

**КЎП ТИЛЛИ ПАРАЛЛЕЛ КОРПУСЛАР ХУСУСИДА**DOI: <https://doi.org/10.53885/edinres.2022.70.62.008>*Жўраева Нилуфар Шухрат қизи**Бухоро давлат университети немис ва француз тили кафедраси ўқитувчиси*

*Аннотация: Жаҳон тилишунослигида компьютер ва корпус лингвистикаси муаммоларини ўрганиши XX асрнинг 40-йилларида бошланди, 60-йилларда мазкур жараён жадаллашди, XXI аср бошларида ўзида миллионлаб сўзларни акс эттирувчи юзлаб тил корпуслари пайдо бўлди. Автоматик таржима, тезаурус, электрон лугатлардан фойдаланиши имконияти кенгайди, илмий-назарий асослари яратилди, амалиётда қўллана бошланди, тил корпусларининг мақсадига кўра фарқ қилувчи кўплаб турлари пайдо бўлди. Бу янгиланишлар ахборот технологияларини тилишуносликка татбиқ этиши билан боғлиқ истиқболли илмий йўналишлар пайдо бўлишига йўл очди. Дунё тилишунослигида XXI асрга келиб, корпус лингвистикасини илмий-назарий жиҳатдан ўрганиши ҳаракати кучайди. Илгарилаб бораётган компьютер лингвистикаси йўналишида автоматик таржима сифатини яхшилаш, таржима дастурларида параллел корпуслардан фойдаланиши таржимашуносликда долзарб масалага айланди. Мазкур мақолада параллел корпус, унинг тузилиши, таркиби ва имконияти; ўзбек тилишунослигида параллел корпус тузиши борасидаги илк тадқиқотлар таҳлил қилинган. Хусусан, параллел корпуслар таркибидаги материал тилига кўра икки ва кўп тилли параллел корпусларга бўлиниши, икки тилли корпуслар таржима ва параллел корпусдан иборат бўлса, кўп тилли корпуслар аслиятдаги матннинг бир неча таржимасидан тузилиши ҳақида фикр юритилган.*

*Калит сўзлар: лингвистик корпус, миллий корпус, параллел корпус, параллел матн (битекст), фрагмент, синтактик бирлик, бир тилли корпус, икки тилли корпус, кўп тилли корпус, таржима корпус, қиёсий корпус, битекст дастур, тескари таржима.*

**О МНОГОЯЗЫЧНЫХ ПАРАЛЛЕЛЬНЫХ КОРПУСАХ***Джурова Нилуфар Шухрат кизи**преподаватель кафедры в немецкого и французского языков в Бухарском государственном университете*

*Аннотация: в данной статье анализируются первые исследования параллельного корпуса, его структура, состав и возможности; формирование параллельного корпуса в узбекском языкознании. В частности, считается, что параллельные корпуса делятся на двуязычные и многоязычные параллельные корпуса в зависимости от языка материала, при этом двуязычные корпуса состоят из перевода и параллельного корпуса, а многоязычные корпуса состоят из нескольких переводов исходного текста. Изучение проблем компьютерной и корпусной лингвистики в мировой лингвистике началось в 1940-х годах, процесс ускорился в 1960-х, а в начале 21 века появились сотни языковых корпусов, представляющих миллионы слов. Расширилось использование автоматического перевода, электронных словарей. Созданы научно-теоретические основы, применяемые на практике, появилось множество типов языковых корпусов, различающихся по своему назначению. Эти инновации привели к появлению перспективных научных направлений, связанных с применением информационных технологий в языкознании. В мире языкознания в XXI веке усилилось движение за научно-теоретическое изучение корпусной лингвистики. Повышение качества автоматического перевода в области передовой компьютерной лингвистики, использование параллельных корпусов в программах перевода стало актуальным вопросом в переводоведении.*

*Ключевые слова: лингвистический корпус, национальный корпус, параллельный корпус,*

параллельный текст (битекст), фрагмент, синтаксическое единство, одноязычный корпус, двуязычный корпус, многоязычный корпус, переводческий корпус, сравнительный корпус, текстовая программа, обратный перевод.

