

**O‘ZBEKISTON RESPUBLIKASI
OLIY TA’LIM, FAN INNOVATSIYALAR VAZIRLIGI
ALISHER NAVOIY NOMIDAGI
TOSHKENT DAVLAT O‘ZBEK TILI VA ADABIYOTI
UNIVERSITETI**

**“KOMPYUTER LINGVISTIKASI: MUAMMOLAR,
YECHIM VA ISTIQBOLLAR”**

mavzusidagi IV an’anaviy xalqaro ilmiy-amaliy konferensiya

(2024-yil 20-aprel)

Toshkent – 2024

UO'K: 004.8+81'32

KBK: 81.1

Kompyuter lingvistikasi: muammolar, yechim va istiqbollar / Xalqaro ilmiy-amaliy konferensiya to'plami. Elektron nashr / ebook. – Toshkent: ToshDO'TAU, 20.04.2024. – 411 b.

Mazkur to'plamdan kompyuter lingvistikasi sohasiga doir tadqiqot natijalari, filologik dasturiy ta'minotlar, ularning imkoniyatlari va afzalliklari, tabiiy tilni qayta ishlashga oid masalalar, semantik tizim bazalari, til va adabiy korpuslar, lingvistik ta'minot manbalari, mashina tarjimasida lingvistik muammolar, lingvodidaktik dasturlar va platformalar, terminologik lug'atlarni yaratish masalalariga oid materiallar joy olgan.

Ushbu ilmiy to'plamdan kompyuter lingvistikasi sohasi vakillari, tadqiqotchilar, mustaqil izlanuvchilar, magistrantlar va kelajak sohalariga qiziquvchi talabalar foydalanishlari mumkin.

Mas'ul muharrir:

Sirojiddinov Sh.S. Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti rektori, akademik

Taqrizchilar:

Jamoldinova O. Toshkent davlat o'zbek tili va adabiyoti universiteti, professor, p.f.d.

Eʃref Adali. Istanbul texnika universiteti, professor, t.f.d.

Vladimir Benko. PhD, Slovakiya fanlar akademiyasi, L.Shtur nomidagi Tilshunoslik instituti katta ilmiy xodimi.

TASHKILIY QO'MITA

Sirojiddinov Shuhrat Samariddinovich (Toshkent davlat o'zbek tili va adabiyoti universiteti)

Jamoliddinova Odinaxon Rustamovna (Toshkent davlat o'zbek tili va adabiyoti universiteti)

Eshref Adali (Turkiya, Istanbul texnika universiteti)

Vladimir Benko (Slovakiya fanlar akademiyasi)

Suleymanov Djavdet Shevketovich (Tatariston Fanlar akademiyasi Amaliy semiotika instituti)

Gilmullin Rinat Abrekovich (Tatariston Fanlar akademiyasi Amaliy semiotika instituti)

Gatiatullin Ayrat Rafizovich (Tatariston Fanlar akademiyasi Amaliy semiotika instituti)

Elov Botir Boltayevich (Toshkent davlat o'zbek tili va adabiyoti universiteti)

Jo'raqo'ziyev Nodir Imomqo'ziyevich (Toshkent davlat o'zbek tili va adabiyoti universiteti)

Murat Orxun (Istanbul Bilgi universiteti)

Hamroyeva Shahlo Mirdjonovna (Toshkent davlat o'zbek tili va adabiyoti universiteti)

Abdullayeva Oqila Xolmo'minovna (Toshkent davlat o'zbek tili va adabiyoti universiteti)

TAHRIR HAY'ATI

Eshref Adali

Elov Botir Boltayevich

Hamroyeva Shahlo Mirdjonovna

Abdullayeva Oqila Xolmo'minovna

Xusainova Zilola Yuldashevna

Xudayberganov Nizomaddin Uktambay o'g'li

O‘ZBEK TILINI RAQAMLASHTIRISH – DAVR TALABI

Assalomu alaykum, hurmatli konferensiya ishtirokchilari, qadrlı mehmonlar!

Mamlakatimizda amalga oshirilayotgan keng ko‘lamli islohotlar jarayonida davlat tilining hayotimizdagi o‘rni va nufuzi tobora oshib bormoqda. O‘zbek tili siyosiy, ijtimoiy-iqtisodiy, ma‘naviy-ma‘rifiy jabhalarda faol qo‘llanib, xalqaro minbarlarda baralla yangramoqda. Xorijiy mamlakatlarda tilimizga e‘tibor va uni o‘rganishga qiziqish kuchaymoqda. Bugungi globallashuv davrida har bir xalq, har qaysi mustaqil davlat o‘z milliy manfaatlarini ta‘minlash, bu borada avvalo o‘z madaniyatini, azaliy qadriyatlarini, ona tilini asrab-avaylash va rivojlantirish masalasiga ustuvor ahamiyat qaratishi tabiiy.

O‘zbekiston Respublikasi Prezidentining 2019 yil 21 oktyabrdagi “O‘zbek tilining davlat tili sifatidagi nufuzi va mavqeini tubdan oshirish chora-tadbirlari to‘g‘risida”gi PF-5850-sonli Farmoni, 2020 yil 20 oktyabrdagi “Mamlakatimizda o‘zbek tilini yanada rivojlantirish va til siyosatini takomillashtirish chora-tadbirlari to‘g‘risida”gi PF-6084-sonli Farmonlari bu vazifalarni amalga oshirishni yanada jadallashtirdi.

Bugungi kunda mamlakatimiz ijtimoiy hayotining barcha sohalarida zamonaviy va innovatsion texnologiyalarning imkoniyatlaridan to‘liq va to‘g‘ri foydalanishga erishish; ta‘lim muassasalarida davlat tilini o‘rgatish va uni bilish darajasini baholash tizimini yanada takomillashtirish, davlat tilining ilm-fan tili sifatidagi nufuzini oshirish; davlat tili sofligini saqlash, uni boyitib borish va aholining nutq madaniyatini oshirish; davlat tilining axborot va kommunikatsiya texnologiyalari, xususan, Internet jahon axborot tarmog‘ida munosib o‘rin egallashini ta‘minlash, o‘zbek tilining kompyuter dasturlarini yaratish bo‘yicha bosqichma-bosqich tizimli ishlar amalga oshirilmoqda.

Bugun universitetimizda tashkil etilgan “Kompyuter lingvistikasi: muammolar, yechim va istiqbollar” mavzusidagi IV xalqaro konferensiya ham davlat tilining axborot va kommunikatsiya texnologiyalari, xususan, Internet jahon axborot tarmog‘ida munosib o‘rin egallashini ta‘minlash, o‘zbek tilining kompyuter dasturlarini yaratish kabi vazifalar ijrosiga bag‘ishlangan. Konferensiyaning barcha ishtirokchilarini tabriklayman, konferensiya ishiga omad tilayman!

Sirojiddinov Shuhrat Samariddinovich,
Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti rektori,
akademik



KOMPYUTER LINGVISTIKASI: TA'LIM YO'NALISHI VA ILMIY IZLANISH SOHALARI HAQIDA

Assalomu alaykum, hurmatli anjuman ishtirokchilari! Avvalo, o'zbek tilining nufuzini oshirish borasida mamlakatimizda keyingi yillarda olib borilayotgan islohotlar birin-ketin o'z samarasini ko'rsatmoqda. O'zbekiston Respublikasi Prezidenti mamlakatimizda o'zbek tilini yanada rivojlantirish va til siyosatini takomillashtirish chora-tadbirlari to'g'risidagi 6084-sonli farmonining 6-bandida Oliy va o'rta maxsus ta'lim vazirligi oliy ta'lim muassasalarining filologiya ta'lim yo'nalishlarida “Kompyuter lingvistikasi”, “Amaliy filologiya” yo'nalishlarini tashkil qilish, o'zbek tiliga oid barcha ilmiy, nazariy va amaliy ma'lumotlarni o'zida jamlagan elektron ko'rinishdagi o'zbek tili milliy korpusini yaratish vazifalari qo'yilgan edi. Buning amaliy natijasi sifatida jamg'arma tomonidan o'zbek tili milliy korpusini yaratish bo'yicha amaliy loyihalar e'lon qilindi.

Bundan tashqari Toshkent davlat o'zbek tili va adabiyoti universiteti olimlari tomonidan keng miqyosda milliy til korpusini ishlab chiqish ishlari olib borilmoqda. “O'zbek tilining ta'limiy korpusini yaratish”, “Turkiy tillarning lingvodidaktik platformasi”, “O'zbek tilining morfologik analizatori” amaliy loyihasi doirasida o'zbek tilining korpusi, o'zbek tilining morfologik analizatori, turkiy tillarni o'rgatishga mo'ljallangan platforma ishlab chiqildi. Olimlarning bunday tashabbuskorligini biz, albatta, qo'llab-quvvatlaymiz! Korpus katta hajmli lug'atlarni tuzish uchun manba vazifasini o'taydi. Vaqt o'tishi bilan korpus turli lingvistik yo'nalishlar uchun ahamiyatli bo'lishi bilan kuchli informatsion resursga aylandi. Korpus asosida kompyuter yordamida lug'atlar avvalgiga nisbatan tezlik bilan tuziladi va qayta ishlanadi. Shu yo'l bilan ish boshlanishidan tugash jarayonigacha (nashrgacha) tilni aks ettirib turadi, lug'at maqolasi “eskirish”ga ulgurmaydi. Tilning zamonaviy holatini aks ettiruvchi korpus turli davrlarda yashab ijod etgan ijodkorlarning mualliflik korpuslari uchun asos bo'lib xizmat qiladi. Korpusning lug'atshunoslik sohasidagi ahamiyati shundaki, so'zning qo'llanish davri va chastotasini aniqlashda hech qanday vosita korpusga tenglasha olmaydi. Korpus lingvistikasining keyingi taraqqiyot bosqichida statistik tadqiq metodi kompyuter tarjimai, nutqni sintezlash va tanish, aniqlash, orfografik tekshiruv kabi lingvistik amallarni bajarishda qo'l kela boshladi. Yaqin orada til o'rganayotgan o'quvchi yoki uning biror jihatini tadqiq etayotgan kishining bugungi kunda lug'atga bo'lgan ehtiyoji korpusga ko'chishi shubhasiz. Konferensiya ishiga omad tilayman!

Jamoldinova Odinaxon Rasulovna,
Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti,
Ilmiy ishlar va innovatsiyalar bo'yicha prorektor



KOMPYUTER LINGVISTIKASINI RIVOJLANTIRISH: EHTIYOJ VA ZARURAT

Dunyo kompyuter lingvistikasida oxirgi yillarda milliy korpuslar, avtomatik tarjima platformalari, tezaurus, elektron lug'atlardan foydalanish, ularni ilmiy-nazariy jihatdan o'rganish harakati avj oldi. Shu jarayonda korpusni tuzishning dasturiy va lingvistik tamoyillarini asoslash, tabiiy tilni qayta ishlash (NLP), mashina tarjimasi, matnga qayta ishlov berish tizimlari, kompyuter lingvodidaktikasi, kompyuter leksikografiyasi, semantik tizimlar va lingvistik ontologiyalar ustida ishlash va korpus grammatikalarini yaratishga keng imkon yaratadi.

Mamlakatimizda global ilmiy-texnik islohotlar amalga oshirilayotgan bugungi kunda kompyuter lingvistikasi doirasida avtomatik tarjima, o'zbek tilini tushunish va qayta ishlash, uni boshqa tillar bilan chog'ishtirishda sun'iy intellektning rolini belgilashga kuchli ehtiyoj sezilmoqda. Shu ma'noda "...davlat tilining sofliğini saqlash, uni boyitib borish va aholining nutq madaniyatini oshirish; davlat tilining zamonaviy axborot texnologiyalari va kommunikasiyalariga faol integrasiyalashuvini ta'minlash" kabi ustuvor yo'nalishlar tahlil qilinayotgan til faktlarini tizimlashtirish, akademik maqsadlarda milliy va xorijiy til ko'nikmalarini egallash, tilni real tasvirlash, shuningdek, amaliy tilshunoslikda til qo'llanishi etalonlarini optimallashtirishga keng yo'l ochadi.

Kompyuter lingvistikasi istiqbolli ilmiy yo'nalish ekanligini inobatga olgan holda o'zbek tili milliy korpusini yaratish, lingvistik modellarni tuzish singari masalalarni zamonaviy ilmiy tamoyillar asosida tadqiq etish fanimiz oldida turgan dolzarb vazifa hisoblanadi. Shu nuqtai nazardan o'zbek tilining formal grammatikasini yaratish, korpus lingvistikasi, tabiiy tilni qayta ishlash (NLP), mashina tarjimasi, matnga qayta ishlov berish tizimlari, kompyuter lingvodidaktikasi, kompyuter leksikografiyasi, semantik tizimlar va lingvistik ontologiyalar, kompyuter lingvistikasida terminologiya va terminografiya, kompyuter lingvistikasida sun'iy intellekt kabi sohalarni rivojlantirish nihoyatda muhim.

Bugun universitetimizda tashkil etilgan "Kompyuter lingvistikasi: muammolar, echim va istiqbollar" mavzusidagi xalqaro konferensiya mazkur masalalarni mutaxassislar va yosh tadqiqotchilar bilan muhokama qilish, bu sohalarda qilingan ishlar bilan tanishishga imkon beradi.

Xoliqov Bahodir Aliqulovich,
**O'zbekiston Respublikasi Vazirlar Mahkamasining Davlat tilini rivojlantirish
departamenti mutaxasisi**

O‘ZBEK TILINING ELEKTRON RESURSLARINI YARATISH – ZAMON TALABI

Assalomu alaykum, hurmatli anjuman ishtirokchilari!

O‘zbekiston Respublikasi Prezidentining “Mamlakatimizda o‘zbek tilini yanada rivojlantirish va til siyosatini takomillashtirish chora-tadbirlari to‘g‘risida”gi 6084-sonli farmonida Internetda o‘zbek tilidagi kontentni ko‘paytirish, o‘zbek tilini raqamlashtirish, o‘zbek tiliga zamonaviy axborot texnologiyalarini joriy etish kabi vazifalar belgilangan. Buning amaliy natijasi sifatida jamg‘arma tomonidan o‘zbek tili milliy korpusini yaratish bo‘yicha amaliy loyihalar e‘lon qilindi, o‘zbek tilining o‘quv elektron lug‘atlarini yaratish bo‘yicha qator loyihalar amalga oshirildi. Ammo bu loyihalar bilan barcha vazifalarni qamrab olish mumkin emas. Shu sababli o‘zbek tili korpuslari, o‘zbek tiliga avtomatik ishlov berish vositalari, o‘zbek tili imlo tekshiruv va tahrir dasturi, o‘zbek tili sinxron tarjimonlari, o‘zbek tilida turli sohalarga oid savol-javob tizimlari, shuningdek, turli elektron lug‘atlarni ishlab chiqish hali ham dolzarbligicha qolmoqda.

Toshkent davlat o‘zbek tili va adabiyoti universiteti olimlaridan iborat ishchi guruh tomonidan keng miqyosda milliy til korpusini ishlab chiqish ishlari olib borilmoqda. “O‘zbek tilining ta’limiy korpusini yaratish”, “Turkiy tillarning lingvodidaktik platformasi”, “O‘zbek tilining morfologik analizatori” amaliy loyihasi doirasida o‘zbek tilining korpusi, o‘zbek tilining morfologik analizatori, turkiy tillarni o‘rgatishga mo‘ljallangan platforma ishlab chiqildi. Bunday loyihalarning yana davom etishida bizning jamoamizda yetarlicha salohiyat mavjud. Shu asosda aytish mumkinki, kelgusida bizning jamoamiz professor-o‘qituvchilari, yosh olimlar, doktorant va magistrantlar ishtirokida mana shunday elektron resurslarni ishlab chiqish bo‘yicha ilmiy izlanish va amaliy loyihalarning amalga oshirilishiga ishonamiz va konferensiya ishiga omad tilaymiz!

Normamatov Sultonbek Ermamatovich,
Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti,
O‘quv ishlari bo‘yicha prorektor

YALPI YIG‘ILISH MA’RUZALARI

Türkçenin Mantığı

Eşref ADALI

Prof. Dr.

İstanbul Teknik Üniversitesi
Bilgisayar ve Bilişim Fakültesi

Türkiye

adali@itu.edu.tr

Öz

Türkçe, sanki bir bilim kurulu tarafından tasarlanmış bir dildir. Dilde kurallar belirgin ve saydamdır. Türkçenin bu özelliğini kanıtlamak için, ses bilimi, biçim bilimi, söz dizimi özelliklerine ilişkin kurallar çıkarılmış, her kural için mantık bağıntısı yazılmış, akış diyagramı çizilmiştir. Akış diyagramına uygun olarak geliştirilen programlar ile Türkçenin kuralları sınanmıştır. Sınamaların sonucu olarak, Türkçenin kurallarının bozulmamış olduğu kanıtlanmıştır. Son olarak, bir cümleyi yapay olarak oluşturacak bir çalışma yapılmış ve program hazırlanmıştır.

Anahtar Sözcükler: Türkçe, Mantık, Ses Bilimi, Biçim Bilimi, Söz Dizimi

1 - Giriş

Dil ve dilin yazıya dökülmesi insanlık tarihindeki en önemli gelişmeler olarak değerlendirilebilir. İnsanların anatomik yapısı incelendiğinde konuşmaya yaklaşık 160.000 yıl önce başladıkları öne sürülmektedir. Yakın zamana kadar yazının geçişinin günümüzden 6.000 yıl öncesine gittiği söylenmekteydi ancak Göbeklitepe'nin kazı başkanı Müslüm Ercan, Göbeklitepe'de yapılan çalışmalarda bulunan bir dikilitaş üzerindeki sahnenin bir olayı konusal olarak anlatmasından dolayı "ilk resim yazısı" olabileceğini söylemiştir. Göbeklitepe'nin günümüzden 12.000 yıl önce yapıldığı hesaplanmaktadır. İnsanların dillerini nasıl oluşturdukları konusunda karşıt görüşler bulunmaktadır. Bu görüşlerden bazıları dili tanrının öğrettiğini söylerken bazıları dilin zaman içinde kendiliğinden geliştiğini söylerler.

Toplumların gelişmesinde dil ve yazının çok önemli bir katkısı olduğu yadsınamayacak bir gerçektir. Dil ve toplum ilişkisini Konfüçyüs (MÖ 551- MÖ 479) şöyle açıklamaktadır:

“... Bir ülkenin yönetimini ele alsaydım, yapacağım ilk iş, hiç kuşkusuz dilini gözden geçirmek olurdu. Çünkü dil kusurlu ise sözcükler düşüncüyü iyi ifade edemez. Düşünce iyi ifade edilemezse görevler ve hizmetler gereği gibi yapılamaz. Görev ve hizmetin gerektiği şekilde yapılamadığı yerlerde âdet, kural ve kültür bozulur. Âdet, kural ve kültür bozulursa adalet hatalı yollara sapar. Adalet yoldan çıkarsa şaşkınlık içine düşen halk ne yapacağını, işin nereye varacağını bilemez. İşte bunun içindir ki, hiçbir şey dil kadar önemli değildir !..”

Prof. Dr. David Cuthell'in Türkçe için şunları söylemektedir:

- “Türkçe öyle bir dildir ki, yüz yüksek matematik profesörü bir araya gelerek, Türkçeyi yaratmışlar sanki.
- Türkçe öyle bir dildir ki, başlı başına bir duygu, düşünce, mantık ve felsefe dilidir.
- Diğer diller sözcüklerin durağan olan anlamlarını öğrenmeye, Türkçe ise bu anlamları bulup çıkarmaya, yani dinamik anlamlandırmaya dayalıdır.”

Prof. Dr. Johan Vandewalle'nin Türkçe için söyledikleri kendi ağzından şöyledir [12, 13]

- “...Ana dili Türkçe olan bir kişinin kısa cümlelerle düşündüğü, konuşma anında ise bu kısa cümleleri çeşitli yollarla birbirine bağlayarak karmaşık yapılar kurduğu görüşündeyim. Farklı dil gruplarına ait birçok dili incelediğim halde şimdiye kadar hiçbir dilde beni Türkçe'deki karmaşık cümle yapıları kadar büyüleyen bir yapıya rastlamadığımı söyleyebilirim.
- Biraz duygusal olmama izin vererseniz, bazen kendime ‘Keşke Chomsky de gençliğinde Türkçe öğrenmiş olsaydı... ‘ diyorum. Eminim o zaman çağdaş dilbilim İngilizce'ye göre değil, Türkçe'ye göre şekillenmiş olurdu...”

Max Müller, 1862 yılında yayımladığı "Lectures on The Science of Language" adlı kitabında Türkçe için şu değerlendirmede bulunmuştur [14] :

- “Türkçe dil bilgisini kullanmak için isteğiniz olmasa bile, Türkçe dil bilgisini okumak gerçekten büyük bir zevktir. Çok sayıdaki dil bilgisi biçiminin ortaya çıkarılış biçimi, çekim sistemindeki düzenlilik, tüm yapının saydamlığı ve anlaşılabilirliği, dilde parıldayan insan zekasının o harika gücünü hisseden herkesi etkilemelidir.
- Türk dillerinin dil bilgisi ise tam tersine, saydam bir yapıya sahiptir ve kristal bir arı kovanındaki peteklerin yapılışını izler gibi iç işleyişini inceleyebileceğimiz düzene sahiptir. Tanınmış bir doğu bilimci, “**Türkçenin bazı seçkin bilginler topluluğunun çalışmalarının sonucu olduğunu hayal edebiliriz**” der. Ancak Tataristan bozkırlarında kendi halinde yaşayan topluluklar yalnızca doğuştan gelen yasaları veya içgüdüsel güçleri ile bu kadar mükemmel bir şeyi tasarlayamazdı.”

2 - Türkçenin Ses Özellikleri

Dilin en temel bileşeni, ses ve bu seslere karşılık düşen harflerdir.

Bugünkü anlamda ilk abecenin Sami ırkından olan denizci Fenikeliler tarafından geliştirildiği bilinmektedir. MÖ 1000 yılında kullanılmaya başlandığı sanılan bu abecede 22 harf bulunmaktadır ve bunların hepsi ünsüz harflerdir. Fenike abecesi, bu denizci ulus tarafından Akdeniz'e kıyısı olan uluslara öğretilmiştir. Bu aktarmalar

- Sert qarşılıǵı bulunmayan yumuşak bir ünsüzden sonra, ikinci sırada sert, üçüncü sırada sert qarşılıǵı bulunan yumuşak ünsüz gelebilir. Öncelikle sert qarşılıǵı bulunmayan bir ünsüz gelir.
- Bir sözcüğün içindeki ünsüzler genellikle aynı kümeden seçilmektedirler.
- Dış ünsüzlerinden sonra yüksek olasılıkla damak daha sonra dudak ünsüzlerinden biri gelebilir.
- Damak ünsüzlerinden sonra yüksek olasılıkla dış daha sonra dudak ünsüzlerinden biri gelebilir.

2.2 – Ses Deǵışmeleri

Ses deǵışmeleri içinde süreksiz sert ünsüzlerin yumuşaması önemli görülür. Yumuşamanın temel kuralları şöyledir:

- Süreksiz sert ünsüzle (**p, ç, t, k**) biten sözcüklere ünlü ile başlayan bir ek ulandığında set ünsüzler yumuşar ve (**b, c, d, g**)'ye dönüşürler.
- **k** ünsüzü iki ünlü arasına düşerse **k, ğ** olur.

Birden çok seslemden kurulu sözcükler için kurallar

- Sonu **-it** ile biten sözcüklerin son harfi deǵışmez.
- Sonu **-ut** ile biten sözcüklerin son harfi **t** deǵışmemektedir ancak *angut* kural dışıdır.

Birden çok seslemden kurulu olan bazı sözcükler yukarıdaki kurala uymazlar

- Sonu **-çüt** ile biten sözcüklerin son harfi **t** deǵışmemektedir ancak *öğüt, söğüt* sözcüklerinin son harfi **d** 'ye dönüşmektedir. Ölçüt sözcüğündeki deǵışmezliǵı sondan üç önceki sert ünsüz belirlemektedir. Ölçüt tek örnektir.
- Sonu **-it** ile biten bir tek *kelkit* sözcüğü bulunmuştur. Bu sözcük halk dilinde kullanılmaktadır. Dolayısıyla genel kurallara alınması doğru olmaz.

Son harfin deǵışmediǵi diǵer durumlar:

- Sözcüğün sonunda yapı ve çatı eki bulunuyor ise, ek ünlü ile başlasa bile sözcüğün son harfi deǵışmez.

Yabancı dillerden alınan sözcükler zaman içinde Türkçenin ses özelliklerine uyum sağlarlar ancak bazıları uymazlar.

İki veya üç harf içeren tek seslemlı sözcükler için kurallar

- İki ünlü arasında bulunan **ç, k, p, t** harfler deǵışmez kuralı iki veya üç harfli olan tek seslemlere uygulanan genel bir kuraldır.
- Altı ad ve beş eylem genel kurallara uymamaktadır:
uç ⇒ ucu, but ⇒ budu, tat ⇒ tadı, gök ⇒ göǵü, dip ⇒ dibi, kap ⇒ kabı
tat-, et-, dit-, git- güt-

Dört harf içeren seslemlere ilişkin yumuşama kuralları

Dört harf içeren seslemlere ilişkin yumuşama kuralları şöyledir:

- Aynı seslem içinde iki ünsüz peş peşe yer alabilir, ancak bu ünsüzler şu kurala uyarlar:
- Öncül ünsüz **l, n, r, s** veya **ş** olmalıdır; **l** den sonra **ç, k, p** veya **t**; **n** den sonra **ç, k, t, r** den sonra **ç, k, p, s** veya **t, s** ve **ş** den sonra **t** gelebilir.

- Dört harfî olan tek seslemlerde, **şekil-2: Sert ünsüzlerin Yumuşamasına ilişkin** bu ön koşullara uyularak son harf yumuşar ancak son iki harfî **rk** ve **nt** ise yumuşamaz.

Yumuşama kuralları Şekil-2’de görülmektedir.

2.3 – Seslem Yapısı

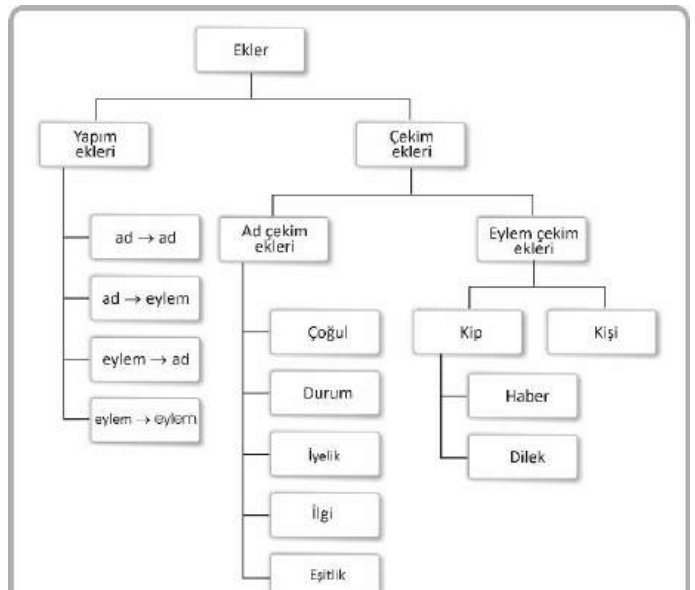
Türkçenin seslem yapısı kurallıdır ve aşağıdaki gibidir:

[S]	(o)	[SZ]	(at)	[SZZ]	(üst)	[ZS]	(su)	[ZSZ]	(tam)	[ZSZZ]	(Türk)
[S + ZS]		(aba)		[S + ZSZ]		(ocak)		[S + ZSZZ]		(avurt)	
[SZ + ZS]		(ayna)		[SZ + ZSZ]		(aşkın)		[SZ + ZSZZ]		(aldanç)	
[ZS + ZS]		(baba)		[ZS + ZSZ]		(tabak)		[ZS + ZSZZ]		(kazanç)	
[ZSZ + ZS]		(bohça)		[ZSZ + ZSZ]		(bostan)		[ZSZ + ZSZZ]		(başkurt)	
[ZSZZ + ZS]		(Türkçü)		[ZSZZ + ZSZ]		(Türklük)					

3 - Türkçenin Biçim Bilimsel Özellikleri

Türkçe sondan eklemeli bir dildir ve kök sözcüğün sonuna yapım ve çekim ekleri ulanmaktadır. Çekim ekleri sözcükler arası bağları oluşturmakta ve dilin anlaşılabilirliğine çok önemli katkı sağlamaktadır. Yapım ekleri ise yeni sözcükleri türetmeyi sağlamaktadır. Çekim ekleri içinde yer alan durum ekleri eylemin yönünü göstermesi açısından önemli katkı sağlamaktadır.

Adların çekimi, ad soylu sözcüklere çoğulluk, iyelik ve ilgi eklerini kazandırarak sözcükler arası bağlantıları açık biçimde



Şekil-3: Yapım ve çekim ekleri

tanımlamaktadır. Eylemlere eklenen çekim ekleri, bir eylemin kim tarafından, ne zaman ve hangi durumlarda gerçekleştirildiğini göstermektedir.

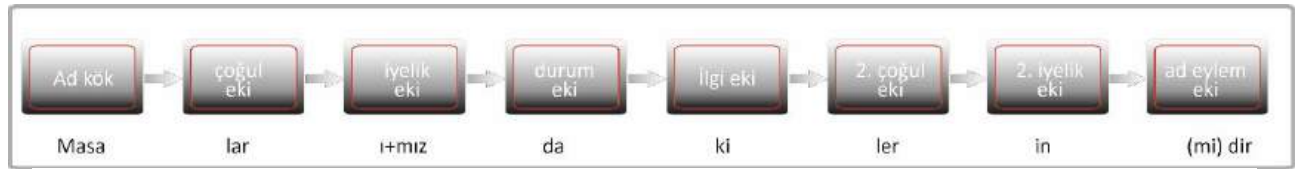
Yapım ekleri kök veya gövde sözcüklere yeni anlamlar kazandırmaktadır. Böylece yeni sözcükler türetilmektedir. **Yeni türetilen sözcüklerde kök değişmediği için, türetilen sözcüğün anlamı konuşan tarafından anlaşılmaktadır.**

Ekler birbirini izleyebilir ve sayısında sınırlama yoktur. Yapılan araştırmalar Türkçede tüm sözcüklerin yaklaşık **0,94 ek** aldıklarını; ek alan sözcüklerin **1,85 ek** aldıklarını göstermektedir.

Türkçede 170 kadar yapım ekinin olduğu ve bunlardan 70 kadarının sıkça kullanıldığı bilinmektedir. Şekil-3'te çekim ve yapım ekleri gösterilmiştir.

3.1 – Ad Çekimi

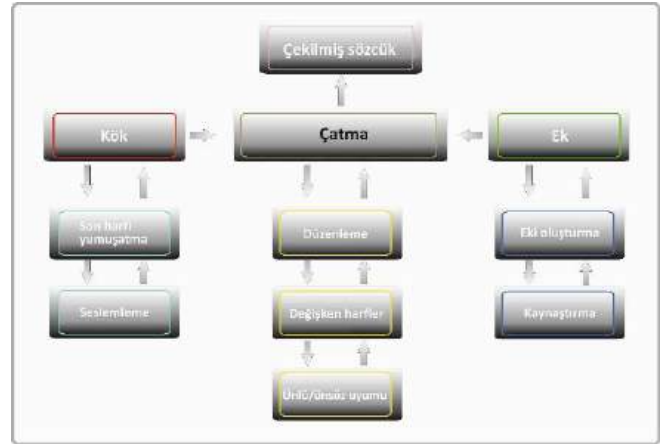
Ad çekim kuralı Şekil-4'te görülmektedir:



Şekil-4: Ad çekim kuralı

Bir eklemin kendisinden sonra gelen eklemeye ulanabilmesi için Türkçenin ses özelliklerine uygun bazı uyarlamaların yapılması gerekir. Çatma süreci adını verdiğimiz bu aşamada,

- ünlü ünsüz uyumları,
- kaynaştırma harflerinin eklenmesi,
- sert ünsüzlerin yumuşatılması ve
- bağlayıcı ünlüler uyumu kuralları uygulanır.



Kaynaştırma süreci Şekil-5'te **Şekil-5: Kaynaştırma süreci** gösterilmiştir.

3.1.1 Çoğul Eki

Çoğul eki +lar biçimindedir. A, kökün son harfine göre a veya e olur. Çoğul ekinin mantık bağıntısı şöyledir:

$$A = (e \vee i \vee ö \vee ü) \wedge (a \vee e)$$

Buna göre şu yazılır

$$söz + \begin{cases} S(a \vee i \vee ö \vee ü) \Rightarrow +lar \\ S(e \vee i \vee ö \vee ü) \Rightarrow +ler \end{cases}$$

3.1.2 Soru Eki

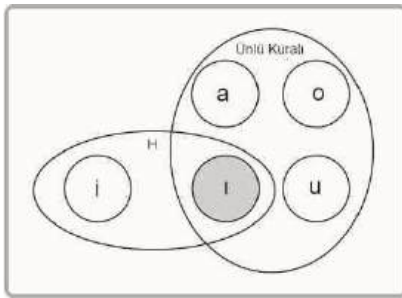
Soru eki $+mH$ biçimindedir. H, kökün son harfine göre değişir. Soru ekinin mantık bağıntısı şöyledir:

$$\begin{aligned} H &= (a \vee i) \wedge (i \vee i \vee u \vee \bar{u}) = i \\ H &= (o \vee u) \wedge (i \vee i \vee u \vee \bar{u}) = u \\ H &= (e \vee i) \wedge (i \vee i \vee u \vee \bar{u}) = i \\ H &= (\bar{o} \vee \bar{u}) \wedge (i \vee i \vee u \vee \bar{u}) = \bar{u} \end{aligned}$$

$$söz + \begin{cases} S(a \vee i) \Rightarrow 'mi?' \\ S(e \vee i) \Rightarrow 'mi?' \\ S(o \vee u) \Rightarrow 'mu?' \\ S(\bar{o} \vee \bar{u}) \Rightarrow 'mü?' \end{cases}$$

3.1.3 İyelik Eki

$$H = (a \vee i \vee o \vee u) \wedge (i \vee i) = i \quad \text{Kaş} + (H) \text{iyelik eki}$$



$$BAĞ = \begin{cases} a \vee i \Rightarrow i \\ e \vee i \Rightarrow i \\ o \vee u \Rightarrow u \\ \bar{o} \vee \bar{u} \Rightarrow \bar{u} \end{cases} \quad \text{ya da } BAĞ(s) \quad \text{Ünlüleri bağlayan kural}$$

Ek	Açıklama	Örnek
+(H)m	1.tekil	kuş <u>um</u>
+(H)n	2.tekil	kuş <u>un</u>
+(s)H	3.tekil	kuş <u>u</u>
+(H)mHz	1.çoğul	kuş <u>umuz</u>
+(H)nHz	2.çoğul	Kuş <u>unuz</u>
+lArH	3.çoğul	

$$söz = \begin{cases} BR 1.tek \Rightarrow YUM(söz) \left\{ \begin{array}{l} SH(söz) = \bar{U}n \Rightarrow +'''' \\ SH(söz) = Sz \Rightarrow +BAĞ(söz) \end{array} \right\} +m \\ BR 2.tek \Rightarrow YUM(söz) \left\{ \begin{array}{l} SH(söz) = \bar{U}n \Rightarrow +'''' \\ SH(söz) = Sz \Rightarrow +BAĞ(söz) \end{array} \right\} +n \\ BR 3.tek \Rightarrow YUM(söz) \left\{ \begin{array}{l} SH(söz) = \bar{U}n \Rightarrow +s \\ SH(söz) = Sz \Rightarrow +'''' \end{array} \right\} +BAĞ(söz) \\ BR 1.çoğ \Rightarrow YUM(söz) \left\{ \begin{array}{l} SH(söz) = \bar{U}n \Rightarrow +'''' \\ SH(söz) = Sz \Rightarrow +BAĞ(söz) \end{array} \right\} +m + BAĞ(söz) +z \\ BR 2.çoğ \Rightarrow YUM(söz) \left\{ \begin{array}{l} SH(söz) = \bar{U}n \Rightarrow +'''' \\ SH(söz) = Sz \Rightarrow +BAĞ(söz) \end{array} \right\} +n + BAĞ(söz) +z \\ BR 3.çoğ \Rightarrow I + \left\{ \begin{array}{l} S(söz) = Ka \Rightarrow a \\ S(söz) = In \Rightarrow e \end{array} \right\} +r + BAĞ(söz) \end{cases}$$

Şekil-6: İyelik eklerinin

3.1.4 Durum Ekleri

Türkçede 6 ana ve üç özel durum eki vardır:

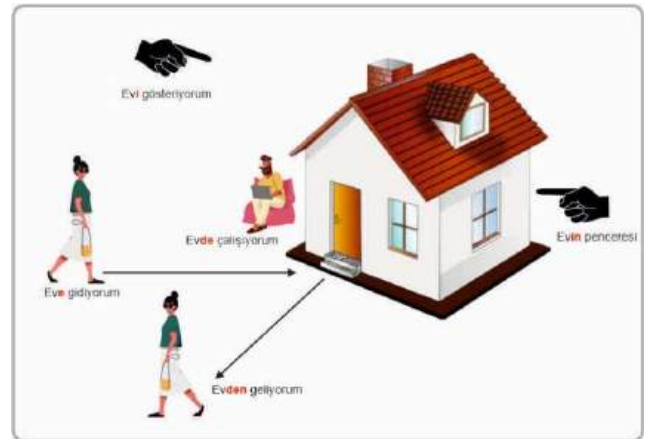
Yalın durum

- +(y)H (I durumu: belirtme durumu)
- +(y)A (e durumu: yönelme durumu)
- +DA (de durumu: bulunma durumu)
- +DAn (den durumu: ayrılma durumu)
- +(n)Hn (in durumu: tamlayan durumu)

ev
evi
eve
evde
evden
evin

Ayrıca şu durum ekleri de bulunur:

- +DAki (ilgi eki) kapı kapıdaki
- +(y)lA (araç veya birliktelik) uçak uçakla
- +CA (eşitlik) ben bence



Şekil-7: Durum eklerinin anlamı

- Adlara durum ekleri eklenirken Türkçenin ses uyum kurallarına uyulur.
- Bir ünsüz ile biten cins adlar durum eklerini ünlü uyumu ve ünsüz benzeşmesine göre alırlar.
- Ayrıca sözcük sonunda bulunan ünsüzler için yumuşama kuralları uygulanır.

