



Dildor O‘TAJONOVA,
Chirchiq davlat pedagogika universiteti dots. v/b., PhD
E-mail: dildorotajonova201@gmail.com

THEORETICAL BASIS OF CORPUS FORMATION

Annotation

This article is devoted to the study the formation of corpus and reasons of originating corpus and its types. It mainly focuses on the steps of originating of corpus and the ways in which it has formed.

Key words: translatology, corpus linguistics, parallel corpus, comparative corpus.

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ФОРМИРОВАНИЯ КОРПУСА

Аннотация

Данная статья посвящена изучению формирования корпусов и причин возникновения корпусов и их типов. Основное внимание уделяется этапам возникновения корпуса и способам его формирования.

Ключевые слова: переводоведение, лингвистика корпуса, параллельный корпус, сравнительный корпус

KORPUS SHAKLLANISHINING NAZARIY ASOSLARI

Аннотация

Ushbu maqola korpusning shakllanishi va vujudga kelish sabablari va uning turlarini o‘rganishga bag‘ishlangan. U asosan korpusning paydo bo‘lish bosqichlari va uning shakllanish usullariga qaratilgan.

Kalit so‘zlar: tarjimashunoslik, korpus lingvistikasi, parallel korpus, qiyosiy korpus.

Kirish. Qadimda tafakkur va muloqot jarayonida tilga oid bilimlarning qanday ahamiyat kasb etganligini idrok etish maqsadida tilni tabiiy hodisa sifatida o‘rganilishiga e‘tibor qaratilganligi bejiz emas. Asrlar davomida insoniyat tomonidan o‘rganilgan boshqa sohalar o‘rtasidagi uzviy aloqalarning o‘rnatilganligini idrok etish barobarida tilshunoslik sohasi anchayin rivojlandi. XXI asrning dastlabki yillarida, tabiiy tildagi o‘ziga xos birikmalarni zamonaviy va qulay bo‘lgan usullardan foydalangan holda tez o‘rganishning yangicha yo‘llarini tadbiq etish mukammal lingvistik nazariyalarni ishlab chiqishni taqozo etadi. Korpus lingvistikasi ma‘lum mezonlarga muvofiq elektron shaklda to‘plangan matnlar jamlanmasini o‘z ichiga oladi [1]. Bu o‘z-o‘zidan tadqiqot sohasi emas, balki tilni baholash, terminologiya, tarjima me‘yorlari, uslub va pragmatik kabi tilning barcha jihatlarini bo‘yicha tadqiqotlar uchun qo‘llanilishi mumkin bo‘lgan qayta ishlash metodologiyasi. Boshqacha qilib aytganda, korpus lingvistikasi nazariy gipotezalarni qo‘llab-quvvatlash yoki rad etish uchun empirik ma‘lumotlar taqdim etish vositasidir. U tilning turli jihatlarini, jumladan, leksikografiya, tilning ijtimoiy tadqiqi, tarixiy tilshunoslik, hisoblash lingvistikasi, terminologiya, tarjima va avtomatlashtirilgan yozuv kabilarni o‘rganish uchun muhim sanaladi.

Xorijiy tillarga bo‘lgan qiziqish, yuqori darajada bilishga bo‘lgan talab tarjima sohasining birmunchayin ilgariylashiga turtki bo‘ldi desak, mubolag‘a bo‘lmaydi. Kundan kunga taraqqiy etayotgan tarjimaning turli jabhalari, nafaqat tilga xos bilimlar, balki, IT sohasiga oid ko‘nikmalarni ham chuqur o‘zlashtirish kabi masalalarning dolzarbligidan dalolat beradi. So‘nggi yillarda, tilshunoslik jabhalarida, jumladan, leksikografiya, terminologiya, kompyuter lingvistikasi, korpus lingvistikasi sohalarida bir qator tadqiqotlar olib borilayotganligiga qaramasdan, nutqiy faoliyat va unda ifodalangan so‘z ma‘nolari bilan bog‘liq masalalar to‘laligicha o‘z yechimini topgani yo‘q. Korpus leksik ma‘lumotlar massivi sifatida muayyan davrda tilning leksik tarkibining o‘zgarish jarayonlari dinamikasini o‘rganishga, turli janrlarda va mualliflarda leksik hamda grammatik

xususiyatlarning qo‘llanilishini tahlil qilishga imkon beradi. Shuningdek, korpuslar, turli xil tarixiy va zamonaviy lug‘atlarni tayyorlashda ko‘p qirrali leksikografik tadqiqotlarning manbai va vositasi bo‘lib xizmat qiladi. Zero, ushbu korpuslardan ta‘lim jarayonida tilni o‘rgatish uchun grammatikani qurish va takomillashtirishda foydalanish ham mumkin. Korpus matnlarni statistik tahlil qilish va qoidalarni asoslash, lingvistik gipotezalarni va muqim me‘yorni yuzaga kelish holatlarini tekshirish uchun foydalaniladigan tizim hisoblanishi bilan muhim ahamiyat kasb etadi.