## ON MULTILINGUAL PARALLEL BUILDINGS

*Juraeva Nilufar Shuhrat kizi,*

*Lecturer of the Department of German and French of the Bukhara State University*

*Abstract.: This article analyzes the first studies of the parallel corpus, its structure, composition and capabilities; formation of a parallel corpus in Uzbek linguistics. In particular, parallel corpora are considered to be divided into bilingual and multilingual parallel corpora depending on the language of the material. While bilingual corpora consist of a translation and a parallel corpora, multilingual corpora consist of several translations of the original text. The study of the problems of computer and corpus linguistics in world linguistics began in the 1940s, the process accelerated in the 1960s, and at the beginning of the 21st century, hundreds of language corpora representing millions of words appeared. The use of automatic translation and electronic dictionaries has expanded. Scientific and theoretical foundations that are applied in practice have been created, many types of language corpora have appeared that differ in their purpose. These innovations have led to the emergence of promising scientific areas related to the use of information technology in linguistics. In the world of linguistics in the 21st century, the movement for the scientific and theoretical study of corpus linguistics has intensified. Improving the quality of automatic translation in the field of advanced computational linguistics, the use of parallel corpora in translation programs has become a topical issue in translation studies.*

*Keywords: linguistic corpus, national corpus, parallel corpus, parallel text (bitext), fragment, syntactic unity, monolingual corpus, bilingual corpus, multilingual corpus, translation corpus, comparative corpus, text program, reverse translation.*

Кириш. Параллел корпус (инглизча parallel corpora) – таржима матнларининг электрон аналоглари; унда аслиятдаги матн ва унинг бир ёки бир неча таржимаси тизимли йиғилади.

Корпусдаги электрон матнлар аслият матн ёки унинг бир қисми бўлиши мумкин. Бошқа манбада параллел корпусга қуйидагича таъриф келтирилиди: “Параллел корпус – оригинал матн ва унинг таржимаси йиғилган корпус”. Параллел корпусларнинг назарий асосланиши ва ўрганилганлик даражасини кузатишларимиз шуни кўрсатдики, параллел корпуснинг тузилиши, таркиби ва имконияти ҳақида Д.О.Добровольский, Ю.Гао, В.Захаров, А.А.Кокорева, Е.П.Сосниналар махсус тадқиқотлар олиб боришган. Параллел корпус тузиш назарияси ва амалиёти билан шуғулланган олим – Д.Добровольский параллел корпусни шундай таърифлайди: “Параллел корпус – асл ва таржимадаги электрон матнлар йиғиндисидан иборат корпус. Асл ва таржима матнлар шунчаки ёнма-ён қўйилмайди, балки бу икки матндаги гап(систактик бирлик)лар мазмунан мос келиш даражасида бўлакка бўлиниб, мана шу бирликлар ёнма-ён, бир-бирига ишора қилиши, ҳавола этиш ҳолатида туради. Аслиятдаги матн фрагментига мос келувчи таржимадаги фрагмент белгилаб қўйилган бўлади. Айнан мана шу ҳолат бундай корпуслар ёрдамида турли лингвистик амаллар бажаришга имкон яратади”.

**Адабиётлар таҳлили ва методологияси.** Ўзбек тилшунослигида параллел корпус тузиш борасидаги илк тадқиқотлар Б.Менглиев, Р.Каримов ишлари орасида кузатилади. Улар параллел корпусларни қуйидагича изохлашади: “Корпуслар таркибидаги материал тилига кўра, дастлаб, бир тилли ва кўп тилли корпус турларига бўлинади. Икки ёки кўп тилли корпуслар эса (1) параллел ёки таржима корпуслари; (2) қиёсий корпус каби қўринишга эга”.

Матн ва унинг таржимаси ёнма-ён келадиган электрон тўпламнинг бошқа кўринишлари ҳам мавжуд. Битекст дастури шундай воситалар сирасига киради. Параллел матн (битекст) – бошқа тилдаги таржимаси билан бериладиган матнлар тўплами. Битекстлар «мослаштирувчи асбоб» (alignment tool) ёки «битекст учун мослама» (bitext tool) деб номланувчи махсус компьютер дастурлари орқали амалга оширилади. Бу дастурлар оригинал ва таржима матн мазмунини турли синтактик бирликлар, асосан, содда гап шаклида мослаштириб беради. Битекстлар йиғиндиси битекст маълумотлар базаси ёки икки тилли корпус деб номланади ҳам турли боғланишларни казатиш имконини берувчи маълумот базаси (справочник) вазифасини бажаради.