Durum eklerine ilişkin mantık bağıntısı şöyledir: Bu bağıntıya ilişkin açıklama şöyledir:

- Sert (St) : sert ünsüzler (ç, f, h, k, p, s, ş, t)
- Yumuşak (Yu): yumuşak ünsüzler (b, c, d, g, ğ, j, v, z, l, m, n, r, y)

Durum eklerine ilişkin mantık bağıntısı Şekil-8’de görülmektedir.

3.2 – Eylem Çekimi

Eylemler iki kümede incelenir: Ek eylemler ve asıl eylemler.

3.2.1 Ek Eylemlerin Çekimi

Türkçede ek eylemler ad soylu sözcüklerin yüklem olarak kullanılmasını sağlarlar. Ekler zaman kipine uygun olarak biçimlenirler ve dört kip için tanımlıdır ve şu aşamalardan geçilerek çekilir:

1. Kökün olumsuzunu oluşturma
2. Eki oluşturma
3. Yumuşatma
4. Ek içindeki değişken harfleri düzenleme
5. Eylemi oluşturma

$$\begin{aligned}
 & \text{Yalın ise } +''' \\
 & \text{Belirtme ise } \left\{ \begin{array}{l} SH(söz) = Ün \Rightarrow +y \\ SH(söz) = Sz \Rightarrow +''' \end{array} \right\} + BA\check{G}(söz) \\
 & \text{Yönelme ise } \left\{ \begin{array}{l} SH(söz) = Ün \Rightarrow +y \\ SH(söz) = Sz \Rightarrow +''' \end{array} \right\} + \left\{ \begin{array}{l} S(söz) = Ka \Rightarrow +a \\ S(söz) = İn \Rightarrow +e \end{array} \right\} \\
 & \text{Bulunma ise } \left\{ \begin{array}{l} SH(söz) = St \Rightarrow t \left\{ \begin{array}{l} S(söz) = Ka \Rightarrow +a \\ S(söz) = İn \Rightarrow +e \end{array} \right\} \\ SH(söz) = Yu \Rightarrow d \left\{ \begin{array}{l} S(söz) = Ka \Rightarrow +a \\ S(söz) = İn \Rightarrow +e \end{array} \right\} \end{array} \right\} \\
 & \text{Ayrılma ise } \left\{ \begin{array}{l} SH(söz) = St \Rightarrow t \left\{ \begin{array}{l} S(söz) = Ka \Rightarrow +a \\ S(söz) = İn \Rightarrow +e \end{array} \right\} \\ SH(söz) = Yu \Rightarrow d \left\{ \begin{array}{l} S(söz) = Ka \Rightarrow +a \\ S(söz) = İn \Rightarrow +e \end{array} \right\} \end{array} \right\} + n \\
 & \text{Tamlayan ise } \left\{ \begin{array}{l} SH(söz) = Ün \Rightarrow +n \\ SH(söz) = Sz \Rightarrow +''' \end{array} \right\} + BA\check{G}(söz) + n \\
 & \text{İlgi ise } \left\{ \begin{array}{l} SH(söz) = St \Rightarrow t \left\{ \begin{array}{l} S(söz) = Ka \Rightarrow +a \\ S(söz) = İn \Rightarrow +e \end{array} \right\} \\ SH(söz) = Yu \Rightarrow d \left\{ \begin{array}{l} S(söz) = Ka \Rightarrow +a \\ S(söz) = İn \Rightarrow +e \end{array} \right\} \end{array} \right\} ki \\
 & \text{Araç ise } \left\{ \begin{array}{l} SH(söz) = Ün \Rightarrow +y \\ SH(söz) = Sz \Rightarrow +''' \end{array} \right\} + I \left\{ \begin{array}{l} S(söz) = Ka \Rightarrow +a \\ S(söz) = İn \Rightarrow +e \end{array} \right\} \\
 & \text{Eşitlik ise } c \left\{ \begin{array}{l} S(söz) = Ka \Rightarrow +a \\ S(söz) = İn \Rightarrow +e \end{array} \right\}
 \end{aligned}$$

Şekil-8: Durum eklerinin mantık bağıntısı

Kipler	Zaman eki	Birey Ekleri	Örnekler
+ken eki	+'''	+(y)ken	öğrenciyken
+mışcasına eki	+'''	+(y)mHşcAsHnA	öğrencimışcasına

Kipler	Zaman eki	Birey Ekleri	Örnekler
Belirsiz geçmiş zaman	+(y)mHş	+Hm, +sHn, +''', +Hz, +sHnHz, +IAr	öğrenciymişim
Belirli geçmiş zaman	+(y)DH	+m, +n, "+", +k, +nHz, +IA	öğrenciydim
Şimdiki ve Geniş zaman	+'''	+(y)Hm, +sHn, +DHR, +(y)Hz, +sHnHz, +IAr	öğrenciyim
Dilek - Koşul kipi	+(y)s(A)	+m, +n, +''', +k, +nHz, +IAr	öğrenciysem

3.2.2 Eylemlerin Çekimi

Türkçe eylemlerin çekimi kurallıdır ve hiçbir kural dışı örnek yoktur. Buna karşın İngilizce ve Fransızcada özellikle sık kullanılan ve sayıları yüzden den fazla olan

eylemlerin çekimleri kuralsızdır. Kuralsız eylemlerin çekimleri ancak ezberlenerek kullanılır. Türkçede eylem çekimlerinin kurallı olması, dile yeni bir eylem girdiğinde kuralara uygun olarak çekilebilmesine olanak sağlar.

Türkçe eylemlerin **yön belirtme özelliği vardır** ve **mantıklıdır**.

Eren Bartu'ya söylüyor

Bartu Eren'i dinliyor

Türkçede bir eylemin çekiliş kalıbı şöyledir:



Bir eylemin çekilmesi aşağıdaki aşamalardan geçilerek yapılır:

1. Çekim ekleri için ön hazırlık
2. Eylem çatısının oluşturulması
Dönüşlü eylem kökünün oluşturulması
İşteş eylem kökünün oluşturulması
Ettirgen eylem kökünün oluşturulması
Edilgen eylem kökünün oluşturulması
3. Yardımcı eylemlerin
Tezlik yardımcı eylemi
Yeterlik yardımcı eylemi
Sürerlik yardımcı eylemi

Kipler	Zaman eki	Birey Ekleri	Örnekler
Belirsiz geçmiş zaman	+mHz	+Hm, +sHn, + ^{III} , +Hz, +sHnHz, +IAr	sevmişim
Belirli geçmiş zaman	+D	+Hm, +Hn, +H, +Hk, +HnHz, +HIAr	sevdim
Şimdiki zaman	+(H)yör	+Hm, +sHn, + ^{III} , +Hz, +sHnHz, +IAr	seviyorum
Gelecek zaman	+(y)AcAk	+Hm, +sHn, + ^{III} , +Hz, +sHnHz, +IAr	seveceğim
Geniş zaman	+(A/H)r	+Hm, +sHn, + ^{III} , +Hz, +sHnHz, +IAr	severim
Belirsiz geçmiş zamanın Hikayesi	+mHzD	+Hm, +Hn, +H, +Hk, +HnHz, +HIAr	sevmiştim
Belirli geçmiş zamanın Hikayesi	+DHyo	+Hm, +Hn, +H, +Hk, +HnHz, +HIAr	sevdiydin
Şimdiki zamanın Hikayesi	+(H)yörD	+Hm, +Hn, +H, +Hk, +HnHz, +HIAr	seviyordum
Gelecek zamanın Hikayesi	+(y)AcAkD	+Hm, +Hn, +H, +Hk, +HnHz, +HIAr	sevecektim
Geniş zamanın Hikayesi	+(A/H)rD	+Hm, +Hn, +H, +Hk, +HnHz, +HIAr	severdim
Belirsiz geçmiş zamanın rivayeti	+mHzmHz	+Hm, +sHn, + ^{III} , +Hz, +sHnHz, +IAr	sevmişmişim
Şimdiki zamanın rivayeti	+(H)yörmHz	+Hm, +sHn, + ^{III} , +Hz, +sHnHz, +IAr	seviyormuşum
Gelecek zamanın rivayeti	+(y)AcAkmHz	+Hm, +sHn, + ^{III} , +Hz, +sHnHz, +IAr	sevecekmışim
Geniş zamanın rivayeti	+(A/H)rmsA	+Hm, +sHn, + ^{III} , +Hz, +sHnHz, +IAr	severmişim
Belirsiz geçmiş zamanın koşulu	+mHzsA	+m, +n, + ^{III} , +k, +nHz, +IAr	sevmişsem
Belirli geçmiş zamanın koşulu	+DHysA	+m, +n, + ^{III} , +k, +nHz, +IAr	sevdiysen
Şimdiki zamanın koşulu	+(H)yorsA	+m, +n, + ^{III} , +k, +nHz, +IAr	seviyorsan
Gelecek zamanın koşulu	+(y)AcAksA	+m, +n, + ^{III} , +k, +nHz, +IAr	seveceksem
Geniş zamanın koşulu	+(A/H)rsA	+m, +n, + ^{III} , +k, +nHz, +IAr	seversen

Şekil-9: Bildirme kipleri

4. Eylemin olumsuzunu oluşturma
5. Eylem kökünde düzeltme
6. Zaman ve birey eklerinin oluşturulması ve birleştirilmesi
7. Eylem kökünü yumuşatma
8. Eylemdeki değişken harfleri düzeltme
9. Eylemin soru biçimini oluşturma
10. Çekilmiş eylemi oluşturma

Bir eylem köküne zaman ve birey eki eklenirken kök eylemin son harfi değiştirilir. Bu değişiklik için uygulanan kural şöyledir:

- İlk aşamada kökün son harfinin a veya e olması durumu incelenir. Son harf a ise ı ve e ise i olarak değiştirilir.
- İkinci aşamada kökün son harfi ünsüz ise son ünlüye bakılır. Son ünlü a veya ı ise köke ı, e veya i ise i, o veya u ise o, ö veya ü ise ü eklenir. Bir anlamda bağlayıcı ünlüler kuralı uygulanır.
- Türkçe eylem kökleri içinde sonu t ile biten dört eylem kökünün (dit, et, git, güt, tat) son harfinde değişiklik yapılması gerekmiştir. Bu eylemlerin ŞİZ, ŞİZH, ŞİZR, ŞİZK, GEZ, GEZH, GZR, GEZR, GNZ, GNZH, GNZR ve GNZK kiplerinde son harfleri olan t $\Rightarrow$ d olarak değiştirilmiştir. Bunun nedeni zaman ekinin ilk harfinin ünlü ile başlamasıdır.

Kipler	Zaman eki	Birey Ekleri	Örnekler
Dilek - koşul	+sA	+m, +n, + ^{'''} , +k, +nHz, +lAr	sevsem
İstek	+A	+yHm, +sHn, + ^{'''} , +lHm, +sHnHz, +lAr	seveyim
Gereklik	+mAlH	+yHm, +sHn, + ^{'''} , +yHz, +sHnHz, +lAr	sevmeliyim
Emir	+ ^{'''}	+ ^{'''} , + ^{'''} , +sHn, + ^{'''} , +HnHz, +sHnAr	sev (2.tek)
Dilek - koşulun Hikayesi	+sAyD	+Hm, +Hn, +H, +Hk, +HnHz, +HlAr	sevseydim
İsteğin Hikayesi	+AyD	+Hm, +Hn, +H, +Hk, +HnHz, +HlAr	seveydim
Gerekliğin Hikayesi	+mAlHyD	+Hm, +Hn, +H, +Hk, +HnHz, +HlAr	sevmeliydim
Dilek - koşulun rivayeti	+sAymHz	+Hm, +sHn, + ^{'''} , +Hz, +sHnHz, +lAr	sevseymişim
İsteğin rivayeti	+AymHz	+Hm, +sHn, + ^{'''} , +Hz, +sHnHz, +lAr	seveymişim
Gerekliğin rivayeti	+mAlHymHz	+Hm, +sHn, + ^{'''} , +Hz, +sHnHz, +lAr	sevmeliymişim
Gerekliğin koşulu	+mAlHysA	+m, +n, + ^{'''} , +k, +nHz, +lAr	sevmeliysen

Şekil-10: İsteme kipleri

Bir eylem köküne zaman ve birey eki eklenirken eklerde yumuşama işlemi uygulanır. Türkçenin bildirme ve isteme kipleri Şekil-9 ve Şekil-10'da gösterilmiştir.

3.2 – Yapım Ekleri

TDK kaynakları Türkçede kök eylem sayısının 758 ve ana eylem sayısının yaklaşık 6523, ad soylu temel sözcüklerin sayısının yaklaşık 15.468 olduğunu göstermektedir. Türkçenin yapısı değerlendirilmeden üstünkörü bir yorumla Türkçenin söz varlığının cılız olduğunu söyleyenler vardır. Bu düşüncede olanlar yabancı dillerde olan birçok sözcüğün karşılığının Türkçede olmadığını söylerler. Eklemeli bir dil olan Türkçede bir kökten çok sayıda yeni sözcük türetilabilir ve türetilen her yeni sözcüğü ilk duyan kişi anlamını çıkarabilir. Çizelge-2'de göz kökünden türetilen sözcükler görülmektedir.

ad + ek $\Rightarrow$ ad	eylem + ek $\Rightarrow$ eylem
ad + ek $\Rightarrow$ eylem	eylem + ek $\Rightarrow$ ad
ad + ek $\Rightarrow$ önad	eylem + ek $\Rightarrow$ önad
ad + ek $\Rightarrow$ belirteç	eylem + ek $\Rightarrow$ belirteç

Çizelge-2: Göz kökünden türetilmiş sözcükler

Türkçe	İngilizce	Sınıfı	Türkçe	İngilizce	Sınıfı
Göz	Eye	Ad	Gözlemci	Observer	Önad
Gözlük	Eyeglasses	Ad	Gözlemcilik	Observation	Ad
Gözlükçü	Optician	Önad	Gözde	Favourite	Önad
Gözlükçülük	Opticians	Ad	Gözü	Mirror	Ad
Gözcü	Watchman	Önad	Gözümlük	Mirror stand	Ad
Gözcülük	Ophthalmology	Ad	Gözücü	Mirror maker	Önad
Gözlem	Observation	Ad	Gözücülük	Mirror makers business	Ad
Gözlemlemek	To observe	Eylem	Gözü	Who have eyes	Önad

Bir kök sözcüğe bir ya da birden çok ek eklenerek yeni sözcük türetilebilir. Yeni sözcük türetmek için 36 özellik belirlenmiştir. Bu özelliklerden biri seçilerek yeni sözcük türetilebilmektedir. Şekil-11'de 36 özellik ve türetilmesi istenen konu gösterilmiştir.

Yapım eki aşağıdaki aşamalardan köke geçilerek eklenir:

1. Hazırlık aşaması: Sözcük Türkçe mi?; Yüklenecek anlam; Kökün sınıfı
2. Eki oluşturma aşaması: Ses uyum kuralları; kaynaşma harfi
3. Yumuşatma aşaması
4. Eki köke ekleme aşaması

Örneğin

Kök: **arpa**;

Ek: **lık**

Ekin ilk harfi: l, değişken değil. Anlık ek: l

Ekin ikinci harfi H: değişken; ek tek seslemli, dolayısıyla aşağıdaki mantık bağıntısı kullanılarak H harfinin yeni değeri ı olarak bulunur. Anlık ek: lı olur.

$$\text{Seslem\_sayısı(ek)} = 1 \text{ ise } \begin{cases} S(\text{Kök}) (a \vee ı) \text{ ise } H \rightarrow ı \\ S(\text{Kök}) (e \vee i) \text{ ise } H \rightarrow i \\ S(\text{Kök}) (o \vee u) \text{ ise } H \rightarrow u \\ S(\text{Kök}) (\ddot{o} \vee \ddot{u}) \text{ ise } H \rightarrow \ddot{u} \end{cases}$$

Ekin son harfi: k, değişken değil. Anlık ek: **lık**

• Yer adı • Eyleme ilişkin konum • Geçilen yer • Eyleme uygun yer, bölge • Niteliğini anımsatan yer adı • Sevimli yerleşim yeri • Görece küçük yerleşim yeri • Barındırdığı özellik nedeniyle verilen ad • Yaşayanları anımsatan yer adı • Kullanım amacına ile ilgili yer adı • Kullanım amacına uygun yer adı • Belirtilen yer • Araç adı • Eyleme katkı sağlayan araç • Eyleme yardımcı olan araç • Eylemde yararlanılan araç • Eylemi uygulamaya yarayan araç • Nesne ile ilgili araç • Alet adı • Eylemde kullanılan Alet • Aygıt adı • Eylem ile ilgili aygıt • Meslek veya meslek aşamasının adı • Meslek adı • Meslek aşaması • Meslektaşlar topluluğu • Meslek sahibinin yaptığı iş • Yaptığı iş ile ilgili meslek adı • Uzmanlık derecesinde meslek adı • Bir düşünce akımını benimseme, taraftan olma • Bir düşünceyi benimseme • Düşünceyi temel alan görüş • Aynı düşünceye olanlar • Birliklilik • Birlikte olma • Benzer şeyler kümesi • Eşitlik, ortaklık, Bağlılık • Ortak özellikleri taşıma • Ortak özellikleri taşıyanlar topluluğu • Ortak nitelikleri taşıma • Aynı yer, kök veya soydan gelme • Bir nesneyi andıran, bir nesneye benzeyen • Üveylik durumu • Üveylik durumu • Huy, davranış, alışkanlık	• Benzerlik • Benzeme andırma • Benzer nitelik • Benzer davranış • Andıran davranış • Hastalık • Hastalık adı • Bitki • Bitki adı • Organ • Organ adı • Üleştirme • Üleştirme sayıları • Sıralama • Sıra sayıları • Eylemi sıralı yapma • İki eylemi birlikte yapma • Zaman • Belli bir an • Belli bir dönemde • Belli bir zaman aralığında • Belli bir zaman • Belli bir anda • Belli bir dilimi • Yapma bildiren eylem • Adın belirttiği yapma eylemi • Eylemin belirttiği yapma eylemi • Eylemsinin belirttiği yapma eylemi • Yansımalı adın belirttiği yapma eylemi • Önadın belirttiği yapma eylemi • Eylemden yapma eylemi • Yansımalı eylemden yapma eylemi • Olma bildiren eylem • Adın belirttiği olma eylemi • Eylemi yapan • Eylemi yapan kişi veya nesne • Eylemi umursamadan yapan kişi • Eylemi düşünmeden yapan kişi • Eylemi yerine getiren kişi • İş görev olarak yapan kişi • Ara sıra yapılan • Ara sıra yapılan eylem	• Huy, davranış, alışkanlık niteliği • Huy ve davranış alışkanlık biçiminde sürdürme • Eylem bağlantılı sonuç huy • Eylem bağlantılı huy, davranış • Küçültme, acıma, sevgi, aşığılama • Küçümseme • Küçük görme • Acıma ve sevgi • Görece zayıf • Aşığılama • Görece küçük • Dil ve lehçe • Dil veya lehçe • Süregelen • Süregelen eylem sonucunu tanımlayan ad • Bilinen durum • Dua ve dilek • Dua ve dilek • Yemek adı • Adını yiyenlerin verdiği yemek adı • Adını pişirme yönteminden alan yemek adı • Yön • Yön • Seven, Düşkün • Seven, yakınlık gösteren, düşkün • Bir özelliği taşıma • Ad ile ilgili niteleme önađı • Adın belirttiği niteliklere sahip olma • Bir şeye benzeme özelliği • Adın belirttiği konu ile ilişkili olma • Ad ile ilgili niteleme önađı • Adın belirttiği organ, nesne ile ilişkili olma • Önadın belirttiği niteliğe yaklaşma • Eylemin sonucunu niteleme • Eylemin sonucunu açıklama	• Eylem sonucu oluşan • Eylem sonucu oluşan yer, ortam • Eylem sonucu oluşan araç • Eylem sonucu oluşan istek • Eylemin edilgen öznesi • Eylem sonucu oluşan belge veya olanak • Eylemi andıran işlev • Eylem ile ilgili soyut kavram • Eylem ile ilgili kavram • Eylemde kullanılan Alet, araç • Eylemi çağırstıran sonuç kavram • Eylem sonucu ortaya çıkan sonuç ürün • Eylemi anımsatan teknik terimler • Eylemi yerine getiren özne • Eylemin gerçekleşmesi ile ortaya çıkan birim, parçası • Eylemin gerçekleşmesi ile ortaya çıkan nesne • Eylemin gerçekleşmesi ilgili işlem • Eylemin gerçekleşmesi ile ortaya çıkan durum • Eylem ile ilgili işin yapılışı • Eylem ile ilgili işin yapma işi • Eylemin uygulandığı yer • Eylem sonucu ortaya çıkan nesne • Eylem sonucu oluşan yapı, ürün • Eylemi çağırstıran yansımalı olay • Eylemi çağırstıran olay • Eylem ile ilintili sonuç • Eylemi çağırstıran yapı, konu • Eylemi çağırstıran yansımalı durum • Bazı nitelikleri taşıma • Bazı nitelikleri bulundurma • Nesne veya kişinin görünümünü andıran ad • Yapı • Yapı

Şekil-11: Yapım eklerinin amaca göre sınıflandırılması

4 - Türkçenin Söz Dizim Özellikleri

Bir dildeki söz varlığı sınırlıdır. Ayrıca her bir dilin tümce yapısı ve kuralları bellidir. Bu gerçeklere dayanarak bir dilde yazılabilecek tümcelerin yapay olarak türetilbileceği ve sayılarının sınırlı olacağı söylenebilir. Ancak insanın yetisi, dilin kurallarına uygun olarak sonsuz sayıda tümce üretilbileceğini göstermektedir. Çünkü tümceleri üretenler bireylerdir ve bireylerin değişik tür tümceler üretebildikleri görülmektedir. Üretilen her yeni tümcenin bir anlamı vardır ve bunu dinleyenler anlamaktadır. Bu yeteneğin insanın yaratıcılığından kaynaklandığı kabul edilmektedir.

Dil bilimci Wilhelm von Humboldt ve Noam Chomsky bir dildeki tümce sayısının sonsuz olduğunu da söylemektedirler. İnsanın yeni tümce üretme ve yeni tümceleri anlama yetisi N. Chomsky tarafından Language Acquisition Device (LAD) (Dil Edinim Aygıtı) olarak adlandırılmıştır.

EDB yaklaşımı her bireyin konuştuğu dilin yapısını belirleyen ilkeler ve değişkenleri doğuştan bildiğini varsayar. İlkeler soyuttur ve dilden dile değişmezler. Her dilin bir yapısı ve kuralları vardır, dolayısıyla bir şeyi anlatmak üzere kurulan tümcede sözcükler rastgele sıralanamazlar. Diğer bir deyişle bir şeyi anlatmak için

sözcükler belli sırada dizilerek tümce oluşturulur. İnsanlar bu kuralları bildiği için söyleneni anlayabilir.

Chomsky'nin EBD kuramını önerirken Türkçeyi yeterince incelemediği düşünülmektedir. **Çünkü Türkçede, özellikle söz dizimi esnektir. Diğer bir deyişle tümceler kalıp biçiminde değildir. Esnek kurulan bir tümcenin anlamını yitirmemesi Türkçenin durum eklerinin Türkçeye kazandırdığı bir yetenektir.**

Arabayı sabunla yıka.
Arabayı yıka sabunla.
Sabunla arabayı yıka.
Sabunla yıka arabayı.
Yıka arabayı sabunla.
Yıka sabunla arabayı.

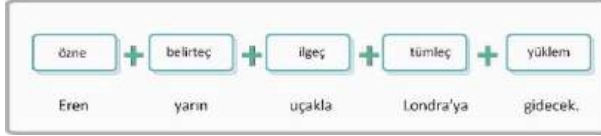
Türkçede tamlayan ve tamlanan sözcükler arasında bağ kurulurken, İngilizcede bu bağ kurulmamaktadır. Türkçe biçim bilimsel ve söz dizimsel özellikleri tümce içindeki sözcükler arasında sıkı bir bağ kurulmasını sağlamaktadır. Kurulan bağ açık biçimde görülebilmektedir. Bağın kuruluşu ve bağı kuran ögeler bellidir ve belli krallara uyularak sözcüklere eklenirler. **Bu özellikler Türkçeye tümce içinde sözcüklerin yerlerini değiştirebilme esnekliği sağlar ancak tümcenin anlamını bozmaz.**

5 - Türkçenin Söz Dizim Özellikleri

Türkçenin kurallı ve mantıklı olmasının sonucu olarak bir tümce bilgisayar tarafından türetilebilir. Bunun için Türkçe tümcenin;

- temel yapısı,
- özne özelliklerini
- tümce özellikleri ve
- yüklem özelliklerini bilmek yeterlidir.

Ölçünlü Türkçede tümce yapısı Şekil-12'deki gibidir.



Şekil-12: Ölçünlü tümce yapısı

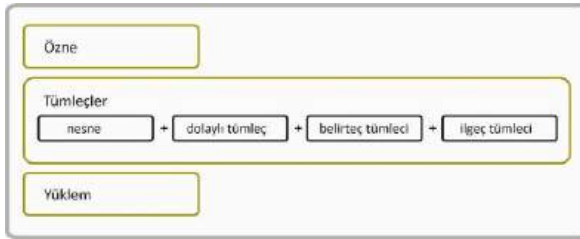
Özne yalın biçimde veya ad tamlaması ya da önad tamlaması biçiminde kurulabilir. Çeşitli özellikleri barındıran ad tamlaması biçiminde oluşturulan özne yapısı Şekil-13'te verilmiştir. Şekil-13'te görüldüğü gibi özne;

- Belirtili ad tamlaması
- Belirtisiz ad tamlaması
- Zincirleme ad tamlaması

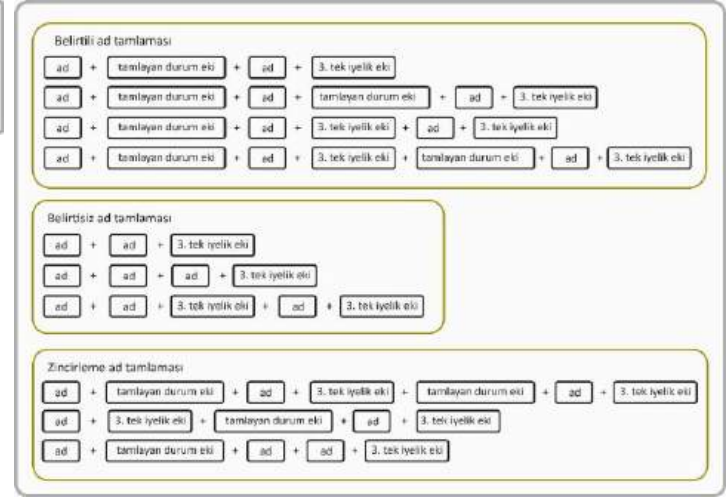
biçiminde kurulabilir.

Önad tamlaması biçimindeki özne yapısı Şekil-14'te verilmiştir. Şekil-14'te görüldüğü gibi dört farklı önad tamlaması ile özne kurulabilmektedir.

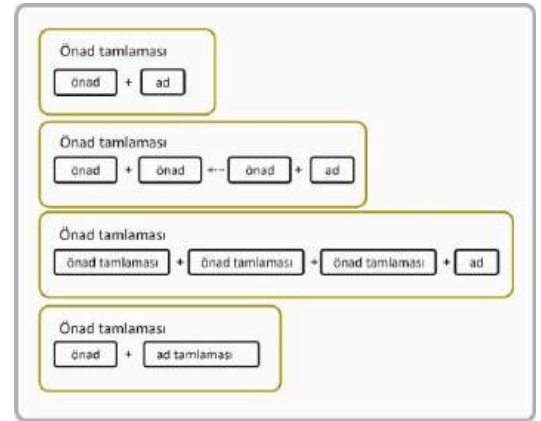
Tümleç yapısı Şekil-15'te çizilmiştir.



Şekil-15: Tümleç yapısı

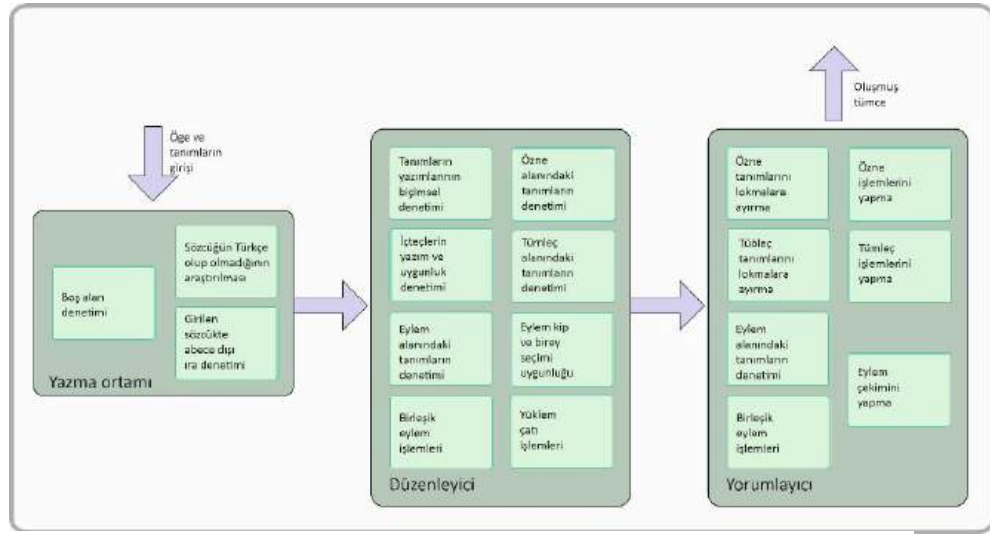


Şekil-13: Ad tamlamalı öznenin



Şekil-14: Önad tamlamalı öznenin yapısı

Tümce oluşturma yazılımı şu yazma ortamı, düzenleyici ve yorumlayıcıdan kuruludur. Bu yapı Şekil-16'da görülmektedir.



Şekil-16: Tümce oluşturan programın yapısı

Düzenleyici

Bu ortamda özne ve tümleç içinde yer alması gereken öğelerin tanımı yapılır. Ayrıca eylem kökü ve eylem ile ilgili olarak çatı ve birleşik eylem işlemlerinin seçimine olanak sağlanır. Doğal olarak eylemin zamanı, bireyi, olumlu ve olumsuz durumları belirlenir. Düzenleyici tanım ve seçimlerin programa girilebilmesini sağlamasının yanı sıra girilmiş olanların yazım ve mantık kurallarına uygun olup olmadığını denetler. Bu kapsamda şu denetimleri yapar.

- Tanım ve seçim yapılacak alanlara giriş yapıp yapılmadığı,
- Girilen sözcükler içinde abece dışında ıranın olup olmadığı,
- Girilen sözcüklerin Türkçe olup olmadığı,
- Tanımların yazımlarının biçimsel denetimi,
- İşleçlerin doğru yazılıp yazılmadığı ve yerinde kullanılıp kullanılmadığı,
- Özne içinde yer alan tüm tanımların doğru yapıp yapılmadığı,
- Tümleç içinde yer alan tüm tanımların doğru yapıp yapılmadığı,
- Yüklem Türkçe olup olmadığı,
- Çatı işlemleri için belirlenen seçeneklerin uygunluğu,
- Birleşik eylem için belirlenen seçenekler uygunluğu,
- Eylem kipi seçilmiş olup olmadığı ve birey seçiminin bu kip için uygunluğu

Yorumlayıcı

Tümce oluşturma ile ilgili gerekli bilgiler girildikten sonra yorumlama aşamasına geçilir. Bu aşamada, ilk olarak özne ve tümleç alanlarına girilen

tanımların çözümlemesi yapılır. Çözümleme işlemi kapsamında, ilgili alanlara girilen tanımlar lokmalarına ayrılır. A[Eren kedi ben] OY[sarışın kız] OY[güzel kedi] AY[okul yol] OT[Alpay güzel kız] OT[Alpay güzel kız -] ATD[kedi göz] OT[biz büyük ev -] ATD[geniş kapı] ZA[han kapı kol] ZB[han kapı kol]

6 - Sonuç ve Değerlendirme

Türkçenin dil bilgisi kuralları binlerce yıl önce oluşturulmuş ve günümüze kadar kurallar bozulmamıştır. Türkçede kural dışı bir durum yoktur. Dilin yapısının kurallı olması, kuralları mantık bağıntıları ile yazabilmeye olanak sağlamaktadır. Mantık bağıntılarından yararlanılarak akış çizgesi çizilebilmekte ve ardından program yazılabilmektedir. Yazılan programlar Türkçenin kurallarına göre aşağıda sıralanan programlar yazılmıştır:

- Sesleme
- Sözcüğün Türkçe olup olmadığını sına
- Sözcüğün son harfini yumuşatma ve eki buna göre düzenleme
- Bir adı tüm olasılıklara göre çekimleme
- Ek eylemi çekimleme
- Eylem çekimleme
- Amaca uygun olarak bir kök sözcüğe ek ekleme
- Tümce oluşturma

Yazılan programlar üzerinde Türkçenin tüm kök sözcükleri (ad ve eylem soylu) sınanmış ve Türkçenin kurallarının sağlamlığı gösterilmiştir.

