

УЎК (УДК, UDC): 81'33

DOI: 10.36078/1564985802

КОРПУС ЛИНГВИСТИКАСИНИНГ АСОСИЙ ХУСУСИЯТЛАРИ



Нозимжон Бобожон ўғли АТАБОЕВ

Докторант (PhD)

Ўзбекистон давлат жаҳон тиллари университети

Тошкент, Ўзбекистон

anb929292@gmail.com

Аннотация

Мақола замонавий тилшуносликнинг корпус лингвистикаси ва компьютер лингвистикаси соҳалари ўртасидаги мавжуд фарқларни аниқлашга бағишланган. Ушбу фанларнинг бир-бирига боғлиқ бўлмаган мустақил соҳалар эканлиги кўрсатилган. Келтирилган фараз ва фикр-мулоҳазалар хорижий адабиётлар асосида тадқиқ этилган ҳамда хулосалар берилган. Соҳалар ўртасидаги ўзаро фарқли жиҳатларга аниқ мезонлар асосида белгилаб кўрсатилган.

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

О СПЕЦИФИКЕ КОРПУСНОЙ ЛИНГВИСТИКИ

Нозимжон Бобожонович АТАБОЕВ

Докторант (PhD)

Узбекский государственный университет мировых языков

Ташкент, Узбекистан

anb929292@gmail.com

Аннотация

В статье рассматривается выяснение различий между сферами корпусной лингвистики и компьютерной лингвистикой в современной науке. Утверждается, что оба направления независимы друг от друга. Эти гипотезы и мнения были в значительной степени исследованы в зарубежной литературе. Различия между названными направлениями были уточнены в соответствии с конкретными критериями.

Ключевые слова: корпусная лингвистика; компьютерная лингвистика; корпус; компьютерный перевод; аутентичные и полуаутентичные тексты.

ON SPECIFIC FEATURES OF CORPUS LINGUISTICS

Nozimjon Bobojon ugli ATABOEV

Ph.D. student,

Uzbekistan State University of World Languages

Tashkent, Uzbekistan

anb929292@gmail.com

Abstract

The article deals with the clarification of the differences between the spheres of corpus linguistics and computer linguistics in the modern linguistics. It is stated that both of the sciences are independent from one another. These hypotheses and opinions have been largely investigated by foreign scholars and the final conclusions have been made. Differences between the fields have been clarified according to specific criteria.

Keywords: corpus linguistics; computer linguistics; corpus; computer translation; authentic and semi authentic texts.

Компьютер технологияларининг ривожини натижасида замонавий тилшуносликда корпус ва компьютер лингвистикаси терминларини семантик таҳлил, корпус лингвистикаси, критик дискурс таҳлил, шунингдек, дескриптив, кийимий 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 (4) корпус лингвистикасига оид терминларнинг ишлатилиши юзасидан ўз қарашларини ифода этган, Ж. Синклиер (18) эса корпус лингвистикасидан фойдаланишнинг аҳамиятли томонларини муҳокама қилган. Д. Бибер, С. Конрад ва Р. Реппенлар (8) корпус лингвистикасининг тил ўрганишдаги ўрнини тадқиқ қилишган, Х. Линдквист (15), Н. Лич (14), Н. С. Даш (9) корпусларнинг классификациясини ишлаб чиққан, Ф. Мейер (11) корпус лингвистикасига оид умумий тушунчаларни ўрганиб чиққан, Т. Грайс (12) гап тузилиши ва сўзларнинг қўлланилиш кўламига тил корпуслари нуқтаи-назаридан ёндашишни таклиф этган, С. Хунстон (13) лисоний корпус таркибидан қидирув натижалари асосида боғловчиларнинг систематик таҳлиliga урғу берган, Т. МкЭнерий, З. Хиао ва Ю. Тоно (17) корпус таҳлилларидан тил ўқитишда фойдаланишнинг ижобий

