



## КОРПУС ВА КОМПЬЮТЕР ЛИНГВИСТИКАСИНИНГ ФАРҚЛИ ЖИХАТЛАРИ

*Атабоев Нозимжон Бобожон ўгли,  
ЎзДЖТУ 1-босқич таянч докторанти*

*Калит сўзлар:* корпус лингвистикаси, компьютер лингвистикаси, корпус, компьютер таржима, аутентик ва ярим-аутентик матнлар.

Компьютер технологияларининг ривожини натижасида замонавий тилшуносликда корпус ва компьютер лингвистикаси терминларини семантик таҳлил, корпус лингвистикаси (BNC), критик дискурс таҳлил (CDA), шунингдек, дескриптив, қиёсий ва компонент таҳлил методлари асосида тадқиқ этиш тобора долзарб аҳамият касб этиб бормоқда. Тилшунослик нуқтаи назаридан ёндашсак, компьютер лингвистикасининг пайдо бўлиши бугунги кунда фаннинг бир қанча муаммоларига ечим топиш имкониятини беради. Компьютер технологияларини тилшуносликнинг илмий-назарий, фалсафий муаммолари: тил ва нутқ, система ва структура, семиотика, синтагматика ва парадигматика, типология, луғатшунослик каби муҳим ҳамда мураккаб масалалари бўйича олиб борилаётган илмий тадқиқотларда қўллашнинг ижобий томонларига таянган ҳолда корпус лингвистикаси соҳаси яратилди. Компьютер ва корпус лингвистикаси бир-бирига ўхшаш ва баъзи ҳолларда бир хил деган нотўғри қарашлар учраб туради. XX асрнинг 90-йиллари биринчи ярмида корпус тилшунослиги тил тўғрисидаги фаннинг алоҳида қисми сифатида шаклланди. Шу билан бирга, у компьютер тилшунослиги билан яқиндан ҳамкорлик қилади, унинг ютуқларидан фойдаланади ва, ўз навбатида, уни бойитади.

Корпус лингвистикасининг назарий ва амалий томонлама муҳим бўлган муаммолари чет эл олимлари томонидан ўрганилган ва улар сирасига қуйидагиларни киритиш мумкин: И.Ф.Ганиева корпус лингвистикасига оид терминларнинг ишлатилиши юзасидан ўз қарашларини ифода этган, Ж.Синклиер эса

корпус лингвистикасидан фойдаланишнинг аҳамиятли томонларини муҳокама қилган. Д.Бибер, С.Конрад ва Р.Реппенлар корпус лингвистикасининг тил ўрганишдаги ўрнини тадқиқ қилишган, Х.Линдквист, Н.Лич, Н.С.Даш, Ф.Мейер, Т.Грайс, Ф.Карен, С.Хунстон, Т.Мк Энерий, З.Хиао, Ю.Тоно, А.Вилсон<sup>23</sup> ва бошқалар мазкур муаммони турлича ёндашув асосида ўрганишган. Ушбу соҳада ўзбек тилшунослигида сезиларли даражада тадқиқот олиб борилмаган.

Қайд этилган манбалар корпус лингвистикасининг хусусиятларини ўрганишда муҳим илмий ва амалий аҳамиятга эга. Компьютер лингвистикаси ҳам ўз муаммо ҳамда ютуқлари нуқтаи назаридан ўрганиб чиқилган. А.Н.Баранов ўз ишларида сунъий онг муаммосини очиб беришга тўхталган, Б.Орехов семантик

