

TIL KORPUSIDA SODDA GAPLARNI AVTOMATIK ANIQLASH

Nizomova Fotima Bo'rixo'ja qizi

ToshDO'TAU Kompyuter lingvistikasi yo'nalishi 3-kurs talabasi

Qodirov Kamoliddin Qodirjon o'g'li

ToshDO'TAU Kompyuter lingvistikasi yo'nalishi 3-kurs talabasi

Annotatsiya. Ushbu maqolada til korpuslari haqida ma'lumot berilgan. Sodda va qo'shma gap turlari va ularning strukturalari orasidagi farqlar va ularni ajratish uchun qoidalar ishlab chiqilgan. Til korpusidan sodda gaplarni avtomatik aniqlash uchun farqlar asosida qoidalar ishlab chiqish mumkinligini izohlab misollar keltirilgan, kompyuter lingvistikasidagi model va model tushunchasiga izoh, til korpuslaridan sodda gaplarni aniqlashda tokenizatsiyaning maqsadlari haqida aytib o'tilgan.

Abstract. Information about language corpora is given in three. Help is provided for the types and structures of simple sentences and their differences. The production of simple sentences from the corpus of the language for automatization based on the differences is explained with examples, the model of computational linguistics and an explanation of the model, the sentences of the token in the simplified methods from the corpus of the language are mentioned.

Аннотация. Информация о языковых корпусах представлена в трех. Помощь оказывается по типам и структурам простых предложений и их различиям. Построение простых предложений из корпуса языка для автоматизации на основе различий поясняется на примерах, модели компьютерной лингвистики и пояснениями к модели, предложения токена в упрощенных методах из корпуса языка упомянут.

Kalit so'zlar: *Til korpuslari, sodda va qo'shma gap strukturasi, model, tokenizatsiya*

Kirish

Til korpusi – ta'lim samaradorligi va milliy ma'naviyatimiz ravnaqi garovi. Milliy til korpusi – milliy til birliklarining xususiyatlarini aniqlash maqsadida elektron qidiruv imkoniyati mavjud tizim, tabiiy tilning raqamlashgan yozma va og'zaki matnlari jamlanmasi. Milliy korpus – tilni tadqiq etish (til birligining o'zgarishi, eskirishi, yangilarining paydo bo'lishi, ma'nosining kengayishi va torayishi; yangi iboralarning paydo bo'lishini kuzatish), o'rganish, an'anaviy va zamonaviy lug'atlar tuzishda keng imkoniyatli dasturlashtirilgan tizim. Shu bois Milliy korpus milliy til xazinasidir.

Milliy korpusdan lingvistlar, leksikograflar, kompyuter lingvistlari, dasturchilar, muharrirlar, tarjimonlar, jurnalistlar, noshirlar, olimlar, o'qituvchilar, ta'lim oluvchilar va boshqa har qanday soha mutaxassisi keng foydalanish

imkoniyatiga ega bo'ladi. Asrimizda korpussiz na nazariy, na amaliy filologiya taraqqiy etadi.

Foydalanishda deyarli kasbiy tabaqalanishga yo'l qo'ymaydigan til korpuslari barcha fan sohasi vakillarini birday qiziqtirishi tabiiy. Ayniqsa, ona tili va chet tillarini o'qitish va o'rganish borasida korpuslarning ahamiyati beqiyos. Bugungi kunda dunyo miqyosida til o'rgatish tizimi korpuslarga yo'naltirilayotganligi ham – fikrimizning dalili. Shuning uchun ham ta'lim korpuslari, sheva matnlari korpuslari, mualliflik korpuslari, poetik matnlar korpuslari, og'zaki, ilmiy, rasmiy matnlar korpuslari, gazeta matnlari korpuslari kabi qator mikrokorpuslarning tuzilayotganligi ahamiyatli.

O'zbek tili milliy korpusining ilk shakli interfeys, qidiruv tizimi, 200 mln ga yaqin so'zshakl, 100 mingdan ortiq matnlar bazasi va 20 dan ortiq elektron lug'at manbasidan iborat. Shuningdek, o'zbek tili milliy korpusi saytining birinchi sahifasida korpus va uning tuzuvchilari haqida asosiy ma'lumot, o'ng tomondagi menyuda istalgan sahifaga o'tish imkoniyati mavjud. Korpus menyusi to'rt qismdan iborat. Bosh sahifa, saytning qidiruv oynasi, matn haqida unga biriktirilgan qo'shimcha ma'lumot ilovasi, qidiruv uchun til birliklariga izoh yozish parametrlari, elektron leksikografik manbalar tavsifi, oxirgi blok korpus tuzuvchilari jamoasi, foydalanilgan dastur, matnlarning mualliflik huquqi haqida to'liq ma'lumotlar bazasidan iborat.

