.

2) ba'zi morfologik xususiyatlar, masalan, kelishik, egalik yoki sonda (boy morfologiyali tillarda mavjud) farqlanadi;

3) bir-biridan faqat mazmun jihatdan farqlanadi (bu holat semantik farqlanishda uchraydi).

Tilning morfologik sathini o'zida qamragan universal tizim (korpus) mavjud bo'lmagan holda ko'pgina tillarda so'zlarning turlanish va tuslanish jarayonlarida LTga kiritilgan morfologik lug'atlarga murojaat qilinadi. Shu bois turli xil lug'atlar uchun omonimlik (noaniqlik) tushunchasi biroz farq qilishi mumkin.

Morfologik lug'atlar bilan bir qatorda korpus, deb ataladigan to'liq matnlar to'plami mavjud¹⁰⁷. Korpus ta'minotining shakllantirilishiga qarab omonim shakllarning nafaqat morfologik, balki semantik va sintaktik xususiyatlari ham namoyon bo'ladi¹⁰⁸. Axborot olishning qulay imkoniyatlari yaratilishi natijasida korpus til haqidagi mukammal ma'lumot manbaiga aylanadi, masalan, omonim shakllarning eng keng tarqalgan turi statistikasini bilish mumkin.

Bugungi kunda lingvistik tadqiqot va amaliy topshiriqlar yechimi uchun til korpuslari zamonaviy tilshunoslikning inkor etib bo'lmas ish quroliga aylandi, chunki korpus turli lingvistik topshiriqlarni yechishga xizmat qiladi¹⁰⁹.

Korpus muayyan elektron manbaga leksik, morfologik, grammatik, semantik belgilar asosida ishlov berilganligi sababli lingvistik tadqiqotlarda foydalanish uchun juda qulay imkoniyatni beradi. Sh.Xamroyeva ta'kidlaganidek, til korpusi elektron kutubxonadan farqli ravishda muayyan tilni tadqiq qilish, o'rganish va o'rgatish uchun zaruriy, foydali va qiziqarli matnlar to'planishini nazarda tutadi.¹¹⁰

1-jadvalda SinTagRus¹¹¹ va OpenCorpora¹¹² rus tili korpuslariga asoslanib, rus tilida omonim hodisalar statistikasi aks ettirilgan.

¹⁰⁷www.dialog-21.ru/media/2138/zakharov.pdf Захаров В.П. Корпуса русского языка.; <https://www.myfilology.ru/177/istoriya-korpusnoj-lingvistiki>. История корпусной лингвистики.; <http://www.ruscorpora.ru>.

¹⁰⁸ Ochiq rus milliy korpusi saydi – <http://opencorpora.org>.

¹⁰⁹ Недошивина Е.В. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. Учебно-методическое пособие. – СПб, 2006. – С.26.

¹¹⁰ Xamroyeva Sh. O'zbek tili mualliflik korpusini tuzishning lingvistik asoslari: Fil.fan. bo'yicha falsafa dokt. (PhD) diss. avtoref. – Qarshi, 2018. – 53 b.

¹¹¹ Богуславский И.М., Иомдин Л.Л. и др. Разработка синтаксически размеченного корпуса русского языка. // Доклады научной конференции «Корпусная лингвистика и лингвистические базы данных». – СПб: изд-во Санкт-Петербургского университета, 2002. – С. 40-50.

¹¹² Bocharov V.V., Alexeeva S.V., Granovsky D.V., Protopopova E.V., Stepanova M.E., Surikov A.V. Crowdsourcing morphological annotation // Компьютерная лингвистика и

Soʻzlar	SinTagRus %	OpenCorpora %
Omonimlikka ega boʻlmagan soʻzshakllar	47.58	70.84
Muayyan morfologik xususiyatlarga ega omonimlar	25.58	16.31
Nutq (matn)da voqelangan koʻrinishdagi omonimlar	12.40	11.91
Omonim lemma turkumi	11.70	0.26
Omonim lemma morfologik xususiyatlari bilan	2.26	0.00
Sintagmatik munosabatdagi omonim lemma	0.48	0.67

Jadvaldan turli korpuslardagi omonimlik turlicha ekanini koʻrish mumkin. Bu morfologiyaning nafaqat turli modellaridan foydalanilishiga, balki matnlar uslubi va maʼlumotlar miqdoriga ham nisbatan bogʻliqdir. Rus tili korpusiga birlashtirilgan matnlar tarkibida omonimlikka ega soʻzshakllar koʻp uchraydi. Shu bois xorijiy va rus tilshunosligida omonimlikni farqlash masalasi (“snyatiya omonimii”) alohida tadqiq qilinadi¹¹³.

Omonimlikni bartaraf etish uchun har bir soʻzni “tasniflash” kerak, yaʼni uni lemma – gap boʻlagi va morfologik xususiyatlar majmui bilan taqqoslash mumkin, ular qulaylik uchun bir tegga qoʻshiladi. Barcha mumkin boʻlgan teglarni oʻrganish uchun morfologik lugʻatdagi soʻzlarga tegishli havolalarni topish yoki MyStem¹¹⁴ kabi morfologik analizatorni

интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 29 мая – 2 июня 2013 г.). Вып. 12 (19). – Москва: РГГУ, 2013.

¹¹³ Кобрицов Б.П. Модели многозначности русской предметной лексики: глобальные и локальные правила разрешения омонимии. Автореф... канд. филол. наук. Москва: РГГУ, 2004.; Зеленков Ю.Г., Сегалович И.В., Титов В.А. Вероятностная модель снятия морфологической омонимии на основе нормализующих подстановок и позиций соседних слов (Электрон ресурс). http://www.dialog-21.ru/media/2444/zelenkov_segalovich.pdf; Кобрицов Б.П. Методы снятия семантической неоднозначности. НТИ, Сер.2, Вып. 3, 2004.; Кобрицов Б.П., Ляшевская О.Н., Шеманаева О.Ю. Снятие лексико-семантической омонимии в новостных и газетножурнальных текстах: поверхностные фильтры и статистическая оценка (Электрон ресурс). http://elar.urfu.ru/bitstream/10995/1388/1/IMAT_2005_03.pdf; Hearst M.A. Noun homograph disambiguation using local context in large text corpora // Proceedings of the 7th conference, UW Centre for the New OED & Text Research Using Corpora, 1991.; Yarowsky D. Word sense disambiguation using statistical models of Roget's categories trained on large corpora // Proceedings of the 14th International Conference on Computational Linguistics, COLING'92, 23-28 August, Nantes, France, 1992. – P. 454-460.

¹¹⁴ <http://tech.yandex.ru/mystem> – MyStem морфологик таҳлил дастури сайти.

ishlatish etarli bo'lib, u so'z teglarini topishda yordam beradi. Shundan so'ng bir nechta teglar orasidan faqat tegishli tegni tanlash kerak bo'ladi.

Ushbu muammoni hal qilishda kompyuterni o'rganish (kompyuter yordamida intensiv ta'lim), ayniqsa, tasniflash usullari yordam berishi mumkin. Tasnifni boshlashdan avval tasniflovchi, ya'ni klassifikator belgilangan holatda bo'lgan haqiqiy misollar bo'yicha o'qitilishi kerak. O'quv mashqi obyektning xususiyatlari (qo'shni so'zlarning gap bo'lagi, tinish belgisi)ni va teg o'rtasidagi munosabatni aniqlashdan iborat. Dasturning muvaffaqiyatli amalga oshirilishi tanlangan metodga bog'liq bo'ladi.

Quyida omonim so'zlarshakllarni matn tarkibida aniqlashning keng qo'llaniladigan usullari haqida so'z yuritildi.

Grammatikaga oid manbalardan ma'lumki, omonimiyani farqlash (ya'ni noaniqlikni bartaraf etish)ning deyarli barcha usullari ikki guruhga bo'linadi¹¹⁵:

1. Grammatik me'yorlarga asoslangan usullar. O'z navbatida, ular quyidagicha guruhlanadi:

- texnologiyalarning aralashuvisiz qo'lda kiritilgan metodlar;
- qoidalarni avtomatik yaratish usullari.

2. Statistika asoslangan usullar.

Ushbu guruhlarning har birida afzallik va kamchiliklar mavjud. Shunday holat bo'ladiki, har ikkala guruhga tegishli metodlar kombinatsiyasidan yuzaga kelgan metod yaxshiroq natijani ko'rsatishi mumkin. Bunday gibrid metod bilan tanishishdan oldin har bir turning vakillarini ko'rib chiqish maqsadga muvofiqdir.

Brill usuli. Qoidalarni avtomatik generatsiya qilish metodining yaqqol namunasi – amerikalik lingvist Erik Brill metodidir¹¹⁶. Ish uslubi quyidagicha:

1. Boshlash: Har bir so'z o'sha so'zning tez-tez ishlatib turiladigan tegi (nutqning asl qismida) bilan bog'lanishi lozim. Noma'lum so'zlar ot so'z turkumi sifatida qabul qilinadi. Boshlash bosqichidan nafaqat o'rganish jarayoni boshlanadi, balki omonimlarni bartaraf etish metodi ham harakatga keladi.

2. Tez-tez uchraydigan xatolik uchun o'zgarish (qayta ishlash) qoidasini yaratish.

¹¹⁵ Рысаков С.В. Методы борьбы с омонимией. <http://samag.ru/archive/article/3059>.

¹¹⁶ Brill E. Transformation-Based Error-Driven Learning and Natural Language Processing: A Case Study in Part-of-Speech Tagging // Computational Linguistics. Vol. 21. 1995. – P. 543-565. <http://acl.ldc.upenn.edu/J/J95/J95-4004.pdf>

3. Istalgan minimal xatoga erishguncha ikkinchi bosqichni takrorlash.

O‘tkazish (qayta ishlash) qoidalarida “eski teg, yangi teg, shart” to‘plamlari o‘z ifodasini topadi va qoidada belgilangan shart bajarilganda eski teg yangisi bilan almashtiriladi. Ushbu metodning kamchiligi qoidalar sonining ko‘payishi bilan aniqlik darajasining pasayishida kuzatiladi¹¹⁷, bu Paretoning prinsipiga mos keladi: 80% harakat 20% natijani ta‘minlaydi. Shu bilan birga, prinsip aksincha jarayonda ham ishlaydi: boshlash bosqichining faqat bitta qadamini bajarish natijasida omonimlikni bartaraf etishda yuqori aniqlikka erishish mumkin. SinTagRus korpusida o‘tkazilgan test natijalaridan ma‘lum bo‘ldiki, ushbu metod har bir so‘zning gapdagi o‘rnini 97,4%, morfologik xususiyatlarning to‘liq majmuini 87,6% aniqlik bilan topish imkonini beradi.

1-jadvalga binoan, SinTagRus korpusidagi so‘zlarning deyarli yarmi shakldosh emas. Bu holat o‘z navbatida dasturda morfologik xususiyatlar va gap bo‘laklarini aniqlashda omonimlikni farqlash metodi emas, balki morfologik lug‘at, ya‘ni morfologik analizatorning mukammal ishlab chiqilgani bilan bog‘liqligini ko‘rsatadi. Shakldosh so‘zlarni bartaraf etish metodining foydali ishini baholash uchun statistikani barcha so‘zlar bilan emas, balki faqat omonimlar bo‘yicha hisoblash kerak. Bundan aniq bo‘ladiki, nutqning bir qismini belgilash vazifasi korpusning 24,10 % ga to‘g‘ri keladi, gap bo‘lagi va morfologik xususiyatlarni aniqlash esa 51,94 % ni tashkil qiladi. Brill usulining qisqartirilgan versiyasi bo‘lgan yuqorida aytilgan chastota metodi uchun foydali aniqlik 89,1 % bo‘lsa, gap bo‘lagi va morfologik xususiyatlarni aniqlash ko‘rsatkichi 76,2 % ni tashkil qiladi.

Yashirin Markov modeli. Baum L.E. va uning hamkasblari tomonidan ishlab chiqilgan¹¹⁸ mazkur model omonimlikni aniqlashning statistik metodi statistik jarayonda yuzaga keladigan barcha variantlar ehtimolligini hisobga olishga yordam beradi. Masalan, ma‘lum bir matnda ot turkumiga oid so‘zlar bog‘lovchiga nisbatan tez-tez va ko‘p uchrasa unda ayni kontekstda mavjud omonim katta ehtimollik bilan bog‘lovchi emas, ot turkumiga oid so‘z bo‘ladi, keyingi ehtimollikda bog‘lovchi sifatida hisobga olinadi. Kontekstni tavsiflash uchun N-grammadan foydalaniladi. N-gramma – matnlarga avtomatik ishlov berishda keng qo‘llaniladigan matematik hisob vositasidir. O‘zbek kompyuter

¹¹⁷ Brill E. A simple rule-based part of speech tagger //Proceedings of ANLC. 92, – C. 154.

¹¹⁸Baum, L. E.; Sell, G. R. Growth transformations for functions on manifolds. Pacific Journal of Mathematics. 27 (2) 1968. – P. 211-227.; https://en.wikipedia.org/wiki/Hidden_Markov_model.

lingvistikasida S.Rizayev harf birikmalarini bigramm, trigramm terminlari bilan ifodalagan¹¹⁹.

N-gramma – soʻzlar yoki teglar kabi N-identifikator elementlarning ketma-ketligini ifodalaydi. Ikki element ketma-ketligi – bigramma, uch element ketma-ketligi esa trigramma, deyiladi. Masalan, *old qoʻshimcha+ot* holati bigrammaga misol boʻladi.

Omonimlikni aniqlashning oddiy statistik metodi va boshqa shu kabi metodlarning tavsifini keltirish uchun quyidagi usullar ishlatiladi:

– w_i – jumladagi i -oʻrinda joylashgan soʻz, t_i – ushbu soʻzning identifikatori (tegi).

– $D_{(w)} = \{t_1^w, t_2^w, \dots, t_k^w\}$ w soʻzining barcha mumkin boʻlgan belgilar majmui. Ushbu maʼlumotlarni morfologik lugʻat yordamida olish mumkin. Agar soʻz lugʻatda boʻlmasa uni Brill usulida bajarilganidek, ot soʻz turkumi sifatida hisoblash mumkin, ammo lingvistik taʼminot ishonchli boʻlishi uchun barcha mumkin boʻlgan teglarni qoʻyib chiqish kerak.

– C – korpusdagi muayyan holatlar soni (n -gramma). Bunda $C(t)$ – t teglar soni; va $C(t_1, t_2)$ – bigrammalar soni (t_1, t_2) .

– $C_t(w, t)$ – t tegli w soʻzlar soni.

– $F(w, t)$ – w soʻzida t tegi mavjudligi ehtimoli. Tavsiflar quyidagi formula boʻyicha hisoblanadi:

$$F(w, t) = \frac{C_t(w, t)}{C(t)}$$

– $P(t_i|t_{i-1})$ - bu t_{i-1} tegidan keyin t_i tegining kelish ehtimollik holati. Bunda $i = 1$ boʻlganda t_i tegi gapda birinchi teg hisoblanadi. Hisoblash formulasi quyidagicha:

$$P(t_i|t_{i-1}) = \frac{C(t_{i-1}, t_i)}{C(t_{i-1})}$$

Yashirin Markov modeliga asoslanib, omonimlikni aniqlashning statistik metodi yordami bilan ishlash natijasida n uzunlikdagi jumlada $T_i \in D(w_i)$ boʻlganda $T = \{T_1, T_2, \dots, T_n\}$ teglarning ehtimoliy ketma-ketligi topiladi.¹²⁰

Modellar modifikatsiyasi. Lingvistik korpusdagi statistikaga asosan ehtimollik hisob-kitobida baʼzi kam uchraydigan teglar juftligi uchun R ehtimolliqi nolga teng boʻladi. Bu arzimasek tuyulgan xatolik qolgan ehtimolliklardan qatʼiy nazar, ketma-ketlik natijasining nolga tushishiga olib kelishi mumkin. Ushbu kamchilikni bartaraf etish maqsadida

¹¹⁹ Rizayev S. Oʻzbek tilshunosligida lingvostatistika asoslari. – Toshkent: Fan, 2006. – B. 18.

¹²⁰http://www.academia.edu/15517740/Анализ_статистических_алгоритмов_снятия_морфологической_омонимии_в_русском_языке.

statistikada silliqlash jarayoni¹²¹ bajariladi. Oddiy misol, P funksiyasida Laplasning silliqlanishida P ning qiymati noldan kattaroq bo'lishi talab qilinadi:

$$P(t_i|t_{i-1}) = \frac{C(t_{i-1}, t_i) + 1}{C(t_{i-1}) + 1}$$

Albatta, silliqlash – qulay usul emas, shuning uchun unga juda ko'p tayanilmaydi. Lingvistik korpus hajmi qanchalik katta bo'lsa, to'plangan ma'lumotlar shonchli va kam uchraydigan lingvistik holatlarni o'tkazib yuborish ehtimoli kamroq bo'ladi¹²². Odatda korpusning hajmi cheklangan bo'ladi, ammo statistika yig'ish uchun xususiyatlarni tanlash erkinligi mavjud. Bigrammalar o'rniga trigrammalar hisobga olinishi mumkin. Bu holda qidiruv formulasi quyidagi shaklni oladi:

$$P(t_i|t_{i-1}, t_{i-2}) = \frac{C(t_{i-2}, t_{i-1}, t_i)}{C(t_{i-2}, t_{i-1})} \text{ holatida}$$

$$\operatorname{argmax}_T \prod_{i=1}^n P(T_i|T_{i-1}, T_{i-2}), \text{ bo'ladi.}$$

Lingvistik ta'minotni yaratishda omonimlarni aniqlash metodida tekshiriladigan belgilarning til xususiyatlariga tinish belgilarini ham qo'shish mumkin, ammo bu holatda tinish belgilari N-grammada hisobga olinishi muammoligicha turadi. Masalan, agar punktuatsiya alohida teg sifatida hisobga olinsa bigramma statistikasida orasiga tinish belgisining qo'yilishi natijasida juft teglar alohida-alohida teglar, deb qabul qilinadi. Teglardagi ketma-ket juftlikni saqlab qolish uchun bigramma (teg, teg)ni trigramma (teg, punktuatsiya, teg)ga almashtirish maqsadga muvofiqdir. Bunda qidiruv holati quyidagicha tus oladi:

$$\operatorname{argmax}_T \prod_{i=1}^n P(T_i|T_{i-1}, pm_{i-1}),$$

$$P(t_i|t_{i-1}, pm_{i-1}) = \frac{C(t_{i-1}, pm_{i-1}, t_i)}{C(t_{i-1})},$$

pm_i – i so'zidan keyingi tinish belgisi. Bu yerda va oldingi misollarda mahraj boshqa ko'rinishga ega bo'lishi mumkin: i so'zning istalgan shaklida barcha zanjirlar uchun mahraj bir xil bo'lsa maksimal izlanishga ta'sir qilmaydi. Chastotali modelda bo'lgani kabi so'zning belgisi sifatida foydalanish mumkin.

¹²¹ Рысаков С.В. Методы борьбы с омонимией. <http://samag.ru/archive/article/3059>.

¹²² Клышинский Э.С., Рысаков С.В. Статистические методы снятия омонимии // «Новые информационные технологии в автоматизированных системах». 2015. – С. 556-557.

Amaliyot shuni ko'rsatdiki, yashirin Markov modelidagi teglar chastotasi to'g'risidagi ma'lumotlardan aniqlikka erishish mumkin. Tinish belgilarisiz bigrammalarni ishlatganda qidirish formulasi quyidagi holatda bo'ladi:

$$\operatorname{argmax}_T \prod_{i=1}^n P(T_i|T_{i-1}) \cdot F(w_i, T_i)$$

barcha so'zlarni alohida guruhlarga ajratib, hisob-kitobni yanada aniqroq qilish mumkin. Buni amalga oshirish uchun $G(w)$ funksiyasi kiritiladi va har bir so'z o'zi tegishli bo'lgan guruh identifikatori bilan bog'lanadi. Guruhlarning kiritilishi bilan R funksiyasi quyidagi shaklga keladi:

$$P_G(t_i|t_{i-1}; g_i) = \frac{C_{g_i}(t_{i-1}, t_i)}{C(t_{i-1})},$$

bu yerda $g_i = G(w_i)$, va C_g oxirgi so'z g guruhiga to'g'ri keladigan n -gramm sonini bildiradi. So'zlarni guruhlashda qo'llaniladigan metodning tanlanishiga bog'liq ravishda metodning aniqligini oshirish yoki kamaytirish ham mumkin. Muallif tomonidan taklif qilingan modifikatsiyada har bir guruhda so'z qabul qilishi mumkin bo'lgan mos keluvchi teg varianti ro'yxati, ya'ni $G(w) = D(w)$ mavjud. Buning natijasida qidirish formulasi tubandagicha bo'ladi¹²³:

$$\operatorname{argmax}_T \prod_{i=1}^n P_G(T_i|T_{i-1}; D(w_i))$$

Yuqorida omonimlarni aniqlash bo'yicha ishlab chiqilgan manbalar va formulalar asosida o'zbek adabiy tili matnlaridagi leksemalarni tekshirish formulasi yaratildi.

S – Rasmiy va ilmiy uslubdagi matnlar majmuasi.

T_c – S matnlar majmuasi (korpus)dagi teglar soni.

w_i – biror matnning i – o'rnidagi so'z, t_j – w so'zining S korpusdagi tegi. n – t_j teglar soni.

Matndagi w_i so'zi va undan keyin kelgan w_{i+1} so'zlari bigramma sifatida olinadi.

$W_i = \{w_i, w_{i+1}\}$ – matndagi i – o'ringidagi so'z va undan keyingi so'zlardan tashkil topgan bigramma.

$Y_j = \{w(t_j), w(t_{j+1})\}$ – S korpusdagi t_j – tegli j – o'rinda turgan va undan keyingi so'zlardan tashkil topgan bigramma.

¹²³ Рысаков С.В. Методы борьбы с омонимией. <http://samag.ru/archive/article/3059>.

Endi berilgan w_i – so‘zining to‘g‘riligi, affikslarning to‘g‘ri qo‘shilganligi mavjud S matnlar to‘plamiga tayangan holda tekshirish quyidagi formula yordamida amalga oshiriladi.

$$\prod_{j=1}^n (W_i, Y_j) \Rightarrow W_i$$

O‘zbek milliy korpusining yaratilmaganligi, rasmiy va ilmiy uslubdagi matnlar bazasi to‘la shakllantirilmaganligi e‘tiborga olinib, o‘zbek tilidagi omonim shakllarni tahlil qilishda N-grammga bog‘liq ravishda omonim-tahlil texnologiyasi qo‘llanildi. Buning uchun o‘zbek tilidagi omonimlar LTga ID bilan kiritildi, bu jihat kontekstda omonimlarning grammatik shakllanishini tahlil qilish imkoniyatini beradi. Tahlil aniqligini oshirish maqsadida omonimshaklning turkumi aniqlandi va unga ramziy qiymat berildi. Shundan so‘ng omonimlar bazasi (List of homonym) so‘z birikmalari modeli (Models of CompWords)ga bog‘landi (List of homonym=> Models of CompWords). Natijada kiritilgan M – omonim o‘zi bilan kelgan N – so‘zga bog‘liq ravishda LTdagi M – omonimning ID raqamiga to‘g‘ri keluvchi Q – affikslarni oladi, ya’ni, M[id]+Q[id].

Mazkur texnologiyani misol orqali tahlil qilamiz: $yeng_1 \rightarrow N/ot$ (kiyimning qo‘lni qoplab turadigan qismi), $yeng_2 \rightarrow V/fe'l$ (g‘olib kelmoq, ustun kelmoq), $yeng_3 \rightarrow V/fe'l$ (yeyishga undovchi buyruq fe’li) omonimshakllar matnda sintaktik modulda yaratilgan so‘z birikmalari modelidagi so‘z turkumlari ramzlariga mos ravishda o‘zidan keyin kelgan so‘z (bigramma – ikki belgi ketma-ketligi) bilan sintaktik bog‘lanishda qabul qiladigan grammatik vositasi tahlil qilinadi. Demak, bu yerda N va V bilan bog‘liq modellarga murojaat qilinadi: $N_{com-ni} + V$ (*yengini kesmoq*), $N_{com-ga} + V$ (*yengiga qadamoq*), $Pron_{ni(dir)} + V$ (*hammani yengmoq*), $Pron_{dan(dir)} + V$ (*mendan yengilmoq*), $Adj_{dan} + V$ (Vh) (*kuchsizdan yengilmoq*) va h.k.

Umuman, omonim-tahlil texnologiyasi muayyan omonim so‘zga boshqa so‘zning ma’lum grammatik ko‘rsatkich orqali bog‘lanishi bilan tushuntiriladi va bu jarayon tahlil imkoniyati hamda tahlilning aniqligini oshiradi.

2.5. Lingvistik protsessorning sintaktikoldi tahlil jarayoni

Formal tillar va formal grammatika nazariyasi XX asrning 50-yillarida N.Xomskiyning generativ nazariyasidan kelib chiqdi. Ushbu nazariya sintaktik tuzilmalarini tahlil qilish uchun asos bo'ldi¹²⁴. Avtomatik sintaktik tahlil (parsing¹²⁵) kompyuter lingvistikasi nazariyasining eng ko'p o'rganilayotgan sohasidir. Sintaktik analizatorlar ma'lumotlar bazasi interfeyslarini loyihalashtirish, sun'iy intellekt, matnlarga avtomatik ishlov berish, avtomatlashtirilgan ma'lumot-qidiruv tizimi sohasi, mashina tarjimasida keng qo'llanilmoqda. Ta'kidlash kerakki, avtomatik sintaktik tahlil (ST) usullari ishi tabiiy tilning formal sintaktik nazariyalaridan kelib chiqadi.

Sintaksis deganda tilshunoslikda asosiy birligi gap va eng kichik birligi so'z birikmalari bo'lgan sathlar nazarda tutilgan. Unda gap tarzida muayyan hukmni ifoda qiladigan grammatik izchil bog'langan so'zlar zanjiri tushuniladi. Grammatik izchil zanjir – so'zshakllarning o'zaro ma'lum bir grammatik munosabatlarda bog'lanishidir. So'zshakl yoki grammatik so'z leksemaning o'ziga biriktirishi mumkin bo'lgan grammatik shaklda kelishidir. Shunday qilib, gapning sintaktik tuzilishi so'zshakllarning o'zaro sintaktik bog'langan zanjiridan tashkil topadi.

STda asosiy maqsad (birinchi qadamda) so'zlarning o'zaro bog'lanish me'yorlarini e'tiborga olish. Bir so'zshakl bir qancha so'zning grammatik formasiga to'g'ri kelishi mumkin. Bunday holatda so'zlarni farqlash uchun matn kirituvchiga shunday so'zlar varianti ko'rsatiladi va matn teruvchi so'zlardan muvofiqini tanlab qo'yadi. Ushbu jarayon MTda ham kuzatiladi, ya'ni so'zning xato yozilgani aniqlangach, matn teruvchiga tanlab qo'yish imkonini beruvchi so'zlar varianti taklif etiladi. Masalan:



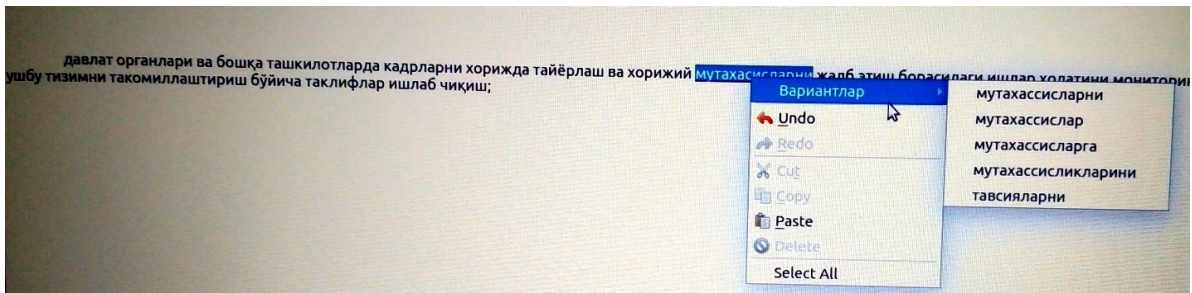
давлат органлари ва бошқа ташкилотларда кадрларни хорижда тайёрлаш ва хорижий мутахассисларни жалб этиш
ушбу тизимни такомиллаштириш бўйича таклифлар ишлаб чиқиш;

10-rasm. Gap tarkibidagi *mutaxassislarni* so'zshaklining xato yozilgani qizil to'lqinli chiziq bilan chizilib ko'rsatilmoqda.

¹²⁴ Большакова Е.И. Автоматическая обработка текстов: задачи, подходы, ресурсы (электрон ресурс).

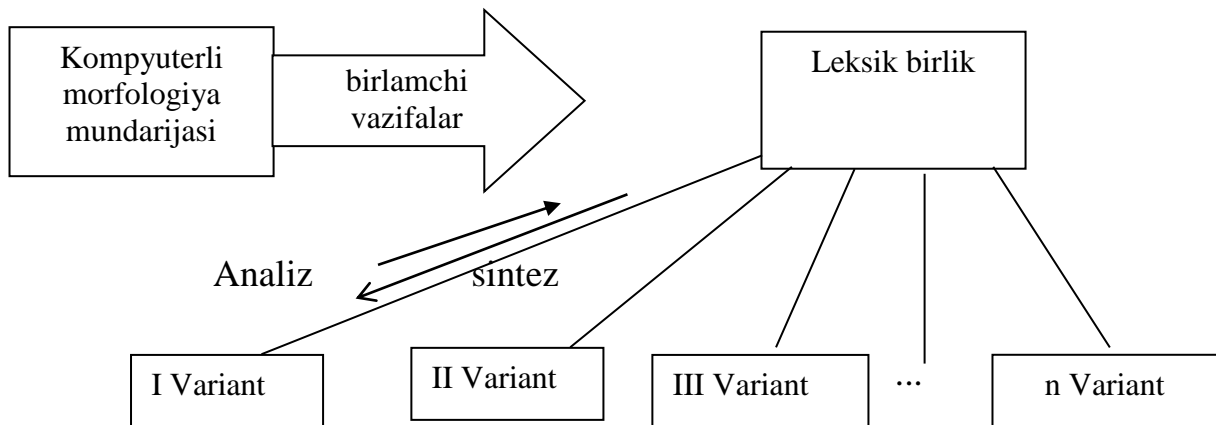
https://www.hse.ru/data/2017/07/24//КЛ_задачи_подходы_ресурсы.

¹²⁵ Rekers, Jan, and Andy Schürr. Defining and parsing visual languages with layered graph grammars. *Journal of Visual Languages & Computing* 8.1 (1997): – P. 27-55.



11-rasm. Xato yozilgan soʻzga ehtimoldagi toʻgʻri variantlar taklif etilmoqda.

Matn tahlilining chizmadagi aksi quyidagi holatda boʻladi:



12-chizma. Tahlil jarayoni.

Sintaktikoldi tahlil jarayoni matni sintaktik tahlil bosqichiga tayyorlab berish vazifasini bajaradi. Ushbu jarayonda matndagi murakkab tarkibli birliklar boʻlaklarga boʻlinib tahlil qilinadi, sodda birliklar esa maʼno jihatdan guruhlanadi. Bu jarayonni kishi miyasida soniya ichida turli fikrlarning aylanishiga oʻxshatish mumkin, zotan, sintaktikoldi tahlil jarayoni koʻzga tashlanmaydi, tahlil faqat dastur sistemasida lingvistik maʼlumotlar asosida olib boriladi.

Sintaktikoldi tahlil jarayoni birinchi galda matndagi birliklar: gap, alohida kelgan soʻzlar va tinish belgilarni aniqlab oladi. Mashina tarjimai tizimidan farqli oʻlaroq, ushbu jarayonda dialogik havolalar, kirish gaplar va birikmalar eʼtiborga olinmaydi. Har bir punktuatsion va raqamli birliklarni, turli belgilarni tekshirish GTning, soʻzshakllarni tahlil qilish esa MTning vazifasi sanaladi. Sintaktikoldi tahlil jarayonida shunday muammolar yuzaga keladiki, ular lingvistik yondashuv orqali oʻz yechimini talab qiladi.

Odatda gap nuqta, soʻroq, undov belgilari, ayrim hollarda koʻp nuqta bilan tugaydi. Ammo bizga maʼlumki, aynan shu belgilar gap oʻrtasida ham qoʻllaniladi. Nuqta koʻp hollarda qisqartmalarda ishlatiladi: *v.h.*, *va boshq.*, *K.R.Rahimov, Toshkent sh., Mirobod tum., Fayzobod koʻch., 2018-*

y. kabi. Undov va so‘roq belgilaridan ko‘chirma gaplarda his-hayajonni anglatish va savol berishda foydalaniladi: *U nima ham qilsin?, nimayam desin?, nimani tushuntirsin?...* Misollardan ko‘rinib turibdiki, yuqorida keltirilgan tinish belgilari hamma vaqt gapni tugallovchi belgilar vazifasini bajarmaydi.

Matnda bosh harflar masalasi ham muammoni yuzaga keltiradi. *Asal judayam ajoyib* (1). *Biznikiga karimalar kelishdi* (2). *Bizga ma’lumki, Yer o‘z o‘qi atrofida aylanadi. Uygaga kelib qaragam, dadam yerga bosh qo‘yib yotgan ekan* (3).

Pragmatika yoki kontekstning ishtirokisiz yuqorida keltirilgan misollardagi ajratib ko‘rsatilgan so‘zlar qanday ma’noda kelayotganini anglash qiyin: (1) gapda *asal* so‘zi ismni anglatayaptimi yoki shirinliknimi? (2) gapdagi *karimalar* so‘zi Karima ismli qiz bilan birgalikdagi shaxslarni anglatayapti, ammo avtomatik tahrir dasturi ism kichik harf bilan yozilgan, deya so‘zni xato yozilgan deb hisoblaydi. (3) misolning 1-gapida Yer sayyorasi haqida fikr yuritilgan, 2-gapda esa ona zamin nazarda tutilgan. Lekin sintaktikoldi tahlil jarayonida 1-gapdagi Yer so‘zi Yer sayyorasi birikmasi tarzida yozilmagani uchun ushbu so‘zda ona zamin haqida fikr yuritilayotganini hisobga olib, uning xato yozilganligini ko‘rsatadi. Bu holatda dasturning ma’lumotlar bazasiga ushbu qoidalarni kiritish talab etiladi:

1. Turdosh otlar atoqli ot vazifasida ham keladi yoki aksincha: *t.ot*→*at.ot*. Ushbu qoidaning muqim bo‘lishi uchun morfologik tahlil modullarida o‘zbek adabiy tilidagi turdosh va atoqli otlar guruhlarini to‘liq va aniq ko‘rsatilishi lozim.

2. Atoqli otlarga *-lar* shakl yasovchi affiksi qo‘shilganda atoqli otlar bosh harf bilan emas, kichik harf bilan yoziladi: *at.ot+lar=t.ot* → *farhodlar, salimalar, komronbeklar*.

3. Toponimlarga *-lik* so‘z yasovchi affiksi qo‘shilganda joy nomlari kichik harf bilan yoziladi: $N_{top.}+lik=N_{top.}$ →*toshkentlik, navoiylik, nurobodlik, chimkentlik, qoraqamishlik, urgutlik* va h.k.

4. Har bir uslubning o‘ziga xos xususiyatlari dastur ta’minotiga kiritiladi. Kiritilgan ma’lumotlar asosida yozilayotgan gapning qaysi uslubga xosligi avtomatik tarzda aniqlanadi, natijada gapda ishtirok etayotgan so‘zlar imlosi tekshiriladi. Masalan: yuqorida keltirilgan (3) misoldagi 1-gap ilmiy uslubga oid bo‘lib, unda ishtirok etgan *Yer* so‘zi bosh harf bilan yoziladi. 2-gap badiiy uslubga xos bo‘lgani bois *yer* so‘zi kichik harfda yozilgan.

O‘zbek tili agglyutinativ tabiatga ega. Shu bois so‘zshakllar asos so‘zga affikslarning ketma-ket (*asos+so‘z yasovchi affiks+so‘z shaklini hosil qiluvchi affiks+so‘z o‘zgartiruvchi affiks*) birikishi natijasida yuzaga keladi. Masalan: *bil+im+don+lar+ga, joy+lar+dagi, xona+dosh+lar+im+niki+da* kabi. Bir so‘zga affikslar va kombinatsiyalarning birikib kelishi natijasida minglab so‘zshakllar yuzaga chiqadi. Bunday ulkan hajmli so‘zshakllarni morfologik lug‘atda saqlash qiyin. Ushbu muammoni bartaraf etish uchun tahrir dasturining morfologik lug‘atiga asos so‘zlar va fonetik o‘zgarishga uchragan so‘zlar grammatik va semantik ma’lumotlari bilan kiritiladi, affikslar esa parametrlari, ya’ni so‘zga qo‘shilib, qanday ma’no ifodalashi, qanday turkumdagi so‘zlarga birikishi, qanday tur affikslari bilan birga kelishi beriladi. Ushbu vazifalar bajarilmasa, *uy+dagi+lar+im+dagi+lar+ning* kabi so‘zshakllarni to‘g‘ri yozilgan, deb hisoblaydi.

