

2-25-2020

## THE ROLE OF THE LINGUISTIC LAYOUT IN THE CREATION OF THE UZBEK NATIONAL CORPS

Guli Toirova Ibragimovna

*Dosent of the Department of Uzbek linguistics of Bukhara State University*

Follow this and additional works at: <https://uzjournals.edu.uz/buxdu>

---

### Recommended Citation

Ibragimovna, Guli Toirova (2020) "THE ROLE OF THE LINGUISTIC LAYOUT IN THE CREATION OF THE UZBEK NATIONAL CORPS," *Scientific reports of Bukhara State University*. Vol. 4 : Iss. 1 , Article 3.  
Available at: <https://uzjournals.edu.uz/buxdu/vol4/iss1/3>

This Article is brought to you for free and open access by 2030 Uzbekistan Research Online. It has been accepted for inclusion in Scientific reports of Bukhara State University by an authorized editor of 2030 Uzbekistan Research Online. For more information, please contact [sh.erkinov@edu.uz](mailto:sh.erkinov@edu.uz).

шоир ўша вариантлардан бирини ёки умумлаштирган ҳолда шеърга солди дегани эмас. Шоир уни ижодий қайта ишлайди, воқеанинг янги маъно қирраларини очадики, бу шеърнинг яқунловчи бандида акс этади:

*Қондошликнинг миллион йиллик  
Ришталари узилди.  
Ўртага бир эшак тушиб  
Оралари бузилди. (I,327)*

Агар ривоятнинг халқ ижодидаги вариантларида: 1) “эл оғзига элак тутиб бўлармиди, нимаки қилма, ундан ўян қидирувчилар ҳамиша топилади” ва 2) “одам ўзгаларнинг нима дейишига қараб эмас, ақл-фаросати билан иш қилмоғи керак” деган маънолар келиб чиқса, А.Орипов маъно урғусини кишилар орасини бузувчиларга кўчиради, уларни эшакка менгзайди. Шоирнинг кашфи туфайли ривоят ўзининг янги умрини бошласа, ажаб эмас. Чунки энди “ўртага бир эшак тушиб...” дейилсаёқ кўпчилик ёдига шу шеър келади, демак, у энди “масал”, яъни мос ҳаётий ҳолатларнинг ечими сифатида мулоқотда фаол яшайверади. Кўряпмизки, “Ҳангома” билан “Шарқ ҳикояти” шеърлари кўп жихатдан ўхшаш, уларнинг орасидаги фарқ кейингисидан кўчма маънога очиқ ишора борлиги билан белгиланади.

А. Орипов ижодида 80-йилларга томон аллегория характеридаги шеърлар ва шеър-аллегориялар салмоқлироқ ўрин тута бошлайдики, бу ижтимоий шароитнинг ўзгариб бориши, яъни сўз айтиш имконининг кенгайиб бориши ва адабиёт зиммасидаги ташвиқий функциянинг долзарбланиши билан боғлиқдир. Аксинча, шоирнинг шеър-метафоралари кўпроқ сўз-фикр эркинлиги бўлмаган давр маҳсулидир. Кейинги йилларда яратган шеърларини таҳлил қилиш, ўрганиш галдаги вазифамиз бўлиб қолаверади.

#### АДАБИЁТЛАР

1. Арасту. Поэтика. Ахлоқи кабир. Риторика. – Т.: Янги аср авлоди, 2011. – 352 б.
2. Аристотель. Риторика. Поэтика. – М.: Лабиринт, 2000. – С. 264.
3. Боров Ю. Эстетика. – М., 1987. – С. 212.
4. Орипов Абдулла. Танланган асарлар. Тўрт жилдлик. 1-4-жилд. – Т.: Адабиёт ва санъат, 2000. – 346 б.
5. Потебня А. Эстетика и поэтика. – М., 1976. – С. 429.
6. Қуроноф Д. Адабиётшуносликка кириш. – Тошкент, 2007. – 224 б.