Korpus bo'yicha qidiruv foydalanuvchiga quyidagilarni aniqlash imkonini beradi:

- 1) muayyan so'zning barcha shakllari misollar massivi bilan;
- 2) misollarning qaysi manbadan olinganligini ko'rish;
- 3) so'zning filologik lug'atlardagi mohiyati va izohini;
- 4) qidiruvga berilgan so'z bilan uning o'ng va chap tomonidan birikish imkoniyatiga ega so'zlar ro'yxatini;
- 5) u yoki bu muallifning ayni so'zdan foydalanish chastotasi yoki statistikasini;
- 6) so'zning o'z va ko'chma ma'nolarini;
- 7) so'z qo'llanishining yashirin modeli (imkoniyati)ni;
- 8) til taraqqiyotining turli davrida qo'llanish holatini.

Til korpuslari – til bo'yicha tadqiqot va amaliy topshiriqlar yechimi uchun zarur ish quroli. U oddiy elektron kutubxonadan farqlanadi. Elektron kutubxonaning maqsadi – xalqning ijtimoiy-siyosiy, ma'naviy, iqtisodiy hayotini aks ettiruvchi badiiy va publitsistik asarlarni nisbatan to'liq qamrab olish. Elektron kutubxona matnlari til nuqtai nazaridan ishlov berilmaganligi sababli tadqiqotlar uchun noqulaylik tug'diradi. Chunki elektron kutubxona ilmiy tadqiqot materiali bazasini tayyorlash maqsadida tuzilmaydi, balki milliy-ma'naviy merosni to'plashni maqsad qilgan bo'ladi. Til korpusi esa elektron kutubxonadan farqli o'laroq, tilni o'rganish va tadqiq qilish uchun zarur, foydali va qiziqarli matnlarni to'plashni nazarda tutadi.

Til tadqiqi hayotdan uzilib qolmasligi uchun muntazam ravishda matnlar bazasi yangilanib boriladigan korpusga tayanish maqsadga muvofiq. Shu bois korpus lingvistikasi laboratoriyasini tashkil etib, doimiy ekspertlar jamoasi faoliyatini yo'lga qo'yish bilan Milliy korpusning taraqqiy etishi va manbalari boyib borishiga zamin yaratiladi.

Milliy til korpusi – ona tilimizning jahonga yuz tutishiga erishish milliy ma'naviyatni takomillashtirish va yuksaltirishning ham asosiy vositalaridan biri.

Gaplarning strukturasiga ko'ra turlari

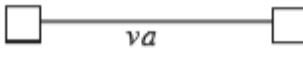
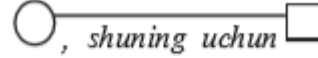
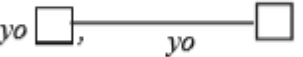
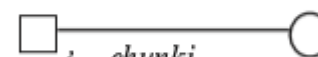
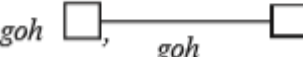
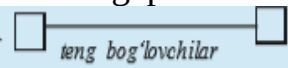
Har qanday gapning tuzilish jihatidan turlarini uning grammatik asosi belgilab beradi. Gapning grammatik asosini bosh bo'laklar: ega va kesim tashkil etadi. Gap tarkibidagi grammatik asos (ega va kesim) ning ishtirokiga ko'ra gaplar tuzilish jihatidan ikki turga bo'linadi: sodda gaplar va qo'shma gaplar. Bitta grammatik asosiga ega bo'lib, ma'lum bir fikr va his-hayajonni ifodalovchi gaplar sodda gaplar deyiladi. Ikki va undan ortiq grammatik asosdan iborat bo'lib, nisbatan murakkabroq fikrni ifodalovchi gaplar qo'shma gap deyiladi.