Kaynakça

E. Adalı, Türkçenin Mantığı, Akademik Bilişim Yayınları, 2024

O'ZBEK SHEVALARI LEKSIKASINI KORPUSGA JOYLASHTIRISHNING DASTURIY TA'MINOTI

Ashirboyev Samixon

F.f.d., ToshDO'TAU professori

Turdialiyev Anvarbek

ToshDO'TAU o'qituvchisi

Annotatsiya. Jahon tilshunosligida shevalarni o'rganishga e'tibor kuchaydi. To'g'ri, o'zbek tilshunosligida shevalarni o'rganish masalasi uzoq tarixni bosib o'tgan bo'lsa-da, keyingi 30 yil ichida bu yo'nalishdagi ishlar ancha sustlashdi, lekin shevashunosligimizning bir tarmog'i bo'lgan lingvistik geografiya bo'yicha qator ishlar paydo bo'ldi. Ayniqsa, o'zbek shevalarining areal tadqiqotlari amalga oshirilganligi va bu sohaga doir darsliklari yaratilganligini alohida ta'kidlash mumkin. O'zbek tilshunosligida bugungi kunda dialektologik tadqiqotlar bo'yicha qilingan bir qancha tadqiqotlar bajarilyapti. O'zbek shevalarini o'rganish va asrab qolish yo'llarida amaliy loyihalar amalga ishirlilyapti. Dialektal korpus yaratish yo'lida qilingan ishlar ham bunga misol.

Abstract. In world linguistics, attention to the study of dialects has increased. True, although the issue of studying dialects in Uzbek linguistics has a long history, in the next 30 years, the work in this direction slowed down, but linguistic geography, which is a branch of our dialectology a number of works appeared. In particular, it is possible to emphasize that area studies of Uzbek dialects have been carried out and textbooks in this field have been created. In Uzbek linguistics, a number of dialectological researches are currently being promoted. Practical projects are being implemented to learn and preserve Uzbek dialects. An example of this is the work done to create a dialectal corpus.

Аннотация. В мировом языкознании возросло внимание к изучению диалектов. Правда, хотя вопрос изучения диалектов в узбекском языкознании имеет давнюю историю, в последующие 30 лет работа в этом направлении замедлилась, но по лингвогеографии, являющейся отраслью нашей диалектологии, появился ряд работ. В частности, можно подчеркнуть, что проведены региональные исследования узбекских диалектов и созданы учебники в этой области. В узбекском языкознании в настоящее время продвигается ряд диалектологических исследований. Реализуются практические проекты по изучению и сохранению узбекских диалектов. Примером тому является работа по созданию диалектного корпуса.

Kalit so'zlar: *dialekt, dialektal sheva, areal lingvistika, razmetkalash, subdialektal korpus.*

Dunyo tilshunosligi zamon bilan hamnafas rivojlanish bosqichida bo'lib, zamon bilan hisoblashmagan til egalari shu tilning kelajagini yo'qotib qo'yadi. Buni

dunyo tilshunosligida qilinadigan tadqiqotlardan ham kuzatish mumkin. Har qanday tilning kelib chiqish asosi va uning milliyligini saqlab turgan lingvistik bazasi mavjud bo'ladi. Lingvistik bazalar esa bu har bir tilning o'z shevasi (shevalari) hisoblanib, sheva sofligi qanchalik yuqori saqlansa, tilning ham yashovchanligi o'sib boradi. Hozirgi axborot texnologiyalar rivojlangan dunyoda tillar xilma-xil bo'lishiga qaramay u doimiy kamayib bormoqda, chunki mahalliy nutq shakllari kichik til oilalari, rivojlanmagan davlatlar tili asosiy muloqot vositasi bo'lgan rivojlangan (davlatlar) tillar ta'sirida o'lib bormoqda. Shu nuqtayi nazardan Amerika tilshunoslik maktabi (LSA) saytida tillarning yo'qolishi haqida quyidagicha jumlar keltiriladi, “*Tilni yosh bolalar o'rganishdan to'xtasa, uning (o'lim)kuni aniq sanaladi*”¹. Demak, bundan shunday xulosaga kelish mumkinki, har qanday tilni asrab qolish uchun til ma'lumotlarini bir joyga to'plash uni qayta ishlash tahlil qilish va uni keng omma foydalanishi uchun qulay tizimga tushirish kerak. Dialektologik korpuslar yo'qolish arafasidagi shevalarni asrashga mo'ljallangan dasturlardan biridir.

Dialektologik korpuslarning maqsadi dialektologiya, sotsiolingvistika, nutq madaniyati, fonetika-fonologiya, leksikologiya, grammatikaga oid tadqiqotlar hamda mahalliy aholining urf-odatlarini, tarixi va boshqa ko'plab sohalar uchun tadqiqotlar maydonini qulaylashtirish, kelajak avlodga qimmatli ma'lumotlarni yetkazish vazifasini bajarishdir.

Dialektologik korpuslarni tuzish ahamiyati shundaki, dunyoda tan olingan tillardan tashqari bir necha 10 ming so'zlovchiga ega bo'lsada iqtisodiy rivojlanishning pastligi sababli yo'qolib borayotgan til va ularning shevasini saqlab qolish hamda tadqiq qilishdir. Dialektlarni aniqlash va tasniflashning turli usullari mavjud bo'lib, bugungi kunda dunyo aholisi 7000 ga yaqin tillarda so'zlashayotgani taxmin qilinmoqda². Ushbu tillarning har biri ijtimoiy guruh yoki boshqa omillarga qarab farq qilishi mumkin bo'lgan bir nechta dialekt (sheva)lar bo'lishi mumkin.

Dunyoda qancha dialekt borligini aniq aytish qiyin. Bu qiyinchilik qisman tilshunoslarning dialekt nimadan iborat ekanligi haqidagi kelishmovchiligidan kelib chiqadi – ba'zi tilshunoslar tilning ikki shaklini turli dialekt deb bilishadi, boshqalari esa bunday fikrda emas. Shu o'rinda aytish joizki, dunyodagi dialektlarning miqdori juda katta bo'lib ana shu taxmin qilingan 7000 tildan faqat bittasi bo'lgan ingliz tilida 160ga yaqin lahjalar borligi taxmin qilinadi³. Ba'zi ma'lumotlarga ko'ra, dialektlarning umumiy soni 10 000 dan ortiq. Ammo shuni ta'kidlash kerakki, bu lahjalarning aksariyati rasman tan olinmagan yoki ro'yxatga olinmagan. Bunga misol sifatida Hindiston davlatini olishimiz mumkin. 2001-yildagi aholini hisoblash davrida Hindistonda 1 mlngacha so'zlashuvchiga ega bo'lgan 30ta til, 10 000 nafargacha so'zlashuvchiga ega bo'lgan 122ta til, 1599ta kam sonli so'zlashuvchiga

¹ <https://www.linguisticsociety.org/content/how-many-languages-are-there-world>

² <https://www.quora.com/How-many-languages-and-dialects-are-there>

³ <https://www.translateday.com/how-many-dialects-does-english-have/>

ega tillar borligi qayd qilinadi. 2011-yildagi hisob kitoblarda esa 121ta til borligi aniqlanadi va shu tillardan 22tasi Hindiston konstitutsiyasida til sifatida rasman tan olinadi⁴. Dialekt va lahjalarning miqdorini aniqlashning imkoni yo'qligi birgina hind tili misolida ham ko'rishimiz mumkin. Turkiy tillarda mavjud lahja va dialektlar ham shular sirasiga kiradi, chunki aholi migratsiyasi, iqtisodiy va ijtimoiy omillar bu farqlarni o'sishida hamda tillarning yo'qolishiga sabab bo'ladi. Korpus dialektologiyasining ikkinchi bir ahamiyati shu yo'qolish arafasidagi til(sheva)larni saqlab qolishdan iboratdir.

Jahon tilshunosligida yaratilgan dialektal korpuslarning yuzaga kelishiga nazar tashlaydigan bo'lsak: Xelsinki Britaniya ingliz dialekti korpusi (HD) Sharqiy Angliya, Janubiy-g'arbiy Lankashir orollari aholisidan to'plangan orfografik transkripsiyalangan audio yozuvlar to'plamini uchratishimiz mumkin.

Shevalar korpusini yaratish bosqichida shevashunosning tadqiqot maydoni doirasida matnlarni fonetik, leksik, grammatik jihatdan razmetkalash bosqichi dastlabki vazifa hisoblanadi. Razmetkalash deganda matn va uning komponentlariga maxsus teglarni biriktirib chiqish nazarda tutiladi. Maxsus teglar ikki xil bo'ladi: lingvistik teglar va ekstralingvistik (tashqi) teglar. Lingvistik teglar matn elementlarining leksik, grammatik va shunga o'xshash boshqa xususiyatlarini tavsiflaydigan ma'lumotlardan iboratdir. Ekstralingvistik teglar esa muallif haqidagi va matn haqidagi ma'lumotlarni (muallif, nomlanishi, nashr yili va joyi, janr, tematika va hokazo) tavsiflaydi⁵. Yangi tuzilgan korpuslar uchun razmetkalash qo'l mehnati asosida bajariladigan ish hisoblanadi. Razmetkalash uch yo'nalishda olib boriladi: kuzatish, fonetik razmetkalash, shevaga oid (dialektal) matnlarni razmetkalash. Kuzatish bosqichida matn olinadigan shahar, qishloq, aholining etnik tarkibi, respondentning kelib chiqishi, kasbi, ma'lumoti (o'rta, oliy) kabilar o'rin olsa, fonetik razmetkalashda unli (old qator, orqa qatorlik, singarmonizm, urg'u olish, olmaslik), undosh (portlovchi-sirg'aluvchi, jarangli-jarangsiz) tovushlar doirasidagi o'zgarishlar; dialektal matnlarni razmetkalashda⁶ sheva materiallarining leksik metarazmetkasi ham muhim hisoblanib har bir shevadagi so'zlar alohida teglanib adabiy til bilan bog'lab chiqiladi, dastlabki dialektal korpuslarda har qanday tildagi korpuslarda bo'lganidek har bir so'z (variantlari) sinonimi, omonimi, antonimi, paronimlari alohida teglanadi. Shevalar o'rtasidagi so'zlarni leksik-semantik teglash murakkab jarayon hisoblanadi. Grammatik razmetkalash yagona standartlashtirilgan grammatik birliklar mavjudligi shevalararo bog'liqliklar sabab ozmi ko'pmi bir-birini tushunishga olib keladi va bu jarayon yuqoridagi uch bosqishdan yengilroq sanaladi. Bundan tashqari, avtomatik teglash (razmetkalash) ham mavjud bo'lib bunda teglashni “tegger [Abdullayeva, 2022: 103] (teglovchi

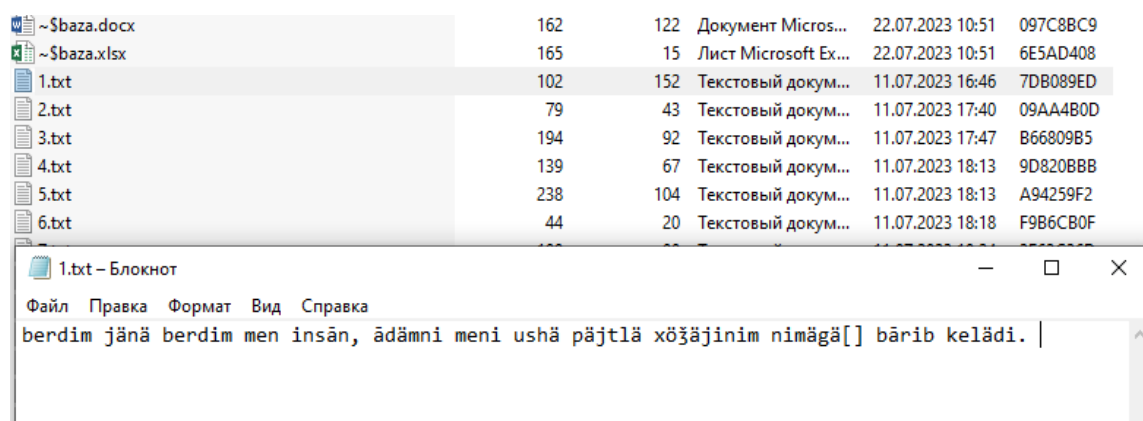
⁴https://en.wikipedia.org/wiki/Eighth_Schedule_to_the_Constitution_of_India

⁵ <http://compling.navoiy-uni.uz/index.php/conferences2/article/view/167/160>

⁶ Matnlarni razmetkalashda: nasriy matn, she'riy matn va aralash (she'riy-nasriy) matnlar ko'proq tadqiq qilingan bo'lib Dialektal korpusda tabiiy va ishonchliligini ta'minlash uchun dialog va monolog (audio) matnlarni tanlab oldik va shularni izohladik.

dastur)” bajaradi. Yuqorida sanalgan korpuslarda qisman razmetkalash amalga oshirilganligi bizning ishimizda ham o‘z o‘rnida foydalanildi. Til korpuslarida sheva ma’lumotlar bazasi shakllantirildi. Bunda avval Denov shevalari haqidagi ma’lumotlar jamlanib tahlil qilindi. Denov tumanida o‘zbek shevalarining qipchoq, qarluq lahjasida so‘zlashuvchi aholi va tojik tilida so‘zlashuvchi aholi vakillari yashaydi. Denov tumanida joylashgan 98ta (MFY) aholi yashash joylarini kuzatib, yashaydigan odamlar bilan suhbatlashish chog‘ida yig‘ilgan audio matnlar tahlil qilindi.

Subkorpusda Denov shevalari ma’lumotlar bazasining shakllantirilishi



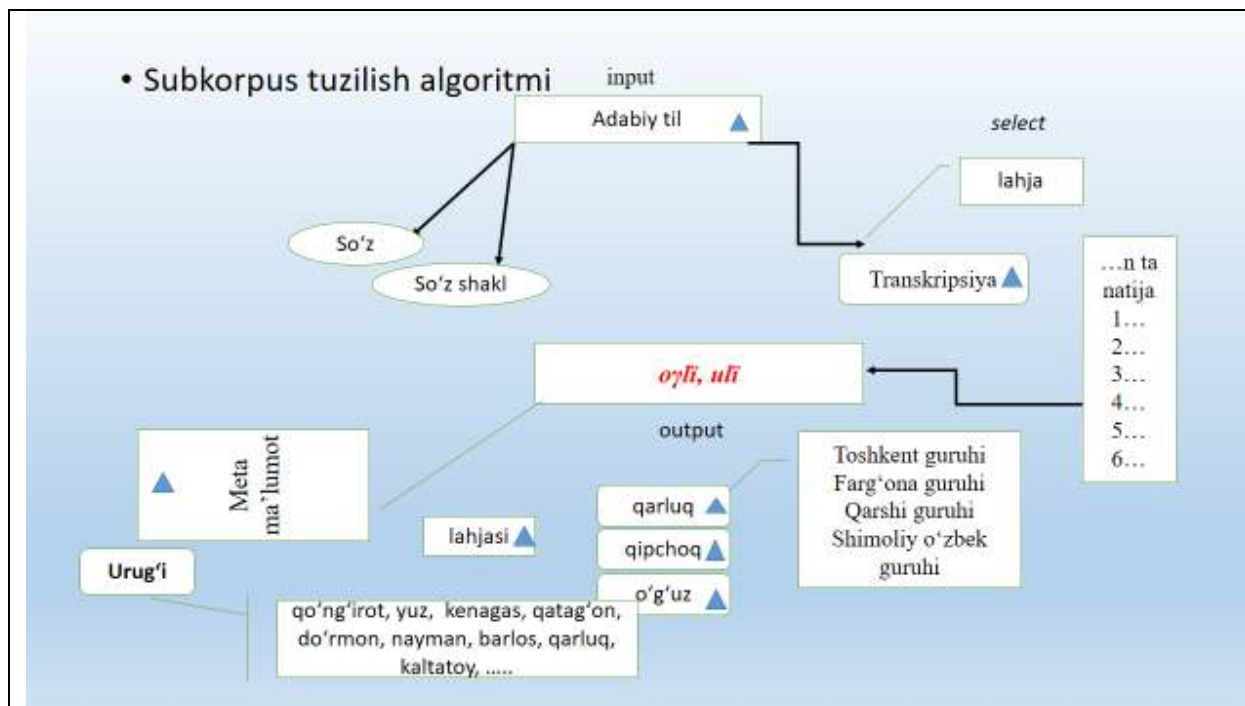
quyidagicha amalga oshirildi. Birinchi bosqichda diktafon, telefon so‘zlashuvlari so‘rovnoma o‘tkazish, anketa to‘ldirish usullari orqali to‘plangan dala ma’lumotlari (audio materiallar) transkripsiyada⁷ yozib chiqildi.

To‘plangan materiallar asosida dialektologik lug‘at tuzildi. Lug‘atda o‘zbek adabiy tilda uchramaydigan, adabiy tilda talaffuzi bilan farqlanadigan so‘zlar alohida meta ma’lumotlari bilan joylashtirib chiqildi. Lug‘atdan o‘rin olgan har bir so‘z uchun maxsus kompyuter dasturlari o‘qishi uchun mo‘ljallangan yagona yunikot tizimiga ega bo‘lgan *UTF-8 (txt)* faylda matnlar tayyorlandi (1-rasm)

Denov shevasining sheva ma’lumotlarini til korpuslari qidiruv tizimida joylanishi va uni amalda qo‘llanish bosqichlarini ishlab chiqishda shu kungacha yaratilgan dunyo korpusshunoslik va o‘zbek tilshunoslari tomonidan yaratilgan til korpuslari metodologiyasiga asoslanildi.

Yaratilgan subkorpus interfeysining **Input** qismida so‘zning adabiy tildagi varianti masalan: *o‘g‘li* so‘zi kiritilganda qidiruv tizimida joylashtirilgan ma’lumotlar (**select**) lahja bo‘limiga buyruqni uzatadi va **transkripsiyada** berilgan sheva materiallaridan nta (cheklanmagan) qidiruv natijalarini mavjud bo‘lgan matnlarni interfeys (**output**) ga chiqaradi. Olingan natija *o‘g‘li* so‘zining qipchoq va qarluq shevalaridagi barcha variantlarini so‘z shakllari bilan birgalikda aks ettiradi. Olingan natija haqida yuqorida ta’kidlaganimizdek meta ma’lumotlarni ham olishimiz mumkin.

⁷ Lotin alifbosidag transkripsiyasidan foydalandik.



Denov kichik dialektal zonasi shevalarining asosiy xususiyatlari, shevalarning bir-biriga ta'siri alohida o'rganildi va baza lug'atiga kiritildi. Subdialektal korpus tuzishda so'zlarning transkripsiyasi va adabiy tildagi transliteratsiyasi alohida izohlar bilan berildi. Har bir so'zning shevadagi varianti va o'zbek adabiy tilidagi sinonim va omonimlari alohida izohlar orqali teglandi. Bundan tashqari korpus imkoniyatidan kelib chiqib ma'lumotlar bazasi doimiy ravishda to'ldirib borishga imkoniyat yaratildi. Keyingi o'rinlarda korpusga joylangan audio materiallar avtomatik ravishda razmetkalanish ham imkoniyati ishlab chiqildi.

Xulosa o'rnida shuni aytish mumkinki, o'zbek shevalar korpusining dastlabki tajribasida quyidagi vazifalar bajarilganini ko'rishimiz mumkin. Korpusni sintaktik teggash va annotatsiyalashda universal tobelik nazariyasi metodologiyasi asosida o'zbek tilining sintaktik modellari aniqlanganligini hisobga olish zarur. Shuningdek, dialektal zona materiallari asosida subkorpus ma'lumotlar bazasi shakllantirilib, qidiruv oynasida natijalarni aks ettirilishi tajribalarda isbotlandi. Korpusni grammatik teggash bosqichida ontologik modellash usulining ilmiy asoslari isbotlangan. Korpusda qidirilayotgan so'zning imloviy xatolarini aniqlash va unga yaqin bo'lgan so'zlar koeffisientini belgilashda qidiruv interfeysida ilova qilishda yuqorida ta'kidlangan algoritmlarni qo'llashning maqsadga muvofiq ekanligi ta'kidlandi.

Denov shevasi ma'lumotlarining til korpuslari qidiruv tizimida joylashtirilishi va uning amalda qo'llanish bosqichlarini ishlab chiqish shu kungacha yaratilgan dunyo korpusshunoslik va o'zbek tilshunoslari tomonidan yaratilgan til korpuslari metodologiyasiga asoslanishi maqsadga muvofiq deb topildi.

Foydalanilgan adabiyotlar:

1. Ashirboyev S. O'zbek dialektologiyasi. – Toshkent: Nodirabegim, 2021. – B. 176.
2. Abdullayeva O. O'zbek tilining internet axborot matnlari korpusini shakllantirishning nazariy va amaliy asoslari: Filol. fan. bo'yicha falsafa doktori (PhD) diss. – Toshkent, 2022. – B. 103.
3. Abdurahmonova N. Kompyuter lingvistikasi. – Toshkent: Nodirabegim, 2021. – 394 b.
4. Холиёров Ў. Ўзбек тили таълимий корпусини тузишнинг лингвистик асослари: Филол. фан. бўйича фалсафа доктори (PhD) дис. автор. – Термиз, 2021. – 52 б.
5. Xidraliyeva Z. O'zbek tilining iqon shevasi. Filol. fan. bo'yicha falsafa doktori (PhD) diss. – Toshkent, 2023. – 166 b.
6. Холова М. Ўзбек миллий шевалари корпуси тадқиқи (Бойсун тумани “ж” ловчи шевалари мисолида). – Термиз: ТерДУ НММ, 2022. – 141 б.
7. Khidralieva Z. R. Vowel harmony features in ikan dialect. // ACADEMICIA: An International Multidisciplinary Research Journal, Vol. 10, Issue 12, December 2020. pp. 1177-1182.
8. Turdaliyev A. Denov kichik dialektal zonasining xususiyatlariga doir. Oltin bitiglar — Golden cript. ISSN 2181-9238// 2022/3

O‘ZBEK TILI PARALLEL KORPUSINING LINGVISTIK MASALALARI

Abduraxmonova Nilufar Zaynobiddin qizi

filologiya fanlari doktori, professor
O‘zMU Kompyuter lingvistikasi va amaliy
tilshunoslik kafedrasini mudiri

Vasliddinova Kamola

O‘zDJTU tayanch doktoranti

Annotatsiya. Tabiiy tilni qayta ishlash jarayoni uchun korpus texnologiyasini rivojlantirish dolzarb masalalardan biridir. Mashina tarjimasi, savol javob tizimi, tarjima nazariyasi va amaliyoti uchun parallel korpus texnologiyasini yaratishda lingvistik baza muhim resurs hisoblanadi. O‘zbek tili korpusi uchun ko‘p tilli parallel korpus yaratishning lingvistik masalalarni tizimli tadqiq etish va til bilan bog‘liq leksikografik tadqiqotlar haqidagi nazariy mulohazalar ushbu maqolada tahlilga tortilgan.

Abstract. Development of corpus technology for natural language processing is one of the important issues. A linguistic database is a significant resource in building parallel corpus technology for machine translation, response systems, and theory and applied translation. Systematic research of linguistic issues of creating a multilingual parallel corpus for the Uzbek language corpus and theoretical considerations of language-related lexicographic research are analyzed in this article.

Аннотация. Развитие корпусной технологии обработки естественного языка является одним из важных вопросов. Лингвистическая база данных является важным ресурсом в создании технологии параллельных корпусов для машинного перевода, систем реагирования, а также теории и практики перевода. Лингвистическая база данных является важным ресурсом в создании технологии параллельных корпусов для машинного перевода, систем реагирования, а также теоретического и прикладного перевода. В статье анализируются системное исследование лингвистических вопросов создания многоязычного параллельного корпуса корпуса узбекского языка и теоретические соображения языковых лексикографических исследований.

Kalit so‘zlar: *parallel korpus, kompyuter leksikografiyasi, mashina tarjimasi tizimlari, korpusga asoslangan terminologik ma’lumotlar bazasi.*

Bizning tadqiqotimiz uchun asos bo‘luvchi korpus strukturasi va uning kontentida parallel korpuslar ham o‘rin olgan. Shu bois parallel korpus va uni o‘zbek tili uchun yaratish masalalari ishimiz asosini tashkil qiladi. Ma’lumki, korpus faqat bir til doirasida yaratilgan bo‘lsa, monolingval korpus, boshqa tillar doirasida yaratilgan bo‘lsa parallel korpus deb nomlanadi. Parallel korpusda asliyatdagi matn

va uning muayyan tildagi tarjimasiga doir *paragraf, jumla yoki so'zlarning o'zaro muqobil juftliklari* bazaga kiritiladi.

Parallel korpus termini ostida mualliflar tomonidan muayyan tavsifiy xarakterdagi turlarga egaligi qayd etiladi [Kudela, 2016]:

- parallel korpus bir nechta tillardagi hujjatlarning sifatli tarjimalarini o'z ichiga olgan korpus hisoblanadi. Bunday korpuslarda tarjima muqobillari aniqlanib, gap darajasida ularni moslash nisbatan oson;

- “shovqinli” parallel korpus – bu u darajada sifatli tarjimaga ega bo'lmagan hamda tarjima muqobillari ham mukammal darajada muvofiqlashmagan bo'ladi. Shunday bo'lsa-da, ulardagi tarjimalarni korpus tarkibiga kiritish mumkin;

- qiyosiy parallel korpus – mavzu jihatdan mushtarak, biroq aynan matnning sinaktik va semantik jihatdan bir-biridan farqli bo'lgan korpus turiga tegishlidir. Ushbu holatda hujjatlar mavzu jihatdan umumiy guruhga tegishli bo'ladi. Muayyan mazmundagi asliyatdagi matn va uning tarjimalari baza sifatida korpusga kiritiladi. Masalan, qiyosiy korpus sifatida olingan turli OAVdagi yangiliklar muayyan vaqtga tegishli voqealarni aks ettiradi. Ushbu holatda matnlarning o'zaro muqobillik belgilari emas, balki so'zlarni mavzuni yoritishda qo'llash mahorati va matn strukturasini shakllantirish uchun qiyosiy-chog'ishtiruv aspektida o'rganiladigan sohalarning obyekti vazifasini bajaradi;

- kvazi-qiyoslash mumkin bo'lgan korpus – bu xilma-xil mavzudagi va grammatik jihatdan o'zaro parallel bo'lmagan hujjatlarni o'z ichiga oladi.

Opus platformasida ingliz-shved qiyosiy korpusi wikipediya olingan ma'lumotlar bo'yicha yaratilgan. Shuningdek, dunyoning turli mintaqalarida ingliz tilida so'zlashuvchi Britaniya ingliz tili, Avstraliya yoki hind ingliz tilining bir million so'zlik subkorpusdan iborat xalqaro korpusi bir xil namuna asosida jamlangani bois u *qiyosiy korpus* sifatida qabul qilinadi [Kubler, Zinsmeister, 2015].

Parallel korpusning mazmuniy tomonini hisobga olib, V.Zaxarov mazkur korpuslarni shartli ravishda ikkiga ajratadi:

- i.muayyan tildagi asliyat matnning ikkinchi tildagi aynan tarjimas;

- ii.biror sohaga oid mavzulashtirilgan guruhlar bo'yicha bir tasnifga kiritilgan matnlar.

Bugungi kunda parallel korpus yaratishning avtomatik texnologiyalari ishlab chiqilgan bo'lib, ularning aksariyati Yevropa tillari uchun instrumentariy vazifasini bajaradi. Shular qatoriga web sahifalardagi korpuslari generatsiya qiluvchi *Bitextor* instrumentariysini keltirish mumkin [Abdurakhmonova, Alisher, Toirova, 2022]. U yaxlit ravishda web-saytni yuklab olib, fayllardagi tillarni aniqlaydi va o'xshash mazmundagi matnlarni muvofiqlashtiradi, so'ng ularni html yoki xml formatda bazaga saqlaydi. Hujjatlar hajmi va muqobil tarjima birliklari bo'yicha tarjima xotirasiga (Translation Memory eXchange -TMX) uzatiladi [Kudela, 2016]. Bitextorning keyingi versiyalari yanada takomillashtirilib, URL (Uniform Resource Locator) va so'zlar to'plami (bag-of-words) modeli asosida filtrlash imkoniyatiga ham ega bo'ldi [Abdurakhmonova, Tuliye, Ismailov, Abduvahobov, 2022].

Avtomatik usulda parallel korpus yaratishning yana bir turi *PaCo2* instrumentariysi orqali ko'rsatish mumkin. Mazkur texnologiya internetdan parallel korpuslarni yig'ishga mo'ljallangan to'liq avtomatlashtirilgan tizim hisoblanadi [Kudela, 2016]. U ham tanlagan tillar bo'yicha bilingval kontentli web-saytlardan ma'lumotlarni yuklash imkoniyatiga ega. Tilni tanlash ixtiyoriy bo'lgani uchun ham ushbu instrument barcha tillar uchun moslasha oladi [Abdurakhmonova, 2019]. U uch asosiy bosqichga ega: 1- bosqichda bilingval kontentli web-saytlarni muayyan til bo'yicha qidiradi. 2- bosqichda dastlabki bosqichda aniqlangan saytdagi parallel web-sahifalar ajratiladi. So'nggi bosqichda parallel web-sahifalar parallel korpus sifatida har bir gapning muqobil juftlari bo'yicha guruhlanadi.

STRAND parallel korpus instrumentariysi hisoblanib, tarjimalarni strukturaviy aniqlash imkoniyatiga ega tizim hisoblanadi [Kudela, 2016]. Ushbu texnologiya web-sahifalardagi parallel juftlarini aniqlashga qaratilgan. U oddiy saytlardagi mazmuniy o'xshashligi bo'lgan matnlarni avtomatik tahlil qiladi

Bugungi kunda mashina tarjimasi tizimlari uchun ochiq platformali parallel korpus bazalari yaratilmoqda. Opus parallel korpus platformasi (<http://opus.nlpl.eu/>) lingvistik resurslar, korpus interfeysi va parallel matnlar bazasidan tashkil topgan. Qayd etish lozimki, ilk bor o'zbek tili uchun *Tanzil* deb nomlangan elektron resursda Qur'oni Karimning o'zbek tilidagi tavsiri parallel matn sifatida bazaga kiritilgan [Abdurakhmonova, Tuliye, Gatiatullin, 2019].

Shu o'rinda *CLARIN infrastrukturasida* 86 turdagi bilingval korpuslarning bazasi yaratilganini alohida qayd etish o'rinli. Undagi mavjud parallel korpuslar Yevropa va boshqa til oilalariga mansub til juftliklarining *hajmi, lingvistik razmetkasi, foydalanish imkoniyati, muqobillik konkordansi, matnlarning uslubiy ko'lami va undagi janrlar xilma-xilligiga* ko'ra farqlanadi.

Ochiq platformali parallel matnlar muhitini yaratish uchun CAT (computer assisted translation- kompyuter yordamidagi tarjima) texnologiyasi yordamida matnlarga ishlov berishning dastlabki bosqichi amalga oshiriladi. Mashina tarjimasining CAT texnologiyasi parallel matnlar asosida lingvistik resurs yaratishning eng maqbul usuli sifatida dunyo olimlari tomonidan tan olingan. Tarjima xotirasi muhiti yordamida ikki tildagi gaplarning o'zaro muqobillik (*alignment*) holati aniqlanadi, tarjima instrumenti SDL sifatida foydalaniladi. Instrumentariylar yordamida tarjimon "tarjima xotirasi" bankini yaratadi. Natijada professional tarjimon muayyan sohada tarjimoni amalga oshirishda tarjima qilingan matnlarni qayta ishlatishi, avval tarjima qilingan segment birliklarni mos tartibga ajratish bosqichi avtomatik tarzda amalga oshirishi, o'xshash tarjimalar yoki terminlarni ilova qilish imkoniyatiga ega bo'ladi. Tarjima instrumentariylari onlayn va kompyuter platformasiga moslashishi lozim. Kompyuter platformasi uchun pulli foydalanishga mo'ljallangan *Trados studio, memoQ, Wordfast, Memsource, Déjà Vu, Across, bepul xizmatlardan Omega T, Café TranEspresso, pulli onlayn xizmatlardan Memsource, Wordbee, XTM Cloud, tekin foydalanish imkoniyatiga ega Smartcat, Matecat va Wordfast Enywhere* kabilarni keltirish mumkin.

Tarjimonlik sohasida ham bir qancha lingvistik instrumentlar mavjud. Ular parallel tarzda tarjima muqobillarini serverda to'plash hamda bilingval terminologik ma'lumotlar bazasini yig'ishga mo'ljallangan. Parallel korpusga asoslangan bilingval terminlarni boyitib borish o'tgan asrning 80-yillariga borib taqaladi. Shu davrda yaponiyalik olim Makato Nagao namunaga asoslangan (example based translation) mashina tarjimasi texnologiyasini ilgari surgan. Endilikda ushbu turdagi texnologiyalar tarjima xotirasi (*TM-translation memory*) nomini oldi.

Korpusga asoslangan terminologik ma'lumotlar bazasini yaratish internet tizimidagi ma'lumotlarning parallel ravishda berilishi hamda mehnat resursining tejamkorligi jihatdan yuqori bo'lganligi bois g'arbda tadqiqot usuli sifatida an'anaga aylangan mazkur texnologiyadan maqsadli foydalanib kelinmoqda. Ushbu texnologiya mashina tarjimasi uchun mehnat samaradorligi va uni izchil ravishda boyitib borish imkoniyatiga egaligi bilan ahamiyatlidir. Shunday instrumentlardan biri *Paraconc* texnologiyasidir. U bilingval va ko'p tilli dastur bo'lib, quyidagi imkoniyatlarga ega [Ketiwe, <http://www.literator.org.za>]:

1. Turli usullarda so'zlarni saralash va hisoblash;
2. Tarjima qilingan parallel matnlarni moslashtirish;
3. Matndagi mavjud birliklarni oson topish va ularni ko'rsatish;
 - 1) terminlarni qidirish;
 - 2) tarjima ekvivalentlarini aniqlash;
 - 3) tarjimalardagi muhim o'rinlarni aniqlash;
 - 4) so'z birikmasi sifatida so'zlar o'ramini tahlil qilish;
 - 5) so'zlar chastotasi.