УДК: 81.31

### ЎЗБЕК ТИЛИ МИЛЛИЙ КОРПУСНИ ЯРАТИШДА ЛИНГВИСТИК РАЗМЕТКАНИНГ ЎРНИ

#### РОЛЬ ЛИНГВИСТИЧЕСКОЙ УСТАНОВКИ В СОЗДАНИИ УЗБЕКСКОГО НАЦИОНАЛЬНОГО КОРПУСА

#### THE ROLE OF THE LINGUISTIC LAYOUT IN THE CREATION OF THE UZBEK NATIONAL CORPUS

*Тоирова Гули Ибрагимовна  
БухДУ ўзбек тилишунослиги каф. доц.*

**Аннотация.** Мақолада тилнинг Интернет, компьютер технологиялари тилига айланиши математик лингвистика, унинг давоми бўлган компьютер лингвистикасининг шаклланиши ва ривожланиши даражаси билан боғлиқлиги, айниқса, сунъий интеллект учун табиий тилларни моделлаштириши муаммоси хусусида мушоҳада юритилган. Хусусан, разметкалаш матнларга ва уларнинг таркибий қисмларига махсус тегларни лингвистик ва экстралингвистик жиҳатдан ажратиши масаласи тадқиқ этилади. Матн ҳақидаги муҳим маълумотни кодлаш талаблари аниқлаб берилади. Ҳар бир сўз туркумини разметкалаш учун махсус лингвистик модель шакллари ишлаб чиқилиши кераклиги хусусида мулоҳазалар билдирилган. Матнларни разметкалаш формати, матн ҳақидаги муҳим маълумотни кодлаш талаблари ўрганилган ҳамда корпус разметкасининг мавжуд стандартлари эътиборга

олинган. Ҳозирда матнга автоматик ишлов бериш ва матндан турли белгилар асосида қидирувни амалга ошириш тизими мавжуд эмаслигини инобатга олган ҳолда, корпус яратиш учун разметка асосий вазифа эканлиги таъкидланган.

**Таянч сўзлар:** корпус, разметка турлари, морфологик разметка, семантик разметка, синтактик разметка, маълумотларини кодлаш стандартлари, автоматик текшириш, лемма, теглаш.

**Аннотация.** В статье рассматривается язык в сети Интернет, формирование языка компьютерных технологий как продолжение вычислительной лингвистики, ее уровни развития, особенности влияния языка на моделирование нового языка для искусственного интеллекта. Особое внимание уделяется вопросам компоновки текста, распределения специальных тегов с позиций лингвистических и экстралингвистических норм. Уточняет идентификацию и кодирование данных. Освещены вопросы разработки маркировочных моделей для лингвистической типологии. Разметка, текст и формат кодирования также рассматриваются отдельно. Он учитывает стандарты тега для маркировки текста. Кроме того, говорится, что нет автоматической обработки текста. Исходя из этого, указывается, что маркировка в настоящее время является основной функцией лингвистики для создания корпуса.

**Ключевые слова:** корпус, виды разметки, морфологическая разметка, семантическая разметка, синтаксическая разметка, стандарты кодирования данных, автоматическая проверка, лемма, тег.

**Annotation.** The article deals with the language on the Internet, the formation of the language of computer technology as a continuation of computational linguistics, its levels of development, especially the influence of language on the modeling of a new language for artificial intelligence. Particular attention is paid to the issues of text layout, the distribution of special tags from the standpoint of linguistic and extralinguistic norms. Clarifies the identification and coding of the data. The article deals with the development of marking models for linguistic typology. Markup, text, and encoding format are also discussed separately. It takes into account body standards for text marking. In addition, it is said that there is no automatic text processing. On this basis, it is indicated that marking is currently the main function of linguistics to create the corpus.

**Key words:** corpus, enclosure, markings, morphological markup, semantic markup, a syntactic markup, standards, data encoding, auto checking, lemma, tag.

**Кириш.** Жаҳон тилшунослигида муайян тилни сақлаб қолиш, унинг ўрганилиш доирасини кенгайтириш, тил имкониятларини намоён этишнинг асосий воситаси интернет тизимида тил корпусларини яратишдан иборат. Зеро, бугунги кунда XX асрнинг буюк кашфиёти бўлган компьютер технологиялари бошқа соҳалар каби тилшунослик фани учун ҳам кенг имкониятлар эшигини очди, ҳам унинг зиммасига компьютер тили билан боғлиқ улкан вазифаларни юкладди, компьютер лингвистикаси йўналишининг пайдо бўлиши, ривожланиши бевосита шу билан характерланади. Сунъий интеллект учун табиий тилларни моделлаштириш компьютер лингвистикасининг асосий вазифаси. Чунки бугун XXI асрнинг иккинчи ўн йиллигига келиб, компьютер технологияси табиий тилларнинг тараққиёти ҳамда яшаб қолиши учун ҳал қилувчи омилга айланиб улгурди.

Компьютер технологиялари воситасида ахборотни қайта ишлаш, машина таржимаси, электрон луғатшуносликни ривожлантириш, тезауруслар тузиш, тил корпусини яратиш ишларини дунё лингвистикаси цивилизациясига эришган тиллар аллақачон амалга оширди. Инглиз, рус, араб, француз, немис, испан, тожик каби тиллар шулар жумласидандир. Мазкур тилларда интернет тизимида тил корпусини яратиш масаласининг илмий-назарий жиҳатлари ҳам ишлаб чиқилганки, бу ўзбек тилини ҳам интернет “тушунадиган” тилга айлантириш (яъни тил корпусини яратиш) саъй-ҳаракатларини жадаллаштириш заруратини кун тартибига қўяди. Бу борада давлатимиз раҳбари томонидан “...илмий ва ижодий изланишларни ҳар томонлама қўллаб-қувватлаш, улар учун зарур шарт-шароитлар яратиш вазифа”[1,56]

сифатида белгиланиши, шу мақсадда ҳукуматимиз томонидан аниқ чора-тадбирлар ишлаб чиқилиши ва амалга оширилиши, ҳар бир фан соҳасида, жумладан, тилшуносликда ҳам чуқур изланишлар олиб бориш зарурлигини кўрсатади.

