

ПАРАЛЛЕЛ МАТНЛАР – ТАРЖИМА МУҚОБИЛЛАРИНИ ТАҲЛИЛ ҚИЛИШДА ЗАМОНАВИЙ ВОСИТА

Анарбаев Орзубек Рахманович
ЖДПИ ўқитувчиси
orzubekananarbaev@gmail.com
995508708

Аннотация. Ушбу мақолада параллел матнлар корпусига берилган таърифлар, уларнинг ўзига хос хусусиятлари ва унинг ҳозирги кун нуқтаи назаридан долзарблиги ҳақида сўз юритилган, яъни параллел матнлар корпуси автоматик таржима ривожини учун ҳам хизмат қилади, компьютер лексикографиясининг тараққиётини таъминлайди ва ёрдамида махсус конкордансер дастурлар ишлаб чиқилади ва турли хил ихтисослик луғатлари тузиш имконияти юзага келади. Шунингдек, параллел матнларнинг пайдо бўлиш сабаблари ва унга тегишли муаммо, ютуқ ва камчиликлар ҳақида мутахассислар томонидан олиб борилган изланиш натижалари келтирилган.

Калим сўзлар: параллел матнларнинг корпуси, таржима корпуси, *bitext*, тенглаштириши, параллел матн, таққосланадиган корпуслар, икки тилли, таржима хотираси, қиёсий корпус.

Annotation. This article discusses the definitions of parallel text corpus, their specific features and its relevance in today's context, that is, the parallel text corpus also serves for the development of automatic translation, provides the development of computer lexicography and special concordance programs are developed and used in various specialties and there is an opportunity to create dictionaries. It also presents the results of research conducted by experts on the causes of the emergence of parallel texts and the problems, successes and shortcomings associated with it.

Keywords: *corpus of parallel texts, translation corpus, bitext, alignment, parallel text, comparable corpora, bilingual, translation memory, comparative corpora.*

Аннотация. В данной статье обсуждаются определения параллельного корпуса текстов, их особенности и актуальность в современном контексте, то есть корпус параллельных текстов также служит для развития автоматического перевода, обеспечивает разработку компьютерной лексикографии, а также разрабатываются и используются специальные программы конкорденс в различных областях и есть возможность создавать словари. Также представлены результаты проведенного экспертами исследования причин появления параллельных текстов и связанных с этим проблем, успехов и недостатков.

Ключевые слова: *корпус параллельных текстов, корпус перевода, битекст, выравнивание, параллельный текст, сопоставимое корпус, двуязычный, память переводов, сравнительное корпус.*

Корпус тилшунослиги лингвистик тадқиқотнинг фаол ривожланаётган соҳаси бўлиб, тилшуносликнинг бошқа бўлимлари учун жуда катта миқдордаги эмпирик материалдан фойдаланиш имкониятини беради. Корпус лингвистикасида параллел матнлар корпуси муҳим аҳамият касб этади. Шу билан бирга тил соҳаларида кўп тиллиликнинг аҳамияти тобора ортиб бораётганлиги сабабли, глобал бозорлар ва бутун дунё бўйлаб маълумот алмашинуви билан параллел матн корпусларидан

фойдаланиш истиқболли технологиялардан бири бўлиб, унинг келажаги жуда порлоқ ҳисобланади.

Параллел корпуслар ёрдамида бир сўз, жумла, абзац, суперсинтактик бутунликларнинг турли тиллардаги вариантларини билиш мумкин. Параллел корпуслар маданиятлараро мулоқот кенг ёйилган ҳозирги давр учун муҳим воқелик ҳисобланади. Параллел корпуслар орқали турли тил муҳитидаги, маданиятларидаги универсалиялар ҳамда тилларнинг ўзига хос ментал хусусиятлари, реалия ва лакунар бирликларни аниқлаш мумкин бўлади. Параллел матнлар корпуси автоматик таржима ривож учун ҳам хизмат қилади, компьютер лексикографиясининг тараққиётини таъминлайди. Параллел матнлар корпуси ёрдамида махсус конкордансер дастурлар ишлаб чиқилади ва турли хил ихтисослик луғатлари тузиш имконияти юзага келади.

Параллел корпуслар кўпинча Бирлашган Миллатлар Ташкилоти, Европа Иттифоқи каби кўп тилли жамоаларда ва Канада каби расмий икки тилли мамлакатларда ишлатиладиган матнлардан яратилади.