Ta‘kidlash joizki, leksemalar va matnlar korpuslarini yaratish hamda ulardan foydalanishning o‘ziga xos jihatlarini o‘z ob‘ekti sifatida belgilab olgan soha – korpus lingvistikasi bugungi kunda ko‘plab lingvistik tadqiqotlarga mo‘ljallangan til ma‘lumotlarining yirik majmualarini, ulardan foydalanishning nazariy asoslari va amaliy mexanizmlarining shakllanishi uchun zamin yaratadi. Til korpusi elektron kutubxonadan farqli ravishda tilni o‘rganish uchun zaruriy, foydali va qiziqarli matnlarni to‘plashni nazarda tutadi. Korpusni elektron kutubxonadan ajratib turuvchi birinchi omil undagi matnning xususiyati va qo‘shimcha ma‘lumotlar bilan boyitilganligi hisoblanadi va bu belgi korpusning alohida qismi – korpus birliklariga yozilgan izohni tashkil etadi. *Korpus* leksemasiga keltirilgan bir qator tavsiflar mavjud bo‘lib, uning asosiy vazifasini yoritib berishga qaratilgandir.

Ma‘lumki, korpus so‘zi lotincha *corpus* – *qism, tana* ma‘nolarini anglatib, uning zamonaviy ingliz tilida paydo bo‘lishi bilan bog‘liq bo‘lgan bir-biridan farqli ikkita qarashlar mavjud:

- 1) *corpse* eski fransuz tilidagi *cors* so‘zidan kelib chiqqan;
- 2) *corps* 18 asrdagi keng zamonaviy fransuz tilida keng qo‘llanilgan *corps* so‘zi yordamida yasalgan.

Birinchi, ya‘ni *corpse* so‘zi ingliz tiliga XIII asrda *cors* shaklida kirib kelgan bo‘lib, XIV asr mobaynida lotincha ‘p’ harfi so‘z tarkibiga kiritildi. Dastlab, “*tana*” ma‘nosini, keyinchalik, XIV asr oxirlarida esa “o‘lik tana” tushunchasi ommalashgan. Biroq, asl lotincha *corpus* termini ingliz tili leksik qatlamida XIV asrdayoq keng qo‘llanila boshlangan.

Korpus tilshunosligi doirasida “korpus” tushunchasi “tilga oid ma'lumotlarning katta hajmdagi to'plamini, yozma matnlarni yoki yozib olingan nutq tamskripsiyasini lingvistik tavsifning boshlang'ich nuqtasi yoki til haqidagi farazlarni tasdiqlash vositasi sifatida ishlatilishi mumkin” [3]. Demak, elektron shaklga ega bo'lgan, tilga xos ma'lum bir jihati yoki qo'llanishini ifodalash maqsadida ilmiy asoslangan yozma va og'zaki matn namunalarning katta to'plami nazarda tutilgan. Aniqlik, korpus – til namunasi sifatida undan foydalanishga oid bo'lgan ma'lum bir lingvistik mezonlarga muvofiq tarzda tartibga solingan birliklar jamlanmasi. Ya'nikim, korpus uslubiy jihatdan turli demografik o'zgarishlarni va tabiiy tilning qo'llanishidagi ko'p qirralikni namoyon etuvchi matnlardan tuzilgan, millionlab so'zlarni o'z ichiga olish uchun mo'ljallangan dasturiy ta'minotdir. Korpus lingvistikasida korpusdan tashqari *concordance* termini ham keng qo'llanilib, *concordance* so'zining etimologik talqiniga nazar tashlaydigan bo'lsak, *concordantia* lotincha *cum*, ya'ni, “bilan” va *cor* esa “yurak” ma'nolarini anglatgan.