Yana bir muhim muammo – tire va chiziqcha muammosi. Ma’lumki, tire (–) tinish belgi, chiziqcha esa “-” diakritik belgi hisoblanadi. Ikkisi ham o‘z vazifalariga ega. Ammo ko‘p hollarda matnlarda ikki belgining o‘rni almashtirib qo‘llanganiga duch kelamiz. Buni bartaraf etish uchun lingvistik ma’lumotlar bazasiga chiziqcha va tire bajaradigan vazifalar algoritmi kiritiladi, shuningdek, chiziqcha bilan birikadigan qo‘shimchalar, juft va qo‘shma so‘zlar ham tayyor so‘zshakllar sifatida kiritiladi.

Matnda keltirilgan raqamlar borasida ham muammolar mavjud. Vazirlar Mahkamasining 1995-yil 24-avgustdagi “O‘zbek tilining asosiy imlo qoidalarini tasdiqlash haqida”gi qarori bilan qabul qilingan “O‘zbek tilining asosiy imlo qoidalari”ning 56-bandida shunday qoida yozilgan: tartib son arab raqamlari bilan yozilsa, *-(i)nchi* affiksi o‘rniga chiziqcha (–) qo‘yiladi: *7-sinf, 5-«A» sinfi, 3-, 7-, 8-sinf o‘quvchilari, 60-yillar, 1991-yilning 1-sentabri* kabi¹²⁶. Tartib sonni ko‘rsatuvchi rim raqamlaridan keyin chiziqcha yozilmaydi: *XX asr, X sinf* kabi. Demak, lotin yozuviga asoslangan yangi o‘zbek alifbosi asosida yozilgan matnlarda yil va oy raqamlaridan so‘ng *-(i)nchi* ma’nosini ifodalash maqsadida chiziqcha “-” qo‘yiladi. Masalan: *2012-yil, 18-aprel*. Ammo bu hamma vaqt ham to‘g‘ri kelavermaydi. Masalan: *Navoiy kon-metallurgiya kombinatining 60 yilligi. O‘zbekiston Respublikasining mustaqillikka erishganiga 26 yil to‘ldi. Komronbekning sportda erishgan bu yutug‘iga 3 oy bo‘ldi*. Misollardan ko‘rinib turibdiki, dasturning lingvistik ma’lumotlar bazasi

¹²⁶ Rahmatullayev Sh. O‘zbek tilining yangi alifbosi va imlosi. – Toshkent: Universitet, 2002. – B.46.

yaratilayotganda yuqoridagi kabi istisnoli holatlar e'tiborga olinishi zarur. Mana shu kabi qoidani dastur ta'minotiga kiritish maqsadga muvofiq:

Qoida: *yil* so'zidan so'ng *bo'ldi, to'ldi* so'zlari kelganda *yil* so'zidan oldin kelgan raqamdan so'ng chiziqcha "-" qo'yilmaydi.

Ma'lumki, har qanday dastur algoritmlar asosida ishlaydi, algoritmlar esa ma'lumotlar manbaiga tayanadi. Matnlarni qayta ishlovchi lingvistik dasturlar ham milliy tilning lingvistik qoidalariga tayanib tuzilgan algoritmlar asosida o'z vazifasini bajaradi. Shuning uchun dasturning lingvistik ma'lumotlar ta'minoti mukammal qilinishi, har bir lingvistik tahrir bosqichlarining modullari ham lingvistik, ham pragmatik jihatdan aniq qoidalar bilan boyitilishi lozim.

2.6. Dasturning sintaktik tahlil qilish moduli

Sintaktik tahlil moduli (STM) sintaktik tahlil qilish dasturi yoki lingvistik dasturning STni amalga oshirishga mo'ljallangan qismi. Bugungi kunda avtomatik STMLarini yaratish KLning dolzarb masalalaridan hisoblanib, buning yechimi turli amaliy maqsadlarni yuzaga keltirish uchun til sistemasini formallashtirishda yuqori sifatli natijalar imkonini beradi: til va nutq birliklarini avtomatik tekshirish sistemasini yaratish, Internetning qidiruv sistemasini mukammallashtirish shular jumlasidandir.

Qayd etish joizki, sintaktik analizator ishi avtomatik MT ishiga nisbatan bir muncha murakkab jarayon hisoblanadi, chunki morfologiya sohasida ko'plab ishlanmalar, me'yorlar, tadqiqotlarning amalga oshirilgani, imlo lug'atlari, morfemik lug'atlari kabi orfografiya, morfologiya, morfemika bo'limlarida manbalarning yaratilgani MT dasturining lingvistik ta'minotini boyitishda muhim o'rin tutadi.

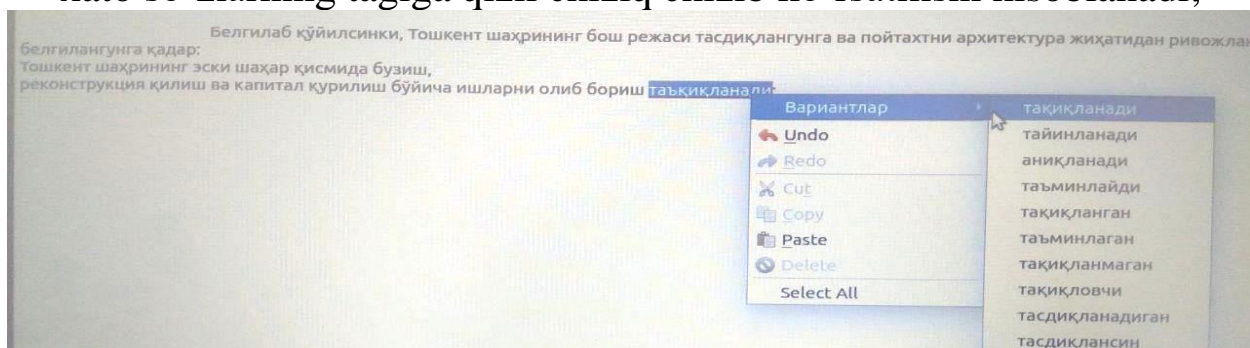
STdan maqsad sintaktik strukturalar va sintaktik bog'lanish me'yorlarini kuzatishdir, biroq barcha tillarda ham STM yaratish murakkab ish hisoblanadi, chunki amaliy tilshunoslikda yetarlicha nazariy va amaliy tadqiqotlar mavjud emas; til va nutq birliklarining sintaktik bog'lanishi individual xilma-xillikka ega. Shunga qaramasdan, ushbu bobda o'zbek tilidagi matnlarni ST qilish dasturining modullari berildi va ST qilish bosqichi tuzilmasi yoritildi. Ishning avvalida sodda sintaktik strukturalarning modullari yaratildi. Biz taqdim etayotgan sintaktik tahlil moduli o'zbek adabiy tilidagi matnlarning grammatik tahliliga yo'naltirilgan Uztextanalysis matn analizatorining qismidir. Sintaktik struktura morfologik me'yorning eng oddiy belgisi bo'lgan so'zshakllar zanjirini tekshiradi, jumlani mantiqiy hosilalarga binoan tahlil qiladi.

Ma'lumki, gap miqdori cheksiz va u individual tuzilish hamda qo'llanilishga bog'liq. Shu bois STda nisbatan kichik birlik hisoblanmish so'z birikmalari tahlili e'tiborga olindi, ya'ni MT modullarini yaratish jarayonida barcha leksemalar so'z turkumlariga ajratilganiga asoslanib, so'zlarning bog'lanish yo'llari modellashtirildi va "so'z birikmalari kategoriyasi" yaratildi. So'z birikmalari kategoriyasi (SBK) bir modulga ega bo'lgan so'z birikmalari guruhidir, ya'ni so'zlararo sintaktik bog'lanishning yagona moduli beriladi, jumladagi sintaktik bog'lanishlar mavjud modullarga tayanib mantiqiy tahlil qilinadi. Jahon KLda ham sintaktik tahlilda, asosan, so'zlarning bog'lanishi tahlil qilinadi. Gap tahlili esa gap bo'laklarini topishda muhim sanaladi¹²⁷. Mazkur tadqiqotda gap bo'laklari tahlili qamrab olinmagan.

Demak, avtomatik tahlil algoritmi jumladagi so'z birikmalarini aniqlash va ular o'rtasidagi aloqalarni qidirishga tushadi.

Белгилаб кўйилсинки, Тошкент шаҳрининг бош режаси тасдиқлангунга ва пойтахтни архитектура жиҳатидан ривожлантиришнинг истиқ
Тошкент шаҳрининг эски шаҳар қисмида бузиш, реконструкция қилиш ва капитал қурилиш бўйича ишларни олиб бориш таъқиқланади;

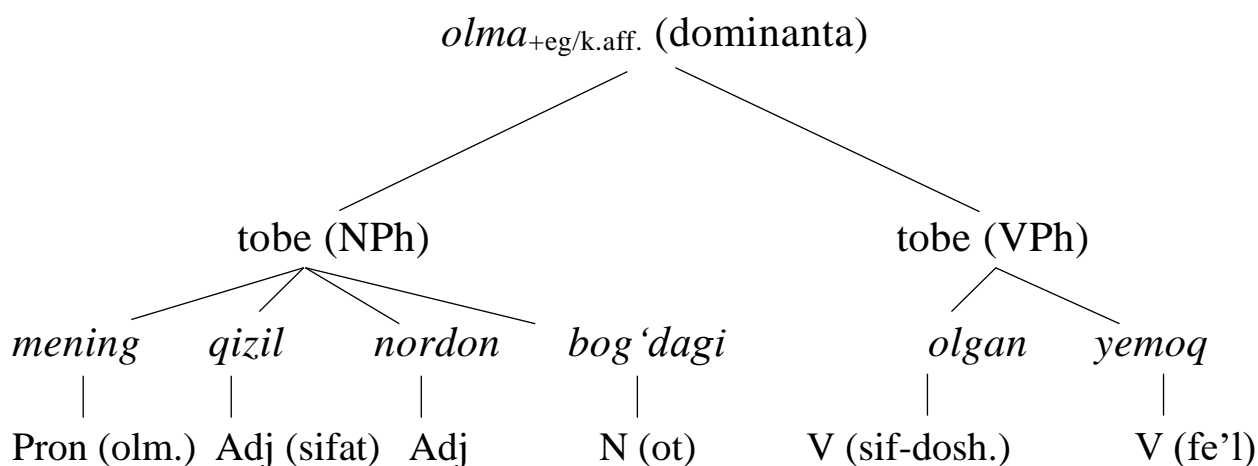
13-rasm. Mazkur tadqiqot ishi natijasida erishilgan 1-amaliy yutuq – bu xato so'zlarning tagiga qizil chiziq chizib ko'rsatilishi hisoblanadi;



14-rasm. 2-katta yutuq: xato so'zga uning to'g'ri yozilish shakllarining taqdim qilinishiga erishildi.

Matematika nuqtayi nazardan har qanday jumla yo'naltirilgan grafik, to'g'rirog'i, daraxt sifatida ifodalanishi mumkin. Daraxtning eng tepasida tobe bog'lanish mavjud. Agar so'zshakllararo munosabat $X \in Y$ ka bog'liq bo'lsa, u holda $X \in Y$ ni ergashtiradi, Y esa X ga bog'liq bo'ladi. X dominanta (hokim) hisoblanadi, Y tobe. Masalan:

¹²⁷ Джураева Н. Разработка формальной модели грамматики официально-делового стиля узбекского языка и её программная реализация: Автореф. дис. ...канд.филол.наук. – Тошкент, 2012.



Avtomatik ST modulini yaratishda yuqori sifatga erishish uchun matnlar korpusidan foydalanish maqsadga muvofiqdir. O‘zbek tili korpusi hali yaratilmagani hisobga olinib, tadqiqotda an’anaviy tilshunoslik me’yorlariga tayanildi. Aytish joizki, matn korpusi asosiy talabni bajaradi: so‘z birikmalarini identifikatsiyalash (bir xillashtirish, muayyan guruhga birlashtirish) va ular o‘rtasidagi sintaktik bog‘lanishni aniqlashni amalga oshiradi.

Ilmiy va rasmiy uslubdagi matnlar tahliliga tayanib, matnlarda uchraydigan sintaktik bog‘lanishlar quyidagicha kategoriyalashtirildi: fe’li birikmalar (FB), otli birikmalar (OtB), sifatli birikmalar (SB), sonli birikmalar (SonB), ravishli birikmalar (RB), olmoshli birikmalar (OB). Har bir sintaktik kategoriya (guruh) o‘z dominantasi (bosh / hokim so‘z)ga ega. Bosh so‘z qaysi so‘z turkumiga oid bo‘lsa, sintaktik bog‘lanish o‘sha turkumga mansub kategoriyaga kiradi. Jumladan, FB dominantasi fe’l, OtB dominantasi ot, SB dominantasi sifat, SonB dominantasi son, RB dominantasi ravish, OB dominantasi olmoshdir. Guruhlardagi so‘zlarning bog‘lanishi va o‘zaro munosabatlari modellashtirildi va an’anaviy tilshunoslikka asosan bitishuvli, moslashuvli hamda boshqaruvli birikmalar tarzida bazasi yaratildi.

Bitishuv yo‘li bilan bog‘langan so‘z birikmalari hech qanday grammatik ko‘rsatkichlarsiz birikkanligi bilan ajralib tursa (*ikki do‘st, tez yurmoq*), moslashuvda hokim va tobe so‘zlarning qaratqich va qaralmish munosabatlariga asoslaniladi (*mening kitobim, yoshlik zavqi*). Boshqaruvda tobe so‘z hokim so‘zga muayyan kelishik yoki ko‘makchi vositasida bog‘lanadi (*do‘sti haqida suhbatlashmoq, uydan baland*). Tobe so‘zning kelishik affikslari bilan birikib kelishi maxsus lug‘atlarda aks etsa maqsadga muvofiq bo‘lar edi.

Bitishuvda birikkan so‘zlarning o‘zaro aloqasi tartib va ohang bilan belgilanadi, ya’ni Y + X (tobe so‘z + hokim so‘z).

Quyida bitishuv yo‘li bilan bog‘lanish modeli berildi:

turdosh ot – N_{com}, fe‘l – V, sifat – Adj, son – Num, ravish – Adv, olmosh – Pron.

№	Bitishuvli birikmalar modeli	misollar
1.	N _{com} + N _{com}	<i>tilla uzuk, ijodkor talaba</i>
2.	N _{dagi} + N _{com}	<i>ishdagi shijoat, xodimdagi mas‘uliyat</i>
3.	N _{dagi} + Num	<i>qarordagi sakkiz..., ilovadagi uchinchi jadval</i>
4.	Adv _{dagi} + N _{com}	<i>quyidagi misollar, yuqoridagi fikrlar</i>
5.	Num + N _{com}	<i>ikkita shartnoma, to‘rt kafedra</i>
6.	Pron + N _{com}	<i>barcha ish, qanday tajriba</i>
7.	Adj + N _{com}	<i>mas‘uliyatli xodim, obod mahalla</i>
8.	Adv + N _{com}	<i>ko‘p ish, yakkama-yakka jang, to‘xtovsiz kurash</i>
9.	Vh + N _{com}	<i>taqdirlash kuni, nishonlash kuni</i>
10.	N _{day/dek} + Adj	<i>kundek ravshan, paxtadek momiq</i>
11.	Vs + Adj	<i>(ta‘lim) olayotgan talabalar, fidoyi yoshlar, tuzalgan bemor</i>
12.	Adj + V	<i>to‘g‘ri so‘zlamog‘, yaxshi bormog‘</i>
13.	Adv + V	<i>eskicha qarash, o‘z-o‘zidan gapirmog‘, yangicha yondashuv</i>
14.	Vr + V	<i>bilib-bilmay gapirmog‘, yakunlanguncha ulgurish</i>

Qoida: qaratqich kelishigidagi tobe so‘zning egalik affiksini olgan hokim so‘zga birikib (qaratqich + qaralmish) kelishi moslashuv hisoblanadi: *yurtimning jamoli, hokimning buyrug‘i, xodimning tushuntirish xati* va v.h.

Rasmiy matnlarda uchragan tobe so‘zshaklda qaratqich kelishigi, hokim so‘zshaklda egalik affiksining ifodalanishi quyidagi modelda berildi:

- 1) Y_{ning} + X_{egalik aff.}: *boshliqning xonasi, talabaning yotoqxonasi;*
- 2) Y₀ + X_{egalik aff.}: *mustaqillik bayrami, Konstitutsiya kuni;*
- 3) Y_{ning} + X₀: *bizning Vatan, bizning yurt.*

Qoida: tobe so‘z tarkibidagi qaratqich kelishigi quyidagi hollarda doimo qo‘llaniladi: a) tobe so‘z va hokim so‘z o‘rtasida boshqa so‘z kelsa, ya‘ni Y_{ning} + L + X_{egalik aff.}: *mudirning xizmat safari, qarorning ikkinchi bandi* va h.k.

b) tobe so‘z atoqli ot yoki olmosh bilan ifodalanganda, ya‘ni N_{prop} + X_{egalik aff} va Pron + X_{egalik aff}: *Ahmadning ishi, uning imzosi* va h.k.

Qaratqich va qaralmish munosabatida tobe so‘z o‘rnida turli mustaqil so‘z turkumlari kelishi mumkin. Tadqiqot davomida rasmiy va ilmiy matnlarda uchragan moslashuvli birikmalar modeli yaratildi:

№	Moslashuvli birikmalar modeli	misollar
1.	N _{com-ning} + N _{com-egalik aff.}	<i>qarorning bandi, buyruqning ijrosi</i>
2.	N _{com-ning} + Adj _{egalik aff}	<i>ishning kattasi, muqovaning yupqasi</i>
3.	N _{com-ning} + Num _{egalik aff}	<i>stolning bittasi, hujjatning ikkitasi</i>
4.	Adj _{ning} + N _{com-egalik aff.}	<i>kattaning o‘rni, kichikning hurmati</i>
5.	Adj _{ning} + Adj _{egalik aff.}	<i>kattaning olasi</i>
6.	Num _{ning} + Num _{egalik aff}	<i>bashning yarmi, ikkitaning biri</i>
7.	N _{com-ning} + Vh _{egalik aff} (Vh – harakat nomi)	<i>rektorning qaytishi, mavzuning yoritilishi, ijroning kechiktirilishi</i>
8.	Vh _{ning} + N _{com-egalik aff.}	<i>izlanish(ning) vaqti, yiqilishning oqibati</i>
9.	Vh _{ning} + Vh _{egalik aff.}	<i>olmoqning bermog‘i</i>
10.	N _{com-ning} + V _{Segalik aff} (Vs– sifatdosh)	<i>olimning izlangani, vaqtning o‘tgani</i>
11.	V _{Sning} + N _{com-egalik aff.}	<i>o‘qiganning tafakkuri, so‘raganning aybi</i>
12.	V _{Sning} + Pron _{egalik aff.}	<i>aytilganlarning hammasi</i>
13.	N _{com-ning} + Adv _{egalik aff.}	<i>ishning ko‘pi, so‘zning kami</i>
14.	Adv _{ning} + Adv _{egalik aff.}	<i>bugunning ertasi</i>
15.	Pron _{ning} + N _{com-egalik aff.}	<i>mening yurtim, uning vazifasi</i>
16.	Adv _{ning} + N _{com-egalik aff.}	<i>hozirning huzuri, kechaning intiqomi</i>

Tobe so‘zning hokim so‘zga tushum, jo‘nalish, o‘rin-payt va chiqish kelishigi affikslari yoki ko‘makchilar yordamida bog‘lanishi boshqaruv hisoblanadi: *o‘zgalarni hurmat qilmoq, hujjatlarni rasmiylashtirmoq, vaziyatdan chiqmoq, tinchlik uchun harakat* va h.k.

Nutqda leksemaning mohiyati hech qachon to‘la-to‘kis voqelanmaganligi kabi sintagmatik imkoniyatning ham faqat zaruri voqelikka aylanadi¹²⁸. Boshqaruvli birikmalarda modellar miqdorining oshib borishi e‘tiborga olinib, bunday birikmalarni quyidagi umumiy model orqali ifodalash maqsadga muvofiq:

¹²⁸ Sayfullayeva R., Mengliyev B., Boqiyeva G. va boshq. Hozirgi o‘zbek adabiy tili. – Toshkent: Fan va texnologiya, 2009. – B. 90.

$\left. \begin{array}{l} N_{com} \\ Adj \\ Num \\ Pron \\ Vh \end{array} \right\} +-ni +V$	$\left. \begin{array}{l} N_{com} \\ Adj \\ Num \\ Adj \\ Vh \end{array} \right\} +-ga+V$	$\left. \begin{array}{l} Num \\ N \\ Pron \\ Vh \end{array} \right\}$	$\left. \begin{array}{l} N_{com} \\ Adj \\ Num \\ Pron \\ Vh \end{array} \right\} +-da+$	$\left. \begin{array}{l} Num \\ N \\ V \\ Adj \\ Vh \end{array} \right\}$
--	--	---	--	---

$\left. \begin{array}{l} N \\ Adj \\ Num \\ Pron \\ Vh \end{array} \right\} +-dan +$	$\left. \begin{array}{l} Num \\ N \\ V \\ Adj \\ Vh \\ Adv \end{array} \right\}$	$\left. \begin{array}{l} N \\ Adj \\ Num \\ Pron \\ Vh \end{array} \right\} + (kel.aff) + ko'makchi +$	$\left. \begin{array}{l} N \\ V \end{array} \right\}$
--	--	--	---

Quyida boshqaruvli birikmalar modellariga misollar keltirildi:

№	Boshqaruvli birikmalar modeli	misollar
1	$N_{com-ni} + V$	<i>arizani yozmoq, murojaatni ko'rib chiqmoq</i>
2	$Adj_{ni} + V$	<i>qizilini tanlamoq</i>
3	$Num_{ni} + V$	<i>uchni to'ldirmoq, ikkisini birlashtirmoq</i>
4	$Pron_{ni(dir)} + V$	<i>hammani sog'inmoq, kimnidir ko'rmoq</i>
5	$Vh_{ni} + V (Vh)$	<i>o'qishni boshlamoq, yozishni o'rgatish</i>
6	$Vh_{ga} + V$	<i>terishga ketmoq</i>
7	$N_{com-ga} + V$	<i>maqsadga o'tmoq</i>
8	$Adj_{ga} + V$	<i>balandga ko'tarilmoq, oqqa ko'chirmoq</i>
9	$Num_{ga} + V$	<i>birga kirmoq, beshga o'tmoq</i>
10	$Pron_{(lar)ga(dir)} + V (Vh)$	<i>hammadan so'ramoq, kimgadir ko'maklashish</i>
11	$Vh_{ga} + V$	<i>ta'lim olishga ko'maklashish</i>
12	$N_{com-da} + V$	<i>uyda o'tirmoq, idorada ishlamoq</i>
13	$Num_{da} + V$	<i>sakkizda kelmoq, to'rtda ketmoq</i>
14	$Pron_{da} + V$	<i>unda tajriba o'tkazmoq</i>
15	$Vh_{da} + V$	<i>(mutolaa) qilishda faollashmoq</i>
16	$N_{com-dan} + Vh$	<i>fikrdan voz kechish, moziyga qaytish</i>
17	$Adj_{dan} + V (Vh)$	<i>sarig'idan olmoq</i>
18	$Num_{dan} + V$	<i>ikkisidan qolishmasdan</i>
19	$Pron_{dan(dir)} + V$	<i>kimdandir so'rash</i>
20	$Vh_{dan} + V$	<i>harakatlanishdan to'xtamoq</i>
21	$Vr + V$	<i>shoshih so'zlamog</i>
22	$N_{com(egalik aff.)} + Ko'm.(ko'makchi) + N_{com}$	<i>xalq bilan muloqot</i>
23	$N_{com(egalik aff.)} + Ko'm.(ko'makchi) + N_{com}$	<i>onam uchun sovg'a</i>

24	Pron + Ko'm.(ko'makchi) + Adj	<i>men uchun aziz</i>
25	N _{com-dan} + Adj	<i>otdan baland, oydan go'zal, temirdan qattiq</i>
26	N _{dan-da} + Adj	<i>kundan-da go'zal</i>
27	Adj _{-dan} + Adj	<i>shirindan shirin</i>
28	Num _{dan} + Num	<i>beshdan ikki</i>
29	Num _{ga} + Num	<i>uchga uch</i>
30	N _{com-qa} + N _{com}	<i>xalqqa murojaat</i>
31	N _{com-ga} +	<i>oqimga qarshi</i>
32	Vh + N _{com}	<i>nuragan uy</i>
33	N _{com-dek} + N _{com}	<i>farishtadek odam</i>
34	Pron _{-dan} + Adv	<i>hammadan o'zgacha</i>
35	Pron _{-dan} + Adj	<i>hammadan chiroyli</i>
va h.k.		

Aslida kompyuter lingvistikasida so'zlarning birikishida grammatik vositalarning mavjudligi bo'yicha so'z birikmalari modellarini quyidagi ikki guruhda yaxlitlash maqsadga muvofiq:

1) grammatik vositali so'z birikmalari: *buyruqning ijrosi, ish bilan band bo'lmoq, birdan gapirmoq*;

2) grammatik vositasiz so'z birikmalari: *katta auditoriya, kichik ilmiy xodim*.

Umuman olganda, so'z birikmalari modellari bizga 2 maqsadda kerak:

1) N-grammali so'zlar birikmasini tahlil qilish;

2) Omonimshakllarning lingvo-tahlil texnologiyasini amalga oshirish.

Kelajakda avtomatik tahlil qilish ehtiyojlari uchun maxsus moslashtirilgan lug'atlar yaratilishi lozim. Masalan, so'z birikmalari lug'ati, juft so'zlar va qo'shma so'zlar lug'ati.

2.7. Gaplarning sintaktik tuzilishini avtomatik tahlil qilish algoritmi tavsifi

Har bir tilning sintaktik analizatori (parser) o'sha til tabiati va qonuniyatlariga binoan yaratiladi. Parser ishi so'zlar kategoriyasi (turkumlar) ma'lumotlari bazasiga asoslanadi va asosiy vazifasi ayrim tillarda jumlada gap bo'laklarini aniqlash (mas., rus tilida) hisoblansa, muayyan tillarda, masalan, ingliz tilida N.Xomskiy nazariyasiga asosan otli guruh va fe'lli guruh tarzida generativ tahlil qilinadi. Gap tarkibini bevosita va bilvosita guruhlariga ajratish yuqoridagi bo'limda ko'rsatib berildi, mazkur bo'limda jumla tuzilishi gap bo'laklari bilan tushuntiriladi.

Ikki holatda ham bazada tildagi har bir soʻzning oʻz turkumligi aniqlanishi zarur sanaladi.

Maʼlumki, anʼanaviy tilshunoslikda gap deb, biror voqea-hodisa haqidagi xabar, soʻroq yoki buyruqni ifodalovchi asosiy sintaktik butunlikka aytiladi. Ammo har bir belgi hisobga olinishi va belgilarning tegga birlashtirilishi inobatga olinib, lingvistik taʼminotda gap quyidagicha qoidalashtirildi:

Qoida: sintaktik analizatorida gap (jumla) deb, ikki koʻrsatkich (tinish belgilari) orasida joylashgan matnning qismiga aytiladi. Bu koʻrsatkichlar quyidagilar: *nuqta / undov belgisi / soʻroq belgisi + boʻshliq va nuqta / undov belgisi / soʻroq belgisi.*

Sxemasi:

[. / ! / ?] + [probel] [. / ! / ?]

Misol:

[. / ! / ?] + [probel] *Suvni tejab ishlatish lozim.*

[. / ! / ?] + [probel] *Hurmatli musobaqa ishtirokchilari*[. / ! / ?]

Sintaktik tahlil jarayoni MT bosqichi soʻzning morfologik xarakteristikasi va uning gap boʻlagi vazifasini aniqlashi bilan boshlanadi. Soʻngra joriy “soʻzlar kategoriyalari” (SK) haqidagi farazlarning shakllanishi boshlanadi. Har qanday SBK quyidagi tarzda ifodalanishi mumkin:

$XG = (x; G)$ – SK nomi, x – dominanta (hokim), YG unga tobe. qiymatlari $Y = 0$ boʻlishi mumkin, bu yerda 0 boʻsh qiymat boʻladi.

Nazariy jihatdan YG ning tarkibiy qismlarga boʻlinishi cheksiz boʻlishi mumkin. Buni rus tilidagi jumla orqali koʻrib chiqamiz:

$1[кот, 2[который пугает и ловит синицу, 3[которая часто ворует пшеницу, 4[которая 5[в тёмном чулане]5 хранится 6[в доме, 7[который построил (Джек)]]1]2]3]4]5]7^{129}$

Keyingi bosqichda SK taʼminoti boʻyicha tahlil qilinayotgan soʻzshakllar kirishi mumkin boʻlgan SK turlari aniqlanadi. Bu SKning kontekstdagi soʻzning morfologik xossalari boʻyicha tekshirilishi bilan hal etiladi.

SK bilan bogʻliq farazlarni shakllantirgandan keyin analizator keyingi soʻzshaklni izlash va tahlil qilishga oʻtadi. Shuningdek, analogiya metodi¹³⁰ bilan dominanta va unga bogʻliq soʻzlar qidiriladi. $XG = (x; YG)$

¹²⁹ Маршак С. Дом, который построил Джек. <https://www.culture.ru/poems/42755/dom-kotoryi-postroil-dzhek>

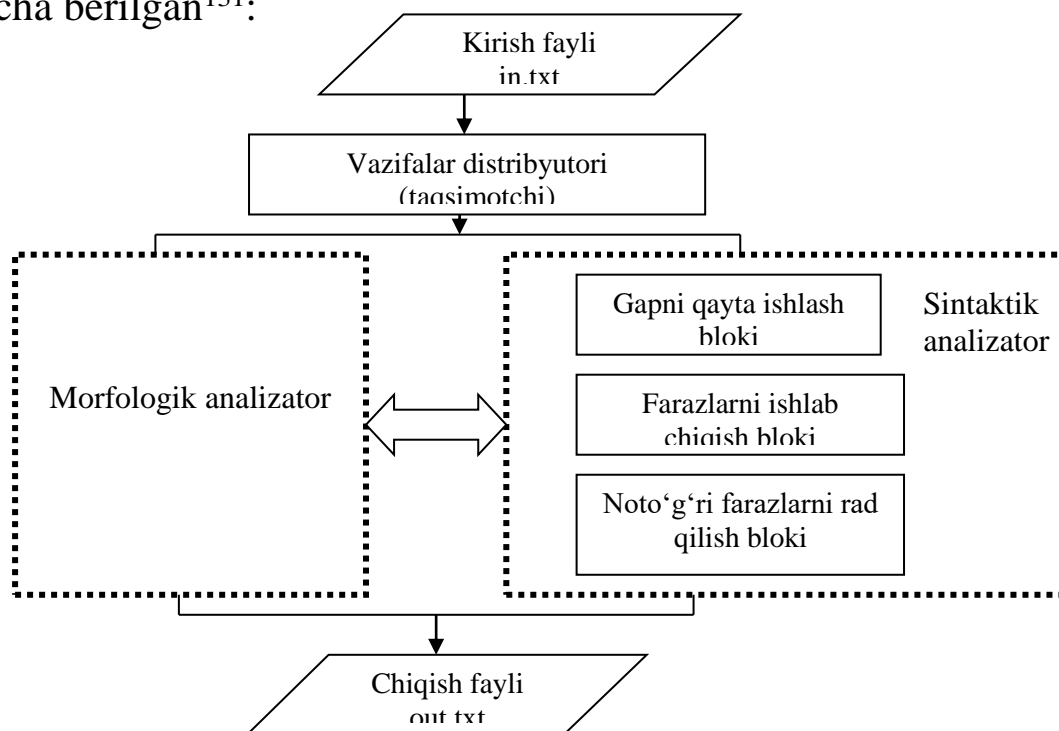
¹³⁰ Analogiya (yunoncha soʻz boʻlib, “moslik”, “oʻxshashlik” maʼnosini anglatadi) munosabatlarning tengligi, muvofiqligi; hodisa va jarayonlarning muayyan xossalari qiyoslash yoʻli bilan anglashda oʻxshash jihatlarini eʼtiborga olish, shuningdek, obyektlar orasidagi

modeli bo'yicha qurilgan jumlar ikki guruhga bo'linadi: ega guruhi (EG) va kesim guruhi (KG). Bundan tashqari, agar A so'zshakl bosh kelishikdagi ot bo'lsa uning dominantasi bo'lmaydi va u ega hisoblanadi. Agar V so'zshakl harakat nomi shaklidagi fe'l bo'lsa u holda uning dominantasi bo'lmaydi va ega bilan shaxs-sonda moslashgan kesim sifatida qabul qilinadi.

ST jumladagi barcha so'zshakllar bir-biri bilan bog'langunga qadar davom etadi. Gap strukturasi tarkibidagi ishtirokchilarni aniqlashning bunday jarayoniga bevosita ishtirokchilarga ajratish metodi deb yuritiladi. Bevosita ishtirokchilarga ajratish sintaktik sath birliklaridan boshlanib, morfologiya, morfemika, fonologiya sathlarigacha davom ettiriladi. Ma'lum bo'ladiki, parsing, ya'ni sintaktik analizatorda morfologik tahlil moduli muhim ahamiyat kasb etadi.

Sintaktik tahlil qilish bosqichi algoritmi. Xorij kompyuter lingvistikasida morfologik va sintaktik tahlil borasida ko'plab izlanishlar olib borilgan. Bunda, asosan, EHM mutaxassislari tadqiqotlari diqqatga sazovordir.

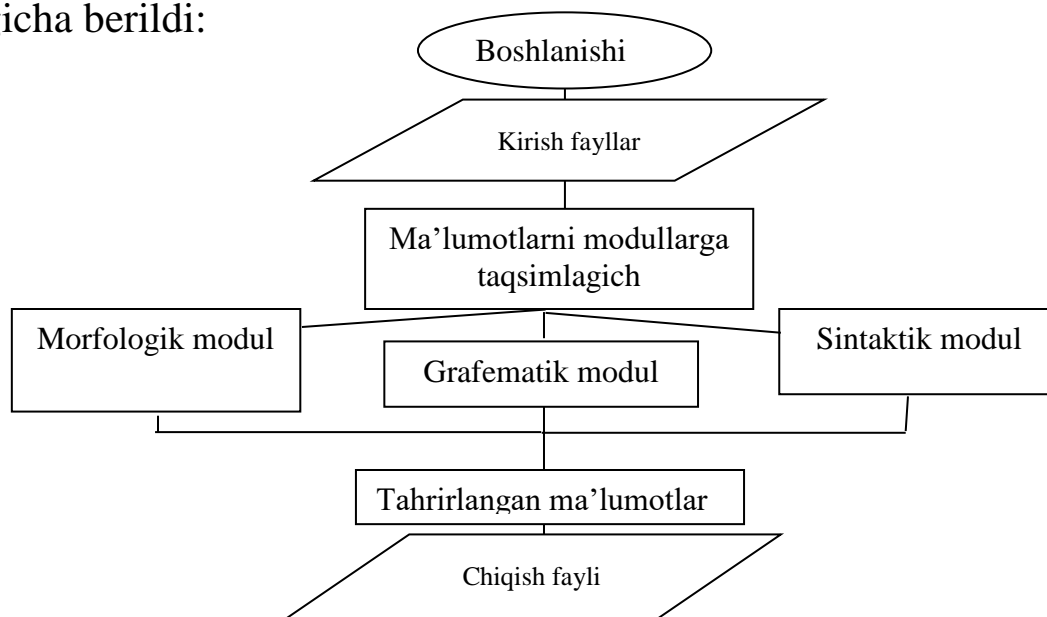
A.Leonteva va I.Kagirovlar tomonidan analizator dasturi blok-sxemasi quyidagicha berilgan¹³¹:



o'xshashlik munosabatlari, o'xshashlik bilan mulohaza yuritish, bir hodisaning xususiyatlarini boshqa hodisa yoki jarayon o'xshashligi bilan tadqiq etish demakdir. Masalan, ingliz tilida *-er, -ness, -dom* kabi, rus tilida *-ночь, -щик, -ник* kabi, o'zbek tilida *-shunos, -chi, -lik* qo'shimchalarga ega bo'lgan so'zlar yuqori ehtimollik bilan ot so'z turkumi hisoblanadi.

¹³¹ Леонтьева А., Кагиров И. Автоматический синтаксический анализ русских текстов // Труды 10-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2008, Дубна, Россия, 2008. – С. 399.