Параллел корпуснинг асосий камчилиги шундаки, матнлар манба тили таъсиридан, таржима хатоларидан ва таржимоннинг индивидуал услубидан холи эмас. Номутаносибликнинг ўхшаш муаммолари бир тилли корпусларда ҳам йўқ эмас. Ҳар қандай ҳолатда ҳам ечим битта: корпус қанчалик катта ва хилма-хил бўлса, алоҳида матннинг таъсири шунчалик кам бўлади.

Параллел матнлар корпусининг илк намунаси сифатида 1799 йил Розетта шаҳри яқинидаги Нил дельтасида топилган милоддан аввалги 196 йилга бориб тақаладиган, Миср ибодатхоналари томонидан икки тилда (юнон ва миср) ва учта ёзув тизимида (Миср иероглифлари ва демотик тилда ёзилган) шоҳ Птоломей Vга берилган шарафлар ҳақида сўз юритилган тош ҳисобланади¹.

Д.О.Добровольскийнинг фикрича “*параллел матнларнинг корпуси* нафақат таржима назарияси, икки тилли лексикография ва тилни таққослаш билан боғлиқ бошқа тил соҳаларида, балки ҳар бир таққосланадиган тилларнинг лексик бирликларининг семантикасини янада батафсил тавсифлаш учун фойдалидир”².

Ҳозирги кунда параллел матнлар корпусига қуйидагича таърифлар берилмоқдаки, уларнинг барчаси параллел корпуснинг ўзига хос хусусиятларини намоён қилади:

- *Параллел матнлар корпуси* бадий асар, қўлланма, оммавий ахборот воситалари, турли хил ҳужжатларнинг икки ёки ундан ортиқ тиллардаги электрон кўринишларидир.

- *Параллел корпус* – бу корпуснинг ўзига хос тури бўлиб, унда бирор тилидаги матн ушбу матнни бошқа тилга таржимасини таққослайди ёки аксинча, чет тилидаги матн унинг ўша тилидаги таржимасини таққослайди. Тенглаштириш деб номланган амалиёт ёрдамида асл ва таржима қилинган матн қисмлари ўртасида тенглаштириш ўрнатилади. Тенглаштирилган параллел корпус илмий тадқиқотлар учун самарали воситадир, жумладан таржима назарияси ва амалиёти учун ҳам.

- *Параллел матнлар корпуси* илмий ва амалий мақсадларда (хусусан, чет тилларини ўқитиш учун) шакллантирилади. Ўзининг тузилиши бўйича бир ёки бир

¹J. Veronis (ed.), *Parallel Text Processing* Kluwer Academic Publishers., 2000. p. 1.

²Добровольский, Д.О. Корпус параллельных текстов в исследовании лексической семантики [Текст] / Д.О. Добровольский // Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог 2004». - М.: Наука, 2004а. - ст. 79.

нечта матнларнинг манба тилидаги матнларнинг мақсадли тилларга таржимаси ҳисобланиб, манба тилидаги матнларнинг тўпламидир.

○ *Параллел корпус* ёки *таржима корпуси* (*ingl. parallel yoki translation corpus*) - бу битта тилдаги матнлардан ва бошқа тилга ёки тилларга таржимасидан иборат бўлган корпус. Бундай параллел жуфтликлар ёки қаторлар битехт (*ingl. bitext*) деб номланади. Параллел корпусни яратиш учун унинг асл нусхаси ва унинг таржимасининг бўлиши етарли ҳисобланмайди. Асл нусхага мос келадиган парчани таржимада топилишини таъминлаш керак бўлади. Шу мақсадда тенглаштириш (*ingl. alignment*) жараёнидан фойдаланилади, натижада параллел матнларнинг бир хил қисмлари бир-бири билан таққосланади.

Матнларнинг параллел корпуси таржима изланишлари учун энг мос келадигани бўлиб, у таржима муқобилларини таҳлил қилиш учун замонавий восита сифатида тадқиқотлар ўтказишда юқори репрезентативлик ва ишончликка эришишга имкон беради.

Корпус мутахассиси (асосан, таржимон)ни доим бир неча тилли корпус яратиш қизиқтириб келган. 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.

