

КОМПЬЮТЕР ЛИНГВИСТИКАСИ — СУНЪИЙ ИНТЕЛЛЕКТНИНГ БИР ТАРМОҒИ СИФАТИДА

Зарипбоева Нафосат Маратовна

Алишер Навоий номидаги Тошкент
давлат ўзбек тили ва адабиёти университети
“Таржима назарияси ва амалиёти факультети” талабаси
[e-mail: nafosat2502@mail.ru](mailto:nafosat2502@mail.ru)

Аннотация. Ушбу мақолада компьютер лингвистикаси, унинг тараққиёт босқичлари, корпуснинг аҳамияти, унинг татбиқий, амалий электрон ресурс воситаси сифатидаги ўрни ва вазифаси ёритилган.

Калит сўзлар: *компьютер лингвистикаси, корпус лингвистикаси, миллий корпус, таълимий корпус, электрон кутубхона, лугат, репрезентативлик*

Annotation. This article describes the importance of the educational corps, its role and function as an applied, practical electronic resource. It is scientifically proven that the educational corps, as one of the intensive didactic tools, has a valuable role as a key component of the national corps.

Keywords: *corpus linguistics, national corpus, educational corpus, electronic library, dictionary, representative*

Сир эмаски, жаҳон тилшунослигида тилнинг амалий, нутқ, жамият, маданият, миллий тафаккур, ҳатто сунъий интеллект билан узвий муносабатини текшираётган янги йўналишларнинг аллақачон шаклланиб бўлганлиги фанимиз олдида янги вазифаларни қўймоқда. Мавжуд ахборотни топиш, сақлаш, қайта ишлаш ва бошқаларга етказишнинг қулай усуллари бўлган эҳтиёж кун сайин ошиб бормоқда. Бу эса XX асрнинг буюк кашфиёти бўлган компьютер технологиялари соҳаси учун катта вазифаларни вужудга келтирди. Кейинги йилларда тилшунослик ва бошқа фанларнинг ”чорраҳаси”да юзага келган фанлар социолингвистика, психолингвистика, этнолингвистика, математик лингвистика, компьютер лингвистикаси каби соҳалар билан боғлиқ илмий тадқиқотларни олиб боришга эҳтиёж кучаймоқда. Манбаларда “Компьютер лингвистикаси” атамаси билан бир вақтда унга синоним равишда ”ҳисоблаш лингвистикаси”, ”математик лингвистика” каби атамалар ҳам фаол ишлатилганлигини кузатиш мумкин. Мутахассислар аслида компьютер лингвистикаси математик лингвистиканинг мантиқий давоми ҳисобланишини қайд этишади.

Математик лингвистика — фан сифатида XX асрнинг 50-йилларида (аниқроғи 1952 йилда) юзага келган бўлиб, бу фаннинг шаклланишида Копенгаген структурал тилшунослик мактабининг асосчиси, глоссематика¹⁰нинг муаллифи, даниялик олим Л.Ельмслевнинг¹¹ ғоялари ўзига хос туртки вазифасини ўтаган. Математик лингвистика — математиканинг табиий тиллар билан баъзи жиҳатдан ўхшаш бўлган мавҳум структураларини ўрганувчи бир бўлими бўлиб, у табиий тилларни тадқиқ этишда математик усуллардан фойдаланади. Аниқроқ айтганда, математик лингвистика тилшуносликнинг бир соҳаси сифатида табиий тил ходисаларини мавҳумий-семиотик моделлаштириш билан шуғулланади; математик лингвистика фан сифатида эса ана шу моделларнинг энг умумий хусусиятларини тадқиқ этади ва уларнинг тузилиш усулларини ўрганади. Ҳар иккала маънодаги математик лингвистика айтишда бир тушунчавий аппаратдан фойдаланади, улар орасида шунчалик яқин боғлиқлик мавжудки, уни маълум шартлар билан ягона семиотик¹² фан соҳаси деб ҳисоблаш мумкин. Ахборот технологиялари соҳасида бўлган сўнгги ютуқлар ва уларнинг тарққиёти натижасида амалий лингвистика соҳасидаги масалаларга муносабат ҳам ўзгарди. Аниқ ва табиий фанлар қаторида тилшунослик соҳасида ҳам компьютердан самарали фойдаланишга бўлган кизиқиш тобора кучая бошлади. Математик лингвистика тадқиқ этаётган масалалар

