



## **PRINCIPLES OF REPRESENTING SYNONYMS IN KOREAN AND UZBEK LINGUISTIC CORPORA**

Sagatova Mukhabbat

Teacher, Department of Korean Philology  
Uzbekistan State World Languages University

### **Abstract**

This article analyzes the principles of representing, identifying, tagging and searching synonyms in Korean and Uzbek linguistic corpora. Corpus linguistics is a modern field that enables the study of language units on the basis of authentic texts. It plays an important role in identifying not only the lexical similarity of synonyms, but also their frequency, collocational environment, genre distribution, stylistic value and pragmatic compatibility. Based on the Sejong Corpus and the Uzbek National Corpus, the article discusses scientific and practical criteria for representing synonyms in electronic linguistic resources. The importance of lexical identification, dominant-centered description, semantic differentiation, stylistic tagging, collocational description and confirmation through contextual examples is also substantiated.

**Keywords:** Corpus linguistics, synonym, linguistic corpus, Sejong Corpus, Uzbek National Corpus, tagging, collocation, frequency, concordance, electronic lexicography.

### **Introduction**

#### **KOREYS VA O‘ZBEK TILI LINGVISTIK KORPUSLARIDA SINONIMLARNI BERISH PRINSIPLARI**

Sagatova Muhabbat Hasanovna

O‘zbekiston davlat jahon tillari universiteti

Koreys filologiyasi kafedrası o‘qituvchisi



## **Annotatsiya**

Mazkur maqolada koreys va o‘zbek tili lingvistik korpuslarida sinonimlarni berish, belgilash, teglash va qidirish prinsiplari tahlil qilinadi. Korpus lingvistikasi til birliklarini real matnlar asosida o‘rganishga imkon beruvchi zamonaviy yo‘nalish bo‘lib, sinonimlarning faqat lug‘aviy ma‘nodoshligini emas, balki ularning chastotasi, kollokatsion muhiti, janrga mansubligi, uslubiy qiymati va pragmatik mosligini aniqlashda muhim ahamiyatga ega. Maqolada koreys tilidagi Sejong korpusi va O‘zbek tili milliy korpusi misolida sinonimlarni elektron resurslarda ifodalashning ilmiy-amaliy mezonlari yoritiladi. Shuningdek, sinonimlarni korpusda berishda leksik identifikatsiya, dominanta markazlashuvi, semantik differensiallashuv, uslubiy teglash, kollokatsion tavsif va kontekstual misollar bilan tasdiqlash kabi prinsiplarning ahamiyati asoslanadi.

**Kalit so‘zlar:** korpus lingvistikasi, sinonim, lingvistik korpus, Sejong korpusi, O‘zbek tili milliy korpusi, teglash, kollokatsiya, chastota, konkordans, elektron leksikografiya.

## **Kirish**

XXI asr tilshunosligi nazariy qarashlar bilan bir qatorda raqamli texnologiyalar, elektron resurslar va katta hajmdagi matn bazalari bilan ham chambarchas bog‘liq holda rivojlanmoqda. Til birliklarini faqat grammatik qoida yoki lug‘aviy izoh asosida emas, balki real nutq materiallari asosida tahlil qilish zamonaviy lingvistik tadqiqotlarning muhim yo‘nalishiga aylandi. Ana shunday yo‘nalishlardan biri korpus lingvistikasidir.

Korpus lingvistikasi ma‘lum prinsiplar asosida tanlangan, elektron shaklda saqlangan va lingvistik ma‘lumotlar bilan boyitilgan matnlar majmuasi asosida til birliklarini o‘rganadi. Korpus orqali so‘zlarning real qo‘llanishi, chastotasi, birikish imkoniyatlari, uslubiy xususiyatlari, janrga mansubligi va kontekstdagi vazifasi aniqlanadi. Shu jihatdan korpus oddiy elektron matnlar to‘plami yoki raqamli kutubxonadan farq qiladi. U til birliklarini avtomatik qidirish, saralash, teglash va tahlil qilish imkoniyatiga ega bo‘lgan ilmiy-amaliy tizimdir.