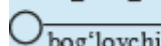
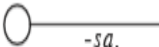
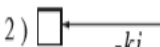
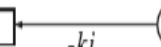
Gapning tuzilishiga ko'ra turlari jadvali:

Predikativ markaz-larning miqdoriga ko'ra - Sodda gap - Qo'shma gap	Bosh bo'laklarning ishtirokiga ko'ra - Bir bosh bo'lakli gap - Ikki bosh bo'lakli gap	Ikkinchi darajali bo'laklarning ishtirokiga ko'ra - Yig'iq gap - Yoyliq gap	Zaruriy bo'laklar ishtirok etmasligiga ko'ra - To'liq gap - To'liqsiz gap	Bo'laklarga bo'linish-bo'linmasligiga ko'ra - Bo'laklarga bo'linadigan gaplar - Bo'laklarga bo'linmaydigan gaplar (so'z gaplar)
---	---	---	---	---

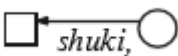
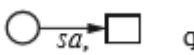
Sodda va qo'shma gap strukturalari orasidagi farqlar va ularni aniqlash uchun qoidalar

Sodda gap	Qo'shma gap
Bitta kesimlik belgisiga ega	Birdan ortiq kesimlik belgisiga ega
Mazmuniy va ohang tugalligiga ega bo'lgan gap	Ikki yoki undan ortiq sodda gapning o'zaro grammatik va mazmuniy munosabatidan tashkil topgan, ohang tugalligiga ega bo'lgan butunlik
Sodda gapda to'liq shakllangan kesim bitta bo'ladi.	sodda gaplar qo'shma gap qismlari sanaladi.
Ega va hol to'g'ridan to'g'ri kesimga bog'lanadi. Aniqlovchi ot bilan ifodalangan har	1. Bog'lovchilar: a) teng bog'lovchilar (va, hamda, ammo, lekin, biroq, yo, yo..., yo, dam..., dam...); b) ergashtiruvchi bog'lovchilar. 2. Bog'lovchi vazifasidagi vositalar: a) bo'lsa, esa, deb; b) –u (-yu), -da, -ki yuklamalari; d) nisbiy so'zlar; e) shuning uchun, shu

qanday bo‘lakka bog‘lanaveradi. To‘ldiruvchi esa fe‘l bilan ifodalanagan bo‘laklarga bog‘lanadi	tufayli, shu sababli singari ko‘makchili qurilmalar; f) shart mayli qo‘shimchasi (-sa). 3. Ohang. 1)  4)  2)  5)  3) 
Kesim ifodalanishiga ko‘ra fe‘l kesim va ot kesimga ajraladi. Fe‘l kesim sof fe‘l va ba‘zan ravishdosh bilan ifodalanadi:	Qismlari teng bog‘lovchilar, bo‘lsa, esa so‘zlari, -u (-yu), -da yuklamalari yordamida bog‘langan gaplar bog‘langan qo‘shma gaplar hisoblanadi. Ularning qolipi  orqali ifodalanadi.
Kesimlar tuzilishiga ko‘ra ikki xil bo‘ladi: sodda kesim va murakkab kesim	Zidlov bog‘lovchilari qo‘shma gapning ikkinchi qismi boshida qo‘llanadi va qo‘shma gap qismlari vergul bilan ajratiladi.
Bir so‘z bilan ifodalgan kesim sodda kesimdir: Paxta va tilladan yurt bezanadi. Rost gapir. (O‘.Hoshimov). Haqiqat achchiq.	-u(-yu), -da yuklamalari bog‘langan qo‘shma gap qismlarini bog‘lash bilan birga ular o‘rtasida zidlash, ketma-ketlik munosabati mavjudligini bildiradi.
Yordamchi so‘zlar (ko‘makchi fe‘l, bog‘lama, nisbiy so‘z va b.) ham murakkab kesimni shakllantiradi.	Inkor bog‘lovchisi-bog‘langan qo‘shma gap qismlarini bog‘lash bilan birga, ularda ifodalgan harakat-holat yoki belgining inkorini ham bildiradi. Shuning uchun bunday bog‘langan qo‘shma gaplarda kesimlar bo‘lishli shakldagi fe‘llar yoki bor, mavjud kabi so‘zlar orqali ifodalansa-da, mazmun inkor tarzida anglashiladi. Yozuvda takror qo‘llanayotgan bog‘lovchidan oldin vergul qo‘yiladi.
Ot kesim tarkibida kelib, uni ega bilan bog‘lovchi so‘z bog‘lama deyiladi; edi, ekan, emish, bo‘lmoq, hisoblanmoq, sanalmoq, kerak, zarur, darkor, lozim	Bo‘lsa, esa so‘zlari bog‘langan qo‘shma gap qismlarini bog‘lash bilan birga, ular o‘rtasida qiyoslash va zidlash munosabati mavjudligini ifodalaydi. Bo‘lsa, esa so‘zlari qiyoslanuvchi bo‘lakdan so‘ng keladi.