¹⁰ Глоссематика (юн. glossematos — тил) — умумлингвистик назария. XX асрнинг 30-йилларида Копенгаген университетида Л.Ельмслев ва бошқалар томонидан яратилган. Глоссематика мохиятан Ф.де Соссюр таълимоти билан алоқадор. Глоссематикадаги лингвистик таҳлил методикаси дедукция сифатида характерланади. У таҳлил қилинаётган матн асосида системани белгилашга олиб келиши керак. Бунинг учун дастлабки маълумотлар — яхлит матн изчиллик билан майда қисмларга (босқичларга, гапларга, сўзларга, бўғинларга ва фонемаларга) бўлинади. Ҳар босқичда бирликни белгилашнинг асосий усули таҳлил бўлади. Глоссематикадаги таҳлилнинг иккинчи муҳим босқичи тил бирликлари ўртасидаги функцияларни ҳисобга олишдир. Ўзаро боғлиқликдан келиб чиқадиган тил ходисаларини расман таҳлил ва баён қилишда глоссематика самарали натижалар беради..

¹¹ Луи Тролле Ельмслев (Louis Trolle Hjelmslev; 3. 10.1899-30.05.1965) Копенгагенлик лингвист. Унинг фикрича, тил семиотик системанинг структурал равишдаги хусусий бир холи бўлиб, уни математика, мантик ва семиотиканинг талаблари руҳида катъий формаллаштириш мумкин. Л.Ельмслев ҳатто тил ходисаларини математик баёнда тушунтирадиган фаннинг номини ҳам таклиф этган. Олимнинг фикрича, бу фан “тил алгебраси” деб аталиши лозим эди.

¹² Семиотика (юн. semeiotikos — белгилар ҳақидаги таълимот). Лингвистик семиотикадаги асосий тушунча лисоний белги бўлиб, у муайян матнлардаги ва муомала вазиятидаги шакл ва мазмуннинг диалектик бирлиги сифатида намоён бўлади. Семиотиканинг “белгилар ҳақидаги фан” сифатида умумий қоидалари, хусусиятлари дастлаб америкалик мантиқшунос Ч.С.Пирс ва швейцариялик тилшунос Ф.де Соссюр асарларида бир вақтнинг ўзида ва мустақил равишда кайд этилган. Бу тушунча учун Ч.С.Пирс “семиотика” терминини, Ф. Де Соссюр эса “семасиология” терминини қўллаган. Семиотика ғояси кейинчалик америкалик файласуф Ч.У.Моррис томонидан тартибга солиниб ривожлантирилди. Бир қанча европа ва рус олимлари ҳам семиотиканинг ривожига ҳисса қўшганлар. Семиотика муаммоларини ишлаб чиқиш математик мантик, математик лингвистика, информатика ва ҳоказолар учун ҳам муҳим аҳамиятга эга.

доирасида энди нафакат математик-лингвистик моделларни яратиш, балки, мавжуд моделларни қандай қилиб компьютерга татбиқ этиш масаласи кўрила бошланди. Шу тарика ”математик лингвистика” атамаси ўрнига ”компьютер лингвистикаси” атамаси истеъмолга кириб келди.

Англашиладики, компьютер лингвистикаси — бу сунъий интеллектнинг муҳим йўналишларидан бири бўлиб, у ўз олдига табиий тилларни электрон кўринишда ифодалашга мўлжалланган математик моделлардан фойдаланиш масаласини қўяди. Компьютер лингвистикасининг асосий мақсади табиий тилнинг математик моделини яратиш, қайта ишлаш ва лингвистик муаммоларни ҳал этишга оид компьютер дастурларини ишлаб чиқишдан иборат. Компьютер лингвистикасининг фан сифатида дунёга келиши энг аввало, компьютер воситасида автоматик таржиманинг муваффақиятли амалга оширилиши билан белгиланади. Янги авлод компьютерлари ҳамда дастурлаш тилларининг яратилиши машина таржимасига оид янги илмий соҳа учун бир туртки бўлган эди. Бир тилдан бошқасига механик таржима қилиш ғояси илк бор немис математиги Г.Лейбниц¹³ томонидан илгари сурилган эди. Бу ғоя таъсирида шаклланган кўпгина ихтирочилар кейинчалик механик таржима қурилмаларини яратишга амалий жиҳатдан муваффақ бўлдилар, жумладан, 1924 йили А.Вахер (Эстония), 1933 йили Пётр Троянский (Россия), худди шу йили Георгий Арцруни (Франция) каби олимлар томонидан механик таржимонлар (лингвистик арифмометр) яратилди. Бир тилдан иккинчи тилга таржимани амалга оширишда ЭХМлардан фойдаланиш масаласи биринчи марта 1946 йилда Эндрю Бут (Andrew D. Booth), Уоррен Уивер (Warren Weaver) ва Норберт Винер (Norbert Wiener) лар томонидан илгари сурилди ва уларни амалга ошириш усуллари аниқ кўрсатиб берилди. Шундан сўнг, 1949 йили АҚШда машина таржимаси муаммолари билан шуғулланувчи илмий жамоа ташкил этилди. 1952 йили Массачусетс технология институтида айнан машина таржимаси мавзусига бағишланган I Халкаро конференция ўтказилган. 1954 йили АҚШнинг Жоржтаун университети¹⁴да илк бор рус тилидан инглиз тилига машина таржимаси бўйича илмий тажриба ўтказилди ва бу оламшумул воқеа бутун дунеда катта шов-шувга сабаб бўлди. Шу тарика 1954 йил компьютер