Sinonimlarni korpusda o‘rganish ayniqsa dolzarbdir. Chunki an‘anaviy lug‘atlarda sinonimlar ko‘pincha bir-biriga yaqin ma‘noli birliklar sifatida



ro'yxat shaklida beriladi. Ammo bunday ro'yxat ularning real nutqda qanday ishlatilishini, qaysi so'zlar bilan birikishini, qaysi uslubda faol ekanini va qaysi kontekstda bir-birining o'rnini bosa olmasligini to'liq ko'rsatib bera olmaydi. Korpus esa aynan shu jihatlarni aniqlash imkonini beradi.

Koreys va o'zbek tillari agglyutinativ tuzilishga ega bo'lgan, boy leksik qatlamlar va murakkab semantik munosabatlarga ega tillardir. Har ikki tilda sinonimlar ko'plab semantik, uslubiy va pragmatik farqlarga ega. Shuning uchun ularni korpusda to'g'ri belgilash, sinonimik qatorlar shaklida tizimlashtirish, har bir birlikning qo'llanish xususiyatini ko'rsatish muhim ilmiy va amaliy ahamiyat kasb etadi.

Mazkur maqolaning maqsadi koreys va o'zbek tili lingvistik korpuslarida sinonimlarni berish prinsiplarini tahlil qilishdan iborat. Ushbu maqsad asosida korpus tushunchasi, sinonimlarni elektron resurslarda belgilash mezonlari, Sejong korpusi va O'zbek tili milliy korpusining imkoniyatlari, shuningdek, sinonimlarni teglash va qidirish masalalari yoritiladi.

## **Asosiy qism**

### **1. Lingvistik korpus tushunchasi va uning sinonimiya tadqiqidagi o'rni**

Lingvistik korpus — bu ma'lum tilning yozma yoki og'zaki matnlarini tizimli ravishda jamlagan, elektron shaklda saqlangan va lingvistik izlanishlar uchun moslashtirilgan matnlar bazasidir. Korpuslar umumiy, maxsus, parallel, o'quv, diaxron, og'zaki, yozma yoki annotatsiyalangan bo'lishi mumkin. Ularning asosiy vazifasi til birliklarining real qo'llanishini empirik asosda ko'rsatishdir.

Korpusning asosiy afzalligi shundaki, u tilshunosning subyektiv kuzatuviga emas, balki real matn materiallariga tayanadi. Masalan, tadqiqotchi ma'lum bir so'zning qanchalik ko'p ishlatilishini, qaysi janrda faol ekanini, qaysi so'zlar bilan ko'proq birikishini va qanday grammatik shakllarda uchrashini korpus orqali aniqlay oladi. Bu imkoniyat sinonimlarni o'rganishda ayniqsa muhimdir.

Sinonimlar ko'pincha lug'atlarda bir-biriga yaqin ma'noli birliklar sifatida beriladi. Biroq ikki sinonimning lug'aviy yaqinligi ularning nutqda har doim bir-birining o'rnini bosa olishini anglatmaydi. Masalan, o'zbek tilida “kuchli shamol” deyish tabiiy, ammo “qattiq shamol” ham ayrim kontekstlarda uchrashi mumkin; “qattiq sovuq” tabiiy bo'lsa, “kuchli sovuq” ham ishlatiladi, lekin har ikkisi bir



xil uslubiy qiymatga ega emas. Bunday farqlarni korpus yordamida aniqroq ko‘rish mumkin.

Koreys tilida ham sinonimlarning kontekstual farqi kuchli. Masalan, 시작하다 va 개시하다 umumiy “boshlamoq” ma’nosiga ega bo‘lsa-da, 시작하다 kundalik va neytral nutqda keng ishlatiladi, 개시하다 esa rasmiy va kitobiy kontekstlarda faolroq uchraydi. Agar bu birliklar faqat lug‘atda “boshlamoq” deb berilsa, o‘rganuvchi ularning uslubiy farqini sezmasligi mumkin. Korpus esa ularning real matnlardagi farqini ko‘rsatadi.

Demak, sinonimlarni korpus asosida o‘rganish ularning ma’nodoshligini emas, balki haqiqiy qo‘llanish tizimini aniqlashga xizmat qiladi. Bu esa leksikografiya, tarjimashunoslik, til o‘qitish va tabiiy tilni qayta ishlash uchun muhim ilmiy asos yaratadi.

