

TEZAURUS VA KOMPYUTER TEXNOLOGIYALARIGA DOIR

Suyunov Baxodir Turdiyevich

suyunovbahodir4@mail.ru

Mirzo Ulug'bek nomidagi O'zbekiston Milliy universiteti
2-kurs doktoranti, filologiya fanlari bo'yicha falsafa doktori (PhD)

Annotatsiya. Mazkur maqolada, tezaurus va uning turlari, shuningdek lug'atchilikka kompyuter texnologiyalarini tatbiq etish xususida fikr-mulohaza yuritilgan.

Muallif mavzuni yoritishda o'zbek va rus tillaridagi ilmiy manbalar hamda turli xildagi lug'atlar va ilmiy tadqiqot ishlaridan maqsadli foydalangan. O'zni bilan ma'lum tushuncha va hodisalarni tasdiqlovchi misollar va dalillarga murojaat qilib, mavzu yuzasidan tegishli ilmiy xulosalar chiqargan hamda amaliy tavsiyalar bergan.

Ayniqsa, maqolada tezaurus tushunchasini boshqa lug'at turlaridan farqlab olishga alohida e'tibor qaratilgan. Bu esa uning ilmiy-nazariy va amaliy ahamiyatini oshirishga xizmat qiladi.

Kalit so'zlar: *tezaurus, omonim, sinonim, antonim, paronim, leksika, semantika, giponim, giperonim, adekvat, sistem, kompyuter, lingvistika, korpus, injenering.*

Tezaurus – hozirgi kundagi zamonaviy tilshunoslikda leksik birliklar o'rtasidagi semantik aloqalar – omonim, sinonim, antonim, paronim, giponim, giperonim va boshqa tushunchalarni ko'rsatadigan umumiy yoki maxsus lug'at, shunday lug'atlarning maxsus turi hisoblanadi. Bu lug'at boshqa lug'atlardan farqli o'laroq, biror bir ilm-fan sohasiga oid lug'aviy birliklar yoki biror mavzu tarkibida joylashtirilgan ana shunday birliklar o'rtasidagi semantik munosabatlar aks ettirilgan ideografik lug'atdir.

Unda kerakli so'zlar tushunchaga qarab qidiriladi. Nazariy jihatdan tezaurus leksik-semantik tizimning ehtimoliy modellaridan biridir. Amalda undan individual lug'atni boyitish va tezkor qidiruv vositasi sifatida foydalaniladi. Shu ma'noda tezaurus lug'atning aksi hisoblandi.

Odatda, biror-bir so'zning ma'nolarini emas, balki so'zning o'zini qidirilganda, tezauruslarga murojaat qilinadi. Ya'ni bu yerda tushuncha ma'lum bo'lsa-da, biroq shu tushuncha ostidagi so'zlar guruhi yoki so'z shakllari haqida ma'lumotlar olinadi. Bu tizim ichida turgan bo'lishi mumkin-u, ammo so'zning

o'zi nimaligini bilmasligimiz mumkin. Shu bois, manbalarda ta'kidlanganidek, tezaurus – bu lug'at bo'lib, so'zlar uchun omborxona vazifasini o'taydi.

Tezaurus umumiy ma'noda – maxsus terminologiya, ya'ni lug'at, ma'lumotlar yig'indisi, korpus yoki jamlanma, ma'lum sohadagi bilimlar yoki faoliyat sohasidagi tushunchalar, ta'riflar va terminlarni to'liq o'z ichiga oladi.

Tezaurus – yunoncha so'zdan olingan bo'lib, “*xazina*”, “*boylik*”, “*zaxira*” demakdir. Mazkur tushuncha maxsus bilimlar sohasi yoki faoliyat sohasining tushunchalari, ta'riflari va atamalarini matnlarda ishlatish misollari bilan to'liq qamrab oluvchi ma'lumotlar to'plami hisoblanadi.

Tezaurus muayyan tilda barcha so'zlarni qamrab oladigan, ularning matnda qo'llanish holatlarini to'la-to'kis aks ettiradigan lug'at. Muayyan yozuv yodgorliklaridan leksik birliklarni yoppasiga terib olishga asoslangan (masalan, yunon, lotin tillarida tuzilgan) lug'atlar shunday lug'atlardan hisoblanadi. So'z tanlash tamoyillariga ko'ra yozuvchilar yoki ularning ijodiga mansub biror asar tili bo'yicha tuzilgan lug'atlar ham tezaurus hisoblanadi.

