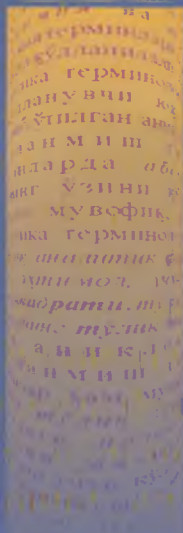


ЛИНГВИСТ

УНАВНІСІ

III



ISBN 978-9943-389-49-6



9 789943 389496

М.А.Абжалова,
тадқиқотчи

КОМПЬЮТЕР ЛИНГВИСТИКАСИДА ЛИНГВИСТИК МОДУЛЬ ТУШУНЧАСИ

Илм-фаннинг ривожланиши ва янгиликларнинг кўпайиши, инновацион технологияларнинг тараққий этиши натижасида, мавжуд терминларга ҳам муносабат ўзгармоқда. Терминларнинг маъно кенгайиши ёхуд маъно торайиши кузатилмоқда. *Модуль* термини лотинча *modulus* сўзидан олинган бўлиб, «кичик ўлчам» деган маънони англатади. Кўп ҳолларда ушбу термин нарсанинг бутунлигига зарар етказмасдан унга қўшимча ишлов бериш имконини берадиган моддийликка нисбатан ҳам қўлланади.

Дунёда кўп фойдаланувчиларига эга «Википедия» сайтида модуль терминининг юқорида биз келтирган таърифидан ташқари маълум соҳалар доирасидаги аҳамиятини ҳам ўқишимиз мумкин.

Информатикада: 1) модуль – дастур файли; 2) модуль – кодни ташкил қилувчи объект; 3) модуль – компьютер совутадиغان система жамланмаси; 4) MOD мусиқий файл формати.

Математикада: 1) абсолют баландлик; 2) вектор модули; 3) автоморфизм модули; 4) бир системадаги логарифмни бошқа системадаги логарифмга ўтказиш коэффициенти, шунингдек, катталиқнинг абсолют қиймати ва ҳ.к.

Механикада: 1) Юнг модули; 2) эластиклик модули; 3) силжиш модули ва ҳ.к.

Бошқа соҳаларда: модуль – дастурнинг тугалланган функционал қисми; модулли ўқитиш – замонавий таълим, яъни билим даражаси бўйича босқичма-босқич ўқитиш; модуль – реклама босмасида графиклар ўлчами ва ҳ.к.

Тилшуносликнинг компьютер лингвистика соҳасида лингвистик модуль термини бугунги кунда муҳим аҳамият касб этмоқда. Зеро, табиий тилнинг компьютер тилига ўтказилиши, яъни компьютер тизими орқали матнга ишлов бериш турли йўллارининг яратилиши кузатилмоқда. Бунинг учун эса чет тилларнинг лингвистик

дастурлари ишлаб чиқилган, ҳозирда улар такомиллаштирилмоқда. *Лингвистик модуль* ана шундай лингвистик дастурларнинг яратилишига асос бўлувчи лингвистик формула ҳисобланади. Лингвистик таъминот эса ушбу элементлар мажмуини ташкил этади. Лингвистик модуллар қанчалик аниқ бўлса, дастурнинг мукамаллиги шунчалик чуқурлашади. Лингвистик дастурларнинг лингвистик модулларини яратишда адабий тил меъёрлари асос бўлади.

Маълумки, тил ўзгарувчан ва ҳар бир тилнинг ўз луғат бойлиги мавжуд. Аммо ҳажм қанчалик катта бўлса, аниқлик шунчалик мавҳумлашади. Ҳар бир сўзнинг қўлланиш шаклларини модуль қилиб дастурга киритиш мушкул вазифадир. Ўзбек адабий тилида саксон мингга яқин сўз мавжуд. Биргина от туркумига оид сўзларнинг шаклларини яратиш тилшуноздан анча диққат ва меҳнатни талаб этади. Мақсадга тезроқ эришиш учун ҳар бир туркум доирасида сўзларнинг лингвистик модули яратилади, яъни сўз-шакллар моделлаштирилади. Сўнгра ўша сўз каби киритилган қўшимчалар комбинациясини олиш эҳтимоли бўлган сўзлар гуруҳланади.

От туркумини моделлаштиришда ўзак сўзларга қўшиладиган қўшимчалар қуйидагича белгилаб олинди:

- кичрайтириш-эркалаш қўшимчалари: $k_e[7] = \{-cha, -choq, -jon, -xon, -oy, -bek, -loq\}$;
- «илиқ муносабатни ифодалаовчи» қўшимча: $i_m = -gina$;
- қарашлилиқ қўшимчаси: $q_q = -niki$;
- ўрин-жой оти қўшимчаси: $u_j = -dagi$;
- чегаралаш қўшимчаси: $ch_q[3] = \{-gacha, -kacha, -qacha\}$;
- кўплик аффикси: $k_a = -lar$;
- келишиқ қўшимчалари (вариантлари билан): $k_q[7] = \{-ning, -ni, -ga, -ka, -qa, -da, -dan\}$;
- эгалиқ қўшимчалари: $e_q[9] = \{-m, -im, -ng, -ing, -lari, -miz, -imiz, -ngiz, -ingiz\}$;
- субстантив шакл ясовчи қўшимча: $sh_y = -lik$.

