



O‘ZBEK VA INGLIZ TILLARIDA JURNALISTIK DISKURSNING STILISTIK XUSUSIYATLARINI KORPUSGA ASOSLANGAN TAHLILI

Mirzarahimov Bahodir Shavkat o‘g‘li
O‘z MU Kompyuter lingvistikasi va amaliy
tilshunoslik kafedraasi 2-bosqich magistranti

Annotatsiya:

Ushbu maqola o‘zbek va ingliz tillaridagi jurnalistik diskursning stilistik xususiyatlarini korpus lingvistikasiga asoslangan holda o‘rganishga bag‘ishlangan. Maqolada gazeta matnlari korpuslari yordamida jurnalistik uslubning o‘ziga xos xususiyatlari, metaforik vositalar, sintaktik tuzilmalarning o‘zgarishi va axborotni taqdim etish strategiyalari tahlil qilinadi. Shuningdek, maqola doirasida turli davr va mintaqalarga mansub gazeta matnlari solishtirilib, milliy va madaniy kontekstning jurnalistik nutqqa ta’siri o‘rganiladi. Olingan natijalar lingvistik tadqiqotlar, tarjima amaliyoti va ommaviy axborot vositalari tahliliga hissa qo‘shishi mumkin.

Kalit so‘zlar: jurnalistik diskurs, stilistika, korpus lingvistikasi, metaforik tahlil, sintaktik tuzilma, o‘zbek tili, ingliz tili, matnshunoslik, axborot taqdimoti, diskursiv strategiyalar.

Gazeta matnlari korpusini yaratish bo‘yicha tadqiqotlar bugungi kunda lingvistik va matnshunoslik yo‘nalishidagi dolzarb masalalardan biri hisoblanadi. Gazeta matnlari korpusi lingvistik tadqiqotlarda muhim ahamiyat kasb etib, u til o‘zgarishlarini kuzatish, uslubiy xususiyatlarni aniqlash hamda ijtimoiy kontekstni tahlil qilish uchun keng qo‘llaniladi. Gazeta matnlari korpusi gazetalarda chop etilgan yangiliklar, maqolalar, sharhlar, intervyular, reklama matnlari va boshqa turdagi jurnalistik materiallarni o‘z ichiga oladi. Ushbu turli janrlardagi matnlar orqali tilning turli funksional uslublarini o‘rganish, jurnalistik tilning o‘ziga xos jihatlarini aniqlash va diskursiv tadqiqotlarni amalga oshirish imkoniyati yaratiladi. Korpus tarkibiga kiritilgan materiallar maqsad va tadqiqot yo‘nalishiga qarab tanlab olinadi, bu esa lingvistik tadqiqotlarning natijalarini yanada aniqroq chiqarishga xizmat qiladi.



International Conference on Economics, Finance, Banking and Management

Hosted online from Paris, France

Website: econfseries.com

24th February, 2025

Gazeta matnlari korpusini yaratishda quyidagi muhim jihatlarga e'tibor qaratiladi. Birinchidan, matn janrlari va mavzularining keng qamrovli bo'lishi lozim. Ijtimoiy, iqtisodiy, siyosiy va madaniy sohalarini o'z ichiga olgan matnlar korpusning ilmiy ahamiyatini oshiradi. Ikkinchidan, vaqt oralig'ining keng qamrovli bo'lishi talab etiladi. Har xil davrga mansub gazeta matnlarining jamlanishi tilning sinxron va diakron tahlilini amalga oshirish imkoniyatini beradi. Uchinchidan, geografik omillar ham muhimdir. Mahalliy, milliy va xalqaro gazetalar korpusga kiritilishi orqali tilning mintaqaviy va global xususiyatlari tadqiq qilinadi. Gazeta matnlari korpusi lingvistik tahlillarda turli maqsadlar uchun foydalaniladi. Leksik-semantik tadqiqotlar orqali gazeta tilidagi yangi so'z va iboralar, terminologik birliklar hamda metaforik ifodalarning shakllanish jarayonlari o'rganiladi. Grammatik tadqiqotlar esa jurnalistik matnlardagi sintaktik tuzilmalar, so'z birikmalari va gap qurilishi xususiyatlarini aniqlash imkonini beradi. Bundan tashqari, stilistik tahlillar orqali jurnalistik uslubning o'ziga xos belgilari aniqlanadi va turli ommaviy axborot vositalari til uslublari o'zaro taqqoslanadi. Gazeta matnlari korpuslari ta'lim jarayonida ham katta ahamiyat kasb etadi. Filologiya va jurnalistika yo'nalishida tahsil olayotgan talabalar uchun real til namunalaridan foydalanish ularning lingvistik bilimlarini oshirishga xizmat qiladi. Shuningdek, tilning rivojlanish tendensiyalarini o'rganish va yangi til birliklarining matnlarda qo'llanishini kuzatish orqali ilmiy natijalarga erishish mumkin.