¹³ Г.Лейбниц (Leibniz Gottfried Wilhelm) — 1646 йилда Лейпцигда туғилиб, 1716 йил 14 ноябрда Ганноверда вафот этган бу немис файласуфи математика, физика, мантик, тарих ва дипломатия соҳаларига ҳам илмий жиҳатдан салмоқли ҳисса қўшган қомусий олимлардан саналади.

¹⁴ Жоржтаун университети (инг. Georgetown University) — Вашингтонда жойлашган. Университетга 1789 йили Жон Кэрролл томонидан асос солинган.

лингвистикаси дунёга келган йил сифатида эътироф қилина бошланди. Шундан кейин жаҳоннинг Англия, Франция, Германия, Хитой, Япония ва Болгария каби кўпгина мамлакатларида олимлар машина таржимаси билан махсус шуғуллана бошладилар.

1962 йилда Кембридж шаҳрида жаҳон тилшуносларининг IX конгресси ўтказилди. Бу анжуманда структурал тилшунослик ва математик лингвистика муаммоларини муҳокама қилишга алоҳида ўрин ажратилди. 1967 йили Бухарестда жаҳон тилшуносларининг X конгресси ўтказилиб, бунда ҳам математик лингвистика ва машина таржимаси муаммолари махсус муҳокама қилинди. 1955 йилга келиб, Москвада ҳам машҳур математик ва кибернетик А.Ляпунов¹⁵ раҳбарлигида машина таржимаси борасидаги ишлар фаол йўлга қўйилди. Унда асосан инглиз тилидан рус тилига ва аксинча, рус тилидан инглиз тилига йўналишда машина таржимасига эътибор қаратилган бўлиб, бу жараёнга рус олимларидан П.Гаврен, Л. Достерт, П.Шеридан, И.Бельская, Л.Королев, С.Разумовскийлар ҳам фаоллик билан иштирок этишган эди. Компьютер лингвистикаси бўйича рус компьютер тилшунослигидаги тадқиқотлар диққатга сазовордир. Унда қуйидаги йўналишларни алоҳида таъкидлаб кўрсатиш мумкин:

1. Машина таржимаси.
2. Автоматик таҳрирлаш.

Тил миллий маънавият, маърифат ва маданиятнинг кўзгусидир. Ҳадиси шарифларда "Кишининг зебу зийнати, гўзаллиги унинг тилидадир" дейилади. Қадим қадим замонлардан бери ҳар бир миллат, ҳар бир қавм ўз тилига буюк ҳурмат билан қарайди. Зотан, тил миллат деган бирликнинг асоси. У бўлмаса, миллат ҳам бўлмайди. Тилнинг расмий мақоми миллат ёки халқ учун жуда улкан ижтимоий-сиёсий, маънавий-маърифий, тарихий, ҳуқуқий аҳамиятга молик масала. Чунки расмий тил соҳибларининг миллий-маънавий яхлитлигини таъминловчи асосий восита ҳисобланади. Шу маънода миллий яхлитликка эришиш, ватанпарварлик миллий тилга эътибордан бошланиши бежиз эмас. Сир эмаски, XXI аср бошларида ўзида миллионлаб сўзларни акс эттирувчи юзлаб тил корпуслар пайдо бўлди. Сунъий интеллектнинг автоматик таржима, компьютер таҳлили, таҳрири, тезаурус, электрон луғат сингари имкониятлари кенгайди, илмий-назарий асослари яратилди, амалиётда қўллаш мумкин бўлган илк намуналари қўлланила бошлади.

¹⁵ Ляпунов Алексей Андреевич (25.09.1911-23.06.1973) — Физика-математика фанлари доктори, профессор. Россия ФА нинг мухбир аъзоси.

