

## ЛИСОНИЙ КОРПУСЛАР ТАСНИФИДА ДИАХРОНИК КОРПУСЛАРНИНГ ЎРНИ

Атабоев Нозимжон Бобожон ўғли

филология фанлари бўйича фалсафа доктори (PhD), доцент БухДУ Хорижий тиллар факультети декани

*Аннотация.* Дунёда лисоний тадқиқотлар сифати статистик таҳлилларга асосланар экан, соҳада тил корпусларини яратиш амалиётининг самарали натижаси матн таҳлили жараёнида сарфланадиган вақтнинг тежалишига имкон яратади. Ушбу мақола ўзининг олдига тил корпусларини таснифлашга доир кўплаб илмий қарашларни таҳлил қилиш орқали бу борадаги назарияни ишлаб чиқиш ва таснифлашга оид янги ёндашушни таклиф этишни ммақсад қилиб олган.

*Калит сўзлар:* корпус лингвистикаси, тасниф, корпус, диахроник корпус

## THE PLACE OF DIACHRONIC CORPUS IN THE CLASSIFICATION OF LANGUAGE CORPORA

Ataboyev Nozimjon Bobojon o'g'li

Doctor of philosophy (PhD) in philology, associate professor, Dean of the Faculty of Foreign Languages of BukhSU

*Abstract.* Since the quality of linguistic research in the world is based on statistical analysis, the effective result of the practice of creating language corpora in the field allows saving the time spent in the process of text analysis. This article aims to propose a new approach to theorizing and classifying linguistic corpora by analyzing several scientific approaches to the classification of language corpora.

*Key words:* Corpus linguistics, classification, corpus, diachronic corpus.

## МЕСТО ДИАХРОНИЧЕСКОГО КОРПУСА В КЛАССИФИКАЦИИ ЯЗЫКОВЫХ КОРПОРА

Атабоев Нозимжон Бобожон ўғли

Кандидат наук (PhD) по филологии, доцент, Декан факультета иностранных языков БухГУ

*Аннотация.* Поскольку качество лингвистических исследований в мире основано на статистическом анализе, эффективный результат практики создания языковых корпусов в полевых условиях позволяет сэкономить время, затрачиваемое на процесс анализа текста. Целью данной статьи является предложить новый подход к теоретизированию и классификации языковых корпусов путем анализа нескольких научных подходов к классификации языковых корпусов.

*Ключевые слова:* Корпусная лингвистика, классификация, корпус, диахронический корпус.

*Кириш.* Корпуснинг таснифланиши қайси мақсадларда қўлланилиши ва ундан қандай натижа кутиш мумкинлиги тўғрисидаги назарияларни ҳам кўриб чиқишга ундайди. Зеро, корпус таҳлилларининг аҳамиятини англаш орқали унинг аниқ турларини кўрсатиш бир қадар осонлашади. Н. Владимовнинг фикрича, корпус таҳлилларни ўтказишни ният қилган ҳар қандай тадқиқотчи жараённинг қуйидаги тартибда боришига эътибор бериши зарур [13; С. 26]: а) киритилган сўз ёки сўз бирикмасини корпусда қидирувга бериш; б) алоҳида майдонда сўзнинг семантик майдонини ҳисобга олган ҳолда қидирув натижаларини кўрсатиш; в) сўзларнинг корпусда қўлланилганлик частотасини чиқариш; г) аниқ параметрлар асосида қидирув натижаларини тартиблаш. Лингвистик тадқиқотларда лисоний корпуслардан кўплаб мақсадларда фойдаланишади. Хусусан, С. Вадяев корпус таҳлиллари қуйидаги вазифаларга ечим топишга қаратилганини таъкидлаган [12; С. 28]:

- 1) сўзларнинг маъносини аниқлаш;
- 2) ўзаро зиддиятли сўз формалари ва қурилмаларининг қўлланиши бўйича частоталарни кўрсатиш;
- 3) синтактик моделлар яратиш;
- 4) частотали луғат тузиш;
- 5) стилистик бирликлар тадқиқи.

