



MATNLARNI QAYTA ISHLASH MANBALARI

Abjalova Manzura Abdureshetovna

ToshDO‘TAU Kompyuter lingvistikasi
va raqamli texnologiyalar kafedrası dots.v.b.,
filologiya fanlari bo‘yicha PhD
abjalova@navoiy-uni.uz

Rashidova Umida Mansurovna

SamDU dotsenti, filologiya fanlari bo‘yicha PhD
umida_rashidova@bk.ru

Abdumuminov Baxtiyor Soatmurotovich

ToshDO‘TAU Kompyuter lingvistikasi mutaxassisligi 2-kurs magistranti

Annotatsiya. Mazkur maqolada kompyuter texnologiyalari asosida matnlarni qayta ishlashga mo‘ljallangan dasturlarning lingvistik ta‘minotida bo‘lishi zarur sanalgan lisoniy manbalar haqida so‘z yuritildi.

Kalit so‘zlar: *lingvistik ta‘minot, filologik dasturlar, matnlarni qayta ishlash, manba.*

SOURCES OF TEXT PROCESSING

Abstract. This article discusses linguistic sources that should be necessary for processing texts based on computer technology.

Key words: *linguistic support, philological programs, processing of texts, source.*

Istalgan tildagi matnlarni qayta qilish, umuman, filologik dasturlarni yaratish uchun kompyuter xotirasida muayyan lingvistik ta‘minot (LT)ning bo‘lishi talab qilinadi. Ta‘minot asosini esa lingvistik qoidalar va me‘yorlar tashkil etadi. Shuningdek, LTdan muayyan tilning filologik lug‘at boyligi ham o‘rin oladi. Shu bois lingvistik protsessorni yaratishda qayta ishlanayotgan tabiiy til haqida yetarli darajada ma‘lumotga ega bo‘lish talab qilinadi. Bu o‘z navbatida lingvistik protsessorning mukammal bo‘lishini ta‘minlaydi.



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.

Morfologik lug'at so'zshaklning, ya'ni ismlarning turlanishi va fe'llarning tuslanishini shartli belgilashning maxsus sistemasi orqali aks ettiradigan leksikografik manba hisoblanadi [1]. 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 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:



ж 8а —47 | св (нсв) 2 —92

ВАТЬ

отретуши́ровать	св 2а	стасова́ть	св 2а
варьи́ровать	нсв 2а	растасова́ть	св 2а ⊗ I
интервьюи́ровать	св-нсв 2а	фасова́ть	нсв 2а
проинтервьюи́ровать	св 2а	расфасова́ть	св 2а ⊗ I
кровать	ж 8а	всова́ть	св 2b ⊗ I (см.)
дива́н-кровать	м, склоняются обе част- ти (косв. формы и оп- ределения избегаются)	подсова́ть	св 2b ⊗ I (см.)
кресло-кровать	с, склоняются обе част- ти (косв. формы и оп- ределения избегаются)	колесова́ть	св-нсв 2а
собб́ровать	св-нсв 2а	адресова́ть	св-нсв 2а
особб́ровать	св 2а	переадресова́ть	св 2а ⊗ I
ворова́ть	нсв 2а	пересова́ть	св 2b
наворова́ть	св 2а ⊗ I	ни́тересова́ть	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 etilgan [3,12,16].

Imlo lug‘atlaridagi leksemalarning qaysi so‘z turkumiga oidligini aniqlashda “O‘zbek tilining izohli lug‘ati” [19] 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 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” [7] 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 nomlanadi [4,17,20]. 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* singarilar esa badiiy matnga tegishlidir. Bundan ko‘rinib turibdiki, frazeologizmlar lug‘ati [11,15] 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 majoziylik 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 ($N_0+N_{e.a.} \Rightarrow$ otli birikma)+V: *Mening sabr kosam to'ldi; Sabr kosang to'lgandir; Uning sabr kosasi ham to'ldi; Sabr kosamiz to'lib bormoqda*. Shu bois LTda 1393 frazema bazasi yaratildi, frazema so'zga teng birlik [frazema=so'z] deb olindi va morfologik tahlil moduli uchun ularda grammatik shakllarning o'zgarish o'rni belgilandi.

Matni yozish jarayonida nutqda talaffuzi bir-biriga yaqin leksemalarni yanglish almashtirib yoxud aytilishi bo'yicha yozish hollari ham uchramoqda. Masalan, *tanbur* o'rniga *tambur*, *adil* o'rniga *adl*, *teri* o'rniga *tire*, *yonilg'i* o'rniga *yoqilg'i*, *sudxo'r* o'rniga *sutxo'r* qo'llash kabi. Ko'pchilik AYoQSh abbreviatsiyasining kengaytmasi *avtomobil yonilg'i quyish shoxobchasi* ekanligiga ahamiyat bermaydi. Shuning uchun *avtomobil yoqilg'i quyish* (yoxud *qo'yish*) *shaxobchasi* deb o'ylashadi. Vaholanki, *yonilg'i* – gaz, kerosin, benzin kabi yonuvchi suyuqliklar majmui; *yoqilg'i* – o'tin, ko'mir kabilar majmui [2, 9]. Shuni nazarda tutib, bunday leksemalarning ma'nolarini farqlash maqsadida A.Ma'rufovning “Paronimlar lug'ati”ga murojaat qilamiz. Matni kirituvchi shaxs leksemaning to'g'ri kiritilganiga shubha qilayotgan paytida dastur protsessoriga joylashtirilgan paronimlar lug'atiga murojaat qiladi va leksemaning anglatgan ma'nosiga qarab uni o'z o'rnida qo'llaydi.