<sup>23</sup> Ганиева И.Ф. Об использовании корпусов в лингвистических исследованиях// Филология и искусствоведение. Т.12. – БГУ, Россия, 2007, № 4. С.104–106.; Sinclair J. 1991. Corpus and text-basis principles. – Oxford: Oxford University Press.; [Biber D.](#), [Conrad S.](#), [Reppen R.](#) Corpus Linguistics: Investigating Language Structure and Use. – Cambridge University Press, 1998. – P. 300; Lindquist H. Corpus Linguistics and the Description of English. – Edinburgh: Edinburgh University Press, 2009.; Leech G. Corpora and theories of linguistic performance. In J. Svartvik (ed.), 1992. – P. 105–122.; Dash N. S. Corpus Linguistics: A General Introduction. – CIIL, Mysore, 2010.; Meyer C.F. English Corpus Linguistics. – Cambridge: Cambridge University Press, 2002; Gries S.Th. Introduction to S. Gries and A. Stefanowitsch (eds.). Corpora in cognitive linguistics: Corpus-based approaches to syntax and lexis, 1–18. – Berlin, 2006.; Gruyter M. de. Geoffrey Th., Hunston S. (eds.). System and corpus: Exploring connections. – London: Equinox, 2006.; McEnery T., Xiao R.Z., Yukio T. Corpus-based language studies: An advanced resource book. – London: Routledge, 2005.; McEnery T. and Wilson A. Corpus linguistics. – Edinburgh: Edinburgh University Press, 1996.



анализ, техник таржима ва автоматик морфология томонидан фанни ўрганиб чиққан бўлса, И.А.Чесебиев компьютер лингвистикасининг методлари ривожланиш тарихини ўрганиб чиққан. Шу ўринда, Н.Хомский (Chomsky)нинг синтактик структуралар ва универсал грамматика назарияси фан ривожига катта хисса қўшган. Ч.Филмор (Fillmore) эса ушбу назарияга қўшимча сифатида тиллардаги предлог каби ноўхшашликлар ва уларнинг тил структурасида бошқа тушунчалар орқали ифода этилиши нуқтаи назардан ёндашган. У.Вудс (Woods)<sup>24</sup> фанга процессуал семантика ғоясини киритди.

Бир гуруҳ олимлар уни компьютер тилшунослигининг таркибий қисми сифатида баҳолашади: корпус тилшунослиги – компьютер техникаси ёрдамида тилшунослик корпусини қуриш ва улардан фойдаланишнинг умумий тамойилларини ишлаб чиқадиغان бирликларни яратиладиган компьютер тилшунослиги. Шу билан бирга, одатда, компьютер асбоб-ускуналари, дастурлари, ташкилотнинг компьютер технологиялари ва маълумотларини қайта ишлашнинг кенг доирада ишлаш механизмини моделлаштириш тарзида тушунилади.

Компьютер тилшунослиги корпус тилшунослиги учун воситаларни (яъни дастурларни) яратади. Шу маънода улар бир-бирини тўлдирядилар. Масалан, корпус тилшунослари корпусда сўзларни автоматик тарзда белгилаш учун махсус воситаларга муҳтож. Агар юз миллиондан

ортиқ фойдаланадиган матн бирлиги мавжуд бўлиб, ҳар бир сўз учун нуткнинг бир қисмини белгилаш керак бўлса, бу жараёни тўлалигича қўлда амалга ошириб бўлмайди, шу боис тизимли дастурий таъминот талаб қилинади. Унга кўра сўзларни таснифий белгиларига кўра белгилаш автоматик равишда амалга оширилади.

Компьютер тилшунослигини қўллашнинг айрим жиҳатлари қуйидагилар:

- 1) автоматик таржима;
- 2) аутентик ва ярим-аутентик матнлардан автоматлаштирилган маълумотларни олиш;
- 3) инсон ва машина ўртасида қулай интерфейсларни яратиш;
- 4) табиий тилларда алоқанинг *кантитатив*<sup>25</sup> тавсифи.