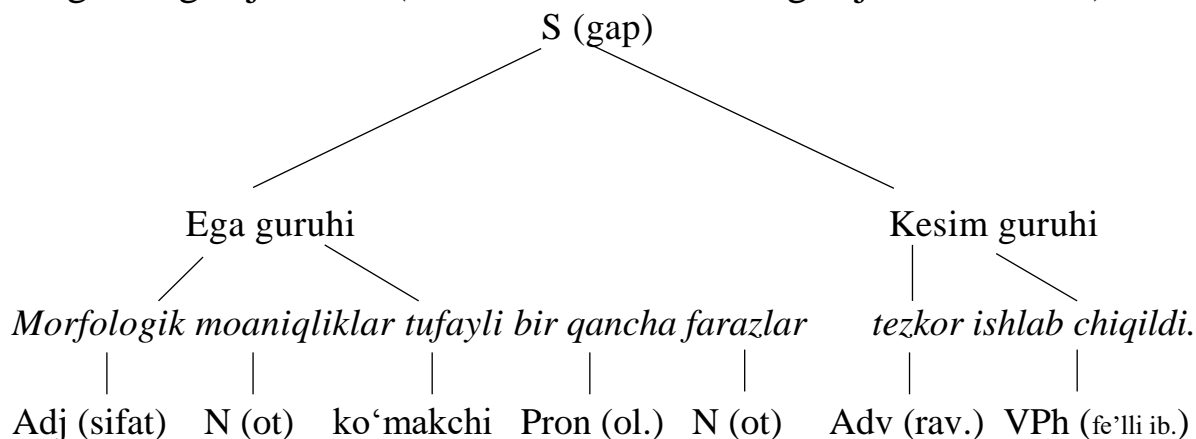
O‘zbek tilidagi matnlarni analiz va sintez qilish dasturi blok-sxemasi tubandagicha berildi:



Yuqorida aytilganidek, kirish ma’umotlari sodda gaplar ro‘yxatidan iborat bo‘ladi. Vazifalarni taqsimlovchi bu ma’lumotlarni sintaktik tahlil qismiga o‘tkazadi. Bu qismda gap so‘zma-so‘z tahlil qilinadi. Joriy so‘zshakl morfologik tahlil bosqichiga uzatiladi,¹³² unda so‘zshaklga mos keluvchi barcha lemma (asos) ko‘rinishlari va grammatik ko‘rsatkichlar tanlanadi, har bir gap so‘zma-so‘z tekshiriladi, shundan so‘ng so‘zshakl MTga uzatiladi. Natijada muayyan so‘zshaklga asos bo‘ladigan barcha asos va mos keladigan grammatik ko‘rsatkichlar aniqlanadi. Agar so‘zshaklga faqat bitta asos mos kelsa, unda u farazlarni yaratish jarayoniga kiradi. So‘z turkumlari va grammatik ko‘rsatkichlarga qarab tegishli sintaktik guruhlar ajratiladi. Dastlabki sintaktik ishlovdan so‘ng so‘zshakl farazlarni ishlab chiqish qismiga o‘tkaziladi. Bu qism asosiy ahamiyatga ega hisoblanadi. Unga kirish tarzida so‘zshakl qabul qilinadi. Agar u gapda birinchi so‘z bo‘lsa, u holda so‘z turkumlariga muvofiq sintaktik guruh aniqlanadi. So‘z turkumlariga bog‘liq ravishda guruhlar birlashtirilishi yoxud variantlar ham yaratilishi mumkin. Masalan, kiritilgan so‘z ot bo‘lsa, unda gapdagi birinchi guruh otli guruh hisoblanadi. Bunday vaziyatda guruhning tartib raqami xotirada saqlanadi va keyingi so‘z qayta ishlanadi. Agar gapdagi birinchi so‘z ravish bo‘lsa u holda bu so‘z ham sifat guruhiga, ham kesim guruhiga tegishli bo‘lishi mumkin. Bunday vaziyatda guruhni aniqlashtirish uchun navbatdagi so‘z talab qilinadi.

¹³² Kagirov I.A., Leontyeva An.B. Grammar-Based Speech- and Word-splitting // Proceedings of 3rd Language & Technology Conference. October 5-7, Poznac, Poland. Poznac: Fundacja Uniwersytetu im. A. Mickiewicza, 2007. 578 s. – P. 413-417.

Gapdagi ikkinchi soʻzdan boshlab muayyan soʻzning qaysi turkumga xosligi emas, balki ayni damda ajratilgan guruh yoki guruhlar haqidagi maʼlumotlar muhim oʻrin egallaydi. Kirishda soʻzshakl joriy sintaktik guruhga tegishli boʻlishi yoki boshqa sintaktik guruhga kirishi ham mumkin. Bunda qoʻshimcha faraz paydo boʻladi va ikkala variant ham koʻrib chiqiladi. Nihoyat gap formal grammatika asosida ega guruhi va kesim guruhiga ajratiladi (bevosita ishtirokchilarga ajratish metodi).



Taʼkidlash joizki, bir guruh, masalan, ot guruhi oʻz tarkibiga boshqa guruhlarini ham olishi mumkin, bunda guruhlarda soʻzlarni tanlash jarayoni indeks shaklida saqlanib qoladi. Shunday qilib, morfologik va sintaktik noaniqliklar tufayli muayyan gap uchun bir qancha farazlarni tuzish mumkin.¹³³ Shundan soʻng ushbu farazlar algoritm blok-sxemasining notoʻgʻri farazlarni tekshirish blokiga oʻtadi. Mazkur blokda ikki bosqichli tekshirish mavjud: birinchi bosqichda bir gapdagi sintaktik guruhlar mosligi tekshiriladi. Bunday moslik guruh dominantasining grammatik xususiyatlaridan kelib chiqib aniqlanadi. Chiqish fayli oʻzida tarkibi sintaktik guruhlariga boʻlingan gaplar roʻyxatini namoyon qiladi. Agar gapda lugʻatga kirmagan soʻz boʻlsa u holda ushbu soʻz tahlil qilinmaydi.

IKKINCHI BOB BOʻYICHA XULOSA

1. Axborot asriga kelib avtomatik matn tahriri va tahlili jahon tilshunosligining shiddat bilan rivojlanayotgan sohasiga, shunday vazifaga ega dasturning oʻzi esa beqiyos zamonaviy ish quroliga aylandi.

2. Model til haqidagi bilimlarni obyektiv ifodalash vositasi boʻlganligi bois lingvistik hodisalarni modellashtirish KLda eng samarali usul hisoblanadi. Boshqa tarafdin, oʻzbek tili ilmiy va rasmiy uslubdagi

¹³³ Попов Э.В. Общение с ЭВМ на естественном языке. – Москва: Эдиториал УРСС, 2004. – С. 260.; Mel'chuk I. A. Dependency syntax: Theory and practice. – Albany, NY: SUNY Press, 1988. – P. 428.

matnlar tahliliga tayangan holda matnlarda uchraydigan sintaktik bogʻlanishlar quyidagicha kategoriyalashtirildi: feʼlli birikmalar (FB), oʻtli birikmalar (OtB), sifatli birikmalar (SB), sonli birikmalar (SonB), ravishli birikmalar (RB), olmoshli birikmalar (OB). Har bir sintaktik kategoriya (guruh) oʻz dominantasi (bosh soʻz)ga ega. Dominanta soʻz qaysi soʻz turkumiga oid boʻlsa, sintaktik bogʻlanish oʻsha turkumga mansub kategoriyaga kiradi. Jumladan, FB dominantasi feʼl, OtB dominantasi ot, SB dominantasi sifat, SonB dominantasi son, RB dominantasi ravish, OB dominantasi olmoshdir. Guruhlardagi soʻzlarning bogʻlanishi va oʻzaro munosabatlari bitishuvli, moslashuvli va boshqaruvli birikmalar modellari orqali amalga oshiriladi.

3. Lingvistik protsessorni yaratishda lingvistik modulning oʻrni va ahamiyati oʻta muhim. Shu bois oʻzbek kompyuter lingvistikasida matnlarga ishlov beruvchi lingvistik dasturlarga mukammal modullarning yaratilishi muayyan maqsad yoʻlida tayyor lingvistik mahsulotning yuzaga kelishiga zamin yaratadi. Oʻzbek tiliga oid matn tahlili jarayonida modullar quyidagilarni amalga oshirish mumkin: 1) leksik-morfologik – soʻzshakldan uning leksik-grammatik xususiyatlariga oʻtish; 2) sintaktik jummalarni ifodalovchi leksik va grammatik xususiyatlar zanjiridan uning sintaktik tuzilishiga oʻtish; 3) semantik – sintaktik tahlil qilingan lingvistik birlikdan semantik ishlovga oʻtish. Umuman, matnlarni tahrir va tahlil qiluvchi dasturning mukammal boʻlishi uchun quyidagi lingvistik modullar ishlab chiqilishi lozim: 1) grafematik tahrir modullari; 2) morfologik tahlil modullari (morfoanalizator); 3) sintaktik tahlil modullari (sintaktik analizator); 4) semantik tahlil modullari.

4. Bugungi kunda lingvistik tadqiqot va amaliy topshiriqlar yechimi uchun til korpuslari zamonaviy tilshunoslikning inkor etib boʻlmas ish quroliga aylandi, chunki korpus turli lingvistik topshiriqlarni yechishga xizmat qiladi. Korpus muayyan elektron manbaga leksik, morfologik, grammatik, semantik belgilar asosida ishlov berilganligi sababli lingvistik tadqiqotlarda foydalanish uchun juda qulay imkoniyatni beradi.

5. Dastlabki joriylanishda rasmiy-idoraviy va ilmiy uslubdagi matnlarni tahrir va tahlil qilishga moʻljallangan lingvo-sistemaning mukammal ishlashi va natijasi samarali boʻlishi uchun mazkur sistema grafematik, morfologik va sintaktik tahlil bosqichlariga ajratilib, modullashtirilishi maqsadga muvofiq. Grafematik bosqichda tahriri amalga oshiriladigan birliklarning belgilab olinishi tahrir aniqligini taʼminlaydi. Morfologik tahlil prinsiplari ot turkumi misolida yoritib berilishi boshqa turkum modullarini tuzish tamoyillarini yaratishga zamin

bo‘ladi. Sintaktik tahlil jarayonida gap miqdori cheksiz va u individual tuzilishi yoxud qo‘llanilishga bog‘liq bo‘lgani sababli STda nisbatan kichik birlik hisoblanmish so‘z birikmalariga tayanilgan ma‘qul, ya’ni, MT modullarini yaratish jarayonida barcha leksemalar so‘z turkumlariga ajratilganiga asoslanib, so‘zlarning bog‘lanish yo‘llari modellashtiriladi va “so‘z birikmalari kategoriyasi” yaratiladi. So‘z birikmalari kategoriyasi (SBK) bir modulga ega bo‘lgan so‘z birikmalari guruhidir, ya’ni so‘zlararo sintaktik bog‘lanishning yagona moduli beriladi, jumladagi sintaktik bog‘lanishlar mavjud modullarga tayanib mantiqiy tahlil qilinadi. Sintaktik tahlilda, asosan, so‘zlarning bog‘lanishi tahlil qilinadi. Gap tahlili esa gap bo‘laklarini topishda muhim sanaladi. Sintaktik tahlil qilish modulini yaratish turli amaliy maqsadlarga erishish uchun til tizimini formallashtirishda yuqori sifatli natijalarni qo‘lga kiritish imkonini beradi.

6. Dunyo kompyuter lingvistikasida omonimlikni aniqlash usullari o‘rganilganida, bu xususdagi tajribadan foydalanib o‘zbekcha matnlardagi so‘zshakllarning tegishli tekshirish formulasi yaratildi. Omonimlikni bartaraf etish uchun har bir so‘zni “tasniflash” kerak, ya’ni uni lemma va morfologik xususiyatlar majmui bilan taqqoslash mumkin, ular qulaylik uchun bir tegga qo‘shiladi. Barcha mumkin bo‘lgan teglarni o‘rganish uchun morfologik lug‘atdagi so‘zlarga tegishli havolalarni topish yoki MyStem kabi morfologik analizatorni ishlatish yetarli bo‘lib, u so‘z teglarini topishda yordam beradi. Shundan so‘ng bir nechta teglar orasidan faqat tegishli tegni tanlash kerak bo‘ladi.

7. Omonim so‘zshakllarni tahlil qilishda qo‘llanilgan texnologiya o‘zbek tilidagi matnlarda uchraydigan omonimshakllarni korpusga bog‘liq bo‘lmagan ravishda tahlil qiluvchi optimal lingvistik usul hisoblanadi.

III BOB. SO‘Z TURKUMLARI MODULLARI

3.1. Mustaqil so‘z turkumlari moduli

So‘z turkumlari til va nutq dixotomiyasining asosiy birligi hamda olamning semiotik tizimi bo‘lmish tildagi lisoniy aksi bo‘lgan so‘zlarning asosiy sinflari (guruhlari, razryadlari) bo‘lib, ular: a) umumiy abstrakt, kategorial (leksik) ma‘noning mavjudligi; b) turkumga xos bo‘lgan ma‘lum grammatik kategoriyalar tizimining mavjudligi; v) ma‘lum sintaktik vazifa(lar)ga xoslanganlikning mavjudligi; g) so‘z va so‘zshakl yasashning maxsus tur va usullarining mavjudligiga ko‘ra turkumlarga ajraladi. Har qanday mustaqil ma‘noli so‘z turkumlari tarkibiga kiruvchi so‘zlar, agar ular yasama so‘zlar bo‘lsa, morfemalardan iborat bo‘lishi lozim. Morfemalar esa ikki xil bo‘ladi:

I. O‘zak morfema – so‘zning tub (atash) ma‘nosini bildirib, mustaqil qo‘llanadi. Masalan: *savat, kitob, daftar, chizg‘ich* va h.k.

II. Affiksial morfema – o‘zakka (asosga) qo‘shilib, turli ma‘nolarni ifodalaydi, so‘zlarni bog‘lash uchun xizmat qiladi.

Affiksial morfema (keyingi o‘rinlarda *affiks*)lar vazifasi va so‘zga qo‘shilib anglatadigan ma‘nosiga ko‘ra ikki turga bo‘linadi:

1. So‘z yasovchi affikslar – asosga qo‘shilib, asosdan o‘sib chiquvchi yangi ma‘noli so‘z yasaydi: *kitob – kitobxon, kuch – kuchli, ma‘no – ma‘nosiz, davlat – nodavlat, ek – ekin, iste‘mol – iste‘molchi* kabi.

2. Shakl yasovchi affikslar. Vazifasi va ma‘nolariga ko‘ra ikki guruhga bo‘linadi:

a) lug‘aviy shakl yasovchi affikslar asosga qo‘shilib, ma‘noni bir oz o‘zgartiradi. Ular sirasiga quyidagi affikslar kiradi:

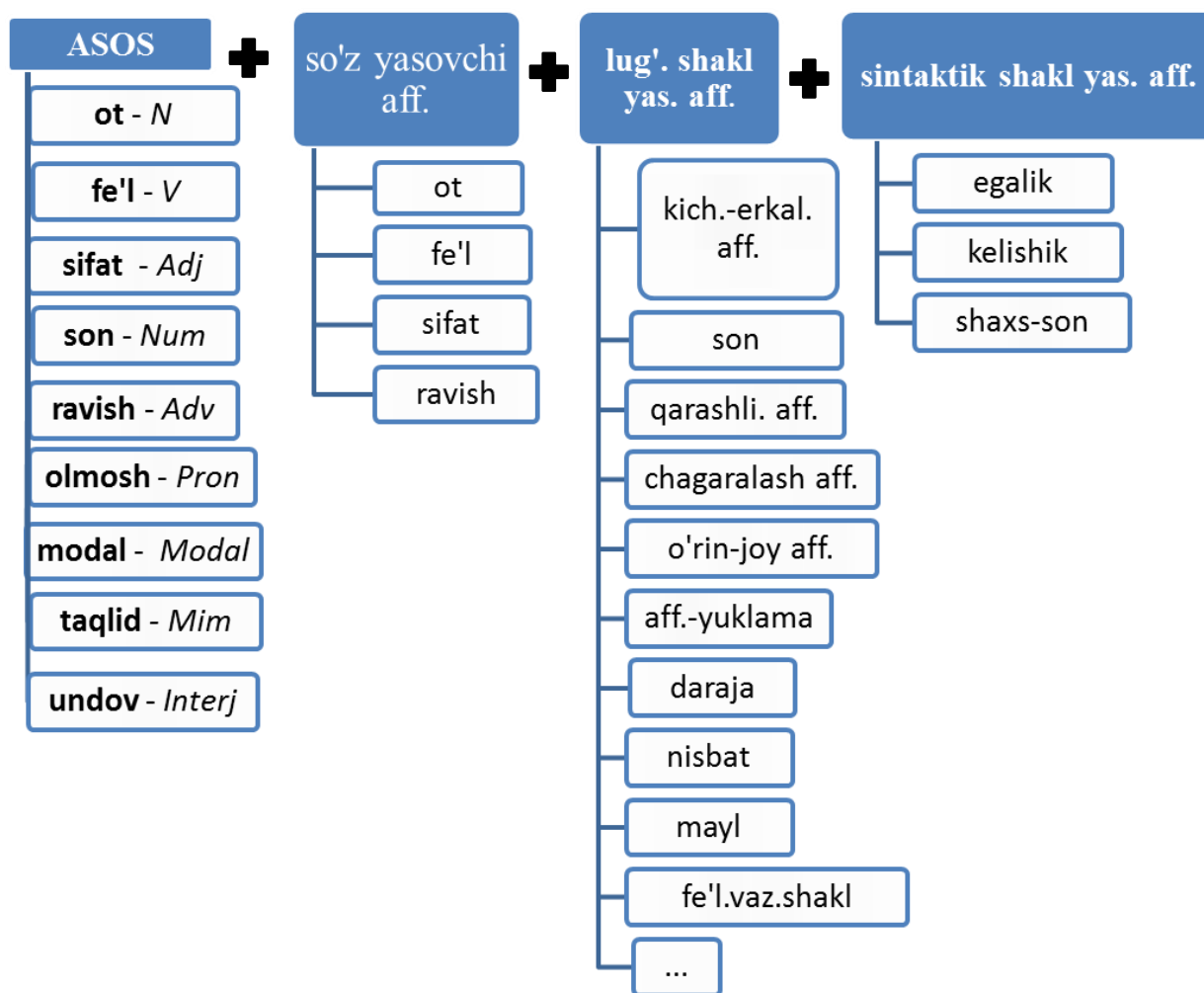
- qarashlilik affiksi: *-niki*;
- o‘rin-joy oti affiksi: *-dagi*;
- chegaralash affiksi: *-gacha*;
- ko‘plik affiksi: *-lar*;
- sifat darajasi affikslari: *-roq; -ish, -sh; -imtir, -mtir*;
- fe‘lning vazifadosh shakllari affikslari, nisbat va mayl affikslari: *-v, -uv, -moq, -mak; -sh, -ish; -r, ar, -gan, -kan, -qan; -adigan, ayotgan; -b, -ib, -a, -y, -gach, -kach, -qach, -guncha, -kuncha, -quncha, -gancha, -kancha, -qancha; -gani, -kani, -qani*;
- affiksli yuklamalar: *-mi, -chi, -gina, -kina, -qina, -dir, -u, -yu, -da, -a, -ya, -mish, -ov, -yov, -mikan*;
- inkor shaklni yasovchi affiks: *-mas, -siz*;
- o‘xshatish-solishtirish affiksi: *-chalik, dek, -day*;

b) sintaktik shakl yasovchi affikslar so‘zlarni bir-biriga bog‘lashda xizmat qiladi. Bularga kelishik, egalik, shaxs-son affikslari kiradi.

Matni ATT qiluvchi dasturning algoritmini tuzishda yuqorida keltirilgan shakl yasovchi affikslarning har biriga ma’lum belgi qo‘yish talab qilinadi. Ramzga ega affikslar Ai va Ei orqali ifodalandi:

Ai	Belgi izohi	Ramzi	Lemma turkumi
A1	ko‘plik affiksi	k_a	Ism asosli shakllar
A2	egalik affiksi	e_a	
A3	kelishik affiksi	ke_a	
A5	o‘rin-joy ot affiksi	õ_j	
A6	qarashlilik affiksi	q_a	
A7	chegaralash affiksi	ch_a	
A8	sifat darajasi affikslari	Adj_a	
A9	affiksli yuklamalar	aff.Part	
A10	inkor shaklni hosil qiluvchi aff.	[inkor]	
A11	ajratish (-gina)	aj_a	
A12	tegishlilik (-ligi)	teg_a	
A13	o‘xshatish, solishtirish (-chalik, -day, -dek)	õx_a	
E1	fe‘lning vazifadosh shakllari affikslari	V _i	
E2	nisbat affikslari	x	
E3	shaxs-son affikslari	y	
E4	zamon affikslari	z	
E5	mayl affikslari	m	

Dasturning LT wordform_set_coord maydonida so‘z turkumlarining grammatik shakllanish tartibi berildi:



Matnli manbalar ustida kompyuter orqali ishlash bilan bog'liq ko'pgina muammolarni hal qilishda jumlar tuzilishi haqida aniq va batafsil ma'lumot talab qilinadi. Ushbu ma'lumotlar mashina tarjima sistemasini yaratish, intellektual (semantik) axborot qidiruvini shakllantirish, to'g'ri va xolis ma'lumotlarni (data mining) uzatish va fikrlarni yig'ishda (opinion mining), matnlarni avtomatik tasniflash hamda avtomatik tahlil sistemasini yaratishda muhim sanaladi. Jahon KLda mazkur ma'lumotlar matnlarni avtomatlashtirilgan tarzda tushunish (NLU) bilan bog'liq sistemalarni yaratish orqali hal etib kelinmoqda. Bu jarayonda *qoidalarga asoslangan usullar* (rule-based methods) – til grammatikasi va semantikasining matematik formallashtirilishi, *statistik metodlar* (statistical methods), *mashinali o'qitish* (machine-learning: matnlar to'plamining sintaktik hodisalari belgilanishiga asoslaniladi) kabi bir necha usullar qo'llaniladi¹³⁴. Jumladan, ATT dasturi morfologik va sintaktik modullarini yaratishda qoidalarga asoslangan usuldan foydalanish

¹³⁴ Добров В.А. Прикладная и компьютерная лингвистика // Николаев И. С., Митренина О. В., Ландо Т. М. (Ред.). – Москва: URSS, 2016. – 320 с.

va affikslar bilan bog‘liq statistik ma’lumotlarning o‘rganilishi ish samarodligini oshiradi.

3.1.1. Ot turkumini formallashtirish moduli

Matnni morfologik tahlil qilish prinsiplari tadqiqotning 2.3-bandida berildi. Shunga binoan avval o‘zbek adabiy tilining imlo lug‘atida mavjud so‘zlarning keng iste’moldagilari ajratib olindi. Galdagi vazifa komyuterli lug‘atdagi so‘zlarni turkumlarga ajratib chiqishdan iborat. Olti mustaqil so‘z turkumi (ot, sifat, son, olmosh, ravish, fe‘l), uch yordamchi so‘z turkumi (bog‘lovchi, ko‘makchi, yuklama), alohida olingan so‘zlar turkumi (modal, taqlid, undov so‘zlar), so‘z-gaplar va aniq bir turkumga kiritilmagan so‘zlar guruhi ajratiladi. Bunda so‘zlarning unli yoki undosh bilan tugallanganligi e’tiborga olinadi. Chunki affikslarning so‘zlarga birikib kelish varianti so‘z oxiridagi tovushga bog‘liq bo‘ladi. Masalan, *uka* va *uy* so‘zlari quyidagicha egalik affikslarini oladi:

soni shaxsi	birlik	ko‘plik	birlik	ko‘plik
I	<i>uka+m</i>	<i>uka+miz</i>	<i>uy+im</i>	<i>uy+imiz</i>
II	<i>uka+ng</i>	<i>uka+ngiz</i>	<i>uy+ing</i>	<i>uy+ingiz</i>
III	<i>uka+si</i>	<i>uka+lari</i>	<i>uy+i</i>	<i>uy+lari</i>

Ishni bajarish, ya’ni matnni ATTini amalga oshiruvchi dasturning lingvistik bazasini yaratishda har bir so‘zning affikslar bilan kelish holati alohida-alohida ko‘rib chiqish, katta yutuqlarga erishish garovidir. Ammo bunga ko‘p vaqt, fizik mehnat talab qilinishi hisobga olinib¹³⁵, ishning optimal varianti o‘ylab chiqildi. Bunda har bir turkum doirasidagi so‘zlar, ya’ni turkum tarkibi ma’no guruhlariga ajratildi. Shuni nazarda tutib, ot turkumidagi so‘zlarning umumiy semantikasi va affikslar kombinatsiyasini o‘ziga biriktirishdagi yaqinliklari hisobga olinib, quyidagi semantik guruhlariga taqsimlandi:

1. Fan, sohalarga oid termin va tushunchalar.
2. Tashkilot, muassasa nomlari.
3. Adabiyot terminlari.
4. Shaxsni bildiruvchi otlar.
5. Narsa va buyum nomlari.
6. Ko‘plik morfemasini oluvchi mavhum otlar.
7. Faqat birlikda qo‘llanuvchi mavhum otlar.

¹³⁵ Bu haqdagi manba: Abjalova M., Toshimov R. Matnlarni lingvistik tahrirlashning psixologik asoslari. “O‘zbekistonning mustaqil taraqqiyot va ijtimoiy-falsafiy tafakkur yangilanishi” (Yosh olimlar ilmiy konferensiyasi materiallari), II qism – T. 2011. – B. 155-158.

8. Oziq-ovqat nomlari.
9. Odam tana a'zosi nomlari.
10. Kimyoviy modda nomlari.
11. O'simlik, tog' jinslari nomlari.
12. Millat, elat, qabila, urug' nomlari.
13. Jarayon nomlari.
14. Baliq nomlari.
15. Hayvon nomlari.
16. Voqea-hodisa nomlari.
17. Vaqt, faslni bildiruvchi nomlar.
18. Soha nomlari.
19. Qushlar, hashoratlar nomi.
20. Kasallik nomlari.
21. Joy, bayram nomlari.
22. Holat (*yo'qsil, megajin, onaxotin, jazman, ofatijon,...*), laqab.
23. Faqat inkor shaklli otlar.

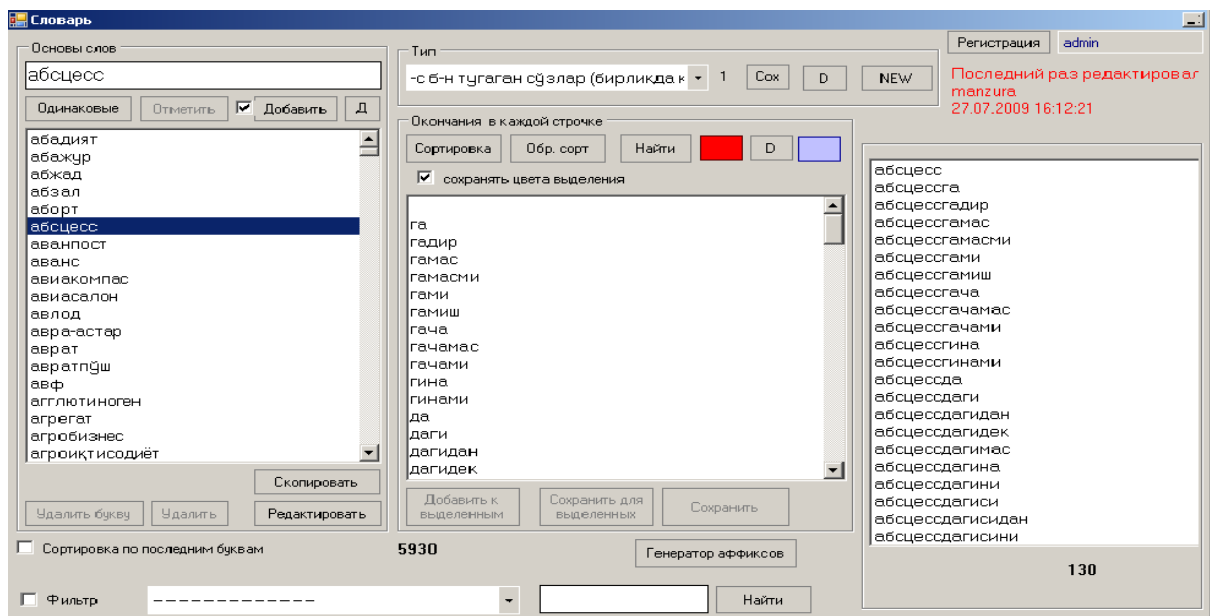
Turdosh otlar lisoniy maydon guruhiga yer osti va yer usti kon boyliklari nomlari (*oltin, kumush, qum, suv*), o'rin-joy nomlari (*qishloq, shahar, sahro*), o'simlik nomlari (*daraxt, buta, o't*), jihoz va buyum nomlari (*mebel, deraza, eshik, kitob, qalam*), vaqt nomlari (*soat, minut/daqiqqa, sekund/soniya, yil, oy, kun, tong, kech*), miqdor nomlari (*kilogramm/kg, gramm/gr, litr/l, botmon, misqol*), tashkilot, muassasa, idora nomlari (*maktab, idora, do'kon, maskan*), shaxs tavsifi nomlari (*chol, yigit, boy, bek, kampir*) kiritildi.

Guruhlarga ajratilgan so'zlarga affikslar kombinatsiyasi qatori tuziladi va ular so'zlarga birlashtirib chiqiladi. Bunda affikslarning semantik guruhlardagi so'zlarga birlashtirishning mantiqiy imkoniyatlariga asoslaniladi. Quyidagi tasvirda *absess* so'ziga affikslarning birlashtirilgan holati keltirildi. Birinchi bo'limda leksemaning so'zlar qatoridan tanlab olinishi, ikkinchi bo'limda tanlab olingan leksemaning amalda o'ziga birlashtirishi mumkin bo'lgan affikslar kombinatsiyasi, nihoyat, uchinchi bo'limda affikslarning tanlab olingan leksemaga birlashtirilgan holati berildi. Natija quyidagicha ko'rinish oldi:

Tanlab olingan leksema: *absess*

Affikslar kombinatsiyasi: 129 ta

Dasturga kiritilish holati: 130 ta (129+ 1(asos/lemma))



15-rasm. Lingvistik bazaning tahrirlash oynasi.

Bir soʻzshakl bir qancha soʻzning grammatik formasiga toʻgʻri kelishi mumkin. Bunday holda soʻzlarni farqlash uchun matn kirituvchiga shunday soʻzlar varianti koʻrsatiladi va matn teruvchi soʻzlardan muvofiqini tanlab qoʻyadi. Ushbu jarayon MTda ham kuzatiladi, yaʼni soʻzning xato yozilgani aniqlangach, matn teruvchiga tanlab qoʻyish imkonini beruvchi soʻzlar varianti taqdim etiladi.

Yuqorida keltirilgan fikrlar va tasvirlar amaliyot jarayoni sifatida berildi.

Ot turkumini modellashtirishda oʻzak soʻzlarga quyidagi affikslar qoʻshilishi eʼtiborga olindi:

- qarashlilik affiksi: $q_a = -niki$;
- oʻrin-joy oti affiksi: $\tilde{o}_j = -dagi$;
- chegaralash affiksi: $ch_a[3] = \{-gacha, -kacha, -qacha\}$;
- koʻplik affiksi: $k_a = -lar$;
- kelishik affikslari (variantlari bilan): $ke_a[7] = \{-ning, -ni, -ga, -ka, -qa, -da, -dan\}$;
- egalik affikslari: $e_a[6] = \{-m, -ng, -si, -miz, -ngiz, -lari\}$;
- egalik affikslari: $e_a'[6] = \{-im, -ing, -i, -imiz, -ingiz, -lari\}$;
- substantiv shakl yasovchi affiks: $sh_y = -lik$;
- inkor shaklni hosil qiluvchi affiks: $[inkor] = -siz$;
- affiks-yuklama: $aff.Part[14] = -chi, -a, -ya, -ku, -da, -u, -yu, -mi, -gina, -kina, -qina, -oq, -yoq, -dir$.

Ushbu affikslar quyidagi ketma-ketlikda kombinatsiya hosil qiladi:

Asos – $Lemma_{ism} = L_{ism}$

S – so‘z yasovchi affiks

L_{ism}= kitob

L _{ism} + A1	derivation> kitoblar
L _{ism} + A2	derivation> kitobim
L _{ism} + A3	derivation> kitobni
L _{ism} + A5	derivation> kitobdagi
L _{ism} + S+A6	derivation> kitobxonniki
L _{ism} + A7	derivation> kitobgacha
L _{ism} + A8	derivation> sariqroq
L _{ism} + A9	derivation> kitobmi
L _{ism} + A10	derivation> kitobsiz
L _{ism} + A11	derivation> kitobgina
L _{ism} + A12	derivation> kitobligi
L _{ism} + A13	derivation> kitobdek
L _{ism} + A1+ A2	derivation> kitoblarim
L _{ism} + A1+ A2+A3	derivation> kitoblarimdan
L _{ism} + A1+ A2+A9	derivation> kitoblarimku
L _{ism} + A1+ A2+A3+A9	derivation> kitoblarimniku
L _{ism} + A1+ A2+A3+A10	derivation> kitoblarimnimas
L _{ism} + A1+ A2+A3+A10+A9	derivation> kitoblarimdanmasmi
L _{ism} + A5+ A1+A2+A3+A10+A9	derivation> kitobdagilarimdanmasdir
L _{ism} + S+A5+ A1+A2+A10+A9	derivation> kitobxondagilardanmasku

Algoritmlar ketma-ketligi shu tartibda davom ettiriladi.

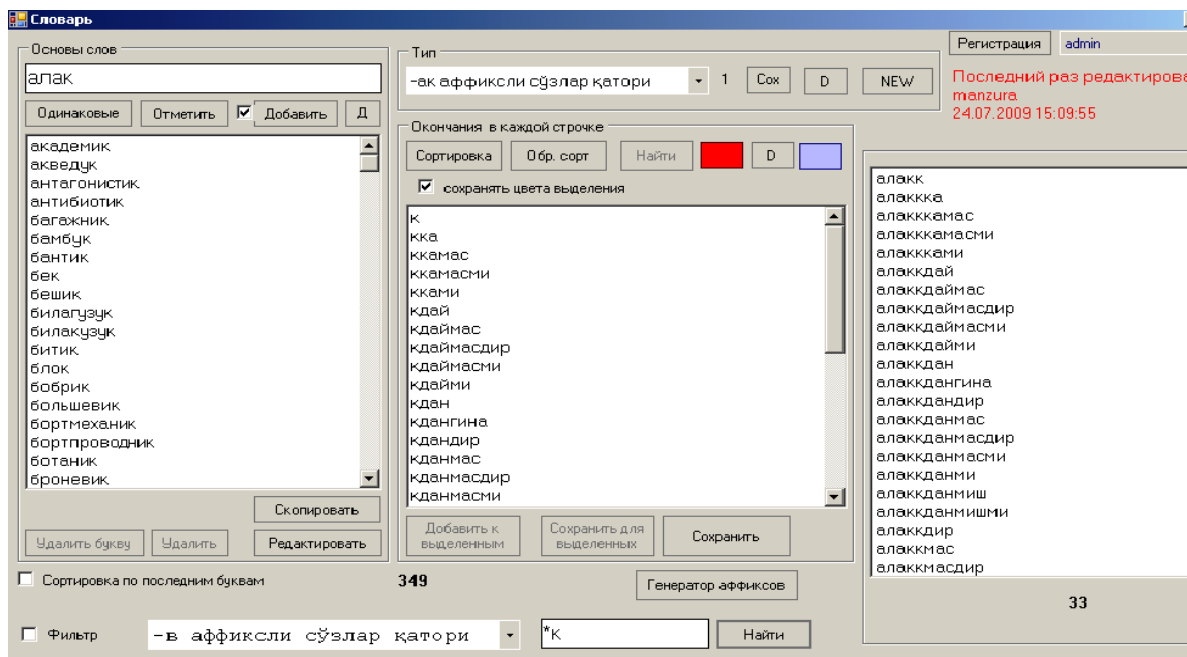
Orfografik qoidalar

Qoida: “i” unlisi bilan boshlanadigan lug‘aviy shakl yasovchi affikslar o‘zidan oldingi fonemaning undosh bo‘lishini talab qiladi. Masalan: *kitob+im, ko‘z+ing, nutq+imiz, stol+ingiz*

-k va -q tovushlari bilan tugaydigan so‘zlarga -k va -q tovushlari bilan boshlanadigan (-ka, -qa, -kacha, -qacha) affikslarning qo‘shilish algoritmi:

Qoida. L – sifatida A asos harflari soni olinadi (L=lenght(A)). S – L o‘lchamli A asosning harflarini o‘zida saqlovchi massiv (S[L] – A asosning oxirgi harfini o‘zida ifodalaydi).

Qoida. Agar (S[L]=k) tenglik o‘rinli bo‘lsa, ya’ni asosning oxirgi harfi *-k* tovushi bilan tugasa, A – asosga *k*- tovushi bilan boshlanadigan affikslar qo‘shiladi (16-rasm).



16-rasm. Dasturda k harfi bilan tugaydigan so‘zlarga affiks qo‘shilishi.

Qoida. Agar S[L]=q tenglik o‘rinli, ya’ni asosning oxirgi harfi *-q* tovushi bilan tugagan bo‘lsa, A – asosga *q*- tovushi bilan boshlanadigan affikslar qo‘shiladi. M-n: *bilak* ∪ *kacha* ⇒ *bilakkacha*, *buloq* ∪ *qacha* ⇒ *buloqqacha*, *ko‘k* ∪ *ka* ⇒ *ko‘kka*, *kurak* ∪ *ka* ⇒ *kurakka*, *hashak* ∪ *ka* ⇒ *hashakka*

-k / -q tovushlari bilan tugaydigan so‘zlarga egalik affikslarining qo‘shilish algoritmi

Qoida. Agar (S[L]=k) tenglik o‘rinli bo‘lsa, ya’ni asosning oxirgi harfi *-k* tovushi bilan tugasa A – asosning oxirgi harfi *-g* tovushiga o‘zgaradi. M-n: *bilak* ∪ *im* ⇒ *bilag* ∪ *im* = *bilagim*, *tilak* ∪ *im* ⇒ *tilag* ∪ *im* = *tilagim*, *chelak* ∪ *im* ⇒ *chelag* ∪ *im* = *chelagim* va h.k.