ATT qilish dasturini yaratishdagi lingvistik muammolardan biri omonimlik hodisasini avtomatik aniqlash masalasidir. Ushbu masala birmuncha dolzarb bo'lib, uning echimini topishda, albatta, maxsus modullar muhim ahamiyatga molikdir [1]. Lingvistik modullar dastur bazasidagi mavjud omonim so'zlar uchun mo'ljallangan bo'ladi. Shu bois bazaga Sh.Rahmatullayevning “O'zbek tili omonimlarining izohli lug'ati” [14] kitobida keltirilgan, shuningdek, keyinchalik nutqda paydo bo'lgan jami 1638 ta omonim birliklar jamlandi.



Tezaurus va ontologiyalar lingvistik manbalarning o'ta muhim turlaridan hisoblanadi. Tezaurus (yunoncha *thesaurós* – *xazina, boylik*) lug'atning o'ziga xos maxsus ko'rinishi bo'lib, unda lug'aviy birliklar o'rtasidagi semantik munosabat (sinonimlar, antonimlar, paronimlar, 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 (semantik munosabati)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 beradi [22].

Tezaurus tushunchasi bilan ontologiya tushunchasi chambarchas bog'liq [6]. Ontologiya (yunoncha *ontós* – *borliq*, *logós* – *ta'limot*) – borliq haqidagi falsafiy ta'limot bo'lib, tilning umumiy leksik ma'lumotlar bazasi hisoblanadi. Ontologik lug'atlarning lingvistik hisoblanishining sababi ular ma'lum bir tilning so'z boyligi asosida tuziladi.

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.

WordNet [18] 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 iyerarxik munosabatdagi so'zlar soni esa 15000 tani (7-8) tashkil etadi. Iyerarxiyaning 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 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 farqlanadi [8]. 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 keltiradi [13]. E.I.Bolshakova o'z izlanishlarida lingvistik protsessor lingvistik ma'lumotlar modeliga asoslanishini, ya'ni dasturiy ta'minotning lingvistik resurslarini quyidagilar tashkil etishini ta'kidlaydi: kompyuter lug'atlari, tabiiy til grammatikasi, so'z birikmalari bazasi, tezaurus va ontologiyalar, matnlar majmuasi va korpus [21].

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 hisoblanadi [5,10].



1-rasm. Avtomatik tahrir va tahlil manbalari.

Matnlarni avtomatik tahrir va tahlil 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.

Adabiyotlar:

1. Abjalova M. Tahrir va tahlil dasturlarining lingvistik modullari. [Matn] : monografiya / M.A. Abjalova. - Toshkent: Nodirabegim, 2020. – 176 b.
2. Abjalova M. O'zbek tilining paronimlar lug'ati [Matn]: o'quv-uslubiy lug'at. – Termiz, 2022. – 64 b.
3. G'ulomov A. va boshq. O'zbek tilining morfem lug'ati. – Toshkent. 1977 – 463 b.
4. Hamrayeva Y. O'zbek tilining o'zlashma so'zlar o'quv izohli lug'ati. – Toshkent: Yangi asr avlodi, 2007.
5. Hearst, M.A. Automated Discovery of WordNet Relations. In: Fellbaum, C. (ed.) WordNet: An Electronic Lexical Database. MIT Press, Cambridge, 1998, – P. 131-151.
6. Hirst, G. Ontology and the Lexicon. In.: Handbook on Ontologies in Niformation Systems. Berlin, Springer, 2003.
7. Hojiyev A. O'zbek tili sinonimlarining izohli lug'ati. – Toshkent: O'qituvchi, 1974. – 308 b.
8. http://gramota.ru/slovari/types/17_5
9. Ma'rufov A. Paronimlar lug'ati. – Toshkent: O'qituvchi, 1974. – B.20.
10. Matsumoto Y. Lexical Knowledge Acquisition. In: The Oxford Handbook of Computational Linguistics. Mitkov R. (ed.). Oxford University Press, 2003, – P. 395-413.
11. Mengliyev B. va boshq. O'zbek tili iboralarining o'quv izohli lug'ati. – Toshkent: Yangi asr avlodi, 2007.
12. Mengliyev B., Bahriddinova B. O'zbek tilining so'z tarkibi o'quv lug'ati. – Toshkent: Yangi asr avlodi, 2007.
13. Po'latov A. Kompyuter lingvistikasi. – Toshkent: Akadernashr, 2011. – B. 213.
14. Rahmatullayev Sh. O'zbek tili omonimlarining izohli lug'ati. – Toshkent: O'qituvchi, 1984. – 215 b.
15. Rahmatullayev Sh. O'zbek tilining izohli frazeologik lug'ati. – Toshkent: O'qituvchi, 2001. – 407 b.
16. Samad A. “X” va “H”li so'zlarning imloviy lug'ati. – Toshkent, 2007. – 346 b.
17. Usmon O., Doniyorov R. Ruscha-internatsional so'zlar izohli lug'ati. – Toshkent, 1965.



18. Word Net: an Electronic Lexical Database / Edit. by Christiane Fellbaum. Cambridge, MIT Press, 1998.
19. 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.
20. O'zbek tili o'zlashma so'zlarining urg'uli lug'ati [Matn]: o'quv-uslubiy lug'at / M.Qurbonova, M.Abjalova, N.Axmedova, R.To'laboyeva. – Toshkent: Nodirabegim, 2021. – 988 b. *ISBN 978-9943-6940-9-5*
21. Большакова Е.И. и др. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика. Учебное пособие. – Москва: МИЭМ, 2011. – С. 97-99.
22. Лукашевич Н.В., Салий А.Д. Тезаурус для автоматического индексирования и рубрицирования: разработка, структура, ведение // НТИ, Сер. 2, №1, 1996. – С.1-6.