Dunyo miqyosida gazeta matnlari asosida yaratilgan korpuslar bir necha turga bo'linadi. British National Corpus ingliz tilidagi gazeta matnlarini qamrab olgan bo'lib, leksik va grammatik tadqiqotlar uchun katta ahamiyatga ega. Corpus of Contemporary American English esa AQShdagi zamonaviy gazeta matnlarini o'z ichiga olib, ingliz tilining zamonaviy rivojlanish jarayonlarini o'rganish uchun ishlatiladi. Shuningdek, Gutenberg Project tarixiy gazeta va matnlar bazasi bo'lib, qadimgi matnlarni tadqiq qilish imkonini beradi. Gazeta matnlari korpusini yaratish va uni tadqiq qilish tilshunoslikda katta ahamiyatga ega bo'lib, matnshunoslik, stilistika, sotsiolingvistika va korpus lingvistikasi doirasidagi izlanishlarga beqiyos hissasini qo'shadi. Ushbu tadqiqot yo'nalishi orqali gazeta tili rivojlanishining muhim bosqichlari aniqlanadi, ommaviy axborot vositalari tilining lingvistik xususiyatlari o'rganiladi hamda jurnalistik diskursning zamonaviy tendensiyalari



International Conference on Economics, Finance, Banking and Management

Hosted online from Paris, France

Website: econfseries.com

24th February, 2025

tadqiq qilinadi. Shunday qilib, gazeta matnlari korpusi lingvistik tadqiqotlar uchun ishonchli manba bo‘lib xizmat qiladi va turli ilmiy yo‘nalishlarda foydalanish uchun katta imkoniyat yaratadi.

Gazeta korpuslari neologizmlar va boshqa leksik hodisalarni o‘rganishda boy manba hisoblanadi.¹ Gazeta matnlari korpuslari zamonaviy tilshunoslik tadqiqotlarida muhim rol o‘ynaydi, chunki ular tilning dinamik o‘zgarishlarini kuzatish, leksik va grammatik xususiyatlarini tahlil qilish hamda til texnologiyalarini rivojlantirish uchun asosiy manba hisoblanadi. Hozirda dunyo bo‘ylab ko‘plab gazeta korpuslari yaratilgan bo‘lib, ular turli davrlardagi matnlarni jamlash orqali til taraqqiyotini tahlil qilish imkonini beradi.

Xronologik gazeta korpuslarini yaratish gazeta materiallarining aniq sanalar bilan bog‘liq bo‘lishi sababli ayniqsa muhimdir. Bunday korpuslar yordamida tilning davriy o‘zgarishlarini kuzatish, muayyan yillarda ommaviy axborot vositalarida qo‘llangan til birliklarini tahlil qilish hamda tilning stilistik va diskursiv jihatlarini tadqiq qilish mumkin. Gazeta matnlariga asoslangan korpuslar sotsiolingvistik tadqiqotlar uchun ham katta ahamiyatga ega bo‘lib, jamiyatdagi madaniy va ijtimoiy o‘zgarishlarni aks ettiruvchi til vositalarini tahlil qilish imkoniyatini yaratadi. Gazeta korpuslari orasida Umumiy Til Resurslari va Texnologiya Infratuzilmasi (CLARIN) muhim o‘ringa ega. Ushbu infratuzilma 34 ta gazeta korpusiga kirish imkonini taqdim etadi va lingvistik tadqiqotlar uchun keng imkoniyatlar yaratadi. CLARIN tarkibidagi gazeta matnlari ko‘pincha tarixiy bo‘lib, XVIII asrdan boshlangan qadimiy nashriyotlarni o‘z ichiga oladi. Ushbu infratuzilmada joylashgan gazeta korpuslarining yettitasi ko‘p tilli bo‘lib, turli tillardagi matnlarni solishtirish imkonini beradi. Qolganlari esa bir tilli bo‘lib, chex, fin, arab, nemis, fransuz, polyak, norveg, shved, yunon va italyan tillaridagi matnlarni qamrab oladi.