U leksik-semantik, korporativ kommunikatsiya (bir fan yoki kasb orqali o'zaro bog'liq bo'lgan shaxslarning muloqotda bir-birlarini tushunishlari) uchun xizmat qiladi. Tezauruslar ma'lum bir fanni talqin qilishda muhim vositalardan biridir.

Tezaurus tushunchasini yuqoridagi sifatlari bilan birgalikda, lug'atning bir turi deb olsak, unda quyidagicha qiyosiy fikr yuritishimiz mumkin:

Lug'at va tezaurus tushunchalarining har ikkisi so'z va uning ma'nolarini bilish uchun xizmat qilsa-da, biroq ularning so'z ma'nolari haqida ma'lumot berish usullari har xil, ya'ni o'zaro farqlanadi.

Lug'at va tezaurus so'zlari lug'aviy ma'nosiga ko'ra, ot hisoblanadi. Lug'at ko'proq til o'rganuvchilar tomonidan so'zlarning ma'nolari, talaffuzi va orfografiyasini aniqlash uchun ishlatilsa, tezaurus tadqiqotchilar tomonidan so'zning sinonim, antonim, omonim va boshqa ma'no shakllarini topish uchun ishlatiladi.

Tezaurus – so'z yoki terminning tegishli tushunchalar guruhidagi ro'yxati. U murojaat qilinayotgan so'z bilan bir xil bo'lgan boshqa so'zlar to'g'risida ham keng qamrovli ma'lumotlar beradi.

Aksariyat tadqiqotchilar sinonim, antonim, paronim va boshqa shu kabilarni bilish uchun tezaurusdan foydalanishadi. Odatda, tezaurus tilda so'zlarning kelib chiqishiga unchalik bog'liq bo'lmaydi, xususan, so'zlarning etimologiyasi haqida qo'shimcha ma'lumotlar bermaydi. Ammo, tezaurus so'zlarning boshqa shakllari,

masalan, nominal shakllari, sifatlar va adverbial shakllar haqida juda ko'plab ma'lumotlar beradi.

Lug'at – bu tildagi so'zlarni (odatda alifbo tartibida) ro'yxatga oladigan va ularning ma'nosini izohlab beradigan yoki shunday so'zlarni boshqa tilda beradigan, ko'pincha, to'g'ri talaffuz haqida ma'lumot beradigan kitob yoki elektron manba hisoblanadi. Boshqacha qilib ifodalasak, lug'at tildagi so'zlar haqida bilim beradigan etimologik va grammatik vositadir.

Lug'atda so'zning jinsi va nutqda ishlatilishi haqida qo'shimcha ma'lumotlar mavjud. U tildagi birorta so'z bilan bir qatorda, unga qo'shimcha ravishda ma'lumotlar ham berishi mumkin. Lug'at bir nechta aniq tillarda tuzilishi mumkin, masalan, ingliz, fransuz, rus va boshqa tillarda. Bir tilda tuzilgan lug'atda boshqa tillardagi shu so'zning ekvivalent ma'nolari ham beriladi. Odatda, bu sifatlar tezaurusga tegishli emas.

Birinchi zamonaviy ingliz tezaurusi Piter Mark Roger tomonidan 1805-yilda yaratilgan. Mazkur tezaurus 1852-yilda Angliyada nashr etilgan, taxminan 15 mingta tushunchani o'z ichiga oladi va shu vaqtdan beri foydalanib kelinadi. Bunday lug'atlarning asosiy maqsadi matn yozishda tegishli so'zlarni tanlashga yordam berish, mavzu doirasida atama yoki termin o'rtasidagi munosabatlarni tavsiflovchi ma'lumotlarni qidirib topishdan iborat. Bunday lug'atlar ma'lum bir soha mutaxassislari tomonidan yaratilgan va mavzu doirasidagi ma'lumotni qidirib topish uchun mo'ljallangan.

1870-yillarda tezauruslardan ma'lumot izlash ishlarida faol foydalanila boshlandi. Shunday lug'atlardan yana biri Vikilug'at deb ataladi. U ko'p tilli bepul yangilanadigan lug'at va tezaurus bo'lib, “Vikimediya” fondi loyihasi doirasida yaratilgan va 2004-yildan hozirgi kungacha ishlatilib kelinadi. Unda 250 mingta tushuncha va 67 mingta semantik munosabatlar o'z aksini topgan.

Rus tilidagi xuddi shunday tezauruslardan biri “RuTez” deb nomlangan. Bu lug'at 1997-yilda Axborot tadqiqotlar markazi tomonidan avtomatik indeksatsiya vositasi sifatida yaratilgan. U hozirgi kunga qadar ishlab chiqilgan 45 mingta tushuncha, 107 mingta so'z va iboralar, 177 mingta semantik munosabatlarni qamrab oladi.

