



**O‘ZBEK TILI KORPUSIDA SO‘Z YASALISH MODELLARI: NAZARIY
TAMOYILLAR VA AVTOMATIK MODELLASHTIRISH
IMKONIYATLARI**

Babayeva Gulchexra Tashpulatovna

Termiz shahar 16-sonli informatika va axborot texnologiyalariga ixtisoslashtirilgan
umumiy o‘rta ta’lim maktabining oliy toifali o‘qituvchisi

Annotatsiya: Ushbu maqolada o‘zbek tili korpusida so‘z yasash modellari nazariy jihatdan tahlil qilinib, ularning avtomatik modellashtirish imkoniyatlari ko‘rib chiqiladi. So‘z yasashining morfologik va derivatsion tamoyillari, ularning kompyuter lingvistikasi nuqtayi nazaridan formal modellari tavsiflanadi. Korpus asosida avtomatik tahlil qilishda affiksatsiya, konversiya, qo‘shma so‘z yasash kabi jarayonlarni aniqlovchi algoritmlar ishlab chiqish zarurligi ta’kidlanadi. Tadqiqotda o‘zbek tili agglutinatив tuzilishining korpus tahliliga ta’siri, morfologik analizatorlardan foydalanish samaradorligi hamda sun’iy intellekt asosida derivatsion modellarni yaratish istiqbollari yoritilgan.

Kalit so‘zlar: o‘zbek tili korpusi, so‘z yasashi, derivatsiya, morfologik analizator, avtomatik modellashtirish, lingvistik algoritim.

Annotation: This article analyzes the theoretical principles of word formation models in the Uzbek language corpus and explores their potential for automatic modeling. The morphological and derivational mechanisms of word formation are described from the perspective of computational linguistics. The study emphasizes the necessity of developing algorithms that can automatically identify affixation, conversion, and compounding processes within the corpus. It also highlights the impact of Uzbek’s agglutinative structure on corpus analysis, the effectiveness of morphological analyzers, and the prospects for building derivational models based on artificial intelligence.

Keywords: Uzbek language corpus, word formation, derivation, morphological analyzer, automatic modeling, linguistic algorithm.

Аннотация: В данной статье рассматриваются теоретические основы моделей словообразования в национальном корпусе узбекского языка и анализируются возможности их автоматического моделирования. Описаны морфологические и



деривационные принципы словообразования с точки зрения компьютерной лингвистики. Отмечается необходимость разработки алгоритмов, позволяющих автоматически выявлять процессы аффиксации, конверсии и сложения слов. В статье также обсуждается влияние агглютинативного строя узбекского языка на корпусный анализ, эффективность морфологических анализаторов и перспективы создания деривационных моделей на основе искусственного интеллекта.

Ключевые слова: корпус узбекского языка, словообразование, деривация, морфологический анализатор, автоматическое моделирование, лингвистический алгоритм.

O‘zbek tili korpusi — bu til tizimining barcha qatlamlarini raqamli shaklda ifodalovchi va avtomatik tahlil imkonini beruvchi lingvistik platformadir. Korpus asosida so‘z yasaliş jarayonlarini o‘rganish, o‘zbek tilining derivatsion tizimini formal modellashtirish va bu modellarni avtomatik tarzda qayd etish imkonini beradi. Korpus lingvistikasi doirasida so‘z yasaliş modellari tahlili, bir tomondan, morfologik tuzilmani aniqlashga, ikkinchi tomondan esa, semantik o‘zgarishlarni kuzatishga xizmat qiladi [4, 174].

So‘z yasaliş nazariyasida derivatsion jarayonlar asosan **affiksatsiya, konversiya, kompozitsiya va reduplikatsiya** kabi mexanizmlar orqali yuzaga keladi. O‘zbek tilida bu jarayonlar, ayniqsa, affiksatsiya orqali faol kechadi. Sh. Rahmatullayevning fikricha, “affikslar semantik yuklamasi bilan so‘zning grammatik rolini belgilaydi va derivatsion munosabatni shakllantiradi” [5, 46]. Ushbu yondashuv korpus asosida morfologik modellarni shakllantirishda nazariy asos sifatida muhimdir.

Korpus tahlilida so‘z yasaliş birliklarini avtomatik qayd etish uchun **morfologik analizator** va **lemmatizator** texnologiyalari keng qo‘llaniladi. Ergasheva M. o‘z tadqiqotida o‘zbek tili morfologik analizatorini so‘z yasaliş jarayonlarini tahlil qilishda qo‘llash imkoniyatlarini ko‘rsatib, “analizator yordamida yasovchi affikslarning formal strukturasi avtomatik aniqlash mumkin” deydi [3, 41–42].