Ахборот-коммуникацион тизимнинг ривожланиши фанда “Корпус лингвистикаси” каби янги соҳаларнинг пайдо бўлишига олиб келди. Бу муаммони ўзбек тилшунослигида биринчилардан бўлиб Б. Менглиев [14] ва унинг шогирдлари олиб чиқди [16,52]. Корпус – интернет тизимидаги электрон кутубхона, луғат ва лингвистик грамматикадир. У табиий (реал) тилнинг электрон шаклдаги, қидирув дастурига жойлаштирилган матнлар йиғиндиси.

**Мақсад ва вазифа.** Ушбу мақолада қўйилган масалани ҳал этишдан мақсад **миллий корпус яратиш учун матнларни тайёрлаш жараёнини тавсифлаш ва разметканинг аҳамиятини** ёритиб беришдан иборат. Миллий корпус яратишда лингвистик базани **морфологик разметкалаш** алгоритмини ишлаб чиқиш ҳамда **семантик разметкалаш** билан таъминлаш йўллари аниқлаш олдимизга қўйилган вазифадир.

**Натижа ва ютуқлар.** Жаҳонда рус тилининг миллий корпуси, Замоनावий америкача инглиз тили корпуси, Оксфорд инглиз корпуси, Чехия миллий корпуси, Венгрия миллий корпуси яратилган. Ўзбекистонда лингвистик база вазифасини бажарувчи корпус яратилмаган. Ҳозирда Ziyonet электрон кутубхона мавжуд бўлса-да, у матнга автоматик ишлов бериш ва матндан турли белгилар асосида қидирувни амалга ошириш тизимда ишламайди. У тил ўрганиш, луғат тузиш дастурига мўлжалланмаган. Матнни овозли эшитиш имкони йўқ. Миллий корпус дастури, яъни маълумотлар базаси матнга автоматик ишлов бериш ва матндан турли белгилар асосида қидирувни амалга ошириш тизимига эгадир. Ундан жуда осонлик билан кам ишлатиладиган сўз, ибора ва бирикмани топиш, қўлланиши ва ёзилиши(орфографияси)ни ўрганиш мумкин.

Бу орқали таълимда йўналтириб ўқитишнинг имконияти ёритиб берилади, ўрганувчи матнни овозли эшитиш имконига эга бўлади. Корпус учун разметка ёки теглаш (лингвистик таҳлил) асосий вазифани бажаради. Разметкалаш матнларга ва уларнинг таркибий қисмларига махсус тегларни лингвистик ва экстралингвистик жиҳатдан ажратишдир. Ҳозирда разметкаларнинг қуйидаги турлари бор: морфологик, семантик, синтактик, анафорик, просодик ва бошқалар[10,26]. Экстралингвистик разметка қуйидаги хусусиятлар билан фарқланади: матн форматининг ўзига хослигини акс эттирувчи (боб, хатбоши, қисм ва ҳ.) разметка ва матн, унинг муаллифига тегишли маълумотни ифодаловчи разметка.

Разметка кўпгина замонавий тиллари SGML/XML га асосланган бўлиб, унда белгиланган матн иккита параллел маълумотлар қатламини қамраб олади: кўринарли (матннинг ўзи) ва яширин (тегланган ёки разметкаланган)[10,26]. Бунда маълумотнинг яширин қисми матн ичига жойлаштирилади, лекин махсус маркерлар <...> ичига олинади, улар эса, ўз навбатида, кўринарли матндан ажратиб туради. Аннотация ёзишнинг ташқи усуллари (масалан, изоҳлар)дан фарқли ўларок, разметка доимо матнга инкорпорация қилинади ва унинг ажралмас қисми ҳисобланади. Баъзи корпусларга кейинги таркибий таҳлил даражалари қўлланилади. Хусусан, баъзи бир кичик корпуслар тўлиқ синтактик таҳлил асосида боғланган бўлади.