Бу жараёнларнинг барчаси жуда тез фурсатларда компьютер программалари ёрдамида амалга оширилади. Конкорданса қидирув натижалари қисқа вақт оралиғида кечади. Ушбу тезкорлик корпус фойдаланувчига (асосан таржимонларга) ўзларига керак бўлган сўз ёки иборани тез ва осон топиш имконини беради.

Асосий қисм.

Илк корпус таснифи сифатида корпуснинг автоматлаштирилганлиги ёки йўқлигидан келиб чиқилса, Ч. Мейер Pre-electronic corpora – пре-электрон корпуслар терминини ишлатиб, улар сирасига компьютер эрасигача яратилган корпусларни мисол қилиб киритиш мумкинлигини айтган (Бровн корпусигача яратилган корпуслар) [9]. Бугунги кунда яратилаётган корпуслар эса electronic corpora – электрон корпуслар сифатида таърифланади. Корпусларни тўлиқ-автомат ва ярим-автомат жараёнли турларга бўлиш мумкин, чунки корпуслар таркибига киритиладиган лисоний бирликларни теглаш жараёни бугунги кунда икки усулда ташкил этилади.

Биринчи усулда матн таркиби тўлиқ автоматлаштирилган тартибда тегланади. Масалан Sketch Engine мухитида яратиладиган лисоний корпуслар фақат матнларни киритиш орқали амалга оширилади, бунда тил ва N-грам дастури танланиши етарли. Баъзи корпусларда эса автоматлаштирилган теглаш жараёни тузувчилар томонидан қўлда бошқатдан кўриб чиқилади.

В.П. Захаров очик фойдаланиш имкониятининг йўқлиги боисбаъзи инглиз тили корпусларини «ёпик» корпус сифатида эътироф этади. Бундан келиб чиқадики, корпуслар очик фойдаланишда бўлса, улар «очик» корпус сифатида таснифланиши мумкин [15].

Кейинги талқин М. Davies томонидан таклиф этилган бўлиб, унда корпусларни миқдорий ҳажмига кўра таснифлаш таклиф этилган. Улар сирасига куйидагиларни киритган [5]:

1. Кичик ҳажми: 1-5 миллион сўзли, илк-авлод корпуслари киради. Мисол, Brown Corpus (ва «Бровн туркуми»)га кирувчи LOB, Frown ва FLOB кабилар).

2. Ўрта ҳажми: иккинчи авлод, жанрига кўра тартибга солинган корпуслар киради. Мисол, 100 миллион сўзли British National Corpus.

3. Каттароқ ҳажми: замонавий, жанр жиҳатдан тартибга солинган корпуслар. Масалан, 450 миллион сўзли Corpus of contemporary American English (COCA).

4. Катта матн архиви: Lexis-Nexis.

5. Ўрта катта ҳажми: Google Books.

6. The Web – кўринишидаги бутун қидирув тизимини ўз ичига оладиган корпус каби. Масалан, Google-based searches.

7. Интернет тармоғи ҳажмидаги корпора: Sketch Engine.

8. Google Books учун кенгайтирилган интерфейс: google-books.byu.edu.

Корпусларни ҳажмига кўра таснифлаш ва қарор қилишни нисбий ёндашув сифатида хулосалаш мумкин. Чунки 1996 йилда Ж. Свартвик [11], 2000 йилда Л. Бурнард [3] BNCни жуда катта корпус дея баҳолашади. М. Девиснинг таснифида эса (2015 йил) BNC ўрта ҳажмли корпус сифатида зикр этилган. Бугунги кунда миллиардлаб сўзли матний бирликларни ўз ичига олган корпусларнинг яратилаётгани сир эмас, албатта, бу тасниф ҳам кейинги ўн йилликда яна ўзгаришига ишончимиз комил.