Korpus lingvistikasi tilshunoslarning ta'riflariga binoan, “muvofiqlik asosida hosil qilingan satrlar va so'z ro'yxatini ekrandan qidirish bilan kompyuter dasturlari orqali katta matnlar yoki kichik matnlarning katta to'plamdagi hodisalarni tushunishga urinish”ning eng oson usuli sifatida ularning ongida shakllangan. Bu kabi matn izohlash usuli katta hajmga ega matnlardagi bir nechta kontekstda ifodalangan so'zlar va iboralarni to'liq qidirishga asoslangan bo'lib, ushbu usul yordamida XIII asrda injilshunos olimlar va ularning izdoshlari tomonidan injil sahifalari satrlarga va so'zlarga ko'chirilgan. Shuningdek, konkordans yohud muvofiqlik asosida injildagi so'zlarning alifbo tartibida, iqtiboslari hamda qaysi matnlarda qo'llanilganligini ko'rsatishning amaliy ehtiyojidan kelib chiqqan holda boshqa olimlar tomonidan amalga oshirilgan. Shuningdek, injildagi to'plangan matnlarning *concordance* deb nomlanishi ushbu so'zning lug'aviy ma'nosi bilan bog'liq holda injilning shunchaki matnlardan tashkil topgan kitob emas, balki ushbu mashhaqqatli urinishning asl g'oyaviy asosi uning ilohiyatga uyg'unlashgan xabar ekanligi haqidagi da'vonni ta'kidlashdan iborat bo'lgan.

Mavzuga oid adabiyotlar tahlili. Korpus tarixiy silsilasi shundan dalolat beradiki, korpuslarni yaratish va korpus usullaridan foydalanish ko'p asrlik davrga ega bo'lib, birinchi davr elektronlashishgacha bo'lgan davr deb nomlanadi, ya'ni *pre-electronic*. Ushbu davr ikki bosqichdan iborat bo'lib, 1) miloddan avvalgi davrdan XVIII asrgacha; 2) XVIII asrdan XIX asrning 60-yillarigacha bo'lgan mobaynni o'z ichiga oladi. Svartvikning nazariyasiga binoan korpus lingvistikasining shakllanib, rivojlanish jarayoni uch davrga bo'linadi:

- 1) *kompyuterlashtirilishgacha bo'lgan korpus lingvistikasi* – XIII asrdan XX asr o'rtalarigacha;
- 2) *kompyuterlashtirilgan korpus lingvistikasi* – 1960-yillardan 1970-yillargacha;
- 3) *korpus lingvistikasi* – 1980/1990 yillardan to hozirgi kunga qadar.

Birinchi bosqich – miloddan avvalgi V-IV asrlarda Panini tomonidan yaratilgan mashhur qadimgi hind grammatikasi dastlabki korpus namunasi sifatida og'zaki shakli saqlangan. Shuningdek, Sanskrit tilini ifodalovchi Veda matnlari korpusi Panini grammatikasiga asoslangan va boshqa ko'plab elektronlashishgacha bo'lgan korpuslar turli diniy kitoblardan ham bog'liq bo'lgan. Ular orasida Injil matnlari korpusi eng mashhur bo'lgan va eng ko'p o'rganilgan. Masalan, Antonio Padua injilning birinchi anonim konkordansi *Concordantiae Morales* bilan bog'liq *Vulgate* – injilning V asrga tegishli bo'lgan, lotin tilidagi muqobiliga asoslangan. Shu davrning o'zida anchayin mukammallashtirilgan nusxasi 1230 yil, Avliyo James

monastirida, Karolik Kardinal *Hugo* tashabbusi bilan tashkil etilgan 500 kishidan iborat Dominikan rohiblari jamoasi tomonidan “*so'zlar indeksi*” tuzilganligini kuzatish mumkin. Shundan buyon, injil konkordanslarining ko'plab boshqa tubdan o'zgargan muqobillari, jumladan, 1737 yil *Crudenning* “Muqaddas yozuvlarga to'liq muvofiqlik” va 1890 yil Strongning “Injilning to'liq muvofiqlik” kabi konkordanslari yaratilgan. Mazkur muqaddas kitobdagi so'zlar ro'yxatida keltirilgan havolalar simfoniya yoki konkordans deb nomlangan bo'lib, unga bog'liq ravishda ilk konkordanslar XIII asrning boshlarida paydo bo'lgan.

Shuningdek, Shekspirning asarlari ham konkordansning asosiy vositasi sifatida ushbu sohaning rivojlanishiga hissa qo'shgan. Misol tariqasida keltirib o'tish joizki, Becketning [2] 1787 yilda Shekspirning asarlari asnosida yaratgan konkordansida so'z va uning lingvistik konteksti hamda joylashuvi berilgan bo'lib, adabiyot shinavandalari uchun qimmatli ma'lumotlarni taqdim etgan. Lindquistning tadqiqotlariga muvofiq, “Shakespeare korpusi” *pre-electronic corpora* namunalaridan hisoblanib, korpus sifatida – “biror bir yozuvchining asarlari yoki muayyan matnlar to'plami” nazarda tutilgan.