Bitext атамаси дастлаб 1988 йил Брайн Харрис томонидан таржима тадқиқотларида фойдаланиш учун ҳужжатларни ва уларнинг бошқа тилларга таржималарини назарда тутиш учун ишлаб чиқилган³. Ўшандан бери *bitext* кўплаб бошқа дастурларни ҳисобга олган ҳолда кенг жамоатчиликда катта қизиқиш уйғотди. Ҳозирги вақтда *bitext* атамасини нафақат асл ҳужжатлар ва уларнинг тўғридан-тўғри таржималари, балки кўпроқ параллел манбаларга мурожаат қилиш учун ишлатиш одатий ҳолдир. Бундай манбалар умумий манбадан турли хил таржималарни ўз ичига олиши мумкин, баъзида ҳатто охириги таржималарни яратиш учун ишлатилган бошқа оралик тиллардан ҳам фойдаланиш мумкин.

Компютер тилшунослигида *параллел матн* атамаси кўпинча *bitext* термини билан синоним сифатида ишлатилади. Афсуски, бу Жан Веронис (2010) таъкидлаганидек таржима назарияси ва терминологияси билан чалкашликлар келтириб чиқаради. Таржима тадқиқотларида *параллел матн* атамаси одатда бир хил соҳадаги матнлар тўпламига ишора қилади, лекин бир хил ҳужжатларнинг таржималарини ўз ичига олмайди. Бироқ, компютер лингвистикасида бундай корпуслар одатда (икки тилли) (*bilingual*) таққосланадиган корпуслар (*comparablecorpora*) деб юритилади⁴.

Компютер тилшунослигида *bitext*лар бўйича бир неча ўн йиллик изланишлардан сўнг, ҳозирги вақтда *bitext*лар тўпламини номлаш учун *параллел корпус* атамаси яратилган⁵.

Соҳанинг мутахассислари параллеллик мезонига кўра корпуслар бир тилли, икки тилли ва кўп тилли турларига бўлишади. Бир тилли корпусларда лаҳжалар, тилининг вариантлари таққосланади.

Икки, кўп тилли корпуснинг яна бир кўринишида оригинал матн ҳамда таржима матни мавжуд бўлади. Корпуснинг ушбу тури қиёсий чоғиштирма

³Brian Harris. Bi-Text, a new concept in translation theory. Language Monthly, 54:8–10, 1988.

⁴Reinhard Rapp. Identifying word translations in non-parallel texts. In Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics (ACL), 1995. DOI: 10.3115/981658.981709 Cited on page(s) 1, 34

⁵Jörg Tiedemann. Bitext Alignment. Morgan & Claypool Publishers series. 2011. p. 16.

тадқиқот олиб боришда, таржима назарияси ҳамда компьютер таржимасини ўрганишда муҳим, фойдали манба вазифасини бажаради.

Икки тилли ва кўп тилли корпусларни иккита асосий турга бўлиш мумкин:

1) ҳар қандай манба тилида ёзилган кўплаб асл матнларни ва ушбу асл матнларнинг бир ёки бир нечта бошқа тилларга таржималарини акс эттирувчи корпуслар;

2) бир хил мавзудаги матнларни бирлаштирган, мустақил равишда икки ёки ундан ортиқ тилда ёзилган корпуслар;

Бундай корпуслар терминологияда ёрдам беради ва кўпинча таржимонлар томонидан қўлланилади.

Ҳар икки турдаги корпуслар ҳам таржима қилишнинг самарали усулларини ишлаб чиқиш, шу жумладан машинада таржима қилиш учун икки тилли ва кўп тилли терминологик луғатларни тузиш учун, шунингдек, тилларни қиёсий тадқиқ қилиш учун (лексикология, грамматика, стилистика, таржимашунослик ва бошқалар соҳасида) ишлатилади.

Биринчи типдаги корпус — параллел корпус (*parallel corpora*) деб номланиб, маълум бир таржиманинг турли аспектини ўрганиш учун қўлланилади. Бундай корпуслар сирасига Канада парламенти йиғини (инглиз/француз) матнлари корпусини киритиш мумкин. Параллел корпус ўз навбатида яна 2 турга бўлинади. Бу тенгланган (*aligned*), тенгланмаган (*not aligned*) корпусдир. *Тенгланган* атамаси корпусда таржима бирликлари орасида бир-бирини тақозо этувчи аниқ алоқа мавжудлигини билдиради. Бундай корпусдан у ёки бу гапнинг қандай таржима қилинганини дарров топамиз. Бу турдаги корпус таржимон учун аҳамиятли, чунки унда ноёб ресурс — таржима хотираси (*translation memory*) мавжуд бўлади.