Корпусларни таснифлаш мақсадида уларни махсус принципларга асосланиши зарурлиги сабабли, бир неча олимлар ўз қарашларини куйидагича таъкидлашган:

сўзловчи нуқтаи назаридан: шева, ижтимоий иқтисодий келиб чиқиши, ёш, жинс, таълим ва бошқа факторлар [8];

мулоқот мақсадига кўра: матн тури, функционал фарқланиш, рўйхатнинг шаклланишига кўра ва бошқалар [2];

корпуснинг мақсадига кўра: умумий, махсус, ўрганувчи, педагогик, тарихий, параллел, қиёслаш ва кузатув кабилар [1];

мулоқот турига кўра: сўзлашув, ёзма, компьютер [10].

К. Кеннеди корпусларни куйидагича таснифлайди [7]:

1. Умумий корпора ва махсус корпус (масалан, BNC = British National Corpus ва Bank of English) бутун тилни ёки суб-тилни ифода этади.

2. Тарихий ва замонавий корпуслар (масалан, Helsinki Corpus ва ARCHER) тилнинг тарихий ёки замонавий ҳолатини ифода этишига кўра белгиланади.

3. Худудий корпус ва турли худудларнинг корпуси (масалан, WCNEZ = Wellington Corpus of Written New Zealand English) муайян бир худуднинг тил хусусиятларини ифода этишга қаратилган корпуслар тушунилади.

4. Ўрганувчи корпуси ва тил соҳиби корпуси (масалан, ICLE = International Corpus of Learner English) тил ўрганувчилар ёки тил соҳиблари томонидан қўлланилган тил бирликларининг йиғиндиси; кўп тилли ва бир тилли корпуслар (масалан, COCA – бир тилли) бир ёки бир неча тиллардан ташкил топган корпусларни киритиш мумкин; сўзлашув ва аралаш корпуслар (масалан, LLC = London-Lund Corpus of Spoken English) тилнинг бир ёки бир неча ҳолатини кўрсатиш учун қўлланилади.

И.Ф. Ганиева корпусларни репрезентативлик ва матнлар танловига кўра иккига бўлишни таклиф этади: а) бутун тилни қамраган корпус – тилнинг барча нутқ бирликларини ифода этиш учун яратилади [14]. Мисол, British National Corpus (<http://thetis.bl.uk>), Mannheim corpus of German language (<http://corpora.ids-mannheim.de/~cosmas>), Russian National Corpus (<http://ruscorpora.ru>); б) режалаштирилган корпуслар – жамиятнинг муайян гуруҳини акс эттирувчи матнлар жамланмасига айтилади. Бундай корпуслар махсус мақсадларга кўра йиғилган корпуслар сирасига киради.

Ж. Беннет [1] ҳам юқорида айтиб ўтилганидек, корпусларнинг таснифланишига ўз қарашларини билдириб ўтган, шулардан икkitаси бошқалардан фарқ қилади: 1) ўрганувчилар корпуси – тил ўрганувчилари томонидан қўлланилган лисоний бирликлардан ташкил топган матнлар йиғиндиси; 2) педагоглар корпуси – дарс жараёнини шакллантиришда ўқитувчилар томонидан қўлланиладиган лисоний бирликларни таркибига олган корпуслар тушунилади, улар ўз ичига дарсликлар ҳамда ўқитувчи нутқ бирликларини олиши мумкин.