Shuningdek, o‘zbek tilida **semantik tarmoq va ontologik modellar** yordamida so‘z yasaliş ma’nolarini avtomatik tasniflash bo‘yicha izlanishlar olib borilmoqda. Toshpo‘latova G. ta’kidlashicha, “leksik-semantik maydon nazariyasi asosida yasalgan so‘zlarning o‘zaro paradigmatic bog‘liqligini aniqlash, ularning ma’nodosh va antonimik qatorlarini tarmoqlashtirish imkonini beradi” [9, 62].

So‘z yasaliş modellari avtomatik modellashtirishda **sun’iy neyron tarmoqlar, statistik modellashtirish** va **ta’limli algoritmlar** asosida qayta ishlanmoqda. Xususan,



xorijiy tadqiqotlarda BERT, Word2Vec va GPT modellarining morfologik tahlil uchun moslashtirilgan versiyalari ishlab chiqilgan [10, 6]. Ushbu yondashuv o'zbek tiliga ham tatbiq etilmoqda, ammo agglutinatив tuzilish algoritmik tanib olishda qo'shimcha murakkablik tug'diradi.

Tojiboeva M. o'z maqolasida o'zbek tili milliy korpusida derivatsion jarayonlarni modellashtirishda **so'z shakllarining variantliligi** va **morfemalarning birikish chastotasini** hisobga olish lozimligini ta'kidlaydi [8, 69–70].

Tilshunos A. Abduazizov esa avtomatik tahlilda **antroposentrik yondashuvning** ahamiyatini qayd etadi. Unga ko'ra, "so'z yasaliş jarayonlari faqat grammatik struktura emas, balki inson tafakkuri va nutq faoliyati bilan bog'liq psixolinguistik mexanizmdir" [1, 114–115].

Mualliflik mulohazasi: o'zbek tili korpusida so'z yasaliş jarayonlarini avtomatik tahlil qilish bo'yicha ilg'or xorijiy modellardan foydalanish, lekin ularni o'zbek tili morfologiyasi va semantikasiga moslashtirish zarur. Chunki o'zbek tilining agglutinatив xususiyati, shuningdek, semantik polisemiyaga boyligi modellarni o'zgartirishni talab etadi. Shu bois, milliy korpus uchun **gibrid modellashtirish** — ya'ni, morfologik, statistik va semantik tahlilni birlashtirgan yondashuv eng maqbul variant sifatida qaraladi.

Korpus tahlilining yana bir muhim jihati — **so'z yasaliş paradigmasining avtomatik aniqlanishidir**. Bu borada Biber so'z yasaliş turlari va ularning matndagi pragmatik funksiyasini ajratish uchun korpus asosidagi statistik ko'rsatkichlarni qo'llashni taklif etadi [2, 86]. Shu yondashuv asosida o'zbek tilida ham derivatsion modellarni avtomatik shakllantirishda statistik chastotani semantik qiymat bilan uyg'unlashtirish ilmiy jihatdan dolzarb hisoblanadi.

Umuman olganda, o'zbek tili korpusida so'z yasaliş modellari nafaqat morfologik birliklarni, balki ularning semantik va funksional o'zgaruvchanligini ham ifodalashi lozim. Korpus asosida avtomatik modellashtirish tizimlari tilshunoslikda yangi bosqichni boshlab beradi, chunki bu tizimlar orqali tilning dinamik rivojlanish qonuniyatlarini kompyuter modeli shaklida ifodalash imkoniyati yaratiladi.

O'zbek tili korpusida so'z yasaliş jarayonlarini avtomatik modellashtirishda leksik-semantik tarmoqlar va kognitiv modellarning qo'llanilishi ham muhim yo'nalishlardan biridir. Kognitiv lingvistika yondashuvi asosida yaratilgan korpus modellarida so'zlarning ma'no maydoni, konseptual tuzilishi va semantik ierarxiyasi formal algoritmlar yordamida ifodalanadi. Sh. Safarovning ta'kidlashicha, "so'z yasaliş jarayonida inson tafakkuri, konseptual asos va semantik motivatsiya o'zaro bog'liq jarayonlardir" [6, 77–78]. Bu qarash korpusda avtomatik derivatsion modellashtirishda



ham asosiy tamoyillardan biri sifatida qaraladi.

Hozirgi kunda o‘zbek tili korpusida yasovchi affikslar bazasi, so‘z yasash daraxtlari (derivatsion trees) va morfemik belgilash tizimlari yaratilmoqda. Bu tizimlar so‘zlarning morfemal tarkibini avtomatik tahlil qilish, yasama so‘zlarni tanib olish va yangi so‘z yasash tendensiyalarini aniqlash imkonini beradi. Tojiboeva M. o‘z tadqiqotida korpusga asoslangan so‘z yasash tahlilini “o‘zbek tilidagi innovatsion birliklar, yangi terminlar va texnolektlarning shakllanish jarayonlarini kuzatishda samarali vosita” deb baholaydi [7, 52].