Binobarin, Xristian dini rohibi Roberto Busa (Jesuit priest Roberto Busa) tomonidan 1950-yillarda boshlangan va 1970-yillarning oxirlarida yakunlangan, avliyo Tomas Akvinas (St Thomas Aquinas)ning to'liq asarlarini o'z ichiga qamrab oluvchi “*Index Thomisticus / Tomas indeksi*” nomli qisqa elektronik raqamlashtirilgan bazasining yaratilishi konkordansning elektron usulining vujudga kelishiga turtki bo'ldi. Shu o'rinda, yana ikkita asosiy manbani sanab o'tish kerak, masalan, leksikograflar va Chomskiydan oldin faoliyat yuritgan struktur tilshunoslarning tadqiqotlari ham zamin bo'la olgan. Har ikkala usul ham tadqiqotchilar uchun ma'lumotlarni to'plashda muhim ahamiyat kasb etgan. Shuningdek, doktor Samuel Jonson [13] (Dr Samuel Johnson)ning 1755 yilda chop etilgan birinchi “Ingliz tilining izohli lug'ati” ko'p yillar, jumladan, 1560-1660 yillar mobaynida qo'llanilgan qog'ozlarga tushirilgan misollarni yig'ish asnosida qog'ozda korpus yaratishning mahsuli bo'lib xizmat qildi. Ehtimol, uch milliondan oshiq qog'oz bo'laklari korpusida mujassamlashgan so'zlarning qo'llanishini amalda ta'minlovchi lug'at, ya'ni, 1880-yillarda Oxford English Dictionary tarkibiga kiritilgan. Dunyoning mashhur lug'atlaridan birini yaratishda ushbu millionlab qog'oz parchalaridagi so'zlar tahrirlangan holda matnning mazmunga ega bir qismiga aylantirildi.

Leechning [5] ta'kidlashicha, 1950-yillarda amerikalik strukturalistlar, jumladan, Harris, Fries and Hillarning tadqiqotlarida “ma'lumotlarni yig'ish” tushunchasi keng qo'llanilganligini kuzatish mumkin. Shuningdek, ilk injilshunos olimlar va ingliz yozuvchilarining so'zlarni qidirish va indekslash borasidagi olib borgan chora-tadbirlari nafaqat, tilga oid ma'lumotlarni to'plash, balki, ona tilga xos bilimlarni tadqiq etish nuqtai nazaridan strukturalistlarning konkordans ilmi rivojida peshqadamligidan dalolat beradi. Ona tilini o'zlashtirishga bo'lgan qiziqish transkripsiyalangan ma'lumotlar – 1960-yillarda yaratilgan CHILDES [14] til bazasiga asoslangan bo'lsada, ushbu loyiha 1984 yildagina rasmiy tasdiqlangan. Shu o'rinda, badiiy asarlarning foydali korpusini yaratishda adabiyotshunoslarning qo'shgan hissalarini e'tirof etish joiz.

Tadqiqot metodologiyasi. Darhaqiqat, bu davrda korporalardan foydalanishning mukammal usullarini ishlab chiqish borasida yutuqlarga erishildi. Ushbu davrning rivojlanishi 1980 yillarda kompyuter texnologiyalarining ommalashishi bilan bog'liq bo'lib, elektron hisoblash mashinalarining ixtirosi va keng tarqalishi bilan elektronlashishgacha bo'lgan davrda yaratilgan korpuslar o'z ahamiyatini yo'qota boshladi. Negaki, korpusni

avtomatlashtirish bilan bog'liq tadqiqotlar shu davrda amalga oshirildi. 1980-yillardan boshlab, ushbu tadqiqotlardan olingan natijalar asnosida korpusni yangi usullar va vositalar yordamida o'rganishga yo'naltirilgan ilmiy ishlar sababli, korpus lingvistikasi rivojlandi. McEnry and Wilsonlarning so'zlariga ko'ra, korpus lingvistikasi hozirda ham taraqqiy etib, metodologik jihatdan tobora yetuklashib va shuningdek, o'rganiladigan tillar soni har yili o'sib bormoqda. Biroq, tilshunoslikning ba'zi sohalarida qog'oz kartotekalar, qayin po'stlog'i yoki sopol bo'laklardagi matnlar bilan ishlash tadqiqot ishining muhim qismi bo'lib, ko'pgina tillar va matnlar uchun korpuslarning turlari hamda sonlarida katta tafovut mavjud edi.