Келтирилган қўшимчаларни ўзакка бириктириш модулига қуйидаги мисолларни келтириш мумкин ($A=асос$, $N=ҳосила$ сўз):

1. $N=A \cup k_e[7]$; $bolajon = bola \cup jon$ ($N=bola$)
2. $N=A \cup i_m$; $bolagina = bola \cup gina$
3. $N=A \cup q_q$; $bolaniki = bola \cup niki$

4. $N=A\cup u_j$; boladagi= $bola\cup dagi$
5. $N=A\cup ch_q[3]$; bolagacha= $bola\cup gacha$
6. $N=A\cup k_a$; bolalar= $bola\cup lar$
7. $N=A\cup k_q[7]$; bolaning= $bola\cup ning$
8. $N=A\cup e_q[9]$; bolam= $bola\cup m$
9. $N=A\cup k_e[7] \cup k_a$; bolajonlar= $bola\cup jon\cup lar$
10. $N=A\cup k_e[7] \cup e_q[9]$; bolajonim= $bola\cup jon\cup im$
11. $N=A\cup k_e[7] \cup k_a \cup e_q[9]$; bolajonlarim= $bola\cup jon\cup lar\cup im$
12. $N=A\cup k_e[7] \cup e_q[9] \cup k_q[7]$; bolajonimga= $bola\cup jon\cup im\cup ga$
13. $N=A\cup k_e[7] \cup k_a \cup e_q[9] \cup k_q[7]$; bolajonlarimga= $bola\cup k_q[7]; jon\cup k_q[7]; lar\cup k_q[7]; im\cup k_q[7]; ga$
14. $N=A\cup k_e[7] \cup e_q[9] \cup u_j$; bolajonimdagi= $bola\cup jon\cup im\cup dagi$.

Моделлаштириш шу тартибда давом эттирилади.

Сўз туркумларининг лингвистик модулини яратишда амалий натижага эришиш мақсадида, аввало, уларга шакл ясовчи қўшимчалар ва уларнинг комбинацияси сўзларга бириктирилиб, лингвистик базага киритишни оптимал йўл деб олдик. Сабаби сўз ясовчи қўшимчалар салмоқли. Ишнинг дастлабки босқичи тугатилгач, келажакда сўз ясовчи қўшимчалар ва уларнинг аранжировкаси ҳам қамраб олиниши кўзда тутилди. Аммо бу ясама сўзлардан воз кечилди дегани эмас. **Саноқ** (*сон+оқ*), **ўроқ** (*ўр+оқ*), **ўқувчи** (*ўқи+в+чи*), **англамоқ** (*онг+ла*) каби ясама сўзлар «автоматик ҳосила»¹ сифатида лингвистик таъминотга киритилди. *Сон* от туркуми, *саноқ* сифат туркуми сифатида, *ўр* феъл туркуми, *ўроқ* от туркуми, *ўқи* феъл туркуми, *ўқувчи* от туркуми, *онг* от туркуми, *англа* феъл туркумига оид сўзлар сифатида олинади. Бунда, сўзларга ясовчи қўшимчалар қўшилганда, ўзақда рўй берадиган товуш ўзгаришларига алоҳида эътиборнинг камайиши ва иш ҳажмининг кенгайишининг олдини олади.

Ишни бажаришда, яъни матн таҳририни амалга оширувчи дастурнинг лингвистик базасини яратишда ҳар бир сўзнинг қўшимчалар билан келиш ҳолатини алоҳида-алоҳида кўриб чиқиш катта

¹ «Автоматик ҳосила» сўзларга қўшимчалар қўшиш жараёнида ҳосила сўзда рўй берадиган товуш ўзгаришларининг лингвистик таъминотда айнан сақланишидир.

ютуқларга эришишга имкон яратади. Аммо бунга кўп меҳнат талаб қилиниши ҳисобга олиниб ишнинг оптимал варианты ўйлаб чиқилди. Бунда ҳар бир туркум доирасидаги сўзлар, яъни туркум таркиби маъно гуруҳларига ажратилади. Шунинг назарда тутиб от туркумидаги сўзларнинг семантикаси ва аффикслар комбинациясини ўзига бириктиришдаги яқинликлари ҳисобга олиниб маъно гуруҳларига бўлинди.

Ҳозирги кунда матнларни таҳрир қилувчи турли дастурий таъминотлар мавжуд. Бу дастурий таъминотлар чет давлатларда яратилган ва ўша давлат тилининг лингвистик базасига эга ҳамда ўша давлат тилининг меъёрларидан келиб чиққан ҳолда тайёрланган. Ўзбек тилидаги матнларни таҳрир қилиш ҳозирги кунда инсон меҳнати билан амалга ошириб келинмоқда, бу эса, ўз навбатида, матнларни таҳрир қилишга кетадиган вақтнинг кўп талаб этилиши, иқтисодий харажатларнинг ошиши, инсон томонидан қилинадиган хатоларга йўл қўйилиши каби масалаларнинг ечимлари устида бош қотиришга олиб келади. Шу боисдан ўзбек тилидаги матнларни таҳрир қилувчи дастурий таъминот яратиш долзарб масалалардан бири ҳисобланади.

Фойдаланилган адабиётлар:

1. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: Учеб. пособие / Под ред. Е.И.Большакова и др. – М.: МИЭМ, 2011.
2. Коваль С.А. Лингвистические проблемы компьютерной морфологии. – СПб., 2005.
3. Крылов С.А., Старостин С.А. Актуальные задачи морфологического анализа и синтеза в интегрированной информационной среде STARLING // Тр. междунар. конф. Диалог'2003 (<http://www.dialog-21.ru/Archive/2003/Krylov.htm>)
4. Леонтьева Н.Н. Автоматическое понимание текстов: системы, модели, ресурсы. – М., 2006.
5. Новое в зарубежной лингвистике. Вып. XXIV. Компьютерная лингвистика. – М.: Прогресс, 1989.
6. <http://en.wikipedia.org>. Computational linguistics.
7. <http://www.redaktoram.ru>