Kompyuter yordamidagi mashina tarjimasi texnologiyasi (CAT –computer assisted translation) ikki qismga ajralgan:

- 1) *Tarjima xotirasi* (Translation memory)
- 2) *Terminologik boshqaruv* (Terminology management)

Parallel matnlar asosida korpus yaratish u darajada oson emas. Chunki ikki tilning mos matn fragmentlari o'zaro juftlanadi. Ushbu masalani avtomatik tarzda hal qilish uchun turli metodlar ishlab chiqilgan, buning uchun ikki tildagi matndagi grammatik konstruktsiya (gaplar va so'z birikmalari), so'zlar muvofiqlashtiriladi. A.V.Zubov va I.I.Zubovalarning darsligida gaplar darajasida gaplarning muqobilligini aniqlash bo'yicha fikrlar o'rtaga tashlangan. Ushbu moslashtirish olti usulda amalga oshirilishi mumkin [3axapob, 2013]:

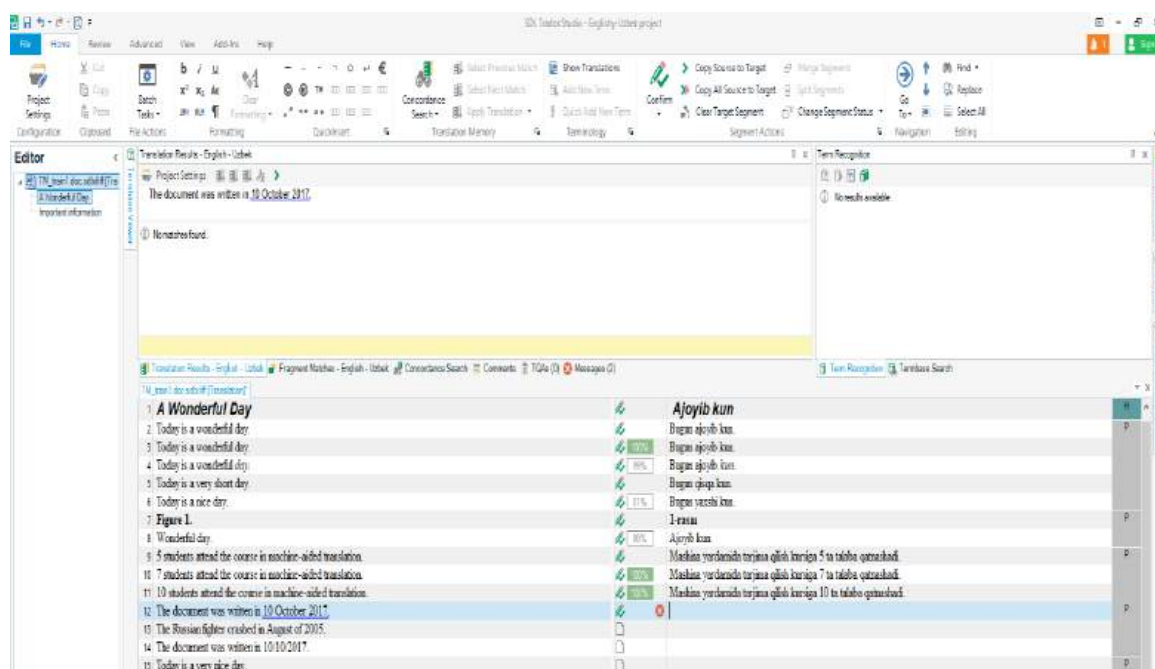
- 1) asliyatdagi bir matn ikkinchi tilga aynan hajmda tarjima qilinishi;
- 2) asliyatdagi ikki matn ikkinchi tilga bir matn bo'lib tarjima qilinishi;
- 3) asliyatdagi bir matn ikkinchi tilga ikkita matn sifatida tarjima qilinishi;
- 4) asliyatdagi ikki gap ikkinchi tilga ham ikkita jumla tarzida tarjima qilinishi;
- 5) asliyatdagi jumla ikkinchi tilga umuman tarjima qilinmasligi;
- 6) asliyatdagi matnning tarjima qilinadigan tilda mos ekvivalentining mavjud emasligi.

Parallel matnlarning mos juftliklarini tuzish uchun inson tomonidan tahrir qilingan matnlar kiritilib, lug‘at orqali matnda uchragan segment birliklarga qarab avtomatik tarzda tarjima birliklari aniqlanadi. Buning uchun ikki tilga mos leksik va grammatik jihatdan ishlab chiqilgan variantlar lug‘atga kiritilgan bo‘lishi kerak. Parallel korpus quyidagi maqsadlarda qo‘llanilishi mumkin [Захаров, 2013]:

- ikkit yoki ko‘p tilli tarjima lug‘atlar tuzishda;
- mashina tarjima tizimlari uchun lug‘at yaratish va uni doimiy tarzda to‘ldirib borish;
- kontekstda uchraydigan ko‘p ma’noli so‘zlarning kompyuter tahlili orqali leksik birliklarni polisemiya o‘ldirishini aniqlash;
- matndagi terminologik va frazeologik birliklarni tarjima qilish;
- korpusdan foydalangan holda tarjimalarning mos variantlarini kompyuter xotirasiga yuklash va shu orqali mashina tarjimasi tizimi uchun to‘liq avtomatik tarjimani amalga oshirish.

Leksikani avtomatik tasniflash matnni avtomatik tushunish uchun asosiy amallardan biri hisoblanadi [Беляева, 2004].

Tarjima instrumentlariga tarjima xotirasi texnologiyasiga (translation memory) asoslangan Trados dasturi hamda WinAlign tarjima texnologiyasi kiradi (CAT –computer assisted translation). Ushbu dasturlar *Translator’s workbench* va *MultiTerm* bazasi asosida yaratilgan. Bunda jumlar sifatida nuqta yoki gap uchun ishlatiladigan boshqa ishoraviy belgilargacha bo‘lgan qator gap sifatida segmentlarga ajratiladi [Abdurakhmonova va b., 2022]. Professional tarjimonlar orqali qilingan tarjima matnlar parallel ravishda kompyuter xotirasiga yuklanadi. Keyingi safar avvalgi qilingan tarjimaga o‘xshash matn yoki ibora uchrasa, kompyuter foiz ko‘rsatkichida segment o‘xshashlikni ilova qiladi. Bu esa tarjimonga muayyan yo‘nalishda tarjima qilinayotgan asarni qayta tarjima qilmay, bazada



mavjud ekvivalentlikni topish ixtiyorini voqelantiradi. Bunda ma'lumotlarni foydalanuvchilar tomonidan boyitib borilgan tarjimalar hisobiga kompyuter xotirasidagi asos va tarjima til uchun yaratilgan tayyor tarjima qoliplari hosil bo'ladi. Manbalarda qayd etishicha, ushbu tarjima uskunlaridan foydalanib inglizcha ilmiy matnlar nemis tiliga tarjima qilinganda, 28% tezroq mehnat samaradorligiga erishish mumkin [Marcos, Mihaela, 2014].

Quyida Trados dasturi orqali inglizcha matnni o'zbek tiliga tarjima qilish namunasi ko'rsatilgan (1-rasm). Tarjima xotiraning afzallik jihat – bir vaqtning o'zida tarjima qilingan so'zlar bilingval lug'at sifatida xotirada saqlanadi. So'ngra undan kontekstga asoslangan tarjima lug'at sifatida foydalanish mumkin. Tradosning yana bir ahamiyatli jhati – tarjima qilinayotgan matn .doc hamda HTML, SGML, XML, ASP, JSP, Tagged RTF, TRADOSTag kabi formatlarni ham o'qishga mo'ljallangan *TagEditor* dasturining mavjudligidadir. Bu esa asarni ayni aslyatdagi shaklida tarjima qilish imkoniyatini yaratadi

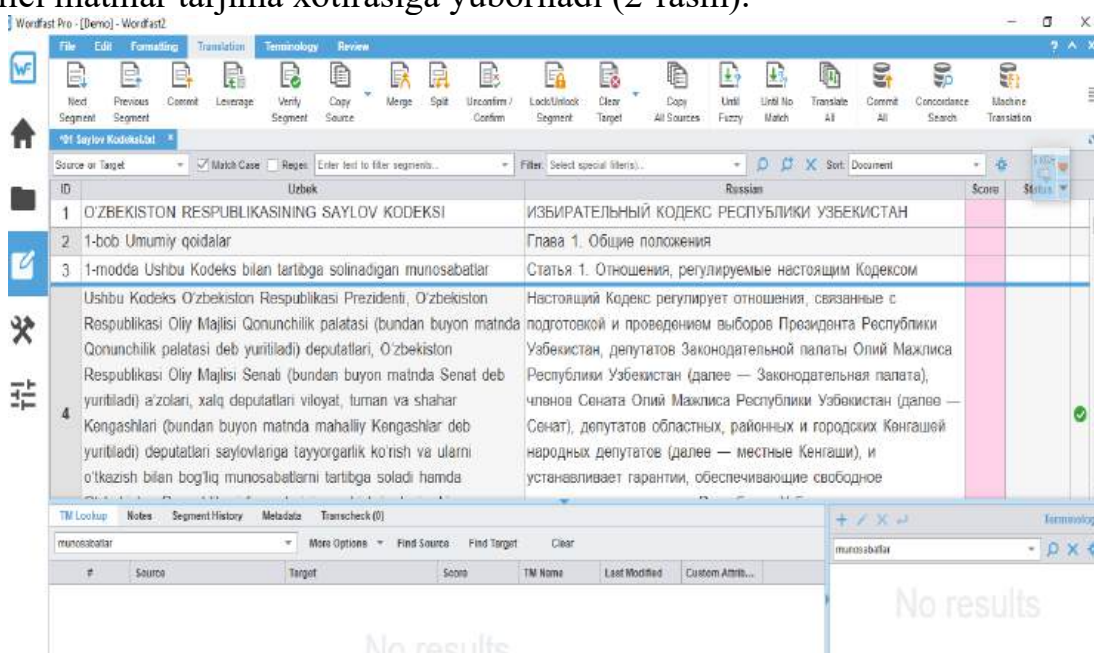
Mazkur tadqiqotda Wordfast Pro versiyasidan foydalanib, malakali tarjimonlar tomonidan o'zbek tilidan rus va ingliz tiliga tarjima qilingan matnlarning elektron shakldagi (.txt formatda) yuzdan ortiq ilmiy va rasmiy uslubdagi [<https://president.uz/uz> va <http://ziyonet.uz/>] parallel matnlar jamlandi (1-rasm).

ID	Russian ()	Score
1	ИЗБИРАТЕЛЬНЫЙ КОДЕКС РЕСПУБЛИКИ УЗБЕКИСТАН	-1
2	Глава 1. Общие положения	-1
3	Статья 1. Отношения, регулируемые настоящим Кодексом	-1
4	Настоящий Кодекс регулирует отношения, связанные с подготовкой и проведением выборов Президента Республики Узбекистан, депутатов Законодательной палаты Олий Мажлиса Республики Узбекистан (далее — Законодательная палата), членов Сената Олий Мажлиса Республики Узбекистан (далее — Сенат), депутатов областных, районных и городских Кенгашей народных депутатов (далее — местные Кенгаши), и устанавливает гарантии, обеспечивающие свободное волеизъявление граждан Республики Узбекистан.	-1
5	Статья 2. Законодательство о выборах	-1
6	Законодательство о выборах состоит из Конституции Республики Узбекистан, настоящего Кодекса и иных актов законодательства Республики Узбекистан.	-1

Ushbu matnlar tibbiyot, matematika, axborot xavfsizligi, tarjima, siyosat, filologiya kabi turli sohalarini o'z ichiga qamrab olgan. Ma'lumot o'rnida shuni aytish mumkinki, internetning imkoniyatlaridan unumli foydalanib, Webcrawler yordamida ko'p tilli saytlar yoki wikipediylarni ham tarjima xotirasiga yuklash mumkin. Manbalarda ta'kidlanishicha, yuklab olingan bilingval saytdagi muqobil parallel matnlar avtomatik generator Betextor yordamida TMX faylga to'planadi [http://lrec-conf.org/workshops/lrec2018/W3/pdf/5_W3.pdf]. Biroq generator faqat kam ta'minlangan resursli tillar uchun amal qilib, turkiy tillarning boshqa til juftliklari uchun mo'ljallangan emas.

O'zbek tilining o'ziga xos xususiyatlarini hisobga olib, internetda mavjud portallardan sohaga doir parallel matnlarni to'plash jarayonning dastlabki bosqichlaridan biri sanaladi.

Tarjimaning nechog'lik sifatli tarjima qilinishi so'z, so'z birikmasi yoki iboralarning uslub jihatdan to'g'ri tanlanganligi tahlil qilinib, segmentlash yoki tarjima jarayonida yo'l qo'yilgan xatolar tarjimon va tilshunoslar tomonidan qayta tahrirlanish jarayoniga ham bog'liqdir. So'z konkordanslarini aniqlash va ilmiy uslubdagi matnda qo'llangan terminlar uchun maxsus dastur yordamida ma'lumotlar bazasining shakllantirilishi, ajratib olingan parallel matnlardagi til va nutqiy birliklar alohida jadval shaklga keltirilishi tarjima dasturi uchun dastlabki bosqichlardan biridir. Segmentlash jarayoni to'liq yakuniga yetkanidan so'ng barcha parallel matnlar tarjima xotirasiga yuboriladi (2-rasm).



Shuningdek, ajratilgan gap segmentlaridan terminlar bazasi, so'z va so'z birikmalari avtomatik ravishda aniqlanib, inson tomonidan uning muqobillik darajasi tahrir qilinadi.

Ilmiy matnlardagi doimiy ishlatiladigan me'yoriy qoliplar sifatida quyidagi kalit so'zlarga e'tibor beriladi: so'z *asosi* => *derivatsion shakli* => *so'z shakli* => *ko'p qo'llanilgan so'z, fraza yoki terminlarning chastotasi* => *konkordansi* => *tarjimasi*. Namunada ko'rsatilgan matn “uzun” jumlati matn turiga kiradi, chunki unda tinish belgilar, qolaversa, predikativ elementlar miqdori ko'p bo'lib, ushbu matn tarjimon tomonidan ikki qismga ajratilib tarjima qilingan. Yirik matnli matndan ajratib olingan parallel jumalar generatsiyasi quyidagi bosqichlarda amalga oshirilishi mumkin [Копотев, 2014]:

I. Parallel matndan olingan tokenlashgan so'zning muqobil juftlari haqida ma'lumot olinadi;

II. Uzun jumlati matn muayyan tinish belgilari “,”, “;”, “:” bo'lgan qismigacha alohida segment gaplarga ajratiladi;

III. Asos – tarjima qilingan gaplarning segmentlari tarkibining miqdori aniqlanib, muvofiqlik holati belgilanadi;

IV. Tarjima – asos matnlarning muqobillik holati aniqlanadi;

V. Segmentlar o'rtasida birga ko'p yoki ko'pga ko'p munosabati kuzatilganda birga bir munosabat shaklida bir nechta segment birliklar o'zaro bog'lanadi yoki birlashtiriladi.

Parallel korpuslar yordamida so'z va so'z birlashtirishining o'zaro valentliklarini belgilash muayyan guruhga kiruvchi leksik-semantik guruhdagi so'zlarning semantik maydonini aniqlash, qolaversa, mashina tarjimai uchun tarjima sifatini oshirishga xizmat qiladi.

Foydalanilgan adabiyotlar:

1. Kudela J. Mining Parallel Corpora from the Web. (master's thesis). Prague, 2016. – P. 13.
2. Kübler S., Zinsmeister H. Corpus linguistics and linguistically annotated corpora. – New York: Bloomsbury Academic, 2015. - P. 21.
3. Kudela J. Mining Parallel Corpora from the Web. (master's thesis). Prague, 2016. – P.4 .
4. Kudela J. Mining Parallel Corpora from the Web. (master's thesis). Prague, 2016. – P. 7.
5. Ketibe N. Using ParaConc to extract bilingual terminology from parallel corpora: A case of English and Ndebele Literature / Journal of Literary Criticism, Comparative Linguistics and Literary Studies ISSN: (Online) 2219-8237, 0258-2279 <http://www.literatur.org.za>
6. Захаров В.П., Богданова С.Ю. КОРПУСНАЯ ЛИНГВИСТИКА. (Учебник) - Санкт-Петербург, 2013, - С. 20.
7. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In *2022 7th International Conference on Computer Science and Engineering (UBMK)* (pp. 73-75). IEEE.
8. Abdurakhmonova, N., Tuliyeu, U., Ismailov, A., & Abduvahobo, G. (2022). Uzbek electronic corpus as a tool for linguistic analysis. In *компьютерная обработка тюркских языков. Turklang 2022* (pp. 231-240).
9. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In *of the International Conference on Language Technologies for All (LT4All)*.
10. Abdurakhmonova, N., Tuliyeu, U., & Gatiatullin, A. (2021, November). Linguistic functionality of Uzbek Electron Corpus: [uzbekcorpus. uz](http://uzbekcorpus.uz). In *2021 International Conference on Information Science and Communications Technologies (ICISCT)* (pp. 1-4). IEEE.
11. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In *2022 7th International Conference on Computer Science and Engineering (UBMK)* (pp. 73-75). IEEE.



12. Беляева Л.Н. Лексикографический потенциал параллельного корпуса текстов // Труды международной конференции «Корпусная лингвистика – 2004» – СПб., 2004. – С. 55-64.
13. Marcos Z., Mihaela V. Quantifying the Influence of MT Output in the Translators' Performance: A Case Study in Technical Translation / Workshop on Humans and Computer-assisted Translation, Gothenburg, Sweden, 2014. – P. 93–98.
14. Копотев М. Введение в корпусную лингвистику. – Прага, 2014. – С. 129-132.
15. <https://president.uz/uz> va <http://ziyonet.uz/> manzilidagi yozma materiallar

ORONIMIK QURILMALARNING MILLIY KORPUSDA BERILISH MASALASI

Toirova Guli Ibragimovna

tugulijon@mail.ru

Filologiya fanlar doktori,
BuxDU professori

Annotatsiya. Maqolada XXI asr ilm-fani bu borada elektron korpuslar misolida keng tadqiq imkonini oronimlarning milliy lisoniy tabiati, ularni lug'atlashtirish hamda milliy korpusda berish masalasiga e'tibor qaratilgan. Toponimika sohasida amalga oshirilgan ilmiy tadqiqotlar, manbalarda oronimlar va oronimika tushunchasi, shuningdek, uning o'rganish obyektni belgilash masalasida turlicha fikrlarni uchratishi sababli yagona milliy korpusda ularni aks ettirish muammosi yoritilgan. Hozirda oronimlarni avtomatik aniqlashning quyidagi latent-semantik tahlil (Latent semantic analysis), parallel matnlarni tenglashtirish (Bitext word alignment) hamda o'zaro aloqador axborotni baholash (Mutual informatsion) metodlari orqali aniqlash mumkinligi taklif sifatiga ko'rsatilgan. Maqolada **oronims.uz** – web ilovasi o'zbek milliy korpusidagi oronimlar bazasini yaratish muammosi o'rta tashlangan.

Abstract. In the article, the science of the 21st century allows extensive research on the example of electronic corpora, focusing on the issue of the national linguistic nature of Oronim, their lexicography, and their inclusion in the national corpus. The problem of their reflection in a single national corpus is highlighted due to the scientific research carried out in the field of toponymy, the concept of oronims and oronymics in the sources, as well as meeting different opinions on the issue of defining the object of its study. Currently, the possibility of automatic identification of phonemes through latent-semantic analysis (Latent semantic analysis), parallel text alignment (Bitext word alignment) and mutual information evaluation (Mutual informative) methods has been shown to the quality of the proposal. The article highlights the problem of creating a database of oronims in the Uzbek national corpus of oronims.uz - a web application.

Аннотация. В статье наука XXI века позволяет проводить масштабные исследования на примере электронных корпусов, уделяя особое внимание вопросу национальной языковой природы оронимов, их лексикографии, включения в национальный корпус. Проблема их отражения в едином национальном корпусе освещается благодаря научным исследованиям, проведенным в области топонимики, понятию оронимов и оронимики в источниках, а также встрече различных мнений по вопросу определения объекта ее изучать. В настоящее время для качества предложения показана возможность автоматической идентификации фонем методами латентно-семантического анализа (Latent Semantic Analysis), параллельного

выравнивания текста (Bitekt word выравнивание) и оценки взаимной информации (Mutual информативный). В статье освещена проблема создания базы данных оронимов в узбекском национальном корпусе oronims.uz – веб-приложении.

Kalit soʻzlar: *oronim, toponim, gidronim, speleonim, relef, korpus lingvistikasi, latent-semantik tahlil, kollokatsiya, parallel matnlarni tenglashtirish, latent-semantik tahlil, web ilova.*

Kirish

Jahon tilshunosligida onomastika, xususan, toponimiyaning bir turi sifatida oronimiya ham xalq tarixi va til tarixi bilan chambarchas bogʻliq boʻlib, milliy til, xalqning iqtisodiy-ijtimoiy hayoti, madaniy-maʼnaviy munosabatlarining oʻziga xos xususiyatlarini aks ettiruvchi lisoniy manba hisoblanadi. Tilning sistematiy birligi boʻlgan mazkur lingvomadaniy birliklar tadqiqi ayniqsa, leksikologiya, leksikografiya, korpus lingvistikasida tez-tez koʻzga tashlanmoqda: XXI asr ilmi-fani bu borada elektron korpuslar misolida keng tadqiq imkonini yaratgan boʻlsa-da, oronimlarning milliy lisoniy tabiati ularnigʻatlashtirish hamda milliy korpusda berish ishlarini bir qadar murakkablashtiradi. Binobarin, Oʻzbekiston toponimik tizimidagi oronimik qurilmalarni qayta ishlanuvchi, lingvistik tasniflangan, mashinada oʻqiladigan metalingvistik maʼlumotlar formatida taqdim etish muammosi hanz dolzarbligicha qolmoqda.

Oʻzbek tilshunosligida toponimikaga oid ilmiy tadqiqotlarda oronimika tushunchasi, uning qamrovi va bu soha oʻrganadigan masalalar koʻlami aniq belgilanmagan va koʻrsatilmagan. Ilmiy adabiyotlarda bu masalaga doir quyidagi izohlarni kuzatish mumkin: “Oronimika – toponimikaning maʼlum bir til yoki region oronimlari majmuyini oʻrganadigan boʻlimi” [Dʼusimov, Tillaeva, 2002]; “Oronimika – toponimikaning oronimlarni vujudga kelishi, taraqqiyot qonunlarini va amal qilishini oʻrganadiga kichik boʻlimi” [Korayev, 2005]. E.Begmatov va N.Uluqovlar esa oronimikaga “Toponimikaning oronimlarni paydo boʻlishi va vazifaviy xususiyatlarini oʻrganuvchi tarmogʻi” deb izoh beradilar [Begmatov, Uluqov].

Yuqorida keltirilgan fikrlarni umumlashtirgan holda shuni taʼkidlash oʻrinliki, toponimiyaga oid ilmiy adabiyotlar va ilmiy tadqiqotlarda oronim va oronimiya tushunchalariga nisbatan quyidagi talqinlar uchraydi: 1) faqat togʻ (relefning musbat shakllari) nomlari; 2) relefning musbat va manfiy shakllari nomlari; 3) relefning musbat va manfiy hamda oraliq (nol) shakllari nomlari.

Oronim termini asli leksik jihatdan “togʻ nomi” maʼnosini anglatga-da, lekin uning maʼno doirasi ancha keng. Oronimlarga togʻlar, togʻ tizmalari, qirlar, dovonlar, tepalar, balandliklar, qoyalar, choʻqqilar, toshlar, adirliklar va togʻliklarda joylashgan boshqa musbat xarakterdagi orografik obyektlarning atoqli otlari bilan bir qatorda, oʻzining qaysi bir jihati bilan shu obyektlar bilan chambarchas bogʻliq

bo'lgan dara va uning turli ko'rinishlari, jar va uning turli ko'rinishlarini o'z ichiga oluvchi relefning manfiy shakllari nomlarini kiritish maqsadga mubofiq bo'ladi.

Oronimika onomastikaning ilmiy yo'nalishi sifatida tilshunoslikning dialektologiya, etimologiya, terminologiya, morfemika, til tarixi kabi bo'limlari bilan uzviy aloqador.

Hozirda oronimlarni avtomatik aniqlashning quyidagi metodlari orqali aniqlash mumkin:

1. Latent-semantik tahlil (Latent semantic analysis).
2. Parallel matnlarni tenglashtirish (Bitext word alignment).
3. O'zaro aloqador axborotni baholash (Mutual informatsion)[15].

Qo'llanilgan dasturlash tillari:

- PHP v7.2 (Laravel Framework v6.2)
- HTML
- SASS/CSS (Bootstrap v4)
- Javascript (Vuejs v2.5, Jqueru v3.4)

Qo'llanilgan ma'lumotlar bazasi:

- MuSQL v5.5

(Dasturning kod qismidan ba'zi lavhalar:

oronimController.php:

```
<?php
```

```
namespace App\Http\Controllers;
```

```
use App\Example;use App\oronim;
```

```
use App\oronimLang;
```

```
use Illuminate\Http\Request;
```

```
class oronimController extends Controller{public function__construct()  
{ $this->middleware('auth')->except(['index','search','show']); } }
```

An'anaga ko'ra parallel matnlarni birlashtirish uchun statistik usullardan foydalaniladi. Ular bir tildagi matnlar uzunligi va boshqa tildagi matnlar uzunligi o'rtasida so'zlar yoki belgilar bilan o'lchanadigan taxminiy muvofiqlikning mavjudligiga asoslanadi. Shuning uchun, bitta tildagi matnlarning uzunligiga qarab, boshqa tilda matnning mos parchalari qayerdan boshlanishi kerakligini hisoblash mumkin. Faqatgina nisbatan ko'proq mos keladigan muvofiqlikni tanlash qoladi. Ushbu usullarfaqat so'zlar soni yoki belgilar sonini hisoblashga asoslangani uchun tezkor usullar hisoblanadi.

Qayd etish kerakki, Facebook, Pinterest, Twitter ijtimoiy tarmoqlari semantik razmetkasiz ish ko'rmaydi. Jumladan, ushbu web-xizmatlar mavjud maqolalarning bir qismini oladi va ularni o'z platformalarida namoyish etadi. Agar biz semantik HTML dan foydalansak, ular bu ishni yaxshiroq, sifatliroq bajarishlari mumkin.

Semantik HTML beb-saytlarga kirish imkonini oshiradi.

Ekrandan o'qish imkonini beruvchi kabi yordamchi texnologiyalar mavjud HTMLni tahlil qiladi va sharhlaydi (talqin qiladi). Semantik HTML yordamida maxsus ehtiyojlari bo'lgan odamlar tegishli maqolalarni o'qishlari va yo'nalish olishlari osonroq bo'ladi. Biroq bu hammasi emas. HTMLni tushunishga harakat qiladigan son-sanoqsiz boshqa mashinalar ham mavjud. Ular internetning katta qismini tashkil qiladi va ularni yanada mazmunli ma'lumotlar bilan ta'minlash katta hajmli mehnatni talab qiladi.

Bu taxminan quyidagicha manzaraga ega:

```
<!DOCTYPE html>
<html itemscope itemtype="https://schema.org/Article" lang="" dir="">
<head>
<title itemprop="name"></title>
<meta itemprop="description" name="description" content="">
<meta itemprop="author" name="author" content="">
</head>
<body>
<article>
<header>
<h1 itemprop="headline"></h1>
<time itemprop="datePublished" datetime=""></time>
<p><a itemprop="author" href=""></a></p>
</header>
<div itemprop="about"></div>
<div itemprop="articleBodu">
<!-- maqolaning asosiy matni bu yerda bo'lishi kerak-->
</div>
<footer>
<time itemprop="dateModified" datetime=""></time>
<section itemscope itemtype="http://schema.org/WebPage">
<!-- Sarlabha-->
<h2></h2>
<p><a itemprop="relatedLink" href=""></a></p>
</section>
</footer>
</article>
</body>
</html>
```

Orografik obyektlarni ifodalovchi atamalar va oronimlarning tarkibida indikator vazifasida qo'llanuvchi quyidagi so'zlarni ko'rsatish mumkin: tog', tosh, qoya, qiya, cho'qqi, qir, tepa, dobon, bel, sirt, adir, ko'tal, ort, do'ng, cho'l, gaza, dara, tangi, zob, jar, ko'h, kamar, to'da, sang, dasht, mo'la, nura, o'r, o'ra, maydon, aydar, mas//mos, taxta, alang, qo'l, shoti, ayri, suyri, aqba//ag'ba, yaylov, yayloq,

chobra, bitob//pitob, xuchchi, tubak, qaloq, do'm, bo'ktar, zayak, hayat, qashqa, qat, qatob, qoboq, singir, sala, jun//jo'na, to'qay, chag'il//chag'al, chag'at, supa, yona, shiram, chuqur, xandaq, g'urum, qum, chamber, qaqir, taqir, dagar, jol, to'la, bayir, sho'r, bet, habol, g'or, poya, qo'ruq, tuz, chalqa, tumshuq, qapug', darboza, darband, qiz//quz, ti//tig'//teg', baroz//barzu, shirdon, bo'z, jira, qayir, bo'yin, bo'ynoq, boldir, changal//jangan, pushta, qasoba//qasag'a, qasnaq, orqa, orqabul, burun, oshub, qayqi, o'y, toqcha, qapchig'ay, yobon, xobar, jobuz, zarang, sahro, chortoq, sangir, bodiya, mag'ora, barg'//barq, dapsan, tarma, tirsak, chosh//shosh, chek, ungur, chordara, chig'anaq va b.

Shablonning HTML-razmetkasi article, header, footer singari semantik elementlardan iborat.

Bilamizki, korpus lingvistikasi:

1. Nutq tadqiqiga doir natija va xulosalarni aniqlash;
2. Empirik nutqiy materialning qamroviga oid tizimli va keng ko'lamli lingvistik tadqiqotlarni amalga oshirishni maqsad qiladi.

Shunga ko'ra korpusda qamralayotgan material birliklarining razmetkasi lingvistik va tashqi lingvistik turlariga bo'linadi. Bugungi kunda razmetkaning standartlashgan vositalarini ishlab chiqishga mo'ljallangan maxsus xalqaro Text Encoding Initiative (TEI) loyihasi amalga oshirilgan bo'lib, bunda elektron hujjatlarning razmetkasi uchun SGML xalqaro razmetka tili yoxud XMLning turlari qo'llaniladi.

Oronimik konversiya usulida hosil bo'lgan oronimlarni ikki guruhga ajratish mumkin:

1. Tub asosdan hosil bo'lgan oronimlar: a) orfografik terminlar oronimlashadi: Ag'ba (Qm., dovon), Kamar (tog'li hududlarda, joy), Gaza (Bs., joy), Kachar (G'z., adir), Sangar (Mk., yaylov), Tarpi (Chr., balandlik), Taqir (Do., balandlik); Ko'tal (Chr., tog', q-q.), Chimgan (G'z., adir), Qayg'i (Bs., adir), Qasaba (Do., balandlik), G'urum (Do., balandlik), G'urim (Do., balandlik), Hasar (Chr., balandlik); b) etnonimlar oronimlashadi: Barmoq (Qm., yaylov, balandlik), Lochin (Shs., cho'qqi), Manqa (Do., adir), O'lmas (Ko'kabuloq, tog'), Uyqur (Do., dara); b) gidrografik terminlar oronimlashadi: Kariz (Kt., adir), Orna (G'z., jar).

2. Yasama so'zlardan hosil bo'lgan oronimlar. Bu guruhga kiruvchi oronimlarni tarkibiy tuzilish xususiyatiga ko'ra yana ikki ichki guruhga ajratish mumkin: a) sodda tarkibli yasama so'zlardan hosil bo'lgan oronimlar: Abrikli (CHr., adir), Andizli (Chr., Qm., Bs., dara), Arnali (G'z., adir), Ahalli (Kt., adir), Bedak (Kt., joy), Burg'onli (Yab., adir), Dubolli (Bs., dara), Jo'jali (Do., dara), Irg'ayli (Qm., adir), Isbanni (Qm., Yab., yaylov, joy), Kapali (G'z., SHs., Mb., yaylov, dasht), Ko'hak (Qo'ng'irtog'ning qadimgi nomi), Surxak (Kt., balandlik), Xabdonli (G'z., yaylov), Chaqishli (SHo., adir), Cho'yanli (YAv., Bs., adir), Sherak (SHo., adir), O'rdali (Jq., tepa), Qazma (G'z., yaylov), Qayroqli (Sho., balandlik), Qarg'ali (Shs., tepa; So., adir), Qarg'anli (Shs., adir) ba b.;

b) qo'shma tarkibli so'zlardan hosil bo'lgan oronimlar: Abg'ontaxta (Do., balandlik), Ajinazob (Qm., balandlik), Ayg'irtepa (G'z., Do., Bs., tepa), Aylanatumshug' (Kt., balandlik), Archamaydon (Shs., tog'), Asqartepa (Yab., tepa), Asqarcho'qqi (Yab., cho'qqi), Baland qo'ton (Chr., adir), Bag'ishtepa (Os., tepa), Beshqiz (Qm., tepa), Beshqo'l (Do., dara), Boyirgaza (Chr., gaza), Bo'rijar (Shs., jar), Bo'g'masoy (Do., dara), Go'rdara (Do., dara), Do'g'irtosh (Dn., balandlik), Etimtaxta (Do., tog') va b.;

b) izohli birikmalardan hosil bo'lgan oronimlar: Bog'ibaland, Go'rimor, Go'riqalmoq, Darayiturko, Darixitoy, Dashtikalta, Dashtichinor, Dahnidarg'oz, Zaxakimaron, Kamariturk, Kamarichiljubut, Ko'hisafed, Ko'hisiyoh, Ko'hichosh, Xamikalon, Chashmiganda, Shibarikalon va b.

g) predikativ birikmalardan hosil bo'lgan oronimlar: Xontushdi (tepa), Lo'limurda (adir), Shuturmurda (dara), Xarduzd (tog'), Darg'amurda (adir), Asppar (dara), Xamixukkushtagi (adir), Ayiquchdi (dara).

Biz taklif qilayotgan milliy oronimlar bazasining dasturiy ta'minoti quyidagilarni o'z ichiga oladi: **oronims.uz** – web ilovasi o'zbek milliy korpusidagi oronimlar bazasini tashkil etish va undan qulay foydalanish imkonini beruvchi tizim hisoblanadi. U o'zida oronimlar bazasini to'ldirib borish va so'ralgan oronimni tushunarli va qulay tarzda batafsil ma'lumotlar asosida taqdim etish funksiyalarini mujassam etgan. Ilova tarmoqqa (internet) ulangan server vazifasini o'tay oladigan kompyuterda ishlashga mo'ljallangan. Unga HTTP protokoli orqali murojaatlar yuborilishi va javob olinishi ko'zda tutilgan. Barcha web ilova (sayt)lar kabi **oronims.uz** web-ilovasi ham o'zi o'rnatilgan server kompyuterda web-server paketlari ishlashini taqozo etadi. Ushbu bazaning texnik bosqichini yaratishda quyidagi dasturlash tillari tanlandi.

Dastur turi: Web ilova (web application).

Qo'llanilgan dasturlash tillari:

PHP v7.2 (Laravel Framework 6.2) HTML SASS/CSS (Bootstrap v4)
Javascript (Vuejs v2.5, Jqueru v3.4)

Qo'llanilgan ma'lumotlar bazasi: MuSQL v5.5

Xulosa

Xulosa o'rnida shuni aytish joizki, til korpusidagi oronimlar bazasi ta'limga, ayniqsa, til ta'limiga doir quyidagi tipik xatolarni bartaraf etish imkonini beradi. Jumladan, taqdim etilgan so'zlar qatoridan so'zlarni to'g'ri tanlashni bilmaslik; oronimik birliklarni bilmaslik; mashhur oronimlardan bexabarlik.