Мутахассислар тузилиши, матннинг жойлашуви, бирликларининг мослаштирилишига кўра параллел корпуснинг бир неча кўриниши мавжуд:

- 1) бир йўналишли (инглиз тилидан рус тилига таржима қилинган матн);
- 2) икки йўналишли (инглиз тилидан рус тилига таржима қилинган матн ва рус тилидан инглиз тилига (тескари таржима));
- 3) кўп йўналишли (инглиз тилидан рус, немис, француз тилларига таржима қилинган матн).

Корпуслар таркибидаги материал тилига кўра, дастлаб, бир тилли ва кўп тилли корпус турларига бўлинади. Икки ёки кўп тилли корпуслар эса қуйидаги кўринишга эга:

- 1) параллел ёки таржима корпуслари;
- 2) қиёсий корпус.

Параллел корпуслар қандай мақсадни қўзлашидан қатъий назар таркибидаги бирликлар мослаштирилган бўлиши лозим. Мослаштириш юқорида айтилганидек, битекст дастури асосида амалга оширилади.

Ш.Ҳамроева лингвистик корпусларни турли параметрлар асосида таснифлар экан, параллеллигига кўра бир тилли, икки тилли, кўп тилли корпусларни фарқлайди. Шунингдек, параллел корпусларни қуйидагича таърифлайди: “Корпуснинг энг асосий вазифаси тил материалини тўplash, тадқиқотларга асос вазифасини ўташ бўлганлиги сабабли унга тилшуносликнинг қайси аспектида мурожаат қилмайлик, барча саволларимизга жавоб топа оламиз. Табиийки, бу адабиёт таълимига ҳам тегишли. Миллий корпуслар таркибида ички корпус ҳам мавжуд бўлиб, улар алоҳида мақсадга йўналтирилади. Бадиий асарлар корпуси ёрдамида маълум давр ёки бир ижодкор асарлари тил хусусиятларини ўрганиш, параллел корпусдан фойдаланиб асар матннинг қиёсий таҳлилини амалга ошириш мумкин”.

Икки тилли корпуснинг 2 тури мавжуд:

- 1) бир-бирининг таржимаси бўлган матнли корпус;
- 2) бир мавзуга оид икки тилдаги матнли корпус.

**Муҳокама ва натижа.** Биринчи типдаги корпус “параллел корпус” (parallel corpora) деб номланиб, маълум бир таржиманинг турли аспектини ўрганиш учун қўлланилади. Параллел корпус ўз навбатида яна 2 турга бўлинади. Бу мосланган (aligned), мосланмаган (not aligned) корпусдир. “Мосланган” атамаси корпусда таржима бирликлари орасида бир-бирини такозо этувчи аниқ алоқа мавжудлигини билдиради. Бундай корпусдан у ёки бу гапнинг қандай таржима қилингани топилади. Бу турдаги корпус таржимон учун аҳамиятли, чунки унда ноёб ресурс – “таржима хотира”си (translation memory) мавжуд.

Мосланмаган корпус “қиёсий” корпус деб ҳам аталади. Матнни унинг таржимаси билан тўғрилаш бу унинг таржимада қайси бирликка тўғри келишини кўрсатиб туришдан иборат. Мослаш автоматик равишда ёки қўлда бажарилиши мумкин. Биринчи усул осон, лекин хатолари кўп. Масалан, таржима жараёнида содда гап қўшма гап шаклида берилиши мумкин. Бундай пайтда қайси қурилиш оригинал эканлигини аниқлаш қийин бўлади. Кўп тилли “мосланган” корпус намунаси сифатида Европа Иттифоқининг Acquis

Communautaire маълумотлар базасини келтиришимиз мумкин. Параллел корпуснинг қиммати унинг ҳажми, тилларнинг миқдори билан белгиланади. Acquis Communautaire – дунёдаги энг катта параллел корпус. Унинг икки муҳим жиҳати корпусдан фойдаланишнинг бепуллиги, мальти-эстон, словен-фин каби кам учрайдиган тиллар жуфтлигининг мавжудлиги билан баҳоланади.

Мутахассислар ушбу корпусларнинг аҳамиятини қуйидагича белгилашади:

- 1) типик таржима усуллари, трансформацияни юзага келтириш;
- 2) автоматик таржима тизими статистикасини ўрганиш;
- 3) бир ва кўп тилли луғатлар яратиш;
- 4) маълумотни сақлаш, узатиш дастурларини ўрганиш, баҳолаш;
- 5) таржиманинг тўғрилигини автоматик тарзда текшириш;
- 6) эквивалент танлаш имкони кенглиги орқали таржимон меҳнатини осонлаштириш.