томонларини очиб берган, А. Вилсон (16) 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. Шунингдек, олиб борилган ва борилаётган изланишлар соҳа муаммоларининг мавжудлиги ҳамда уларга ечим топишдаги ёндашувнинг ўзига хослигини кўрсатиш билан бир қаторда, унинг бошқа фанлардан мустақил эканлигини кўрсатиб бермоқда. Қолаверса, тилшуносликка йўналтирилган кўплаб олий ўқув юртларида корпус лингвистикасига ихтисослаштирилган кафедра ва йиллик кесимида ўтказиладиган халқаро конференцияларнинг ташкил этиб келиниши, қолаверса, корпус таҳлилларини қўллашга доир мақолаларни чоп этиб келаётган қатор илмий журналларнинг мавжудлиги ушбу янги тармоқнинг бугунги кундаги аҳамиятини яққол кўрсатиб турибди.

Компьютер лингвистикаси ҳам ўз муаммо ҳамда ютуқлари нуқтаи назаридан ўрганиб чиқилган. А. Н. Баранов (3) ўз ишларида сунъий онг муаммосини очиб беришга тўхталган, Б. Орехов семантик анализ, техник таржима ва автоматик морфология томонидан фанни ўрганиб чиққан бўлса И. А. Чесебиев (7) компьютер лингвистикасининг методлари ривожланиш тарихини ўрганиб чиққан. Шу ўринда, Н. Хомский (6) (Chomsky)нинг синтактик структуралар ва универсал грамматика назарияси фан ривожига катта ҳисса қўшган. Ч. Филлмор (Fillmore) (11) эса ушбу назарияга қўшимча сифатида тиллардаги предлог каби ноўхшашликлар ва уларнинг тил структурасида бошқа тушунчалар орқали ифода этилиши нуқтаи назардан ёндашган. У. Вудс (Woods) (19) фанга процессуал семантика ғоясини киритди. Юқорида номлари келтирилган олимлар ўз ишларида тил ва онг тушунчаларини техникада жамлашни ўз олдларига мақсад қилиб олишган. Зеро, компьютер лингвистикасининг энг долзарб вазифаларидан бири – бу инсон ва компьютер ўртасида автоматик мулоқот интерфейсини яратиш. Бундан келиб чиқадики, ушбу соҳада асосий мақсад тил ўрганувчилари ва изланувчиларини махсус техник қурилма ёки программалар билан таъминлаш.

В. Мамонтова (5) ва Н. Хомский (6) каби бир гуруҳ олимлар корпус лингвистикасини компьютер тилшунослигининг таркибий қисми сифатида баҳолашади. Уларнинг кўпчилиги куйидагича таърифни ўртага суришади: корпус тилшунослиги – компьютер техникаси ёрдамида тилшунослик корпусини қуриш ва улардан фойдаланишнинг умумий тамойилларини ишлаб чиқадиган birlikларни яратиладиган компьютер тилшунослиги. Шу билан

бирга, одатда, компьютер асбоб-ускуналари, дастурлари, корпуснинг 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 тўлдирадилар. 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 нутқнинг 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.

И. А. Чесебиев ўз изланишларида компьютер лингвистикасидан фойдаланишнинг айрим жиҳатларига тўхталиб, хусусан, қуйидагиларни эслатиб ўтади (7):

- 1) автоматик таржима;
- 2) аутентик ва ярим-аутентик матнлардан автоматлаштирилган маълумотларни олиш;
- 3) инсон ва машина ўртасида қулай интерфейсларни яратиш, табиий тилларда алоқанинг *квантитатив* (муаллифдан: Quantitative – (инг. – сонга асосланган) ихтиёрий танланган манба хусусидаги хулосани аниқ рақамлар асосида чиқариш) тавсифи.