Qoida. Agar S[L]=q tenglik o‘rinli bo‘lsa, ya’ni asosning oxirgi tovushi *-q* tovushi bilan tugasa, A – asosning oxirgi tovushi *-g*‘ tovushiga o‘zgaradi. Misol sifatida quyidagilarni ko‘ramiz: M-n: *buloq* ∪ *i* ⇒ *bulog* ∪ *i* = *bulog‘i*, *buzoq* ∪ *i* ⇒ *buzog* ∪ *i* = *buzog‘i*, *tuzoq* ∪ *imiz* ⇒ *tuzog* ∪ *imiz* = *tuzog‘imiz*, *suvoq* ∪ *i* ⇒ *suvog* ∪ *i* = *suvog‘i* va h.k.

Bunday tovush o'zgarishiga uchraydigan so'zlar guruhlashtiriladi va amalda qo'llaniladigan affikslar kombinatsiyalari algoritm orqali ularga qo'shiladi.

Qoida. Shunday so'zlar borki, ular q/k tovushi bilan tugashiga qaramasdan tovush almashinuviga uchramaydi, $S[L]=q$ ($S[L]=k$) tenglik o'rinli bo'lsa, ya'ni asosning oxirgi harfi $-q$ ($-k$) tovushi bilan tugasa ham, A – asosning oxirgi harfi o'zgarishsiz qoladi. M-n: $ravnaq+i \Rightarrow ravnaqi$, $axloq+ing \Rightarrow axloqing$, $ishtirok+imiz \Rightarrow ishtirokimiz$, $ocherk+i \Rightarrow ocherki$.

$[nok / chok / park / erk / ko'rk / idrok / huquq / nutq / zavq] \cup [e_a[6]=\{-im, -ing, -i, -imiz, -ingiz, -lari\}] = S[L]=q$ ($S[L]=k$). Bunday so'zlar alohida guruhga jamlanib, affikslar kombinatsiyasi biriktiriladi.

Egalik affikslari qo'shilganda asosda tovush tushirilishi algoritmi

Qoida. A – asosga bevosita egalik affikslari quyidagi qoida asosida qo'shiladi:

A asosning $L-1$ – harfi, ya'ni $S[L-1]$ harf o'chiriladi va oxirgi L – harfi o'chirilgan $L-1$ – harf o'rniga qo'yiladi. Natijada L uzunlikdagi asos $L-1$ uzunlakka ega bo'ladi, egalik affiksi qo'shilgandan keyin N natijaviy so'zshakl hosil bo'ladi. Misol: $shahar \cup im \Rightarrow shahr \cup im \Rightarrow shahrim$, $burun \cup i \Rightarrow burn \cup i \Rightarrow burni$, $qorin \cup i \Rightarrow qorni$ va h.k.

$[o'g'il / o'rin / egin / og'iz / bo'yin / bag'ir / qo'yin / singil / ko'ngil / bo'g'iz / yarim / zahar] \cup [e_a[5]=\{-im / -ing / -i / -imiz / -ingiz\}] \rightarrow S[L-1]$

Qoida. $parvo$, $obro'$, $avzo$ leksemalariga egalik affikslari qo'shilganda, “y” tovushi orttiriladi, ya'ni $[parvo / obro' / avzo] \cup [e_a[5]=\{-im / -ing / -i / -imiz / -ingiz\}] \rightarrow S[L+u]$.

Qoida. $mavzu$ leksemasiga egalik affikslari qo'shilganda, “y” tovushi orttiriladi, ya'ni $[mavzu] \cup [e_a[4]=\{-im / -ing / -imiz / -ingiz\}] \rightarrow S[L+y]$: $mavzuyim$, $mavzuying$, $mavzuyimiz$, $mavzuyingiz$. Qolgan III shaxs egalik affikslarini o'zgarishsiz qabul qiladi, ya'ni $[mavzu] \cup [e_a[2]=\{-si / lari\}] \rightarrow N$: $mavzusi$, $mavzulari$ va h.k.

Ot turkumi paradigmatic va noparadigmatic shakllarining ketma-ket kelish me'yorlari

Qoida. $A \cup$ lug'aviy shakl yasovchi affikslar \cup sintaktik shakl yasovchi affikslar. Misol: $uka \cup jon \cup im$, $ona \cup gina \cup m \cup ni$, $qo'zi \cup choq \cup qa$

Qoida. A asosga shakl hosil qiluvchi affikslar hamisha qo'shilaveradi.

Qoida. e_a[11] A asos yoki lug‘aviy shakl hosil qiluvchi affikslardan so‘ng qo‘llaniladi: *kitob*∪*im*, *dada*∪*m*, *sumka*∪*cha*∪*ng*, *buvi*∪*jon*∪*ingiz* va h.k.

3.1.2. Fe’lni formallashtirish moduli

Fe’l turkumi boshqa so‘z turkumlariga nisbatan boy va murakkab grammatik kategoriyalar, shakllar sistemasiga ega. Ushbu izlanishda mavjud nazariy adabiyotlarga tayangan holda (avtomatik tahrir dasturlari uchun) fe’llarni formallashtirishga e’tibor qaratildi (ilovada 6-jadval).

Ma’lumki, adabiyotlarda fe’llar kategorial va funksional shakllarga ajratiladi. Kategorial shakllar fe’lning ma’lum bir grammatik kategoriyasiga xos grammatik ma’nolarni ifodalaydi: nisbat, mayl, zamon va shaxs-son kategoriyalariga xos shakllar shular sirasidandir. Funksional shakllar fe’lning turli gap bo‘laklari vazifasida kelishiga xoslanadi. Ularga sof fe’l shakli, harakat nomi shakli, sifatdosh va ravishdosh shakli kiradi.

Asos fe’l (lemma, o‘zak, etakchi morfema) V ning kategorial shaklini hosil qilishda to‘rt turdagi morfemalar guruhi e’tiborga olindi:

1. Nisbat kategoriyasi (x).
2. Shaxs-son kategoriyasi (u).
3. Zamon kategoriyasi (z).
4. Mayl kategoriyasi (m).

Fe’l asoslariga qo‘shilib, gapda uni kesim vazifasida qo‘llanilishiga xoslovchi shakllar fe’lning munosabat yoki sintaktik shakl yasovchi shakllari hisoblanadi, ya’ni bunday shakllardagi fe’l gapda kesim vazifasida keladi (sintaktik tahlil moduli bosqichi). Yuqoridagi to‘rt turdan shaxs-son, zamon, mayl kategoriyalari munosabat shakllaridir. Ushbu shakllardan holi fe’l asosi fe’lning noaniq shakli, ya’ni fe’l lemma hisoblanadi.

Formal grammatika nuqtayi nazaridan kategorial shakllarni quyidagicha formallashtirish mumkin:

1. Nisbat kategoriyasi – x

1.1 Aniq nisbat – x_1 , \emptyset morfema: *ishla*, *qil*, *o‘qi*, *ko‘r*, *kel*, *och*, *chiq*, *bo‘ya*, *cho‘z*, *yuv* va h.k.

- 1.2 O‘zlik nisbat – x_2

$x_2(1.1) = -n$; $x_2(1.2) = -in$: *taran(di)*, *yuvin(di)*;

$x_2(2.1) = -l$; $x_2(2.2) = -il$: *bo‘yal(di)*, *cho‘zil(di)*.

- 1.3 Orttirma nisbat – x_3

$x_3(1) = -tir$: *buyurtir*, *o‘stir*;

$x_3(2) = -dir$: *qildir*, *ochdir*;

$x_3(3) = -ir$: *ochir, ko'chir*;

$x_3(4) = -ar$: *qaytar, chiqar*;

$x_3(5) = -iz$: *oqiz, tomiz*;

$x_3(6) = -giz, -g'iz, -g'az$. Jarangli undosh¹³⁶ bilan tugagan fe'lga qo'shiladi: *ko'rgiz, yurg'iz, turg'az*;

$x_3(7) = -kiz, -qaz, -kaz$. Jarangsiz¹³⁷ undosh bilan tugagan fe'lga qo'shiladi: *ketkiz, o'tqaz, o'tkaz*;

$x_3(8) = -qiz$: *yutqiz*;

$x_3(9) = -sat$: *ko'rsat*;

$x_3(10) = -(i)t$: *to'lat, ishlat, o'qit, kirit, yurit*.

1.4 Birgalik nisbati – x_4

$X_4(1.1) = -sh$; $X_4(1.2) = -ish$: *ishlash(di), ko'rish(di), turish(di)*.

1.5 Majhul nisbat – x_5

$x_5(1.1) = -n$; $x_5(1.2) = -in$: *ishlan(di), qilin(di)*;

$x_5(2.1) = -l$; $x_5(2.2) = -il$: *o'qil(di), ko'ril(di)*.

(Fe'lning munosabat shakllari)

2. Shaxs-son kategoriyasi (y). Hozirgi o'zbek adabiy tilida shaxs-son affikslarining uch asosiy turi bo'lib, ilmiy tadqiqot ishida ular to'rt tur sifatida (uchinchi turning inkor shakli bilan) olindi.

y_1 ning shakllari:

Shaxs-son asosi	shaxs	birlik		ko'plik	
		affiksi	modeli	affiksi	modeli
1) o'tgan zamon xabar mayli	I	-man	$y_1\text{bir}1$	-(i)miz	$y_1\text{ko'p}1$
	II	-san	$y_1\text{bir}2$	-(i)siz	$y_1\text{ko'p}2$
2) hozirgi zamon xabar mayli 3) kelasi zamon xabar mayli	III	-di	$y_1\text{bir}3$	-(di)lar	$y_1\text{ko'p}3$

y_2 ning shakllari:

Shaxs-son asosi	shaxs	birlik		ko'plik	
		affiksi	modeli	affiksi	modeli
1) o'tgan zamon xabar maylining -di shakli	I	-(i)m	$y_2\text{bir}1$	-(i)k	$y_2\text{ko'p}1$
	II	-(i)ng	$y_2\text{bir}2$	-(i)ngiz	$y_2\text{ko'p}2$
2) -sa kelasi zamon shart mayli	III	Ø	$y_2\text{bir}3$	-(di)lar	$y_2\text{ko'p}3$

¹³⁶ Illovada 2-jadval

¹³⁷ O'sha jadvalda

Qoida. Mazkur shakllar fe'l (V)ning o'tgan zamon $z_1(1)=-di$ va $z_3(9)=-sa$ shakllaridan so'ng qo'llanadi, ya'ni $V z_1 \cup [y_2[9]=\{-m/-im, -ng/-ing, -k/-ik, -ngiz/-ingiz, -lar\}] \rightarrow Vy_2: ishladim, ishlading, ishladik, ishladingiz, ishladilar$; $V z_3 \cup [y_2[9]=\{-m/-im, -ng/-ing, -k/-ik, -ngiz/-ingiz, -lar\}] \rightarrow Vy_2: ishlasam, ishlasang, ishlasa, ishlasak, ishlasangiz, ishlasalar$ va h.k.

y_3 ning shakllari:

Shaxs-son asosi	shaxs	birlik		ko'plik	
		affiksi	modeli	affiksi	modeli
kelasi zamon buyruq-istak mayli (tub fe'l/lemma)	I	-(a)y	y_3bir1	-(a)yluk	$y_3ko'p1$
	II	-gin, Ø	y_3bir2	-(i)ng, -(i)nglar	$y_3ko'p2$
	III	-sin	y_3bir3	-sin(lar)	$y_3ko'p3$

y_4 ning shakllari:

shaxs	birlik		ko'plik	
	affiksi	modeli	affiksi	modeli
I	-im yo'q	y_4bir1	-imiz yo'q	$y_4ko'p1$
II	-ing yo'q	y_4bir2	-ingiz yo'q	$y_4ko'p2$
III	-i yo'q	y_4bir3	-lari yo'q	$y_4ko'p3$

3. Zamon kategoriyasi – z

z_1 – o'tgan zamon

$z_1(1)=-di: ishladi, qildi, o'qidi, ko'rdi, keldi$ va h.k.;

$z_1(2)= -gan/ -kan/ -qan: ishlagan, ekkkan, chiqqan$;

$z_1(3)= -gan edi: ishlagan edi, qilgan edi, o'qigan edi, ko'rgan edi, kelgan edi$;

$z_1(4)= -gandi: ishlagandi, qilgandi, o'qigandi, ko'rgandi, kelgandi$;

$z_1(5)= -ar edi: ishlar edi, qilar edi, o'qigan edi, ko'rgan edi, kelgan edi$;

$z_1(6)= -mas edi: ishlamas edi, qilmas edi, o'qimas edi, ko'rmas edi, kelmas edi$;

$z_1(7)= -(a) yotgan edi: ishlayotgan edi, qilayotgan edi$;

$z_1(8)= -(a)yotgandi: ishlayotgandi, qilayotgandi, ko'rayotgandi, kelayotgandi$;

$z_1(9)= -moqda edi: ishlamogda edi, qilmoqda edi, ko'rmoqda edi, kelmoqda edi$;

$z_1(10)= -moqchi edi: ishlamogchi edi, qilmoqchi edi, ko'rmoqchi edi, kelmoqchi edi$;

$z_1(11)= -moqchi emas edi: ishlamogchi emas edi, qilmoqchi emas edi$;

$z_1(12)$ = -moqchi emasdi: *ishlamoqchi emasdi, qilmoqchi emasdi, kelmoqchimasdi*;

$z_1(13)$ = -b/ -ib: *ishlabman, ishlabsan, ishlabdi, ishlabmiz, ishlabsiz, ishlabdilar*.

z_2 – hozirgi zamon

$z_2(1)$ = -yap: *boryapman, boryapsan*;

$z_2(1^1)$ = -yapti: *ishlyapti, kelyapti, ko‘ryapti*;

$z_2(2)$ = -ayotib / -yotib: *borayotib, qilayotib, ishlayotib, boshlayotib*;

$z_2(3)$ = -moqda: *ishlamoqda, kelmoqda, qilmoqda, ko‘rmoqda*;

$z_2(4)$ = -yotir: *o‘qiyotir, qilayotir, berayotir*;

$z_2(5)$ = -(a)yotgan+(shaxs-sonning u_2 shakl affikslari) _yo‘q: *ishlayotgani yo‘q, kelayotganingiz yo‘q, qilayotganing yo‘q, o‘qiyotganimiz yo‘q*;

$z_2(6)$ =-(a)yotganlik (III shaxs ko‘plikda)+k_a+e_a /-(a)yotganligi: *ishlayotganligi, ishlayotganligim, ishlayotganliging, ishlayotganligimiz, ishlayotganligingiz, ishlayotganliklari*.

z_3 – kelasi zamon

$z_3(1)$ = -y: *ishlay, boray, o‘qiy*;

$z_3(1^1)$ = -a: *bora()di, ko‘ra(man), o‘qita(di), qurdira(miz)*;

$z_3(2)$ = -(a)r: *ishlar, o‘qir, kesar; kelar, borar*;

$z_3(3)$ = -moqchi: *ishlamoqchi, kelmoqchi, o‘qimoqchi, qilmoqchi, bo‘yamoqchi*;

$z_3(4)$ = -mas: *kelmas, ishlamas, qaytmas*;

$z_3(5)$ = -jak: *borajak, qilajak*;

$z_3(6)$ = -ur: *borur, kelur*;

$z_3(7)$ = -gusi / -g‘usi: *kelgusi, bo‘lg‘usi*;

$z_3(8)$ = -gay: *kelgay, borgay*;

$z_3(9)$ = -sa: *ishlasa(ng), kelsa(m), borsa(m)*.

4. Mayl shakllari (m).

Fe’l asosidagi harakat-holat maylning quyidagi to‘rt turi orqali ifodalanadi:

1. Xabar mayli: Ko‘rsatkichi yo‘q – \emptyset : *keldi, kelyapti, keladi*.

2. Buyruq-istak mayli. Fe’l-lemmaga -(a)y, -(a)yin, -gin, -(i)ng, -sin, -(a)lik, -(i)nglar affikslarini qo‘shish yo‘li bilan hosil qilinadi.

Qoida: undosh tovush bilan tugagan fe’l-lemmaga buyruq-istak mayli affikslarining -ay, -ayin, -gin, -ing, -sin, -alik, -inglar shakllari qo‘shiladi: *boray, borayin, borgin, boring, borsin, boraylik, boringlar*.

Qoida: unli tovush bilan tugagan fe’l-lemmaga buyruq-istak mayli affikslarining -y, -yin, -gin, -ng, -sin, -lik, -nglar shakllari qo‘shiladi: *sanay, sanayin, sanagin, sanang, sanasin, sanaylik, sananglar*. Inkor

shakli esa fe'l-lemmaga *-ma* affiksi qo'shib yasaladi: *borma, sanama* kabi.

3. Shart mayli fe'l-lemmada shaxs-son shaklidan oldin kelgan *-sa* (odatda fe'lshakldan so'ng *ham* va *-da* yuklamasi keladi), *-sa edi, -saydi* affikslari yordamida hosil qilinadi: *borsa, borsa edi, borsaydi*.

Fe'lning vazifa shakli. Fe'lning gapda turli vazifada kelishini ta'minlaydigan grammatik vositalar fe'lning vazifa shakli deyiladi. Fe'lning vazifa shakli 4 ta:

- 1) sof fe'l (fe'l-kesim);
- 2) ravishdosh (hol);
- 3) sifatdosh (sifatlovchi-aniqlovchi);
- 4) harakat nomi (ega, to'ldiruvchi).

O'z navbatida ularni ikkiga bo'lish mumkin:

- 1) kesimlik shakli: zamon, shaxs-son, mayl, bo'lishli-bo'lishsizlik;
- 2) nokesimlik shakli: ravishdosh, sifatdosh, harakat nomi.

Nokesimlik shakli. Fe'lning nokesimlik shakli bo'lgan ravishdosh ravish, sifatdosh sifat, harakat nomi ot vazifasini bajaradi.

V(r) – ravishdosh.

V(r)_{holat}: *-(i)b, -a/-y; -gancha/-kancha/-qancha; -gudek/-kudek/ -qudek: o'qib, yozib, sevinib, yura-yura, yig'lay-yig'lay, eb qo'ygudek, bitir- ib yuborgudek, shoshgancha.*

Qoida: *-a/-y* ko'rsatkichli ravishdosh takror holda qo'llanadi: *bora-bora, chopacha-chopacha*. Uning bo'lishsiz shakli yakka qo'llanaveradi.

V(r)_{payt}: *-gach/-kach, -guncha/-kuncha: ko'rgach, tikkach, kelguncha; kelgach, chiqqach, borguncha, kechikkuncha*

V(r)_{maqsad}: *-gani/-kani/-qani (ba'zan -gali/-kali/-qali): maslahatlashgani.*

Qoida: ravishdoshning barcha ko'rsatkichi uchun *-may (in)/ -masdan* bo'lishsiz shakl. Faqat maqsad ravishdoshi bo'lishsiz shaklga ega emas.

V(s) – sifatdosh.

V(s)_{o'tgan}: *-gan/-kan/-qan (kelgan odam, kechikkan o'quvchi, namiqqan gilam).*

Qoida: *-mish* o'tgan zamon sifatdoshining kam uchraydigan shakli: *O'zbek adabiy tilining asoschisi hisoblanmish Navoiy...*

Qoida: hozirgi va kelasi zamon sifatdoshi quyidagicha hosil bo'ladi:

1) *-a* shaklli ravishdoshga *-yotgan* affiksini qo'shish bilan. Bunda *-y* tushib qoladi: *o'qiyotgan bola, chopayotgan it;*

2) -a/-y shakliga *-digan* affiksini qo‘shish orqali: *bor-a-digan*, *o‘qi-y-digan*;

3) -(a)r, -mas: *oqar suv*, *borar joyimiz*, *so‘nmas hayot*;

4) -(g)(u)vchi: *ishlovchi*, *tanuvchi*, *xohlovchi*, *boruvchi*, *uyg‘onguvchi*;

5) -g‘lik, -rlik, -gulik: *yasatig‘lik xontaxta*, *yopig‘lik qozon*, *qorin to‘ydirarlik bir kasb*, *arzigulik ish*.

Qoida: -(u)vchi, -g‘lik, -rlik, -gulik shakli uchun *-ma+gan* ko‘rsatkichi bo‘lishsiz shakl: *yasatig‘lik – yasatilmagan*, *to‘ydirarlik – to‘ydirmaydigan*; sifatdoshning -(g)(u)vchi, -g‘lik, -rlik, -gulik shakli bo‘lishsiz qo‘llanmaydi.

V(s)_{istak}: -gur/-kur, -qur/-g‘ur: *bo‘yning uzilgur bola*, *qurg‘ur yigit*, *bo‘lmag‘ur gap*.

V(h) – harakat nomi.

Qoida: harakat nomi otning shakl yasovchisini olib, otga xos sintaktik vazifani bajaradi.

Grammatik shakllari:

-moq: *o‘qimoq*, *yoqmoq*, *tanishmoq* (*lug‘atlarda fe‘lning noaniq shaklini beradi*);

-(i)sh: *o‘qish*, *yoqish*, *tanishish*;

-(u)v: *o‘quv* (*o‘qi+v*), *yoquv*, *tanishuv*.

-r/-ar: *ucharga qanot*, *turarga joy*, *ko‘rarga ko‘z*.

-r/-ar qo‘shimchasi sifatdoshdagi shunday qo‘shimcha bilan shakldosh.

-maslik (*o‘qish - o‘qimaslik*) – bo‘lishsizlik shakli

Qoida: -moq va -(i)sh qo‘shimchasi -lik so‘z yasovchisi bilan qo‘llanadi: *o‘qimoqlik*, *borishlik*, *bormaslik*, *ketishlik*, *kelmoqlik*.

Fe‘l shaklining umumiy yasaliş qoidasi

Fe‘llarni ikki guruhga bo‘lish mumkin:

1) tub fe‘llar: *yurmoq*, *o‘qimoq*;

2) yasama fe‘llar: *tozala*, *gapirmoq*, *sotib olmoq*.

Yasama fe‘llarni hosil qiluvchi affikslar 3.4-bo‘limda jadval asosida berildi.

Qoida: 1-2-3-4-turdagi morflar doimo fe‘l shaklini hosil qilishda ishtirok etadi.

Istisno qoidasi: u formaning {-(a)y, -gin, -sin, -(a)ylik, -(i)ng, -(i)nglar, -sin(lar)} shakllari}=u₃ qo‘llanganda 3-turdagi morflar (zamon affikslari) ishtirok etmaydi: *ishlay*, *boray*, *kelsin*, *ketgin*, *quraylik*, *tuzing*, *ishlanglar*, *buyursinlar* va h.k.

Qoida: 1-tur morflari (x) doimo asos fe'l (V)dan keyin qo'shiladi: $to'la(V) \cup -t(x_3(10)) = to'lat; orttir, bildir, ko'rsat, beril, taran$ va h.k.

Qoida: Orttirma nisbat – x_3 morflari V ga birdan ortiq qo'shilishi mumkin: $to'la+t+tir, o'qi+t+tir, to'l+dir+giz$ va h.k.

Qoida: 2-tur morflari doimo 3-tur morflaridan keyin qo'shiladi.

Qoida: Gapda ishtirok etadigan fe'lni quyidagi qolipda ifodalash mumkin:

$$V * x \cup z \cup u = Vx \cup z \cup u$$

$$Vx \cup z \cup u = V$$

Qoida: Ko'makchi fe'lli so'z qo'shilmasida 27 ta ko'makchi fe'ldan 25 tasi $-(i)b$, 11 tasi $-a/-y$, 9 tasi har ikkala, 16 tasi faqat $-(i)b$, 2 tasi faqat $-a/-y$ shakli yordamida (yoz, boshla) etakchi fe'lga birika oladi (7-jadval).

Qoida: Fe'lning murakkab shakli quyidagicha yasaladi:

Murakkab shakl	Tarkibi	Ma'nosi	Misol
-sa kerak	shart mayli shakli + modal so'z	gumon, taxmin	<i>borsa kerak tugatsam kerak</i>
-gan bo'l	sifatdosh shakli + <i>bo'l</i> fe'li	yuzaki bajarish	<i>qatnagan bo'ldi achinayotgan bo'ldi</i>
-ganim bo'lsin	sifatdosh shakli + egalik affiksi + <i>bo'l</i> fe'li	jazm qilish	<i>borganim bo'lsin qilmaganim bo'lsin</i>
-(a)y de	buyruq shakli + <i>de</i> fe'li	harakat chegarasi	<i>tugay dedi uchiray dedi</i>
-gisi kel	-gi affiksi + egalik shakli + <i>kel</i> fe'li	harakat istagi	<i>borgisi keldi ko'rgisi keladi</i>
-gan	sifatdosh va ravishdosh takrorlanishi	surunkali davomiylik	<i>o'qigani o'qigan yurgani yurgan kengaygandan kengaydi</i>

I. Fe'l shaklini hosil qilishda morflarning mos kelish va mos kelmaslik qoidasi

Qoida: 2-tur morfining $u_1 = \{-man, -san, -di; -miz, -siz, -di(lar)\}$ shakli 3-tur morfining $z_1(2), z_1(13)$ shakllari, z_2 zamonning $z_2(1), z_2(2), z_2(3), z_2(4)$ shakllari va z_3 zamonning $z_3(1), z_3(1^1), z_3(2), z_3(3), z_3(4), z_3(5), z_3(6), z_3(8)$ shakllaridan keyin qo'llanadi: *taqdirlanganman, ko'ribsan; kurashyapmiz, borayotibmiz, kelmoqdasiz, qurayotirlar; ishlayman, borasan, kelarman, tashlamoqchisan, qaytmasman, borajakmiz, borurmiz, kelgaymiz* va h.k.

Qoida: 2-tur morfining $u_2 = \{-(i)m, -(i)ng, -(i)k, -(i)ngiz, -dilar\}$ shakli 3-tur morfining $z_1(1), z_1(4), z_1(5), z_1(6), z_1(7), z_1(8), z_1(9), z_1(10), z_1(11), z_1(12)$ shakllari, z_2 zamonning $z_2(5), z_2(6)$ shakllari va z_3 zamonning $z_3(9)$ shakllari bilan birika oladi: *keldim, kelgandim, kelar eding, kelmas edik, kelayotgan edingiz, kelayotgandim, kelmoqda edilar, kelmoqchi eding, kelmoqchi emas edik, kelmoqchi emasdim; kelayotganliging, kelayotganim yo'q; kelsang* va h.k.

Qoida. $z_1(13) = -b/ -ib$ morfi bilan yasalgan o'tgan zamon shaklidan so'ng shaxs-son kategoriyasining z_1 shakli qo'llaniladi, ya'ni $V_{u_1(13)} = [-b/ -ib] \cup z_1$: *ishlabman, ishlabsan, ishlabdi, ishlabmiz, ishlabsiz, ishlabdilar* va h.k.

Qoida: edi, ekan, emish, emas to'liqsiz fe'llaridan so'ng 2-tur morfning y_2 va y_1 shakllari qo'llanishi mumkin, ya'ni $V \cup V_{to'liqsiz} \rightarrow V \Delta [edi, ekan, emish, emas] \cup u$: *ishlagan ekansan, sudlangan ekansiz, qurar emishsiz, turar ekanmiz* va h.k.

Qoida: 2-tur morfning variantlari (z_1, z_2, z_3, z_4) hech qachon bir-biri bilan birikmaydi.

Qoida: 1-tur morfning x_4 (birgalik nisbati) va x_5 (majhul nisbat) affikslari doim 2-tur hamda 3-tur morflari bilan birga qo'shiladi:

$V_{x_4} \cup z; V_{x_4} \cup z \cup u$: *ko'r+ish+di, qur+ish+gan; kel+ish+di+m, qur+ish+di+k* va h.k.

$V_{x_5} \cup z; V_{x_5} \cup u; V_{x_5} \cup z \cup u$: *ishla+n+gan, buyur+il+di; qor+il+sin, ber+il+sin* va h.k.

Qoida: 1-tur morfning x_1 (aniq nisbati), x_2 (o'zlik nisbat) va x_3 (orttirma nisbat) affikslari 2-tur hamda 3-tur morflarisiz qo'llana oladi: *ko'r, qur, ishla; ko'rin, taran, yasan; ko'rsat, ishlat, qildir* va h.k.

Qoida: 1-tur morfning x_2 (o'zlik nisbat), x_3 (orttirma nisbat) va x_5 (majhul nisbat) affikslari fe'l yasovchi affikslardan oldin qo'shilishi mumkin:

$V_{x_2,3,5} \cup V_{morfems} \rightarrow$ *maqta+n+choq, kuy+in+chak; qur+il+ma, qaz+il+ma; achi+t+qi, ko'r+gaz+ma* va h.k.

Qoida: 1-tur morfning variantlari (x_1, x_2, x_3, x_4, x_5) ketma-ket birikadi: *beza+n+tir+il(di)*, *ishla+t+tir+ish(di)*, *qur+dir+tir+il(gan)* va h.k.

II. Morfonologik va fonologik qoidalar

Qoida: Morflarning ayrim shakllari qavs ichida berilgan yoki “/” belgisi bilan ajratilgan *i* va *a* unilari o‘zidan oldingi fonemaning undosh bo‘lishini talab qiladi. Masalan: *bit+ir*, *bor+ayotgan*, *qil+in+ayotgan*, *ko‘r+il(di)* va h.k.

Qoida: z_3 ning $z_3(1)$ formasi (ya’ni *-y*) so‘z yasovchi *-la* affiksi bilan yoki unli fonema bilan birikadi: *ishlay*, *gullay*, *to‘lay*, *bitay*, *tugatay* va h.k.

Qoida: *-la* bilan tugagan fe’llar hech qachon $z_3(1')$ bilan birikmaydi.

Qoida: z_3 ning $z_3(1')$ formasi (ya’ni *-a*) faqat undosh harf bilan birika oladi: *qiynal+a(di)*, *bor+a(man)*, *ko‘r+a(di)* va h.k.

III. Orfografik qoidalar

Qoida. unli tovush bilan tugagan (*ishla*, *o‘qi*, *kuyla*, *boshla* kabi) tub fe’l (V)ga y_1 ning shakllari qo‘shilganda, *y* tovushi orttiriladi, ya’ni [*ishla/o‘qi*] \cup [$y_1[6]=\{-man, -san, -di, -miz, -siz, -dilar\}$] \rightarrow V[L+ *y*]: *ishlayman*, *ishlaysan*, *ishlaydi*; *ishlaymiz*, *ishlaysiz*, *ishlaydilar* va h.k.

Qoida: *edi*, *ekan*, *emish*, *emas* to‘liqsiz fe’llari ajratib yoziladi, ya’ni $V \cup$ to‘liqsiz fe’llar \rightarrow $V\Delta[edi, ekan, emish, emas]$: *borgan edi*, *ishlar ekan*, *tinmas emish*, *ko‘rkam ekan* va h.k.

Qoida: To‘liqsiz fe’llar boshidagi *e* unlisi tushib qolganda, har ikkala fe’l qo‘shib yoziladi va ular $V+V_1$ ning ekvivalenti hisoblanadi. Ekvivalent holatlar ishda (...) tarzida berilgan.

Qoida: Ravishdosh bo‘lgan biror felning birikishidan yasalgan qo‘shma fe’llar ajratib yoziladi. Uning qolipi quyidagicha:

$$\left. \begin{array}{l} V + -a \cup V_2 + \dots \\ V + -y \cup V_2 + \dots \\ V + -b \cup V_2 + \dots \\ V + -ib \cup V_2 + \dots \end{array} \right\} \longrightarrow V_4$$

Istisno qoidasi: qo‘shma fe’lning ikkinchi qismi (V_2) *olmoq* fe’li bo‘lsa, birinchi qism (V) *a* tovushi bilan tugagan bo‘lsa u holda *a* tushib qolishi va ikkita fe’l qo‘shilib yozilishi mumkin. Uning qolipi quyidagicha:

$$V + a \cap \cup ol + \dots \Rightarrow V_3$$

Misol: $V = yoz; yoz + a \wedge ol + di + m \Rightarrow yozoldim$

Qoida. *yemoq, demoq* fe'l shakllariga *-il* majhul nisbat va *-ish* birgalik nisbat affikslari qo'shilganda asosdan so'ng bir *-y-* tovushi orttiriladi, ya'ni $V(e, de) \cup [-il/-ish] \rightarrow [\{e/de\} + y \cup \{il/ish\}] \rightarrow yeyil, yeyish, deyil, deyish$.

Qoida. *a* va *i* unlilari bilan tugagan fe'llardan *-v* affiksi yordamida harakat nomi yasalganda, *a* tovushi *o* tovushiga va *i* tovushi *u* tovushiga almashadi. Misol: *sayla* → *saylov*, *o'qi* → *o'quv*, *to'qi* → *to'quv* va h.k.

Qoida. Agar $(V[L]=k)$ tenglik o'rinli bo'lsa, ya'ni asosning oxirgi harfi *-k* tovushi bilan tugasa V – asos fe'lga *-gan* sifatdosh yasovchi affiks, *-gach*, *-guncha*, *-gudek*, *-gancha*, *-gani* ravishdosh yasovchi affikslarining *k-* tovushi bilan boshlanadigan varianti birikadi, ya'ni $(V[L]=k) \cup [kan / kach / kuncha / kudek / kancha / kani]$. Masalan: *ekkan*, *birikkach*, *tikkancha* va h.k.

Qoida. Agar $(V[L]=q)$ tenglik o'rinli bo'lsa, ya'ni asosning oxirgi harfi *-q* tovushi bilan tugasa, yuqorida ko'rsatilgan affikslarning *q-* tovushi bilan boshlanadigan varianti birikadi, ya'ni $(V[L]=q) \cup [qan / qach / quncha / qudek / qancha / qani]$. Masalan: *boqqan*, *yoqquncha*.

3.1.3. Sifatni formallashtirish moduli

Sifat turkumi (Adj)ga xos leksemalar, asosan predmetning, qisman harakatning belgisini bildiradi va darajalanuvchi leksemalar hisoblanadi: *oq ko'ylak*, *oqroq mato*, *oppoq paxta*.

Boshqa so'z turkumlari nutq jarayonida belgini ifodalasa sifat barqaror va turg'un belgi ifodalashi jihatidan ulardan ajralib turadi¹³⁸.

Sifatning belgi bildirishi, darajalanishi uning morfologik xususiyati hisoblansa, gapda, asosan, sifatlovchi-aniqlovchi, ba'zan kesim, ayrim hollarda hol vazifasida kelishi sintaktik xususiyati sanaladi. Bu xususiyatlar sifatlarning morfologik va sintaktik tahlilida muhim ahamiyat kasb etadi.

Belgi xususiyatiga ko'ra sifatning asliy va nisbiy sifatlarga ajratilishi ta'minotda o'z o'rniga ega bo'lib, bu holat darajalanuvchi sifatlar bazasini yaratishga imkon beradi, ya'ni mavjud sifatlar ikki: asliy va nisbiy sifatlar guruhiga kiritiladi. Ma'lumki, asliy sifatlar narsa-predmetning turg'un, nisbatlanmaydigan va darajalanadigan belgisini ifodalaydi: *qizil*, *sariq*, *kichik*, *oz*¹³⁹. Shundan so'ng asliy sifatlarning darajalanish modullari

¹³⁸ Sayfullayeva R. va boshq. Hozirgi o'zbek adabiy tili. – T., Fan va texnologiya, 2009. – B.237.

¹³⁹ Sayfullayeva R. va boshq. Hozirgi o'zbek adabiy tili. – T., Fan va texnologiya, 2009. – B.237.

beriladi. Nisbiy sifatlar darajalanish, qiyoslanishga ega bo'lmagani va belgini qandaydir bir predmet, belgi, miqdor, o'rin va harakatga nisbatan ifodalashi e'tiborga olinib, darajalanmaydigan sifatlar tarzida modullashtirildi. Imlo lug'atga kiritilmagan, matnda uchraydigan sifatlar (Adj) dissertatsiya ilovasining 8-jadvalida berildi.

Asliy sifat belgining darajasini ifodalashga binoan quyidagi darajalarga ajratildi:

1)Oddiy daraja. Hech qanday vositasiz – Ø

2)Qiyosiy daraja. Yasovchi vositasi: -roq. Adj + [-roq] = qiyosiy sifat

3)Orttirma daraja. Yasalishi:

Qoida: *eng, juda, g'irt, jiqqa, tim, naq, zap, behad, nihoyatda, beqiyos, bag'oyat, o'ta* so'zlarini sifatlardan oldin keltirish orqali: *eng qisqa, juda katta, nihoyatda shinam* va h.k. Bunday so'zlar sifatlardan ajratib yoziladi.

Qoida: Rang-tusni bildiruvchi sifatning boshlang'ich tovushlarini olib, unga *p, m, s* tovushini qo'shish va yonidan chiziqcha bilan sifatning o'zini keltirish orqali: *qop-qora, sap-sariq, ko'm-ko'k, bus-butun, yam-yashil qip-qizil* va h.k.