Фандаги бу янгиланишлар ахборот технологияларини тилшуносликка татбиқ этиш билан боғлиқ истиқболли илмий йўналишлар пайдо бўлишига йўл очди. Бу эса корпус, корпус лингвистикаси, унинг шаклланиши, тараққиёти, бугунги ҳолати ва корпус тузишнинг умумий тамойилларини ўрганишнинг долзарблигини белгилайди. Жаҳон тилшунослигида компьютер ва корпус лингвистикаси муаммоларини ўрганиш XX асрнинг 60-йилларидан бошланиб, бу соҳада дастлабки илмий фаразлар айтилди. Ишонарли лингвистик маълумотлар катта массивли матнлар мажмуасидангина олинishi мумкин деган қараш ўтган асрнинг 60-йилларида Р.Г.Пиатровский томонидан айтилган бўлса-да, корпус соҳасидаги мақсадли тадқиқотлар 40-йилларда Блумфильд, Фрайс ва Бонджерслар томонидан бошланган. Браун корпуси (1961-1964) тузувчилари Нилсон Френсис ва Генри Кучера илк марта корпус тузиш тамойилларини ишлаб чиққан. Бу борада инглиз тили банки (1980) лойиҳаси муаллифи Жон Синклер ишлари ҳам эътиборга сазовор. Рус тилшунослигида В.П.Захаров, А.Б.Кутузов, Е.В.Недошивина, В.В.Риков, В.Плунгянлар корпус, унинг турлари, ўзига хос хусусияти, корпуснинг ижтимоий аҳамияти, корпус тузиш тамойиллари борасида тадқиқот олиб боришган. Муаллифлик корпуслари О.В.Кукушкина, А.А.Поликарпов, Е.В.Суровцевалар томонидан тадқиқ этилган. Ўзбек тилшунослигида компьютер лингвистикаси, матнга лексикографик ишлов бериш ва лингвостатистик таҳлил этиш борасида бирмунча тадқиқотлар амалга оширилган. Аммо бу тадқиқотлар матнни инновацион ёндашув компьютер лингвистикаси ютуқлари ёрдамида лексикографик ва лингвостатистик тадқиқ этишнинг замонавий усуллари тавсия этганлиги билан долзарблик касб этган бўлса-да, уларнинг ҳеч бирида ўзбек тили корпусларини яратиш масаласи кун тартибига қўйилган эмас. Ўзбек тили дунёнинг кўплаб мамлакатларида, жумладан, АҚШнинг Вашингтон, Калифорния университетларида, Германиянинг Берлин Гумбольдт университетида, Туркиянинг Анқара университетида, Хитой, Жанубий Корея, Япония университетларида ўқитилмоқда. Такмиллашиб бораётган компьютер лингвистикаси йўналишида автоматик таржима сифатини яхшилаш, тилни лингвистик моделлаштириш, ҳар бир тилга оид сўзларни леммалаш назарияси, алгоритминини яратиш ҳамда муайян тилнинг кўп асрлик миллий-маданий меросдан фойдаланиш имконини ошириш мақсадида уларни электронлаштириш жаҳон тилшунослигида долзарб масалага айланди. Тилшуносликда, хусусан, компьютер лингвистикаси соҳасида



корпус яратиш, мавжуд корпуслар ҳажмини кенгайтириш, матнни автоматик қайта ишлайдиган дастурларни ишлаб чиқиш кабилар ечимини кутаётган муҳим масалалардан бири бўлиб турибди. Ўзбек тилининг миллий корпуси, унинг ўзига хос хусусияти, ижтимоий, лексикологик, таълимий ва бошқа соҳалардаги аҳамияти, корпус лингвистикаси тарихи, корпус турлари, таълимий корпусининг лингвистик қимматини ўрганиш, ўзбек тили миллий корпусини яратишнинг лингвистик асосларини ишлаб чиқишнинг асосий мезонидир.

Фойдаланилган адабиётлар рўйхати:

1. Блумфилд Л. Язык.-М.: Прогресс, 1968.608 с.; Fries Ch.C. The structure of English. An introduction to the construction of English sentences. – L.,1969.;
2. Andrew Wilson, Dawn Archer, Paul Rayson Language and computers studies in practical linguistics No 56/Corpus linguistics around the world New York, 2006. 242 p.
3. В.П. Захаров, С.Ю. Богданова Корпусная лингвистика Учебник Санкт-Петербург 2013.
- 4.www.sal.tohoku.ac.jp/ling/corpus1/1early.html
- 5.<http://corpus.leeds.ac.uk/tools/>
- 6.<http://tugantel.tatar/?lang=ru>
7. <http://www.badennet.org/>