Шу билан бирга, компьютер тилшунослиги “одатда, компьютер воситалари (дастурлари, маълумотларни ташкил қилиш ва қайта ишлаш учун компьютер технологиялари)ни муайян шароитларда, вазиятларда, муаммоли соҳаларда ва тиллардаги моделларнинг кўламини нафақат тилшуносликда, балки бошқа фанларга ҳам қўллашни назарда тутуди”[1.13]. А.Н.Баранов “тилни компьютер моделлаштириш тилшуносликдаги дастурлаш назарияси (компьютер моделларини қуриш ва компьютер графикаси фани)ни қўллаш соҳаси сифатида қаралиши мумкин. Чунки ушбу муаммо кейинги йилларда фақатгина тилшунослик масаласи сифатида эътироф этилмоқда”, деб таъкидлаган[1]. Корпус тилшунослиги компьютерни аниқ восита сифатида ишлатади ва уларсиз, албатта, у бажарадиган вазифаларни бажаролмайди. Бироқ бу замонавий билимларнинг деярли ҳар бир соҳасига тааллуқли бўлиб, уларни компьютер фанининг ажралмас қисмларига айлантормайди.

Корпус тилшунослиги ўз материалларини яратади, анироғи, уни мустақил равишда тузади. Бу эса, мустақил

<sup>24</sup> Баранов А.Н. Категории искусственного интеллекта в лингвистической семантике. Фреймы и сценарии. М., 1987; <https://postnauka.ru/video/66255>; Чесебиев И.А. Компьютерное распознавание и порождение речи / И.А. Чесебиев. – М.: Спорт и культура, 2008 – 128 с.; Хомской Н. Синтаксические структуры. – США, 1957.; Fillmore Ch. Frame semantics and the nature of language// In Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech, 1976. –P. 20-32.; Woods W. “Procedural Semantics as a Theory of Meaning” in A. Joshi, B.L.Webber and I. Sag (eds.), Elements of Discourse Understanding, Cambridge University Press, 1981.

<sup>25</sup> Quantitative – (инг. – сонга асосланган) ихтиёрий танланган манба хусусидаги хулосани аниқ рақамлар асосида чиқариш.



тилшуносликни ўз ичига олади – у ишлатилган оғзаки материалларнинг ўзига хос хусусияти ва воситалари, яъни корпуслар таҳлили учун дастурларга эгадир. Илм-фаннинг мустақиллиги таҳлил материаллари ёки тадқиқот усуллари мавжудлиги билан белгиланади. Корпус тилшунослиги унисига ҳам, бунисига ҳам эга [4].

Корпус лингвистикасига тилшуносликнинг лисоний корпуслар яратиш ва улардан фойдаланишни ўрганадиган бўлими сифатида қараймиз ва бу танлаб олинган матнлар асосида ифодаланган тилларни ўрганиш демакдир. Баъзи олимлар эса корпус лингвистикасини компьютер лингвистикаси қамровидаги соҳа дея эътироф этишади: Корпус лингвистикаси – компьютер лингвистикаси бўлими бўлиб, компьютер технологиялари ёрдамида тил корпусларини яратиш ва қўллаш муаммолари билан шуғулланади. “Компьютер лингвистикаси – компьютер технологиялардан тил фаолиятларини моделлаштиришда маълумотлар йиғиш, саралаш ва кенг қўламда фойдаланишдир. Баъзи муаммоли вазиятларга кўра, тилнинг компьютер модели нафақат тилшуносликда, балки у билан чамбарчас боғлиқ бўлган фанларда ҳам қўллаш назарда тутилади” [2]. А.Н.Баранов таъкидлаганидек, “компьютер лингвистикасига тилга асосий урғу берилмасдан, балки энг сўнгги ўринларда эътибор қаратилиб, бунда тилнинг компьютер моделлаштируви тилшунослик соҳаларидаги программалаштиришга оид техник таъминот назарияларигагина аҳамият берилади” [1]. Корпус лингвистикаси эса компьютерлардан восита сифатида фойдаланади. Тан олиш керакки, агар компьютерлар бўлмаганида у ҳозирдаги фаолиятини амалга ошира олмас эди. Фанларнинг тараққиёти натижасида корпус лингвистикаси барча фанларда татбиқ этилади, лекин бу орқали ушбу фанлар компьютер соҳаси (computer science)нинг бир қисми бўлиб қолмайди.