Бундай ҳолатлар одатда чуқур изоҳланган ёки синтактик тузилишга эга бўлади. Чунончи, синтактик разметка ўзича катта бир дарахтдир. Матнларни қўлда таҳлил қилиш қимматли ва катта вақт талаб қиладиган вазифа ҳисобланади. Ҳозирда рус ва хориж сайтларида очиқ (тўғридан тўғри) кириш мумкин бўлган, таҳлил учун турли дастурий воситалар тақдим этилган. Улар яқка, яъни мустақил ва веб-сайтларга бўлинган. Бу ҳолатда охириги йилларда ишлаб чиқувчиларнинг веб-иловаларга йўналтирилганлиги диққатга сазовордир. Ушбу тизимларнинг қуйидаги афзалликлари мавжуд: бир вақтнинг ўзида бир нечта фойдаланувчи томонидан битта ҳужжат таҳлил(разметкалаш) қилиниш имкониятига эга, қўшимча дастурларни ўрнатишни талаб қилмайди, аммо браузер бундан мустасно, кириш ҳуқуқи чегараланган, разметкалаш жараёнини кузатиш мумкин. Жумладан, Абдулла Қаҳҳорнинг “Нутқ” ҳикоясидан олинган матннинг таҳлил жараёнига эътиборни қаратсак.

Матн: “- Сиз мени яхши кўрмайсиз... Турмушимиздан ҳурсанд эмассиз... Шу соатгача, шу минутгача кутдим, бир оғиз айтмадингизки, бир ёстиққа бош қўйганимизга бир йил бўлди... Нотиқ ҳақиқатан буни унутган эди, лекин гап отди...”

Юқорида келтирилган матн куйидаги хусусиятларга кўра ажратиб чиқилади:

1)

№	Гапнинг тузилишига кўра тури	
1.	[содда гап]	<СГ>, </СГ>
2.	[уюшган гап]	<УГ>, </УГ>
3.	[қўшма гап]	<КГ>, </КГ>

2)

№	Гапнинг ифода мақсадига кўра тури	
1.	[дарак гап]	<дг>
2.	[сўрок гап]	<сг>
3.	[буйруқ гап]	<бг>

3)

№	Гап лисоний қурилишида эганинг ифодаланиш ёки ифодаланмаслигига кўра			
1.	[эғали гап]	<Е+>		
2.	[эғасиз гап]	<Е->	[шахси номаълум гап]	<ш.н.г>
			[атов гап]	<а.г>
			[семантик-функционал шаклланган гап]	<с.фш.г>

4)

№	Бош ва иккинчи даражали бўлакнинг иштирокига кўра	
1.	[ййғиқ гап]	<йг>
2.	[ёйиқ гап]	<ёг>

5)

№	Гап билан грамматик алоқага киришмайдиган бўлакларнинг мавжудлигига кўра	
1.	[ундалма]	<у>, </у>
2.	[киритма]	<к>, </к>

Морфологик разметка тизимига сўзшакл, лемма ва тег киради. Сўзшакл – танланган матндаги морфологик бирлик. Сўзшаклни разметкалашнинг биринчи босқичи уни леммалаш, яъни сўзшаклнинг лексема шаклини келтиришдан иборат. Флектив тилларни разметкалашда энг қийин босқич бу леммалаш (лемматизация), яъни сўзнинг лексема шаклини сўзшаклга тег сифатида бириктириш ҳисобланади. Чунки флектив тилларда сўзшаклдаги грамматик маъно сўз ўзагига қоришиб кетган бўлади. Флектив тиллардан фарқли ўларок, агглютинатив тилда леммалаштириш жараёни анча осон[4]. Дастлаб сўз шакллари учун таҳлил вариантлари рўйхат кўринишида берилади, бунда тўғри вариантни танлаш ёки мавжуд бўлган вариантни таҳрирлаш лозим. Муҳаррир матн бўйича бемалол кўчиб юришни ҳамда глобал алмаштиришлар ва ўзгартиришни амалга ошириш имкониятини яратиб беради. Шундай қилиб, разметкалаш иловаси ўзига таниш муҳитга тушиб, ушбу муҳаррирнинг барча имкониятларидан самарали фойдаланилади. Визуал ажратиш мақсадида, матнинг турли хил элементлари турли ранг ва услубларда безатилади, хусусан:

— разметканинг таҳлили ва буйруқлар варианты яширин матн кўринишида расмийлаштирилади ва одатда оддий режимда кўринмайди;

— сўз шакллари таҳлил вариантнинг сонига кўра турли ранглар билан расмийлаштирилади: ноль, бир ёки бир неча.

Сўзшаклнинг грамматик шаклсиз қисми – ўзак ёки негиз леммага тенг. Разметкада лемма <\*> белгиси ичида берилади. Барча сўз туркумларида леммалаш шу асосда, яъни “сўзнинг ўзак-негиз қисми леммага тенг” тамойилига асосланса, феъл туркумида феъл-лемма II шахс буйруқ-истак майли шаклида берилади. Луғат мақолаларида феъл ҳаракат номи шаклида берилади: <бормоқ>. Аммо бу шакл корпус учун мос эмас, чунки корпусдаги матнда сўзнинг <бормоқ> шакли эмас, <бор> шакли қидирилади. Шундан келиб чиқиб, феъл-лемма ўқитди <ўқи>, бўлмади <бўл>, кўрсатди <кўр>, олди <ол> шаклида берилади

[16, 22]. Разметкаш жараёнида ҳар бир сўзшаклга 5 тадан 10 тагача, баъзан ундан ҳам кўпроқ морфологик тег (изоҳ) ёзиш талаб қилинади.