Тенгланмаган корпус — *қиёсий корпус* деб ҳам аталади. Матнни унинг таржимаси билан тўғрилаш бу унинг таржимада қайси бирликка тўғри келишини кўрсатиб туришдан иборат. Тенглаш 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 — тенгланган корпус намунаси сифатида Европа Иттифоқининг *Acquis Communautaire* маълумотлар базасини келтиришимиз мумкин. Параллел корпуснинг қиймати унинг ҳажми, тилларнинг миқдори билан белгиланади. *Acquis Communautaire* – дунёдаги энг катта параллел корпус ҳисобланади. Унинг икки муҳим жиҳати корпусдан фойдаланишнинг бепуллиги, мальти-эстон, словен-фин каби кам учрайдиган тиллар жуфтлигининг мавжудлиги билан баҳоланади.

Тенглаштириш (*alignment*) - асл матн парчалари ва таржима матни ўртасида тенглаштириш муаммосини ҳал қилиш учун матнни автоматик равишда тенглаштиришнинг турли усуллари - гаплар, клаузлар (грамматик тузилмалар), иборалар ва сўзларга мувофиқ қўлланилади.

Гаплар усулида тенглаштиришда А.В. Зубов ва И.И. Зубоваларнинг дарслигида кўрсатилган иккала матн жумлалари орасидаги олтига амалга ошиш имкони бўлган мувофиқлик қўлланилиши мумкин⁶:

- 1) битта асл гап битта жумлага таржима қилинади;
- 2) иккита асл жумла битта жумла билан таржима қилинади;
- 3) битта асл жумла икки жумла билан таржима қилинади;

⁶А. В. Зубов, И. И. Зубова. Информационные технологии в лингвистике: Учеб.пособие для студентовлингв., фак-тов высш. учеб. ЗаведенийМ.: Издательскийцентр «Академия», 2004. - 208 с.

4) иккита асл жумла икки жумла билан таржима қилинади, аммо асл жумла ва таржима қилинган матндаги ушбу жумлаларнинг ички чегаралари мос келмайди;

5) асл матннинг жумласи таржима қилинмайди;

6) таржима матнидаги жумланинг асл матнида эквиваленти бўлмайди.

Амалда, жумлаларнинг нисбий узунликларига мос келиш, матнни параграфларга ажратиш, тиниш белгиларини, ташқи сўз бойлигини ва бошқа омилларни таҳлил қилиш асосида матнларни автоматик равишда тенглаштирадиган турли хил тенглаштириш дастурлари мавжуд. Кўпинча, ушбу дастурлар автоматик равишда тенглаштириш натижаларини таҳрирлашдан сўнг, инсон томонидан ишлаб чиқарилган машинада қўлланилади.

Иккинчи хилдаги корпус — таржима корпуслари (*translation corpora*) деб аталиб, айна 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 ҳамдир.

Параллел корпусларни, шу жумладан иккинчи турдаги корпусларни ўрганишда сўзларни автоматик таснифлаш воситаларидан муваффақиятли фойдаланиш мумкин.

Биринчи турдаги параллел корпусларда эса асл ва таржима матнларини параллел қайта ишлаш жараёнида олинган кластер элементларининг нисбати тўғрисидаги маълумотлар таржиманинг етарлигини аниқлашда ва чоғиштира тадқиқотлар ўтказишда катта аҳамиятга эга.

Сўзларни автоматик таснифлаш модулларидан фойдаланиш параллел корпусларда қидирув самарадорлигини оширади, кўп тилли луғатларни тўлдириш ва тузатиш учун маълумот олиш, таржима машиналарининг тизимлари ва уларни ўқитишнинг сифатини текшириш имконини беради.

Тилларни автоматик қайта ишлашда параллел матнларни мунтазам равишда қўллаш 1980-йилларда бошланди. Уларни машина таржимаси учун ишлатишга бир неча бор уринишлар 1950-йилларнинг охирларида қилинган эди, аммо бу йиллар давомида компьютерларнинг чекланган хотираси ва қайта ишлаш қуввати, катта миқдордаги матнли маълумотларни киритиш қийинлиги ўша пайтда корпусдан чекланган фойдаланиш сабаби ҳисобланади.