Иккинчи хилдаги корпус “таржима корпуслари” (translation corpora) деб аталиб, айни бир фикрнинг турли тилдаги ифодасини ўрганиш учун муҳим. Бундан ташқари, икки тилли корпус тилга оид бир қанча кузатув натижасини (таржима назарияси тадқиқи ёки таржима маҳсулини баҳолаши) бера олганлиги сабабли тилшунослик йўналишини танлаган талаба учун малакавий битирув ишини бажаришнинг жуда фойдали воситаси саналади.

Параллел корпусларни кузатишимиз асосида бир неча параллел корпусларни қамраб олган корпус сифатида The International Corpus of English корпуси ҳақида қуйидаги маълумот топилди: “Халқаро инглиз тили корпуси (The International Corpus of English (ICE)) бир неча ўнлаб университетнинг ҳамкорликдаги лойиҳаси, ҳар бири миллионлаб сўз қўллаш ҳолатини қамраб олган, 20 та параллел ички корпусдан иборат. Инглиз тили давлат тили ёки иккинчи тил ҳисобланган давлатларнинг (Австрия, Канада, Янги Зеландия ва х.) ўзига хос хусусиятини ўрганиш мумкин. Бу корпусни таҳлил қилиш учун махсус мукамал дастурий таъминот ишлаб чиқилган”.

Илк авлод корпуси давридаёқ инглиз, фин, француз, немис, грек, норвег, испан, швед ва бошқа тиллар учун икки тилли корпус пайдо бўла бошлади. Бундай корпус bitexts деб ҳам аталади. Корпусни икки, уч, тўрт, ундан ортиқ тилли қилишга ҳеч қандай тўсиқ йўқ. Соҳанинг бошқа бир мутахассислари эса параллеллик нуктаи назаридан корпусни бир, икки, кўп тилли каби турларга бўлишади. Бир тилли корпусда тил варианты, шевалар бир-бирига қарама-қарши қўйилади. Масалан, инглиз тили она тили ва хорижий тил сифатида алоҳида ўрганилади. Янги технологиялар пайдо бўлгунча инглиз тили хорижий тил сифатида ўрганилмай қолганлиги ҳақида фикрлар мавжуд. Икки, кўп тилли корпус бир мавзу доирасида турли тилда ёзилган матндан ташкил топади.

**Хулоса.** Хулоса сифатида айтиш жоизки, параллел корпуслар таркибидаги материал тилига кўра икки ва кўп тилли параллел корпусларга бўлинади. Икки тилли корпуслар таржима ва параллел корпусдан иборат бўлса, кўп тилли корпуслар аслиятдаги матннинг бир неча таржимасидан тузилади. Параллел корпус тузиш учун махсус битекст дастурлар талаб қилинади.

#### Фойдаланилган адабиётлар:

1. Соснина Е.П. Параллельные корпусы в обучении языку и переводу. – (Электрон ресурс): [http://ling.ulstu.ru/linguistics/resources/literature/articles/corpus\\_education\\_translation/](http://ling.ulstu.ru/linguistics/resources/literature/articles/corpus_education_translation/)

2. Добровольский Д. Какие задачи решают параллельные корпусы? (Электрон ресурс): <https://postnauka.ru/video/>; Добровольский О.Д., Кретов А.А., Шаров С.А. Корпус параллельных текстов: архитектура и возможности использования // Национальный корпус русского языка: 2003-2005. М.: Индрик, 2005. – С. 263-296.