Замонавий тилшуносликда ҳар иккала фаннинг ўрганилиш даражаси ва

мазкур соҳалар бўйича дунё олимларининг изланишлари ва илмий қарашларига таянган ҳолда компьютер лингвистикаси корпус лингвистикасидан қуйидаги хусусиятларига кўра кескин фарқланиши бўйича хулосалар чиқарилди:

1) номланишидаги фарқ;

2) мазкур икки соҳанинг белгиланган мақсадлари жиҳатидан:

а) компьютер лингвистикаси тилга оид программаларни ишлаб чиқишни ўз олдига асосий мақсад қилиб қўяди. Таъкидлаш жоизки, бу бевосита тилшунослик олдидagi муамоларни ҳал этишга қаратилмаган;

б) корпус лингвистикаси эса, фаннинг мақсади сифатида мавжуд тилнинг корпусини яратиш ва унинг ёрдамида тилдаги сўзлар, иборалар ҳамда бошқа лисоний бирликларнинг қўлланилиш частотасини аниқлашни белгилаб қўйган. Бу эса корпус лингвистикасини айнан тилшуносликка оид эканлигини кўрсатади.

3) компьютернинг барча фанларда фойдаланиш ўрни ва аҳамияти томонидан:

а) компьютер лингвистикасида компьютер қурилмалари асос бўлиб, ушбу фандаги барча изланишлар компьютерлар ва уларнинг афзалликларини оширишга қаратилади;

б) корпус лингвистикаси эса компьютер қурилмаларидан фақатгина восита сифатида фойдаланади, бунда матнларни йиғиш ва саралаш амалиётларини бевосита улар ёрдамида амалга оширади.

Ҳар қандай юксалаётган фан ўз амалиётида компьютерлар тараққиётининг ютуқларидан фойдаланади, лекин ушбу фанлар (корпус лингвистикаси, компьютер лингвистикаси, компьютер математикаси, компьютер физикаси) компьютершуносликнинг бир қисми сифатида қабул қилиниши жоиз эмас. Мазкур фанлар таҳлил объектини тадқиқ этишда компьютердан восита сифатида фойдаланади, холос.

Юқоридаги таҳлиллардан келиб чиқиб, шуни айтиш мумкинки, корпус лингвистикаси компьютер



лингвистикасининг алоҳида бир қисми ёки бўлими эмас. Улар йўналиш ҳамда

тадқиқот объекти, мақсад ва муаммолари турлича бўлган мустақил фанлардир.

#### Адабиётлар:

1. Баранов А.Н. Компьютерная лингвистика // Баранов А.Н. Введение в прикладную лингвистику: учебное пособие. – М.: Едиториал УРСС, 2003. – С. 13.
2. Баранов А.Н. Введение в прикладную лингвистику: Учебное пособие. – М.: Эдиториал УРСС, 2001. – 360 с.
3. Мамонтова В.В. Корпусная лингвистика. Особенности перевода сложносоставных слов с английского языка на русский: на материале корпуса публицистических текстов: Дисс. ... канд. филол. наук. – Ставрополь, 2008. – 139 с.
4. <https://www.myfilology.ru/177/osnovnye-zadachi-i-napravleniya-korpusnoj-lingvistiki-vzaimodejstvie-korpusnoj-lingvistiki-i-kompyuternoj-computational-lingvistiki/>

*Атабоев Н. Различительные свойства корпусной и компьютерной лингвистики. В статье рассматриваются различия между сферами корпусной лингвистики и компьютерной лингвистикой. Утверждается, что обе науки независимы друг от друга.*

*Ataboyev N. Differentiating features of corpus and computer linguistics. The article deals with the clarification of the differences between the spheres of corpus linguistics and computer linguistics. It is stated that both of the sciences are independent from one another.*