Алан Мелбий фикрига кўра, намунавий таржималарни кейинчалик қайта ишлатиш учун сақлаш ғояси етмишинчи йилларнинг охирларидатурли хил тадқиқот марказларида мустақил равишда пайдо бўлган. Матнни автоматик равишда параллел тенглаштиришнинг биринчи усули 1987 йилда Мартин Кай ва Мартин Росчеисен (Кай & Росчеисен, 1988) лар томонидан ишлаб чиқилган ва бундан кўп ўтмай, турли даражадаги матнларни тенглаштириш учун жуда кўп усуллар таклиф қилинган (яъни бир-бирини таржима қиладиган бирлик таққослашлари). Кўриб чиқилган бирликларга параграфлар, жумлалар, сўзлар ва иборалар қиради.⁷

Параллел матнларнинг қўлланилиши жуда хилма-хилдир ва таржима хотираларини тузиш, луғатлар ва икки тилли терминологиялар рўйхатларини яратиш, тиллараро маълумотларни қидириш учун билимларни олиш ва компьютер ёрдамида ўқитиш ёки қиёсий тилшунослик учун имтихон маълумотларини олиш ва ва ҳақозаларни ўз ичига олади. Қисман фирмаларда электрон ҳужжатларни сақлаш

⁷J. Veronis (ed.), *Parallel Text Processing* Kluwer Academic Publishers., 2000. p.2.

ва архивлашнинг яхшиланганлиги ва бир томондан кўп тилли ҳужжатларни етказиб берувчиси, тобора ўсиб бораётган Интернет тармоғи сабабли параллел матнларнинг сони кўпайиб бормокда⁸.

Таржима муаммоларини ўрганишда корпусдан фойдаланиш турли хил усуллар билан амалга оширилиши мумкин. Улардан бири Н.В. Владимовнинг диссертациясида тасвирланган⁹. Чет тилига таржима қилишнинг фундаментал имконияти тезисига асосланиб, у энг мақбул таржима вариантларини топиш муаммосини "субъектнинг (таржимоннинг) тил саводхонлигининг юқори даражасини таъминлаш ёки корпус ёндашувидан фойдаланган ҳолда" ҳал қилиш мумкин деб ҳисоблайди. Корпус ёндашуви билан у "матнларнинг электрон аннотацияланган корпуси" дан фойдаланган ҳолда чет тилига таржима қилиш жараёнида юзага келадиган қийинчиликлар ечимини тушунади.

Муаллиф чет тилига таржима қилиш жараёнида юзага келадиган таржима муаммоларини ҳал қилишда қуйидаги кетма-кетликни таклиф этади:

- 1) муаммони таҳлил қилиш;
- 2) таржима вариантини яратиш;
- 3) матнли корпусда ушбу имкониятнинг мавжудлигини текшириш;
- 4) энг мос вариантни аниқлаш учун "тест - ҳаракат - синов - чиқиш" амалиёларини даврий бажариш.

Н.В. Владимов томонидан ўтказилган тажриба қуйидагича эди. Экспериментал таржимонлар гуруҳи (юқори курс талабалари ва эндигина битирганлар), уларнинг ихтиёрида компьютер, электрон ва оддий китоб луғатлари рус тилидан инглиз тилига таржима қилинган. Иккинчи босқичда, таржималардаги хато ва камчиликларни кўрсатиш учун эксперимент иштирокчилари билан услубий-кўрсатма дарси ўтказилган ва кейинчалик Бритиш Национал Корпус (*BritishNationalCorpus*) ёрдамида бу хатолар қандай бартараф этилиши кўрсатилган. Муҳокама давомида шарҳлар ёзиб олинган, сўнг баённомалар тузилган. Шундан сўнг, матнлар ушбу корпус ёрдамида таҳрир қилинган. Охири вариантлар эксперт баҳосини олган. Якуний босқичда эксперимент ҳужжатлари (дастлабки таржима вариантлари, таҳрирланган таржима, ўзгартириш киритилган баённомалари ва эксперт баҳолари) доимлилигини аниқлаш мақсадида таҳлил қилинган. Белгиланган қийинчиликлар орасида (камайиш тартибида): лексик ва грамматик мувофиқлик, тиниш белгилари, грамматик қурилишни танлаш, имло, артикль ва бошқалар келтирилган.¹⁰