Qoida: Ozaytirma daraja. *-ish, -g'ish, -mtir, -imtir, -tob, -gina, -kina, -qina* affikslari sifatlarga biriktirib, shuningdek, *och, nim, tim, sal* so'zlarini rang-tus bildiruvchi leksemalar oldidan qo'llash orqali: *ko'kish, qizg'ish, oqimtir, qoramtir, yashiltob, kattagina; och qizil, sal qora, nim sariq*. Ayrim hollarda birdaniga har ikkala usul yordamida hosil qilinadi: *sal qoramtir*.

3.1.4. Sonni formallashtirish moduli

Son turkumi (Num)ga xos leksemalar, asosan, predmet-hodisa, qisman harakat-holatning aniq miqdori, sanog'i, tartibi, shuningdek, mavhum sanog'ini bildiradi va *necha, nechta, nechanchi, qancha, qanchadan, qanchalab* kabi so'roqlarga javob bo'ladi.

Sonlar raqamlarda quyidagicha ifodalanadi va ular tahrir tizimining grafematik tahrir bosqichida o'z aksini topadi:

1)arab raqami bilan: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 – Num_{arab};

2)rim raqami bilan – Num_{rim}:

a) birlik: I, II, III, IV, V, VI, VII, VIII, IX;

b) o'nlik: X, XX, XXX, XL, L, LX, LXX, LXXX, XC;

v) yuzlik: C, CC, CCC, CD, D, DC, DCC, DCCC, CM;

g) minglik: M, MM, MMM...

Qoida: rim raqamlaridan so'ng chiziqcha qo'yilmaydi.

Qoida: sonlar ot turkumiga xos lug‘aviy va sintaktik shakl yasovchi affikslarni hamda ularning sintaktik vazifasini bajaradi: *o‘ninchisini, ikkinchigacha. O‘nlarcha(ni), uchala(ga)* sonlari bundan mustasno.

Sonlarning 0 dan 9 gacha bo‘lgan o‘n sanog‘idan hosil qilingan kombinatsiyasi cheksiz bo‘lganligi bois lingvistik ta‘minotga grammatik xususiyatiga ko‘ra miqdor va tartib sonlarning yasaliş qoidalari kiritildi.

Qoida: sanoq sonning ko‘rsatkichi yo‘q – \emptyset , ya‘ni Num \emptyset

Qoida: dona son sanoq songa *-ta* affiksini affiksini qo‘shish bilan hosil qilinadi, ya‘ni Num + [ta] = Num_{dona}: *ikkita, uchta, to‘rtta*.

Qoida: *-ta* affiksi bir so‘ziga qo‘shilganda *r* tovushi *t* bilan almashadi, ya‘ni Num bir + [ta] = bi[r→t]+[ta] = Numbitta: *o‘n bitta, yigirma bitta, o‘ttiz bitta* va h.k.

Qoida: chama son sanoq songa *-tacha, -lab, -larcha* affikslarini qo‘shish bilan yasaladi, ya‘ni Num+[-tacha/-lab/-larcha]=Num_{chama}: *ikkitacha, yuzlab, o‘nlarcha*.

Qoida: jamlovchi son sanoq songa *-ov, -ala, -ovlon* affikslarini qo‘shib yozish bilan yasaladi, ya‘ni Num+[-ov/-ovlon/ -ala] = Num_{jamlovchi}: *ikkov, beshovlon, uchala*.

Qoida: *-ov, -ala* affikslari faqat ikkidan o‘ngacha bo‘lgan sonlarga qo‘shiladi, ya‘ni Num [ikki/uch/to‘rt/besh/olti/etti/sakkiz/to‘qqiz/o‘n]+[-ov/ -ala] = Num_{jamlovchi}

Qoida: *-ov, -ala* affikslari unli (Voc) bilan tugagan sanoq sonlar (ikki, yetti, olti)ga qo‘shilganda asos qismidagi *i* unlisi tushib qoladi, ya‘ni Num asosning L-1 – harfi, ya‘ni S[L-1] harf o‘chiriladi va oxirgi L – harfi o‘chirilgan L-1 – harf o‘rniga qo‘yiladi. Natijada L uzunlikdagi asos L-1 uzunlakka ega bo‘ladi, jamlovchi son affiksi qo‘shilgandan keyin Num_{jamlovchi} natijaviy so‘zshakl hosil bo‘ladi. M-n: *ikki ∪ ala ⇒ ikk ∪ ala ⇒ ikkala, olti ∪ ov ⇒ olt ∪ ov ⇒ oltov, yetti ∪ ov ⇒ yett ∪ ov ⇒ yettov*.

Qoida: taqsim son sanoq songa *-tadan* affiksini qo‘shish orqali yasaladi, ya‘ni Num+[-tadan] = Numtaqsim: *ikkitadan, beshtadan, oltitadan*.

Qoida: tartib son sanoq songa *-inchi, -nchi* affiksini qo‘shish orqali yasaladi, ya‘ni Num+[-inchi/-nchi]=Num_{tartib}: *birinchi, ikkinchi, yetmish ikkinchi*.

Qoida: tartib sonlar arab (Num_{arab}) raqami bilan yozilganda *-inchi, -nchi* affiksi o‘rnida arab raqamidan so‘ng chiziqcha yoziladi, ya‘ni Num_{arab} ∪ [-inchi / -nchi → “-”] ⇒ Num_{arab}[1] ∪ [-inchi / -nchi → “-”] uy ⇒ 1-uy.

Istisno qoidasi: kirill yozuviga asoslangan imlo qoidasiga binoan sanani yozishda yil va oy nomlaridan oldin chiziqcha qo'yilmaydi, lotin yozuviga asoslangan yangi o'zbek yozuvi imlo qoidasiga asosan chiziqcha qo'yiladi. Masalan, *1 mapm, 2018 йил; 1-yanvar, 2019-yil* va h.k.

Qoida: tartib sonlar rim raqamlari bilan ifodalanganda chiziqcha qo'yilmaydi, ya'ni $\text{Num}_{\text{rim}} \cup [-\text{inchi}/-\text{nchi} \rightarrow \text{"Ø"}] \Rightarrow \text{Num}_{\text{rim}}[\text{XI}] \cup [-\text{inchi}/-\text{nchi} \rightarrow \text{"Ø"}] \text{ asr} \Rightarrow \text{XI asr}$.

Qoida: *kishi, fuqaro, odam, inson* kabi shaxs otlaridan oldin *nafar* hisob so'zi qo'llaniladi yoki sanoq son bilan shaxs otlari to'g'ridan-to'g'ri ishlatiladi, ya'ni Num nafar [*kishi, fuqaro, odam, inson, o'qituvchi, talaba, o'quvchi, chol, kampir*]: *ikki nafar odam, ikki nafar chol* va h.k.

Num [*kishi, fuqaro, odam, inson, o'qituvchi, talaba, o'quvchi, chol, kampir*]: *bir prezident, ikki hokim, yetti bola, uch qiz, bir deputat* va h.k.

Qoida: hayvonlarning turdosh otlari oldidan *bosh* so'zi qo'llaniladi, ya'ni Num bosh [*qo'y, mol, ho'kiz, echki, qo'zichoq*]: *bir bosh qo'y, ikki bosh ho'kiz* va h.k.

Qoida: narsa-buyum, predmet otlari oldidan *dona* hisob so'zi ishlatiladi, ya'ni Num dona [*narsa-buyum, predmet*]: *bir dona daftar, yetti dona piyola, uch dona kitob* va h.k.

Qoida: sonlarning tuzilishiga ko'ra juft va takror turlari chiziqcha bilan yoziladi, ya'ni $\text{Num}_x \cup \text{Num}_y \Rightarrow [\text{Num-Num}]_{\text{juft}}$; $\text{Num}_x \cup \text{Num}_x \Rightarrow [\text{Num-Num}]_{\text{takror}}$: *bir-ikki, besh-olti; ming-ming(lab), beshta-beshta* va h.k.

Qoida: sonlar otlashadi va ot turkumidagi so'zlarga birikadigan affikslarni oladi: $\text{Num} \rightarrow \text{n} \Rightarrow \text{Num}_n \cup \text{N}_{\text{aff}}$: *birdan, ikkinchini, beshalasi*.

3.1.5. Ravishni formallashtirish moduli

Ravish harakat-holatning bajarilish tarzi, payti, o'rni, daraja-miqdori kabi belgilarini bildiradi va qanday? qayer? qachon? qancha? so'roqlariga javob bo'ladi.

Qoida: ravishlar otlashadi va ot turkumidagi so'zlarga birikadigan affikslarni oladi, ya'ni $\text{Adj} \rightarrow \text{n} \Rightarrow \text{Adj}_n \cup \text{N}_{\text{aff}}$: *ko'pdan, ozidan, ertasiga*.

Qoida: ravishlar tarkibida *-ga, -dan* kelishik qo'shimchalari, *-i / -si* III shaxs birlik son egalik qo'shimchasi, *-lar* ko'plik qo'shimchasi yaxlitlanib qolgan. Shuning uchun ularga turlanish sifatida qaralmaydi, lingvistik bazaga bunday ravishlar butun lemma (asos) tarzida kiritiladi, lemmitizatsiya jarayonida ular morfemalarga ajratilmaydi: *birdan, zimdan, qo'qqisdan, yangidan, birdaniga, harbiychasiga, kechasi, kunduzi,*

uzoqdan, oldinda, tubdan, rostdan, avvallari, kechalari, teskarisiga, uzunasiga va h.k.

Qo'shma ravishlar quyidagicha hosil qilinadi:

1) *u/bu/shu/o'sha* (ko'rsatish olmoshi) + *yer/yoq/ora/o'rta* oti: *u yerga, bu yerda, shu yerdan, o'sha yoqdan, bu orada, shu o'rtada*;

2) *shu/o'sha* (ko'rsatish olmoshi) + *kun/zamon/vaqt/on/asno* oti: *shu kuni, o'sha zamon, shu zamon, shu zahoti, shu onda, shu asnoda*;

3) *har* (belgilash olmoshi) + bosh/o'rin kelishigidagi so'z: *har kuni, har yili, har yoqqa, har yerda, har taraf, har dam, har zamonda, har gal, har yoqdan, har lahza*;

4) *bir* (son) + bosh/o'rin kelishigidagi so'zlar: *bir yo'la, bir kuni, bir dam, bir zumda, bir nafas, bir pas, bir on, bir onda, bir boshdan, bir zamon, birtekis, biroz, birmuncha, birtalay, birnav*;

5) *hamma* (olmosh) + *yoq/vaqt/zamon/joy/yer*: *hamma yoq, hamma vaqt, hamma zamonda, hamma joy, hamma yer*;

6) *qay* (olmosh) + *vaqt/choq/zamon/payt/mahal*: *qay vaqt, qay choq, qay zamon, qay payt, qay mahal*;

7) *alla* (olmosh) + *mahal/nechuk/payt/zamonlar*: *allamahal, allanechuk, allapayt, allazamonlar*;

Matnlardan ravishning 487 ko'rinishi aniqlandi (9-jadval).

3.1.6. Olmoshni formallashtirish moduli

Olmosh (Pron) boshqa mustaqil so'z, so'z birikmasi va gap o'rnida qo'llanadi, ularga ishora qiladi yoki so'rog'ini bildiradi. Ma'lumotlar bazasiga olmoshlarning grammatik jihatdan yetti turiga mansub barcha olmoshlar kiritiladi, fonologik hamda morfonolik qoidalari modullashtiriladi.

Olmosh mustaqil turkum bo'lsa-da, o'ziga xos so'rog'i yo'q. U qaysi turkumga xos so'zni almashtirsa, shu turkum so'rog'ini oladi. o'z-o'zidan ma'lum bo'ladiki, olmoshlarning qanday qo'shimchalarni biriktirib kelishi uning qaysi turkumga ishora qilayotganiga bog'liq bo'ladi. Misol:

Qoida: olmosh qaysi turkumga xos so'zni almashtirsa shu turkumning morfologik xususiyatiga ega bo'ladi.

Qoida: olmosh otning kichraytirish shaklini olmaydi (-gina erkalash shakli bundan mustasno): *Shuginani o'ylamay bo'ladimi*.

Qoida: olmosh o'zi almashtiradigan so'z turkumining sintaktik xususiyatiga ham ega bo'ladi.

Qoida: olmoshlar shaxs-son kategoriyasining 4-tur morfi $u_1 = \{-man, -san, -miz, -siz, \}$ varianti shakllari bilan birikadi: *menman, sensan, bizmiz*,

sizsiz, kimman, kimsiz, hammamiz, hech kimmiz.

Qoida: kishilik olmoshlari egalik kategoriyasi affikslarini biriktirmaydi ($man \cup im \rightarrow manim$ badiiy ifodasi bundan mustasno).

Qoida: o'zlik olmoshlari egalik kategoriyasining $e_a'[6]=\{-im, -ing, -i, -imiz, -ingiz, -lari\}$ varianti formalari bilan birikib keladi: $o'z \cup im, o'z \cup ing, o'z \cup i; o'z \cup imiz, o'z \cup ingiz, o'z \cup lari$.

Qoida: olmoshlar $Pron \cup$

- “iliq munosabatni ifodalaovchi” affiks: $i_m = -gina: mengina$;
- qarashlilik affiksi: $q_a = -niki: meniki$;
- o‘rin-joy oti affiksi: $\tilde{o}_j = -dagi: sendagi$;
- chegaralash affiksi: $ch_a[3] = \{-gacha, -kacha, -qacha\}: mengacha$;
- ko‘plik affiksi: $k_a = -lar: senlar$;
- kelishik affikslari (variantlari bilan): $ke_a[5] = \{-ning, -ni, -ga, -da, -dan\}: uni, menga, sendan, bizda, sizning$;
- gumon shakli: $gum_a = -dir: mendir, sendir, bizdir$ va h.k.;
- affiksli yuklamalar: $yuk_a = \{-mi, [-chi], [-ku], [-da], [-a], [-ya], -mikan, -mish\}: menmi, sen-chi, biz-ku, siz-da, sen-a, u-ya, bizmikan, senmish$;
- substantiv shakl yasovchi $sh_y = -lik+i \rightarrow \{-ligi\}$ ($sizligi+ngiz$) affiks bilan birikib keladi.

Qoida: olmoshlar kelishik kategoriyasi bilan birikadi: $Pron \cup ke_a[7] = \{-ning, -ni, -ga, -ka, -qa, -da, -dan\}$

Qoida: *men* va *sen* olmoshiga *-ning, -ni, -niki* affikslari qo‘shilsa bir *-n* tovushi tushib qoladi, ya’ni $Pron(men/sen) \cup [-ning / -ni / -niki] \rightarrow mening, meni, meniki; sening, seni, seniki$.

Qoida: *u, shu, bu, o‘sha* olmoshlariga *-ga, -da, -dan, -day, -cha* affikslari qo‘shilsa bir *-n* tovushi orttiriladi, ya’ni $Pron(u/ shu/ bu / o‘sha) \cup [-ga / -da / -dan / -day / -cha] \rightarrow u/ shu/ bu \cup n \cup -ga / -da / -dan / -day / -cha$.

Qoida: gumon olmoshi turini yasovchi *alla-* prefiksi so‘roq olmoshlarining oldidan qo‘shib yoziladi, *-dir* affiksi so‘roq olmoshlaridan keyin qo‘shib yoziladi, ya’ni $[alla-] \cup Pron_{so‘roq} \rightarrow allaPron_{so‘roq}$; $Pron_{so‘roq} \cup [-dir] \rightarrow Pron_{so‘roq}dir: allakim, allanima; kimdir, nimadir, qanchadir$ va h.k.

Olmosh (Pron)ning 126 ko‘rinishi aniqlandi (10-jadval).

3.2. Yordamchi va alohida olingan soʻz turkumlari moduli

Yordamchi soʻzlar atash maʼnosiga ega boʻlmagan, mustaqil ravishda semantik maʼno anglatmaydigan, maʼlum soʻroqlarga javob boʻlmaydigan va gapda maʼlum gap boʻlagi vazifasida kelmaydigan soʻzlardir. Ular mustaqil soʻzlarni yoki gaplarni bir-biriga bogʻlash, qoʻshimcha maʼno yuklash vazifalarini bajaradi.

Yordamchi soʻzlarga koʻmakchi, bogʻlovchi, yuklama kiradi.

Avvalo har uch turga mansub soʻzlar guruhi shakllantiriladi, yaʼni yordamchi soʻzlar turkumiga oid soʻzlar lingvistik bazaga kiritiladi.

Koʻmakchilarni formallashtirish. Koʻmakchi ot, olmosh, sifat-dosh va harakat nomidan keyin kelib, ularni hokim soʻzga bogʻlash uchun xizmat qiluvchi yordamchi soʻzdir. Oʻzbek tilshunosligida koʻmakchilar tabiatiga koʻra ikki turga boʻlinadi: sof koʻmakchilar va koʻmakchi vazifasidagi soʻzlar (11-jadval). Tahrir va tahlil bazasi uchun lemma va uning oʻziga biriktiradigan affikslar hamda affikslar kombinsiyasi muhim ahamiyatga egaligi bois bazaga koʻmakchilarning quyidagi holatda kiritilishi maqsadga muvofiqdir:

Qoida: Affiks qabul qilmaydigan sof koʻmakchilar, yaʼni \downarrow SofK \emptyset [sari, sayin, kabi, singari, yangligʻ, qadar, dovur, to, doir].

Qoida: Affiks qabul qiladigan sof koʻmakchilar, yaʼni \uparrow SofK + [-mi / -ligi / -ligidan / [-a]] \rightarrow [bilan / uchun /oid] + [-mi / -ligi / -ligidan] : *bilanmi/uchunmi/oidmi; bilanligi/uchunligi/oidligi; bilanligidan/uchunligidan/oidligidan; bilan-a/uchun-a/oid-a.*

Qoida: Affiks qabul qilmaydigan koʻmakchi vazifasidagi soʻzlar (KoʻmV), yaʼni \downarrow KoʻmV = [qaramay / qaraganda / qaramasdan / koʻra / uzra / osha / boʻylab / ichra / boʻyi / chamasi / qarata / burun / vajidan / misoli / tortib] \emptyset .

Qoida: Affiks qabul qiladigan koʻmakchi vazifasidagi soʻzlar, yaʼni 1-guruh: \uparrow KoʻmV = [qarab/oʻrtasiga/oʻrtasida/oʻrtasidan/boʻlak/tashqari/sababli/tufayli/ soʻng/beri/ boʻyicha/ chogʻi / yarasha / asosan / binoan / muvofiq / buyon / ilgari / boshlab / badaliga / evaziga] \cup [-mi/[-a /-ya] /-dir].

1-guruh: \uparrow KoʻmV [boshqa / qarshi / oʻzga] \cup [-mi / [-ya] / -si / -dir].

2-guruh: \uparrow KoʻmV [haqida/toʻgʻrisida/holda/yoʻsinda/borasida/bobida] \cup [-gi / [-ya] / -mi].

3-guruh: \uparrow KoʻmV [tomon/ osti / usti / tepasi / oldi / orqasi / yoni / ichi / orasi / boshi / oʻrtasi / tagi / oldin / keyin / davomida] \cup [-ga / -da / -dan / -i / -ma / [-a] / -dagi].

Qoida: *uchun* va *bilan* ko‘makchilarining qisqargan shakllari chiziqcha bilan yoziladi, ya‘ni $\uparrow\text{SofK}(\text{uchun} / \text{bilan}) \rightarrow [-\text{chun}] / [-\text{la}]$.

Bog‘lovchilarni formallashtirish. Bog‘lovchilar qo‘shma gap tarkibidagi sodda gaplar va uyushgan bo‘lak o‘rtasidagi munosabatlarni ifodalashda qo‘llaniladigan yordamchi so‘zlardir.

Bog‘lovchilar qo‘llanilishiga ko‘ra ikki turga bo‘linadi (12-jadval):

1) yakka qo‘llanadigan bog‘lovchilar: *a, va, hamda, ammo, lekin, biroq, balki, holbuki, vaholanki, gar/agar/agarda, basharti, chunki, garchi/ garchand, go‘yo/go‘yoki, -ki/-kim, esa, ya‘ni, shuning uchun, yo, yoki, yoxud, yoinki.*

2) takroriy qo‘llaniladigan bog‘lovchilar: *dam..., dam..., ba‘zan..., ba‘zan..., ham..., ham..., bir..., bir..., xoh..., xoh..., goh..., goh...*

Qoida: [*a / va / hamda / esa / yo / yoki / yoxud / yoinki*] bog‘lovchilarining har ikki tomonida ham tinish belgisi ishlatilmaydi, ya‘ni $\Delta[a / va / hamda / esa / yo / yoki / yoxud / yoinki]\Delta$.

Qoida: [*,*] Δ [*ammo / lekin / biroq / balki / holbuki / vaholanki / gar/agar/agarda / basharti / chunki / garchi / garchand / go‘yo/go‘yoki / ya‘ni / shuning uchun*] Δ (ya‘ni ushbu bog‘lovchilardan oldin vergul qo‘yiladi).

Qoida: *-ki/-kim* ergashtiruvchi bog‘lovchilari bilan kelgan so‘z (W)dan so‘ng vergul tinish belgisi qo‘yiladi: $[W] \cup \{-ki, -kim\} [,]$

Yuklamalarni formallashtirish. Yuklama – ayrim so‘z yoki sodda gapga qo‘shimcha ma‘no yuklash uchun qo‘llanadigan yordamchi lingvistik vosita (13-jadval).

Qoida: *-mi, -gina, -kina, -qina, -oq, -yoq, -dir* qo‘shimcha-yuklamalar so‘zlarga qo‘shib yoziladi: $[W] \cup \{ mi, -gina, -kina, -qina, -oq, -yoq, -dir \}$

Qoida: lotin yozuviga asoslangan o‘zbek alifbosida *-u, -yu* qo‘shimcha-yuklamalar juft so‘zlar orasida bog‘lovchi bo‘lib kelsa birinchi so‘zga chiziqcha orqali birikadi, undan so‘ng probel qo‘yiladi, ya‘ni $[W_1] \cup \{ -u, -yu \} \Delta [W_2]$: *katta-yu kichik, sen-u men* va h.k.

Qoida: kirill yozuviga asoslangan o‘zbek alifbosida *-u, -yu* qo‘shimcha-yuklamalar juft so‘zlar orasida bog‘lovchi bo‘lib kelsa birinchi so‘zga chiziqcha bilan qo‘shib yoziladi, undan so‘ng kelgan so‘z bir chiziq qo‘yib yoziladi, ya‘ni $[W_1] - \{ -y, -ю \} - [W_2]$: *каттаю-кичик, пасту-баланд* каби.

Qoida: *-chi, -a, -ya, -ku, -da, -u, -yu* qo‘shimcha-yuklamalar so‘zlarga chiziqchasi bilan qo‘shib yoziladi: $[W] \cup \{ -chi, -a, -ya, -ku, -da, -u, -yu \}$

Alohida so‘z turkumlari moduli. Lingvistik protsessor bazasiga ilovada keltirilgan modallar (14-jadval), undov so‘zlar (15-jadval) va taqlidlar (16-jadval) jumlada quyidagicha tinish belgilari bilan yoziladi:

Qoida: jumlaning boshida kelsa undan so‘ng vergul qo‘yiladi: $\{\hat{A}=\text{modal, undov}\} \cup [,]$

Qoida: jumlaning o‘rtasida kelsa har ikki tomoniga ham vergul qo‘yiladi: $[,] \cup \{\check{a}=\text{modal, undov}\} \cup [,]$

Qoida: quyidagi holda modal va undovlar gapning oxirida kelgan hisoblanadi va ulardan oldin kelgan so‘zdan so‘ng vergul qo‘yiladi: $[,] \triangle \{\check{a}=\text{modal, undov}\} \cup [,]$

3.3. Lingvistik protsessorada affiksatsiya moduli

Ma’lumki, o‘zbek tilida so‘z yasashning quyidagi usullari e’tirof etiladi:

- 1) affiksatsiya yoki morfologik usul (affiks qo‘shish) – leksema+affiks;
- 2) kompozitsiya yoki sintaktik usul (so‘z qo‘shish) – leksema+leksema;
- 3) konversiya (boshqa turkumga ko‘chish) – $[W_1] \rightarrow [W_2]$;
- 4) semantik yoki leksik-semantik (leksemani yangi ma’noda qo‘llash);
- 5) abbreviatsiya (so‘zni qisqartirish).

Dastlabki bosqichda ishning ko‘lamini chegaralagan holda, birinchi galda morfoanalizatorning lingvistik bazasiga so‘z yasashning affiksatsiya usuli modullari kiritildi.

Affiksatsiya usuli bilan yangi leksema yasash to‘rt so‘z turkumiga xos bo‘lib, ular quyidagilar: ot, sifat, fe’l, ravish.

Faqat mana shu to‘rt so‘z turkumida so‘z yasovchi qo‘shimcha mavjud. Son, olmosh, taqlid, undov va modal dan leksema yasalmasa-da, ular xuddi shu usul bilan ot, sifat, fe’l va ravish yasashda ishtirok etadi.

Ma’lumotlar bazasida hosil bo‘lgan leksema *yasalma*, *yasalma* asosi *lemma* deyildi.

Boshqa turkumlardan ot, fe'l, sifat, ravish lemma yasovchi affikslar miqdori

	ot lemma yasovchi affikslar	fe'l lemma yasovchi affikslar	sifat lemma yasovchi affikslar	ravish lemma yasovchi affikslar
■ jami yasovchi affikslar	114	58	117	48
■ ot lemmadan	37	17	56	18
■ fe'ldan	54	0	40	16
■ sifatdan	11	16	11	5
■ ravishdan	3	7	3	4
■ taqlid so'zlardan	1	8	3	0
■ sondam	1	3	0	0
■ olmoshdan	1	3	2	3
■ undov lemmadan	1	1	0	0
■ modal lemmadan	1	3	2	2

O'zbek adabiy tilida so'z yasovchi affikslar leksemaga quyidagi tartibda qo'shiladi: (prefiks₁)+asos₂+(yasovchi affiks₃) (*xush₁xabar₂; qabul₂xona₃; be₁odob₂lik₃*)

Dasturning lingvistik ta'minotiga o'zbek adabiy tilida mavjud barcha yasovchi affikslar kiritildi. Bu kelajakda tahlil dasturini mukammallashtirish va yangi-yangi so'z yasalishlarni avtomatik tahlil qilishda qulaylik yaratadi.

Ilovada so'z turkumlaridan affiksatsiya usuli bilan so'z yasalish modellari jadval ko'rinishida berildi.

Ot yasovchi affikslar

O'zbek adabiy tilida affiksatsiya usuli bilan ot yasovchi 114 ta affiks aniqlandi. Shundan 37 tasi ot, 11 tasi sifat, 54 tasi fe'l, 3 tasi ravish, 5 taqlid, 1 son, 1 olmosh, 1 undov va 1 modal lemmaga qo'shilib, *yasalma ot lemma* hosil qiladi (17-jadval).

Fe'l yasovchi affikslar

Fe'l yasovchi 58 ta affiks aniqlandi. Shundan 17 tasi ot, 16 tasi sifat, 3 tasi son, 7 tasi ravish, 3 olmosh, 8 taqlid, 3 modal va 1 undov lemmaga qo'shilib, *fe'l lemma* yasaydi (18-jadval).

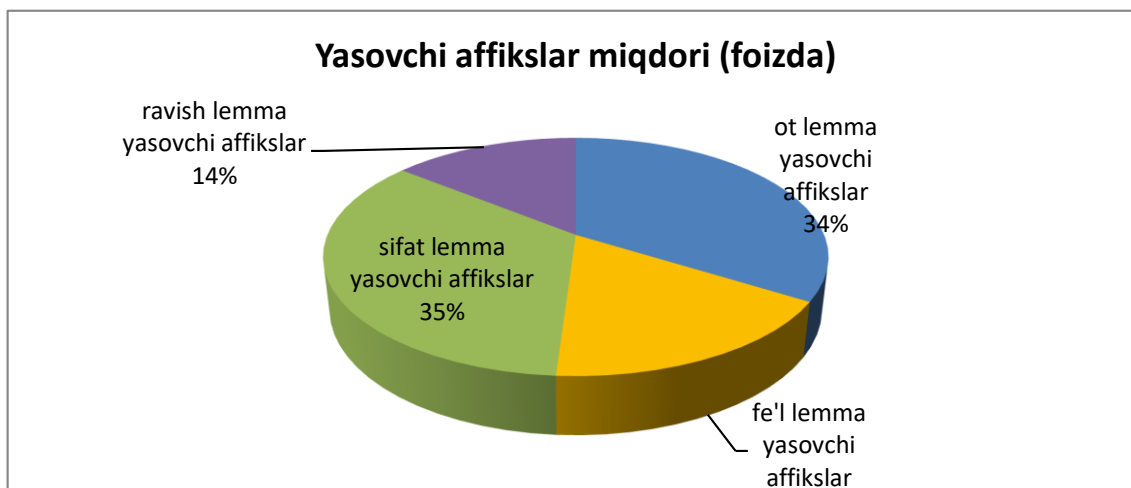
Sifat yasovchi affikslar

O'zbek tilida sifat yasovchi affikslarning 117 ko'rinishi aniqlandi. Shundan 56 tasi ot, 40 tasi fe'l, 11 tasi sifat, 3 tasi ravish, 2 olmosh, 2 modal va 3 taqlid lemmaga qo'shilib, yangi *sifat lemma* yasaydi (19-jadval).

Ravish yasovchi affikslar

Manbalar asosida ravish yasovchi 48 ta affiks aniqlandi. Ulardan 18 tasi ot, 16 tasi sifat, 5 tasi fe'l, 4 tasi ravish, 3 olmosh va 2 modal lemmaga qo'shib, *ravish lemma yasovchi* (20-jadval).

Yasovchi affikslarning foizlardagi ko'rsatkichi tubandagicha:



UCHINCHI BOB BO'YICHA XULOSALAR

Dissertatsiyaning uchinchi bobi so'z turkumlarining lingvistik modulini yaratishga qaratildi. Bundan maqsad ATT dasturining algoritmini tuzish orqali uning lingvistik ta'minotini shakllantirishdir.

1. Lingvistik ta'minot asosini leksik birliklar tashkil etadi. Bunday leksik birliklar muayyan turkumga mansub bo'lgan yoki turkumga ega bo'lmagan birlik sifatida ajratiladi. Bunday jihat avtomatik tahrir va tahlil dasturi bosqichlari texnologiyalarining mukammal ishlashini ta'minlaydi.

2. So'zshakllarni yuzaga keltirishda orfografik, fonologik, morfonologik hodisalar ro'y beradi. Lingvistik modulni yaratishda hozirgi o'zbek adabiy tilining lingvistik me'yorlariga matn tahriri va tahlilida asosiy lisoniy hodisa sifatida qaraladi. An'anaviy tilshunoslik me'yorlari asosida KLda grammatik qoidalarni berish, ularning algoritmini yaratish ustuvor vazifa hisoblanadi.

3. O'zbek tilidagi yasovchi qo'shimchalarni statistik jihatdan hisoblash va ularning leksik bazasi va *lemma+yasovchi affiks* modelini yaratish lingvistik ta'minotning muhim tarkibiy qismi hisoblanadi. Kompyuter uchun affikslarni so'z yasovchi va shakl yasovchi affikslar tarzida alohida bazalarga ajratish, modellarini aniqlash tahlil jarayonida muhim.

4. Matnlarni ATT qilish dasturini innovatsion ta'limning bir ko'rinishi sifatida qayd etib, uning quyidagi imkoniyatlarga ega ekanligini ko'rsatish mumkin:

- 1) turli hajmdagi matnlarni tezkor tahrir qiladi;
- 2) matn teruvchining yozma nutq malakasini takomillashtiradi;
- 3) foydalanuvchiga matnlarni mustaqil ravishda to'g'ri va savodli yozishida ko'mak beradi;
- 4) o'zbek tilining leksik, grammatik me'yorlarini o'rganishda lingvistik manba vazifasini o'taydi.

XULOSA

O‘zbekcha matnlarni tahrir va tahlil qilish dasturining lingvistik ta‘minotini yaratishda quyidagi xulosalarga kelindi:

1. Shu paytgacha amalga oshirilgan ilmiy tadqiqot ishlarida avtomatik tahrir va tahlilga qisman e‘tibor berilgan, xolos. Mavjud izlanishlarda ikki til yoxud o‘zga til morfologiyasi qiyosiy-tipologik uslubda tahlil qilingan. Shunga qaramay, bu masala lingvistik tizim sifatida azaldan tilshunoslarni qiziqtirganini ko‘rish mumkin.

2. Jahon KLdagi matnlar bilan bog‘liq tadqiqotlar, o‘z navbatida, matnning tahriri va tahlili tadqiqi sari yetaklaydi, ulardagi xulosalar va olinadigan natijalarning asosli hamda ishonchli bo‘lishiga ko‘maklashadi. Ana shu g‘oyalarning amalga oshishi kompyuter lingvistikasida matnlarni qayta ishlash tadqiqotlarini yangi pog‘onaga olib chiqadi va uning amaliyotga tatbiq etilishini osonlashtiradi.

3. ATT dasturining lingvistik ta‘minotini yaratish lingvistik me‘yorlar va algoritmgaga ega qoidalarning ishlab chiqilishi, leksikografik manbalarning ta‘minotga kiritilishi, so‘zshakllar va so‘zlarning o‘zaro bog‘lanish modellari tuzilishi bilan belgilanadi. Mazkur jarayonda rasmiy va ilmiy uslubdagi matnlar mohiyati lingvistik tahlil qilinmaydi, balki mazkur uslubdagi matnlar imlosi va grammatikasini tekshiruvchi sistema yoki DT (morfoanalizator)ni yaratish tamoyillari ishlab chiqiladi.

4. Bir so‘zning uslublar doirasida o‘z qo‘llanish variantiga egaligi va ushbu uslublarning me‘yorlarini hisobga olib aytish mumkinki, so‘zning hamma uslubdagi ko‘rinishi hamda affikslar bilan kelish shaklini yaratish murakkab vazifa. Shu bois ATT dasturining LTini yaratishda ilmiy va rasmiy uslublarda qo‘llaniladigan so‘z va affikslarning olinishi ishning samaradorligini oshirishga xizmat qiladi.

5. Har qanday dasturning lingvistik bazasini yaratishda kerakli manba – “O‘zbek tilining morfologik lug‘ati”dir. Bunday lug‘atda bir leksemaning qaysi turkumga oidligi, uning so‘zshakllari keltiriladi. Bu esa dasturning mukammal chiqishini ta‘minlaydi. Shu bois o‘zbek tilidagi matnlarga ishlov beruvchi dasturlarning lingvistik ta‘minotini yaratishda qulaylik va natijaga tezkor erishish maqsadida ushbu lug‘atning ishlab chiqilishi maqsadga muvofiqdir. Morfologik lug‘at tilshunosning vaqtini tejaydi va belgilangan natijaga erishishni tezlashtiradi.

6. Lingvistik protsessorni yaratishda lingvistik modulning o‘rni va ahamiyati o‘ta muhim. KLda modul termini dasturiy ta‘minotning muayyan lingvistik jarayonini qamragan mustaqil tarkibiy qismi sifatida

qo‘llaniladi. Leksik modulda lug‘at qatlami qurshab olinsa, grafematik modulda ramzlar, tinish belgilar, harfiy va boshqa belgilar tahrir qilinadi, orfografik modulda imlo qoidalari, morfologik modulda so‘zshakllar analizi (morfo-tahlil so‘zshakldan leksemaga qadar tahlil) va sintezi (leksemaning grammatik shakllanishi tahlili jarayoni), sintaktik modulda supersintaktik birlik – gap yoki so‘zlarning o‘zaro bog‘lanish hodisasi tahlil qilinadi.

7. O‘zbek tilining ATTSning dastlabki joriylanishida rasmiy-idoraviy va ilmiy uslubdagi matnlarni tahrir va tahlil qilishga mo‘ljallangan lingvo-sistemaning mukammal ishlashi va natijasi samarali bo‘lishi uchun mazkur sistema grafematik, morfologik va sintaktik tahlil bosqichlariga ajratilib, modullashtirilishi maqsadga muvofiq. Grafematik bosqichda tahriri amalga oshiriladigan birliklarning belgilab olinishi tahrir aniqligini ta‘minlaydi. Morfologik tahlil prinsiplari ot turkumi misolida yoritib berilishi boshqa turkum modullarini tuzish tamoyillarini yaratishga zamin yaratadi. Sintaktik tahlil jarayonida gap miqdori cheksiz va u individual tuzilishi yoxud qo‘llanilishga bog‘liq bo‘lgani sababli STda nisbatan kichik birlik hisoblanmish so‘z birikmalariga tayanilgan ma‘qul, ya‘ni, MT modullarini yaratish jarayonida barcha leksemalar so‘z turkumlariga ajratilganiga asoslanib, so‘zlarning bog‘lanish yo‘llari modellashtiriladi va “so‘z birikmalari kategoriyasi” yaratiladi.