3. Жўраева Н. Evfemizm lingvokulturologik fenomen sifatida //центр научных

- публикаций (buxdu.uz). – 2021. – т. 8. – №. 8.
4. Жўраева Н. Evfemizm, yohud yumshoq muomala fransuz ayollari talqinida //центр научных публикаций (buxdu.uz). – 2020. – т. 6. – №. 2.
5. Жўраева Н. Evfemizmlarning gender xususiyatlari xususida //центр научных публикаций (buxdu.uz). – 2021. – т. 8. – №. 8.
6. Жўраева Н. Types des niveaux du langage dans les pays francophones //центр научных публикаций (buxdu.uz). – 2021. – т. 8. – №. 8.
7. Жўраева Н. Zamonaviy fransuz tilshunosligida evfemizm va tabu hodisalari to ‘g’risidagi nazariy qarashlar //центр научных публикаций (buxdu.uz). – 2021. – т. 8. – №. 8.
8. Жўраева Н. TYPES DES NIVEAUX DU LANGAGE DANS LES PAYS FRANCOPHONES //ЦЕНТР НАУЧНЫХ ПУБЛИКАЦИЙ (buxdu.uz). – 2021. – т. 8. – №. 8.
9. Жўраева Н. LES PARTICULARITÉS DE L’ENSEIGNEMENT DES LANGUES ÉTRANGÈRES //ЦЕНТР НАУЧНЫХ ПУБЛИКАЦИЙ (buxdu.uz). – 2021. – т. 8. – №. 8.
10. Жўраева Н. СНЕТ ТИЛИНИ О’ҚИТИШДА ЗАМОНАВИЙ YONDASHUVLAR // ЦЕНТР НАУЧНЫХ ПУБЛИКАЦИЙ (buxdu.uz). – 2021. – т. 8. – №. 8.
11. Кокорева А.А. Корпус параллельных текстов в обучении иностранному языку. – Вестник ТГУ (Тамбовский государственный университет), выпуск 2 (118), 2013. – С. 57-62.
12. Жураева М. MEDIA RELATIONS IN FRENCH DISCOURSE //ЦЕНТР НАУЧНЫХ ПУБЛИКАЦИЙ (buxdu.uz). – 2021. – т. 8. – №. 8.
13. Жураева М. ЁЗМА МАТБУОТ ДИСКУРСИНИНГ ФУНКЦИОНАЛ ХУСУСИЯТЛАРИ //ЦЕНТР НАУЧНЫХ ПУБЛИКАЦИЙ (buxdu.uz). – 2021. – т. 8. – №. 8.
14. Тао Ю., Захаров В.П. Разработка и использование параллельного корпуса русского и китайского языков // Автоматизация обработки текста. (Электрон ресурс): <https://www.researchgate.net/publication/306911853>
15. Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: филол. фанлари бўйича фалсафа доктори диссертацияси. – Бухоро, 2018. – Б. 29.
16. Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: филол. фанлари бўйича фалсафа доктори диссертацияси. – Бухоро, 2018. – Б. 39, 52.
17. BOBOKALONOV R. O. INTERNATIONAL MEDICAL TERMS FROM FRENCH TO UZBEK LANGUAGE //E-Conference Globe. – 2021. – С. 136-144.
18. Bobokalonov R., Mukumov L. Ways to Prevent Conflicts in the Hotel Business //» ONLINE-CONFERENCES» PLATFORM. – 2021. – С. 268-272.
19. Jo’rayeva N. S. Evfemizmlarning lingvomadaniy xususiyatlari //экономика и социум. – 2021. – №. 1-1. – с. 107-110.
20. Mengliev B., Karimov R. The Role of the Parallel Corpus in Linguistics, the Importance and the Possibilities of Interpretation International Journal of Engineering and Advanced Technology (IJEAT) –ISSN: 2249 – 8958, Volume-8, Issue-5S3 July 2019. – P. 388-391.
21. Mengliev B., Karimov R. Theoretical fundamentals of uzbek-english parallel corpus / JOURNAL OF CRITICAL REVIEWS – ISSN- 2394-5125 VOL 7, ISSUE 17, 2020.
22. Mukhammadovna J. M. Media relations in french discourse //Middle European Scientific Bulletin. – 2021. – т. 12. – С. 178-180.
23. Rajab B. English Explanation Of Medical Hybrid Terms With French Prefix» Anti» // International Journal of Progressive Sciences and Technologies. – 2021. – т. 25. – №. 1. – С. 352-356.
24. Rajab B., Savriya K. Structural Analysis of Medical Terms //» ONLINE-CONFERENCES» PLATFORM. – 2021. – С. 94-98.
25. Shukhratovna J. N. udc: 81.373. 49 316.754. 6 The main areas (spheres) of euphemisms and taboos in french linguistics //scientific reports of Bukhara State University. – с. 134.
26. Shavkatovna S. D., Ostonovich B. R., Isroilovna T. G. LINGUOCULTUROLOGICAL ANALYSIS OF IMAGES-SYMBOLS OF LITERARY TEXT //International Engineering Journal For Research & Development. – 2021. – т. 6. – №. TITFL. – С. 236-238.