Foydalanilgan adabiyotlar:

1. Авлакулов Я.И. Ўзбек тили ономастик бирликларининг лисоний тадқиқи: Филол. фан. номз... диссер автореф. – Тошкент, 2012. – Б. 27.
2. Бегматов Э., Улуков Н. Ўзбек ономастикаси терминларининг изоҳли луғати. – Наманган, 2006. – Б. 63.
3. Дўсимов З., Тиллаева М. Топонимика асослари. – Тошкент, 2002. – Б. 82.
4. Исломов И. Ўзбек тилида ер сатҳи кўринишларини ифодаловчи атов бирликлар тизими: Филол. фан. фалсафа докори (PhD) диссертацияси. – Қарши, 2019. – 124 б.
5. Суперанская А.В. Что такое топонимика? – Москва: Наука, 1984. – С. 22.
6. Подольская Н.В. Словарь русской ономастической терминологии. – Москва: Наука, 1978. – С. 105.
7. Пospelов Е. М. Оронимика СССР: состояние и проблемы// Оронимика (Сборник статей). – Москва, 1969. С. 4
8. Темиров Ш. Самарқанд вилояти оронимлакрининг лисоний тадқиқи: Филол. фан. ном. ... дисс. автореф. – Самарқанд, 2019. – 24 б.
9. Каримов С.А., Бўриев С.Н. Ўзбек топонимикаси тараққиёт босқичида. – Тошкент: Фан, 2006. -Б. 44.
10. Корпусная лингвистика как раздел языкознания. — Текст : электронный // MyfilologY.ru – информационный филологический ресурс : [сайт]. – URL: <https://myfilology.ru/177/korpusnaya-lingvistika-kak-razdel-yazykoznanieYa/> (дата обращения: 5.02.2021)
11. Қораев С. Географик номлар маъноси. – Тошкент: Ўзбекистон, 1978. – Б. 193.
12. Qorayev S. Toponimika. – Toshkent: O'zbekiston faylasuflari milliy jamiyati, 2006. – B. 66.
13. Қораев С. Ўзбекистон вилоятлари топонимлари. – Тошкент: Ўзбекистон миллий энциклопедияси, 2005. – Б. 217.
14. Эназаров Т. Номшунослик масалалари. – Тошкент, 2010. – Б.78.
15. Шильнов, А. Г., Миллер, Д., Митчелл, П. Д., & Смокотин, В. М. (2020). Сравнительный анализ идиом, представленных в российских Учебниках английского языка, и идиом, представленных в современных аутентичных источниках. Язык и культура, (49), 157-174. <https://doi.org/10.17223/19996195/49/10>
16. Шоабдирахмонов Ш., Асқарова М. ва б. Ҳозирги ўзбек адабий тили. – Тошкент, 1980. – Б. 182.
17. Ўзбек тили грамматикаси. 1-қисм. – Тошкент, 1975. – Б.34.
18. Toirova G. The Role Of Setting In Linguistic Modeling. International Multilingual Journal of Science and Technology. ISSN: 2528-9810 Vol. 4 Issue 9, September – 2019, -P.722-723



19. Toirova, G. Creation and importance of language corps in Uzbekistan. BIO Web of Conferences, 2024, 84, 04003.
20. Toirova G., Astanova G., Rahimova N. Artistic Expressions of a Situational Pragmatic SYstem. // International Journal of Recent TechnologY and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-3, September 2019. –P.4591-4593
21. Toirova G., Yuldasheva M., Elibaeva I. Importance of Interface in Creating Corpus. // International Journal of Recent TechnologY and Engineering, 2019, 8(2 Special Issue 10), P. 352–355
22. Toirova G, Abdurahmonova N., Ismoilov A., Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. // Proceedings - 7th International Conference on Computer Science and Engineering, UBMK 2022, 2022, P. 73–75
23. O'zbek Milliy Korpusida idiomalar bazasi deponentlanadigan materiallari. Huquq egasi Begmatova Gulchehra Xolmurodovna. Mualliflar: Begmatova Gulchehra Xolmurodovna. Ulug'ov Anvar Abduraimovich. – 2019.

O'ZBEK TILI KORPUSI MATNLARINI POS TEGGLASH USULLARI

Botir Elov Boltayevich

Texnika fanlari bo'yicha falsafa doktori PhD, dotsent

elov@navoiy-uni.uz

ToshDO'TAU Kompyuter lingvistikasi va raqamli
texnologiyalar kafedrası mudiri

Nizomaddin Xudayberganov Uktambay o'g'li

nizomaddin@navoiy-uni.uz

ToshDO'TAU Kompyuter lingvistikasi va raqamli
texnologiyalar kafedrası o'qituvchisi

Annotatsiya. Til korpusi qurilishida lingvistik ta'minot masalasi muhim va murakkab hisoblanadi. Korpuslarda matnlardagi nutq bo'laklariga mos identifikatorini belgilash jarayoni muammolidir, sababi tilni modellash tirish teglash qoidasi va tilda mavjud qonuniyat bilan bog'liq. Teglash, xususan, grammatik teglash yoki PoS tegging o'zbek korpus lingvistikasi uchun ham dolzarb masaladir. Ushbu maqolada jahonda keng qo'llanib kelinayotgan tegsetlardan foydalanib o'zbek tili korpusi matnlarini POS teglash usullari ko'rib chiqiladi.

Abstract. In the construction of the language corpus, the issue of linguistic support is important and complex. The process of assigning appropriate identifiers to speech fragments in texts in corpora is problematic, because language modeling is related to tagging rules and regularities in the language. Tagging, especially grammatical tagging or PoS tagging, is also a relevant issue for Uzbek corpus linguistics. In this article, the methods of POS tagging of texts of the Uzbek language corpus using tagsets that are widely used in the world are considered.

Аннотация. При построении языкового корпуса важным и сложным является вопрос языкового обеспечения. Процесс присвоения соответствующих идентификаторов речевым фрагментам в текстах в корпусах проблематичен, поскольку моделирование языка связано с маркировкой правил и закономерностей в языке. Маркировка, особенно грамматическая маркировка или маркировка PoS, также является актуальной проблемой для узбекской корпусной лингвистики. В данной статье рассмотрены методы POS-разметки текстов корпуса узбекского языка с использованием наборов тегов, которые широко используются в мире.

Kalit so'zlar: *Teg, razmetka, annotatsiya, tegset, NLP, korpus, PoS teglash.*

Kirish

Ma'lumki, istalgan turdagi korpusni tuzish mavjud datani teglashdan boshlanadi. Teglash jarayoni annotatsiya, razmetka kabi so'zlar bilan turli adabiyotlarda yonma-yon, ba'zan sinonim sifatida ishlatilgan.

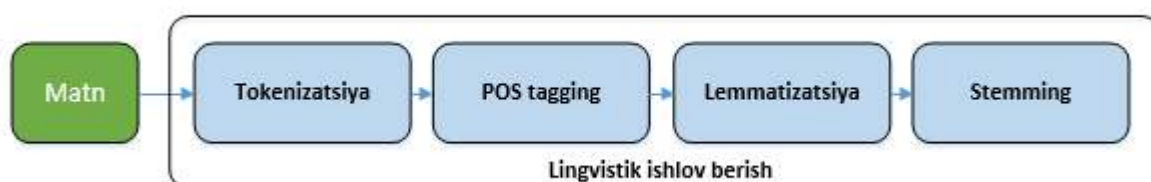
Annotatsiya (annotation) – Korpus lingvistikasi doirasida qaralganda berilgan matnga bevosita aloqasi bo'lmagan, ammo uning qaysidir qismi haqida

lingvistik yoki ekstralingvistik axborot beruvchi umumiy ma'lumot. Annotatsiya o'z ichiga metama'lumot va teglarni qamrab olishi mumkin.

Razmetka (markup) – razmetka va meta ma'lumotni annotatsiyalash jarayonining bir qismi sifatida baholanadi. Jahon amaliyotida razmetkalashning standart prinsiplari ishlab chiqilgan.

Teg – Kompyuter yordamida matn tahlilini amalga oshirish jarayonini tezlashtirish va osonlashtirishga xizmat qiluvchi shartli belgi yoki maxsus kod. Teglar bir necha turlarga bo'linadi: *semantik teg*, *sintaktik teg* va *grammatik teg*. Grammatik teg, shuningdek, PoS (Part of speech) tegging nomi bilan ham mashhur.

POS teglash – bu berilgan gapdagi har bir so'zshaklga uning turkum (*ot*, *fe'l*, *sifat*, *son*, *ravish* yoki *olmosh*)ga mansubligini belgilash (teglash) vazifasidir. POS teglash tabiiy tilni qayta ishlash (Natural Language Processing, NLP)ning asosiy vazifalaridan biri bo'lib, pipeline konveyerining muhim bosqichi hisoblanadi (1-rasm).



1-rasm. Matnga boshlang'ich ishlov berish bosqichlari

Tabiiy tilni qayta ishlash (NLP) vositalari axborot-kommunikatsiya texnologiyalarining jadal rivojlanishi tufayli katta qiziqish uyg'otdi. Natijada bugungi kunda turli xil NLP usullari va vositalari yaratilmoqda. Biroq, tabiiy tillardagi matnlarni qayta ishlaydigan samarali NLP vositalarini ishlab chiqishda lingvistik teglash kabi vazifalarni hal qilish lozim.

Lingvistik teglash tavsiflovchi yoki analitik belgilarni til ma'lumotlari bilan bog'lashni o'z ichiga oladi. Strukturlanmagan (qayta ishlanmagan) ma'lumotlar matnli yoki har qanday manba yoki janrdan olingan yoki vaqt funksiyalari (audio, video va/yoki fiziologik yozuvlar) shaklida bo'lishi mumkin. Teglar barcha turdagi transkripsiyalarni (fonetik xususiyatdan nutqiy qurilmagacha), POS teglash, ma'no belgilari, sintaktik tahlil, NER obyektlar, semantik rol belgilari, vaqt va hodisalarni aniqlash, so'zlarning sintaktik zanjirlarini, nutq darajasini o'z ichiga olishi mumkin.

Til korpuslari ishlab chiqilgandan so'ng ularni teglash lozim. Jumladan, NLP tadqiqotchilari tomonidan **Brown korusi** matnlari POS teglangan [Maverick, 1969: 35(1)]. Ushbu tadqiqot K.V.Gruch [Church, 1989], S.J.Derose [Derose, 1988: 14(1)], R.Garside [Garside, 1987] va B.B.Greenlarning POS teglash bo'yicha ilmiy tadqiqotlarida foydalanilgan [Greene, Rubin, 1971:23]. Brown korpusi singari 70-80-yillarda ishlab chiqilgan korpuslar odatda POS teglangan, ammo samarali

avtomatik usullarning yo'qligi va qo'lda teglashning murakkabligi boshqa til hodisalari uchun teglarni o'z ichiga olgan yetarlicha katta hajmdagi korpuslarni ishlab chiqishga imkon bermagan.

1980-yillarning oxirida yangi katta hajmli til ma'lumotlarining mavjudligi lingvistik teglash tizimlarining ko'payishiga olib keldi. Asosan, POS yoki morfo-sintaktik teglashlashga asoslangan va avtomatik teglash uchun statistik usullar ishlab chiqildi. Ushbu turdagi birinchi yirik sa'y-harakatlar ingliz tilidagi bir million so'zdan iborat **Lancaster-Oslo-Bergen (LOB) korpusi** morfo-sintaktik va sintaktik teglar asosida teglangan [Beale, 1985].

Ushbu ilmiy tadqiqotga asoslanib, Penn Treebank loyihasi asosida Wall Street Journal maqolalarining bir million so'zdan iborat korpus ishlab chiqilgan [Marcus, Santorini, Marcinkiewicz, 1993] va POS hamda sintaktik teglangan [Marcus, Kim, Marcinkiewicz, MacIntyre, Bies, Ferguson, Katz, Schasberger, 1994]. 1990-yillardagi avtomatik teglangan korpuslarga 1994-yilda ishlab chiqilgan 100 million so'zli British National Corpus [Erjaveç, 1998]; MULTEXT ko'p tilli korpusi [Ide, Véronis, 1994: 588-592]; PAROLE va SIMPLE korpuslari [Kolodnytsky, Bernsen, Dybkjaer, 2004] o'n to'rtta Yevropa tilidagi ma'lumotlar va ularning POS teglarini o'z ichiga olgan.

Matn alohida tokenlarga ajratilgandan so'ng har bir token yoki tokenlar to'plamini **so'z turkumi** (ot, fe'l, olmosh va boshqalar) bilan belgilash mumkin. Bu vazifa NLPda **so'z turkumini aniqlash va teglash** yoki **PoS teglash**dir. PoS teglash morfologik tahlil sanaladi, chunki o'zakka qo'shilgan grammatik shakl orqali so'z turkumini aniqlashga yordam beradi. Misol uchun, morfologik teglash jarayoni - *gan*, *-di* qo'shimchalarining fe'llar bilan qo'llaniladigan morfologik birlik ekanligini anglatadi (istisnolar ham mavjud). Biroq morfologik (grammatik) shaklni olmagan so'zlar ham teglanadi (2-jadval). Masalan, *mergan* – **ot**, *bergan* – **fe'l**. Bu so'zlardagi *-gan* alohida birinchi so'zda morfologik birlik emas, balki so'zning bir bo'g'inidir. Bu kabi shakllar o'zbek tilida uchrab turadi. Quyidagi 1-jadvalda shunday birliklarga misol keltiramiz:

1-jadval. O'zbek tilidagi so'zlarni POS teglash namunasi

№	So'z	O'zak	POS teg	Izoh
1.	mergan	mergan	Ot (N)	"mergan" alohida leskema, tarkibiy qismlarga bo'linmaydi
2.	bergan	ber	Fe'l (VB)	"ber"+gan; <i>-gan</i> o'tgan zamon yoki sifatdosh qo'shimchasi; leksema + grammatik shakl
3.	olov	olov	Ot (N)	"olov" alohida leskema, tarkibiy qismlarga bo'linmaydi
4.	ulov	ula	Fe'l (VB)	"ula"+v; <i>-v</i> so'z yasovchi qo'shimcha; leksema + grammatik shakl
5.	ming	ming	Ot (N) Son (Num)	"ming" alohida leskema, tarkibiy qismlarga bo'linmaydi

6.	ahvoling	ahvol	Ot (N)	"ahvol"+ing; -ing egalik qo'shimchasi: leksema + grammatik shakl
7.	boisi	boisi	Ot (N)	"boisi" alohida leskema, tarkibiy qismlarga bo'linmaydi
8.	ukasi	uka	Ot (N)	"uka"+si; -si egalik qo'shimchasi

Mazkur jadvaldagi *mergan*, *olov*, *ming*, *boisi* kabi birliklar alohida leksema sanalib, tarkibiy qismlarga bo'linmaydi, bu so'zlardagi qismlar qo'shimcha emas, o'zakning bir qismi, boshqa holatda o'zakka qo'shilgan grammatik shakl sanaladi. Matn gaplarga bo'lingandan so'ng har bir gapdan *otli birikma* va *fe'lli birikma* kabi tarkibiy so'z birikmalari aniqlanadi. Bu tahlil sintaktik tahlil bo'lib, **lingvistik daraja (sintaksis)** va **annotatsiya vazifasi (sintaktik parsing)** o'rtasida aniq muvofiqlik mavjud. Bugungi kunda lingvistik tahlil bosqichi bo'yicha ko'plab lingvistik nazariyalar mavjud bo'lib, o'zbek tili uchun hozirda keng qo'llaniladigan **so'z birikmalari modeli** yondashuvidan foydalanish maqsadga muvofiq.

POS teglash vazifasida matndagi har bir token yoki tokenlar ketma-ketligi POS yorlig'i bilan belgilanadi. O'zbek tilida 12 ta so'z turkumi mavjud va ularning POS-tegi belgilangan (qarang: 2-jadval).

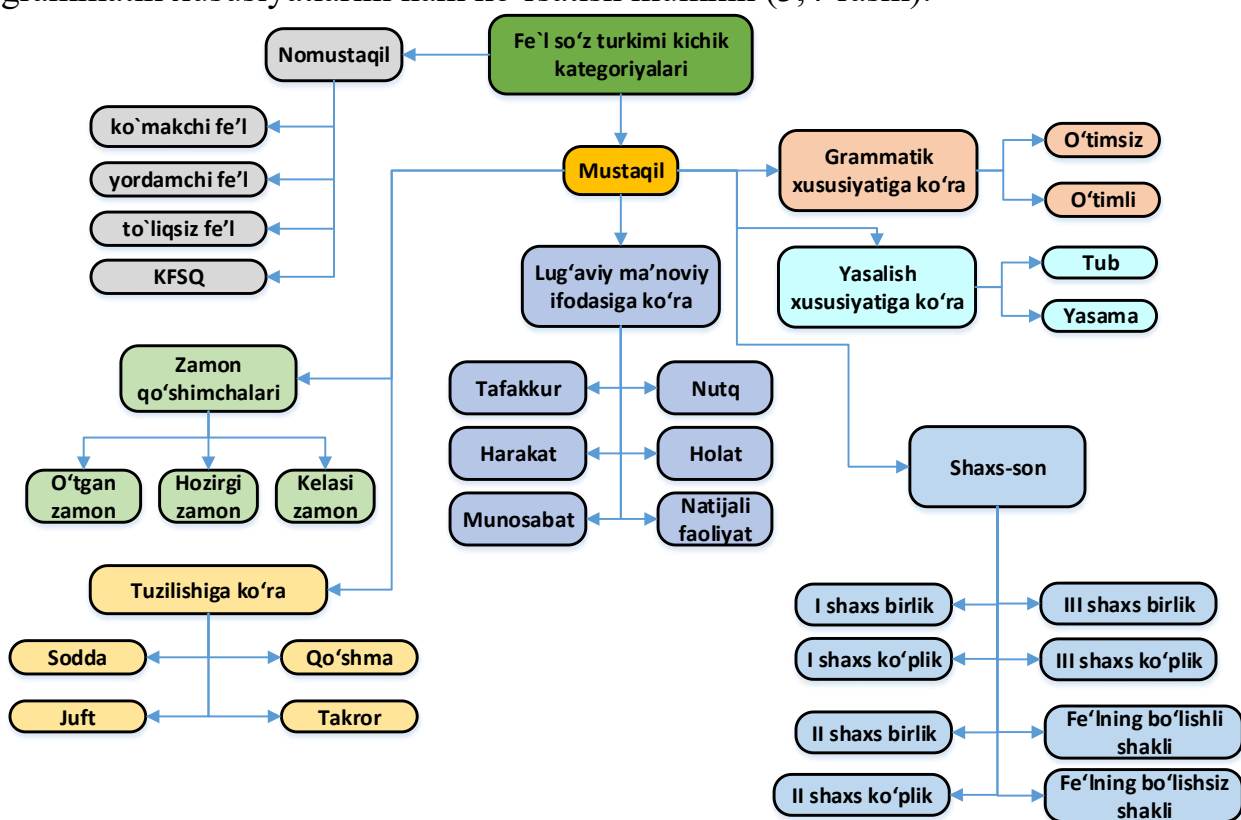
2-jadval. O'zbek tilidagi POS teglar ro'yxati (qisqa variant)

№	POS teg	Ingliz	Belgilanishi
1.	Ot	Noun	N
2.	Sifat	Adjective	JJ
3.	Fe'l	Verb	VB
4.	Son	Number	NUM
5.	Ravish	Adverb	RR
6.	Olmosh	Pronoun	P
7.	Bog'lovchi	Conjunction	C
8.	Ko'makchi	preposition	II
9.	Yuklama	particle	Prt
10.	Modal so'z	modal words	MD
11.	Undov so'z	exclamatory words	UH
12.	Taqlid so'zlar	imitation words	IM

Turli tadqiqotchilar tomonidan boshqa ko'plab muqobil tasniflar taklif qilingan. B.Elov va Sh.Hamroyevalarning “O'zbek tilida POS tegging masalasi: muammo va takliflar” [Elov, Hamroyeva, Abdullayeva, Uzakova, 2022: 51-68] nomli maqolasida o'zbek tili birliklarini POS teglash uchun teg tizimi ishlab chiqilgan keltirilgan, o'zbek tili uchun POS teglarning to'liq (kengaytirilgan varianti) “POS Tagging of uzbek texts using hidden Markov models (HMM) and Viterbi algorithm” [Elov, Hamroyeva, Xudayberganov, Yodgorov, Yuldashev, 2023] nomli maqolada o'zbek tili birliklarini POS teglash uchun teg tizimidan foydalanish

tavsiflangan. Shuningdek, “Agglutinativ tillar uchun pos teglash va stemming masalasi (turk, uyg‘ur, o‘zbek tillari misolida)” [Elov, Hamroyeva, Abdullayeva, Husainova, Xudayberganov, 2023: 6-39] nomli maqolada agglutinativ tillarni teglashda turk, uyg‘ur va o‘zbek tillari teg timizi qiyoslangan. “O‘zbek tilidagi sodda gaplarning sintaktik teglangan bazasi orqali tahlil daraxtini qurish” [Ramatova, 2023] nomli maqolada o‘zbek tili birliklarini POS teglash uchun faqat mustaqil so‘z turkumlari POS teglari taklif qilingan. Quyida o‘zbek tili leksik birliklarini POS teglashda ahamiyatli bo‘lgan grammatik xususiyatlarni izohlaymiz.

Tilshunoslikda bu asosiy kategoriyalar yana kichik kategoriyalarga bo‘linadi. Masalan, fe‘l so‘z turkumini teglashda, uning fe‘l ekanligidan tashqari boshqa grammatik xususiyatlarini ham ko‘rsatish mumkin (3,4-rasm).

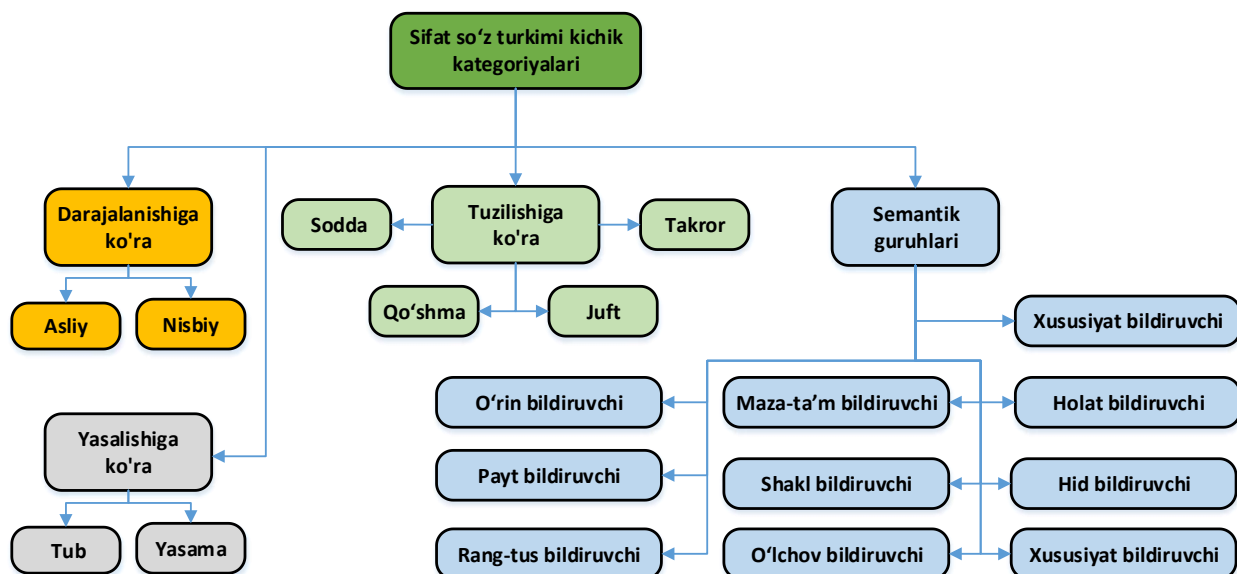


1-rasm. Fe‘l so‘z turkimi grammatik kategoriyalari tasnifi

Fe‘l so‘z turkumini morfoligik (POS) teglash quyidagicha amalga oshiriladi:

Og‘rib qoldi: [fe‘l], [mustaqil fe‘l], [holat fe‘li], [o‘timsiz fe‘l], [bo‘lishli fe‘l], [aniq nisbat], [sodda fe‘l], [tub fe‘l], [o‘tgan zamon], [III shaxs birlik], [har. t.], [yetakchi fe‘l], [ko‘makchi fe‘l].

Bu yerda grammatik kategoriyalar to‘liq keltirildi, ammo teglash jarayonida kengaytirilgan teglar tizimidan foydalaniladi. Kengaytirilgan teglar tizimi “O‘zbek tilida POS tegging masalasi: muammo va takliflar” nomli maqolada batafsil keltirilgan [Elov, Hamroyeva, Abdullayeva, Uzakova, 2022: 51-68].



2-rasm. Sifat so'z turkimi grammatik kategoriyalari tasnifi

Sifatni teglash umumiy grammatik ma'noni belgilashdan boshlanadi. Sifat belgining xususiyatini ifodalashiga ko'ra ikkiga bo'lingan: 1) asliy sifat; 2) nisbiy sifat. Sifat turkumida sintaktik kategoriya keng bo'lmagan voqelanishga ega. Sifat turkumi egalik kategoriyasi UGMsini "keyingi sifatni oldingi so'zga bog'lash, mansublik, xoslik ma'nosini ifodalash" tarzida xususiylashtiradi. Kesimlik kategoriyasi esa sifat turkumida o'z mohiyatini cheklangan darajada namoyon qiladi. Ko'p hollarda sifat turkumida kesimlik kategoriyasi bog'lama vositasida yuzaga chiqadi.

Kengaytirilgan teglar tizimi asosida sifat so'z turkumining teglanishi quyidagicha bo'ladi:

- **o'lik:** [sifat], [nisbiy sifat], [oddiy daraja], [yasama sifat], [sodda sifat], [jismoniy harakat LMG].
- **tuzuk:** [sifat], [asliy sifat], [oddiy daraja], [tub sifat], [yasama sifat], [xususiyat LMG].
- **yiroq:** [sifat], [asliy sifat], [oddiy daraja], [sodda sifat], [tub sifat], [xususiyat LMG].
- **yo'g'onroq:** [sifat], [asliy sifat], [qiyosiy daraja], [sodda sifat], [tub sifat], [shakl LMG].

Xulosa

Til korpusi matnlari ustida o'tkazilgan turli tajribalar shuni ko'rsatadiki, o'zak ma'lumotlarini sintaktik vazifa bilan birlashtirish morfologik jihatdan boy til uchun POS teglash natijasini yaxshilaydi, bu esa NLP vazifasining hal qilish samaradorligini oshirishga xizmat qiladi. O'zbek tili korpusi matnlarini POS teglash vazifasining o'ziga xos jihatlari, POS teglar kategoriyalari va kichik kategoriyalar ro'yxati keltirildi. Jumladan fe'l va sifat so'z turkumlari grammatik kategoriyalari tasnifi va namunalar izohlandi. Maqola keltirilgan POS teglash usullaridan o'zbek

tili morfologik analizatorni ishlab chiqishda va boshqa murakkab NLP vazifalarni hal qilishda foydalanish mumkin.

Foydalanilgan adabiyotlar:

1. Maverick, G. v. (1969). Computational Analysis of Present-Day American English. Henry Kučera, W. Nelson Francis. *International Journal of American Linguistics*, 35(1). <https://doi.org/10.1086/465045>
2. Church, K. W. (1989). Stochastic parts program and noun phrase parser for unrestricted text. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2. <https://doi.org/10.3115/974235.974260>
3. Derose, S. J. (1988). GRAMMATICAL CATEGORY DISAMBIGUATION BY STATISTICAL OPTIMIZATION. *Computational Linguistics*, 14(1).
4. Garside, R. (1987). The CLAWS word-tagging system. *The Computational Analysis of English: A Corpus-Based Approach*.
5. Greene, B.B., Rubin, G.M.: Automatic Grammatical Tagging of English. Brown University, Department of Linguistics (1971)
6. Beale, A. D. (1985). Grammatical analysis by computer of the Lancaster-oslo/bergen (lob) corpus of british English texts. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 1985-July. <https://doi.org/10.3115/981210.981246>
7. Marcus, M. P., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2). <https://doi.org/10.1162/coli.2010.36.1.36100>
8. Marcus, M., Kim, G., Marcinkiewicz, M.A., MacIntyre, R., Bies, A., Ferguson, M., Katz, K., Schasberger, B.: The penn treebank: annotating predicate argument structure. In: *Proceedings of the Workshop on Human Language Technology*, pp. 114–119. Association for Computational Linguistics, Stroudsburg, PA, USA (1994)
9. Erjavec, T., Ide, N.: The MULTEXT-East corpus. In: *Proceedings of First International Conference on Language Resources and Evaluation*, pp. 971–974 (1998)
10. Ide, N., Véronis, J. MULTEXT: multilingual text tools and corpora. In: *Proceedings of the 15th International Conference on Computational Linguistics (COLING 94)*, vol. I, pp. 588–592. Kyoto, Japan (1994)
11. Kolodnytsky, M., Bernsen, N. O., & Dybkjær, L. (2004). A visual interface for a multimodal interactivity annotation tool: Design issues and implementation solutions. *Proceedings of the Workshop on Advanced Visual Interfaces AVI*. <https://doi.org/10.1145/989863.989937>
12. Elov B., Hamroyeva Sh., Abdullayeva O., Uzakova M. O‘zbek tilida POS tagging masalasi: muammo va takliflar / Uzbekistan: Language and Culture. *Applied philology*. 2022/2(5). – 51-68-b.



13. Elov B., Hamroyeva Sh., Xudayberganov N., Yodgorov U., Yuldashev A. Pos tagging of Uzbek texts using hidden Markov models (HMM) and Viterbi algorithm. O'zMU xabarlari. Mirzo Ulug'bek nomidagi O'zbekiston Milliy universiteti ilmiy jurnali. 2023 yil Maxsys son.
14. Elov, B.B., Hamroyeva, Sh.M., Abdullayeva, O.X., Husainova, Z.Y., Xudayberganov, N.U. 2023. "Agglutinatив tillar uchun pos teglash va stemming masalasi (turk, uyg'ur, o'zbek tillari misolida)". O'zbekiston: til va madaniyat 2: 6-39.
15. Ramatova M. O'zbek tilidagi sodda gaplarning sintaktik teglangan bazasi orqali tahlil daraxtini qurish // Educational Research in Universal Sciences. VOLUME 2 | ISSUE 12 | 2023. // <https://zenodo.org/records/10463799>.

TABIIY TILNI QAYTA ISHLASH (NLP)

O‘ZBEK TILI MATNLARINI MORFOLOGIK TAHLIL QILISH UCHUN LINGVISTIK MANBA YARATISH VA LEMMALASH

Sharipov Maqsud Siddiqovich

Texnika fanlari nomzodi, dotsent,

m.sharipov@urdu.uz,

UrDU, Kompyuter ilmlari kafedrası dotsenti

Samandarova Dilfuza Shonazar qizi

dilfuzasamandarova44@gmail.com,

UrDU, Kompyuter ilmlari kafedrası magistranti

Sobirov Og‘abek Ollayor o‘g‘li

sobirov.o@urdu.uz,

UrDU, Kompyuter ilmlari kafedrası tayanch doktoranti

Annotatsiya. Til resurslari Tabiiy Til Jarayonlari (TTJ) vazifalarida hal qiluvchi rol o‘ynaydi va tadqiqotchilarga katta hajmdagi matnlarni samarali tahlil qilish va qayta ishlash imkonini beradi. Ushbu maqolada biz korpus tahlili va lemmalash uchun moslashtirilgan lingvistik manba yaratishga kompleks yondashuvni taqdim etganmiz. Biz korpus lingvistikasida lingvistik manbalarning ahamiyatini muhokama qilganmiz va korpus yaratish jarayonini tasvirlaganmiz. Bundan tashqari, biz lemmalash tushunchasini va uning lingvistik tahlildagi ahamiyatini ko‘rib chiqqanmiz. Maqolada korpus tahlili uchun lingvistik manbalarni yaratish bosqichlari ko‘rsatilgan va lemmalash usullari haqida tushuncha berilgan.

Annotation. Language resources play a crucial role in Natural Language Processing (NLP) tasks, enabling researchers to analyze and process large volumes of text efficiently. In this paper, we present a comprehensive approach to linguistic resource creation tailored for corpus analysis and lemmatization. We discuss the importance of linguistic resources in corpus linguistics and describe the process of corpus creation. Additionally, we delve into the concept of lemmatization and its significance in linguistic analysis. The paper outlines the steps involved in creating linguistic resources for corpus analysis and provides insights into lemmatization techniques.

Аннотация. Языковые ресурсы играют решающую роль в задачах обработки естественного языка (NLP), позволяя исследователям эффективно анализировать и обрабатывать большие объемы текста. В этой статье мы представляем комплексный подход к созданию лингвистических ресурсов, адаптированный для корпусного анализа и лемматизации. Мы обсуждаем важность лингвистических ресурсов в корпусной лингвистике и описываем процесс создания корпуса. Кроме того, мы углубляемся в понятие лемматизации и ее значение в лингвистическом анализе. В статье

описываются шаги по созданию лингвистических ресурсов для корпусного анализа, а также даются сведения о методах лемматизации.

Kalit so‘zlar. *Tabiiy Til Jarayonlari (TTJ), korpus, korpus tahlili, lingvistik manba, lemma, lemmalash, O‘zbek tili.*

Kirish

Har qanday tildagi matnlarni kompyuterda tahlil qilish, umuman lingvistik dasturlar yaratish uchun, ma’lum lingvistik manba bo‘lishi kerak. Xususan, ushbu maqolada o‘zbek tilidagi matnlarni avtomatik tahlil qilish uchun lingvistik manba yaratish va uni aynan Tabiiy Til Jarayonlari (TTJ) [Natural Language Processing (NLP)] muammolarini hal qilishga qo‘llash va lemmalash masalasi haqida alohida so‘z yuritilgan. Bunda lingvistik manba ma’lumotlar bazasi bo‘lishi mumkin, ushbu ma’lumotlar bazasi esa korpus [corpus] deb ataladi va u yozma nutqning har qanday to‘plami bo‘lishi mumkin. Bunda lingvistik tahlil dasturlari ushbu bazalardan tezkor qidirish va tahlil qilish orqali foydalanadi.

Adabiyotlar tahlili. O‘zbek kompyuter lingvistikasida [computational linguistics] lingvistik manbalardan foydalangan holda amalga oshirilgan ishlar bilan tanishib chiqar ekanmiz, quyidagi ilmiy ishlar va ularning natijalarini aytib o‘tish joiz. So‘zlar [Words] va Qo‘shimchalar [Affixes] bazalari shakllantirilgan holda qoidaga asoslangan lemmalash algoritmi ishlab chiqilgan, algoritmini qurishda Chekli Holatlar Mashinasidan (ChHM) [Finite State Machine (FSM)] foydalanilgan [Sharipov, Sobirov, 2022]. So‘z asosini topish uchun qo‘shimchalarning XML fayli yaratilib, ular o‘n beshta sinfga ajratilib, har bir sinf uchun alohida ChHM qurilgan va ular birlashtirilib asosiy yagona ChHM yaratilib u asosida stemming jarayoni amalga oshirilgan [Sharipov, Yuldashov, 2022]. Matnlarda Ot va Olmosh so‘z turkumlarini aniqlash uchun ularning tup shakldagi bazasi va ularning qo‘shimchalar bazasi shakllantirilga [Sharipov, Mavlonova, Sobirov, 2023]. O‘zbek tili fe’l so‘z turkumi uchun stemming algoritmini ishlab chiqish ChHM yordamida amalga oshirilgan [Sharipov, Salayev, Matlatipov, 2021]. Stemmingdan foydalangan holda matnlarni belgilash jarayoni amalga oshirilgan, turli sohalarga tegishli bo‘lgan ~23 000 atrofida so‘zlar testlab ko‘rilgan va 89.78% aniqlikka erishilgan [Sharipov, Kuriyozov, Yuldashov, Sobirov, 2023]. Lug‘atda mavjud va lug‘atda mavjud bo‘lmagan o‘zbek, turk va uyg‘ur tillari uchun so‘z turkumlarini belgilash va stemming bo‘yicha nazariy va amaliy ishlar amalga oshirilib, ularning tahliliy natijalari turli xil jadvallar va grafiklar bilan chop qilingan [Elov, Adali, Hamroyeva, Abdullayeva, Xusainova, 2023]. Kompyuterda TTJ muammolarini hal qilish uchun lingvistik manba tarkibida mavjud bo‘lishi kerak bo‘lgan lingvistik resurslar haqida so‘z boradi [Abjalova, Rashidova, Abdumuminov, 2022]. TF-IDF yordamida matnlarni xulosalash, bunda semantik tahlil muhim bo‘lib, n-gram metodi yordamidan foydalanilgan va yig‘ilgan “Maktab korpusi”dan foydalanilgan [Madatov, Bekchanov, Vichich, 2023].