Разметканинг бошқа тиллари (TeX, RTF)га нисбатан SGML/XMLнинг асосий афзаллиги унда разметка буйруқларининг қатъий синтаксиси, атрибутлари ва элементлари фарқлиниши, элемент чегараларининг аниқ кўрсатмаси, ўз-ўзини ҳужжатлаштириш, ёзувни грамматик жиҳатдан тўғри киритилишини автоматик текшириб бериш имкониятлари мавжудлигидан иборат.

Корпус маълумотларини кодлаш стандартлари ўртасида энг нуфузли стандартлар қуйидагилардир: TEI (Text Encoding Initiative)[5,80 ], CES (XML Corpus Encoding Standard)[6], EAGLES (European Advisory Group on Language Engineering Standards)[8]. Айниқса, батафсил ишлаб чиқилган стандарт сифатида TEI тан олинган бўлиб, турли хил турдаги матнлар ва матнли маълумот элементларининг ифодаланиш қоидаларини аниқлаб беради, бунда қуйидагиларга алоҳида эътибор қаратилади: тузилиш, сарлавҳа, нутқ услублари (наср, шеърят, драма), бетлар, цитаталар, сноска ёки ҳавола (саҳифа ост ичида изоҳ тарзидаги ёзув, изоҳ), тузатишлар, жадваллар, формулалар, махсус символлар (белгилар), лингвистик аннотациялар ва б.к. Стандартнинг махсус сарлавҳаси корпусни кодлаш қоидаларига бўйсунди. Гарчи TEI корпус иловаларига махсус мослашган бўлмасда, кўпгина ҳолларда унга ўхшаш стандартлар билан биргаликда фаолият юритади, масалан, British National Corpus (BNC), Чехия миллий корпуси, Венгрия миллий корпуси ва б.к. XCES стандарти TEI нинг тараққий этган иловаси бўлиб, фақат корпус учун ишлаб чиқилган ва корпусга тегишли ўзига хос тегларни аниқлаш учун мўлжалланган.

Лекин TEI ва XCES универсал стандартларни батафсил ўрганганимизда, матннинг оммавий разметкаланиши учун, улар жуда мураккаб, керагидан ортикча ва ноқулай бўлиши аниқланди. TEI нинг тўлиқ қоидалари жуда кенг кўламли ва ҳар доим ҳам асосли эмас, шунинг учун мазкур стандартнинг барча талабларига амал қилиш етарли даражада қийинчилик туғдиради. Формат ихчамлик билан ажралиб турмайди, шунингдек, мазмунли маълумотга нисбатан разметка кўпайиб кетиши кузатилади. Формат тиниқлик вазифасини йўқотиб қўяди, масалан, мета-атрибуталарни матн кўринишида тег ичида ёзиб қўйиш таклиф қилинади, шунинг учун разметка олиб ташланганда дастлабки матн, асл ҳолига қайтишда хатоликлар юзага келади.

“Ортикча” тегларни инкор этиб, TEI иловалари билан ҳам чегараланса бўлади. Корпусни тақдим этиш учун TEI дан тегларнинг минимал тўплами танланади: Guidelines:<text> — матн, <p> — сатр боши, <s> — гап, <w> — сўз, морфологик таҳлил эса <w ana =...> атрибут кўринишида ёзилади. Лекин бундай кўриниш корпус разметкаси стандартига тўлиқ риоя қилмайди. Мазкур кўриниш HTML нинг содалаштирилган вариантини эслатади.

XML – форматларининг мураккаблиги унинг асосий муаммоси эмас, балки тайёрлаш, қайта ишлаш, индексациялаш ва қидириш каби ҳаммабоп дастурларнинг умуман йўқлиги, катта муаммо ҳисобланади. Тилшуносларнинг ихтиёрида нисбатан оддий дастурлар мажуд. Улардан: XML – анализаторлар, муҳаррирлар, конверторлар, чизиқли қидирув дастурларидан кенг фойдаланиб келинмоқда. Бундан кўринадики, миллионлаб сўз ҳажмига эга корпус учун бундай дастурлар тўплами етарли эмас. Албатта, корпуснинг ички муаммоларини ва разметкасини тайёрлаш каби вазифаларни махсус ёзилган конверторлар, макрослар ва бошқа воситалар ёрдамида бартараф этиш мумкин.

Корпусда маълумотларни тасвирлаш формати, мавжуд бўлган кодлаш стандартлари асосида ишлаб чиқилади (TEI, XCES). HTML тили SGML/XML оиласига мансуб, энг кенг тарқалган формат бўлиб, кўпгина дастурларда ишлаш имкониятига эга[7,12]. Бугунги кунда қидирув тизимлари HTML тегларнинг семантикаси ва тузилишини тушуниш қобилиятига эга.