kabi soʻzlar bogʻlama vazifasida qoʻllanadi.	
Kengaygan birikma va ibora yaxlit holda bitta gap boʻlagi boʻlib keladi va oʻz ichida boshqa boʻlaklarga ajratilmaydi. Shuning uchun ular murakkab kesim hisoblanadi	Ergashgan qoʻshma gaplar: Mazmuni izohlanayotgan gap bosh gap, bosh gap mazmunini izohlayotgan gap ergash gap hisoblanadi. Ergashgan qoʻshma gaplar, asosan,  yoki  bogʻlovchi qolipida ifodalanadi.
Ega doimo bosh kelishik shaklida boʻladi	Ega doimo bosh kelishik shaklida boʻladi
1. Ot bilan: Milt-milt yonib turgan chiroq birozdan soʻng oʻchdi. 2. Olmosh bilan: Sen oʻzingni maqtama, seni birov maqtasin. (Maqol) 3. Son bilan: Ikkalasi ham chiqib ketishdi. 4. Sifatdosh bilan: Oʻqigan oʻqdan oʻzar, oʻqimagan turtkidan shoshar. (Maqol) 5. Sifat bilan: Yaxshilar koʻpaysin, yomon qolmasin. (Qoʻshiqdan) 6. Harakat nomi bilan: Oʻqish jafoli, iqlibati vafoli. (Maqol) 7. Taqlid soʻz bilan: Qars ikki qoʻldan chiqadi. (Maqol)	Ergash gaplar bosh gapga quyidagi vositalar yordamida bogʻlanadi: 1) ergashtiruvchi bogʻlovchilar; 2) koʻmakchili qurilmalar (shuning uchun shu sababli, shu tufayli va boshqalar); 3) yuklamalar (-ki); 4) nisbiy soʻzlar (kim – oʻsha, nima —shu, qayer — shu yer va boshqalar); 5) deb soʻzi va boshqalar. 1) Agar  -sa,  2)  -ki,  3)  chunki  4)  shuning uchun 
Atov gap va soʻz gaplar ham shaxsi nomaʼlum gaplar hisoblanadi, sodda gaplar sirasiga kiradi	Bosh gapda ifodalangan mazmunning sababi, maqsadi, sharti yoki qiyosini bildirgan ergash gaplar bosh gapga sabab (chunki, negaki), maqsad (toki), shart (agar, agarda, basharti, mabodo) ergashtiruvchi bogʻlovchilari, shuningdek, qiyoslash-chogʻishtiruv (goʻyo, goʻyoki, xuddi) yuklama-bogʻlovchilari

	yordamida bog‘lanadi. Sabab, maqsad, qiyoslash-chog‘ishtiruv bog‘lovchilari ergash gap boshida qo‘llanadi va ergash gap bosh gapdan so‘ng keladi. Bunday vaqtda yozuvda ergash gap bosh gapdan vergul bilan ajratiladi.
Atov gap bitta so‘zdan iborat bo‘lishi ham, aniqlovchi bilan kengayib kelishi ham mumkin: 1. Bahor. 2. Erta bahor.	Bosh gapda ifodalangan mazmunning sababini bildiruvchi ergash gaplar bosh gapga shuning uchun, shu sababli, shu tufayli singari ko‘makchili qurilmalar vositasida bog‘lanadi. Ular bosh gap tarkibida keladi
So‘z gaplar: Kishilar orasidagi oddiy muomalada tez-tez ishlatib turiladigan so‘z yoki ibora so‘z-gap vazifasida qo‘llanadigan bo‘lib qolgan: assalomu alaykum, yoqimli ishtaha, marhamat, salomat bo‘ling, qulluq, osh bo‘lsin, yaxshi boring, xush ko‘rdik, labbay, jonim bilan, bajonidil kabi.	Bosh gapda ifodalangan mazmunning maqsadini, sababini bildiruvchi ergash gaplar bosh gapga ko‘pincha deb so‘zi yordamida bog‘lanadi
Ega va kesimdan yoki faqat kesimdan iborat gap yig‘iq gap hisoblanadi. Misollar: Men – o‘qituvchi. O‘qituvchiman.	Bosh gap tarkibidagi ko‘rsatish olmoshi ma’nosini izohlash uchun qo‘llangan ergash gaplar bosh gapga -ki yuklamasi yordamida bog‘lanadi. Bu yuklama bosh gap kesimi tarkibida bo‘ladi va ergash gap bosh gapdan vergul bilan ajratiladi
	Ergash gap tarkibida qo‘llanuvchi kim, nima, qancha, qanchalik, qanday, qayer kabi so‘roq olmoshlari va bosh gap tarkibida unga javob bo‘lib keluvchi shu, o‘sha, shuncha, shunchalik, shunday kabi olmoshlar bir-biriga nisbatan qo‘llanganligi, biri ikkinchisini taqozo etganligi uchun nisbiy so‘zlar hisoblanadi. Ergash gapning kesimi shart mayli shaklidagi fe‘llar bilan ifodalanadi.