Lingvistik manba yaratish. Morfologik tahlilning asosiy maqsadi tilni tushunish va uni qo'llashga yordam berishdir. O'zbek tilidagi morfologik tahlil, so'zlar va ularning qo'shimcha shakllarini aniqlash, funksiyasini belgilash va tarkibiy qurilishini tushuntirishga oid bo'ladi. Morfologik tahlil, tilni tushunish, o'rganish va foydalanishda muhim bir qadamdir. Bu tilning qoidalarini va tuzilmalarini tushunishda va unga oid matnlar va yozuvlardan foydalanishda yordam beradi. Lingvistik manba, tilning tadqiqotlarini, ma'lumotlarini va ilmiy o'rganishlarini taqdim etuvchi muhim manba yoki manbalardir. Bu manbalar tilshunoslikda, yozuvchilarga, tarjimanlarga va boshqa lingvistik soha bilan shug'ullanuvchi o'quvchilar, olimlar va tadqiqotchilar uchun amaliy yordamchi bo'ladi. Lingvistik manba turli xil shaklda bo'lishi mumkin: tarjima lug'atlar, izohli lug'atlar, imlo lug'atlari va hattoki qo'llanmalar ham. O'zbek tilini morfologik tahlil qilish uchun lingvistik manba sifatida biz so'zlarning tegishli belgi bilan belgilangan bazasini yaratdik. Bunda so'zlarni darsliklar, maqolalar, adabiyotlar va boshqa manbalardan oldik [Sapayev, 2009; Mirzayev, Begmatov, Madvaliyev, 2008]. Ushbu manba TTJ vazifalarida hal qiluvchi rol o'ynaydi va tadqiqotchilarga katta hajmdagi matnlarni samarali tahlil qilish va qayta ishlash imkonini beradi. Bu esa barcha so'z turkumlari doirasida, xususan ot, fe'l, sifat, son, ravish, olmosh, oraliq va yordamchi so'z turkumlaridagi so'zlarni qamrab oladi. Manba quyidagicha tuzilishga ega:

- Kitob/xon (dastlab: ot, keyin: ot)
- Qora/ymoq (dastlab: sifat, keyin: fe'l)
- Iz/chil (dastlab: ot, keyin sifat)
- O'q/dek (dastlab: ot, keyin: ravish)
- Bog'/bon (dastlab: ot, keyin: ot)
- Ikki/lanmoq (dastlab: son, keyin: fe'l)

Lemmalash. Matnlar ustida lingvistik amallar bajarishda dastlabki muhim ishlov berish usullaridan biri lemmalashdir. Ya'ni TTJ muammolarini hal qilish uchun dastlab matnlar ustida dastlabki ishlov berish usullari belgilash (teglash), tokenlash, lemmalash va stemming kabi ishlarni amalga oshirish kerak bo'ladi. Lemma tilshunoslik sohasida foydalaniladigan muhim konseptdir. Lemma - so'zning o'zgaruvchan qismi, odatda asosiy yoki standart shakli deb ataladi va lemmalar to'g'ri tarjima qilinadi. Lemmalash – so'zni o'z asos shakliga qaytarish amaliyotidir. Tilning morfologik tahlilini amalga oshirib, so'zni undagi asosiy shakliga aylantirishdir. Lemmalash natijasida olingan asosiy shakl lemma deb ataladi. Masalan, “*kitoblar*”, “*kitobni*”, “*kitoblarga*” kabi so'zlar “*kitob*” lemmasiga (asosiy shakliga) aylantirish jarayonidir.

Natija va tahlil. Ushbu lingvistik manba asosida adabiyot sohasiga tegishli matn tahlil qilib ko'rildi. Natija quidagi jadvalda berilgan [1-jadval].

№	Soʻz	Matnda uchrashlar soni	Matnda uchrash foizi	Lemmasi
1	daraxt	14	1.4%	daraxt
...				
8	uning	7	0.7%	u
...				
999	qoʻydi	1	0.1%	qoʻymoq

1-jadval. Lingvistik manba asosida adabiyot sohasiga tegishli matn tahlili
Matn quyidagicha boshlangan “*Hovlimizdan oqib oʻtadigan ariq labidagi bu daraxt qishlogʻimizda yagona hisoblanardi. ...*”. Ushbu matnni biz yaratgan lingvistik manba yordamida tahlil qilish uchun dastur ishlab chiqib, aynan soʻzlar lemmasini topish uchun qoʻlladik. Lingvistik manbadan foydalangan holda TTJlarining boshqa muammolarini yechishda ham foydalanish mumkin.

Xulosa

Xulosa oʻrnida shuni aytish joizki, lingvistik muammolarni hal qilishda, TTJ vazifalarini bajarishda, tarjima dasturlarini ishlab chiqishda lingvistik manbaning oʻrni katta. Ushbu maqolada aynan lingvistik manba yaratishning nazariy va amaliy maʼlumotlari berib oʻtildi va ushbu maʼlumotlar asosida lingvistik manba yaratilib, u asosida matn tahlil qilib koʻrildi, ushbu tahlil natijalari jadval asosida eʼlon qilingan. Kelajakda ushbu manbani mukammallashtirib, uni turli xil janrdagi matnlarni atroflicha tahlil qilish uchun asosiy manbaga aylantirmoqchimiz.

Foydalanilgan adabiyotlar:

1. Sharipov M, Sobirov O. Development of a Rule-Based Lemmatization Algorithm Through Finite State Machine for Uzbek Language. CEUR Workshop Proceedings, CEUR-WS 2022, 154 – 159.
2. Sharipov M, Yuldashov O. UzbekStemmer: Development of a Rule-Based Stemming Algorithm for Uzbek Language. CEUR Workshop Proceedings, CEUR-WS 2022, 137 – 144.
3. Sharipov M, Mavlonova M, Sobirov O. Oʻzbek tilidagi matnlarda ot va olmosh soʻz turkumlariga tegishli soʻzlarni aniqlovchi algoritmi yaratish. Computer Linguistics: Problems, Solutions, Prospects 2023; 1.
4. Sharipov M, Salaev U, Matlatipov G. Implemented stemming algorithms based on finite state machine for uzbek verbs| computer linguistics: problems. solutions, prospects 2021;
5. Sharipov M, Kuriyozov E, Yuldashev O, Sobirov O. UzbekTagger: The rule-based POS tagger for Uzbek language. arXiv preprint arXiv:230112711 2023;
6. Boltayevich EB, Adali E, Mirdjonovna KS, Xolmoʻminovna AO, Yuldashevna XZ, Nizomaddin Uktamboy OʻGʻli X. The Problem of Pos Tagging



and Stemming for Agglutinative Languages (Turkish, Uyghur, Uzbek Languages). UBMK 2023 - Proceedings: 8th International Conference on Computer Science and Engineering, 2023, 57 – 62.

7. Abjalova M, Rashidova U, Abdumuminov B. Matnlarni qayta ishlash manbalari. computer linguistics: problems, solutions, prospects 2022; 1.

8. Madatov K, Bekchanov S, Vičič J. Uzbek text summarization based on TF-IDF. arXiv preprint arXiv:230300461 2023;

9. Sapayev Q. Hozirgi O'zbek tili (morfemika, so'z yasalishi va morfologiya). Toshkent: Nizomiy nomidagi TDPU, 2009.

10. Mirzayev T, Begmatov E, Madvaliyev A *et al.* O'zbek tilining izohli lug'ati. Toshkent: 'O'zbekiston milliy ensiklopediyasi' Davlat ilmiy nashriyoti, 2008.

MATNDAGI TINISH BELGILARINI TIKLASH MUAMMOLARI VA YECHIMLARI

Ochilov Mannon Musinovich

ochilov.mannon@mail.ru

Texnika fanlari bo'yicha falsafa doktori PhD

TATU dotsenti

Jurayev Dilshod Boymuradovich

dilsamtuit@gmail.com

TATU tayanch doktoranti

Narzullayev Oybek Otabek o'g'li

oybeknarzullaev99@gmail.com

TATU magistranti

Annotatsiya. Ushbu maqolada matndagi tinish belgilarini tiklash muammolari va yechimlari haqida so'z yuritiladi. Maqolada shuningdek, matnlarda tinish belgilarini tiklashning turli usullari va yondashuvlarining afzallik va kamchiliklari ko'rib o'tiladi.

Abstract. This article discusses problems and solutions for restoring punctuation marks in text. The article also reviews the advantages and disadvantages of various methods and approaches to punctuation in texts.

Аннотация. В этой статье рассматриваются проблемы и решения по восстановлению знаков препинания в тексте. Также в статье рассматриваются преимущества и недостатки различных методов и подходов к пунктуации в текстах.

Kalit so'zlar: *Tinish belgilar, statistik usullar, mashinali o'qitish algoritmlari, neyron tarmoqlar, RNN, BiLSTM, Transformer modellari.*

Bugunga kelib o'zbek tili bo'yicha qator ilmiy ishlar olib borilgan bo'lib, ularning aksariyati Tabiiy tilni qayta ishlash(NLP)ga qaratilgan bo'lsa-da, ularning orasida o'zbek tili matnlaridagi mavjud imlo va grammatik xatoliklarni tuzatishga qaratilgani juda kam. Biz ushbu maqolada matndagi turli xildagi xatoliklarni tuzatish bo'yicha yuzaga keladigan muammolar, mavjud usullar tahlili va yechimlari to'g'risida to'xtalib o'tamiz. Matndagi xatoliklarni tuzatish muammolari orasida tinish belgilarni qo'yish muhim ahamiyatga ega. Tinish belgilar bo'yicha ikki turdagi masala ko'riladi. Birinchisi, matnda mavjud tinish belgilarni to'g'ri qo'yilganligini tekshirish. Bu turdagi masalalar juda ko'p uchraydi. Ikkinchisi esa, matnda tinish belgilari umuman mavjud emas vaziyatlarda ularni to'g'ri qo'yish. Keltirilgan ikkinchi masala odatda nutqni tanish tizimlari tomonidan hosil qilingan matnlarda uchraydi. Matndagi tinish belgilarini tiklashning qator muammo va vazifalarini keltirishimiz mumkin. Bu vazifalar murakkab bo'lib, tinish belgilarining to'g'ri joylashishini aniqlash uchun tilning konteksti, semantikasi va grammatik

qoidalarini tushunishi kerak. Bu esa bir qator muammolarni keltirib chiqaradi. Tinish belgilarini tiklash quyidagi vazifalarni o'z ichiga oladi:

1. Nuqta va vergul qo'yish: jumlar yoki alohida bandlarni ko'rsatish uchun nuqta va vergullarni to'g'ri joylashtirishni ta'minlash.
2. So'roq va undov belgilari qo'yish: mo'ljallangan ohangni bildirish uchun savollarga so'roq va buyruq va his-hayajonlar uchun undov belgilarini qo'shish.
3. Qo'shtirnoq belgilari: so'zlovchining to'g'ridan-to'g'ri gapiga nisbatan yoki ba'zi terminlarni matnda ajratish uchun qo'shtirnoqlarni joylash.
4. Ikki nuqta va nuqta-verguldan foydalanish: tegishli jumla tuzilmalari uchun ikki nuqta va nuqta-vergul o'rtasidagi farqni aniqlashtirish.
5. Qavslar: qo'shimcha ma'lumotlarni qamrab olish yoki tarkibni aniqlashtirish uchun qavslardan foydalanishni tuzatish.
6. Chiziqchalarni tuzatish: urg'u yoki uzilish hamda qo'shma so'zlar uchun chiziqchalardan to'g'ri foydalanish.
7. Tutuq belgisini to'g'ri joylashtirish: ba'zi so'zlarga tutuq belgilarini qo'yib ularni to'g'irlash.

Ushbu vazifalarni bajarishda o'z navbatida quyidagi muammolar kelib chiqadi:

➤ **Matn mazmunidagi noaniqlik:** Ba'zi jumlar to'g'ri tinish belgilarini aniqlash uchun kontekstni talab qilishi mumkin. Ayrim iboralar tushunarsiz bo'lishi mumkin, bu esa to'g'ri tinish belgilarini qo'yishni qiyinlashtiradi.

➤ **Matn tarkibining qisman yo'qolganligi:** Tinish belgilarini tiklashga yetishmayotgan ma'lumotlar yoki to'liq bo'lmagan jumlar to'sqinlik qilishi mumkin. Yetarli kontekstsiz, mo'ljallangan tinish belgilarini aniqlash qiyin kechadi.

➤ **Turli xil yozish uslublari:** Turli yozuvchilar tinish belgilaridan boshqacha foydalanishi mumkin va bitta yozish uslubiga o'rgatilgan model boshqa yozish uslubi bilan yozilgan matnlarga moslashtirishni talab etadi. Bu esa o'z navbatida tinish belgilarini qo'yishda katta qiyinchilik tug'diradi.

➤ **Murakkab jumlar:** Bir nechta gap yoki qavs ichidagi iboralar bilan murakkab jumlar tinish belgilarini to'g'ri belgilash qiyin bo'lishi mumkin.

➤ **Domenga xos muammolar:** Ixtisoslashgan domenlarda (masalan, yuridik, ilmiy) tinish belgilarini tiklash noyob tinish belgilari bilan ishlash uchun domenga xos bilimlarni talab qilishi mumkin.

➤ **Ko'p tillilik muammosi:** Tinish belgilari turli tillarda farqlanadi va bir tilda o'rgatilgan model boshqa tilda turli tinish belgilari bilan qo'llanilganda unchalik yaxshi ishlamasligi mumkin.

Matnlarda tinish belgilarini tiklashning turlicha usullari mavjud bo'lib ularni 4 ta umumiy guruhga ajratib olishimiz mumkin:

1. Qoidalariga asoslangan tinish belgilarini tiklash algoritmlari;
2. Statistik modellar asoslangan yondashuv;
3. Mashinali o'qitishga asoslangan yondashuv;
4. Chuqur o'qitish neyron tarmog'iga asoslangan yondashuv.

Quyida ushbu usullarning har biriga batafsil to'xtalamiz.

Qoidalarga asoslangan tinish belgilarini tiklash algoritmlari. Ushbu usul haqida [Päis, Tufiş, 2022] ishda batafsil ma'lumot va turdosh ilmiy izlanishlarni keltirib o'tishgan. Bu usulning boshqa usullardan farqi shundagi undagi algoritmlar an'anaviy dasturlash tamoyillariga muvofiq amalga oshiriladi. Bunda algoritmni tuzib chiqish to'laqonli dasturchining ishi hisoblanadi. Keling, shunday algoritmlardan bittasini sodda misol bilan tuzib ko'raylik. Aytaylik bizga sodda, faqat birinchi shaxsda yozilgan gaplardan iborat, birorta tinish belgisi yo'q, hamda faqat kichik harflardan iborat matn berilsin:

“men talabaman bugun uyga boramanmi men topshiriqlarning hammasini bajardim”

Matn mazmunini bir qarashda sezish mumkin. Unda tinish belgilarini tiklashni amalga oshirish uchun sodda algoritm yozish yetarli. Ushbu algoritmida bor yo'g'i ikkita qoida bor:

1. Agar so'z oxiri “man” bilan tugasa shu so'zning oxiriga “nuqta” qo'y va shu so'zdan keyingi so'zni bosh harf bilan boshla;
2. Agar so'z oxiri “mi” bilan tugasa shu so'zning oxiriga “so'roq belgisi” qo'y va shu so'zdan keyingi so'zni bosh harf bilan boshla;

Algoritmga yana boshqa qoidalarni kiritish mumkin. Tuzilgan algoritm faqat yuqoridagi misolgagina mutlaq to'g'ri ishlaydi. Ammo boshqa tuzilmadagi matnlari uchun bu algoritm to'liq ishlamaydi. Ushbu algoritmlar bugungi kunda tinish belgilarini tiklashda keng qo'llanilmaydi. Chunki ushbu usulda bir qator kamchiliklar bor:

1. Cheklangan moslashuvchanlik: Qoidalarga asoslangan algoritmlar oldindan belgilangan qoidalarga tayanadi. Bu ularni turli xil yozish uslublari yoki kutilmagan o'zgarishlarga kamroq moslashtiradi. Ular o'rnatilgan qoidalardan chetga chiqadigan nostandart til yoki ijodiy yozuv bilan ishlashda qiyinchiliklar keltirib chiqarishi mumkin.
2. Noaniqlik bilan bog'liq qiyinchilik: Qoidalar tilning barcha xususiyatlarini qamrab olmasligi mumkin, bu esa noto'g'ri tinish belgilarini tiklashga olib keladi.
3. Qoidalarning murakkabligi: Tinish belgilarini tiklash uchun keng qamrovli qoidalarni ishlab chiqish murakkab va ko'p mehnat talab qilishi mumkin. Tabiiy tilning barcha nozik tomonlarini, ayniqsa, yozish uslublari va kontekstlarining xilma-xilligini hisobga olgan holda, qoidalar to'plamida butun tilni qamrab olish qiyin bo'lishi mumkin.
4. Tilga bog'liqlik: Qoidalarga asoslangan algoritmlar ko'pincha tilga bog'liq, ya'ni turli tillar uchun alohida qoidalar to'plami ishlab chiqilishi kerak. Bu resurs talab qilishi va ko'p tilli ssenariylarni boshqarish uchun yaxshi miqyosda bo'lmasligi mumkin.
5. Kengayuvchanlik muammolari: Tilning murakkabligi oshgani sayin, barcha mumkin bo'lgan holatlarni qamrab olish uchun talab qilinadigan qoidalar

soni ham oshadi. Bu masshtablilik muammolariga olib kelishi mumkin, chunki ma'lumotlar to'plamining hajmi yoki tilning murakkabligi kengayganligi sababli qoidalarga asoslangan tizimni saqlashni qiyinlashtiradi.

Qoidalarga asoslangan tinish belgilarini tiklash algoritmalaridan farqli o'laroq bugungi kunda mashinali o'qitish modellari hamda neyron tarmoq modellari o'zlarining kengayuvchanligi hamda samarali ekanligi bilan ancha ilg'or hisoblanadi.

Statistik modellarga asoslangan yondashuv. Statistik modellar - bu ma'lumotlar ichidagi munosabatlarni tavsiflash va ifodalash uchun statistik usullardan foydalanadigan matematik modellardir. Statistik modellar yordamida tinish belgilarini tiklash qoidalarga asoslangan algoritmalar nisbatan samaraliroq hisoblanadi. Statistik modellar ma'lumotlardan o'rganishi va turli lingvistik kontekstlarga moslashishi, so'zlar va tinish belgilari o'rtasidagi bog'liqlikni ko'rsatib bera olishi mumkin. Ular yozish uslublari va noan'anaviy tinish belgilaridan foydalanishdagi o'zgarishlarni yaxshiroq talqin etadi. Chunki ular qoidalarga tayanmasdan, ma'lum bir oldindan mavjud bo'lgan tinish belgisi xatosiz ma'lumotlardan o'rganadilar. Statistik modellar turli sohalarda, jumladan, fan, muhandislik, iqtisod va ijtimoiy fanlarda keng qo'llaniladi.

Ko'pchilik ishlarda tinish belgilarini tiklashda Yashirin Markov modellari (HMMs – Hidden Markov models) va Shartli tasodifiy maydonlar (CRFs – Conditional Random Fields) kabi statistik usullardan keng foydalanilgan [Boros va b., 2017].

Ammo statistik usullar ham qoidalarga asoslangan algoritmalaridan samarali bo'lishiga qaramay bir qator kamchiliklarga ega.

1. Matn kontekstni tushunishda cheklanganlik: Yashirin Markov modellari va shartli tasodifiy maydonlarni o'z ichiga olgan statistik modellar uzoq muddatli bog'liqliklarni olish va murakkab kontekstual jumalarni tushunishda cheklovlariga ega bo'lishi mumkin. Tinish belgilarini tiklash ko'pincha kontekstni chuqur tushunishni talab qiladi, ayniqsa murakkab jumla tuzilmalarida.

2. O'zgaruvchanlikni boshqarish qiyinligi: Statistik modellar yozish uslublariidagi o'zgarishlar, norasmiy til yoki noan'anaviy tinish belgilaridan foydalanishda qiyinchiliklarga duch kelishi mumkin. Ular ko'pincha oldindan belgilangan uslublarga tayanadilar va kamroq tuzilgan yoki xilma-xil ma'lumotlarga duch kelganda kutilgan natijani bermasligi mumkin.

3. Katta o'lchamli ma'lumotlar bilan ishlashda qiyinchilik: Statistik modellarni o'rgatish, ayniqsa ko'p sonli parametrlarga ega bo'lgan modellarni o'rgatishni amalga oshirish uchun katta resurs kerak bo'ladi. Katta o'lchamli ma'lumotlar to'plamlari bilan ishlash xotira va qayta ishlash talablari nuqtayi-nazaridan qiyinchiliklarga olib kelishi mumkin.

4. Semantik kontekstga kamroq moslashish: Tinish belgilarini tiklash ko'pincha matnning semantik kontekstini hisobga olishni o'z ichiga oladi. Statistik

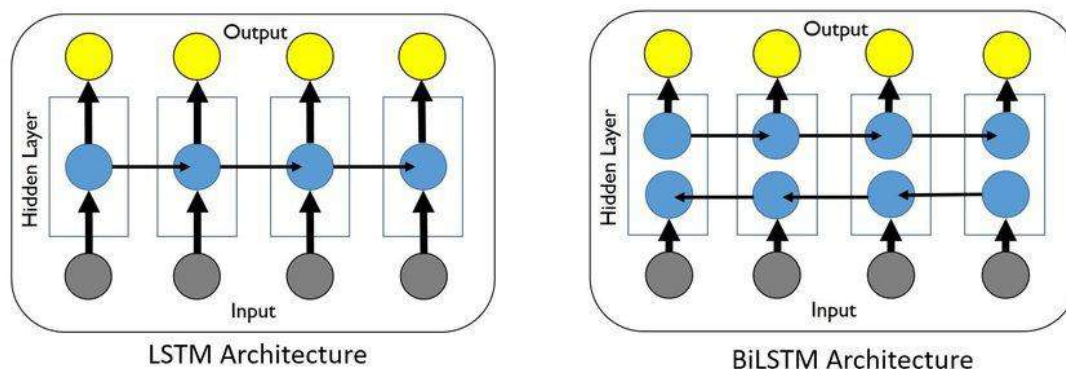
modellar ilg'or mashinali o'qitish modellari yoki neyron tarmoqlar bilan solishtirganda semantik ma'lumotni olishda unchalik yaxshi bo'lmashligi mumkin.

5. Chiziqli bo'lmagan munosabatlarni olishda qiyinchilik: An'anaviy statistik modellar o'zgaruvchilar orasidagi chiziqli munosabatlarni nazarda tutadi. Biroq, tinish belgilarini tiklash kabi til vazifalarida munosabatlar chiziqli bo'lmashligi mumkin va statistik modellar bu murakkab ssenariylarda samarali ishlamasligi mumkin.

Mashinali o'qitishga asoslangan yondashuv. Mashinali o'qitish modellari tabiiy tilni qayta ishlash jarayonlarida, jumladan, tinish belgilarini tiklashda ham ishlatilishi mumkin. Qaror daraxtlari (Decision Trees), tasodifiy o'rmonlar (Random Forest) va qo'llab-quvvatlovchi vektor mashinalari (Support Vector Machines) kabi ko'plab mashina o'rganish algoritmlari statistik asoslarga egadir. Shuning uchun ham mashinali o'qitish modellarini statistik modellashtirishning kengaytmasi sifatida ko'rish mumkin, ammo ular ko'pincha murakkabroq algoritmlarni va kattaroq ma'lumotlar to'plamini o'z ichiga oladi. Biroq bugungi kunda tinish belgilarini tiklash va shu kabi masalalar uchun an'anaviy mashinali o'qitish modellaridan ko'ra chuqur o'qitish neyron tarmoq modellari nisbatan muammoni hal etishda ancha ko'proq samara beradi. Boshqacha qilib aytganda, an'anaviy mashinali o'qitish modellari tinish belgilarini tiklash uchun ishlatilgan bo'lsa-da, ular ilg'or chuqur o'rganish modellari bilan solishtirganda til va kontekstual bog'liqliklarning nozik tomonlarini qamrab olishda sezilarli kamchiliklarga ega.

Chuqur o'qitish neyron tarmog'iga asoslangan yondashuv. Neyron tarmoq modellarining kashf etilishi mashinali o'qitishda yana bir sohaning paydo bo'lishiga olib keldi. Chuqur o'qitish neyron tarmoqlarining kengayishi bilan tabiiy tilni qayta ishlash masalalari, jumladan tinish belgilarini qayta ishlash masalalari o'zining cho'qqisiga chiqdi. Sohada natijalar sezilarli ravishda ortdi. Tabiiy tilni qayta ishlashda chuqur o'qitish neyron tarmoqlaridan RNN (Recurrent neural network), CNN (Convolutional neural network) va Transformer larga asoslangan neyron tarmoq modellaridan keng foydalaniladi.

Takroriy neyron tarmoqlari (RNN) ma'lumotlardagi ketma-ket bog'liqliklarni modellashtirish qobiliyati tufayli tinish belgilarini tiklash vazifalarida ishlatilgan [Tarwani va b., 2017]. Tinish belgilarini tiklash vazifasi berilgan matnning so'zlari atrofidagi boshqa so'zlardan kelib chiqib, to'g'ri tinish belgilarini bashorat qilishni o'z ichiga oladi. Ushbu neyron tarmoqlariga mansub uzoq-qisqa muddatli xotira (LSTM - Long Short-Term Memory) hamda ikki yo'nalishli uzoq-qisqa muddatli xotira (BiLSTM - Bidirectional LSTM) kabi neyron tarmoq arxitekturalarini tinish belgilarini tiklash uchun keng ko'lamda ishlatilganligini misol keltirishimiz mumkin. LSTMlar uzoq muddatli bog'liqliklarni olishda ustunlik qiladi, ikki yo'nalishli LSTMlar esa bu imkoniyatni har ikki yo'nalishdagi kontekstni hisobga olgan holda kengaytiradi va bu bilan tabiiy tilni qayta ishlashda tinish belgilarini tiklash kabi ikki tomonlama kontekstda muhim bo'lgan vazifalarni bajarish mumkin bo'ladi.



1-rasm. LSTM va BiLSTM neyron tarmoq arxitekturalari

Konvolyutsion neyron tarmoqlari (CNN) ko‘proq kompyuter vizualizatsiyasi (Computer vision) vazifalari bilan bog‘liq bo‘lsa-da, ular tabiiy tilni qayta ishlash vazifalarida, shu jumladan tinish belgilarini tiklashda ham qo‘llanilgan [Tundik, Szaszak, 2018]. CNN ma‘lum NLP (Natural Language Processing) vazifalarini hal etishda foydalanib kelinadi. Biroq ular matnlarda uzoq muddatli bog‘liqliklarda LSTM yoki Transformer modellari va takrorlanuvchi neyron tarmoq arxitekturalari kabi samarali bo‘lmashligini ta’kidlash kerak.

Transformerlarga asoslangan oldidan o‘qitilgan (pre-trained) til modellari. Ushbu modellari turli xil tabiiy tillarni qayta ishlash vazifalariga, shu jumladan tinish belgilarini tiklashga bugungi kunda keng qo‘llanildi. Transformerlarga asoslangan til modellari tinish belgilarini tiklash va tabiiy tilni qayta ishlash kabi vazifalarda sezilarli darajada o‘shishni amalga oshirdi. Bularga misol tariqasida BERT [Devlin va b., 2019] va RoBERTa [Liu va b., 2019] keltirish mumkin. Bundan tashqari hozirda shunga o‘xshash bir qancha oldindan o‘qitilgan Transformerlarga asoslangan til modellari ishlab chiqarilgan. Ushbu modellar yordamida hozirda ko‘plab matnlardagi tinish belgilarini tiklash masalalari amalga oshirilmoqda. Bularga misol qilib Tanvirul Alam, Akib Khan, Firoj Alamlar Bangla tili uchun o‘zlarining ilmiy ishida dataset yig‘ib bangla tili uchun tinish belgilarini tiklash modelini ishlab chiqishgan hamda uning natijalarini ingliz tilidagi model natijalari bilan solishtirishgan. Eng yaxshi natijalarni (yangiliklar transkripsiyalari dataseti uchun: Precision:87.4%, Recall:86.6%, F1 score:87.0%) XLM-RoBERTa modelida olishganlar [Alam va b., 2020]. Uygur Kurt, Aykut Chayirlar esa o‘zbek tiliga nisbatan yaqin bo‘lgan til turk tili uchun tinish belgilarni tiklashda 1.8 mln token, 146 mingta gapdan iborat korpus shakllantirib, ushbu dataset yordamida Transformerlarga asoslangan BERT, ELECTRA, ConvBERT kabi oldindan o‘qitilgan modellar yordamida yangi til modelini olishdi hamda eng yaxshi natijani (Precision: 0.823%, Recall: 0.859%, F1-score: 0.839%) ELECTRA modelida qayd etishdi [Kurt, Çayır, 2023].

Bundan ko‘rinib turibdiki Transformerlarga asoslangan oldindan o‘qitilgan til modellari hozirgi kunda tinish belgilarini tiklash kabi tabiiy tilni tanib olish

masalalarida eng ilg'or yondashuv deb hisoblash mumkin. Kelgusida o'zbek tili uchun ham aynan ushbu yondashuvdan foydalanish maqsadga muvofiqdir.

Foydalanilgan adabiyotlar:

1. Păiș, V., Tufiş, D. Capitalization and punctuation restoration: a survey. *Artif Intell Rev* 55, 1681–1722 (2022).
2. Liberman, M.Y. (1991). The Trend towards Statistical Models in Natural Language Processing. In: Klein, E., Veltman, F. (eds) *Natural Language and Speech*. ESPRIT Basic Research Series. Springer, Berlin, Heidelberg.
3. Tiberiu Boros, Stefan Daniel Dumitrescu and Sonia Pipa. *Fast and Accurate Decision Trees for Natural Language Processing Tasks*. 2017.
4. Kanchan M. Tarwani, Swathi Edem. *Survey on Recurrent Neural Network in Natural Language Processing*. 2017
5. Arvind T. Mohan, Datta V. Gaitonde. *A Deep Learning based Approach to Reduced Order Modeling for Turbulent Flow Control using LSTM Neural Networks*.
6. Mate Akos Tundik and Gyorgy Szaszak. *Joint Word- and Character-level Embedding CNN-RNN Models for Punctuation Restoration*. 2018.
7. Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. 2019.
8. Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, Veselin Stoyanov. *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. 2019.
9. Tanvirul Alam, Akib Khan, Firoj Alam. *Punctuation Restoration using Transformer Models for High-and Low-Resource Languages*. 2020.
10. Uygur Kurt, Aykut Çayır. *Transformer Based Punctuation Restoration for Turkish*. 2023.

O'ZBEK TILIDAGI SO'ZLARNING LEMMASINI TOPISH

Sobirov Og'abek Ollayor o'g'li,

sobirov.o@urdu.uz,

UrDU, Kompyuter ilmlari kafedrası tayanch doktoranti

Annotatsiya. Lemma bu so'zning asosiy, ya'ni lug'atda uchraydigan shakli. Lemmani topish (lemmalash) – turli shakldagi so'zlarning lemmasini topish jarayoni. Lemmalash jarayoni o'zbek tilining agglyutinatív va morfologik boyligi jihatidan murakkab jarayondir. Ushbu maqolada o'zbek tili so'zlarining lemmasini topish haqida mulohazalar keltirilgan. Bunda o'zbek tili so'z turkumlarining tuzilishiga ko'ra turlari jadval shaklida keltirilib, barcha so'z turkumlarining qo'shimchalar orqali yasalgan holatidan lemmasiga qaytish qoidalari misollar bilan keltirilgan. Qo'shimchalar turlari, soni va so'zlarga qo'shilish qoidalari haqida ham fikrlar yuritilgan. Lemmalash tabiiy til jarayonlarida muhim birinchi qadamlardan biridir. Bu jarayon Ma'lumot olish va Mashina tarjimasí kabi sohalarda foydalaniladi.

Abstract. Lemma is the fundamental, dictionary form of a word. Lemmatization is the process of finding the lemma of words of different forms. The lemmatization is a complex process in terms of the agglutinative and morphological richness of the Uzbek language. This article contains comments on finding the lemma of Uzbek language words. In this, the types of Uzbek language Part-of-Speech according to their structure are presented in the form of a table, and the rules for returning to the lemma of all part-of-speeches from the state formed by affixes are presented with examples. There are also the types, number, and joining to word rules of affixes. Lemmatization is an important first step in Natural Language Processing. This process is used in fields such as Information Retrieval and Machine Translation.

Аннотация. Лемма – это основная словарная форма слова. Лемматизация – это процесс нахождения леммы слов различной формы. Лемматизация представляет собой сложный процесс с точки зрения агглютинативного и морфологического богатства узбекского языка. В данной статье содержатся комментарии по поиску лемм в словах узбекского языка. При этом в виде таблицы представлены типы частей речи узбекского языка по их структуре, а также правила возврата к лемме всех частей речи из состояния, образованного аффиксами, представлены с примерами. Также различаются типы, количество и правила присоединения к словам аффиксов. Лемматизация – важный первый шаг в обработке естественного языка. Этот процесс используется в таких областях, как поиск информации и машинный перевод.

Kalit so'zlar. *Tabiiy Til Jarayonlari (TTJ), lemma, lemmani topish, juft so'zlar, takror so'zlar, qo'shimchalar, o'zbek tili.*

Kompyuter va turli texnologiyalarning rivojlanishi bilan birgalikda, elektron matnlarning hajmi ham ortib bormoqda. Har kuni minglab, hattoki millionlab terabayt ma'lumotlar vujudga kelmoqda. Bunga misol qilib quyidagilarni aytish mumkin, kitoblarning elektron shaklini hosil qilish, yangiliklar saytidagi yangi postlar, ijtimoiy tarmoqlardagi yozishmalar, yana ko'plab holatlarni misol keltirish mumkin. Bularning barchasi kompyuterda matnlarni qayta ishlashga olib keladi, bu jarayon esa allaqachon soha sifatida mavjud bo'lib, ushbu soha Tabiiy Til Jarayonlari (TTJ) [inglizcha: Natural Language Processing (NLP)] deb ataladi.

O'zbek tili uchun ushbu sohada, ya'ni TTJ sohasida bir qancha ishlar amalga oshirilgan va muhim natijalar olingan, hozirda ham davom etmoqda. Quyida shu ishlar bilan tanishib chiqilgan. Ushbu maqolada lemma topish jarayonining qoidaga asoslangan usuli chekli holat mashinasi yordamida ishlab chiqilgan, bunda qo'shimchalar bazasi ham shakllantirilgan [Sharipov, Sobirov, 2022]. Bu maqolada asos topish jarayonining qoidaga asoslangan usuli ishlab chiqilgan, bunda qo'shimchalar bazasi shakllantirilmagan, aksincha qo'shimchalarning 15 ta guruhi shakllantirilib chekli holatlar mashinasi yordamida bajarilgan [Sharipov, Yuldashov, 2022]. Ushbu maqolada o'zbek tili agglyutativ til ekanligi eslatilib, ot va olmosh so'z turkumlarini matnlarda aniqlash haqida so'z yuritilgan, bunda tub ot va olmosh so'zlar bazasi yaratilgan [Sharipov, Mavlanova, Sobirov, 2023]. Keyingi ishlarda matn tahlili va ma'lumot olish kabi TTJ vazifalari uchun o'zbek tilidagi matnlarda nomuhim so'zlarni aniqlash va avtomatik tahlil qilish borasidagi muhim ishlar va ularning natijalari bayon qilingan [Madatov, Bekchanov, 2023]. Tadqiqotlarda matnlarga dastlabki lingvistik ishlov berish jarayonlari yaxshi yoritilgan bo'lib, o'zbek, turk va uyg'ur tillari uchun asos topish, lemma topish va matnlarni belgilash ishlari amalga oshirilib, muhim natijalar bayon qilingan [Elov va boshqalar, 2023]. Ushbu maqolada matnlarni belgilash va atoqli otlarni tanish uchun o'zbek tili korpusi yordamida Yashirin Markov Modelidan foydalanib belgilash algoritmi ishlab chiqilgan, tahlil qilingan va muhim natijalar e'lon qilingan [Elov va boshqalar, 2023]. Matnlarni belgilash borasida yana bir muhim ishlardan birini ushbu maqolada ko'rishimiz mumkin, matnlarni avtomatik tahlil qilish uchun matnlarni belgilash uslubi haqida so'z borgan [Abjalova va boshqalar, 2021]. So'zlarning lemmasini topish, ushbu jarayonning TTJda muhimligi, qo'llanilish sohasi va o'zbek tili uchun bu jarayon bir muncha murakkabligi, turli misollar bilan yoritib berilgan [Xusainova, 2023].