## **2. Koreys tilida Sejong korpusi va sinonimlarni berish imkoniyatlari**

Koreys tilida Sejong korpusi milliy til resursi sifatida alohida ahamiyatga ega. Ushbu korpus koreys tilining yozma va og‘zaki matnlarini, turli janr hamda uslublarini qamrab olgani sababli leksik birliklarning real qo‘llanishini tahlil qilish imkonini beradi. Sejong korpusi asosida so‘zlarning chastotasi, morfologik shakllari, gapdagi vazifasi, birikish xususiyatlari va kontekstual ma’nolari o‘rganilishi mumkin.

Sinonimlarni o‘rganishda Sejong korpusi an’anaviy lug‘at tahlilidan ancha keng imkoniyat yaratadi. Masalan, koreys tilida 말하다, 이야기하다, 언급하다, 진술하다 birliklari umumiy “gapirmoq/aytmoq” semantik maydoniga mansub. Lug‘atda ularning barchasi “aytmoq” yoki “gapirmoq” tarzida izohlanishi mumkin. Ammo korpus materiallari ularning qo‘llanish muhiti farqli ekanini ko‘rsatadi. 말하다 umumiy va neytral, 이야기하다 suhbat yoki hikoya qilishga yaqin, 언급하다 ma’lum mavzuni tilga olishni bildiradi, 진술하다 esa huquqiy yoki rasmiy bayonot kontekstida qo‘llanadi.

Sejong korpusining yana bir afzalligi shundaki, u koreys tilidagi hurmat shakllari bilan bog‘liq sinonimik munosabatlarni o‘rganishga imkon beradi. 먹다 va 드시다, 자다 va 주무시다, 죽다 va 돌아가시다 birliklari semantik jihatdan yaqin bo‘lsa-da, ularning tanlanishi ijtimoiy munosabat va hurmat darajasiga



bog‘liq. Korpusdagi misollar bunday birliklarning qaysi nutqiy vaziyatlarda, qanday subyektlar bilan va qaysi uslubda ishlatilishini aniqlashga yordam beradi. Koreys tilidagi sinonimlarni korpusda berishda etimologik qatlam ham muhimdir. Sof koreyscha birliklar, xitoycha-koreyscha so‘zlar va zamonaviy o‘zlashmalar bir sinonimik qator ichida uchrashi mumkin. Masalan, 시작하다 va 개시하다, 사람 va 인간 kabi birliklar o‘rtasidagi farq ko‘pincha rasmiylik va uslubiy registr bilan bog‘liq. Korpusda bunday birliklarni “neytral”, “rasmiy”, “kitobiy”, “so‘zlashuv”, “hurmat shakli” kabi teglar bilan belgilash ularning qo‘llanishini aniqroq ko‘rsatadi.

Shu sababli Sejong korpusi koreys tilidagi sinonimlarni nafaqat semantik jihatdan, balki uslubiy, pragmatik va korpusiy parametrlar asosida ham tavsiflash imkonini beruvchi muhim manbadir.

### **3. O‘zbek tili milliy korpusida sinonimlarni tavsiflash zarurati**

O‘zbek tili milliy korpusi o‘zbek tilining zamonaviy holatini elektron matnlar asosida o‘rganishga xizmat qiluvchi muhim resursdir. O‘zbek tilida sinonimik qatorlar juda boy va ko‘p qatlamli bo‘lgani sababli ularni korpusda to‘g‘ri berish muhim vazifalardan biri hisoblanadi. Ayniqsa, o‘zbek tilining turkiy, arabcha, fors-tojikcha, ruscha va xalqona qatlamlari sinonimik munosabatlarni rang-barang qiladi.

O‘zbek tilida sinonimlarni korpusda tavsiflashda, avvalo, har bir birlik lemma sifatida aniq belgilanishi kerak. Masalan, yuz, bet, aft, chehra, jamol, oraz birliklari bir sinonimik qatorga mansub bo‘lishi mumkin. Ammo ularning har biri mustaqil lemma sifatida kiritilishi, so‘ngra umumiy semantik maydon doirasida bog‘lanishi lozim. Bunda yuz neytral, bet so‘zlashuvga yaqin, aft salbiy bo‘yoqli, chehra badiiy, jamol va oraz poetik-kitobiy birlik sifatida teglanishi mumkin.