HTML — разметканинг таркиби ва ҳажм жиҳатдан минимал даражада талабларни тақдим этадиган жуда оддий формат бўлиб, амалда кўп бўлмаган буйруқларни ишлатиш имконига эга. Қўлда қилинадиган таҳрир ва визуал идрок учун жуда қулай ва ихчам формат

хисобланади. Одатда, тил бирликларини намоёиш этишда, стандартнинг ўзида теглар мавжуд бўлмади, лекин HTML ностандарт теглардан фойдаланиш имконини бера олади ва мазкур муаммо қидирув серверининг махсус созлов йўли орқали ўз ечимини топади.

Корпус формати бир нечта HTML тилига эга бўлиб, унда лингвистик бирликлар учун бир мунча махсус теглар бириктирилган. Ушбу формат, матн ҳақидаги муҳим маълумотни кодлаш талабларини аниқлаб беради ва қуйидагиларни ўз ичига олади:

1) метаматнли атрибуталар;

2) матн тузилиш элементлари (сарлавҳа, хат боши, шеърлар, сноска ёки ҳавола (саҳифа ост ичида изоҳ тарзидаги ёзув, изоҳ) ва жадваллар);

3) лингвистик бирликлар (гаплар, сўзлар);

4) лексик маълумот (грамматик, семантик белгилар).

5) матнни расмийлаштириш параметрлари, махсус белгилар ва б.к.[ 10, 98]

Метаматнли атрибуталар матнларда ҳар хил вазиятда ёзиб қўйилади, шунинг учун 2 ва 3 босқичлар параллел ёки ихтиёрий равишда бажарилиши мумкин. Лекин матн идентификацияланган ва қайд этилган файл номига эга бўлиши шарт. Бунда бирор-бир бирикув ёки файл номини ўзгартириш каби амаллар бажарилмайди, чунки бундай амаллар бутун бир тизимнинг ишини бузиб, ўзгартириб қўйиши мумкин. Метамаълумотларни сақлаш мақсадида, олдиндан белгиланган тузилиш бўйича оддий Excel жадвалларидан фойдаланилади, бунда биринчи устунда файл номи (аниқ кўрсатилган йўл), бошқа устунларда эса метаматнли атрибуталар ва технологик маълумот келтирилади. Мазкур амал Excel дастурининг ўрнатилган воситаларидан самарали фойдаланиш имконини беради ва қидирув тизимида анча қулайликлар туғдиради. Масалан, қидирув, филтрация, таҳлил ва маълумотларни қайта ишлаш (амаллар рўйхати, автотўлдириш, статистика). Бунда жадваллар матн форматида сақланиши лозим ва бу форматни Excel тушуниши шарт. Мазкур амал орқали жадвал кўринишда сақланган файл нафақат Excel, балки бошқа жадвалли дастурлар қабул қилиши ва ишлаш имкониятининг самарадорлигини ошириш имкониятини беради.

Назарий жиҳатдан метамаълумот ҳар бир матндан алоҳида ҳолда сақланиши мумкин, лекин HTML қодаларига асосан маълумот файл сарлавҳасида сақланиши керак, шундагина Яндекс-сервер маълумотни индексация қилиш имкониятига эга бўлади. Метамаълумотни алоҳида хотирада сақлаш пайтида синхронизация, мета-жадваллар ва матнларнинг ўзаро келишуви муаммоси доимо келиб чиқади.

Метамаълумотни алоҳида хотирада сақлаш учун қуйидаги усуллар амалга оширилади:

1. *Metas* дастури файл сарлавҳаларидан метаматнли атрибуталарни йиғиб, мета-жадвал сарлавҳаси яратилади. Excel муҳитида қўл ёрдамида ўзгартириб чиқилади. Илк қайта ишлаш босқичидаёқ матнга маълум бир метамаълумот киритилиши мумкин, масалан, муаллиф номи, сарлавҳа ва яратилиш санаси. Охириги босқичда эса *Metas.bat* дастури барча атрибуталарни йиғиб, текширув босқичини яқунлайди.

2. *Meta.txt* дастури метаматнли атрибуталарни тузатилган мета-жадваллардан олиб, мавжуд бўлган матнларга ўтказилади. Ушбу дастур файл мавжудлигини текшириб, сарлавҳани янгилан туради. Жадвалларда атрибуталарнинг кўпчилик амаллари “ ” симболи орқали бўлинади. Матн ўзгартирилганда эса ҳар бир амал алоҳида атрибут кўринишига келади. Шундай қилиб, метаматнли атрибуталар матнлар ва мета-жадваллар орасида эркин ҳаракатланиши мумкин. Метаразметка эса текширувнинг бир неча цикллари билан интерактив бажарилиши мумкин.