So‘z yasash modellari korpusda faqat til tizimini emas, balki jamiyatdagi yangi ijtimoiy, ilmiy va texnologik o‘zgarishlarni aks ettiruvchi leksik yangilanishni ham ifodalaydi. Chunki korpus orqali “raqamlashtirish”, “mobilografiya”, “onlaynlashtirish” kabi yangi hosilalar qanday morfologik shaklda va qanday semantik o‘zgarishlar bilan yuzaga kelgani aniqlanishi mumkin. Bunday tahlil milliy til rivojining tabiiy dinamikasini ilmiy asosda o‘rganish imkonini beradi.

Avtomatik modellashtirish tizimlarida data-driven yondashuvlar (ma’lumotga asoslangan modellar) va rule-based (qoidalarga asoslangan) tizimlarning integratsiyasi yuqori aniqlikni ta’minlaydi. Bunday gibrid tizimlar o‘zbek tili uchun ham istiqbolli, chunki u grammatik qoidalarni formal tarzda ifodalab, bir vaqtning o‘zida statistik tahlil asosida real til qo‘llanilishini ham hisobga oladi [2, 89].

Mualliflik mulohazasi: o‘zbek tili korpusida so‘z yasash modellari ishlab chiqishda faqat morfologik mexanizmlarni emas, balki semantik motivatsiya, konseptual bog‘liqlik va tilning milliy xususiyatlarini hisobga olish zarur. Chunki derivatsion tizim — bu nafaqat grammatik mexanizm, balki milliy tafakkur va dunyoqarashning lingvistik ifodasidir.

Xulosa

O‘zbek tili korpusida so‘z yasash modellari lingvistik tizimning eng faol va murakkab qatlamini tashkil etadi. Ushbu jarayonni avtomatik modellashtirish nafaqat texnik yechim, balki o‘zbek tili morfologiyasining nazariy asoslarini yanada chuqurlashtirishga xizmat qiladi. Tadqiqotlar shuni ko‘rsatadiki, korpus asosida so‘z yasash modellari: derivatsion tizimning formal ifodasini yaratadi; affiks va morfemalarning semantik yuklamasini aniqlash imkonini beradi; yangi so‘z yasash tendensiyalarini avtomatik qayd etadi; tilni kompyuter modeli sifatida ifodalashga zamin yaratadi; korpus lingvistikasini milliy tilshunoslik bilan integratsiyalash imkonini beradi.

Xulosa sifatida ta’kidlash joizki, o‘zbek tili korpusida so‘z yasash modellari bo‘yicha yaratilayotgan tizimlar milliy tilni raqamli tahlil qilish, o‘qitish, tarjima va sun’iy



intellekt asosida qayta ishlash imkoniyatlarini kengaytiradi. Shu bois, bu yo‘nalishda nazariy, metodologik va texnologik integratsiyani yanada rivojlantirish, korpus asosida avtomatik derivatsion analizator yaratish dolzarb vazifadir.

Foydalanilgan adabiyotlar ro‘yxati

1. Abduazizov, A. *Tilshunoslikning yangi yo‘nalishlari*. — Toshkent: Ziyonashriyoti, 2022. — 168 b.
2. Biber, D. *Corpus Linguistics and the Study of Meaning*. — Cambridge: Cambridge University Press, 2015. — 198 p.
3. Ergasheva, M. *O‘zbek tilining morfologik analizatori va uni korpus lingvistikasida qo‘llash masalalari*. — “Tilshunoslikda kompyuter texnologiyalari” Respublika ilmiy-amaliy konferensiyasi materiallari. — Toshkent: TDPU, 2020. — B. 41–42.
4. McEnery, T., & Hardie, A. *Corpus Linguistics: Method, Theory and Practice*. — Cambridge: Cambridge University Press, 2012. — 274 p. — ISBN 978-0-521-89543-0.
5. Rahmatullayev, Sh. *O‘zbek tilining so‘z yasash tizimi*. — Toshkent: Fan nashriyoti, 1982. — 120 b.
6. Safarov, Sh. *Kognitiv lingvistika asoslari*. — Toshkent: Fan va texnologiya nashriyoti, 2015. — 188 b.
7. Tojiboeva, M. *O‘zbek tili milliy korpusi va innovatsion leksika*. — “Zamonaviy filologik tadqiqotlar” ilmiy to‘plami. — Toshkent: TDSHU, 2023. — B. 52..
8. Tojiboeva, M. *O‘zbek tili milliy korpusi va uning lingvistik imkoniyatlari*. — “Filologiya masalalari” ilmiy jurnali. — Toshkent: O‘zbekiston Milliy universiteti, 2021. — B. 69–70..
9. Toshpo‘latova, G. *Leksik-semantik maydon va uning korpus tahlilidagi roli*. — “Zamonaviy lingvistik tadqiqotlar: nazariya va amaliyot” ilmiy to‘plami. — Toshkent: TDSHU, 2021. — B. 62..
10. Vaswani, A. et al. *Attention Is All You Need*. — In: *Advances in Neural Information Processing Systems (NIPS)*. — 2017. — P. 1–11.