O‘zbek korpusida sinonimlarning uslubiy xususiyatlarini berish ayniqsa muhimdir. Chunki o‘zbek tilida ko‘plab sinonimlar aynan uslubiy bo‘yoqdorligi bilan farqlanadi. Masalan, o‘lmoq, vafot etmoq, dunyodan o‘tmoq, jon bermoq birliklari bir umumiy semantik maydonga mansub. Ammo o‘lmoq neytral yoki bevosita, vafot etmoq rasmiy-hurmatli, dunyodan o‘tmoq yumshoq va ehtiromli, jon bermoq esa badiiy-dramatik ohangga ega. Korpusda bunday farqlar ko‘rsatilsa, foydalanuvchi birliklarni to‘g‘ri tanlay oladi.



Korpusda sinonimlarning kollokatsion muhiti ham alohida ahamiyatga ega. Masalan, o‘zbek tilida chuqur bilim, keng dunyoqarash, og‘ir kasallik, keskin tanqid, kuchli shamol kabi birikmalar tabiiy hisoblanadi. Bu yerda sifatlar bir-biriga yaqin bo‘lsa-da, ularni har doim almashtirish mumkin emas. Korpusdagi kollokatsion tahlil sinonimlarning qaysi so‘zlar bilan tabiiy birikishini aniqlash imkonini beradi.

O‘zbek tili milliy korpusida sinonimlarni keng ko‘lamda teglash va qidiruv tizimi bilan bog‘lash kelajakda elektron sinonim lug‘atlar, o‘quv platformalari, tarjima dasturlari va sun‘iy intellekt asosidagi til texnologiyalari uchun muhim zamin yaratadi.

#### **4. Sinonimlarni korpusda berishning asosiy prinsiplari**

Sinonimlarni korpusda berish ko‘p bosqichli va tizimli jarayon hisoblanadi. Bunda birinchi asosiy prinsip leksik identifikatsiya prinsipidir. Har bir sinonim mustaqil leksik birlik sifatida aniqlanishi, uning lemma shakli, so‘z turkumi, grammatik shakllanishi va asosiy ma‘nosi kiritilishi kerak. Bu bosqichsiz sinonimlarni korpusda to‘g‘ri qidirish va solishtirish mumkin emas.

Ikkinchi prinsip dominanta markazlashuvi prinsipidir. Har bir sinonimik qator uchun markaziy, neytral va keng qo‘llanadigan birlik belgilanadi. Masalan, koreys tilida 말하다 qatori uchun 말하다 dominanta bo‘lishi mumkin; o‘zbek tilida chiroyli, go‘zal, suluv, ko‘rkam qatorida chiroyli yoki go‘zal dominanta vazifasini bajarishi mumkin. Dominanta boshqa a‘zolari tushuntirishda asosiy tayanch bo‘ladi.

Uchinchi prinsip semantik differensiallashuv prinsipidir. Har bir sinonimning qator ichidagi farqlovchi semasi ko‘rsatilishi lozim. Masalan, koreys tilida 보다 umumiy qarash, 바라보다 ma‘noli yoki uzoqqa qarash, 쳐다보다 tik yoki yuqoriga qarash, 지켜보다 esa kuzatish va nazorat ma‘nosini bildiradi. O‘zbek tilida yig‘lamoq, bo‘zlamoq, ingramoq, ho‘ngramoq qatorida farq yig‘lashning kuchi, ovoz darajasi va emotsional holati orqali yuzaga chiqadi.

To‘rtinchi prinsip uslubiy teglash prinsipidir. Sinonimlar neytral, rasmiy, badiiy, poetik, so‘zlashuv, salbiy bo‘yoqli, hurmat shaklidagi, kitobiy kabi teglar bilan tavsiflanishi kerak. Bu ayniqsa koreys va o‘zbek tillari uchun muhim, chunki har ikki tilda sinonimlarning uslubiy farqlari kuchli.