	Kesim ergash gap bosh gapga ko'rsatish olmoshi tarkibidagi -ki yuklamasi yordamida bog'lanadi. 
	Bog'lovchisiz qo'shma gap qismlarini yozuvda ajratish uchun vergul, tire, ikki nuqta kabi tinish belgilari ishlatiladi. 1. Qo'shma gap qismlari bir paytda yoki ketma-ket sodir bo'ladigan voqea-hodisalarni ifodalasa, ular orasiga vergul qo'yiladi. (Yigitlar daraxtlarning ostini yumshatdilar, qizlar maktab hovlisini supurdilar.) 2. Qo'shma gap qismlari orasida o'xshatish, zidlash, shart kabi munosabatlar ifodalanganda, ular orasiga tire qo'yiladi. (Hamal keldi – amal keldi.) 3. Qo'shma gap qismlari orasida izohlash munosabati ifodalanganda, ular orasiga ko'pincha ikki nuqta qo'yiladi. (Bizning qarorimiz shu: hasharda barchamiz faol ishtirok etamiz.)
	Ergash gapning kesimi shart mayli shaklidagi fe'llar orqali ifodalanganda, shart mayli qo'shimchasi ergash gapni bosh gapga bog'lovchi vosita ham sanaladi. 
	Ko'chirma gapli qo'shma gaplar doimo ikki qismdan – muallif gapi va ko'chirma gapdan tashkil topadi.
	Muallif gapi (M) va ko'chirma gap (K) bir-biriga nisbatan turlicha joylashishi mumkin, bunda tinish belgilari quyidagicha ishlatiladi: 1) M: «K». M: «K?» M: «K!» 2) M: «K», — m. M: «K?» — m. M: «K!» — m. 3) «K», — m. «K?» — m. «K!» — m. 4) «K, — m, — k».
	O'arkibidagi sodda gaplar mazmunan bir-biri bilan bog'liq bo'lib, maxsus bog'lovchi vositalarsiz, faqat ohang yordamida bog'langan qo'shma gaplar bog'lovchisiz qo'shma gaplar sanaladi.
	O'ahlilda bog'lovchisiz qo'shma gaplar bir-biri bilan to'g'ri chiziq bilan bog'langan ikki to'rtburchak  orqali ifodalanadi. Bu bog'lovchisiz qo'shma gap qolipi sanaladi.
	1. Qo'shma gap qismlari bir paytda yoki ketma-ket sodir bo'ladigan voqea-hodisalarni ifodalasa, ular orasiga vergul qo'yiladi. (Yigitlar daraxtlarning ostini yumshatdilar, qizlar maktab hovlisini supurdilar.)

	2. Qo'shma gap qismlari orasida o'xshatish, zidlash, shart kabi munosabatlar ifodalanganda, ular orasiga tire qo'yiladi. (Hamal keldi – amal keldi.) 3. Qo'shma gap qismlari orasida izohlash munosabati ifodalanganda, ular orasiga ko'pincha ikki nuqta qo'yiladi. (Bizning qarorimiz shu: hasharda barchamiz faol ishtirok etamiz.)
--	--

Til korpusidan sodda gaplarni avtomatik aniqlash uchun farqlar asosida qoidalar ishlab chiqish jarayoni

Birinchi jarayon-matnni tokenlarga ajratish

1. Dastur til korpuslaridan olingan matnni tokenlarga bo'lishi kerak.
2. Har bitta token o'z ichiga har qanday turdagi (**sodda, qo'shma, qo'shma gap turlarini, ko'chirma gapli qo'shma gaplarni**) ajratib oladi.