3. *MetaTest* дастури мета-жадвалнинг тўғрилигини текшириб беради. Бунда норматив жадвалда атрибута амаллари шаблонларда кўрсатилган амаллар билан қиёсланади. Дастур нотўғри амалларни “#” смиволи билан белгилайди ва қўл ёрдамида текшириб, тузатиш имкони пайдо бўлади.

Юқорида барча кўрсатилган дастурлар Perl тилида амалга оширилади.

Қайта ишловнинг якуний босқичида метаразметкали ва морфологик разметка билан тайёрланган матнлар яна бир неча автоматик трансформациядан ўтаказилади. Конвертор

фигурални кавс ичидаги морфологик таҳлилни <w lex = .....gr=...> тўғри форматга айлантириш орқали, қидирув тизимининг сифатини янада ошириш мақсадида, разметканинг баъзи бир хатоликлари текширилади.

Семантик разметка дастури, махсус семантик луғатдан фойдаланиб, сўзларга асосий семантик белгиларни қўшиб беради. Ушбу усул корпусда семантик қидирувни анча енгиллаштириш хусусиятига эга. Семантик луғат жадвал кўринишида расмийлаштирилган бўлиб, биринчи устунда лексема ва сўз туркуми киритилади, қолган устунларда эса семантик белгилар ёзилади. Дастур сўзнинг морфологик белгиларини луғат билан қиёслаб, ўхшашликни топгандан сўнг, <w> тегнинг sem атрибутида семантик белгиларни кўчиради. Кўп белгили сўзларда эса семантик қидирувда турли хил хатоликлар юз бериши мумкин.

Юқорида келтирилган технология корпус учун матнларни тайёрлашда мураккаб амалларни автоматлаштиришда ёрдам беради. Баъзи амаллар умуман автоматлаштирилмайди (матнларни тозалаш, омонимияни олиб ташлаш, метаразметка), лекин шундай амаллар учун сервис воситаларининг мажмуи ишлаб чиқилган бўлиб, анча ишни енгиллаштириш имкони яратилади. Маълумотнинг илк босқичидаёқ кодлаш формати махсус оддий кўринишда ишлаб чиқилади. Натижада мураккаб разметка ишлаб чиқиш формати охириги босқичда автоматик равишда юзага келади.

**Хулоса.** Юқоридаги фикрлардан келиб чиққан ҳолда қуйидаги хулосага келинди: Разметкаларни лингвистик моделлаштириш мақсадга мувофиқ, чунки лингвистик моделда морфологик тег шартли қисқартма шаклини олади. Ҳар бир сўз туркумини разметкалаш учун махсус лингвистик модель шакллари ишлаб чиқилади. Лингвистик базани морфологик разметкалаш алгоритмининг ишлаб чиқиш зарур бўлади. Лингвистик базани семантик разметкалаш билан таъминлаш йўллари аниқлаш керак. Чунки миллий корпус яратишда ва унинг лингвистик базани тузишда лингвистик разметкалашнинг аҳамияти жуда каттадир.

Ўзбек тилининг халқаро миқёсдаги мақомини ошириш, уни жаҳон мулоқот тили даражасига кўтариш, ўзбек тилини чет элларда ўрганиш ва ўргатиш, миллий тилимизнинг имкониятларини кенгайтириш ва сайқаллаш ишлари ҳам бевосита миллий корпус орқали амалга ошишини назарда тутсақ, ишнинг амалий аҳамияти, электрон-маълумотлар базаси сифатида ўзбек тилининг тараққиёти ҳамда яшаб қолиши учун асосий омил бўлади.

#### АДАБИЁТЛАР

1. Мирзиёев Ш.М. *Эркин ва фаровон, демократик Ўзбекистон давлатини биргаликда барпо этамиз. Ўзбекистон Республикаси Президенти лавозимига киришиши тантанали маросимига бағишланган Олий Мажлис палаталарининг қўшма мажлисидаги нутқ.* – Тошкент: Ўзбекистон НМИУ, 2016. – 56 б.
2. Блумфилд Л. *Язык.* – М.: Прогресс, 1968. – 608 с.
3. Bongers H. *The history and principles of Vocabulary control.* – Woerden: WOCOPI, 1947. – С. 74.
4. Ванюшкин А.С., Граценко Л.А. *Оценка алгоритмов извлечения ключевых слов: инструментарий и ресурсы // Новые информационные технологии в автоматизированных системах.* 2017. - № 20. – С. 95-102.
5. Захаров В.П. *Корпусная лингвистика: Учебно-метод. пособие.* – СПб., 2005. – 48 с.
6. Қутузов А.Б. *Корпусная лингвистика.* – (Электрон ресурс): Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс) - //lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf.
7. Leech G. *The State of Art in Corpus Linguistics // English Corpus Linguistics / Aimer K., Altenberg K.(eds.) – London, 1991. – P. 8-29.*
8. Мельчук И.А. *Порядок слов при автоматическом синтезе русского слова (предварительные сообщения).* // Научно-техническая информация. 1985, №12. -С.12-36
9. Николаев И.С., Митренина О.В., Ландо Т.М. *Прикладная и компьютерная лингвистика.* – М. URSS, 2016. – 320 с.
10. Недошвинова Е.В. *Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. Учебно-методическое пособие.* – Санкт-Петербург, 2006. 26 с.
11. Рахилина Е.В., Марушкина А.С. *Корпусные исследования особенностей речи нестандартных говорящих ("херитажный русский") // Acta Linguistica Petropolitana. Труды института лингвистических исследований.* 2015. Т. XI. № 1. – С. 621-639.
12. Fries Ch.C. *The structure of English. An introduction to the construction of English sentences.* – L., 1969. – С. 98.
13. Френсис Н., Кучера Г. *Вычислительный анализ современного американского варианта английского языка.* – М., 1967.