Tilshunoslikda korpus – bu tizimli ravishda yig‘ilgan va tahlil qilish uchun mo‘ljallangan matnlar to‘plami bo‘lib, odatda elektron formatda saqlanadi va qayta ishlanadi. Gazeta matnlari korpusi esa ommaviy axborot vositalari orqali tarqatilgan til namunalari asosida tuzilgani uchun real kommunikativ kontekstni aks ettiruvchi asosiy manbalardan biri hisoblanadi. Ushbu korpuslardan lingvistik tadqiqotlarda,

¹ S. Kübler, H. Zinsmeister. “Corpus Linguistics and Linguistically Annotated Corpora.” Bloomsbury. 312 pp, 2015. ISBN: 978-1-4411-6447-6

International Conference on Economics, Finance, Banking and Management

Hosted online from Paris, France

Website: econfseries.com

24th February, 2025

avtomatik matn tahlili, sun'iy intellekt asosida ishlov berish va tarjima tizimlarini yaratishda keng foydalaniladi. Shunday qilib, gazeta matnlari korpuslarini yaratish va ularni lingvistik tadqiqotlarga jalb qilish zamonaviy tilshunoslikning muhim yo'nalishlaridan biri bo'lib, bu nafaqat tilning tarixiy va zamonaviy holatini o'rganishga, balki turli tillarning o'zaro ta'sirini tahlil qilishga ham xizmat qiladi.² Korpus tilshunoslikda bir tildagi yoki bir nechta tildagi matnlarni statistik tahlil qilish, gipotezalarni tekshirish, tildagi hodisalarni kuzatish yoki nazariy lingvistik qoidalarni sinovdan o'tkazish uchun ishlatiladi. CLARIN infratuzilmasidagi korpuslarning aksariyati ko'p tilli (multilingual) bo'lib³, bir tilli gazeta matnlari korpuslariga misollar quyidagi jadvalda keltirilgan.

1-jadval. CLARIN infratuzilmasiga kirgan bir tilli gazeta korpuslariga misollar

	Korpus nomi	Korpus hajmi	Til	Korpus tavsifi	Korpusdan foydalanish litsenziyasi
1.	An-Nahar gazeta matnlari korpusi [5]	24 million token	Arab	Ushbu korpusda 1995 yildan 2000 yilgacha arabcha "An-Nahar" gazetasidan olingan maqolalar mavjud	ELRA END USER
2.	The Karelian Finnish gazeta matnlari korpusi [6]	500 000 token	Fin	Ushbu korpus Finlyandiyaning Karjalan Sanomat gazetasining 2012 yildan 2014 yilgacha bo'lgan maqolalarini o'z ichiga oladi	CLARIN ACA
3.	TIGER gazeta korpusi [7]	900 000 token	Nemis	Ushbu korpusda Germaniyaning Frankfurter Rundschau gazetasida maqolalari mavjud	CLARIN PUB
4.	Zamonaviy grekcha matnlar korpusi: "Makedoniya" gazetasi korpusi [8]	3 million token	Grek	Ushbu korpusda turli mavzularda (siyosat, iqtisodiyot, sport) gazeta maqolalari mavjud	CC-BY-NC-SA
5.	GP 1994 and 2001-2011 [9]	271 million token	Shved	Ushbu korpus Shvetsiyaning Göteborgsposten gazetasidan 1994 yildan va 2001 yildan 2011 yilgacha bo'lgan maqolalarni o'z ichiga oladi.	CC-BY

² R. Nordquist. "Definition and Examples of Corpus Linguistics," ThoughtCo., 2019. [Online]. Havola: <https://www.thoughtco.com/what-is-corpus-linguistics1689936>

³ S. Wallis, G. Nelson. "Knowledge discovery in grammatically analysed corpora." Data Mining and Knowledge Discovery, 5: 305–335. 2001.