Ikkinchi jarayon-kesimlik belgisiga ko'ra gaplarni bir-biridan farqlash

1. Birinchi navbatda kesimlik belgisi jihatidan gaplarni bir-biridan farqlaydi.
2. Bunda, fe'llar bazasi kerak bo'ladi, bunga sabab, sodda va qo'shma gaplarda kesimlik belgisining eng ko'p qismini shu turkum qamrab olganligidir.
3. Dastur fe'llar bazasida mavjud so'zlarni tokenlar ichidan bir marta topsa, bu gaplarni sodda gaplar bazasiga qo'shadi. Bazadagi so'zlar tokenlar ichida ikki marta qatnashgan bo'lsa ular sodda gaplar to'plamiga qo'shilmaydi.

Uchinchi jarayon-yordamchi so'z turkumlari asosida gaplarni sodda va qo'shma gaplar bazasiga ajratish

1. Sodda va qo'shma gaplarni qoidalar asosida farqlaydigan bo'lsak, sodda gaplarda yordamchi so'z turkumlarining (**biriktiruv, ayiruv, zidlov, inkor bog'lovchilari, bog'lovchi vazifasidagi vositalar, nisbiy so'zlar, ko'makchili qurilmalar**) vositalari yordamida bog'lanmaydi.

1. Yuqorida sanab o'tilgan vositalar esa sodda gaplarni funksional sintaktik “bog'lash” belgisiga ko'ra qo'shma gaplarni hosil qiladi.
2. Shu ikki qoida asosida biz korpusdan sodda gaplarni ajratib olishimiz mumkin.
3. Uchinchi jarayonda dastur tokenlarga ajratilgan gaplarni quyidagi model asosida ajratib olishi kerak.

- 1) **Ega+kesim 2) kesim,+ yordamchi so'z turkumiga oid barcha so'zlar + kesim**

Bunda, birinchi Ega+kesim modeliga mos keluvchi gaplarni ajratib olib, sodda gaplar bazasiga qo'shadi. **2) kesim,+ yordamchi so'z turkumiga oid barcha so'zlar + kesim** modeliga mos keluvchi qo'shma gaplarni esa bizga taqdim etmaydi.

To'rtinchi jarayon-matndan ko'chirma gapli qo'shma gaplarni olib tashlash

1. Ushbu jarayonda dastur tokenlarni quyidagi qoida asosida tahlil qiladi.
Ko‘chirma gapli qo‘shma gaplar doimo ikki qismdan – muallif gapi va ko‘chirma gapdan tashkil topadi.

Muallif gapi (M) va ko‘chirma gap (K) bir-biriga nisbatan turlicha joylashishi mumkin, bunda tinish belgilari quyidagicha ishlatiladi: 1) M: «K». M: «K?» M: «K!» 2) M: «K», — m. M: «K?» — m. M: «K!» — m. 3) «K», — m. «K?» — m. «K!» — m. 4) «K, — m, — k».

Tinish belgilari orqali tokenlarni tahlil qilish. To‘rtinchi jarayonda, dasturning vazifasi tokenlarga ajratilgan gaplarni tinish belgilarining qo‘llanishiga ko‘ra tahlil qiladi.

Gapda 1) M: «K». M: «K?» M: «K!» 2) M: «K», — m. M: «K?» — m. M: «K!» — m. 3) «K», — m. «K?» — m. «K!» — m. 4) «K, — m, — k» belgilarga mos keluvchi qo‘shma gaplar mavjud bo‘lsa, bu turdagi gaplar sodda gaplar bazasiga qo‘shilmaydi.

Model. Model tushunchasi.

Hozirgi kunda kompyuter lingvistikasi terminiga juda katta e‘tibor berilmoqda. Tilshunoslikda modellashtirish metodidan foydalanish yaxshi samara beradi. Shuning uchun ushbu maqolada modellashtirish haqida ayrim mulohazalarni keltirib o‘tdim. O‘zbek tilshunosligida fonetika, grafika, orfografiya, orfoepiya, morfologiya, sintaksis kabi bo‘limlar mavjud. Ayni vaqtda barcha bo‘limlarning o‘z qonun qoidasi mavjud. Lekin bu bo‘limlarni qanday qilib modelllashtirish yoki qaysi metodlar orqali ushbu bo‘limlarni ochib berish haqida aniq ma‘lumotlar mavjud emas.