O'zbek tili uchun TTJ sohasining dastlabki muhim lingvistik jarayoni bo'lmish lemmani topish (lemmalash) jarayoni uchun ushbu maqolada muhim ma'lumotlar va ularga misollar keltirilgan. Lemmani topish (lemmalash) [inglizcha: lemmatization] - turli shakldagi so'zlarning lemmasini topish jarayoni. Lemma bu so'zning asosiy ya'ni lug'atda uchraydigan shakli.

O'zbek tilida so'z turkumlarini morfologiya bo'limi o'rganadi. So'z turkumlari jami 12 ta, mustaqil so'z turkumlari: *ot, fe'l, sifat, ravish, son, olmosh*; yordamchi so'z turkumlari: *bog'lovchi, ko'makchi, yuklama*; oraliq yoki alohida

olingan soʻz turkumlari: *undov, taqlid, modal*. Bulardan mustail soʻz turkumlari tuzilishiga koʻra 4 xil *sodda, qoʻshma, juft* va *takror* boʻladi [1-rasm]. Boshqa soʻz turkumlari bunday tuzilish turlariga ega emas, quyidagi baʼzilaridan tashqari. Undov (qoʻshma: *bor boʻling*, juft: *eh-he*, takror: *hay-hay*); Taqlid (juft: *taqur-turuq*, takror: *yalt-yalt*); Modal (birikma: *mening fikrimcha*).

Tuzilish turi		Ot	Feʼl	Sifat	Ravish	Son	Olmosh
Sodda	Tub	<i>kitob</i>	<i>oʻqi</i>	<i>yashil</i>	<i>tez</i>	<i>besht</i>	<i>men</i>
	Yasama	<i>guldona</i>	<i>ishla</i>	<i>aqlli</i>	<i>eskicha</i>	-	<i>kimdir</i>
Qoʻshma		<i>qorbobo</i>	<i>kashf etdi</i>	<i>mehnatsevar</i>	<i>biroz</i>	<i>oʻn bir</i>	<i>mana bu</i>
Juft		<i>aka-uka</i>	<i>aytdi-qoʻydi</i>	<i>katta-kichik</i>	<i>qishin-yozin</i>	<i>ikki-uch</i>	<i>men-sen</i>
Takror		<i>qon-qoniga</i>	<i>oʻqib-oʻqib</i>	<i>uzun-uzun</i>	<i>oz-oz</i>	<i>uchta-uchta</i>	<i>oʻsha-oʻsha</i>
		Qisqartma <i>BMT</i>	KFSQ <i>aytib yuribdi</i>	Darajalash <i>juda, sal, eng, tim, nim, liq</i>			
			Toʻliqsiz feʼl <i>edi, emish, ekan, emas</i>				

1-jadval. Oʻzbek tili mustaqil soʻz turkumlarining tuzilishiga koʻra turlari

Matnda soʻzlar turli shakllarda uchraydi, uning lugʻatda mavjud variantini topish esa aynan lemmalash jarayoni hisoblanadi. Yuqoridagi jadvaldan foydalanib, ona-tilimiz qonun qoidalariga rioya qilgan holda matnda soʻzlarning turli shakldan uning lemmasini topish uchun qoidalarni koʻrib chiqib, misollar keltiramiz.

1. Sodda soʻz shakllari.

Sodda soʻz shakllari tub va yasama holatda uchraydi, bunda sodda tub soʻzlar hech qanday qoʻshimcha olmaydi va matnlarda shunday holatda uchrashi mumkin (yoki shakl yasovchi qoʻshimchalar qoʻshilgan holatda uchraydi), sodda yasama soʻzlar esa tarkibida soʻz yasovchi qoʻshimchalar olgan holda uchraydi, lemmalash jarayonida esa ushbu soʻz yasovchi qoʻshimchalar olib tashlanmaydi.

Matn	Soʻz	Lemma
<i>Kitobsiz uy – qalbsiz tanaga oʻxshaydi.</i>	<i>Kitobsiz</i>	<i>kitob</i>
<i>Siz oshpaz boʻlib ishlagan boʻlarmidingiz.</i>	<i>ishlagan</i>	<i>ishlamoq</i>
<i>Koʻzingni yumginda uchgacha sana.</i>	<i>uchgacha</i>	<i>uch</i>

2-jadval. Sodda soʻz shakllarini lemmalash jarayoniga misollar

2. Qoʻshma soʻz shakllari.

Qoʻshma soʻz shakllarida *feʼl*, *son* va *olmosh* har doim ajratib yoziladi, *sifat* esa har doim qoʻshib yoziladi, *ot* va *ravish* esa 2 xil holatda: ajratib va qoʻshib ham yoziladi.

Matn	Soʻz	Lemma
<i>Ishxonamiz yaqinda yangicha qiyofaga kiradi.</i>	<i>Ishxonamiz</i>	<i>ishxona</i>
<i>Men sizlarga yaxshi yangilik olib keldim.</i>	<i>olib keldim</i>	<i>olib kelmoq</i>
<i>Gultojixoʻroz urugʻini arzon narxda sotib oling.</i>	<i>Gultojixoʻroz</i>	<i>gultojixoʻroz</i>

3-jadval. Qoʻshma soʻz shakllarini lemmalash jarayoniga misollar

3. Juft soʻz shakllari.

Juft soʻzlar har doim chiziqli bilan qoʻshib yoziladi. Chiziqli va soʻzlar orasida boʻsh joy tashlanmay qoʻshiladi. Ular oʻzaro yaqin munosabatdagi soʻzlardir. Juft soʻzdan qoʻshimcha bilan yasalgan soʻzlar ham chiziqli bilan yoziladi.

Matn	Soʻz	Lemma
<i>Keyin mehr-shafqatdan soʻz ochdi.</i>	<i>mehr- shafqatdan</i>	<i>mehr-shafqat</i>
<i>Imtihon uchun kecha-yu kunduz tayyorlandi.</i>	<i>kecha-yu kunduz</i>	<i>kecha-kunduz</i>
<i>Har kimga oʻn-oʻn beshtadan tegadi.</i>	<i>oʻn-oʻn beshtadan</i>	<i>oʻn-oʻn besh</i>

4-jadval. Juft soʻz shakllarini lemmalash jarayoniga misollar

4. Takror soʻz shakllari.

Takror soʻz shakllari ham juft soʻzlar kabi har doim chiziqli bilan yoziladi. Takror soʻzlar maʼnoni kuchaytirish uchun bitta soʻzni takrorlashdan hosil qilinadi. Lugʻatlarda soʻzlar har doim ham takroriy shakllarda uchramaydi shuning uchun

ularning lemmasi faqat bitta soʻzdan yoki ularning takrorlanishidan iborat boʻlishi mumkin.

Matn	Soʻz	Lemma
<i>Uzoq-uzoqlarga ketib qolgan ekan.</i>	<i>Uzoq-uzoqlarga</i>	<i>uzoq</i>
<i>Eshik oʻzidan-oʻzi ochilib ketdi.</i>	<i>oʻzidan-oʻz</i>	<i>oʻz</i>
<i>Har kimga bitta-bittadan beramiz.</i>	<i>bitta-bittadan</i>	<i>bitta-bitta</i>

5-jadval. Takror soʻz shakllarini lemmalash jarayoniga misollar

Qolgan shakllarda ham oʻzbek tilining tilshunoslik qoidalari asosida lemmalash jarayoni amalga oshiriladi. Bunday shakllar quyidagilardir:

- Otlarda qisqartmalar: *BTMning* – *BMT*
- Feʼllarda KFSQ: *aytib yuribdi* – *aytmoq*
- Toʻliqsiz feʼllar: *borgan ekan* – *bormoq*
- Sifatlarda daraja shakllari: *tim qora* – *qora*

Foydalanilgan adabiyotlar:

1. Sharipov M, Sobirov O. Development of a Rule-Based Lemmatization Algorithm Through Finite State Machine for Uzbek Language. CEUR Workshop Proceedings, CEUR-WS 2022, 154 – 159.
2. Sharipov M, Yuldashov O. UzbekStemmer: Development of a Rule-Based Stemming Algorithm for Uzbek Language. CEUR Workshop Proceedings, CEUR-WS 2022, 137 – 144.
3. Sharipov M, Mavlonova M, Sobirov O. Oʻzbek tilidagi matnlarda ot va olmosh soʻz turkumlariga tegishli soʻzlarni aniqlovchi algoritm yaratish. Computer Linguistics: Problems, Solutions, Prospects 2023; 1.
4. Madatov K, Bekchanov S, Vičič J. Automatic Detection of Stop Words for Texts in the Uzbek Language. Informatica (Slovenia) 2023; 47: 143 – 150.
5. Boltayevich EB, Adali E, Mirdjanovna KS, Xolmoʻminovna AO, Yuldashevna XZ, Nizomaddin Uktamboy oʻgʻli X. The Problem of Pos Tagging and Stemming for Agglutinative Languages (Turkish, Uyghur, Uzbek Languages). UBMK 2023 - Proceedings: 8th International Conference on Computer Science and Engineering, 2023, 57 – 62.
6. Boltayevich EB, Samariddinovich SS, Mirdjanovna KS, Adali E, Yuldashevna XZ. Pos Tagging of Uzbek Text Using Hidden Markov Model. UBMK 2023 - Proceedings: 8th International Conference on Computer Science and Engineering, 2023, 63 – 68.
7. Abdurashetona AM, Ismailovich IO. Methods of Tagging Part of Speech of Uzbek Language. Proceedings - 6th International Conference on Computer Science and Engineering, UBMK 2021, 2021, 82 – 85.
8. Xusainova Z. Methods of Lemmatizing Various Structured Words in Uzbek Languageʻʻ. CAJLPC 2023; 4: 67–72.

O'ZBEK TILI KORPUSIGA MORFOLOGIK ISHLOV BERISH

Nizomaddin Xudayberganov

nizomaddin@navoiy-uni.uz

ToshDO'TAU o'qituvchisi

Annotatsiya. Til korpusini morfologik tahlil qilish uchun uni morfologik teglash eng muhim NLP vazifalaridan biri hisoblanadi, chunki u korpusni sintaktik, semantik va pragmatik tahlil qilish uchun muhim qadam hisoblanadi. O'zbek tilidagi til birliklarining morfologik teglash qo'lda va yarim avtomatik tarzda amalga oshiriladi. Korpus matnlarini teglash chatbotlar, virtual yordamchilar, qidiruv tizimlari va mashina tarjimai kabi tabiiy tilni qayta ishlash (NLP) ilovalarida asosiy amal hisoblanib, u sun'iy intellekt modellariga inson tili va muloqotni yanada samaraliroq tushunishga yordam beradi. Korpus matnlarini annotatsiyalash yoki teglash orqali so'z, ibora, so'z birikmasi shu kabi leksik birliklarni izohlash kabi matnga teg (izoh) qo'shish va ularni belgilashdan iborat. Mashinali o'rgatishda matn annotatsiyasi – bu mashinali o'rgatish modellariga inson tilini tushunish va qayta ishlashda yordam berish uchun matnga metama'lumotlar yozish jarayoni. Yuqori sifatli ma'lumotlar to'plami sun'iy intellektni o'qitishda juda muhim, chunki ular AI modellarining ishlashi, aniqligi va ishonchliligiga bevosita ta'sir qiladi. Matnni teglash orqali kalit so'zlarni ajratib olish evaziga qidiruv tizimi natijasini optimallashtirish mumkin. ToshDO'TAU Kompyuter lingvistikasi va raqamli texnologiyalari kafedrasida olimlari tomonidan til korpusi uchun maxsus teglar to'plami ishlab chiqilgan bo'lib, o'zbek tilidagi so'zlardagi gap bo'laklari va qo'shimchalari birikmalarining eng keng tarqalgan affiksalarini aniqlanib, ularni belgilangan teglar asosida tahlil qilish, til korpusining yanada rivojlanishiga asos bo'lib xizmat qilishi mumkin. Til korpusini morfologik teglash uchun 12 dan ortiq maxsus teglar tanlab olindi va korpusda nutq birliklari teglandi.

Abstract. Morphological tagging of a language corpus for morphological analysis is one of the most important NLP tasks, as it is an important step for syntactic, semantic and pragmatic analysis of the corpus. Morphological tagging of Uzbek language units is done manually and semi-automatically. Corpus text tagging is a key practice in natural language processing (NLP) applications such as chatbots, virtual assistants, search engines, and machine translation, where it helps AI models understand human language and communication more effectively. Annotating or tagging corpus texts consists of adding a tag (explanation) to the text and defining them, such as explaining words, phrases, phrases, and similar lexical units. Text annotation in machine learning is the process of writing metadata to text to help machine learning models understand and process human language. High-quality data sets are critical to AI training, as they directly affect the performance, accuracy, and reliability of AI models. By tagging text, you can optimize search engine results by extracting keywords. ToshDO'TAU scientists of the Department of Computer

Linguistics and Digital Technologies have developed a set of special tags for the language corpus, identifying the most common affixes of phrases and adverbs in Uzbek words, analyzing them based on the specified tags, for the further development of the language corpus. can serve as a basis. More than 12 special tags were selected for morphological tagging of the language corpus and speech units were tagged in the corpus.

Аннотация. Морфологическая разметка языкового корпуса для морфологического анализа является одной из важнейших задач NLP, поскольку является важным этапом синтаксического, семантического и прагматического анализа корпуса. Морфологическая разметка единиц узбекского языка осуществляется вручную и в полуавтоматическом режиме. Тегирование корпуса текста — это ключевая практика в приложениях обработки естественного языка (NLP), таких как чат-боты, виртуальные помощники, поисковые системы и машинный перевод, где они помогают моделям AI более эффективно понимать человеческий язык и общение. Аннотирование или маркировка корпусных текстов заключается в добавлении к тексту тега (пояснения) и его определении, например пояснении слов, словосочетаний, словосочетаний и подобных лексических единиц. Текстовые аннотации в машинном обучении — это процесс записи метаданных в текст, который помогает моделям машинного обучения понимать и обрабатывать человеческий язык. Наборы высококачественных данных имеют решающее значение для обучения AI, поскольку они напрямую влияют на производительность, точность и надежность моделей AI. Помечая текст тегами, вы можете оптимизировать результаты поисковых систем, извлекая ключевые слова. Ученые ТошДО'ТАУ кафедры компьютерной лингвистики и цифровых технологий разработали набор специальных тегов для языкового корпуса, определяющих наиболее распространенные аффиксы словосочетаний и наречий в узбекских словах, анализируя их на основе указанных тегов, для дальнейшего развития языкового корпуса. может служить основой. Для морфологической разметки языкового корпуса было выбрано более 12 специальных тегов и в корпусе были размечены речевые единицы.

Kalit soʻzlar: *Til korpusi, matnni qayta ishlash, lingvistik tahlil, mashinali oʻqitish modellari, korpus turlari, parallel korpus, teglangan korpuslar, POS teg.*

Kirish

Morfologik teglangan korpus ilmiy tadqiqotchilariga turli xil lingvistik tahlil va tabiiy tilni qayta ishlash (NLP)ni amalga oshirish imkonini beradi. Morfologik ishlov berilgan/teglangan korpusdan tabiiy tildagi soʻz shakllari, nutq qismlari (soʻz turkumlari) va grammatik xususiyatlarning chastotasi, tarqalishi va birikmasi haqidagi maʼlumotlarni aniqlash uchun foydalanish mumkin. Shuningdek, morfologik tahlil, generatsiya, segmentatsiya va tildagi noaniqlik kabi vazifalar

uchun modellar va algoritmlarni o'rgatish va sinab ko'rish uchun ham qo'llaniladi. Sintaktik, semantik yoki pragmatik kabi boshqa turdagi korpusni teglash usullari uchun boshlang'ich qadam hisoblanadi.

Korpusni morfologik teglash haqida (Ide, Pustejovsky, 2017) “The Handbook of Linguistic Annotation” nomli kitobda morfologiyaning tamoyillari, usullari va muammolari haqida umumiy ma'lumot bergan [Erjavec, 1998; Kyoto, 1994]. Universal Dependencies loyihasi (Nivre va boshq., 2020)da bog'liqlik sintaksisi va morfologiyasi uchun tillararo teglash sxemalari va resurslarini ishlab chiqish haqida fikr mulohaza yuritilgan [Nivre, 2020]. Shuningdek, Savary va boshq. (2023) PARSEME ilmiy tadqiqotida 18 tildagi murakkab fe'l shakllarining morfologik teglash usullarini keltirilgan [Savary, 2023]. Kurimo va boshq. (2010) Morpho Challenge tadqiqotida turli xil morfologik tahlil va segmentatsiya uchun nazoratsiz mashinali usullarni baholagan [Kurimo, 2010].

Korpusga morfologik ishlov berish

Morfologik ishlov berish korpusda ishlatiladigan so'zlarni avtomatik tahlil qilishdan iborat. Morfologik tahlil jarayonida gap(matn)dagi so'z, qo'shma so'zlar va so'z birikmalari qayta ishlanadi. Morfologik tahlilning asosiy maqsadi – matndagi **so'zlarni aniqlash, uni kontekst qurshovidan ajratib olish, morfemik tuzilishini tahlil qilish, asl ma'nosi va matndagi sintaktik rolini aniqlash**dir. Matnni qayta ishlash natijasida olingan ma'lumotlardan NLPning *so'z ma'nosini aniqlash, lug'at tuzish, tahlil qilish, til o'rganish* kabi vazifalarida foydalanish mumkin. Ingliz va boshqa tillar uchun morfologik tahlilni amalga oshirish bo'yicha Greene va Rubin (1971), Karttunen va Wittenburg (1983), Koskenniemi (1983), deHaan (1984), Garside (1987), Church (1991), deRose (1991), Merialdo (1994) kabi olimlar ilmiy tadqiqotlar olib borishgan.

O'zbek tilida morfologik tahlilni quyidagicha amalga oshirish mumkin:

1. **So'z turkumi.**
2. **Yasalishi.**
3. **Tuzilishi.**
4. **Grammatik ma'nolari.**

Shuningdek, har bir so'z turkumi o'ziga xos morfologik xususiyatlarga ega bo'lganligi sababli har bir so'z turkumining tahlil parametrlari turlicha bo'ladi. Masalan, olmosh so'z turkumi morfologik tahlilida quyidagilarga e'tibor berish maqsadga muvofiq:

Olmosh so'z turkumi

1. Ma'noviy guruhlar:

- 1) kishilik olmoshi = [kish.olm.];
- 2) ko'rsatish olmoshi = [ko'r.olm.];
- 3) o'zlik olmoshi = [o'z.olm.];
- 4) gumon olmoshi = [gum.olm.];
- 5) so'roq olmoshi = [so'r.olm.];
- 6) belgilash olmoshi = [belg.olm.];

7) bo'lishsizlik olmoshi = [bo'l.olm.].

2. Tuzilishiga ko'ra turi:

- 1) sodda olmosh = [s.olm.];
- 2) qo'shma olmosh = [qo'sh.olm.];
- 3) juft olmosh = [j.olm.];
- 4) takror olmosh = [tak.olm.].

3. Yasalishi:

- 1) tub olmosh = [tub olm.];
- 2) yasama olmosh = [yas.olm.].

4. Olmoshlarda kelishik kategoriyasi:

- 1) [bosh kelishik] = [b.k.];
- 2) [qaratqich kelishigi] = [qar. k.];
- 3) [tushum kelishigi] = [tush. k.];
- 4) [jo'nalish kelishigi] = [j. k.];
- 5) [o'rin-payt kelishigi] = [o'.p. k.];
- 6) [chiqish kelishigi] = [ch. k.].

5. Son kategoriyasi:

- 1) [birlik son] = [birl. s.];
- 2) [ko'plik son] = [ko'p. s.].

6. Egalik kategoriyasi:

- 1) [I shaxs] = [I sh.];
- 2) [II shaxs] = [II sh.];
- 3) [III shaxs] = [III sh.].

Tabiiy til bilan shug'ullanadigan mutaxassislar tomonidan amalga oshirilgan morfologik tahlil samaradorligi yuqori hisoblanadi. Bugungi kunda mashinalar yordamida morfologik tahlilni amalga oshirishda 95% dan yuqori samaradorlikka erishilgan [Elov, Hamroyeva, Ahmedova, 2023].

POS teglash (Part-of-Speech Tagging)

Matndagi so'z yoki so'z birikmalarining muayyan xususiyatlarini ko'rsatish uchun ularga maxsus kodlar biriktiriladi va bu amal NLPda **lingvistik annotatsiya** yoki **teglash** deb nomlanadi. Xususiyatlarga tayinlangan kodlar **teg** deb nomlanadi.

POS teglash (so'z turkumini belgilash) – so'zning gapda qo'llanilgan turkumi asosida belgilash hisoblanadi. POS teglash amali uch bosqichda amalga oshiriladi [Elov, Hamroyeva, Abdullayeva, Uzakova, 2022].

- *boshlang'ich tahrirlash*;
- *avtomatik teglash*;
- *qo'lda tahrirlash*.

POS teglashning *boshlang'ich tahrirlash* bosqichda korpusdagi har bir so'z yoki so'z birikmasi POS teglash uchun mos formatga aylantiriladi. Leksik va grammatik omonimiya tufayli bitta so'z bir nechta POS tegiga ega bo'lishi mumkin.

Soʻzning turkumini (POS)ni dastlabki belgilashdan soʻng matnlardagi soʻzlarni ajratish uchun soʻzlarning teglari qoʻlda tuzatiladi.

Oʻzbek tili soʻz turkumlari uchun B.Elov va boshqalar tomonidan quyidagi teglar toʻplami ishlab chiqilgan [Elov;, Hamroyeva, Abdullayeva, Uzakova, 2022]:

1-jadval. Oʻzbek tilidagi POS teglar roʻyxati (qisqa variant)

№	POS teg	Belgilanishi
1.	Ot	N
2.	Sifat	JJ
3.	Feʼl	VB
4.	Son	NUM
5.	Ravish	RR
6.	Olmosh	P
7.	Bogʻlovchi	C
8.	Koʻmakchi	II
9.	Yuklama	Prt
10.	Modal soʻz	MD
11.	Undov soʻz	UH
12.	Taqlid soʻzlar	IM
13.	Atoqli otlar	NER
14.	Iboralar	IB

POS teglashning ikkinchi darajasi **grammatik teglash (grammatical tagging)** deb ataladi. Grammatik teglash jarayonida korpusdagi har bir soʻzga mos barcha grammatik maʼlumotlar belgilanadi. Ushbu jaryondan matnni tahlil qilishda, soʻzlarning grammatik maʼnosini aniqlash/farqlashda foydalaniladi. **Morfologik, leksik, semantik** va **sintaktik** tahlil bilan bogʻliq muammolarni hal qilishda matnda uchraydigan har bir soʻz uchun grammatik maʼlumotlarni oʻz ichiga olgan korpusga ehtiyoj mavjud.

Soʻz maʼnosini belgilash (Word Sense Tagging)

Soʻz maʼnosini teglash yoki **semantik teglash** uchun POS va grammatik teglangan matndan foydalaniladi, teglangan matnlar ustida amallar bajariladi. Matnni avtomatik semantik teglashdan soʻng har bir soʻz toʻgʻri semantik tasnifga ega boʻlishini taʼminlash maqsadida qoʻlda **post-tahrirlash** amalga oshiriladi.

Soʻz maʼnosini belgilashning quyidagi misoli Buyuk Britaniyaning Lankaster universiteti loyihasidan olingan. *I like a particular shade of lipstick.*

2-jadval. Soʻz maʼnosini belgilash

№	POS va grammatik teg	Soʻz	Semantik teg
1.	PPIS1	<i>I</i>	Z8
2.	VV0	<i>like</i>	E2+

3.	AT1	<i>a</i>	Z5
4.	JJ	<i>particular</i>	A4.2+
5.	NN1	<i>shade</i>	O4.3
6.	IO	<i>of</i>	Z5
7.	NN1	<i>lipstick</i>	B4

Bu jadvalda matn pastga qarab o'qiladi, chap tomonda POS va grammatik teglar, o'ng tomonda esa so'z ma'nosini anglatuvchi teglar keltirilgan.

Semantik teglar umumiy nutq maydonini ko'rsatadigan bosh harf, maydonning birinchi bo'linmasini ko'rsatadigan raqam, aniqroq bo'linmani ko'rsatish uchun o'nli nuqtadan keyin keyingi raqam, ko'rsatish uchun bir yoki bir nechta “ortiqcha” yoki “minyslardan” iborat semantik miqyosda **ijobiy** yoki **salbiy** pozitsiya kabi ma'lumotlardan tashkil topadi.

Masalan, **A4.2+** tegi “*umumiy va mavhum so'zlar*” (**A**), “*tasniflash*” (**A4**) toifasidagi, “*alohida va umumiy*” (**A4.2**) va “*alohida*” toifasidagi so'zni ko'rsatadi. “**umumiy**” (**A4.2+**). Xuddi shunday, **E2+** “*hissiy holatlar, harakatlar, hodisalar va jarayonlar*” (**E**), “*yoqtirish va yoqtirmaslik*” (**E2**) toifasiga kiradi va “*yoqmaslik*” (**E2+**) emas, balki “**yoqtirish**”ni anglatadi.

Xulosa

Tabiiy tilni qayta ishlash vazifalarining aksariyatida zamonaviy tizimlar odatda matematik modellarini yaratish uchun mashinali o'rganish usullariga tayanadi. Ushbu tizimlarning aksariyati nazorat ostida o'rganish strategiyalarini qo'llashini hisobga olsak, muammoli soha uchun teglangan korpus muhim ahamiyatga ega. Korpusga morfologik ishlov berishning hozirgi usuli bir nechta mutaxassislarni yollash va ularni qo'lda yoki yordamchi dastur yordamida korpusga teglashdan iborat. Biroq, bu usul qimmat va vaqt talab etadi. Ushbu maqolada biz yuqoridagi muammolarni hal qilishga qaratilgan yangi usul taklif qilindi.

Foydalanilgan adabiyotlar:

1. Erjaveç, T., Ide, N.: The MULTTEXT-East corpus. In: Proceedings of First International Conference on Language Resources and Evaluation, pp. 971–974 (1998)
2. Ide, N., Véronis, J. MULTTEXT: multilingual text tools and corpora. In: Proceedings of the 15th International Conference on Computational Linguistics (COLING 94), vol. I, pp. 588–592. Kyoto, Japan (1994)
3. Nivre, J., de Marneffe, M. C., Ginter, F., Haji, J., Manning, C. D., Pyysalo, S., Schuster, S., & Zeman, F. T. D. (2020). Universal dependencies v2: An evergrowing multilingual treebank collection. LREC 2020 - 12th International Conference on Language Resources and Evaluation, Conference Proceedings.
4. Savary, A., Khelil, C. ben, Ramisch, C., Giouli, V., Mititelu, V. B., Mohamed, N. H., Krstev, C., Liebeskind, C., Xu, H., Jiang, M., Stymne, S., Güngör,



T., Pickard, T., Guillaume, B., Bhatia, A., Butler, A., Candito, M., Gantar, A., Iñurrieta, U., ... Walsh, A. (2023). PARSEME Corpus Release 1.3. 19th Workshop on Multiword Expressions, MWE 2023 - Proceedings. <https://doi.org/10.18653/v1/2023.mwe-1.6>

5. Kurimo, M., Virpioja, S., Turunen, V., & Lagus, K. (2010). Morpho challenge competition 2005-2010: Evaluations and results. Proceedings of the Annual Meeting of the Association for Computational Linguistics.

6. Boltayevich, E. B., Mirdjonovna, H. S., & Ilxomovna, A. X. (2023). Methods for Creating a Morphological Analyzer. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 13741 LNCS. https://doi.org/10.1007/978-3-031-27199-1_4

7. Elov B., Hamroyeva Sh., Abdullayeva O., Uzakova M. O'zbek tilida POS tegging masalasi: muammo va takliflar / Uzbekistan: Language and Culture. Applied philology. 2022/2(5). – 51-68-b.

8. Elov B., Hamroyeva Sh., Xudayberganov N., Yodgorov U., Yuldashev A. Pos tagging of Uzbek texts using hidden Markov models (HMM) and Viterbi algorithm. O'zMU xabarlar. Mirzo Ulug'bek nomidagi O'zbekiston Milliy universiteti ilmiy jurnali. 2023 yil Maxsys son.

9. Elov, B.B., Hamroyeva, Sh.M., Abdullayeva, O.X., Husainova, Z.Y., Xudoyberganov, N.U. 2023. “Agglutinatив tillar uchun pos teglash va stemming masalasi (turk, uyg'ur, o'zbek tillari misolida)”. O'zbekiston: til va madaniyat 2: 6-39.

BERT TIL MODELIDAN FOYDALANGAN HOLDA O'ZBEK TILI LINGVISTIK MUAMMOLARINI YECHISH

Jumanazarov Mardonbek Davronbek o'g'li

jumanazarovmardonbek99@gmail.com,

UrDU magistranti

Madatov Xabibulla Axmedovich

fizika-matematika fanlari nomzodi

khabibulla@urdu.uz,

UrDU dotsenti

Annotatsiya. BERTbek nomi bilan ma'lum bo'lgan o'zbek BERT modelining paydo bo'lishi o'zbek tili uchun tabiiy tilni qayta ishlashda (NLP) muhim bosqichni ko'rsatadi. Ushbu maqola BERTbekning turli lingvistik vazifalarni, xususan, hissiyotlarni tahlil qilish, matnlarni tasniflash, niqoblangan tilni modellashtirish (MLM) va nomuhim so'zlarni olib tashlashdagi imkoniyatlarini o'rganadi. Tadqiqotchilar va ishlab chiquvchilar ushbu oldindan o'rgatilgan modeldan foydalanish orqali o'zbek tilidagi NLP ilovalarining samaradorligi va aniqligini oshirishi, til texnologiyasi va lingvistik tadqiqotlardagi yutuqlarni qo'llab-quvvatlashi mumkin.

Abstract. The advent of the Uzbek BERT model, known as BERTbek, marks a significant milestone in natural language processing (NLP) for the Uzbek language. This essay explores the capabilities of BERTbek in addressing various linguistic tasks, specifically sentiment analysis, text classification, masked language modeling (MLM), and stopwords removal. By leveraging this pre-trained model, researchers and developers can enhance the efficiency and accuracy of NLP applications in Uzbek, fostering advancements in language technology and linguistic research.

Аннотация. Появление узбекской модели BERT, известной как BERTbek, знаменует собой важную веху в области обработки естественного языка (NLP) для узбекского языка. В этом эссе исследуются возможности БЕРТбека в решении различных лингвистических задач, в частности, анализа настроений, классификации текста, моделирования языка в масках (MLM) и удаления стоп-слов. Используя эту предварительно обученную модель, исследователи и разработчики могут повысить эффективность и точность приложений НЛП на узбекском языке, способствуя развитию языковых технологий и лингвистических исследований.

Kalit so'zlar: *NLP, BERT, o'zbek tili, sentiment tahlil, matnlar tasnifi, MLM (maskalangan tilni modellashtirish), nomuhim so'zlar.*

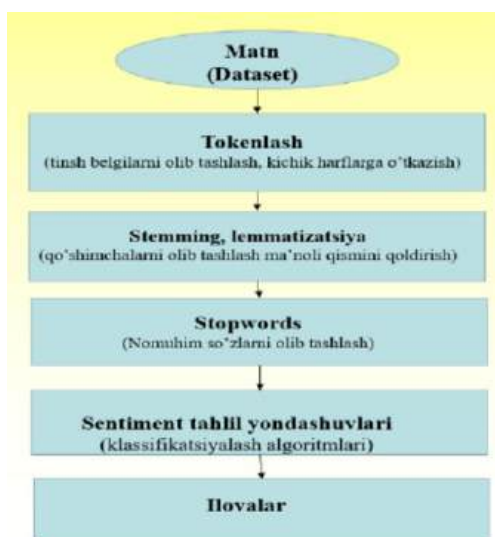
Kirish

BERT (Transformerlardan ikki tomonlama kodlovchi vakilliklari) modellari matnning chuqur kontekstli tasvirlarini taqdim etish orqali NLP sohasida inqilob

qildi. O'zbek tili uchun mo'ljallangan BERTbek dasturining ishlab chiqilishi avval ixtisoslashtirilgan hisoblash resurslari yo'qligi sababli qiyin bo'lgan NLP vazifalari uchun yangi yo'llarni ochadi. O'zbekcha matnning katta jamlanmasida tayyorlangan ushbu model tilning nozik tomonlarini qamrab olgan bo'lib, yanada nozik va samarali lingvistik tahlil qilish imkonini beradi. BERTbekning hissiyotlarni tahlil qilish, matn tasnifi, MLM va nomuhim so'zlarni olib tashlashda qo'llanilishi uning ko'p qirraliligi va o'zbek kontekstidagi lingvistik vazifalarni o'zgartirish imkoniyatlarini ko'rsatadi.

NLP vazifalari

Sentiment tahlil. BERTbek jamoatchilik fikrini tushunish, bozor tadqiqotlari va ijtimoiy media monitoringi uchun muhim bo'lgan matn segmentlarining emotsional ohangini aniqlaydigan hissiyotlarni tahlil qilishda ustundir. BERTbek orqali kiritilgan matnni qayta ishlash orqali his-tuyg'ularni ijobiy, salbiy yoki neytral deb aniq tasniflash mumkin, bu esa o'zbekzabon jamoalarning jamoaviy kayfiyati yoki fikri haqida qimmatli ma'lumotlarni taqdim etish imkonini beradi.



O'zbek tilidagi matnlarni sentiment tahlil qilish bosqichlari dastlab berilgan matnni fundamental NLP qadamlardan, masalan transliteratsiya [Kutlimuratova, Kuriyozov, Tillaeva, 2022: 161-171], tokenlash [Sharipov, Salayev, Matlatipov, 2021], hamda stopwrd(nomuhim so'z)lardan tozalash [Kutlimuratova, 2021] kabi qadamlardan o'tkazib, keyin sentiment tahlil qiluvchi modelga kerakli formatda yetkazib berishdan iborat. Bu bosqichlar quidagi blok-sxemada ko'rsatilgan.

Lingvistik qoidalar qurish usulida salbiy va ijobiy ma'no beruvchi so'zlar bazasidan yoki umumiy lingvistik qoidalardan foydalaniladi. Sentimental gaplarda sifatlar ko'p qo'llanilganligi uchun sifatlarning orttirma darajasini yasovchi so'zlarni ham kiritish orqali matnning salbiy yoki ijobiy ekanligini aniqroq aytishimiz mumkin.