International Conference on Economics, Finance, Banking and Management

Hosted online from Paris, France

Website: econfseries.com

24th February, 2025

Tilshunoslikda korpus lingvistik tadqiqotlar uchun asosiy manbalardan biri bo'lib, u turli tillardagi matnlarni tizimli tarzda yig'ish va tahlil qilish imkonini beradi. Gazeta matnlari korpuslari bu yo'nalishda muhim rol o'ynaydi, chunki ular tilning diaxron va sinxron jihatlarini tahlil qilishda, shuningdek, jamoatchilikka ta'sir ko'rsatgan muhim lingvistik o'zgarishlarni kuzatishda asosiy manba hisoblanadi. Gazeta matnlari korpuslari tuzilishiga ko'ra, bir tilli (monolingual) va ko'p tilli (multilingual) turlarga bo'linadi. Bir tilli korpuslar faqat bitta tildagi matnlarni o'z ichiga olgan bo'lib, lingvistik tadqiqotlar uchun eng keng qo'llaniladigan tur hisoblanadi. Bunday korpuslar, odatda, muayyan tildagi leksik, grammatik va sintaktik xususiyatlarni tahlil qilish uchun ishlab chiqiladi.

Korpusdagi har bir maqola gazetaning nomi, sanasi, mamlakati, maqola turi, sahifa raqami kabi metadata bilan ta'minlangan bo'lib, bu tadqiqotlarning yanada aniq va tizimli o'tkazilishiga yordam beradi. Shuningdek, nemis tilidagi **TIGER Corpus** (2.1 va 2.2 versiyalari) Frankfurter Rundschau gazetasi matnlaridan tashkil topgan bo'lib, taxminan 900 000 token (taxminan 50 000 jumla) dan iborat. Ushbu korpus yarim avtomatik tarzda morfologik va sintaktik jihatdan teglangan bo'lib, unda har bir terminal tugun uchun morfologik va lemma ma'lumotlari mavjud. Bunday turdagi bir tilli korpuslar nemis tili grammatikasi, sintaktik tuzilmalari va matn strukturasini tahlil qilishda muhim ahamiyat kasb etadi.

Ko'p tilli gazeta korpuslari esa ikki yoki undan ortiq tildagi matnlarni o'z ichiga oladi va odatda parallel yoki taqqoslama tadqiqotlar uchun mo'ljallangan bo'ladi. Bunday korpuslar turli tillardagi gazeta maqolalarini o'zaro solishtirish, tarjima tadqiqotlari, kross-lingvistik tahlillar va til o'zgarishlarini kuzatish uchun muhim ahamiyatga ega. Ko'p tilli korpuslar lingvistik tadqiqotlarning yanada kengroq qamrovini ta'minlab, bir necha til tizimlari o'rtasidagi sintaktik, semantik va pragmatik farqlarni tahlil qilish imkonini beradi. Gazeta matnlari korpuslarining xilma-xilligi va ularning turli tillardagi namunalari lingvistik tadqiqotlar doirasini yanada kengaytirishga xizmat qiladi. Xususan, xronologik va ko'p tilli korpuslar global lingvistik o'zgarishlarni kuzatish va turli tillar o'rtasidagi semantik hamda sintaktik munosabatlarni aniqlashda muhim manba sifatida e'tirof etiladi.