O'zbekiston Milliy Ensiklopediyasida tezaurus terminiga quyidagicha izoh berilgan:

“Tezaurus (yun. thesaurus – “xazina”, “boylik”) – 1) muayyan tildagi barcha so'zlarni qamrab oladigan, ularning matnda qo'llanish holatlarini to'la-to'kis aks ettiradigan lug'at” [O'zbekiston Milliy Ensiklopediyasi, 2004: 329].

Tezaurus – bu lug‘atlar tushunchasi va uning birliklari o‘rtasidagi sobit semantik aloqalar bo‘lib, har ikkala mustaqil xususiyat ham tushunchani aniqlash uchun muhim ahamiyat kasb etadi.

“Tezaurus termini, avvalambor tilning lug‘at fondini maksimal darajada to‘liqlik bilan ifodalaydigan xazina sifatida talqin etiladi. Matnlarning adekvat talqini va ma‘lumotlari uning sistem ma‘nolarini o‘zida mujassam etgan va soha vakillari tomonidan e’tirof etilgan tezaurusda aniqlanadi” [Valitova, 2012: 19].

Hozirgi vaqtda ushbu tushunchaga kompyuter texnologiyalarida so‘zlarni qidirishda axborot texnologiyalarining tarqalishi, Internet va mashinalarning tarjimasi bilan bog‘liq lug‘atlar kiradi. Psixologiyada shaxsning tezaurusi ma‘lumotni idrok etish va tushunish bilan xarakterlanadi. Bunda aloqa nazariyasi va uning elementlari o‘zaro ta’sir ko‘rsatadigan murakkab tizimning umumiy tushuncha-tezislari ko‘rib chiqiladi.

Kompyuter lingvistikasining tarkibiy qismi bo‘lgan tezauruslar leksemalarning o‘z va ko‘chma ma‘nolari, kommunikativ xususiyatlari, emotsional-ekspressiv vazifalari haqida keng ma‘lumot berishi bilan qimmatlidir. Shu o‘rinda ta’kidlash joizki, jahon tilshunosligida kompyuter lingvistikasi, korpus tushunchasiga doir ilk ma‘lumotlar XX asrning qirqinchi yillarida yuzaga kelganligi ilmiy manbalarda qayd etilgan [Kutuzov]. O‘zbek tilshunosligida korpus lingvistikasi masalalari keyingi yillarda ilmiy tadqiqot ishlari sifatida keng o‘rganila boshladi.

Xususan, Sh.M. Hamroyevaning “O‘zbek tili mualliflik korpusini tuzishning lingvistik asoslari” nomli dissertatsiyasida o‘zbek tilshunosligida birinchi marta korpus, uning o‘ziga xos xususiyatlari, nazariy asoslari, til korpusining lingvistik hamda nazariy va amaliy ahamiyati yoritib berilgan [Hamroyeva, 2018:15]. Shuningdek, korpus lingvistikasining shakllanish tarixi, taraqqiyot yo‘li, o‘ziga xos xususiyatlari va bugungi holati masalalari atroflicha talqin qilingan.

O‘zbek tilini jahon tillari qatorida rivojlantirish, uni dunyoviy tillar sirasiga kiritish va til o‘rganish hamda o‘rgatishda kompyuter lingvistikasi fani dolzarb ahamiyat kasb etmoqda. Texnika taraqqiyoti mahsuli bo‘lgan kompyuter tizimi barcha sohalarida qulayliklar yaratadi, ma‘lumotlarning tezkor yetkazib berilishini, tarjima, tahrir jarayonlarining mashina yordamida qisqa muddatlarda amalga oshirilishini ta’minlaydi.

Kompyuter lingvistikasi amaliy tilshunoslik yo‘nalishlaridan biridir. “Amaliy lingvistika” termini esa ko‘p ma’noli termin bo‘lib, bu termin umumtilshunoslikda har xil talqin qilinadi. G‘arb tilshunosligida *applied linguistics*,

angewandte linguistik termini, birinchi navbatda, o'qitish metodikasi, grammatik xususiyatlarni qamrab olgan holda, ona tili va chet tillarini o'qitish amaliyoti bilan bog'lanadi.

Aslida mazkur tushunchalar kompyuter texnologiyalarining ishlab chiqilishi va axborotni qayta ishlash tizimi (boshqaruvning avtomatlashgan tizimi, informatsion qidiruv tizimi, matnni qayta ishlashning avtomatik tizimi) shakllanishi jarayonida yuzaga kelgan.