Beshinchi prinsip kontekstual misol bilan tasdiqlash prinsipidir. Har bir sinonimning tavsifi real matndan olingan misol bilan berilishi lozim. Korpusiy misollar soʻzning haqiqiy qoʻllanishini koʻrsatadi. Lugʻaviy izoh foydalanuvchiga umumiy maʼlumot bersa, kontekstual misol birlikning nutqdagi vazifasini ochib beradi.

Oltinchi prinsip kollokatsion muhitni koʻrsatish prinsipidir. Sinonimning qaysi soʻzlar bilan tez-tez birikishi, qaysi birikmalari tabiiy yoki cheklangan ekani korpus asosida aniqlanishi kerak. Yettinchi prinsip esa chastota va janr boʻyicha tavsiflashdir. Yaʼni sinonim qaysi janrda, qaysi matn turida va qanday chastotada uchrashi koʻrsatilishi lozim.

### **5. Teglash va qidiruv mexanizmlarining ahamiyati**

Korpusda sinonimlarni samarali berish uchun teglash tizimi mukammal boʻlishi kerak. Teglash — bu matndagi birliklarga grammatik, semantik, uslubiy yoki pragmatik belgilar biriktirish jarayonidir. Sinonimlarni teglash orqali foydalanuvchi ularni faqat shakl boʻyicha emas, balki maʼno, uslub va vazifa boʻyicha ham qidirishi mumkin.

Masalan, foydalanuvchi koreys korpusida “qaramoq” semantik maydoniga oid birliklarni qidirganda 보다, 바라보다, 쳐다보다, 지켜보다 kabi feʼllar bir guruhda koʻrsatilishi mumkin. Agar ular semantik teglar bilan boyitilgan boʻlsa, foydalanuvchi qaysi birlik umumiy, qaysi biri nazorat, qaysi biri tik qarash, qaysi biri uzoqqa qarash maʼnosini bildirishini koʻra oladi.

Oʻzbek korpusida ham shunday imkoniyat zarur. Masalan, “yuz” semantik maydoni qidirilganda yuz, bet, aft, chehra, jamol, oraz birliklari chiqishi, ularning har biri uslubiy tegi bilan koʻrsatilishi kerak. Shunda foydalanuvchi neytral, salbiy, badiiy yoki poetik birliklarni ajrata oladi.

Qidiruv tizimida sinonimlar orasidagi munosabatlar koʻrsatilsa, korpusdan foydalanish ancha qulaylashadi. Foydalanuvchi bir soʻzni qidirganda uning sinonimik qatori, dominantasi, tipik kollokatsiyalari, kontekstual misollari va janriy faolligini koʻrishi mumkin. Bu oddiy qidiruvdan farqli ravishda, til birliklarining tizimli tahlilini taʼminlaydi.

Bunday tizim til oʻrganuvchilar, tarjimonlar, leksikograflar, tadqiqotchilar va kompyuter lingvistikasi mutaxassislari uchun foydalidir. Ayniqsa, koreys-oʻzbek



ikki tilli korpuslari va elektron lug‘atlari yaratilganda sinonimlarni bunday teglash ilmiy-amaliy jihatdan katta ahamiyat kasb etadi.

## **6. Sinonimlarning korpusiy tavsif modeli**

Koreys va o‘zbek tili korpuslarida sinonimlarni tavsiflash uchun maxsus model ishlab chiqish zarur. Bunday model har bir sinonimni bir nechta parametr asosida ifodalashi kerak. Eng avvalo, birlikning lemma shakli, so‘z turkumi, asosiy ma’nosi va sinonimik qatori beriladi. Keyin dominantaga munosabati, semantik farqlovchi belgisi, uslubiy tegi, kollokatsion muhiti, janriy faolligi va real kontekstdagi misollari ko‘rsatiladi.