1979 yilga kelib, Howard-Hillning tashabbusi bilan kompyuterlashtirilgan, ya'ni, "adabiyotni o'rganish uchun maqsadli ishlovchi dastur" konkordanslari yaratildi [4, 30]. Conradning sakkizta ishi jamlanmasi konkordanslari va shuningdek, Gerard Manley Hopkins va T.S.Eliotlarning konkordanslari 1985 yilda chop etildi. Ilk bor, 1950 yillar oxirlarida, kompyuterlashtirilgan, ya'ni, ma'lumotlarni saqlash uchun mo'ljallangan perfokarta (punched-card technology for storage) yordamida boshqariladigan konkordanslar tuzildi. Perfokartadan foydalanishdagi kamchiliklardan biri, kompyuter xotirasiga ma'lumotlarni kiritishda uzoq vaqt sarflangan, masalan, oltin mingdan ortiq so'zlarni qayta terish uchun 24 soatdan ortiq vaqt ketgan. 1970 yillarda bu jarayon sezilarli darajada mukammallashib, joriy yildan boshlab kalit so'zlar konteksti konkordanslar (Key Word in Context (KWIC)) katalogini indekslovchi kartalarini almashtirish va mavzulashgan matn tahlilini avtomatlashtirish jarayonlariga e'tibor qaratildi. Ushbu usul kutubxonada axborot izlash va turli mavzular tahlilini boshqarishda keng foydalanilgan. Shuningdek, uning afzalligi, asarlardan keltirilgan iqtiboslarni tezkor topish va undan unumli foydalanishdadir.

Tahlil va natijalar. Ma'lumotlar ba'zasini yaratishga sarflangan mehnat va tuzuvchilar e'tibordan chetda qolib, bu go'yo kompyuterning mehnati sifatida talqin etilgan. Biz uchun hozirda ma'lum bo'lgan ma'lumotlar korpusi 80-90 yillarda yaratilgan bo'lib, tilshunoslikning barcha jabhalarida keng foydalanib kelinmoqda. Ilgari korpus tushunchasi bir xil mavzu doirasida yig'ilgan yoki jamlangan asarlarni ifoda etgan bo'lsa, endilikda u lingvistik terminologiyada keng qo'llanila boshlandi.

XVIII asrda korpusning yuqorida keltirilgan ma'noda ishlatilganligi, jumladan, olimlarning "Lotin shoirlari korpusi" yoki "qonunchilik korpusi"lariga murojaat etganliklari Oksford ingliz tili lug'atining yaratilganligi bilan dalillanadi. Oksford ingliz tili lug'atida korpus so'ziga keltirilgan iqtibos – lingvistik manbada ommaga tushunarli bo'lgan ma'no, ya'ni, "lingvistik tahlilga asoslangan yozma yoki og'zaki material to'plami" ifodalangan. Biroq, 1980-yil boshlarigacha bo'lgan davrda korpus lingvistikasi termini ommolashib ulgurmasdan,

ingliz tilidagi ma'lumotlar korpusining yaratilishi bilan amalda tadbiiq etila boshlangan. Shuningdek, korpus doirasida tadqiqotlar olib borilishining yo'lga qo'yilishi 40-yillardayoq boshlangan bo'lib, bunga – Bloomfield, Fritz, Bonders va boshqa tilshunoslarning ishlarini misol tariqasida keltirib o'tish mumkin. Ushbu tadqiqotlarning davomiyligi W.N.Francis, Henry Kuceralarning korpus yaratish bosqichlari bilan bog'liq bo'lgan ishlar orqali ta'minlandi.