Ayrim rus tilidagi adabiyotlarda, ko'p hollarda, “kompyuter lingvistikasi” (“computational linguistics”), “hisoblash lingvistikasi”, “avtomatik lingvistika”, “injener lingvistikasi” tarzida qo'llaniladi [Baranov, 2001:5]. Amaliy lingvistikaning mazkur nomlar bilan yuritilishi uning faoliyati xarakterini belgilaydi. Ushbu terminlarda amaliy lingvistikaning integratsiyaga asoslanishi, avtomatlashish xususiyati o'z ifodasini topgan.

Xulosa. Tezaurus – bu umumiy ma'noda maxsus bilim sohasi yoki faoliyat sohasining tushunchalari, ta'riflari, atama yoki terminlarini to'liq qamrab oladigan maxsus terminologiya, yana-da qat'iy va obyektiv ravishda lug'at, ma'lumotlar to'plami, korpus yoki kod, demakdir.

Tilda tezauruslar elektron formatda, alohida fan sohalarini tavsiflashning samarali vositalaridan biri bo'lib, biror fan yoki kasb-hunar bilan bog'liq bo'lgan munosabatlar va shu munosabatlardagi tushunchalarni rivojlantirishga xizmat qiladi. U nafaqat so'z ma'nolari, balki intellektual tizimlarning ilmiy asoslarini to'ldirishda ishlatilishi mumkin bo'lgan so'zlarni boshqa tushunchalar va ularning guruhlari bilan o'zaro bog'lash orqali ochib berishga imkoniyat yaratadi. Tildagi axborotlar bazasida tezaurus termini subyekt ega bo'lgan barcha ma'lumotlarning umumiylikni ifodalaydi.

Amaliy tavsiyalar. Lingvistik masalalarni tezkor hal etishda, kompyuter texnologiyalari va informatikaning o'рни beqiyosdir. Ularning milliy tilshunoslikka tatbiq etilishi axborotlarning tezkorlik bilan o'zlashtirilishini, shuningdek bajariladigan amallarning aniqligini ta'minlaydi. Tilshunoslik muammolarining kompyuter yordamida hal qilinishi, ayni paytda, tilning qo'llanish doirasi kengayishi, ichki imkoniyatlarning ortishiga xizmat qiladi.

Kompyuter lingvistikasi har bir tilning o'ziga xos tabiatini hisobga olib, muayyan tilning fonetik, leksik, grammatik sathlariga oid masalalarni mashina yordamida hal qilish vazifasini qo'ygandagina lingvistik taraqqiyot omili bo'la oladi.

Shuningdek, turli xil lug'atlar va tezauruslarga ham kompyuter

texnologiyalarini tatbiq etish bugungi kundagi tilshunoslikning dolzarb vazifalaridan biridir.

Foydalanilgan adabiyotlar:

1. Baranov A. Vvedeniye v prikladnuyu lingvistiku. – Moskva, 2001.
2. Kutuzov A.B. Kurs “Korpusnaya lingvistika”. Litsenziya Creative commons Attribution Share-Alike 3.0 Unported - // lab314.brsu.by/kmp-lite/ kmp-video/ CL/ CorporeLingva.pdf
3. Valitova N.R. Formirovaniye professionalnogo tezaurusa u studentov fizkulturnogo vuza. Avtoref. na soiskaniye uchyonoy stepeni kand.ped.nauk. – Omsk, 2012.
4. Mahkamov N. Terminologik tamoyillar va xalqaro terminelementlar. // Xorijiy so‘z va terminlardan foydalanishda me‘yor va milliy-assotsiativ fikrlash muammolari. Toshkent, 2011
5. Mirahmedova Z. O‘zbek tilining anatomiya terminologiyasi va uni tartibga solish muammolari. – Toshkent: “Fan”, 2010
6. Rasulov R., Suyunov B., Mo‘ydinov Q. Nutq madaniyati va notiqlik san’ati. Toshkent, 2010
7. Eshmo‘minov. A. O‘zbek tili milliy korpusining sinonim so‘zlar bazasi. Filol. f. bo‘yicha f.d.(PhD) diss. avtoref. – Qarshi: 2019.
8. O‘zbekiston Milliy Ensiklopediyasi. – Toshkent: 2004
9. O‘zbek tilining izohli lug‘ati. – Toshkent: “O‘zbekiston milliy ensiklopediyasi” – 2006
10. Hamroyeva Sh.M. O‘zbek tili mualliflik korpusini tuzishning lingvistik asoslari. Filol. f. bo‘yicha f.d.(PhD) diss. avtoref. – Qarshi: 2018.