Н. Даш «Корпус тилшунослиги: умумий кириш» [4] китобида «схема» деб номланган мезонларга мувофиқ корпусга тасниф берган. Бу тасниф бугунги кунда бошқа таснифларга нисбатан аниқроқ шаклда ифода этилган, яъни ушбу схема, имкон қадар аниқ ажратилган тоифалар билан корпусни таснифлашнинг оқилона усулини таклиф қилди. Бундан ташқари, корпус, суб-корпора ва уларнинг таркибий қисмларига таснифлашнинг турли мезонлари қўлланилган. Мисол учун, ташқи ва ички объектлар, жумладан, тил мезонлари қўлланилади. Кўриниб турибдики, ташқи мезонлар асосан корпуснинг объектлари, иштироки, ижтимоий жойлашуви, тилнинг коммуникатив функцияси ва бошқалар билан боғлиқ бўлган матн типологиясидан ажралиб туради. Бу масалаларни ҳисобга олган ҳолда Н. Даш корпусни кенг схемалаштирилган ҳолда таснифлайди, турларни қуйидагича мезонларга кўра беришни таклиф этади: матн жанри, базанинг хусусиятлари, матн тури, дизайн мақсади ва қўллаш хусусиятлари.

## 1.1-жадвал

### Н.Даш таснифи

Корпусларни таснифлаш усуллари					
Матн жанри	Базанинг хусусиятлари	Матн тури	Дизайн мақсади	Қўллашнинг хусусияти	
Сўзлашу в корпуси	Умумий корпус	Бир тилли корпус	Изоҳси з корпус	Таржима корпус	
	Махсус корпус			Қўш корпус	
Нутқ корпуси	Суб тил корпуси	Икки тилли корпус		Изоҳли корпус	Параллелкорпус
	Намунавий корпус		Ҳаволали корпус		
Ёзма корпус	Адабий корпус	Кўп тилли корпус	Изоҳли корпус		Қиёсланадиган корпус
	Мониторкорпус				Кенг имкониятли корпус

С. Вадяев корпусларнинг яратилиш принциплари тўғридан-тўғри уларнинг амалий татбикига боғлиқлигини айтиб ўтади [12]. Шу ўринда унинг бу фикрини асословчи С. Хьюстоннинг таснифини келтириш ўринлидир [6]:

1. Махсус корпус (*specialised corpus*) — фақат бир турдаги матн турларини ўз ичига олган корпуслар тушунилади (масалан, газета материаллари, илмий мақолалар, маъруза матнлари ва ҳ.к.). Унинг мақсади танлаб олинган матн турларининг лингвистик хусусиятларини очишга қаратилган бўлиб, тадқиқотчилар ўзлари изланиш олиб бораётган соҳа тил хусусиятларини аниқлаш учун махсус корпуслар яратишади. Бу типга мансуб таниқли корпуслар сирасига қуйидагиларни киритиш мумкин: Британия дискурс корпуси CANCODE (Cambridge and Nottingham Corpus of Discourse in English), 5 миллион сўз ҳажмидаги норасмий сўзлашув улубига хос матнлардан ташкил топган, Академик услубдаги Инглиз тили Мичиган корпуси MICASE (Michigan Corpus of Academic Spoken English).

2. Умумий корпус (*general corpus*) тилнинг турли услубларидаги матнларни ўз ичига олади. Бу турдаги корпуслар аниқ бир соҳага тегишли муаммога ечим топиш учун корпуслар яратиш эмас, балки турли хилдаги матнларни йиғиш орқали уларни умумлаштириб, тилнинг ифодасини акслантириш учун қўлланилади. Бу турдаги корпуслар сирасига Bank of English (400 млн. с.) ва COCA (Corpus Of Contemporary American English) ларни киритиш мумкин.

3. Қиёсий корпус (*comparable corpus*) ўз ичига икки ёки ундан ортиқ турли тилли ёки турли тил вариантли суб-корпусларни қамраб олади. Бунда ҳар бир суб-корпус ўзаро мутаносиб категориялар асосида шакллантирилиши талаб этилади. Масалан, ICE (International Corpus of English), ҳар бир инглиз тили вариантдан 1 миллион сўзли матнларни ўз таркибига киритган.