14. Менглиев Б. Ўзбек тили корпуси яратилляптими? 24 апрель, 2018 йил <http://www.xabar.uz/talim/ozbek-tili-korpusi-yaratilyaptimi>

15. Менглиев Б. Ўзбек тилининг миллий корпуси. 26 апрель, 2018 йил <http://marifat.uz/marifat/ruknlar/fan/1241.htm>

16. Ҳамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари. PhD диссер.автореф. – Қарши, 2018. – 52 б.

17. Ҳамроева Ш. Таълимда корпусдан фойдаланиш “Тил ва адабиёт таълими” журнали. 2017 йил сентябрь, № 9. 49-50-б.

УДК:811.111’362811.512.133’362

## ИНГЛИЗ ВА ЎЗБЕК ТИЛЛАРИДА ТАКРОР ИФОДА УСУЛИНИНГ ХУСУСИЯТЛАРИ

### ОСОБЕННОСТИ ПОВТОРА КАК СРЕДСТВА ВЫРАЖЕНИЯ В АНГЛИЙСКОМ И УЗБЕКСКОМ ЯЗЫКАХ

### PECULIARITIES OF REPETITION AS THE MEANS OF EXPRESSION IN ENGLISH AND UZBEK

*Асқарова Шахноза Камолидиновна*  
ТДШИ Фарбий Европа тиллари  
каф. кат. ўқит., фил.ф.ф.д.,(PhD)

**Аннотация.** Мақола инглиз ва ўзбек тилларида такрор ифода усулининг хусусиятлари таҳлилига бағишланган. Такрор ҳар икки тилга оид манбалардан олинган маълумотлар ҳамда тадқиқотчиларнинг фикрлари асосида таҳлил қилинган. Мақола сўнгида фикрлар умумлаштирилиб, хулоса ясалади.

**Таянч сўзлар:** такрор, восита, таҳлил, тил, халқ, вакил, гап, сўз, маданият, эмоционаллик.

**Аннотация.** Статья посвящена исследованию особенности повтора как средства выражения в английском и узбекском языках. Повтор analyzed on the basis of the sources полученных из источников касающихся, обоих языков и мнений исследователей. В конце статьи обобщается мнения авторов подытоженный по данному вопросу рассматриваемых языков.

**Ключевое слова:** повтор, средства, исследование, язык, народ, представитель, предложения, слова, культура, эмоциональность.

**Annotation.** The article is devoted to the analysis of the peculiarities of repetition as the means of expression in English and Uzbek. Repetition is analyzed on the basis of the sources pertaining to each of two languages as well as the researchers' ideas. At the end of the article thoughts of linguists on the given issue is generalized.

**Key words:** repetition, means, analyze, language, people, representative, sentence, word, culture, emotionality.

**Кириш.** Интенсивлик ҳодисаси борлиқдаги нарса ва ҳодисаларнинг ҳаракати, ҳолати билан бевосита боғлиқ ҳолда намоён бўлади. Тилшуносликда интенсивлик ва эмоционалликни ифодалаш учун бир неча ифода воситалари қўлланади. Такрор маънони кучайтириш, интенсивлик ва эмоционаллик учун хизмат қилади. Бу борада кўплаб тилшунос олимлар изланишлар олиб борганлар. Жумладан, А. Шомақсудов, С. Глазырина, Ш. Балли, Ю. Казакова, Е. Давыдова, И. Туранский каби олимлар такрорни ўзбек, инглиз, рус ва француз тилларида таҳлил этиб, унинг таърифи, турларига алоҳида аҳамият қаратганлар. Шундай бўлса-да, ифода воситаси сифатида такрор тилдаги вазифаси ва турлари юзасидан етарлича тадқиқ этилмаган. Бу жиҳатлар мазкур мавзуда чуқур илмий