Илк корпуснинг яратилиши 1812 йилга бориб тақалади, бунда немис олими Кадлинг ўзининг немис сўзларидаги ундош ҳарфлар дистрибуциясини таҳлил қилган. Ваҳоланки, у даврда ҳали компьютер термининг ўзи ҳам бўлмаган. Кейинчалик, замонавий инглиз тили корпуси намуналаридан бири сифатида Бровн корпуси 1960–1961 йилларда яратилди ва у илк бор босма ҳолда чоп этилди. Орадан бир йил ўтибгина, ушбу корпус электронлаштирилди. Кўриниб турибдики, илк корпуслар компьютер технологияларисиз ҳам мавжуд бўлган ва изланишлар олиб борилган. Кейинчалик фан ривожини қидирув методи(конкорданс)ни электрон форматдаги матнларда ўтказишни тақозо этди ва компьютер лингвистикаси билан соҳалараро муносабатга эҳтиёж туғилди.

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 тутди” (1, 13). А. Н. Баранов “тилни компьютер моделлаштириш тилшуносликдаги дастурлаш назарияси

(компьютер моделларини куриш 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 вазибаларни 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 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. А. Н. Баранов компьютерная лингвистикаси терминига қуйидаги таърифни беради: “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). Олим ўз фикрининг изчиллигини сақлаб қолган ҳолда, бошқа бир асарида шундай уни асослайди, “компьютер лингвистикасида 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 олдидаги 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.

Фойдаланилган адабиётлар

1. Баранов А.Н. Введение в прикладную лингвистику: Учебное пособие. – М.: Эдиториал УРСС, 2001. – 360 с.
2. Баранов А.Н. Компьютерная лингвистика // Введение в прикладную лингвистику: учебное пособие. – М.: Эдиториал УРСС, 2003. – 320 с.
3. Баранов А.Н. Категории искусственного интеллекта в лингвистической семантике. Фреймы и сценарии. М., 1987. – 53 с.
4. Ганиева И.Ф. Об использовании корпусов в лингвистических исследованиях// Филология и искусствоведение. Т.12. – БГУ, Россия, 2007, № 4. С.104–106.

5. 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.
6. Хомской Н. Синтаксические структуры. – США, 1957. – URL:// <https://www.rulit.me/books/sintaksicheskie-struktury-read-504959-1.html>
7. Чесебиев И.А. Компьютерное распознавание и порождение речи / И.А. Чесебиев. – М.: Спорт и культура, 2008 – 128 с.
8. Biber D., Conrad S., Reppen R. Corpus Linguistics: Investigating Language Structure and Use. – Cambridge University Press, 1998. – 300 p.
9. Dash N. S. Corpus Linguistics: A General Introduction. – CIVIL, Mysore, 2010. – 25 p.
10. Meyer C.F. English Corpus Linguistics. – Cambridge: Cambridge University Press, 2002. – URL:// <https://pdfs.semanticscholar.org/c775/4bfab1d0f770e26c0fc7c603e8cf38793be8.pdf>
11. 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.
12. 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. URL:// <https://pdfs.semanticscholar.org/3002/2792e60aefebfda684bf8127c741325491c6.pdf>
13. Gruyter M. de. Geoffrey Th., Hunston S. (eds.). System and corpus: Exploring connections. – London: Equinox, 2006. – URL:// http://sro.sussex.ac.uk/id/eprint/53389/1/what_is_corpus_linguistics.pdf
14. Leech G. Corpora and theories of linguistic performance. In J. Svartvik (ed.), 1992. – P. 105–122.
15. Lindquist H. Corpus Linguistics and the Description of English. – Edinburgh: Edinburgh University Press, 2009. – URL:// https://www.researchgate.net/publication/269836624_Hans_Lindquist_Corpus_Linguistics_and_the_Description_of_English_Edinburgh_University_Press_2009_a_Detailed_Evaluation_In_Corpus_Linguistics_and_the_Description_of_English_Hans_Lindquist_offers_another
16. McEnery T. and Wilson A. Corpus linguistics. – Edinburgh: Edinburgh University Press, 1996. – URL:// <http://clu.uni.no/icame/ij21/meyer.pdf>
17. McEnery T., Xiao R.Z., Yukio T. Corpus-based language studies: An advanced resource book. – London: Routledge, 2005. P. 272-278.
18. Sinclair J. 1991. Corpus and text-basic principles. – Oxford: Oxford University Press. – URL:// https://www.researchgate.net/publication/245802711_Corpus_and_text_-_Basic_principles
19. 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. – P. 300-334.
20. <https://www.myfilology.ru/177/osnovnye-zadachi-i-napravleniya-korpusnoj-lingvistiki-vzaimodejstvie-korpusnoj-lingvistiki-i-kompyuternoj-computational-lingvistiki/>