Tilshunoslikda modellashtirish metodidan foydalanish. So‘z turkumlarini formallashtirish. Gap bo‘laklarini formallashtirish. o‘zbek, rus va ingliz tillaridagi asosiy gap konstruktsiyalari ko‘p yillar davomida tilshunoslikda kuzatish metodlari yagona metod hisoblanib kelindi. Ammo bu metodlar yordamida lingvistik xodisalarning ichki mohiyati ochilmaydi. Bu metod yordamida so‘z shakllari, gap tuzilishi va boshqa struktur xususiyatlarni o‘rganish mumkin bo‘ladi.

Hozirgi kunda kuzatish metodi talabga javob bermayapti.

Yevropa ta‘lim tizimi barcha sohalari modellashtirilgan holatda va deyarli barcha fanlari keys study texnologiyasi asosida olib boriladi. Hamda darslar to‘liq taksanomiyaga asoslangan holatda o‘tiladi. Ya‘ni ularning har bir darsida bilish, tushinish, tahlil, analiz-sintez, xulosa, baholashdan iborat bo‘ladi. Tilshunoslikda til va nutqning tuzilishi murakkab bo‘lganligi sabablari bu metod bilan (kuzatish) ularni o‘rganib bo‘lmaydi. Chunki:

Til –umumiylik xususiyatiga ega, Nutq–xususiylik xususiyatiga ega.

Til –abstrakt, Nutq –konkret. Til–imkoniyat, Nutq –voqelik.

Modellashtirish metodida tadqiqotchi obektning o‘zini emas balki uning modelini o‘rganadi. Original bilan model o‘rtasida o‘zaro bog‘liqlik va mutanosiblik mavjud bo‘ladi. Tilshunoslikda modelning 3 ta turi bor:

1.Original modellar

2.Funksional modellar

3.Struktur modellar

Original modellar –obektning tuzilishini o‘rganadi. Funksional modellar – original modellarning qanday ishlashini o‘rganadi. Struktur modellar –har ikkalasi haqida ma’lumot beradi.

Original modellarga quyidagi talablar qo‘yiladi:

1. Model tabiiy obyektning aynan nusxasi bo‘lishi kerak.

2. Original model o‘zida elementlarning murakkab tuzilishini namoyon eta olishi kerak.

3. Tabiiy obyektning hamma xususiyatlari original modelga to‘g‘ri kelishi kerak.

4. Model evristik funksiyaga ega bo‘lishi kerak. Ya’ni u yangi g‘oyalar bera olishi va uni amaliyotda sinab ko‘rish imkoniyati bo‘lishi zarur.

Tilni model qilish usullari kompyuter lingvistikasida turli maqsadlarga yo‘naltirilgan. Bu usullar tilni avtomatik ravishda tahlil qilish, tarjima qilish, so‘zlar va gaplarni tan olish, tilni model qilish, so‘zlar orasidagi o‘zaro bog‘lanishni aniqlash va boshqa vazifalarni bajarish uchun ishlatiladi. Quyidagi zamonaviy usullar bu sohada muhimdir:

1. **Manfiy bo‘lmagan matritsa faktorizatsiyasi (Non-negative Matrix Factorization, NMF):** Bu usul til korpusini matritsa sifatida ko‘rsatadi va tilni modellashtirish uchun o‘zaro bog‘lanishlarni aniqlashda foydalaniladi.

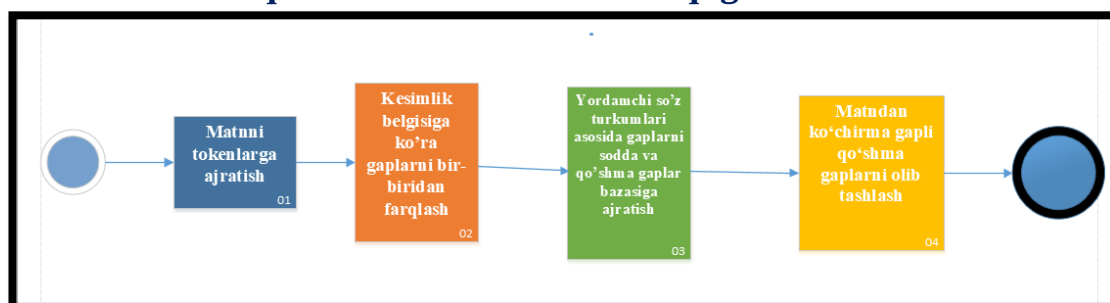
2. **Yashirin semantik taqsimot (Latent Semantic Allocation, LSA):** LSA til korpusini matritsa sifatida ifodalaydi va tilni model qilishda o‘zaro bog‘lanishlarni aniqlashga yordam beradi.