8. So‘zshakllarni yuzaga keltirishda orfografik, fonologik, morfonologik hodisalar ro‘y beradi. Lingvistik modulni yaratishda hozirgi o‘zbek adabiy tilining lingvistik me‘yorlariga matn tahriri va tahlilida asosiy lisoniy hodisa sifatida qaraladi. An‘anaviy tilshunoslik me‘yorlari asosida KLda o‘zbek tili so‘z turkumlari modulida leksik-grammatik kodni shakllantirish maqsadida fonologik, morfonologik va orfografik qoidalar algoritmini tuzish; sintaktik va semantik kodni shakllantiruvchi so‘z birikmalarining lingvistik modelini yaratish grammatik qoidalarni berish, ularning algoritmini yaratish ustuvor vazifa hisoblanadi. Mazkur masala nafaqat lingvo-tahlil jarayonining sifatli olib borilishini ta‘minlaydi, balki o‘zbek tilini o‘qitish uchun ham dolzarbligini ko‘rsatadi.

9. O‘zbek tilidagi yasovchi qo‘shimchalarni statistik jihatdan hisoblash va ularning leksik bazasi va *lemma+yasovchi affiks modelini* yaratish lingvistik ta‘minotning muhim tarkibiy qismi hisoblanadi. Kompyuter uchun affikslarni so‘z yasovchi va shakl yasovchi affikslar tarzida alohida bazalarga ajratish, modellarini aniqlash tahlil jarayonida muhim.

10. Omonim shakllar hamisha tilshunoslik markazidagi hodisalardan biri sifatida qaralgan. Shu bois matnda uchraydigan omoshakllar rus va xorij tilshunosligida alohida tadqiq qilinadi. Shu kunga qadar ishlab chiqilgan omonim soʻzshakllarni matndan aniqlash va ularni farqlash usullaridan Brill usuli, yashirin Markov modeli va boshqa modellar modifikatsiyasi oʻrganilib, oʻzbekcha matnlarda omonim shakllarni tahlil qilish texnologiyasi tatbiq qilindi. Natijada omonimshakllarni matndan aniqlab ularni tahlil qilish asosi yaratildi.

11. Frazologik birikmalar nutqiy emas, balki lisoniy hodisa ekanligi, ulardagi grammatik bogʻlanish turgʻun boʻlganligi bois ularning kategorial modellarini yaratish murakkab masala hisoblanadi. Shu bois tugʻun birikmalar tarkibidagi soʻzlarning bogʻlanish modeli emas, balki grammatik shaklini oʻzgartirish oʻrinlarini belgilash va lingvistik bazaga kiritish avtomatik tahlilning aniqligini taʼminlaydi.

12. Lingvo-analiz dasturi nafaqat tahrir va tahlil jarayonini tezlashtiruvchi vosita, balki oʻzbekcha matnlarni bequsur yozish koʻnikmasini shakllantiruvchi innovatsion sistema sifatida muhim ahamiyatga ega hisoblanadi.

13. Hozirga qadar jamlangan lingvistik qoidalarning keng koʻlamli tadqiqi kelajakda oʻzbek lingvistik korpusining yaratilishiga muhim hissa boʻlib qoʻshiladi.

ILOVALAR

1-jadval

Harflar tizimi¹⁴⁰

№	kirill yozuvi alifbosi				lotin yozuvi harflari ¹⁴¹				lotin yozuvi alifbosi	ingliz tili harflari		ingliz tili alifbosi
	bosh harflar		kichik harflar		bosh harflar		kichik harflar			bosh harflar	kichik harflar	
	bosma	yozma	bosma	yozma	bosma	yozma	bosma	yozma				
1	А	А	а	а	А	А	а	а	Aa	A	a	Aa
2	Б	Б	б	б	Б	В	b	b	Bb	B	b	Bb
3	В	В	в	в	В	V	v	v	Dd	D	d	Cc
4	Г	Г	г	г	Г	G	g	g	Ee	E	e	Dd
5	Д	Д	д	д	Д	D	d	d	Ff	F	f	Ee
6	Е	Е	е	е	Е	Ye (E)	ye (e)	ye (e)	Gg	G	g	Ff
7	Ё	Ё	ё	ё	Ё	Yo	yo	yo	Hh	H	h	Gg
8	Ж	Ж	ж	ж	Ж	J	j	j	Ii	I	i	Hh
9	З	З	з	з	З	Z	z	z	Jj	J	j	Ii
10	И	И	и	и	И	I	i	i	Kk	K	k	Jj
11	Й	Й	й	й	Й	Y	y	y	Ll	L	l	Kk
12	К	К	к	к	К	K	k	k	Mm	M	m	Ll
13	Л	Л	л	л	Л	L	l	l	Nn	N	n	Mm
14	М	М	м	м	М	M	m	m	Oo	O	o	Nn
15	Н	Н	н	н	Н	N	n	n	Pp	P	p	Oo
16	О	О	о	о	О	O	o	o	Qq	Q	q	Pp
17	П	П	п	п	П	P	p	p	Rr	R	r	Qq
18	Р	Р	р	р	Р	R	r	r	Ss	S	s	Rr
19	С	С	с	с	С	S	s	s	Tt	T	t	Ss
20	Т	Т	т	т	Т	T	t	t	Uu	U	u	Tt
21	У	У	у	у	У	U	u	u	Vv	V	v	Uu
22	Ф	Ф	ф	ф	Ф	F	f	f	Xx	X	x	Vv
23	Х	Х	х	х	Х	X	x	x	Yy	Y	y	Ww
24	Ц	Ц	ц	ц	Ц	S (ts)	s (ts)	s (ts)	Zz	Z	z	Xx
25	Ч	Ч	ч	ч	Ч	Ch	ch	ch	O‘o‘			Yy
26	Ш	Ш	ш	ш	Ш	Sh	sh	sh	G‘g‘			Zz
27			ъ	ъ					Shsh			
28			ь	ь					Chch			
29	Э	Э	э	э	Э	E	e	e	ng			
30	Ю	Ю	ю	ю	Ю	Yu	yu	yu				
31	Я	Я	я	я	Я	Ya	ya	ya				
32	Ў	Ў	ў	ў	Ў	O‘	o‘	o‘				

¹⁴⁰ Kirill alifbosi tartibida yozildi

¹⁴¹ Lotin yozuviga asoslangan o‘zbek alifbosi. O‘zbekiston Respublikasining 1995-yil 6-maydagi 71-I-sonli Qonuni

33	Қ	Қ	қ	қ	Қ	Q	q	q				
34	F	F	f	f	F	G'	g'	g'				
35	X	X	x	x	X	H	h	h				
			ҺГ	ҺГ			ng	ng		W	w	
										C	c	

2-jadval

Tovushlar tizimi

		a	o	i	u	e	o'	e	yo	yu	ya							
Unli tovushlar (Voc)																		
Undosh tovushlar (C)	jarangli	b	v	g	d	z	j			l	m	n	r	y				
	jaangsiz	p	f	k	t	s	ch	sh							x	s	q	h
Faqat harf va belgini anglatadi		'		'														

3-jadval

Raqamlar tizimi (Num)

	arab raqamlari	rim raqamlari		arab raqamlari	rim raqamlari
birik	0		yuzlik	100	C
	1	I		200	CC
	2	II		300	CCC
	3	III		400	CD
	4	IV		500	D
	5	V		600	DC
	6	VI		700	DCC
	7	VII		800	DCCC
	8	VIII		900	CM
	9	IX			
o'nlik	10	X	minglik	1000	M
	20	XX		2000	MM
	30	XXX		3000	MMM
	40	XL			
	50	L			
	60	LX		1986	MCMLXXXVI
	70	LXX		21	XXI
	80	LXXX		2019	MMXIX
	90	XC		2020	MMXX

Belgilar tizimi

№	Harf yoki raqamdan so'ng probelsiz yoziladigan belgilar		Harf yoki raqamdan so'ng probel yoki probelsiz yoziladigan belgilar	
	ramzi	nomi	ramzi	nomi
1	.	nuqta	-	chiziqcha
2	,	vergul	—	tire
3	()	ochiq va yopiq qavs	:	ikki nuqta
4	!	undov	№	raqam belgisi
5	?	so'roq	%	foiz ramzi
6	...	ko'p nuqta	*	yulduzcha
7	“ ” « »	ochiq va yopiq qo'shtirmoq	\$	dollar ramzi
8	'	tutuq belgisi	#	reshotka
9	:	ikki nuqta	/	bir qiya chiziq
10	;	nuqtali vergul	//	ikki qiya chiziq
	@	kuchukcha		
	_	pastki chiziq		
	^			
	{ }	figurali ochiq va yopiq qavs		
	[]	kvadrat ochiq va yopiq qavs		
	>	katta yoki kichik belgisi		
	<	katta yoki kichik belgisi		

Affikslar tizimi

№	Lug'aviy shakl yasovchi affikslar												
		1	Kichraytirish -erkalash k_e[7]	-cha	-choq	-loq	-jon	-xon	-oy	-	bek		
2	“iliq munosabatni ifodalaovchi” aff. i_m[3]	-gina	-kina	-qina									
3	qarashlilik aff. q_a	-niki											

4	o‘rin-joy oti aff. ō j	-dagi											
5	chegaralash aff. ch_a[3]	- gacha	- kacha	- qach a									
6	ko‘plik aff. k_a	-lar											
7	substantiv shakl yasovchi aff. sh_y	-lik											
8	sifat darajalari affikslari	-roq	-ish	-sh	-imtir	-mtir	- g‘ish	- tob	- gina				
9	fe‘lning vazifadosh shakllari	-v	-uv	-moq	-mak	-sh	-ish	-r	-ar	-gan	-kan	-qan	
		- adiga n	- ayotg an	-b	-ib	-a	-y	- gac h	- kach	-qach	- gunch a	- kunch a	
		- qunch a	- ganch a	- kanch a	- qanch a	- gani	-kani	qan i	- gude k	- kudek	- qudek	- gali	
		-may	- mayin	- masd an	-vchi	- uvch i			- guvc hi	- maslik			
10	fe‘l nisbatlari affikslari	-n	-in	-l	-il	-tir	-dir	-ir	-ar	-iz	-giz	-g‘iz	
		-g‘az	-kiz	-qaz	-kaz	-qiz	-sat	-t	-it	-sh	-ish		
11	son turlari aff.	-ta	-tacha	-lab	- larch a	-ov	-ala		- ovlo n	-tadan	-inchi	-nchi	
12	affiksli yuklamalar	Leksemaga qo‘shilib yoziladi						Leksemaga chiziqcha bilan birikadi					
		-mi	-gina	-kina	-qina	-dir		-u	-yu	-da	-a	-ya	
		-oq	-yoq					-chi	-ku				
Sintaktik shakl yasovchi affikslar													
1	kelishik affikslari ke_a[7]	-ning	-ni	-ga	-ka	-qa	-da	- dan					
2	egalik affikslari e_a[11]	-m	-ng	-si	-miz	-ngiz	-lari	-im	-ing	-i	-imiz	-ingiz	
3	shaxs-son aff.	-man	-san	-di	-miz	-siz	-di	- imiz	-isiz	-dilar			
		-m	-ng	-k	-ngiz	-di	-im	-ing	-ik	-ingiz	-dilar		
		-y	-gin	-sin	-ylik	-ng	- nglar	-sin	-ay	-aylik	-ing	- inglar	

		<i>-sinlar</i>										
4	fe'l zamonlari aff.	<i>-di</i>	<i>-gan</i>	<i>-kan</i>	<i>-qan</i>	<i>-gan edi</i>	<i>-kan edi</i>	<i>-qan edi</i>	<i>-gandi</i>	<i>-ar edi</i>	<i>-mas edi</i>	<i>-(a) yotgan edi</i>
		<i>(a)yotgandi</i>	<i>-moqda edi</i>	<i>-moqchi edi</i>	<i>-moqchi emas edi</i>	<i>-moqchi emas di</i>						
		<i>-yap</i>	<i>-yapti</i>	<i>-ayotib</i>	<i>-yotib</i>	<i>-moqda</i>	<i>-yotir</i>					
		<i>-y</i>	<i>-a</i>	<i>-moqchi</i>	<i>-mas</i>	<i>-jak</i>	<i>-ur</i>	<i>-gusi</i>	<i>-g'usi</i>	<i>-gay</i>	<i>-sa</i>	
5	ergashtiruvchi bog'lovchilar	<i>-ki</i>	<i>-kim</i>									

6-jadval

Rasmiy va ilmiy matnlarda uchraydigan fe'l shakllari modeli

№	Fe'l qoliplari	V – uchra
1	$V(x) + y_1(1) + z_2 \text{bir}1 \Rightarrow V_3$	uchra+di+m=uchradim
2	$V(x) + y_1(1) + z_2 \text{bir}2 \Rightarrow V_3$	uchra+di+ng=uchrading
3	$V(x) + y_1(1) + z_2 \text{bir}3 \Rightarrow V_3$	uchra+di+Ø=uchradi
4	$V(x) + y_1(1) + z_2 \text{ko'p}1 \Rightarrow V_3$	uchra+di+k=uchradik
5	$V(x) + y_1(1) + z_2 \text{ko'p}2 \Rightarrow V_3$	uchra+di+ngiz=uchradingiz
6	$V(x) + y_1(1) + z_2 \text{ko'p}3 \Rightarrow V_3$	uchra+di+Ø yoki lar=uchradi(lar)
7	$V(x) + ma + y_1(1) + z_2 \text{bir}1 \Rightarrow V_3$	uchra+ma+di+m=uchramadim
8	$V(x) + ma + y_1(1) + z_2 \text{bir}2 \Rightarrow V_3$	uchra+ma+di+ng=uchramading
9	$V(x) + ma + y_1(1) + z_2 \text{bir}3 \Rightarrow V_3$	uchra+ma+di+ Ø =uchramadi
10	$V(x) + ma + y_1(1) + z_2 \text{ko'p}1 \Rightarrow V_3$	uchra+ma+di+k=uchramadik
11	$V(x) + ma + y_1(1) + z_2 \text{ko'p}2 \Rightarrow V_3$	uchra+ma+di+ngiz=uchramadingiz
12	$V(x) + ma + y_1(1) + z_2 \text{ko'p}3 \Rightarrow V_3$	uchra+ma+di+lar=uchramadilar
13	$V(x) + y_1(2) + z_1 \text{bir}1 \Rightarrow V_3$	uchra+gan+man=uchraganman
14	$V(x) + y_1(2) + z_1 \text{bir}2 \Rightarrow V_3$	uchra+gan+san=uchragansan
15	$V(x) + y_1(2) + z_1 \text{bir}3 \Rightarrow V_3$	uchra+gan+ Ø =uchragan
16	$V(x) + y_1(2) + z_1 \text{ko'p}1 \Rightarrow V_3$	uchra+gan+miz=uchraganmiz
17	$V(x) + y_1(2) + z_1 \text{ko'p}2 \Rightarrow V_3$	uchra+gan+siz=uchragansiz
18	$V(x) + y_1(2) + z_1 \text{ko'p}3 \Rightarrow V_3$	uchra+gan+lar=uchraganlar
19	$V(x) + ma + y_1(2) + z_1 \text{bir}1 \Rightarrow V_3$	uchra+ma+gan+man=uchramaganman
20	$V(x) + ma + y_1(2) + z_1 \text{bir}2 \Rightarrow V_3$	uchra+ma+gan+san=uchramagansan
21	$V(x) + ma + y_1(2) + z_1 \text{bir}3 \Rightarrow V_3$	uchra+ma+gan+ Ø =uchramagan

22	$V(x) + ma + y_1(2) + z_1ko'p1 \Rightarrow V_3$	uchra+ma+gan+miz=uchramaganmiz
23	$V(x) + ma + y_1(2) + z_1ko'p2 \Rightarrow V_3$	uchra+ma+gan+siz=uchramagansiz
24	$V(x) + ma + y_1(2) + z_1ko'p3 \Rightarrow V_3$	uchra+ma+gan+lar=uchramaganlar
25	$V(x) + y_1(2) + z_4bir1 \cup yo`q \Rightarrow V_3$	uchra+gan+im \cup yo`q \Rightarrow uchraganim yo`q
26	$V(x) + y_1(2) + z_4bir2 \cup yo`q \Rightarrow V_3$	uchra+gan+ing \cup yo`q \Rightarrow uchraganing yo`q
27	$V(x) + y_1(2) + z_4bir3 \cup yo`q \Rightarrow V_3$	uchra+gan+i \cup yo`q \Rightarrow uchragani yo`q
28	$V(x) + y_1(2) + z_4ko'p1 \cup yo`q \Rightarrow V_3$	uchra+gan+imiz \cup yo`q \Rightarrow uchraganimiz yo`q
29	$V(x) + y_1(2) + z_4ko'p2 \cup yo`q \Rightarrow V_3$	uchra+gan+ingiz \cup yo`q \Rightarrow uchraganingiz yo`q
30	$V(x) + y_1(2) + z_4ko'p3 \cup yo`q \Rightarrow V_3$	uchra+gan+lari \cup yo`q \Rightarrow uchraganlari yo`q
31	$V(x) + y_1(3) + z_4bir1 \Rightarrow V_3$	uchra+gan \cup edi+m \Rightarrow uchragan edim
32	$V(x) + y_1(3) + z_4bir2 \Rightarrow V_3$	uchra+gan \cup edi+ng \Rightarrow uchragan eding
33	$V(x) + y_1(3) + z_4bir3 \Rightarrow V_3$	uchra+gan \cup edi+ $\emptyset \Rightarrow$ uchragan edi
34	$V(x) + y_1(3) + z_4 ko'p1 \Rightarrow V_3$	uchra+gan \cup edi+k \Rightarrow uchragan edik
35	$V(x) + y_1(3) + z_4 ko'p2 \Rightarrow V_3$	uchra+gan \cup edi+ngiz \Rightarrow uchragan edingiz
36	$V(x) + y_1(3) + z_4 ko'p3 \Rightarrow V_3$	uchra+gan \cup edi+lar \Rightarrow uchragan edilar
37	$V(x) + u_1(4) + z_2bir1 \Rightarrow V_3$	uchra+gandi+m \Rightarrow uchragandim
38	$V(x) + u_1(4) + z_2bir2 \Rightarrow V_3$	uchra+gandi+ng \Rightarrow uchraganding
39	$V(x) + u_1(4) + z_2bir3 \Rightarrow V_3$	uchra+gandi+ $\emptyset \Rightarrow$ uchragandi
40	$V(x) + ma + u_1(4) + z_2bir1 \Rightarrow V_3$	uchra+ma+gandi+m \Rightarrow uchramagandim
41	$V(x) + ma + u_1(4) + z_2bir2 \Rightarrow V_3$	uchra+ma+gandi+ng \Rightarrow uchramaganding
42	$V(x) + ma + u_1(4) + z_2bir3 \Rightarrow V_3$	uchra+ma+gandi+ $\emptyset \Rightarrow$ uchramagandi
43	$V(x) + u_1(5) + z_2bir1 \Rightarrow V_3$	uchra+r \cup edi +m \Rightarrow uchrar edim
44	$V(x) + u_1(5) + z_2bir2 \Rightarrow V_3$	uchra+r \cup edi +ng \Rightarrow uchrar eding
45	$V(x) + u_1(5) + z_2bir3 \Rightarrow V_3$	uchra+r \cup edi + $\emptyset \Rightarrow$ uchrar edi
46	$V(x) + u_1(6) + z_2 ko'p1 \Rightarrow V_3$	uchra+mas \cup edi +k \Rightarrow uchramas edik
47	$V(x) + u_1(6) + z_2 ko'p2 \Rightarrow V_3$	uchra+mas \cup edi +ngiz \Rightarrow uchramas edingiz
48	$V(x) + u_1(6) + z_2 ko'p3 \Rightarrow V_3$	uchra+mas \cup edi +lar \Rightarrow uchramas edilar
49	$V(x) + u_1(7) + z_2bir1 \Rightarrow V_3$	uchra+yotgan \cup edi +m \Rightarrow uchrayotgan edim
50	$V(x) + u_1(7) + z_2bir2 \Rightarrow V_3$	uchra+yotgan \cup edi +ng \Rightarrow uchrayotgan eding
51	$V(x) + u_1(7) + z_2bir3 \Rightarrow V_3$	uchra+yotgan \cup edi + $\emptyset \Rightarrow$ uchrayotgan edi
52	$V(x) + u_3(3) + z_1bir1 \Rightarrow V_3$	uchra+moqchi+man \Rightarrow uchramoqchiman
52	$V(x) + u_3(3) + z_1bir2 \Rightarrow V_3$	uchra+moqchi+san \Rightarrow uchramoqchisan

54	$V(x) + u_3(3) + z_1 \text{bir}3 \Rightarrow V_3$	uchra+moqchi+ $\emptyset \Rightarrow$ uchramoqchi
55	$V(x) + u_2(1) + z_1 \text{ko'p}1 \Rightarrow V_3$	uchra+yap+miz \Rightarrow uchrayapmiz
56	$V(x) + u_2(1) + a_1 \text{ko'p}2 \Rightarrow V_3$	uchra+yap+siz \Rightarrow uchrayapsiz
57	$V(x) + u_2(1^1) + z_1 \text{ko'p}3 \Rightarrow V_3$	uchra+yapti+lar \Rightarrow uchrayaptilar
58	$V(x) + u_3(1) + z_1 \text{ko'p}1 \Rightarrow V_3$	uchra+y+miz \Rightarrow uchraymiz
59	$V(x) + u_3(1) + z_1 \text{ko'p}2 \Rightarrow V_3$	uchra+y+siz \Rightarrow uchraysiz
60	$V(x) + u_3(1) + z_1 \text{ko'p}3 \Rightarrow V_3$	uchra+y+dilar \Rightarrow uchraydilar

7-jadval

Matn tarkibida aniqlangan ko'makchi fe'l vazifasidagi mustaqil fe'llar (V):

ber (-ver), bil, bit (bitir), bor, boshla, boq, bo'l, et, yoz, yot, kel, ket, ko'r, ol, sol, tashla, tur, tush, chiq, yubor, yur, o'l, o't, o'tir, qara, qol, qo'y

8-jadval

Imlo lug'atga kiritilmagan, matnda uchraydigan sifatlar (Adj):

Sifat turlari	Sifatlar
Har ikki qismi yakka holda ishlatilmaydigan juft sifatlar	zim-ziyo, ilang-bilang, poyintar-soyintar (gap), dali-g'uli, zim-ziyo, ilang-bilang, aji-buji (yozuv), almoyi-aljoyi, apoq-chapoq, aloq-chaloq (tush), uvali-juvali, o'poq-so'poq
Takroriy sifat	burda-burda, yo'l-yo'l (otdan ko'chgan), oppoq-oppoq, yaxshi-yaxshi, katta-katta, shirin-shirin.

9-jadval

Ravishning 487 ko'rinishi aniqlandi:

ablhlarcha	apir-shapir	ahyon-ahyonda
avval	arang	bag'oyat
avvalgiday	asta	badastir
azza-bazza	asta-asta	bajonudil
aytganidek	asta-sekin	bazo'r
aksincha	astoydil	baqadrihol
allazamonlar	atay	baloday (otdan ko'chgan)
allamahall	ataydan	bamaylixotir
allanechuk	atayin	bamaslahat
allapayt	ataylab	banogoh
ancha	atigi	baravariga
apil-tapil	axir	

baralla	bir pas	butunlay
barvaqt	bir tekis	butunligicha
basma-bas	birakayiga	vaqtincha
batamom	birato‘la	vahshiyona
bafurja	birato‘lasi	vahshiylarcha
bahamjihat	birvarakay	vijdonan
bahodirona	birvarakayiga	galma-gal
baholiqudrat	birvarakasiga	gohida
bahuzur	birga	goho
ba‘zan	birga-birga	gul-gul
ba‘zida	birgalikda	dabdurustdan
beadad	birda	darajama-daraja
beayov	birdaniga	darrov
bevaqt	birin-ketin	darhol
bevosita	biryo‘la	dastavval
bedarak	birma-bir	dastlab
beijozat	birmuncha	daf‘atan
beixtiyor	birmuncha	devoriy
bekami-ko‘st	birnav	dehqonchasiga
bekitqicha	birnavi	deyarli
bekorga	biroz	doim
bekorgayam	biror	doimgidek
bekordan-bekorga	birpas	doimo
bemalol	birtalay	dom-daraksiz
beso‘roq	birtekis	donalab
betinim	bittalab	do‘stlarcha
betinim	bolalarcha	do‘stona
betma-bet	boricha	yovvoyilarcha
beto‘xtov	boshdan-oyoq	yodaki
bexos	boshqacha	yozin-qishin
bexosdan	boshma-bosh	yonlamasiga
bilinar-bilinmas	boya	yoppasiga
bir boshdan	bu yerga	yetarlicha
bir dam	bu yerdan	yoshlayin
bir yoqlama	bu orada	javoban
bir zamon	bo‘g‘inlab	jiq
bir zumda	bugun	jiq-jiq
bir yo‘la	bultur	jiqqa
bir kuni	bunda	jilla
bir nafas	burun	jinday
bir on	butunisicha	jichcha
bir onda	butunlay	juda

juda ham
jo'rttaga
jo'rttaga
zarracha
zimdan
zinhor-bazinhor
zo'raki
zo'rg'a
yigitchasiga
ilalabad
ilgari
ilgariga
ilgarida
ilgaridan
imi-jimida
imi-jimida
indamasdan
indin
indinga
indinda
indindan
istagandek
istagancha
istagancha
itoatkorona
itoatkorona
yo'l-yo'lakay
keyin
keyinga
keyincha
keragicha
keragicha
kech
kecha-kunduz
kechasi
kechin
kechqurun
kechroq
kilolab
ko'ylakchang
kunda
ko'ndalangiga

kundek
kunduzi
kunduzi
ko'p
ko'pincha
ko'plab
ko'r-ko'rona
qadimdan
qay vaqt
qay zamon
qay mahal
qay payt
qay choq
qayta
qayta-qayta
qanaqasiga
qanaqasiga
qardoshlarcha
qasddan
qatorasiga
qat'iyon
qahramonlarcha
qiyosan
qiyg'os
qisqasi
qittak
qishin-yozin
qishloqcha
qishloqchasiga
qonunan
qop-qop
quyiga
quyida
quyidan
qo'qqisdan
qo'qqisdan
mazmunan
mazmuni
mangu
mardlarcha
mardona
mahsichang

millionlab
millionlarcha
minglab
mo'g'ulchasiga
mudom
muqaddam
mo'l-ko'l
mutlaqo
navbatma-navbat
navbatsiz
nari-beri
naridan-beri
noiloj
noilojlikdan
nomiga
oz
ozgacha
ozmi-ko'pmi
oz-moz
ozmuncha
oyda-yilda
oydek
oylab
oysimon
oqilona
oqshomlari
olg'a
oldin
oldinga
oldinma-keyin
oldinma-keyin
oliftanamo
(gerdaymoq)
onda-sonda
ora-sira
ora-chora
osonlikcha
otalarcha
oxiri
ohista
ochiqchasiga
ochin-to'qin

paydar-pay	tezda	o'lgudek
piyoda	tezlikda	o'lda-jo'lda
pinhona	tezroq	umrbod
rasman	tez-tez	o'ngga
rosa	tekinga	o'ngdan
rostakam	telbalarcha	unda-bunda
rostakamiga	teskarisiga	ura-sura
rosti	tikkama-tikka	o'rtaga
ro'y-rost	tikkasiga	o'rtada
ro'para	tiriklay	o'rtadan
ruhlan	tiriklayin	o'sha yoqdan
savat-savat	tirnoqcha	o'sha zamon
sal	tongla	uyalgannamo (yuzini
sal burunroq	tonnalab	chetga burmoq)
sal-pal	tugal	fidoyilarcha
saharlab	to'g'ridan-to'g'ri	fidokorona
sekin	tun-kun	fikran
sekinroq	tunlari	xayolan
sekin-sekin	to'satdan	xafasimon
sidirg'asiga	to'xtovsiz	xiyol
sizcha	u yoqda	xiyla
sizcha	u yoqdan	xolisona
sira	u yoqqa	xomligicha
sira-sira	u yerga	hazilnamo
soldatchasiga	u yerda	hazilsimon
so'ng	u yerdan	hali
so'ngra	o'g'rincha	hali-beri
surunkasiga	o'zbekcha	halitdan
tag'in	o'zgacha	hamisha
takror-takror	uzzukun	hamma vaqt
talay	o'zidan-o'zi	hamma yoq
tappa	uzil-kesil	hamma yer
tarixan	o'zicha	hamma joy
tasodifan	o'zicha	hamma zamonda
tasodifan	uzoq	hamon
taxminan	o'z-o'zidan	hanuz
tashqariga	uzunasiga	har gal
tashqarida	uzundan-uzoq	har dam
tashqaridan	uzundan-uzoq	har yoqdan
ta'kidlaganidek	o'ylamasdan	har yoqqa
tegishlicha	o'qdek	har yerda
tez	o'qtin-o'qtin	har zamonda

har yili	chor-nochor	ertan
har kuni	shaklan	ertan-kechin
har lahza	shapir-shupur	erta-yu kech
har taraf	shaxsan	erta-yu kech
harbiychasiga	shu asnoda	es-es
hartaraf	shu asnoda	eskidan
harflab	shu yerga	eskicha
haftalab	shu yerda	eson-omon
hech yoqda	shu yerdan	es-es
hech yoqdan	shu zamon	ehtiyotkorona
hech yoqqa	shu zahoti	yayov
hozir	shu kuni	yaqinda
hozircha	shu onda	yaqinda
huda-behuda	shu o'rtada	yalpisiga
chaqqon-chaqqon	shunda	yalpisicha
chala-chulpa	elan-qaran	yana-tag'in
chalqancha	emin-erkin	yangiligicha
chalqanchasiga	endi	yangitdan
chamasi	eran-qaran	yangicha
chambarchas	erkalangansimon	yaxshilab
chapga	erkalangansimon	yaxshilikcha
chapda	eronchasiga	yaxshisi
chapdan	erta	yashinday
charchamasdan	erta-indin	yashirincha
chinakamiga	ertalab	
chippa	ertan	

10-jadval

Olmosh (Pron)ning 126 ko'rinishi aniqlandi:

№	Olmosh turi	olmoshlar
1.	Kishilik olmoshi	men [tarixiy shakllari: kamina, banda, qulingiz], sen, u, biz, siz, ular
2.	O'zlik olmoshi	o'z (o'zim, o'zing, o'zi; o'zimiz, o'zingiz, o'zlari)
3.	Ko'rsatish olmoshi	ana shu, ana, anovi, bu, bunaqa, bunday, buncha, mana bu, mana shu, mana, u, unaqa, unday, uncha, shu, o'sha, o'shanaqa, o'shanday, o'shancha
4.	So'roq olmoshi	kim, nima, qayer, qanday, qanaqa, qaysi, qancha, nechta, necha, nechanchi, qachon, qayerda, qayerga, qayerdan, qayoqqa, qayoqda, qayoqdan, nega, nechun

5.	Belgilash-jamlash olmoshi	hamma, barcha, bari, jami, jamiki, yalpi, butun, har kim, har nima, har bir, har qaysi, har qachon, har qayerda
6.	Bo'lishsizlik olmoshi	hech bir, hech narsa, hech kim, hech nima, hech qanday, hech qanaqa, hech qaysi, hech qancha, hech qachon, hech qayerda
7.	Gumon olmoshi	kimdir, nimadir, qandaydir, qanaqadir, qaysidir, qachondir, qayerdadir; allakim, allanima, allaqayer, allaqancha; bir narsa, bir kishi, bir qancha, bir nechta, bir necha; birov, biron(ta), biror(ta), falon(chi), fiston(chi)

So'roq olmoshlarining matnlardagi quyidagi ko'rinishlari aniqlandi:

kim, nima, necha, kim, kimim, kimingiz, kimlar, kimdagi, kimmiki, kimning, kimga, kimdan, kimmiki, nega, nechun, negadir, nimagadir, nimadandir, nechta, nechanchi, nechuk, qanday, qanaqa, qandaydir, qaysi, qaysisi, qaysiniki, qayer, qayerda, qayerga, qayerdan, qayoqqa, qayoqda, qayoqdan, qancha, qanchacha (taxminiy), qanchadan, qachon, qani, qachondan, qachongacha, qachonlar, qanisan, qanisiz, qalay, qalaysan, qalaysiz

11-jadval

Aniqlangan ko'makchilar (RostP):

Ko'makchi turlari	Ko'makchilar
Sof ko'makchi	bilan, uchun, sari, sayin, kabi, singari, yanglig', qadar, dovur, to, doir, oid.
Vazifadosh ko'makchilar	tomon, qarab, qaramay, qaraganda, qaramasdan, boshqa, bo'lak, tashqari, sababli, tufayli, so'ng, oldin, keyin, beri, ko'ra, bo'yicha, uzra, chog'i, osha, bo'ylab, ichra, bo'yi, chamasi, haqida, to'g'risida, holda, yo'sinda, qarshi, yarasha, asosan, binoan, muvofiq, qarata, o'zga, buyon, burun, ilgari, boshlab, badaliga, davomida, vajidan, evaziga, borasida, bobida, misoli, tortib; [ost, ust, tepa, old, orqa, yon, ich, ora, bosh, o'rta, tag, bo'y]+[-i/-si (III shaxs egalik) + -ga/-da/-dan (kelishik qo'shimchalari bilan)]

Bog'lovchilar – Conj

Bog'lovchi turlari		Bog'lovchilar
Qo'llanishiga ko'ra	Yakka qo'llanadigan bog'lovchilar	a, va, hamda, ammo, lekin, biroq, balki, holbuki, vaholanki, gar, agar, agarda, basharti, chunki, garchi, garchand, go'yo, go'yoki, -ki, -kim, esa, ya'ni, shuning uchun, <i>yo, yoki, yoxud, yoinki</i>
	Takroriy qo'llanadigan bog'lovchilar	dam ... dam, ba'zan ... ba'zan, ham ... ham, bir ... bir, xoh ... xoh, goh ... goh, <i>yo ... yo, yoki ... yoki, yoxud ... yoxud,</i>
Vazifasiga ko'ra	Teng bog'lovchilar	va, hamda, bilan (ko'makchi), ham, -u//-yu. -da, ammo, lekin, biroq, esa, balki, holbuki, vaholanki, a, yo//yoki//yoxud//yoinki, yo-yo, goh-goh, dam-dam, ba'zan-ba'zan, bir-bir, xoh-xoh
	Ergashtiruvchi bog'lovchilar	ya'ni, -ki//-kim; chunki, shuning uchun; agar//gar//agarda, basharti, mabodo, garchi//garchand, modomiki; go'yo//go'yoki

Yuklama – Part

Yuklama turlari	Yuklamalar
Affiks-yuklama	-a//-ya, -mi, -chi, -oq//-yoq, -ki, -kim, -gina//-kina//-qina, -dir, -u//-yu, -ku, -da
So'z-yuklama	faqat, axir, hatto, hattoki, na-na, xuddi, naq, nahot, nahotki, ham, xolos
Matnda yuklama vazifasida qo'llanilgan so'zlar	tim, qoq, liq, lim, shilta, g'arq, g'irt

Aniqlangan modallar (Modal):

Modal turlari	Modallar
Sof modallar	albatta, aftidan, darvoqe, tabiiy, mazmuni, mayli, shekilli, chamasi, ehtimol, harqalay, hoynahoy
Modal ma'noli so'zlar	bor, yo'q, kerak, darkor, zarur, lozim, mumkin, iborat
Vazifasiga	aslida, axiri, axiyri, balki, bari bir, vassalom, go'r, go'rga,

ko'ra modal so'zlar	darhaqiqat, deylik, demak, zotan, ishqilib, kel, keling, qo'y, qo'ying, qo'yingki, ma'lumki, muhaqqaq, nafsilambiriga, nihoyat, olaylik, ochig'i, rost, rostdan, rosti, so'zsiz, to'g'ri, to'g'rirog'i, to'g'risi, o'z-o'zidan, umuman, filhaqiqat, xayriyat, xayriyatki, chindan, chog'i, shubhasiz, haytovur, haqiqatan, haqiqatda, har holda, harqalay
---------------------	---

15-jadval

Aniqlangan undovlar (Interj):

Undov turlari	Undovlar
his-hayajon undovlari	ie, eh, o, uh, uf, bay-bay, ho, o'h-ho', ehe, e, hay-hay, a-ha, e-ha, ha-ya, o', i, a, o'h, obbo, iya, he, ie-voy, ura, vo, vodarig', esiz, evoh, vah-vah, ajabo
odamga qaratilgan haydash-chaqirish undovlari	hay, hoy, oy, ay, allo, ey, ma, tss, marsh, qani, jim, tek, fisht, huv
hayvon/parrandaga qaratilgan haydash-chaqirish undovlari	beh-beh, pish-pish, pisht, tu-tu, gah, kuch-kuch, mah-mah, xo'sh, tak, dirr, ish, xix, chu(h), kisht, hayt, xuyt, hayt-hut, hayto-hayt
ko'rsatish, ta'kid undovlari	hu, huv, ha, hmm, hov
rasm-odat undovlari	assalomu alaykum, assalom, astag'firullo, attang, afsus, afsuski, balli, barakalla, bor bo'ling, vaalaykum assalom, omonmisiz, ofarin, marhamat, marhabo, rahmat, salom, tashakkur, uzr, xayr, xo'sh, xo'p, shukur, esonmisiz, qoyil, qulluq, hayfki, hayfkim, hormang

16-jadval

Taqlidlar (Mim):