Svartvikning [9] korpus lingvistikasi bo'yicha bildirgan fikriga muvofiq "Korpus lingvistikasi yetuklikka erishganligini xitob qilishi ushbu sohaning tilshunoslikning bir jabhasi sifatida shakllanib ulguraganligidan dalolat beradi. Keyinchalik, bu sohaga oid olib borilgan tadqiqotlarda yangi yo'nalish haqida bir-biriga nisbatan yaqin bo'lgan tavsiflar keltirila boshlandi. Masalan, E.Fineganning fikriga ko'ra, "korpus – matn yaratilgan vaziyat va unga tegishli ma'lumotlar, jumladan, so'zlovchi, muallif, adresat yoki auditoriya haqidagi ma'lumotlarni o'z ichiga olgan, mashina o'qiydigan formatga ega reprezentativlangan matnlar majmuasi" bo'lsa, McEnery va Wilsonlarning [6] nuqtai nazarida esa "korpus – til modeli sifatida foydalanish uchun aniq til mezonlari bo'yicha tanlangan til fragment (qism / parcha)lari to'plami". Ularning nazariyasiga muvofiq, *corpus* matnning xohlagan qismiga, yoki kompyuter vositasida avtomatlashtirilgan matnning bir qismiga yoki tilning xilma-xilligini ifodalovchi avtomatlashtirilgan matnlarning cheklangan to'plamiga tegishlidir. Demak, korpus shaklida bo'lmagan nutqiy ma'lumot muhim sanalsa-da, faqat korpusga asoslangan holda lingvistik tahlillarni amalga oshirishning imkoni yo'q.

Sinclair ularning fikridan farqli ravishda tashqi omillarga diqqatini qaratgan holda quyidagi xulosaga keladi: "korpus tilni yoki tilning o'zgarishini to'liq aks ettirish uchun tashqi mezonlarga muvofiq tanlangan elektron shakldagi matnlar parchalari to'plami" [10]. Ya'nikim, tilning holatini yoki turini tavsiflash maqsadida tanlangan, nutqimizda tabiiy ravishda mavjud bo'lgan til matnlari to'plami, aniqrog'i, so'zlar ifodasidagi xilma-xillik tilning leksik qatlamlarga ajratilishida qiyinchiliklarni keltirib chiqarishi milliondan ortiq so'zni korpus tarkibiga kiritilishiga sabab bo'ladi.

Renoufning ta'kidlashicha, "korpus tushunchasi lingvistik tadqiqotlar maqsadida kompyuterda saqlanadigan va qayta ishlanadigan yozma yoki og'zaki matnlar to'plami"ni ifodalaydi [8].

Xulosa va taliflar. Xulosa o'rmda keltirish joizki, turli tipdagi ma'lumot til korpusida o'zining tabiiy kontekstual shaklida joylashishi sababli, uni har tomonlama o'rganish uchun asos bo'ladi. Bir marta tuzilgan axborotlar massivi bir necha tadqiqotchi tomonidan ko'p marta turli maqsadlarda qo'llanilishi mumkin. Korpus tilni tadqiq etish, o'rganish, lug'at tuzishda eng zamonaviy va keng imkoniyatli dasturlashtirilgan tizimdir.

ADABIYOTLAR

1. Baker, M. "Corpora in Translation Studies: An Overview and some Suggestions for Future Research." *Target* 7.2: 1995. 223 – 43. Print.
2. Becket A Concordance to Shakespeare. London: Printed for G.G.J. and J.Robinson. viii, 1787. – 470 pages.
3. Crystal, D. *The Cambridge Encyclopaedia of the English Language*. Cambridge: Cambridge University Press. 1995.
4. Howard-Hill, T. H. *British Bibliography and Textual Criticism: A Bibliography*. Oxford: Clarendon Press. 1979.
5. Leech, G. "Corpus linguistics and theories of linguistic performance". In, J. Svartvik (Ed.) *New Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82*. Berlin: Mouton de Gruyter. 1992. Pp. 125-148.
6. McEnery, T. and A. Wilson *Corpus Linguistics*. Edinburgh: Edinburgh University Press. 1996.
7. Pápai, V. "A Universal of Translated Text" *Translation universals: do they exist?* 2004. – 143.
8. Renouf, A. *Corpus development*. In J. Sinclair (ed.) *Looking up*. London: Collins. 1987.
9. Svartvik J. *Corpus linguistics comes of age // Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82 Stockholm, 4 – 8 August 1991 / ed. J. Svartvik*. Berlin; New York: Mouton de Gruyter, 1992. P. 7 – 14.
10. Viana V., Zyngier, S. & G. Barnbrook (Eds). *Perspectives on Corpus Linguistics*. Amsterdam: John Benjamins. 2011.
11. Zakharov V., Bogdanova S.) *Korpusnaya Lingvistika: uchebnik*. 2-y izd., pererab. – SPb.: S.-Peterb. 2011. – 215 s.

12. Zubov A.V., Zubova I.I. Informatsionniye texnologii v lingvistike: Ucheb. Pososbiye. – M.: Akademiya. 2004.
13. https://archive.org/details/johnsons_dictionary_1755
14. <https://childes.talkbank.org/access/>