Муаллиф хулосага кўра, корпусдан фойдаланиш натижасида таржиманинг сифати сезиларли даражада яхшиланганлигини таъкидлайди, аммо корпусдан фойдаланганда ҳам унинг сифатини ошириш учун асос ўзини ўзи бошқариш ва таржимоннинг таклиф этилаётган вариантларга нисбатан танқидий муносабати ҳисобланади. Мавзуни ҳаддан ташқари кенг ёритилиши ва камчиликлар сифатида қайд этилган сўнгги ўзгартиришлар ва қўшимчаларнинг йўқлиги танқидга қарши турмайди, чунки улар корпуснинг хусусий версияси — BNCsampleversion (БМК) намунавий версиясига тегишлидир. Бундан ташқари, қандай қилиб Н.В. Владимов инглиз тилидаги таржима вариантларини таҳлил қилишни таклиф қилиши, ноаниқ

⁸J. Veronis (ed.), *Parallel Text Processing* Kluwer Academic Publishers., 2000. p.3.

⁹Мамонтова В. В. Особенности перевода сложносоставных слов с английского языка на русский (на материале корпуса публицистических текстов) Диссертация на соискание ученой степени кандидата филологических наук. Ставрополь — 2008. ст. 60.

¹⁰ibid

бўлиб қолмоқда. Интернет-версиясидан фойдаланиш, муаллифнинг сўзларига кўра, фойдаланувчи билан боғлиқ бўлган корпус ости (*subcorpus*)ни ўрнатишга, шунингдек, сўнгги ўзгаришларни кузатишга имкон беради¹¹.

Умуман олганда, параллел матн корпуслари катта ҳажмдаги маълумотларни тақдим этади ва ушбу корпуслардан қуйидаги мақсадларда фойдаланиш мумкин бўлади:

- типик таржима усуллари, трансформацияни юзага келтириш;
- автоматик таржима тизими статистикасини ўрганиш;
- автомат таржима тизимлари учун луғатларни яратиш ва уларни тўлдириб бориш;
- бир, икки ва кўп тилли таржима луғатларини яратиш;
- маълумотни сақлаш, узатиш дастурларини ўрганиш, баҳолаш;
- таржиманинг тўғрилигини автоматик тарзда текшириш;
- эквивалент танлаш имкони кенглиги орқали таржимон меҳнатини осонлаштириш.
- гап узунликдагидан ошиб кетган полисемантик сўз контекстини компьютер таҳлили орқали лексик birlikларнинг полисемиясини йўқ қилиш;
- матннинг терминологик ва фразеологик birlikларини таржима қилиш;
- компьютер хотирасида манба матнлари ва уларнинг таржималарининг турли даражаларда турғунлигини тўплаб, таржима хотираси тизимлари деб номланган янги автомат таржима тизимлари доирасида тўлиқ автоматик таржимани амалга ошириш.

Фойдаланилган манбалар:

1. Brian Harris. Bi-Text, a new concept in translation theory. *Language Monthly*, 54:8–10, 1988.
2. J. Veronis (ed.), *Parallel Text Processing* Kluwer Academic Publishers., 2000.
3. Jörg Tiedemann. *Bitext Alignment*. Morgan & Claypool Publishers series. 2011.
- Reinhard Rapp. Identifying word translations in non-parallel texts. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics (ACL)*, 1995. DOI: 10.3115/981658.981709
4. А. В. Зубов, И. И. Зубова. Информационные технологии в лингвистике: Учеб. пособие для студентов лингв., фак-тов высш. учеб. Заведений М.: Издательский центр «Академия», 2004. - 208 с.
5. Добровольский, Д.О. Корпус параллельных текстов в исследовании лексической семантики [Текст] / Д.О. Добровольский // *Компьютерная лингвистика и интеллектуальные технологии: Труды междунар. конференции «Диалог 2004»*. - М.: Наука, 2004а.
6. Мамонтова В. В. Особенности перевода сложносоставных слов с английского языка на русский (на материале корпуса публицистических текстов) Диссертация на соискание ученой степени кандидата филологических наук. Ставрополь — 2008.

¹¹Мамонтова В. В. Особенности перевода сложносоставных слов с английского языка на русский (на материале корпуса публицистических текстов) Диссертация на соискание ученой степени кандидата филологических наук. Ставрополь — 2008. ст. 62.