Taqlid turlari	Taqlidlar
Tovushga taqlid	ak-ak, ang-ang, ar-ar, apshu, baka-bang, baqir-buqur, bag'-bug', biq-biq, bit-bildiq, vang-vang, vov-vov, gangir-gungur, gijbang, guv (shiddatli, bo'g'iq tovush), guldur, gurs, jaz-buz, jiz-biz, inga-inga, miyov-miyov, patir-putur, tapara-tapar, taraqa-turuq, tars, tars-turs, tars-tars, tasira-tusur, taqir-tuqur, taq-taq, taq-tuq, tiq-tiq, xir-xir, xur-xur, chak-chak, chalp-chulp, chilp-chilp, cho'lp-cho'lp, chirs-chirs, chars-churs, shapir-shupur, qars, qars-qars, qars-qurs, qart-qurt, qasir-qusur, qa-qaq, qiy-chuv, qisir, g'art, g'art-g'urt, g'arch, g'arch-g'urch, g'at, g'at-g'at, g'iz, g'irt-g'art, hingir-hingir, ho'ngir-ho'ngir

holat/obrazga taqlid	atak-chechak, alang-jalang, bij-bij, bijir-bijir, vij-vij, vijir-vijir, gar-gar, gir-gir (aylanmoq), guv-guv, gup, dik, dik-dik, dikir-dikir, duv-duv, jig‘, jimir-jimir, jilpang-jilpang, zirq-zirq, yilt-yilt, lapang-lapang, lim-lim, limmo-lim, lov-lov, lop, lo‘q-lo‘q, milt-milt, mo‘lt-mo‘lt, pildir-pildir, shart-shurt, yum-yum (yig‘lamoq), sim-sim (ogrimoq), yalt, yalt-yalt, yalt-yult, yaraq, yarq, g‘ij-g‘ij (aql), g‘im-g‘ij, g‘im-g‘irch, g‘ir-g‘ir (esmoq), hang-mang, hil-hil (pishmoq)
----------------------	--

Egasi olmosh bilan ifodalangan gap kesimining grammatik shakllanishi:

Ega	Kesim	Kesimlik shakli
Men	V+a/ may	-man
Sen		-san
U		-di
Biz		-miz
Siz		-siz
Ular		-dilar

Ega	Kesim	Kesimlik shakli
Men	V+di/-madi	-m
Sen		-ng
U		–
Biz		-k
Siz		-ngiz
Ular		– /-lar

17-jadval

Ot yasovchi affikslar

Asos / lemma	Yasovchi affiks	modeli	Yasalma / so‘zshakl
ot	-bin	N N morfems = N leksema N +	folbin
ot	-bon		darvozabon, soyabon, tarozibon, xazinabon
ot	-boz		masxaraboz, qimorboz, dorboz, xotinboz
ot	-voy		novvoy
ot	-gar		zargar, savdogar, da‘vogar, kimyogar
ot	-kar		miskar
ot	-garchilik		yog‘ingarchilik, odamgarchilik
ot	-gir		jahongir, fazogir

ot	-goh		oromg'oh, saylgoh, qarorgoh, ziyoratgoh
ot	-go'y		maslahatgo'y, nasihatgo'y
ot	-diq		o'rindiq
ot	-don		guldon, kuldon, qalamdon, siyohdon
ot	-dor		muhrdor, chorvador, safdor
ot	-dosh		sinfdosh, maslakdosh, kursdosh, safdosh
ot	-do'z		etikdo'z, mahsido'z, kashtado'z, gilando'z
ot	-zor		olmazor, gulzor, mevazor, lolazor
ot	-iston		guliston, go'riston, O'zbekiston
ot	-kash		aravakash, qalamkash, suratkash
ot	-kor		ganchkor, paxtakor, san'atkor
ot	-kov		go'rkov
ot	-lik		bolalik, vaqtichog'lik, do'stlik
ot	-liq		boshliq
ot	-loq		o'tloq, qumloq, toshloq
ot	-noma		taklifnoma, tabriknoma, pandnoma
ot	-navis		tarixnavis, hikoyanavis, xamsanavis
ot	-paz		oshpaz, kabobpaz, somsapaz
ot	-soz		soatsoz, kemasoz
ot	-furush		nosfurush
ot	-xon		kitobxon, she'rxon
ot	-xona		darsxona, ishxona, darsxona, dorixona, devonxona
ot	-xo'r		merosxo'r, poraxo'r, sudxo'r
ot	-cha		qalamcha, zarracha, sichqoncha
ot	-choq		o'yinchoq
ot	-chi		ishchi, temirchi, terimchi
ot	-chilik		hunarmandchilik, o'zbekchilik, dehqonchilik
ot	-shunos		tilshunos, tabiatshunos, uslubshunos
ot	ham-		hamqishloq, hamshahar, hamyurt
sifat	-a	Adj + N morfems = N leksema	bo'za, quyqa
sifat	-a (2)		xaroba, vayrona
sifat	-archilik		ocharchilik
sifat	-at		ko'kat
sifat	-garchilik		namgarchilik, xafagarchilik, xunobgarchilik, sharmandagarchilik
sifat	-ot		ma'lumot, mushkulot

sifat	-yat		madaniyat, majburiyat
sifat	-lik		yaxshilik, ezgulik, yomonlik, saxiylik
sifat	-cha		qizilcha, olacha
sifat	-chi		qiziqchi
sifat	-chilik		pishiqchilik, arzonchilik
fe'l	-a	V + N morfems = N leksema	yara (yor∪a)
fe'l	-at		ko'chat, o'lat
fe'l	-vchi		o'quvchi (o'qi∪vchi)
fe'l	-uvchi		yozuvchi (yoz∪uvchi), uchuvchi
fe'l	-gi		sezgi, sevgi, supurgi
fe'l	-ki		turtki, tepki
fe'l	-qi		sanchqi
fe'l	-g'i		tomizg'i
fe'l	-g'u		tuyg'u
fe'l	-gich		kulgich, o'tkazgich
fe'l	-kich		ko'rsatkich
fe'l	-qich		yoritqich, tutqich, ochqich
fe'l	-g'ich		to'g'nag'ich, chizg'ich, o'chirg'ich
fe'l	-gin		tizgin
fe'l	-qin		to'lqin, toshqin, to'sqin, bosqin
fe'l	-g'in		yong'in, quvg'in
fe'l	-qun		uchqun
fe'l	-dak		yugurdak, kekirdak
fe'l	-doq		qovurdoq, qo'ndoq
fe'l	-ik		ko'rik, teshik, kekirik, bitik
fe'l	-imlik		o'simlik, ichimlik
fe'l	-in		yig'in, yog'in, ekin, tiqin
fe'l	-un		tugun, tutun
fe'l	-indi		chiqindi, cho'kindi
fe'l	-ndi		chirindi
fe'l	-undi		yuvundi
fe'l	-it		chiqit
fe'l	-ich		cho'mich, cho'kich, o'pich, bog'ich
fe'l	-ish		qarg'ish
fe'l	-iq		chaqiriq, kesatiq, yotqiziq, chopiq
fe'l	-uq		yutuq, buyruq
fe'l	-k		ko'rk, elak, tilak, kurak, bezak, to'shak
fe'l	-kilik	ichkilik	

fe'l	-gilik		ko'rgilik
fe'l	-m		to'plam, ho'plam, chidam, tishlam
fe'l	-im		kechirim, qo'nim, terim, chiqim, bitim, bosim
fe'l	-um		unum, tuzum
fe'l	-ma		surma, o'sma, tortma, suzma, toshma
fe'l	-mak		emak, emak
fe'l	-moq		ilmoq, chaqmoq, topishmoq
fe'l	-machoq		bekinmachoq, quvlashmachoq
fe'l	-mish		o'tmish, kechmish, qilmish
fe'l	-movchilik		anglashilmovchilik, kelishmovchilik, etishmovchilik
fe'l	-on		qiron, to'zon
fe'l	-ch		sevinch, quvonch, yupanch
fe'l	-j		ilinj
fe'l	-inch		qo'rqinch
fe'l	-cha		tushuncha
fe'l	-chak		belanchak
fe'l	-choq		ovunchoq, taqinchoq
fe'l	-chi		suyunchi, tilanchi, tomchi
fe'l	-chiq		suyanchiq, yopinchiq
fe'l	-q		charchoq, taroq, qayroq, qiynoq
fe'l	-oq		o'roq, yotoq
ravish	-ik	Adv + N morfems = N leksema	ko'pik
ravish	-lik		tezlik, sekinlik, birgalik
ravish	-chilik		ko'pchilik, ozchilik
taqlid	-ildoq	Mim+N morfems = N leksema	hiqildoq, chirildoq
taqlid	-a (1)		qahqaha, sharshara, jizza, do'mbira, dag'dag'a
taqlid	-ak		bizbizak, pirpirak, guldirak, varrak, qarsak, xurak
taqlid	-um		qultum
taqlid	-os		uvvos, chuvvos
son	-lik	Num+ N morfems = N leksema	birlik, to'rtlik, o'nlik
olmosh	-lik	Pron+ N morfems = N leksema	o'zlik

undov	-chi	Interj+ N morfems = N leksema	haybarakallachi
modal	-lik	Modal+ N morfems = N leksema	yo'qlik, borlik

18-jadval

Fe'l yasovchi affikslar

Asos / lemma	Fe'l yasovchi affiks – V morfems	modeli	Yasalma / so'zshakl
ot	-a	N + V morfems = Vleksema	sana, tuna, yo'la, yasha, osha, qona, o'yna
ot	-an		kuchan
ot	-i		ranji, changi, avji
ot	-ik		ko'zik
ot	-ir		gapir
ot	-ay		kuchay
ot	-ka		iska
ot	-la		egovla, mog'orla, bolala, gulla, yog'la, ishla, arrala, zangla, egalla (ega∪la), ko'zla, bog'la, boshla
ot	-lan		faxrlan, gavdalan, lazzatlan, zavqlan, jonlan, xavotirlan, ovqatlan, foydalan
ot	-lash		do'stlash, suhbatlash, ahdlash, bahslash, maslahatlash
ot	-lashtir		sahnalashtir, kompyuterlashtir
ot	-sit		aybsit
ot	-sira		suvsira, tuzsira, xayolsira, qonsira, ersira, yog'sira, xavfsira, gumonsira, hadiksira
ot	-q		orziq (orzu∪q),
ot	-iq		yo'liq
ot	-qar		boshqar
ot	-g'ar		jamg'ar
sifat	-a	Adj + V morfems = Vleksema	bo'sha, qiyna
sifat	-i		tinchi, boyi
sifat	-ay		toray, kengay, keksay, qoray, ulg'ay (ulg'∪ay), pasay (past∪ay), susay (sust∪ay)
sifat	-la		maydala, tozala, yaxshila, tikla, tekisla, yomonla
sifat	-lan		shodlan, achchiqlan, sergaklan
sifat	-lash		yiriklash, osonlash, asabiylash,

			qiyinlash, yaqinlash, og'irlash, oydinlash, uzoqlash, qo'llash, go'zallash
sifat	-lashtir		soxtalashtir, qalbakilashtir
sifat	-r		o'zgar, eskir, qisqar
sifat	-ar		oqar, ko'kar, qizar
sifat	-si		garangsi
sifat	-sin		qiziqsin, yotsin
sifat	-sira		yotsira, begonasira
sifat	-t		to'lat
sifat	-it		berkit
sifat	-iq		dimiq, zoriq, namiq, zo'riq
son	-ik	Num + V morfems = Vleksema	birik
son	-lan		ikkilan
son	-lash		birlash
ravish	-ik	Adv + V morfems = Vleksema	kechik
ravish	-ir		ko'pir
ravish	-ay		kamay, ozay, ko'pay
ravish	-la		tezla, sekinla
ravish	-lash		birgalash, tezlash
ravish	-sit		kamsit
ravish	-chi		ko'pchi
olmosh	-la	Pron + V morfems = Vleksema	sizla, senla
olmosh	-si		mensi
olmosh	-sira		sensira, sizensira
Taqlid so'z	-a	Mim+ V morfems = Vleksema	shildira, hilpira, g'ildira, guldura
taqlid	-illa		chirqilla, guvilla, zirilla, lovilla
taqlid	-ur		tupur (tufur)
taqlid	-ira		sharqira, yaltira, zirqira, yarqira, mo'ltira
taqlid	-la		gumburla, shivirla
taqlid	-ra		ho'ngra, mara, ingra
taqlid	-shi		g'ingshi
taqlid	-qir		hayqir
undov	-la	Interj + V morfems = Vleksema	dodla, chuhla
modal	-la	+ V morfems = Vleksema	yo'qla
modal	-ol		yo'qol
modal	-ot		yo'qot

Sifat yasovchi affikslar

Asos / lemma	Yasovchi affiks	modeli	Yasalma / soʻzshakl
ot	-aki	N + Adj morfems = Adj leksema	ogʻzaki, qalbaki, yuzaki
ot	-al		regional, gorizonta, monumental, patriarxal
ot	-aro		xalqaro, davlatlararo
ot	ba-		basavlat, badavlat, bahaybat, bamaʼni
ot	bo-		boadab
ot	bad-		badhazm, badbaxt, badnafs, badbashara
ot	bar-		barqaror, barkamol, barhayot
ot	-baxsh		orombaxsh, hayotbaxsh, shifobaxsh
ot	be-		begʻubor, bedavo, betashvish, bexosiyat, beor
ot	-boz		vaʼdaboz, safsataboz
ot	-bon		mehribon [mehr ∪ i ∪ bon]
ot	-bop		ommabop, qishbop
ot	-vash		parivash, devvash
ot	-vor		umidvor
ot	-gay		kungay
ot	-gar		sitamgar, igʻvogar, hiylagar
ot	-gi		kuzgi, bahorgi, yozgi
ot	-ki		qishki, ichki
ot	-qi		sirtqi, tashqi
ot	-gin		gʻamgin
ot	-gun		gulgun
ot	-goʻy		xushomadgoʻy
ot	-don		bilimdon, gapdon, qarddon
ot	-dor		azador, aloqador, aybdor, guldor, daxldor
ot	-dosh		ohangdosh, fikrdosh, sirdosh, kursdosh
ot	-zada		alamzada, yurakzada
ot	-i		qozoqi (it), qishloqi
ot	-iv	obʻektiv, subʻektiv	
ot	-iy	ilmiy, shaxsiy, axloqiy,	

		oilaviy, diniy (jinoiy, maishiy, lugʻaviy, siyosiy, madaniy kabi soʻzlarda bu affiks ajralmaydi.)
ot	-ik	faktik, geologik, filologik
ot	-kash	zahmatkash, jafokash, dardkash, dilkash, hazilkash
ot	-kor	zulmkor, isyonkor, omilkor, fidokor
ot	-li	aqlli, gulli, kuchli
ot	-lik	yozlik (kiyim), koʻrpalik (mato)
ot	-mand	orzumand, davlatmand
ot	-namo	darveshnamo
ot	no-	noinsof, noumid
ot	-nok	alamnok, gʻannok
ot	-on	charagʻon, zaʻfaron
ot	-an	tuban
ot	-omuz	hayratomuz, haqoratomuz
ot	-parast	mansabparast, xayotparast, maishatparast
ot	-parvar	oilaparvar, xalqparvar
ot	-saro	motamsaro
ot	ser-	sergoʻsht, serhosil, sersomon
ot	-siz	oʻrinsiz, tuzsiz, ishsiz
ot	-simon	naysimon, sharsimon, maymunsimon, oysimon
ot	-xoʻr	poraxoʻr, oshxoʻr, sudxoʻr
ot	xush-	xushmanzara, xushfeʻl, xushsurat, xushqomat
ot	-cha	oʻzbekcha, fargʻonacha (kurash)
ot	-chan	kurashchan, talabchan, harakatchan
ot	-chi	ayirmachi
ot	-chil	xalqchil, epchil, dardchil
ot	-shumul	olamshumul, jahonshumul
ot	-qa	loyqa
ot	ham-	hamnafas, hamfikir, hamshahar

fe'1	-a	to'la, aylana, ko'tara (savdo)
fe'1	-ag'on	bilag'on, topag'on, chopag'on
fe'1	-vuch	iskovuch
fe'1	-ovuch	hurkovuch
fe'1	-gir	sezgir
fe'1	-kir	keskir, o'tkir
fe'1	-qir	topqir, chopqir
fe'1	-g'ir	olg'ir
fe'1	-qur	uchqur
fe'1	-diq	tashlandiq (tashla \cup n \cup diq)
fe'1	-imli	to'yimli, sevimli, yoqimli, o'timli
fe'1	-in	to'kin, to'lin, sog'in, rangin, g'ishtin
fe'1	-inki	ko'tarinki
fe'1	-inch	shoshilinch, tiqilinch
fe'1	-k	chirik, o'ksik
fe'1	-q	quruq
fe'1	-g'	yorug'
fe'1	-ik	o'lik, egik
fe'1	-iq	yopiq, ochiq, bosiq
fe'1	-uq	buzuq, yumuq
fe'1	-kun	tushkun
fe'1	-qin	toshqin, jo'shqin
fe'1	-g'in	so'lg'in, horg'in
fe'1	-g'un	turg'un
fe'1	-l	tugal
fe'1	-m	qaram
fe'1	-ma	qaynatma, ko'tarma, qo'shma, ko'chma, tug'ma
fe'1	-mas	yaramas, o'tmas, qaytmas, qo'rqmas
fe'1	-mon	bilarmon, ustabuzarmon, toparmon
fe'1	-msiq	qarimsiq, achimsiq
fe'1	-ndi	asrandi, tashlandi
fe'1	-ong'ich	so'kong'ich, tepong'ich
fe'1	-ng'ich	tishlong'ich [tishl [a \rightarrow o] \cup ng'ich]
fe'1	-oq	qo'rqoq, baqiroq, qoloq

V + Adj morfems = Adj leksema

fe'1	-rli		arzirli
fe'1	-arli		ishonarli, zerikarli, aytarli
fe'1	-ch		tinch, jirkanch
fe'1	-chak		kuyunchak
fe'1	-choq		tortinchoq, erinchoq
fe'1	-chiq		sirpanchiq, qizg'anchiq
harakat nomi	-chan		yashovchan, sinovchan, o'suvchan
fe'1	-shta		sinashta
fe'1	-q		oqsoq, porloq, quvnoq [oqs [a→o] ∪ q]
fe'1	-qa		qisqa
fe'1	-qi		vaysaqi, sayroqi, o'ynoqi [o'yn [a→o] ∪ qi]
fe'1	-qoq		yopishqoq, urishqoq, uyushqoq, tirishqoq
sifat	-avon	Adj + Adj morfems = Adj leksema	zo'ravon
sifat	-aki		xomaki, zo'raki
sifat	-bin		kaltabin, nekbin
sifat	-vor		ulug'vor
sifat	-kay		terskay
sifat	-lom		sog'lom
sifat	-namo		tentaknamo
sifat	no-		noto'g'ri, noma'lum, nomard
sifat	-xo'r		tekinxo'r
sifat	-chi		a'lochi, bekorchi
sifat	g'ayri-		g'ayritabiiy, g'ayriqonuniy
olmosh	-aro		Pron+ Adj morfems = Adj leksema
olmosh	-bop		hammabop, sizbop
ravish	-ki	Mem+ Adj morfems = Adj leksema	ertalabki
ravish	-gi		avvalgi
ravish	-chil		kamchil
modal	-iy	Modal+ Adj morfems = Adj leksema	zaruriy
modal	-li		kerakli
taqlid	-aki	Mim+ Adj morfems = Adj leksema	shartaki
taqlid	-ildoq		so'lqildoq, likildoq, chiyildoq
taqlid	-os		sharros, gulduros

Ravish yasovchi affikslar

asos / lemma	Yasovchi affiks	modeli	Yasalma / soʻzshakl
ot	-an	N + Adv morfems = Adv leksema	tasodifan, vijdonan, taxminan, fikran, xayolan, ruhan, fikran, rasman
ot	bar-		barvaqt
ot	be-		beijozat, beixtiyor, bevosita, bedarak, bemalol, betinim, bevaqt
ot	-dek		oʻqdek, kundek, oydek
ot	-day		yashinday
ot	-iga		nomiga
ot	-siga		surunkasiga, qatorasiga
ot	-lab		oylab, donalab, haftalab, harflab, boʻgʻinlab, tonnalab
ot	-larcha		qahramonlarcha, bolalarcha, doʻstlarcha, otalarcha
ot	-namo		hazilnamo
ot	-in		qishin-yozin, oldin
ot	-ona		doʻstona
ot	-siz		dom-daraksiz, navbatsiz
ot	-simon		hazilsimon, oysimon
ot	-cha		tirmoqcha, qishloqcha, yaxshilikcha, osonlikcha, oʻzbekcha, oʻgʻrincha [oʻgʻri∪cha], vaqtincha [vaqt∪in∪cha].
ot	-chasiga		qishloqchasiga, dehqonchasiga, yigitchasiga
ot	-chang		mahsichang, koʻylakchang
ot	-aki		yodaki, zoʻraki
sifat	-an	Adj + Adv morfems = Adv leksema	qatʼiyan
sifat	-day		avvalgiday
sifat	-dek		doimgidek
sifat	-iga		rostakamiga, koʻndalangiga, chinakamiga, baravariga
sifat	-siga		uzunasiga, tikkasiga, yoppasiga
sifat	-lab		yaxshilab
sifat	-lay		tiriklay
sifat	-layin		tiriklayin, yoshlayin
sifat	-larcha		vahshiylarcha, mardlarcha, yovvoyilarcha, telbalarcha
sifat	-ligicha		yangiligicha, xomligicha, butunligicha

sifat	-namo		oliftanamo (gerdaymoq)
sifat	-in		ochin-to‘qin
sifat	-ona		mardona, xolisona, oqilona, fidokorona, itoatkorona
sifat	-simon		xafasimon
sifat	-cha		o‘zgacha, boshqacha, yangicha, yashirincha, eskicha
sifat	-chasiga		ochiqchasiga, harbiychasiga
ravish	-iga	Adv + Adv morfems = Adv leksema	birdaniga
ravish	-lab		ko‘plab, saharlab
ravish	-n / -in		ertan-kechin
ravish	-cha		ko‘pincha [ko‘pʊinʊcha], hozircha, keyincha
fe‘l	-dek	V + Adv morfems = Adv leksema	istagandek, aytganidek, ta‘kidlaganidek
sifatdosh	-namo		uyalgannamo (yuzini chetga burmoq)
sifatdosh	-simon		erkalangansimon
sifatdosh	-cha		istagancha
harakat nomi	-siz		to‘xtovsiz
olmosh	-siga	Pron + Adv morfems = Adv leksema	qanaqasiga, yalpisiga
olmosh	-lay		butunlay
olmosh	-cha		sizcha, o‘zicha
modal	-an	Modal+ Adv morfems = Adv leksema	majburan
modal	-cha		keragicha [kerakʊiʊcha], boricha [borʊiʊcha]

GLOSSÁRIY

Avtomatik hosilá – bu soʻzlarga qoʻshimchalar qoʻshish jarayonida hosila soʻzda roʻy beradigan tovush oʻzgarishlarining lingvistik taʼminotda aynan saqlanishidir.

Aranjiróvka (fr. *arranger* – “tartibga solish, tartiblash”) – muayyan musiqiy asarni birinchi nusxasidan farqli shaklda tayyorlash, tinglovchilar auditoriyasiga moslashtirish, uni aʼlo darajada yetkazish sanʼati.

Mazkur tadqiqotda aranjirovka termini etimologiyasiga muvofiq ravishda affikslarni ularning birikib kelish ketma-ketligida tartiblash maʼnosida qoʻllanilgan. Jumladan, *bilimdonlik* yasalmasida soʻz yasovchi affikslar $V+N_{\text{morf}}+Adj_{\text{morf}}+N_{\text{morf}}$ modeliga muvofiq keladi: *bil+im+don+lik*

Artefákt (lot. *artefactum*: arte – sunʼiy, factus – amalga oshirilgan) sunʼiy yaratilgan obyekt yoki predmet, inson faoliyati samarasidir. Madaniy hayotda artefakt sifatida kiyim-kechak, xoʻjalik mollari, uy-joy tushuniladi. Bugungi kunda madaniyatshunoslikda artefakt tushunchasi koʻlami oshib bormoqda. Manbalarda artefakt sinonimlari sifatida *jarayon*, *madaniy meros*, *haykal*, *hodisa* tushunchalari beriladi.

UML modellar tilidagi artefakt dasturiy taʼminotni yaratish jarayonida sistemaning almashtiradigan qismi sifatida platforma darajasida ishlab chiqiladi.

Indeksirlásh (ingl. *Indexing* – koʻrsatkich yaratish) – qidiruv tizimida hujjat, maʼlumot va soʻrovlarni axborot-qidiruv tilida tavsiflash jarayoni. Indeksirlash natijasida har bir maʼlumotga uning semantikasini aks ettiruvchi kalit soʻzlar toʻplami yuzaga keladi.

Shuningdek, indeksirlash kompyuterdagi fayllar, elektron pochta xabarlari va shu kabilarni koʻrish jarayoni va ular haqidagi maʼlumotlar, jumladan, ular tarkibidagi soʻzlarni tasniflash hisoblanadi. Kompyuterda qidiruv amalga oshirilayotganda indeksirlashdan soʻng sistema natijalarni tezroq topish uchun terminlar indeksini tahlil qiladi.

Módul (dasturlashda) – dasturning doimiy, alohida fayl shaklida yaratilgan funksional tugallangan qismi.

Kompyuter lingvistikasida modul termini keng isteʼmolda boʻlib, dasturiy taʼminotning muayyan lingvistik jarayonini qamragan qismi sifatida qoʻllaniladi. Leksik modulda lugʻat qatlami qamrab olinsa,

grafematik modulda ramzlar, tinish belgilar, harfiy va boshqa belgilar tahrir qilinadi, orfografik modulda imlo qoidalari, morfologik modulda soʻzshakllar analizi (soʻzshakldan leksemaga qadar tahlil) va sintezi (leksemaning grammatik shakllanishi tahlili jarayoni), sintaktik modulda supersintaktik birlik – gap yoki soʻzlarning oʻzaro bogʻlanish hodisasi tahlil qilinadi.

Morfologik lugʻat soʻzshaklning, yaʼni ismlarning turlanishi va feʼllarning tuslanishini shartli belgilash maxsus sistemasi orqali aks ettiradigan leksikografik manba hisoblanadi. Lugʻat toʻgʻri yoki teskari alifbo tartibida tuziladi. Teskari (inversion) tartibda soʻzning soʻnggi harfi eʼtiborga olinadi. Lugʻat avvalida turlanish va tuslanish hodisasi namuna tarzida koʻrsatilgan “Grammatik maʼlumotlar” beriladi. Shu bois har bir soʻzda “Grammatik maʼlumot”ga ishora qiluvchi grammatik belgi va indeks boʻladi. Bu jihat foydalanuvchiga muayyan soʻzning shakl oʻzgarish hodisasini aniqlashga yordam beradi. Namunani qidirishni tezlashtirish va qulaylik yaratish uchun lugʻatning har bir sahifasi yuqorisi yoki kolontitulida sahifada uchragan indekslar hisobi va indekslar ifodalagan “Grammatik maʼlumot”lar lugʻatning qaysi sahifasida joylashgani haqida maʼlumot keltiriladi. Quyida taniqli rus leksikografi A.A. Zaliznyakning “Грамматический словарь русского языка” kitobidan fragment berildi:

св (нсв): 1—91; 6—108; 6°—112; △ —133

ДАТЬ

совладать	св нп 1а	недоедать	нсв 1а (есть не до конца) ⊗ недоесть;
падать	нсв нп 1а		нсв нп 1а (голодать)
западать	св нп 1а (начать па- дать)	поедать	нсв 1а ⊗
западать	нсв нп 1а ⊗7	проедать	нсв 1а ⊗ проесть
нападать	св нп 1а [// напасть] (падая, скопиться — о снеге и т. п.) ⊗II	забрeдaть	нсв нп 1а ⊗7(-сти)
¹ нападать	нсв нп 1а (набрасыва- ться; охватывать) ⊗7	набрeдaть	нсв нп 1а ⊗7(-сти)
² нападать	нсв нп 1а (падая, скапливаться — о сне- ге и т. п.) ⊗7 // I	подбрeдaть	нсв нп 1а ⊗7(-сти)
впадать	нсв нп 1а ⊗7	перeбрeдaть	нсв нп 1а ⊗7(-сти)
совпадать	нсв нп 1а ⊗7	взбрeдaть	нсв нп 1а ⊗7(-сти)
подпадать	нсв нп 1а ⊗7	прибрeдaть	нсв нп 1а ⊗7(-сти)
перепадать	св нп 1а (упасть — о многом)	добрeдaть	нсв нп 1а ⊗7(-сти)
перепадать	нсв нп 1а ⊗7	сбрeдaть	нсв нп 1а ⊗7(-сти)
припадать	нсв нп 1а ⊗7	выбрeдaть	нсв нп 1а ⊗7(-сти)
опадать	нсв нп 1а ⊗7	передать	св △b/c[①]: буд. пе- редá м, -шь, -ст, -дйм, -дите, -дут; по- вел. передáй; прош. пéредал, -á, -о, -и [// передáл, -á, -о, -и], передáвший; прич. страд. пéреданный (кф пéредан, -á // -а, -о, -ы) ⊗III
попадать	св нп 1а (упасть — о многих, о многом)		
попалать	нсв нп 1а ⊗7		

Morfologik sintez matnga avtomatik ishlov berish jarayonida yoxud uning soʻnggi – natijalarni foydalanuvchi hukmiga havola qilish bosqichida soʻzlarning turli shakllarini chiqarish, analiz natijalarini

umumlashtirishdan iborat. Leksema shakllarini sintez qilish jarayoni soʻzshakllarni tahlil (analiz) qilish bilan bir xil maʼlumotlar bazasi va algoritmlarga asoslanadi. Shubhasiz, har qanday shaklni sintez qilish uchun taʼminotda leksema va uning grammatik xususiyatlari kiritilishi talab qilinadi, shundagina dastur grammatik parametrlariga ega boʻlgan har qanday leksemaning shakllarini tahlil va sintez qilish imkoniyatiga ega boʻladi.

Sintez jarayoni uchun quyidagi parametrlar (xususiyatlar) zarur:

- 1) otlar uchun jins (oʻzga tillarda), son, kelishik;
- 2) sifat va ravishlarning (mavjud boʻlsa) oddiy, qiyosiy va orttirma darajalari, ular mansub boʻlgan soni va turlanishi;
- 3) feʼl va sifat-doshning zamoni, tuslanishi, shakli.

Meynfréym (yoki meynfreym, ingl. mainframe) – muhim kirish va chiqish resurslariga, katta hajmli operativ va tashqi xotiraga ega va real vaqtda muhim sistemalarda (ingl. mission-critical) foydalanish uchun moʻljallangan ulkan universal yuqori sifatli ishlash qobiliyatiga ega va xatoliklarga chidamli server. Meynfreymnlarni ishlab chiqaruvchi asosiy korporatsiya IBM hisoblanadi.

Meynfreym (M.)larning tarixi 1964-yilda IBM System/360 universal kompyuter sistemasi paydo boʻlishidan boshlanadi, M.ning yaratilishiga IBM 5 mlrd dollar sarfladi. 1960-yillar va 1980-yillarning boshlarida System/360 bozorda liderga aylanadi. Uning klonlari koʻplab mamlakatlarda mavjud. IBM meynfreymnlari dunyo boʻyicha 25 mingdan ortiq tashkilotlarda qoʻllaniladi, shundan 7000 ga yaqini Rossiyadagi tashkilotlar sanaladi. Barcha muhim biznes maʼlumotlarning 70 % meynfreymlarda qayta ishlanadi.

1980-yillar oxiriga kelib M.larga qiziqishning susayishi sabab shaxsiy kompyuterlar va Unixga asoslangan kompyuterlarning rivojlanishiga eʼtibor kuchaydi. Natijada mikrosxemalar yaratilishiga, elektrquvvat sarfining kamaytirilishiga zamin yaratildi. M.lar katta oʻlchamdagi maydonni egallashi va oʻsha davrdagi eskirgan yarimoʻtkazgichlardan foydalanish natijasida uni sovutish uchun suyuqlikning, jumladan, suvning koʻp sarf qilinishi M.lardan foydalanish zaruratini pasaytirdi. Shu sababli 1993-yilga kelib M.larni oʻchirib qoʻyish taklifi ilgari surilgan. Lekin 2013-yildan M.larning yangi versiyalarini ishlab chiqarila boshlandi.

N-grámma – matnlarga avtomatik ishlov berishda keng qoʻllaniladigan matematik hisob vositasidir. N-gramma – soʻzlar yoki

teglar kabi N-identifikator elementlarning ketma-ketligini ifodalaydi. Ikki element ketma-ketligi – bigramma, uch element ketma-ketligi esa trigramma, deyiladi. Masalan, *old qo‘shimcha+ot* holati bigrammaga misol bo‘ladi. Shuningdek, so‘zlar tartibi hisobga olingan modellardan farqli o‘laroq, N-gramma modeli matndagi qo‘shni so‘z qatorlarini aniqlaydi va saqlaydi.

Spell checker (imloni tekshirish tizimi) – kompyuter dasturi bo‘lib, kiritilgan matnning orfografik tekshiruvini amalga oshiradi. Aniqlangan imlo xatolari maxsus belgilanadi, ya‘ni xato yozilgan leksemaning tagiga chiziladi. Ko‘p hollarda matn teruvchiga imloviy xatolarga ishora qilishdan tashqari dastur maxsus eslatmasi sifatida so‘zning to‘g‘ri yozilish variantlarini ham taklif qiladi. Shuningdek, matnga qanday tuzatish kiritish mumkinligiga izohlar ham beriladi.

Lemmatizatsiya so‘zlarning leksikonidan foydalangan holda ularning morfologik tahlilini amalga oshiruvchi aniq jarayon bo‘lib, lemmatizatsiya jarayonida faqat fleksiyaga uchragan affikslar o‘chiriladi va lemma deb atalmish ta‘minotdagi so‘zning asosiy yoki lug‘at shakliga qaytariladi.

Stemming (*stemming*) – so‘zshakldagi affikslarni kesib chiqish (asosga qadar) jarayoni. Bunda topilgan so‘z asosi morfologiyada qabul qilingan so‘z o‘zagiga mos kelishi talab qilinmaydi. Zotan, dastur ta‘minotiga “tayyor qolipli so‘zlar” asos sifatida alohida kategoriya qilib kiritilish ehtimoli yuqori bo‘ladi. Masalan, *huquq, tadqiq* kabi leksemalarga egalik qo‘shimchasi qo‘shilganda leksema so‘ngidagi *q* undoshi o‘zgarishga uchramagani bois, bunday istisnoli leksemalar “tayyor qolipli so‘zlar” kategoriyasiga *huquqi, tadqiqi* tarzida kiritiladi va ular stemmingda asos, deb qabul qilinadi.

Párser yoxud sintaktik analizator – dastur qismi yoki faqat sintaktik tahlilni amalga oshirish uchun yaratilgan maxsus dastur.

Utilíta (ingl. utility) operatsion sistema (OS)ning ishi bilan bog‘liq ravishda maxsus odatiy vazifalarni bajarishga mo‘ljallangan yirik dasturiy ta‘minotning bir qismi bo‘lgan yordamchi kompyuter dasturi.

Fragmént (lot. fragmentum – parcha, qism) – butunning qismi. Dastur ish oynasidan lavha berish, veb-sahifadan namuna ko‘rsatish tushuniladi.

Foydalanilgan adabiyotlar

I. Normativ-huquqiy hujjatlar

1. “O‘zbekiston Respublikasi davlat tili haqida”gi Qonuni. – Toshkent: O‘zbekiston, 1989.
2. Lotin yozuviga asoslangan o‘zbek alifbosi. O‘zbekiston Respublikasining 1995-yil 6-maydagi 71-I-sonli Qonuni.
3. O‘zbek tilining asosiy imlo qoidalari. – Toshkent, 1995-yil 24-avgust, 339-son / <http://lex.uz/docs/1625271>.
4. Ma’naviy-ma’rifiy ishlar samaradorligini oshirish va sohani rivojlantirishni yangi bosqichga ko‘tarish to‘g‘risida. O‘zbekiston Respublikasi Prezidenti Qarori // Xalq so‘zi, 2017-yil, 29-iyul.
5. “Idoraviy normativ-huquqiy hujjatlarni tayyorlash va qabul qilish qoidalarini tasdiqlash to‘g‘risida”gi O‘zbekiston Respublikasi Adliya vazirligining 2014-yil 28-fevraldagi 53-mh-sonli buyrug‘i

II. Ijtimoiy-siyosiy adabiyotlar

6. Mirziyoyev Sh.M. Tanqidiy tahlil, qat’iy tartib intizom va shaxsiy javobgarlik – har bir rahbar faoliyatining kundalik qoidasi bo‘lishi kerak. – Toshkent: O‘zbekiston, 2017. – 104 b.
7. Mirziyoyev Sh.M. Milliy taraqqiyot yo‘limizni qat’iyat bilan davom ettirib, yangi bosqichga ko‘taramiz. I jild. – Toshkent: O‘zbekiston, 2017. – B. 168-174.
8. 2017-2021-yillarda O‘zbekiston Respublikasini rivojlantirishning beshta ustuvor yo‘nalishi bo‘yicha Harakatlar strategiyasi. O‘zbekiston Respublikasi Prezidentining 2017-yil 7-fevraldagi PF-4947-sonli Farmoni.
9. O‘zbekiston Respublikasi Prezidenti Shavkat Mirziyoyevning Oliy Majlisga Murojaatnomasi. – Toshkent: O‘zbekiston, 2018. – 88 b.