Masalan, koreyscha 바라보다 birlikini tavsiflashda quyidagi ma’lumotlar berilishi mumkin: lemma - 바라보다; so‘z turkumi - fe‘l; sinonimik qator - 보다, 쳐다보다, 지켜보다; dominantaga munosabati - 보다 ga nisbatan hosila yoki aniqroq ma’noli birlik; semantik farqi - uzoqqa, ma’noli yoki umid bilan qarash; uslubiy tegi - neytral/badiiy; tipik kollokatsiyasi - 하늘을 바라보다, 미래를 바라보다; misol - 그는 창밖을 바라보았다.

O‘zbekcha chehra birlikini tavsiflashda esa quyidagi model qo‘llanishi mumkin: lemma - chehra; so‘z turkumi - ot; sinonimik qator - yuz, bet, aft, jamol, oraz; dominantaga - yuz; semantik farqi - inson yuzining badiiy va ijobiy ifodasi; uslubiy tegi - badiiy/kitobiy; tipik kollokatsiyasi - nurli chehra, ochiq chehra; misol - Uning chehrasida samimiy tabassum paydo bo‘ldi.

Bu model sinonimlarni oddiy ro‘yxat sifatida emas, balki real nutqda ishlaydigan semantik-uslubiy birliklar sifatida ko‘rsatadi. Natijada korpus foydalanuvchisi so‘zning qaysi vaziyatda ishlatilishini, qaysi birliklar bilan birikishini va boshqa sinonimlardan qanday farqlanishini aniq tushuna oladi.

Korpusiy tavsif modeli, ayniqsa, ikki tilli lug‘atlar va tarjima dasturlari uchun muhim. Chunki sinonimlarni noto‘g‘ri tanlash tarjimada uslubiy xatolarga, ma’no noaniqligiga yoki pragmatik nomuvofiqlikka olib kelishi mumkin. Korpus asosidagi model bunday xatolarni kamaytirishga xizmat qiladi.

## **Xulosa**

Koreys va o‘zbek tili lingvistik korpuslarida sinonimlarni berish masalasi zamonaviy tilshunoslikning muhim yo‘nalishlaridan biridir. Korpus lingvistikasi



sinonimlarni faqat lugʻatdagi maʼnodosh birliklar sifatida emas, balki real nutqda faoliyat koʻrsatuvchi, maʼno, uslub, kollokatsiya, chastota va pragmatik moslik bilan belgilanadigan birliklar sifatida tahlil qilish imkonini beradi.

Sejong korpusi koreys tilidagi sinonimlarning uslubiy, pragmatik va etimologik farqlarini aniqlashda muhim manba boʻlib xizmat qiladi. Oʻzbek tili milliy korpusi esa oʻzbek tilidagi boy sinonimik qatorlarni tizimli ravishda belgilash va ularning real qoʻllanishini koʻrsatish imkonini beradi. Har ikki korpusda sinonimlarni berishda leksik identifikatsiya, dominanta markazlashuvi, semantik differensiallashuv, uslubiy teglash, kontekstual misollar, kollokatsion muhit va chastota-janr koʻrsatkichlari asosiy mezon sifatida olinishi zarur.

Sinonimlarning korpusiy tavsifi elektron leksikografiya, tarjima amaliyoti, til oʻqitish, ikki tilli lugʻatlar yaratish va tabiiy tilni qayta ishlash tizimlari uchun katta amaliy ahamiyatga ega. Shunday qilib, koreys va oʻzbek tillarida sinonimlarni korpus asosida oʻrganish ushbu tillarning leksik-semantik tizimini chuqurroq anglashga hamda ularni raqamli muhitda samarali ifodalashga xizmat qiladi.

### **Foydalanilgan adabiyotlar roʻyxati**

1. Sinclair J. Corpus, Concordance, Collocation. – Oxford: Oxford University Press, 1991.
2. McEnery T., Wilson A. Corpus Linguistics: An Introduction. – Edinburgh: Edinburgh University Press, 2001.
3. Im Ji-ryong. Hangugeo Eohwi Uimiron. – Seul: Taekhaksa, 2002. – 328 b.
4. Kim Gwang-hae. Gugeo Eohwiron Gaeseol. – Seul: Jipmundang, 2008.
5. Kang Beom-Mo, Kim Hung. A Frequency Analysis of Korean by the Sejong Corpus // Journal of Korean Linguistics. – 2004. – Vol. 44.