REFERENCES

1. Baranov A. N. *Kompyuternaya lingvistika* (Computer linguistics), Moscow: Editorial URSS, 2003, 320 p.
2. Baranov A. N. *Vvedenie v prikladnuju lingvistiku* (Introduction to the applied linguistics), Moscow: Jeditorial URSS, 2001, 360 p.
3. Baranov A.N. *Kategorii iskusstvennogo intellekta v lingvisticheskoj semantike. Freymi I tsenarii* (Categories of artificial intelligence in linguistic semantics. Frames and Scripts), Moscow, 1987/

4. Ganieva I.F. *Filologiya i iskusstvoovedeniye*, Vol.12. – БГУ, Russia, 2007, № 4. pp.104–106.
5. Mamontova V. V. *Korpusnaya lingvistika* (Corpus linguistics), candidate's thesis, Stavropol, 2008, 139 p.
6. Chomsky N. *Sintakicheskiye struktury* (Syntactic structures), USA, 1957, available at:// <https://www.rulit.me/books/sintaksicheskie-struktury-read-504959-1.html>
7. Chesebiyev I.A. *Komp'yuternoye raspoznavaniye i porozhdeniye rechi* (Computer recognition and speech generation), Moscow: *Sport i kul'tura*, 2008, 128 p.
8. Biber D., Conrad S., Reppen R. *Corpus Linguistics: Investigating Language Structure and Use*, Cambridge University Press, 1998, 300 p.
9. Dash N. S. *Corpus Linguistics: A General Introduction*, CIVIL, Mysore, 2010, 25 p.
10. Meyer C.F. *English Corpus Linguistics*, Cambridge: Cambridge University Press, 2002, available at:// <https://pdfs.semanticscholar.org/c775/4bfab1d0f770e26c0fc7c603e8cf38793be8.pdf>
11. Fillmore Ch. *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech*, 1976, pp, 20–32.
12. 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, available at:// <https://pdfs.semanticscholar.org/3002/2792e60aefebfda684bf8127c741325491c6.pdf>
13. Gruyter M. de. Geoffrey Th., Hunston S. (eds.). *System and corpus: Exploring connections*, London: Equinox, 2006, available at:// http://sro.sussex.ac.uk/id/eprint/53389/1/what_is_corpus_linguistics.pdf
14. Leech G. *Corpora and theories of linguistic performance*. In J. Svartvik (ed.), 1992, pp. 105–122.
15. Lindquist H. *Corpus Linguistics and the Description of English*, Edinburgh: Edinburgh University Press, 2009, available at:// https://www.researchgate.net/publication/269836624_Hans_Lindquist_Corpus_Linguistics_and_the_Description_of_English_Edinburgh_University_Press_2009_a_Detailed_Evaluation_In_Corpus_Linguistics_and_the_Description_of_English_Hans_Lindquist_offers_another
16. McEnery T. and Wilson A. *Corpus linguistics*, Edinburgh: Edinburgh University Press, 1996, available at:// <http://clu.uni.no/icame/ij21/meyer.pdf>
17. McEnery T., Xiao R.Z., Yukio T. *Corpus-based language studies: An advanced resource book*, London: Routledge, 2005, pp. 272-278.
18. Sinclair J. 1991. *Corpus and text-basic principles*, Oxford: Oxford University Press, available at:// https://www.researchgate.net/publication/245802711_Corpus_and_text_-_Basic_principles
19. 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, pp. 300-334.
20. <https://www.myfilology.ru/177/osnovnye-zadachi-i-napravleniya-korpusnoj-lingvistiki-vzaimodejstvie-korpusnoj-lingvistiki-i-kompyuternoj-computational-lingvistiki/>