2-jadval. CLARIN infratuzilmasiga kirgan multilingual gazeta korpuslari

	Korpus nomi	Korpus hajmi	Til	Korpus tavsifi	Korpusdan foydalanish litsenziyasi
1.	MLCC Ko'p tilli va Parallel Korpus [10]	100 million token	golland, ingliz, fransuz, nemis, italyan, ispan	Ushbu korpusda 1986 yildan 1994 yilgacha golland, ingliz, frantsuz, nemis, italyan va ispan tillaridagi gazetalardagi maqolalar mavjud	ELRA END USER
2.	Gazeta matnlari korpusi [11]	435 million token	Shved, ingliz, fin	Ushbu korpusda turli shved, ingliz va fin gazetalaridagi maqolalar mavjud	-
3.	SETIMES - Bolqon tillarining parallel korpusi [12]	341.83 million token	Rumin, turk, serb, ingliz, bolgar, makedon, xorvat, yunon, alban	Ushbu parallel korpusda SETimes veb-sahifasidan olingan onlayn yangiliklar maqolalari mavjud.	Cheklovlar bilan qayta foydalanish uchun ochiq
4.	Parallel Global Voices [13]	8 million	40 ta til	Ushbu korpus https://globalvoices.org/ veb-saytidagi maqolalarni o'z ichiga oladi, bu yerda ko'ngillilar 40 dan ortiq tillarda yangiliklarni nashr etadilar va tarjima qiladilar	CC BY

Ko'p tilli gazeta korpuslari, ba'zan **parallel korpus** deb ham ataladi, ikki yoki undan ortiq tildagi matnli ma'lumotlardan iborat bo'lib, undagi matnlar bir-birining aynan tarjimasi sifatida muvofiqlashtirilgan bo'ladi. Parallel korpuslarning asosiy xususiyati shundan iboratki, har bir matnning ekvivalent tarjimasi mavjud bo'lib, bu lingvistik tadqiqotlar, mashinaviy tarjima, kontrastiv tahlil va tarjima tadqiqotlarida muhim rol o'ynaydi. Bunday korpuslar segmentlar, jumlar yoki mavzular darajasida moslashtirilishi lozim. Masalan, mashhur asarlar va ularning tarjimalari yoki turli tillarda nashr etilgan gazeta maqolalari shunday parallel korpuslarni shakllantirishda asosiy manba bo'lib xizmat qiladi. Bu turdagi korpus foydalanuvchilarga bitta tildagi so'z yoki iboraning barcha misollarini qidirish imkoniyatini beradi va natijalar boshqa tildagi mos keluvchi jumlar bilan birgalikda taqdim etiladi. Shunday qilib, foydalanuvchi ma'lum bir so'z yoki iboraning tarjima variantlarini kuzatish va tahlil qilish imkoniyatiga ega bo'ladi.



International Conference on Economics, Finance, Banking and Management

Hosted online from Paris, France

Website: econfseries.com

24th February, 2025

2-jadvalda keltirilgan gazeta korpuslari ushbu tamoyil asosida shakllangan bo‘lib, ikki va undan ortiq tildagi matnlarni o‘z ichiga oladi. Bunday korpuslar tadqiqotlar doirasini kengaytirib, tilshunoslik hamda tarjima fanlari bo‘yicha izlanishlarni yanada samarali olib borish imkoniyatini yaratadi.

Xulosa:

Umuman olganda, gazeta korpuslarining yaratilishi lingvistik tadqiqotlar uchun muhim infratuzilma vazifasini o‘taydi. Bunday korpuslar turli tillardagi matnlarni markazlashtirilgan platformada jamlash orqali tadqiqotchilar uchun izlanish jarayonlarini sezilarli darajada yengillashtiradi. Xususan, ko‘p tilli gazeta korpuslari tadqiqot qamrovini kengaytirib, tillararo o‘zaro bog‘liqlikni tahlil qilish hamda tarjima jarayonlarini chuqurroq o‘rganish imkoniyatini taqdim etadi. Shu sababli, ushbu korpuslar tarjima fanlari, kognitiv lingvistika, til o‘rganish va mashinaviy tarjima tadqiqotlari uchun bebaho manba hisoblanadi.

Foydalanilgan adabiyotlar:

1. S. Kübler, H. Zinsmeister. “Corpus Linguistics and Linguistically Annotated Corpora.” Bloomsbury. 312 pp, 2015. ISBN: 978-1-4411-6447-6
2. R. Nordquist. “Definition and Examples of Corpus Linguistics,” ThoughtCo., 2019. [Online]. Havola: <https://www.thoughtco.com/what-is-corpus-linguistics1689936>
3. S. Wallis, G. Nelson. “Knowledge discovery in grammatically analysed corpora.” Data Mining and Knowledge Discovery, 5: 305–335. 2001.