3. **Ehtimoliy yashirin semantik tahlil (Probabilistic Latent Semantic Analysis, PLSA):** PLSA til korpusini ehtimoliy model sifatida tasavvur qiladi va mavzularni aniqlashda foydalaniladi.

4. **Ida2vec:** Bu chuqur o‘rganish modeli tilni model qilishda so‘zlar va mavzular orasidagi o‘zaro bog‘lanishlarni aniqlashda ishlatiladi.

tBERT: Bu model til korpusini tashqi ma’lumotlar bilan birlashtiradi va tilni model qilishda yuqori natijalar olishga imkon beradi. Tilni model qilish uchun katta hajmdagi til korpusi kerak bo‘ladi.

Til korpusidan sodda gaplarni avtomatik aniqlash uchun ishlab chiqilgan qoidalar asosida ishlab chiqilgan model



Tokenizatsiya. Til korpuslaridan sodda gaplarni aniqlashda tokenizatsiya qilishning maqsadlari

Til korpusidagi matnni **tokenizatsiya** qilish orqali gaplarga bo'lib olishning bir nechta maqsadlari mavjud:

1. **Matn tahlili va tilni model qilish:** Tokenizatsiya til korpusidagi matnni so'zlarga yoki tokenlarga bo'lishga yordam beradi. Bu, matn tahlilida so'zlar va ularning tartibini saqlab qolishga imkon beradi. Tokenizatsiya tilni model qilish uchun kerakli formatni tayyorlashda muhimdir.

2. **Tarjima:** Tokenizatsiya matndagi so'zlarni ajratib olishga imkon beradi. Tarjima tizimlari matnlarni so'zlarga bo'lib olish uchun tokenizatsiyadan foydalanadi.

3. **Morfologik tahlil:** Tokenizatsiya til korpusidagi so'zlarni morfologik tahlil uchun tayyorlaydi. Bu, so'zlar va ularning grammatik xususiyatlarini o'rganishda yordam beradi.

4. **Ma'lumotlar bazasiga kiritish:** Tokenizatsiya ma'lumotlar bazasiga ma'lumot kiritish uchun kerakli formatni tayyorlaydi. Tokenlar ma'lumotlar bazasida indeks sifatida ishlatiladi.

5. **Matn tahlili va ma'lumot izlash:** Tokenizatsiya matndagi so'zlarni ajratib olishga imkon beradi. Bu, matn tahlilida so'zlar va ularning tartibini saqlab qolishga imkon beradi. Ma'lumot izlash uchun tokenizatsiya ma'lumotlar bazasiga ma'lumot kiritish uchun kerakli formatni tayyorlaydi.

Tokenizatsiya, kompyuter lingvistikasida matn tahlilini osonlashtirish uchun muhimdir. Bu tarjima, so'zlar orasidagi o'zaro bog'lanishni aniqlash, tilni model qilish va boshqa vazifalarni bajarishda foydalaniladi.

Foydalanilgan adabiyotlar:

1. <http://til.gov.uz/uz/news-and-announcements/news/237>
2. <https://azkurs.org/403-guruh-talabasi-maxammadiyeva-sitora-savollar.html?page=6>
3. [til-ta-limida-o-quv-lug-atchilikni-rivojlantirishning-korpusga-asoslangan-yondashuvi.pdf](#)
4. <https://pshedu.jdpu.uz/index.php/hp/article/view/317/138>
5. [Copilot \(microsoft.com\)](#)
6. Qodirov M., Ne'matov N., Abduraimova M., Sayfullayeva R., Mengliyev B. Ona tili - Umumiy o'rta ta'lim maktablarining 8-sinfi uchun darslik. – Toshkent, 2019.
7. Mahmudov N.M., Nurmonov A., Sobirov A.Sh. Ona tili-Umumiy o'rta ta'lim maktablarining 9-sinfi uchun darslik. – Toshkent, 2014