4. Параллел корпус (*parallel corpus*) ҳам таркибан бир неча суб-корпуслардан ташкил топади. Фарқли томони унда матнларнинг аниқ иккинчи ёки кейинги тиллардаги таржимаси киритилиши шарт. Параллел корпуслар таржимонлар томонидан кенг қўлланилиб, ўзаро тиллараро тушунчаларнинг эквивалентини топиш имконини беради.

5. Ўрганувчи корпуси (*learner corpus*) ўзида тил ўрганувчилар ва тил соҳиблари томонидан қўлланилган нутқ бирликларини акс эттиради. Ушбу турдаги корпуслар тил ўрганувчилар оқсайдиган ўринларни аниқлаб, тил соҳибларининг нутқ фаолияти билан солиштириш имконини беради. Масалан, Инглиз тилининг халқаро

ўрганувчи корпуси ICLE (International Corpus of Learner English). Ҳар бири 20000 сўздан иборат кичик корпусларни ўз ичига олган инглиз тили ўрганувчилари томонидан ёзилган иншооларни ҳар хил давлатлар кесимида киритиб борилади.

6. Педагог корпуси (pedagogic corpus) – чет тилларни ўқитиш жараёнида қўлланиладиган лисоний бирликларни ўз ичига олган матнлар тўплами. Бундай корпус таркибига ўқув қўлланма ва материаллар кирилиши мумкин.

7. Тарихий / диахроник корпус (historical / diachronic corpus) тил тарихидаги ҳар хил даврларни ўзида акслантирадиган матн бирликларидан ташкил топади. Тиллар ривожини тадқиқотида қўл келади. Масалан, Инглиз тилининг Хельсинкий корпуси (Helsinki Corpus), 700 йилдан то 1700 йиллар оралиғида яратилган матнларни ўз ичига олган (1,5 млн. С.).

8. Кузатув корпуси (monitor corpus) тилда содир бўлаётган ўзгаришларни ўрганиш учун қўлланилади. Ҳар хил вақт оралиғида яратилган лисоний бирликлар корпус таркибига давомли қўшиб борилади.

Н. Даш ва С. Хюстонларнинг таснифлари тўлиқроқ қамров доирасига эга эканлигини кўриш мумкин. Юқорида эслатиб ўтилган корпус таснифлари турли нуқтаи назарларга асосланганлиги сабабли бир корпус айна пайтда бир турга ёки бир неча турга мансуб бўлиши мумкин. Масалан, СОСА (тадқиқот объектимиз) бир тилли, умумий, замонавий, ҳаволали, турли ҳудудларни акс эттирувчи (АҚШ таркибидаги штатлар назарда тутиляпти), бутун тилни қамраган ва ҳ.к. турларга мансуб бўлиши мумкин. Бу эса унга қайси ёндашув асосида қаралишига боғлиқ.

Маълум маънода корпусларни қўллаш мумкин бўлган соҳаларга кўра уларни туркумлаш назарияси шартли равишда якуний шаклга келтирилди ва унда мавжуд терминлардан фойдаланилди. Лисоний корпусларни қўллаш мумкин бўлган соҳалар қаторига қуйидагиларни киритиш мумкин: А) таржима амалиёти ва чоғиштирма лингвистика: махсус/ умумий корпус; қиёсий корпус; бир/кўп тилли корпус; оғзаки/ёзма корпус; Параллел корпус ва ҳ.к. Б) лексикография: махсус/умумий корпус; ҳаволали корпус; оғзаки/ёзма корпус; қўш корпус; бир/кўп тилли корпус ва ҳ.к. В) адабиёт назариясида: параллел корпус; муаллифлик корпуси; адабий корпус ва ҳ.к. Г) лингводидактика ва тилларни ўқитиш: махсус/ умумий корпус; ўрганувчи/ педагог корпуси; кузатув корпуси; оғзаки/ёзма корпус ва ҳ.к. Д) этимология ва тил тарихи тадқиқи: махсус/ умумий корпус; тарихий/замонавий; кузатув корпуси; ҳудудий корпус ва ҳ.к. Бу рўйхатни ҳам фойдаланувчи қўллаши мумкин бўлган барча тил хусусидаги назарий ва амалий тадқиқотларда аҳамиятли бўлган корпуслар ва соҳалар билан қаторда давом эттириш мумкин.