III. Ilmiy risola, monografiya, darslik, o‘quv va metodik qo‘llanmalar

10. Do Prado H.A. Emerging Technologies of Text Mining: Techniques and Applications / Ed. by H. A. Do Prado, E. Ferneda. – Idea Group Reference, 2007. – 358 p.
11. Earnest, Les. The First Three Spelling Checkers. Stanford University. Archived from the original (PDF) on 22 October 2012.
12. Earnest, Les. Visible Legacies for Y3K (PDF). Archived from the original on 2012.

13. Harvey, Charlotte Bruce. Teaching Computers to Spell (obituary for Henry Kučera). *Brown Alumni Magazine*. 2010. – P. 79.
14. Hearst, M.A. Automated Discovery of WordNet Relations. In: Fellbaum, C. (ed.) *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, 1998, – P. 131-151.
15. Hirst, G. Ontology and the Lexicon. In.: *Handbook on Ontologies in Niformation Systems*. Berlin, Springer, 2003.
16. Jurafsky D., Martin J. *Speech and Language Processing: An introduction to natural language processing, Computational linguistics, and speech recognition*. Prentice Hall, 2000. – 927 b.
17. Jurafsky, D. *Speech and Language Processing* / D. Jurafsky, J. H. Martin. – 2nd – New Jersey: Prentice Hall, 2008. – 1024 p.
18. Matsumoto Y. Lexical Knowledge Acquisition. In: *The Oxford Handbook of Computational Linguistics*. Mitkov R. (ed.). Oxford University Press, 2003, – R. 395-413.
19. Peterson, James. *Computer Programs for Detecting and Correcting Spelling Errors*. 1980.
20. Po‘latov A.Q., Mo‘minova T., Po‘latova I.O. *Dunyoviy o‘zbek tili (O‘zbek tilida fe‘lning shakllari va ularning rus, ingliz tillarida berilishi)*. – Toshkent: Universitet, 2003. – 404 b.
21. Rahimov A. *Kompyuter lingvistikasi asoslari*. – Toshkent: Akadernashr. 2011. – 160 b.
22. Sayfullayeva R., Mengliyev B., Boqiyeva G. va boshq. *Hozirgi o‘zbek adabiy tili*. – Toshkent: Fan va texnologiya, 2009. – 414 b.
23. Abjalova M. *Esperanto tilini bilasizmi?* – Toshkent: Mumtoz so‘z, 2010. – 64 b. – lug‘at: – 8 b.
24. *Автоматическая обработка текстов на естественном языке и компьютерная лингвистика. Учебное пособие* / Большакова Е.И., Клышинский Э.С., Ландэ Д.В., Носков А.А., Пескова О.В., Ягунова Е.В. – Москва: МИЭМ, 2011. – 272 с.
25. Апресян, Ю.Д. *Исследования по семантике и лексикографии*. Т. I: *Парадигматика*. – Москва: Языки славянских культур, 2009.
26. Арзикулов Х.А., Пиотровская К.Р. *Информатика и переработка текста средствами вычислительной техники (учебное пособие)*. – Самарканд, 1986.
27. Баскакова И.Л., Глухов В.П. *Практикум по психолингвистике / Учебное пособие*. – Москва: Астрель, 2008. – С. 315.
28. Белоногов Г.Г., Котов Р.Г. *Автоматизированные информационно-поисковые системы*. – Москва, 1968.

29. Беляева Л.И., Чижаковский В.А. Тезаурус в системах автоматической переработки текста. – Кишинев, 1983.
30. Большакова Е.И., Клышинский Э.С. и др. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие. – Москва: МИЭМ. 2011. – 272 с.
31. Вежбицкая А. Сопоставление культур через посредство лексики и прагматики. – Москва: Языки славянской культуры, 2001. – 272 с.
32. Добров В.А. Прикладная и компьютерная лингвистика // Николаев И. С., Митренина О. В., Ландо Т. М. (Ред.). – Москва: URSS, 2016. – 320 с.
33. G‘ulomov A., Asqarova M. Hozirgi o‘zbek adabiy tili (Sintaksis). – Toshkent: O‘qituvchi, 1987.
34. Иомдин Л.Л. Автоматическая обработка текста на естественном языке: модель согласования. – Москва: Наука, 1990. – 180 с.
35. Иорданская Л. Н. Автоматический синтаксический анализ. Т. 2. Межсегментный синтаксический анализ / Л. Н. Иорданская. – Новосибирск, 1967. – 231с.
36. Yo‘ldoshev M. Vadiiy matn va uning lingvopoetik tahlili asoslari. – Toshkent: Fan, 2007.
37. Клышинский Э. С. Начальные этапы анализа текста / Э. С. Клышинский // Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учебное пособие / Е. И. Большакова [и др.]. – Москва: МИЭМ, 2011. – 272 с. – С. 106-140.
38. Компьютерный синтаксический анализ: описание моделей и направлений разработок / Г. Д. Карпова [и др.] // Итоги науки и техники. Серия «Вычислительные науки». Т. 6. – Москва, 1991. – 243 с.
39. Котов Р.Г. Якунин Б.В. Язык информационных систем. – Москва: 1989.
40. Котов Р.Г. Лингвистические аспекты автоматизированных систем управления. – Москва: Наука, 1977.
41. Котов Р.Г. Оптимизация речевого воздействия. – Москва: Наука, 1990.
42. Котов Р.Г. Прикладная лингвистика и информационная технология. – Москва: Наука, 1987.

43. Кубрякова Е.С. Память и ее роль в исследовании речевой деятельности // Текст в коммуникации. – Москва: Языкознания, 1991. – С.18.

44. Кубрякова Е.С. Язык и знание. На пути получения знаний о языке: части речи с когнитивной точки зрения. Роль языка в познании мира. Рос. академия наук. Ин-т языкознания. – Москва: Языки славянской культуры, 2004. – 560 с. – (Язык. Семиотика. Культура).

45. Qo'ng'irov R., Begmatov E., Tojiev YO. Nutq madaniyati va uslubiyat asoslari. – Toshkent: O'qituvchi, 1992.

46. Lapasov J. Badiiy matn va lisoniy tahlil. – Toshkent: O'qituvchi, 1995.

47. Леонтьева А., Кагиров И. Автоматический синтаксический анализ русских текстов // Труды 10-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2008, Дубна, Россия, 2008. – С. 399.

48. Леонтьева Н.Н. Автоматическое понимание текста: системы, модели, ресурсы: учеб. пособие. – Москва: Академия, 2006. – 304 с.

49. Леонтьева Н.Н. К теории автоматического понимания естественных текстов. – Москва: МГУ, 2000.

50. Mamajonov A. Tekst lingvistikasi (pedagogika institutlarning o'zbek filologiyasi fakultetlari studentlari uchun maxsus kurs). – Toshkent: TDPI, 1989.

51. Марчук, Ю. С. Компьютерная лингвистика / Ю. С. Марчук. – Москва, 2007. – С. 60-70.

52. Махмудов М.А., Пиотровская А.А., Садыков Т. Система системного анализа и синтеза тюркской словоформы / Переработка текста методами инженерной лингвистики. – Минск, 1982.

53. Махмудиян М. Лингвистика. – Москва, 1985. – С. 173.

54. Мельчук, И. А. Автоматический синтаксический анализ. Том 1. Общие принципы. Внутрисегментный синтаксический анализ. / И. А. Мельчук. – Новосибирск, 1964. – 360 с.

55. Mukarramov M. Hozirgi o'zbek adabiy tilining ilmiy stili. – Toshkent: Fan, 1984. – B.13.

56. Muhamedova S. Kompyuter lingvistikasi (metodik qo'llanma). – Toshkent, 2007. – 98 b.

57. Muhamedova S. O'zbek tilidagi harakat fe'llari asosida kompyuter dasturlari uchun lingvistik ta'min yaratish. – Toshkent, 2006. – 80 b.

58. Недошивина Е.В. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. Учебно-методическое пособие. – СПб, 2006. – С.26.
59. Новое в зарубежной лингвистике. Вып. 24. Компьютерная лингвистика. – Москва: Прогресс, 1989. – С.10.
60. Norov A. Kompyuter lingvistikasi asoslari. – Qarshi, 2017. – 136 b.
61. Пиотровский Р.Г. и др. Методы автоматического анализа и синтеза текста. – Минск: Высшая школа, 1985.
62. Попов Э.В. Общение с ЭВМ на естественном языке. – Москва: Эдиториал УРСС, 2004. – С. 260.
63. Po‘latov A. Kompyuter lingvistikasi. – Toshkent: Akademnashr, 2011. – 520 b.
64. Po‘latov A., Muhamedova S. Kompyuter lingvistikasi (o‘quv qo‘llanma). – Toshkent, 2008. – 98 b.
65. Rahmatullayev Sh. O‘zbek tilining yangi alifbosi va imlosi. – Toshkent: Universitet, 2002. – B.46.
66. Rizayev S. Kibernetika va tilshunoslik. – Toshkent: O‘zbekiston, 1976.
67. Rizayev S. O‘zbek tilshunosligida lingvostatistika asoslari. – Toshkent: Fan, 2006.
68. Rizayev S. O‘zbek tilshunosligida lingvostatistika muammolari (monografiya). – Toshkent: Fan. 2005. – 296 b.
69. Safarov Sh. Kognitiv tilshunoslik. – Jizzax: Sangzor, 2006. – 92 b.
70. O‘rinboyev B., Qo‘ng‘urov R., Lapasov J. Badiiy tekstning lingvistik tahlili. – Toshkent: O‘qituvchi, 1990.
71. Xudoyqulov M. Badiiy publitsistika janrlari. – Toshkent: Universitet, 2010. – B. 26.
72. Shomaqsudov A., Rasulov I., Qo‘ng‘urov R., Rustamov H. O‘zbek tili stilistikasi. –Toshkent: O‘qituvchi, 1983, – 246 b.
73. Ягунова Е.В. Основы теоретической, вычислительной и экспериментальной лингвистики / Е.В.Ягунова // Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учебное пособие / Е. И. Большакова [и др.]. – Москва: МИЭМ, 2011. – 272 с. – С. 7-89

IV. Matbuot materiallari va ilmiy maqolalar

74. Advertisement. Micro Computer Industries, Ltd // Compute! Magazine, Issue 8, Vol. 3, № 1. 1981. – P. 119.

75. Advertisement. The Spelling Bee Is Over // PC Magazine. 1982. – P. 165.

76. Bocharov V.V., Alexeeva S.V., Granovsky D.V., Protopopova E.V., Stepanova M.E., Surikov A.V. Crowdsourcing morphological annotation // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 29 мая – 2 июня 2013 г.). Вып. 12 (19). – Москва: РГГУ, 2013.

77. Brill E. A simple rule-based part of speech tagger // Proceedings of ANLC. 92, – S. 154.

78. de Amorim, R.C.; Zampieri, M. Effective Spell Checking Methods Using Clustering Algorithms. Proceedings of Recent Advances in Natural Language Processing. Hissar, Bulgaria. 2013. – P. 172-178.

79. Hearst M.A. Noun homograph disambiguation using local context in large text corpora // Proceedings of the 7th conference, UW Centre for the New OED & Text Research Using Corpora, 1991.

80. Kagirow I.A., Leontyeva An.B. Grammar-Based Speech- and Word-splitting // Proceedings of 3rd Language & Technology Conference. October 5–7, Poznas, Poland. Poznas: Fundacja Uniwersytetu im. A. Mickiewicza, 2007. 578 s. – P. 413–417.

81. Leonard Bloomfield (1927). On some rules of Pāṇini. Journal of the American Oriental Society. 47. American Oriental Society. – P. 61-70.

82. Schank R.C., Hunter L. The quest to understand thinking // Byte. 1985, apr. – P. 155

83. Schleicher, August. Zur Morphologie der Sprache. Mémoires de l'Académie Impériale des Sciences de St.-Pétersbourg. VII^o. I, N.7. St. Petersburg. 1859. – P. 35.

84. Word Net: an Electronic Lexical Database / Edit. by Christiane Fellbaum. Cambridge, MIT Press, 1998.

85. Yarowsky D. Word sense disambiguation using statistical models of Roget's categories trained on large corpora // Proceedings of the 14th International Conference on Computational Linguistics, COLING'92, 23-28 August, Nantes, France, 1992. – P. 454-460.

86. Zampieri M., de Amorim, R.C. Between Sound and Spelling: Combining Phonetics and Clustering Algorithms to Improve Target Word Recovery. Proceedings of the 9th International Conference on Natural

Language Processing (PolTAL). Lecture Notes in Computer Science (LNCS). Springer. 2014. – P. 438-449.

87. Abjalova M., Toshimov R. Matnlarni lingvistik tahrirlashning psixologik asoslari // O‘zbekistonning mustaqil taraqqiyot va ijtimoiy-falsafiy tafakkur yangilanishi. Yosh olimlar ilmiy konferensiyasi materiallari, II qism. – Toshkent, 2011. – B. 155-158.

88. Белоногов Г.Г., Дуганова И.С., Кузнецов А.Б. Экспериментальная система автоматизированного обнаружения и исправления ошибок в тексте / НТИ. Серия 2, 1984. – №3. – С. 20-25.

89. Берзон В.Е. и др. О разработки подсистемы редактирования выходного текста в рамках системы МП // Проблемы внутренней динамики речевых норм. – Минск, 1992. – С.169-174.

90. Богуславский И.М., Иомдин Л.Л. и др. Разработка синтаксически размеченного корпуса русского языка. // Доклады научной конференции «Корпусная лингвистика и лингвистические базы данных». – СПб: изд-во Санкт-Петербургского университета, 2002. – С. 40-50.

91. Dadaboyev N. Kompyuter texnologiyalarining amaliy tilshunoslikdagi o‘rni. O‘zbek tili ta’limida pedagogik texnologiyalarning o‘rni. – Toshkent, 2011. – B. 247-249.

92. Дадабоев Х. Компьютерная лингвистика Узбекистана: история, современное состояние, перспектива. Материалы международной конференции. 1 Том. – Шымкент, 2014. – С. 2016-219.

93. Dadaboyev N. O‘zbek tilida affiksatsiya usuli bilan yasalgan sifat-terminlar. “O‘zbek tilshunosligi: taraqqiyoti va istiqbollari” mavzusidagi Respublika ilmiy-nazariy anjuman materiallari. – Toshkent, 2016. – B. 193-199.

94. Dadaboyev N. O‘zbek tilida morfologik usul bilan termin yasash. Lingvist. Ilmiy maqolalar to‘plami. VII. – Toshkent, 2016. – B. 4-10.

95. Демьянков В.З. Когнитивная лингвистика как разновидность интерпретирующего подхода // Вопросы языкознания. – Москва, 1994. – № 4. – С. 22.

96. Клышинский Э.С., Рысаков С.В. Статистические методы снятия омонимии // «Новые информационные технологии в автоматизированных системах». 2015. – С. 556-557.

97. Кобрицов Б.П. Методы снятия семантической неоднозначности. НТИ, Сер.2, Вып. 3, 2004.

98. Кубрякова Е.С. Начальные этапы становления когнитивизма: лингвистика, психология, когнитивная наука / Е. С. Кубрякова // Вопросы языкознания. 1994. – №4.
99. Кубрякова Е.С. Семантика в когнитивной лингвистике. // Известия РАН. – Сер. лит. и яз. 1999. – Т. 58. № 5-6. – С. 3-12.
100. Лавошникова Э. К. О компьютерной проверке синтаксических конструкций в текстах на русском языке // Информационные процессы. Том5. 2005. – №3. – С. 201-212.
101. Лавошникова Э.К. О «подводных камнях» в компьютерных системах проверки правописания // Вестник Московского университета. Серия 9. Филология. 2002. – №6. – С.151-163.
102. Лакофф Дж. Лингвистические гештальты. В кн.: Новое в зарубежной лингвистике. – Москва: Прогресс, 1981. Вып. X. – С.350-368.
103. Лукашевич Н.В., Салий А.Д. Тезаурус для автоматического индексирования и рубрицирования: разработка, структура, ведение // НТИ, Сер. 2. 1996. – №1. – С.1-6.
104. Мельчук И.А. Порядок слов при автоматическом синтезе русского слова (предварительные сообщения) / Научно-техническая информация. 1985. – № 12. – С.12-36.
105. Орехов Б.В., Слободян Е.А. Проблемы автоматической морфологии агглютинативных языков и парсер башкирского языка [Текст] // Информационные технологии и письменное наследие: материалы международной научной конференции (Уфа, 28-31 октября 2010 г.) / отв. ред. В. А. Баранов. – Уфа. Ижевск: Вагант, 2010. – С. 167-171.
106. Пулатов А.К., Алиходжаев Б., Джураева Н. Разработка программы компьютерного анализа и синтеза глаголов узбекского языка // ЎзМУ хабарлари. – Тошкент, 2002. – №2. – С. 17-19.
107. Пулатов А.К., Джураева Н. Разработка формальной модели грамматики узбекского языка // Узбекский математический журнал. 2002, №1. – С. 47-54.
108. Сулейманов Д. Ш. Синтаксический анализатор предложений татарского языка / Д. Ш. Сулейманов, А. Р. Гатиатуллин // Труды Математического центра имени Н. И. Лобачевского. – Казань, 1999. – Т. 4. – С. 111-126.
109. Хакимов М.Х. Расширяемый входной язык математического моделирования естественного языка для много-язычной ситуации машинного перевода // ЎзМУ хабарлари, 2009. – № 1. – С.75-80.

V. Dissertatsiya va avtoreferatlar

110. 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.

111. Abdurahmonova N. Inglizcha matnlarni o‘zbek tiliga tarjima qilish dasturining lingvistik ta’minoti: Filol.fan.bo‘yicha falsafa doktori (PhD)...dis. aftoref. – Toshkent, 2018. – 49 b.

112. Бритвин В.Г. Прикладное моделирование синтагматической семантики научно-технического текста (на примере автоматического индексирования): Дисс. канд. филол.наук. – Москва: МГУ, 1983.

113. Джураева Н. Разработка формальной модели грамматики официально-делового стиля узбекского языка и её программная реализация: Автореф. дис. ...канд.филол.наук. – Тошкент. 2012. – 25 с.

114. Исхакова Х.Ф. Исследования в области формальной морфологии тюркских языков (на материале татарского литературного языка в сопоставлении с турецким и узбекским): Дис. канд. филол.наук. – М., 1972.

115. Кобрицов Б.П. Модели многозначности русской предметной лексики: глобальные и локальные правила разрешения омонимии: Автореф... канд. филол. наук. – Москва: РГГУ, 2004.

116. Kodirbekova D. Inglizcha-o‘zbek axborot-kommunikatsiya texnologiyalari terminologiyasi va uning leksikografik xususiyatlari: Filol.fan.bo‘yicha falsafa dokt.diss. avtoref. – Toshkent, 2017. – 42 b.

117. Махмудов М. Разработка системы формального морфологического анализа тюркской словоформы (на материале азарбейджанского языка): Дисс... канд. филол. наук. – Баку, 1982.

118. Мухамедов С.А. Статистический анализ лексико-морфологической структуры узбекских газетных текстов: Автореф. дис. ...канд.филол.наук. – Тошкент, 1980. – 25 с.

119. Ножов И.М. Морфологическая и синтаксическая обработка текста (модели и программы) сегментации русского предложения: Дис. канд. филол.наук. – Москва, 2003.

120. Rizayev S. O‘zbek tilining lingvostatistik tadqiqi: Filol. fan. dok. ...diss. avtoref. – Toshkent, 2008. – 50 b.

121. Сокирко А.В. Семантические словари в автоматической обработке текста: По материалам системы ДИАЛИНГ: Дисс. ... канд. тех. наук. – Москва, 2001. – 120 с.

122. O‘rinboyeva D. Xalq og‘zaki ijodi: janriy-lisoniy va lingvostatistik tadqiq muammolari: Filol. fan. bo‘yicha dokt. (DSc) ... diss. avtoref. – Samarqand, 2019. – 74 b.

123. Xamroyeva Sh. O‘zbek tili mualliflik korpusini tuzishning lingvistik asoslari: Filol.fan.bo‘yicha falsafa doktori (PhD)...dis. aftoref. – Qarshi, 2018. – 53 b.

124. Hamrayeva Yo. O‘zbek tilining ideografik lug‘atini tuzish tamoyillari (harakat va holat tushunchasini ifodalovchi so‘zlar misolida): Filol.fan.bo‘yicha falsafa dokt.diss. avtoref. – Toshkent, 2009. – 25 b.

VI. Lug‘atlar

125. Quathamer, doktor Tobias « «Немецкий медицинский словарь слов». / Doktor Tobias Kuatxamer. 2016.

126. Вежбицкая А. Язык. Культура. Познание. Пер. с англ., отв. ред. М.А. Кронгауз, вступ. ст. Е. В. Падучевой. – Москва: Русские словари, 1996. – 416 с.

127. G‘ulomov A. va boshq. O‘zbek tilining morfem lug‘ati. – Toshkent. 1977. – 463 b.

128. Зализняк А.А. Грамматический словарь русского языка: Словоизменение: около 100000 слов / А.А.Зализняк. – Москва: Русский язык, 1977. – 879 с.

129. Ma‘rufov A. Paronimlar lug‘ati. – Toshkent: O‘qituvchi, 1974. – 56 b.

130. Mengliyev B. va boshq. O‘zbek tili iboralarining o‘quv izohli lug‘ati. – Toshkent: Yangi asr avlodi, 2007.

131. Mengliyev B., Bahriddinova B. O‘zbek tilining so‘z tarkibi o‘quv lug‘ati. – Toshkent: Yangi asr avlodi, 2007.

132. Rahmatullayev Sh. O‘zbek tili omonimlarining izohli lug‘ati. – Toshkent: O‘qituvchi, 1984. – 215 b.

133. Rahmatullayev Sh. O‘zbek tilining izohli frazeologik lug‘ati. – Toshkent: O‘qituvchi, 2001. – 407 b.

134. Розенталь Д. Э. Справочник по русскому языку. Пунктуация. – Москва: ОНИКС XXI век, 2003.

135. Русский орфографический словарь: около 200 000 слов / РАН. Ин-т рус. яз. им. В. В. Виноградова / Под ред. В. В. Лопатина, О. Е. Ивановой. – Изд. 4-е, испр. и доп. – Москва: АСТ-Пресс Книга, 2012. – С. 709. – (Фундаментал. словари рус. яз.). – ISBN 978-5-462-01272-3.

136. Samad A. “X” va “H”li soʻzlarning imloviy lugʻati. – Toshkent, 2007. – 346 b.
137. Oʻzbek tilining izohli lugʻati: 80 000 dan ortiq soʻz va soʻz birikmasi (A. Madvaliyev tahriri ostida). 5 jildli. – Toshkent: Oʻzbekiston milliy ensiklopediyasi, 2006.
138. Usmon O., Doniyorov R. Ruscha-internatsional soʻzlar izohli lugʻati. – Toshkent, 1965.
139. Hamrayeva Y. Oʻzbek tilining oʻzlashma soʻzlar oʻquv izohli lugʻati. – Toshkent: Yangi asr avlodi, 2007.
140. Hojiyev A. Oʻzbek tili sinonimlarining izohli lugʻati. – Toshkent: Oʻqituvchi, 1974. – 308 b.

VII. Internet materiallari, elektron resurslar

141. <http://prezident.uz/uz/lists/view/137>. Oʻzbekiston Respublikasi Prezidenti Shavkat Mirziyoyevning Oliy Majlisga Murojaatnomasi.
142. <http://lcpb.bashedu.ru/index.php?go>. Орехов Б., Галлямов А. Башкирский морфологический анализатор (Электронный ресурс).
143. <http://my.ict.nsc.ru/~salerat/kaz/>. Программа генерации словоформ казахского языка.
144. <http://tamgasoft.kg/morfo/ru/>. Морфология кыргызского языка.
145. www.rusnauka.com/ Ныгметова Б.Д. Когнитивная парадигма в современной лингвистике.
146. <https://cyberleninka.ru/article/v/matematicheskie-modeli-sintaksisa> / Ясулова Х.С. Математические модели синтаксиса (2011).
147. <https://ru.wikipedia.org/wiki/Артефакт>
148. <https://ru.wikipedia.org/wiki/Модуль>
149. <http://web.stanford.edu/~learnest/les/> Earnest, Les. The First Three Spelling Checkers. Stanford University.
150. https://en.wikipedia.org/wiki/Spell_checker
151. https://en.wikipedia.org/wiki/Spell_checker. Peterson, James. [Computer Programs for Detecting and Correcting Spelling Errors](#). 1980.
152. <http://web.stanford.edu/~learnest/les/> Earnest, Les. [Visible Legacies for Y3K](#).
153. <https://web.archive.org/web/20090205140452/http://cled.georgeto wn.edu/faculty/>
154. <http://www.aot.ru/docs/sokirko/Dialog2004.htm>. Сокирко, А. В. Морфологические модули на сайте www.aot.ru (Электронный ресурс)

155. <http://sz.ru/parser/> Russian Link Grammar Parser [Электронный ресурс] / С. Протасов.
156. <http://www.aot.ru/download/LexGroup.rar> Русский морфологический словарь Дмитрия Григорьева [Электронный ресурс].
157. <http://company.yandex.ru/teehnology/products/mystem/mystem.xm!> Парсер Mystem [Электронный ресурс].
158. <https://ru.wikipedia.org/wiki/Lemmatization>
159. <https://ru.wikipedia.org/wiki/Лемматизация>
160. <https://research-journal.org/languages/k-voprosu-o-tokenizacii-teksta/>. Гречачин В. А.
161. <http://www.dialog-21.ru/Archive/2004/Koval.htm/>
162. <http://www.comprice.ru/articles/detail.php?ID=43153> Куц Виктор. Система проверки правописания ОРФО 2004.
163. <https://www.multitran.ru/GIF/ViewWordForms.gif>
164. http://gramota.ru/slovari/types/17_5
165. www.dialog-21.ru/media/2138/zakharov.pdf Захаров В.П. Корпуса русского языка.
166. <https://www.myfilology.ru/177/istoriya-korpusnoj-lingvistiki>. История корпусной лингвистики.
167. <http://opencorpora.org>. Ochiq rus milliy korpusi sayti.
168. [https://en.wikipedia.org/wiki/Morphology_\(linguistics\)](https://en.wikipedia.org/wiki/Morphology_(linguistics))
169. <https://research-journal.org/languages/k-voprosu-o-tokenizacii-teksta/> Гречачин В. А. К вопросу о токенизации текста.
170. http://www.dialog-21.ru/media/2444/zelenkov_segalovich.pdf. Зеленков Ю.Г., Сегалович И.В., Титов В.А. Вероятностная модель снятия морфологической омонимии на основе нормализующих подстановок и позиций соседних слов (Электрон ресурс).
171. http://elar.urfu.ru/bitstream/10995/1388/1/ИМАТ_2005_03.pdf. Кобрицов Б.П. Методы снятия семантической неоднозначности. НТИ, Сер.2, Вып. 3, 2004.
172. <http://tech.yandex.ru/mystem> – MyStem морфологик тахлил дастури сайти.
173. <http://acl.ldc.upenn.edu/J/J95/J95-4004.pdf>. Brill E. Transformation-Based Error-Driven Learning and Natural Language Processing: A Case Study in Part-of-Speech Tagging // Computational Linguistics. Vol. 21. 1995. – P. 543-565.
174. https://en.wikipedia.org/wiki/Hidden_Markov_model. Baum, L. E.; Sell, G. R. [Growth transformations for functions on manifolds](#). Pacific Journal of Mathematics. 27 (2) 1968. – P. 211–227.;

175. http://www.academia.edu/15517740/Анализ_статистических_алгоритмов_снятия_морфологической_омонимии_в_русском_языке.

176. 185. <http://samag.ru/archive/article/3059>. Рысаков С.В. Методы борьбы с омонимией.

177. https://www.hse.ru/data/2017/07/24/.../кл_задачи_подходы_ресурсы

178. Muller, Thomas; Cotterell, Ryan; Fraser, Alexander; Schütze, Hinrich. [Joint Lemmatization and Morphological Tagging with LEMMING](#) (Электрон ресурс).

179. Большакова Е.И.. Автоматическая обработка текстов: задачи, подходы, ресурсы (Электрон ресурс). Rekers, Jan, and Andy Schürr. [Defining and parsing visual languages with layered graph grammars](#). Journal of Visual Languages & Computing 8.1 (1997): – P. 27-55.

180. Кобрицов Б.П., Ляшевская О.Н., Шеманаева О.Ю. Снятие лексико-семантической омонимии в новостных и газетножурнальных текстах: поверхностные фильтры и статистическая оценка (Электрон ресурс).

181. Майорова А. Доверяй, но проверяй. Системы контроля орфографии консервируют язык / «Известия.Ру», izvestia.ru saytidan.

182. O‘zbekiston Respublikasi Prezidentining 2019-yil 29-yanvardagi «Siyosiy fanlar sohasida kadrlarni tayyorlash, fundamental va amaliy tadqiqotlar samaradorligini oshirish chora-tadbirlari to‘g‘risida»gi PQ-4139-son Qarori. <http://www.lex.uz/docs/4182353>.

183. [Медицинская проверка орфографии для Firefox и Thunderbird](#)».e-MedTools. 2017.

VIII. Internet saytlari

184. <http://www.philol.msu.ru/~humlang/articles/PolystyICorp.htm>

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

186. <http://www.yazykoznanie.ru>

187. <http://www.CICLing.ore>

188. <http://cled.georgetown.edu/faculty/>

189. <http://en.wikipedia.org>. Computational linguistics.

190. http://gramota.ru/slovari/types/17_5

191. <http://opencorpora.org>. Ochiq rus milliy korpusi sayti.

192. <http://tech.yandex.ru/mystem> – MyStem morfologik tahlil dasturi sayti

193. <http://www.aot.ru/docs/sokirko/Dialog2004.htm> – morfologik modullar

194. <http://www.bokrcorpora.narod.ru/frqlist/frqlist.html>
195. <http://www.dialog-21.ru/Archive/2004/Koval.htm/>
196. <http://www.philol.msu.ru/~lex/corpus>
197. <http://www.redaktoram.ru>
198. <http://www.ruscorpora.ru> (corpora.yandex.ru)
199. <http://www.rusf.ru/books/analysis>
200. <http://www.russian.slavica.org>
201. <http://www.wikipedia.org/wiki/model.html>
202. <http://www-formal.stanford.edu/jmc/whatisai.html>
203. <https://www.multitran.ru/GIF/ViewWordForms.gif>
204. <https://www.multitran.ru/morphology.htm>
205. www.asknet.ru
206. www.krugosvet.ru

MUNDARIJA

KIRISH.....	5
I BOB. KOGNITIV TADQIQOTLAR PARADIGMASIDA MATN TAHRIRI VA TAHLILI	
1.1. Hozirgi tilshunoslikda kognitiv paradigma.....	11
1.2. Kompyuter lingvistikasida modul.....	15
1.3. Matn tahriri va tahlilini amalga oshiruvchi dasturiy sistemalar...	18
1.4. Matnlarni avtomatik tahrir va tahlil qilish dasturining lingvistik ta'minoti manbalari.....	37
1.5. O'zbekcha matnlarni avtomatik tahrir va tahlil qilish sistemasini yaratishda uslub masalasi.....	44
1.6. Matnlarni avtomatik tahrir qilishdagi muammolar.....	52
Birinchi bob bo'yicha xulosa.....	54
II BOB. MATNNI AVTOMATIK TAHRIR VA TAHLIL QILISH SISTEMASI MODULLARI	
2.1. Matnlarni avtomatik tahlil qilish borasida amalga oshirilayotgan tadqiqotlar yo'nalishi va sistema bosqichlari.....	56
2.2. Grafematik tahrirlash moduli va uning texnologiyasi.....	58
2.3. Morfologik tahlil qilish modulini yaratish prinsiplari va uning algoritmi.....	64
2.4. Omonimlarni tahlil qilish texnologiyasi.....	69
2.5. Lingvistik protsessorning sintaktikoldi tahlil jarayoni.....	78
2.6. Dasturning sintaktik tahlil qilish moduli.....	82
2.7. Gaplarning sintaktik tuzilishini avtomatik tahlil qilish algoritmi tavsifi.....	88
Ikkinchi bob bo'yicha xulosa.....	92
III BOB. SO'Z TURKUMLARI MODULLARI	
3.1. Mustaqil so'z turkumlari modullari.....	95
3.1.1. Ot turkumini formallashtirish moduli.....	98
3.1.2. Fe'lni formallashtirish moduli.....	104
3.1.3. Sifatni formallashtirish moduli.....	113
3.1.4. Sonni formallashtirish moduli.....	114
3.1.5. Ravishni formallashtirish moduli.....	116
3.1.6. Olmoshni formallashtirish moduli.....	117
3.2. Yordamchi va alohida olingan so'z turkumlari moduli.....	119
3.3. Lingvistik protsessorlarda affiksatsiya moduli.....	121

Uchinchi bob bo'yicha xulosa.....	123
XULOSA.....	125
ILOVA.....	128
GLOSSARIY.....	154
FOYDALANILGAN ADABIYOTLAR.....	158

CONTENTS

INTRODUCTION.....	5
 I CHAPTER. TEXT EDITING AND ANALYSIS IN THE PARADISM OF COGNITIVE RESEARCH	
1.1. Cognitive paradigm in modern linguistics.....	11
1.2. Module in computer linguistics.....	15
1.3. Software systems for text editing and analysis.....	18
1.4. Sources of linguistic support for automatic text editing and analysis software.....	37
1.5. Methodological issues in the creation of a system of automatic editing and analysis of Uzbek texts.....	44
1.6. Problems with automatic text editing.....	52
Summary of the first chapter.....	54
 CHAPTER II. AUTOMATIC TEXT EDITING AND ANALYSIS SYSTEM MODULES	
2.1. The direction and system stages of research on automatic analysis of texts.....	56
2.2. Graphic editing module and its technology.....	58
2.3. Principles of creating a module of morphological analysis and its algorithm	64
2.4. Methods for determining homonyms.....	69
2.5. The process of syntactic analysis of a linguistic processor.....	78
2.6. Syntactic analysis module of the program.....	82
2.7. Description of the algorithm for automatic analysis of the syntactic structure of sentences.....	88
Summary of the second chapter.....	92
 CHAPTER III. WORD GROUPS MODULES	
3.1. Modules of independent word groups.....	95
3.1.1. Horse group formation module.....	98
3.1.2. Verb formation module.....	104
3.1.3. Quality formulation module.....	113
3.1.4. Number formulation module.....	114
3.1.5. Form formulation module.....	116
3.1.6. Pronoun formulation module.....	117
3.2. Auxiliary and separate phrases module.....	119
3.3. Affixation module in a linguistic processor.....	121

Summary of chapter three.....	123
CONCLUSION.....	125
APPENDIX.....	128
GLOSSARY.....	154
REFERENCES.....	158

Abjalova Manzura Abdurashetovna

**TAHRIR VA TAHLIL DASTURLARINING
LINGVISTIK MODULLARI**

monografiya

“Nodirabegim” nashriyoti

Nashriyot litsenziyasi AI № 313. 24.11.2017 y.

Bosishga ruxsat etildi: 28.12.2020.

“Times New Roman” garniturası. Qog‘oz bichimi: 60x84 ¹/₁₆

Nashriyot bosma tabog‘i 11. Adadi 200 nusxa.

100129, Toshkent shahri, Shayxontohur tumani,

Navoiy kochasi, 30-uy.

OOO “AKTIV PRINT” bosmaxonasida chop etildi.

Toshkent, Chilonzor 25, Lutfiy 1A.



Abjalova Manzura Abdurashetovna – Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti “Axborot texnologiyalari” kafedrasida katta o‘qituvchisi, filologiya fanlari bo‘yicha falsafa doktori (PhD)

Ilmiy-ommabop nashri: “Esperanto tilini bilasizmi?”

Doktorlik dissertatsiyasi (PhD): “O‘zbek tilidagi matnlarni tahrir va tahlil qiluvchi dasturning lingvistik modullari (2019)

Yutuqlari:

- “Axborot-kutubxona muassasasining eng yaxshi mutaxassisi – 2018” Respublika tanlovi 1-o‘rin g‘olibi

- “O‘zbekiston ayollarining 100 ta eng yaxshi innovatsion loyihasi” III Respublika tanlovining “Eng yaxshi EHM uchun dastur va ma‘lumotlar bazasi” yo‘nalishi g‘olibi (2020)

- O‘zbek tilini o‘qitishning eng yaxshi metodikasi” ilmiy-ijodiy ishlar tanlovi sovrindori (2020)

- “Innovatsion g‘oyalar yetakchisi” ko‘krak nishoni sohibi (2020)

Qiziqishlari sohasi: korpus lingvistikasi, kompyuter leksikografiyasi, tabiiy tilni qayta ishlash (NLP): mashina tarjimasi, kompyuterli morfologiya, sintaktik tahlil, nutq sintezatori, nutqni tanish.

manzura_ok@mail.ru

abjalova.manzura@gmail.com

ISBN 978-9943-6939-0-6



9 789943 693906