Хулоса. Корпус таснифи буйича олиб борилган ишлар қўлами ва қамрови кенг бўлиб, юқоридаги фикр-мулоҳазаларни рад этмаган ҳолда, корпус турларини якуний қолипга солиб бўлмайди, дея хулосалаш ўринли. Чунки корпус тури корпус тузувчининг прагматик мақсадидан келиб чиқиб шаклланади, унинг дизайн ҳамда структураси ҳам айнан тадқиқот вазифаларидан келиб чиқиб яратилади. Шу аснода янги корпус ўзи билан янги тур ва таснифлаш тамойилини ҳам олиб келиши мумкин. Бу эса мукамал корпус таснифини ишлаб чиқишга тўсқинлик қилади.

#### Фойдаланилган адабиётлар рўйхати

1. Bennett G.R. Using Corpora in the Language Learning Classroom: Corpus Linguistics for Teachers. – Michigan: ELT, 2010. – 134 p. Available at: <http://www.press.umich.edu/titleDetailDesc.do?id=371534>
2. Biber D. Corpus Linguistics: Investigating Language Structure and Use / D. Biber, S. Conrad, R. Reppen. – Cambridge: Cambridge University Press, 1998. – 300 p.
3. Burnard L. Reference Guide for the British National Corpus – World Edition, 2000. – 331 p.
4. Dash N.S. Corpus Linguistics: A General Introduction. – CHL, Mysore. 2010. – 1-11 pp.
5. Davies M. Corpora: An Introduction // English Corpus Linguistics edited by D. Biber and R. Reppen. UK. – 2015. – P. 11.
6. Hunston S. Corpora in Applied Linguistics – Cambridge University Press, 2002. – P 14-16.
7. Kennedy C., Tiziana M. An evaluation of intermediate students' approaches to corpus investigation // Language learning and technology 5(3), – 77-90 pp. URL: <http://www.llt.msu.edu/vol5num3/kennedy/>
8. Labov W. The anatomy of style-shifting. In P. Eckert y J. R. Rickford (Eds), Style and Sociolinguistic Variation. – Cambridge: Cambridge University Press, 2001. – 85-108 pp.
9. Meyer C.F. Pre-Electronic Corpora // Corpus Linguistics: International Handbook. Volume 1. 2008. – P 1.
10. Siekmeyer A. Strukturelle Kategorien des sprachlichen Ausbaus bei Jugendlichen mit Deutsch als Erstsprache und Deutsch als Zweitsprache. – Universität Saarbrücken dissertation. – 2011.
11. Svartvik J. Using Corpora For Language Research edited by Jenny Thomas and Mick Short. – London and New York: Longman, 1996. – P 9.
12. Вадяев С.Е. Лингвистические принципы построения и Использование корпуса текстов для Исследования официально-делового стиля Современного немецкого языка (на материале электронного корпуса «DER») диссер... для кандидата филологических наук. – Самара. 2005. – С 29.
13. Владимов Н.В. Корпусный подход к решению переводческий проблем (на материале английского и русского языках). Дис. ...канд. филол. наук. – Москва: РГБ, 2005. – С 26.
14. Ганиева Й.Ф. «Об использовании корпусов в лингвистических исследованиях» // Филология и искусствоведение. – Россия: БГУ, 2007, Т.12, №4, 104-106 ст.
15. Захаров В. П. Корпусная лингвистика: Учебно-метод. пособие / Захаров В. П. – СПб, 2005. – 48 с. – Режим доступа: <http://download.yandex.ru/class/zakharov/plan.pdf>