Bu turdagi dasturga oddiy misol:

import re

```
pos_suzlar="yaxshi, a'lo, chiroyli, aqlli, go'zal, yoqdi, bo'ladi, ajoyib, ko'p,  
qiziqarli"
```

```
neg_suzlar="yomon, rasvo, xunuk, aqlsiz, dabdala, qiziqarsiz, zerikarli,  
yoqmadi"
```

```
kuchaytiruvchi_suzlar="juda, eng, bag'oyatda, g'oyat, nihoyatda, behad "
```

```
gap=input("Matn kiriting: ")
```

```
gap1=gap.split()
```

```
t=0
```

```
for i in gap.split():
```

```
    if re.findall(i,pos_suzlar):
```

```
        t=1
```

```
        if re.findall(j, kuchaytiruvchi_suzlar):
```

```
            print("Bu ijobiy gapligi aniq")
```

```
            break
```

```
        print("Bu ijobiy gap")
```

```
        break
```

```
    elif re.findall(i, neg_suzlar):
```

```
        t=1
```

```
        if re.findall(j, kuchaytiruvchi_suzlar):
```

```
            print("Bu salbiy gapligi aniq")
```

```
            break
```

```
        print("Bu salbiy gap")
```

```
        break
```

```
    j=i
```

```
if t==0: print("Aniqlab bo'lmadi.")
```

Matnlar tasnifi

Matnlarni tasniflashda BERTbek matnga oldindan belgilangan toifalarni tayinlaydi, bu esa kontentni tashkil etish, mavzuni ochish va ma'lumot qidirishni osonlashtiradi. Yangilik maqolalarini saralash, ilmiy maqolalarni tasniflash yoki veb-kontentni tartibga solish bo'ladimi, BERTbekning chuqur o'rganish imkoniyatlari tasniflash topshiriqlarida yuqori aniqlik va dolzarblikni ta'minlaydi, o'zbek tilida ma'lumotlarga kirish va boshqarishni yaxshilaydi.

Niqoblangan tilni modellashtirish (MLM)

BERTning asosiy o'quv komponenti bo'lgan MLM jumlada yetishmayotgan so'zlarni bashorat qilishni o'z ichiga oladi. BERTbekning MLM tilidagi malakasi tilni ilg'or tushunish va yaratish imkonini beradi, matnni to'ldirish, avtomatik tuzatish va til o'rgatish kabi vazifalarni bajarishda yordam beradi. Bu qobiliyat o'zbek tilidan dinamik foydalanishni qo'llab-quvvatlovchi tilga asoslangan intellektual dasturlarni ishlab chiqish uchun juda muhimdir.

Nomuhim so'zlarni olib tashlash

BERT modellarining to'g'ridan-to'g'ri qo'llanilishi bo'lmasa-da, nomuhim so'zlarni olib tashlash NLP-da dastlabki ishlov berishning muhim bosqichidir.

BERTbek kontekst va foydalanish asoslarini tushunish orqali nomuhim soʻzlar roʻyxatini aniqlashtirishga yordam beradi va shu bilan quyi oqim NLP vazifalarining aniqligini oshiradi. BERTbek tegishli boʻlmagan soʻzlarni aniqlash va yoʻq qilish orqali oʻzbek tilidagi lingvistik tahlillarning yoʻnalishi va samaradorligini oshiradi.

Xulosa

Oʻzbek BERT modeli BERTbek oʻzbek tilidagi turli lingvistik vazifalarni hal qilish uchun yangi vosita hisoblanadi. Uning his-tuygʻularni tahlil qilish (sentiment tahlil), matn tasnifi, MLM va nomuhim soʻzlarni olib tashlashda qoʻllanilishi modelning moslashuvi va oʻzbekcha matnni qayta ishlash va tushunishdagi kuchini koʻrsatadi. BERTbek nafaqat lingvistik tadqiqotlar va NLP ilovalarini ishlab chiqishni olgʻa siljitadi, balki oʻzbek tilining texnologik imkoniyatlarini kengaytirishga sezilarli hissa qoʻshadi va oʻzbek tilida soʻzlashuvchilar uchun tilni qayta ishlash yanada qulay, aniq va nozik boʻlgan kelajakni vaʼda qiladi.

Foydalanilgan adabiyotlar:

1. Salaev, U. I., Kuriyozov, E. R., & Matlatipov, G. R. (2023, November). Design and Implementation of a Tool for Extracting Uzbek Syllables. In *2023 IEEE XVI International Scientific and Technical Conference Actual Problems of Electronic Instrument Engineering (APEIE)* (pp. 1750-1755). IEEE.
2. Madatov, K. (2019). A prolog format of uzbek WordNet's entries. *Human Language Technology as a Challenge for Computer Science and Linguistics*, 316-320.
3. Allaberdiev, B., Matlatipov, G., Kuriyozov, E., & Rakhmonov, Z. (2024). Parallel texts dataset for Uzbek-Kazakh machine translation. *Data in Brief*, 110194.
4. Kutlimuratova, B., Kuriyozov, E., & Tillaeva, M. (2022). TEACHING ENGLISH AS A FOREIGN LANGUAGE FOR PRIMARY SCHOOL CHILDREN: LITERATURE REVIEW. *FOREIGN LANGUAGE TEACHING AND APPLIED LINGUISTICS*, 161-171.
5. Sharipov, M., Salaev, U., & Matlatipov, G. (2021). OʻZBEK TILI FEʼL SOʻZ TURKUMI UCHUN CHEKLI AVTOMATLAR ASOSIDA STEMMING ALGORITMINI YARATISH. *COMPUTER LINGUISTICS: PROBLEMS, SOLUTIONS, PROSPECTS*, 1(1).
6. Kutlimuratova, B. (2021). Uzbek students learning English as a foreign language: Error analysis using corpora.
7. Sharipov, M., & Salaev, U. (2022). Uzbek affix finite state machine for stemming. *arXiv preprint arXiv:2205.10078*.
8. Madatov, K., Bekchanov, S., & Vičić, J. (2023). Uzbek text summarization based on TF-IDF. *arXiv preprint arXiv:2303.00461*.
9. Niyazmetova, Kumushoy & Kuriyozov, Elmurod. (2023). Oʻzbek tilidagi matnlarni sentiment tahlil qilish.

MATN TOZALASH TIZIMI AYRIM KATEGORIYALARINING TAVSIFI

Baxodirov Sanjarbek Rahmatali o'g'li

sanjarbahodirov9901@gmail.com,

ToshDO'TAU magistranti

Annotatsiya. Matnni tozalash tabiiy tilni qayta ishlash (NLP) vazifalarida muhim dastlabki ishlov berish bosqichi sanalib, bunda matn ma'lumotlarining sifati va izchilligini ta'minlash uchun maxsus belgilar, kulgichlar (emojis), raqamlar, tinish belgilari va qo'shimcha bo'shliqlar, elektron pochta manzillari, HTML teglari hamda URL (Uniform Resource Locator) manzillari olib tashlanadi. Matnni tozalash tizimlari emojilarni qayta ishlash va bo'shliqlarni olib tashlash kabi kategoriyalar bo'yicha tahlil qilish uchun matn ma'lumotlarini oldindan qayta ishlashda muhim rol o'ynaydi. Emojilar grafik belgilar bo'lib, matnni qayta ishlashni murakkablashtirishi mumkin va matnni tozalash tizimlari ko'pincha ularni o'zgartirish yoki olib tashlash vositalarini o'z ichiga oladi. Bo'shliqlar matnning o'qilishi va izchilligiga ta'sir qilishi mumkin, shuning uchun tizimlar ortiqcha bo'shliqlarni olib tashlash funksiyalarini o'z ichiga oladi. Ushbu kategoriyalar va boshqa dastlabki ishlov berish vazifalarini hal qilish orqali matnni tozalash tizimlari matn ma'lumotlarining yaxshi formatlanganligini va tahlil qilish uchun tayyor bo'lishini ta'minlaydi. Ushbu maqola matnni tozalash jarayonida ushbu elementlarni olib tashlash uchun ishlatiladigan usullar va texnikalariga qaratilgan.

Abstract. Text cleaning is a crucial preprocessing step in natural language processing (NLP) tasks. It involves removing special characters, emojis, numbers, punctuation, extra spaces, email addresses, HTML tags, and URLs (Uniform Resource Locators). Text cleaning systems are essential for preprocessing text data for categorization analysis, such as emoji processing and whitespace removal. Emojis are graphic symbols that can complicate text processing. Text sanitization systems often include tools to modify or remove them. Excess whitespace can affect the readability and consistency of text, so systems include functions to remove it. Text cleaning systems handle these categories and other preprocessing tasks to ensure that text data is well formatted and ready for analysis. This article focuses on the methods and techniques used to remove unnecessary elements during the text cleaning process.

Аннотация. Очистка текста — важнейший этап предварительной обработки в задачах обработки естественного языка (NLP). Он включает в себя удаление специальных символов, смайлик, цифр, знаков препинания, дополнительных пробелов, адресов электронной почты, HTML-тегов и URL-адресов (унифицированных указателей ресурсов). Системы очистки текста необходимы для предварительной обработки текстовых данных для анализа категоризации, например, обработки эмодзи и удаления пробелов. Эмодзи —

это графические символы, которые могут усложнить обработку текста. Системы очистки текста часто включают в себя инструменты для их изменения или удаления. Лишние пробелы могут повлиять на читаемость и согласованность текста, поэтому в системы включены функции для их удаления. Системы очистки текста обрабатывают эти категории и другие задачи предварительной обработки, чтобы гарантировать, что текстовые данные хорошо отформатированы и готовы к анализу. В этой статье основное внимание уделяется методам и приемам, используемым для удаления ненужных элементов в процессе очистки текста.

Kalit so'zlar: *matnni tozalash tizimi, matn ma'lumotlarining aniqligi, ma'lumotlarning qayta ishlash samaradorligi, regular ifoda, emojilar, bo'shliqlar.*

KIRISH

Matnni tozalash tizimi matn ma'lumotlaridan keraksiz va ahamiyatsiz ma'lumotlarni olib tashlash orqali ma'lumotlarning aniqligi va ishonchliligini ta'minlashda muhim vosita hisoblanadi. So'nggi yillarda matnlarni tozalash tizimlariga talab sezilarli darajada oshdi, chunki korxonalar va tashkilotlar ijtimoiy tarmoqlar, mijozlar sharhlari va yangiliklar kabi turli manbalardan yaratilgan katta hajmdagi ma'lumotlar bilan ishlashiga to'g'ri kelmoqda.

Matn tozalash tizimining umumiy tavsifi tozalanmagan matn ma'lumotlarini toza va tartibli formatga aylantirishga qaratilgan turli xil texnika va jarayonlarni o'z ichiga oladi. Matn tozalash jarayonidagi ba'zi bir umumiy qadamlar maxsus belgilarni, tinish belgilarini va nomuhim so'zlarni olib tashlashni, shuningdek, so'zlarni standartlashtirish va ma'lumotlarning o'lchamini kamaytirish uchun stemming va lemmatizatsiyani o'z ichiga oladi. Bundan tashqari, matn ma'lumotlarining sifatini yanada oshirish uchun ko'pincha imlo tekshiruv, tokenizatsiya va nutqning bir qismini teglash qo'llanadi.

Adabiyotlar sharhi

Mashinani o'rganish (machine learning) va data science (ma'lumotlar bilan ishlash) dagi yutuqlar matnni tozalash tizimlarining rivojlanishini yanada kuchaytirdi. Chunki sifatli matn ma'lumotlariga bo'lgan ehtiyoj NLP modellarini o'qitish va sinovdan o'tkazish uchun tobora muhim ahamiyat kasb etdi. Tadqiqotchilar matnni tozalash uchun turli xil yondashuvlarni o'rganishni boshladilar, masalan, nomuhim so'zlarni olib tashlash, HTML teglarini, URL manzillarini olib tashlash, imlo xatolarini tuzatish, formatlarni standartlashtirish, maxsus belgilar bilan ishlash va boshqalar.

Jahon kompyuter tilshunosligi NLP sohasida matnni tozalash borasida bir qator olimlar ilmiy izlanishlar olib borgan. B.Roepke, G.Spiller, J.A.Yeung, N.V.Otten, Pankaj, Vidvatek, N.Pentapalli, D.Silval, R.Vikkery, R.Chandra, S.Kim, K.Pikes, I.A.Xalid, K.Rastogi, I.Roldos kabi olimlarning tadqiqotlarini ana shunday ishlar sifatida qayd etish mumkin. D.Jurgens va I.Augenstein matnni tozalash usullarini o'rgangan va matn sifatini yaxshilash va uni NLP vazifalari uchun ko'proq

moslashtirish uchun dasturiy vositalarni ishlab chiqqan. J.Xu va Ch.Miao hissiyotlarni tahlil qilish va ma'lumotlar olish kabi turli ilovalar uchun matnni tozalash dasturi ustida ishlar olib borishgan.

Tsyurix universiteti qoshidagi Hisoblash tilshunosligi universiteti va Qatar Hisoblash tadqiqot instituti kabi tadqiqot muassasalari ham matn tozalash va dastlabki ishlov berish texnikasi bo'yicha tadqiqotlar olib bordilar. Bundan tashqari, Allen sun'iy intellekt (AI) instituti AllenNLP kutubxonasi kabi matnni tozalash va qayta ishlash uchun vositalar va resurslarni ishlab chiqdi.

Bular matnni tozalash dasturlari sohasida o'z hissalarini qo'shgan tadqiqotchilarning bir nechta misollari. Ko'plab boshqa tadqiqotchilar va tashkilotlar ham matn sifatini va turli ilovalar uchun qulaylikni oshirish uchun ushbu sohada faol ishlashmoqda.

Matnni tozalash tizimlari nomuhim so'zlar, maxsus belgilar, kulgichlar(emojis), raqamlar, tinish belgilari va qo'shimcha bo'shliqlar, elektron pochta manzillari, HTML teglari hamda URL(Uniform Resource Locator) kabi elementlarni olib tashlash orqali matn ma'lumotlarining sifatini oshiradi. Ushbu tizimlar tabiiy tilni qayta ishlash, ma'lumotlarni qidirish, **data mining** dasturlarida keng qo'llanadi. Quyida matnni tozalash turlarining har bir turiga to'xtalib o'tiladi.

1.1 Kulgichlar (emojis)ni olib tashlash: Emojilar – yozma muloqotda his-tuyg'ularni yoki g'oyalarni ifodalash uchun ishlatiladigan kichik raqamli tasvirlar yoki piktogrammalar. Emojilar matnga chuqur emotsiyalarni qo'shishi mumkin bo'lsa-da, ularning biz xohlamagan va noo'rin holatlari ham mavjud.

Matndan kulgichlarni (emojis) olib tashlash uchun **Regular expression**dan foydalaniladi, bu orqali barcha emoji belgilarni bo'sh qator bilan qidirish yoki almashtirish mumkin. Shu bilan bir qatorda, emojilarni matndan avtomatik ravishda olib tashlash mumkin bo'lgan maxsus vositalar va dasturlar mavjud. Matndan bunday ortiqcha kulgichlar (emojis)ni olib tashlashni Python dasturlash tili asosida ko'rib chiqamiz [Yeung, 2022]

```
import re
text = u"O'shanda uchinchi sinfda o'qirdim 😊 😊. Biz oilamiz bilan boshqa
qishloqqa ko'chadigan bo'ldik 😞 😞."
print(text) # with emoji
def remove_emojis(text):
    regex_pattern = re.compile(pattern = "["
        u"\U0001F600-\U0001F64F"
        u"\U0001F300-\U0001F5FF"
        u"\U0001F680-\U0001F6FF"
        u"\U0001F1E0-\U0001F1FF"
        u"\U00002702-\U000027B0"
```

```
u"\U000024C2-\U0001F251"  
    "+", flags = re.UNICODE)  
return regex_pattern.sub(r", text)  
print(remove_emojis(text))
```

#Output

#with emoji: O'shanda uchinchi sinfdan o'qirdim 😊 😊. Biz oilamiz bilan boshqa qishloqqa ko'chadigan bo'ldik 😊 😊.

#remove emoji: O'shanda uchinchi sinfdan o'qirdim. Biz oilamiz bilan boshqa qishloqqa ko'chadigan bo'ldik.

Matndan kulgichlar (emojis)ni olib tashlashni bir qancha sabablari mavjud:

1. *Aniqlik.* Emojilar ba'zan noto'g'ri talqin qilinishi yoki tushunarsiz bo'lishi mumkin, ayniqsa professional yoki rasmiy muloqotda. Emojilarni olib tashlash ma'lumotning aniq tushunilishini ta'minlashga yordam beradi.

2. *Foydalanish imkoniyati.* Ko'zida nuqsoni bo'lgan ba'zi odamlar emojilarni talqin qilishda qiynalishi mumkin, natijada ma'lumotlar tushunilishida qiyinchiliklar yuzaga keladi.

3. *Formatlash.* Emojilar har doim ham turli platformalar yoki qurilmalarda to'g'ri ko'rsatilmaligi mumkin, bu esa tashqi ko'rinishdagi nomuvofiqlikka olib keladi. Emojilarni olib tashlash matnning izchil va professional formatlanishini saqlashga yordam beradi.

4. *Tillar o'rtasidagi nomutanosibliklar.* Kulgichlar har doim ham universal ma'noga ega bo'lavermaydi va ular turli madaniyatlarga mansub shaxslar tomonidan noto'g'ri talqin qilinishi mumkin. Emojilarni olib tashlash til to'siqlari tufayli yuzaga keladigan tushunmovchiliklardan qochishga yordam beradi.

Umuman olganda, emojilarni matndan olib tashlash muloqotning konteksti va auditoriyasiga bog'liq. Emojilar ba'zi holatda muloqotni emotsionalligini oshirishi mumkin bo'lsa-da, boshqa hollarda ularni olib tashlash ma'qul bo'ladi.

1.2 Qo'shimcha bo'shliqlarni olib tashlash. Bo'shliq matn ichidagi so'zlar va belgilarni ajratish uchun ishlatiladigan belgini (masalan, yangi qator, yorliq (tab) va ortiqcha bo'shliqlar) anglatadi. Ba'zi hollarda, bir nechta sabablarga ko'ra matndan bo'shliqlarni olib tashlashga to'g'ri kelishi mumkin:

1. *O'qishni osonlashtirish uchun.* Ortiqcha bo'shliqlarni olib tashlash, matnni o'qishni va tushunishni osonlashtiradi.

2. *Joyni tejash.* Cheklangan joy bilan ishlanganda, masalan, dasturlash yoki veb-sayt dizaynida, keraksiz bo'shliqlarni olib tashlash fayl hajmini kamaytirish va ishlashni optimallashtirishga yordam beradi.

3. *Matnni normallashtirish.* Matnni qayta ishlash ilovalarida bo'shliqni olib tashlash matn formatini standartlashtirishga va izchillikni yaxshilashga yordam beradi.

Matndan ortiqcha bo'shliqlarning olib tashlashning turli xil usullari mavjud:

1. *Trim funksiyasi yordamida*. Ko'pgina dasturlash tillari va matn muharrirlarida “Trim” funksiyasi mavjud bo'lib, u satrdan oldingi va keyingi bo'shliqlarni olib tashlash uchun ishlatilishi mumkin.

2. *Regular ifoda*. “Regular expressions” (re) matn ichidagi bo'shliqlarni qidirish va almashtirish uchun ishlatilishi mumkin. Masalan, “\s+” bo'sh joy belgilarining istalgan ketma-ketligini moslashtirish uchun ishlatilishi mumkin.

3. *Qo'lda tahrirlash (manual editing)*. Ba'zi hollarda keraksiz bo'shliqlarni olib tashlash uchun matnni qo'lda tahrirlash, ayniqsa, kichik hajmdagi matnlar uchun eng samarali yechim bo'lishi mumkin.

Bundan tashqari, matndagi bo'shliqlarni olib tashlash uchun foydalanilayotgan dasturlash tiliga qarab turli usullardan foydalanish tavsiya etiladi. Ba'zi keng tarqalgan usullar qatoriga Pythonda **strip**, **rstrip** yoki **lstrip** usullari, Javada esa, **trim** usuli yoki JavaScriptdagi **replace** usuli kiradi.

Python dasturlash tili yordami bilan turli xil usullarda ortiqcha bo'shliqlarni olib tashlashni ko'rib chiqamiz[Zia, 2024]

1-usul

```
import re
s = " O'zbek tiliga davlat tili maqomini berilishi, o'zbek
tiliga bo'lgan e'tiborni yanada yuksaltirdi. "
s = re.sub(" +", " ", s)
print(s)

#Output: O'zbek tiliga davlat tili maqomini berilishi, o'zbek tiliga bo'lgan
e'tiborni yanada yuksaltirdi.
```

2-usul

```
s = " O'zbek tiliga davlat tili maqomini berilishi, o'zbek
tiliga bo'lgan e'tiborni yanada yuksaltirdi"
s = s.replace(" ", "")
print(s)

#Output:
O'zbektiligadavlattilimaqominiberilishi,o'zbektiligabo'lgane'tiborniyanadayuksalti
rdi
```

2- usulda qo'llangan **replace()** usuli 1-usulga qaraganda samaradorligi past. Ya'ni bu usulda qo'shtirnoq orasiga qanday buyruq kiritgan bo'lsak faqat shu buyruq bilan ishlaydi.

3-usul

```
s = "    O'zbek tiliga davlat    tili maqomini berilishi, o'zbek  
tiliga  bo'lgan e'tiborni yanada yuksaltirdi  "  
s = s.lstrip()  
print(s)  
  
#Output: O'zbek tiliga davlat    tili maqomini berilishi, o'zbek  
tiliga  bo'lgan e'tiborni yanada yuksaltirdi
```

Bu usulda so'ralgan tomondan bo'shliqlar olib tashlanadi, ya'ni **lstrip()** – **leftstrip(chap)** va **rstrip()** – **rightstrip(o'ng)** tomondagi bo'shliqlarni tozalaydi. Bu usulda gap o'rtasidagi bo'shliqlar tozalanmaydi.

Umuman olganda, matndan bo'shliqlarni olib tashlash o'qishni yaxshilash, joyni tejash va turli ilovalar uchun matn formatini normallashtirishga yordam beradi.

XULOSA

Matnni tozalash tizimlarini bir nechta toifalarga bo'lish mumkin, jumladan emojilarni qayta ishlash va matndan bo'shliqlarni olib tashlash. Ushbu tizimlar matnli ma'lumotlarni yanada izchil, aniq va oson tahlil qilish uchun qayta ishlashga mo'ljallangan.

Matnni tozalash tizimlarida keng tarqalgan vazifalardan biri kulgichlar (emojis) bilan ishlashdir. Emojilar - bu matnli xabarlar va ijtimoiy media postlarida his-tuyg'ular, g'oyalar va tushunchalarni ifodalash uchun ishlatiladigan grafik belgilar. Ular matnga ortiqcha ma'no qo'shishlari mumkin, lekin ular matnni qayta ishlash algoritmlari uchun qiyinchiliklarni ham yaratishi mumkin. Matnni tozalash tizimlari ko'pincha emojilarni qayta ishlash vositalarini, masalan, ularni matn ekvivalentlariga aylantirish yoki ularni butunlay olib tashlashni o'z ichiga oladi.

Matnni tozalashda yana bir muhim vazifa matndan oq bo'shliqlarni olib tashlashdir. Bo'shliqlar, yorliqlar va qatorlar orasidagi bo'shliqlar matn ma'lumotlarining o'qilishi va izchilligiga ta'sir qilishi mumkin. Matnni tozalash tizimlari odatda matndan keraksiz oq bo'shliqlarni olib tashlash vositalarini o'z ichiga oladi, bu esa matnning yaxshi formatlanganligini va tahlil qilish oson bo'lishini ta'minlaydi.

Emojilarni qayta ishlash va bo'sh joylarni olib tashlashdan tashqari, matnni tozalash tizimlari matnli ma'lumotlarni oldindan qayta ishlash uchun boshqa xususiyatlarni ham o'z ichiga olishi mumkin. Masalan, bu tizimlar imlo va grammatik xatolarni tuzatish, maxsus belgilarni, HTML teglari, URL manzillarini olib tashlash va matn formatini normallashtirish vositalarini o'z ichiga olishi mumkin. Matn ma'lumotlarini tozalash va qayta ishlash orqali ushbu tizimlar matnni tahlil qilish algoritmlarining aniqligi va samaradorligini oshirishi mumkin.

Umuman olganda, matnni tozalash tizimlari matn ma'lumotlarini tahlil qilish uchun tayyorlashda hal qiluvchi rol o'ynaydi. Emojilar bilan ishlash, bo'shliqlarni olib tashlash va boshqa qayta ishlash vazifalarini bajarish orqali ushbu tizimlar matn ma'lumotlarining toza, izchil va keyingi tahlilga tayyor bo'lishini ta'minlaydi. Matnni tozalash tizimlari tadqiqotchilar va ma'lumotlar bo'yicha olimlardan tortib sotuvchilar va ijtimoiy media tahlilchilarigacha bo'lgan matnli ma'lumotlar bilan ishlaydigan har bir kishi uchun muhim vositadir.

Foydalanilgan adabiyotlar:

1. Chandra R. Text Cleaning in Python: Effective Data Cleaning Tutorial, 2023. [Text Cleaning in Python: Effective Data Cleaning Tutorial – Kanaries](#)
2. Otten N.V. Top 20 Essential Text Cleaning Techniques [Practical How to Guide in Python] / Data Science, 2023, Sep 18. <https://spotintelligence.com/2023/09/18/top-20-essential-text-cleaning-techniques-practical-how-to-guide-in-python>
3. Pankaj. How to Remove Spaces from a String in Python / DigitalOcean, 2022, December 12. How To Remove Spaces from a String In Python | DigitalOcean
4. Roepke B. How to Clean Text Like a Boss for NLP in Python / Data Knows All, 2024, Mart 3. <https://dataknowsall.com/blog/textcleaning>
5. Williams A. Remove emails from Text in Python (methods+examples) / PYtutorial, 2023, Sep 26. <https://pytutorial.com/remove-emails-from-text-in-python-methods-examples>
6. Yeung J.A. How to remove emojis from string in Python /The Web Dev, 2022, April 16. <https://thewebdev.info/2022/04/16/how-to-remove-emojis-from-a-string-in-python>
7. Zia A. How to Remove Whitespaces in Python Strings? /jQuery-AZ /2024. <https://www.jquery-az.com/how-to-remove-whitespaces-in-strings-of-python>

MATNLARNI TASNIFLASH UCHUN BERT MODEL VA TENSORFLOW KUTUBXONASINING O'RNI

Berdiyev Jahongir Botir o'g'li

berdiyevjahongir94@gmail.com,

ToshDO'TAU 1-bosqich magistranti

Annotatsiya. Tabiiy tilni qayta ishlash hozirgi kun ilm-fanining muhim masalalaridan biriga aylanib bormoqda. Tabiiy tilni qayta ishlashning eng muhim vazifalaridan biri matnlarni tasniflashdir. Uning maqsadi kiritiladigan matn ma'lumotlarini avvaldan belgilab olingan sinflarga ajratishdan iboratdir. Buning uchun esa BERT modeli va TensorFlow kutubxonasidan foydalanish juda yaxshi samara beradi. Mazkur maqolada matnni tasniflash nima ekanligi, uning tabiiy tilni qayta ishlashdagi ahamiyati, matnni tasniflash borasida jahon tajribasi va erishilgan yutuqlar kabi masalalar yoritilgan. Bundan tashqari BERT modelining yaratilish tarixi va unga bo'lgan ehtiyoj, uning boshqa modellardan farqi va imkoniyatlari hamda Tensorflow kutubxonasini o'rnatish va undan foydalanish, TensorFlowning o'ziga xos xususiyatlari, uning imkoniyatlari va boshqa kutubxonalardan farqi, sun'iy intellekt uchun ahamiyati haqida ham ma'lumotlar kiritilgan. Qolaversa, matnlarni tasniflashda BERT modeli va TensorFlow kutubxonasining o'rni hamda ulardan kutiladigan natijalar qay darajada ijobiy ekanligi kabi masalalarga e'tibor qaratilgan.

Abstract. Natural language processing is becoming one of the important issues of modern science. One of the most important tasks of natural language processing is text classification. Its purpose is to classify input text data into predefined classes. For this, the BERT model and the TensorFlow library are very effective. This article covers issues such as what text classification is, its importance in natural language processing, world experience and achievements in text classification. In addition, the history of the creation of the BERT model and the need for it, its differences and capabilities from other models, and the installation and use of the Tensorflow library, the characteristics of TensorFlow, its capabilities and differences from other libraries, and its importance for artificial intelligence data is also included. In addition, the role of the BERT model and the TensorFlow library in text classification and how positive the expected results are focused.

Аннотация. Обработка естественного языка становится одной из важных проблем современной науки. Одной из наиболее важных задач обработки естественного языка является классификация текста. Его цель — классифицировать входные текстовые данные по предопределенным классам. Для этого очень эффективны модель BERT и библиотека TensorFlow. В данной статье рассматриваются такие вопросы, как что такое классификация текста, ее значение в обработке естественного языка, мировой опыт и достижения в классификации текста. Кроме того, история создания модели BERT и

необходимость в ней, ее отличия и возможности от других моделей, а также установка и использование библиотеки TensorFlow, характеристики TensorFlow, ее возможности и отличия от других библиотек, а также также включена его важность для данных искусственного интеллекта. Кроме того, акцентируется внимание на роли модели BERT и библиотеки TensorFlow в классификации текста и на том, насколько позитивны ожидаемые результаты.

Kalit so'zlar: *Google, BERT, TensorFlow, sun'iy intellekt (AI), tuyg'ular tahlili, matnni tasniflash.*

Kirish

Zamonlar o'tgani sari insonning ongi ham rivojlanib boradi. Bu esa, o'z navbatida insonning bilimi, talab va ehtiyojlari ham ortib borishiga, odamlarni o'z ustida ishlashga, yanada ko'proq va qulayroq natijalarga erishishga sabab bo'ladi. Tabiiyki, buning natijasi avvalo, ilm-fan taraqqiyotida o'z aksini topadi. Oqibatda, ilm-fanning mavjud sohalarigina emas, avval mavjud bo'lmagan va taraqqiyot natijasida yuzaga kelgan yangi sohalar vujudga keladi. Xuddi shunday sohalaridan biri kompyuter texnologiyalari sohasidir. Vaqt o'tgani sari bu sohaga ehtiyoj tobora ortib bordi va bugungi kunga kelib kompyuter texnologiyalarisiz hozirgi kunni tasavvur qilish qiyin bo'lib qoldi. Uning kirib bormagan biror sohasi qolmadi. Tabiiy tillarni qayta ishlash ham shular jumlasidandir. Tabiiy tillarni qayta ishlash esa, ko'plab masalalarni hal qilishda muhim ahamiyat kasb etadi. Kompyuter texnologiyalariga ehtiyoj tobora ortib borar ekan, turli xil dasturlar, ilovalar, modellar va shu kabi ko'plab vositalar yaratilaveradi.

Adabiyotlar sharhi. Jahonda kompyuter lingvistikasi sohasida BERT modeli va TensorFlow kutubxonasidan foydalanib tabiiy tilni qayta ishlash, jumladan, matnni tasniflash borasida ko'plab tadqiqotlar olib borilgan. N.Baratalipour va boshqalar [Baratalipour va b., 2020], L.V. Motovskix [Мотовских, 2022], I.V.Selevanova [Селеванова, 2020], D.R.Baymurzina [Баймурзина, 2021], N.Kant va boshqalar [Kant va b., 2019], I.F.Putra va A.Purvarianti [Putra, Purvarianti], S.Prabhu va boshqalar [Prabhu va b., 2021], R.Buoy va boshqalar [Buoy va b., 2021], M.Özkan va G.Kar[Özkan, Kar, 2022], D.Samuel va boshqalar [Samuel va b., 2023], X.Fam [Fam, 2023], D.Noza va boshqalar [Noza va b., 2021], M.Baz va boshqalar [Baz va b., 2022] kabi olimlarni ana shunday tadqiqot olib borgan tadqiqotchilar sirasiga kiritish mumkin.

J.Devling va boshqalar tomonidan BERT modelining yaratilishi matnni tasniflashda yangi bosqich bo'ldi. Ular ushbu modelning boshqa til modellaridan ustun tomonlari, qulayliklari, tabiiy tilni qayta ishlashdagi ahamiyati va, umuman olganda, model haqida batafsil ma'lumot beradi. Khang Pham esa matnlarni tasniflash haqida batafsil to'xtalib, matnning tasnifi va BERT modeli haqida sodda va tushunarli fikr, matn tasnifi uchun BERT modeli qanday ishlashi hamda kutubxonalar, ulardan foydalanish haqida ma'lumotlar keltiradi. BERT yordamida matnlarni tasniflash bo'yicha qo'llanma ham berib, matnni tasniflash uchun ilk

qadamdan to natijagacha bo'lgan jarayonni ko'rsatib bergan. Rus tadqiqotchisi L.V.Motovskix internet yangiliklari matnlarini tasniflar ekan, “Matnni avtomatik tasniflash uchun foydalaniladigan usullardan biri TF-IDF matritsasidan foydalanishga asoslangan LSA yashirin semantik tahlil usuli hisoblanadi. Axborot texnologiyalarining rivojlanishi va tasniflash uchun yangi dasturiy yechimlarning paydo bo'lishi LSA yordamida olingan natijalarni, xususan, turli tasniflagichlardan foydalanish orqali sezilarli darajada yaxshilandi” [Мотовских, 2022], deya fikr bildiradi.

Jahonda tabiiy tilni qayta ishlash, jumladan, matnlarni tasniflash ishlari yaxshi o'rganilgan bo'lishiga qaramay, yurtimizda bu borada qilinishi kerak bo'lgan ishlar talaygina ekanligiga ko'zimiz tushadi. Hatto matnlarni tasniflash tizimi bo'yicha olib borilgan ishlarga ko'zimiz tushmaydi. Faqatgina E.Kuriyozov, U.Salayev, S.Matlatipov, G'.Matlatipovlar tomonidan “BERTbek” nomi modelning yaratilganligini ko'rishimiz mumkin. Ushbu tadqiqotchilar tomonidan “O'zbek tili uchun matn tasniflash ma'lumotlar to'plami va tahlili” nomli maqola ham chop etilgan. Mualliflarning fikriga ko'ra, Eng yaxshi samaradorlikka BERTbek modeli erishadi, bu esa transformatorga asoslangan BERT modeli bo'lib, O'zbekiston korpusida o'qitilgan [Kuriyozov va b., 2023]. Ushbu maqolada BERTbek modelining xususiyatlari, imkoniyatlari, erishilgan natijalari, umuman olganda, modelni atroflicha yoritib bergan.

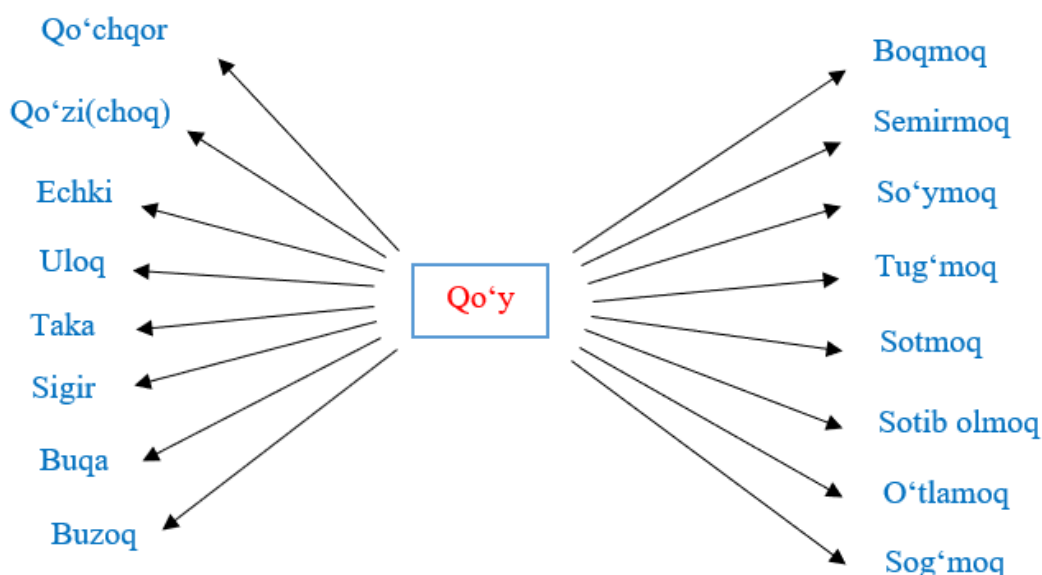
Asosiy qism

BERT modeliga bo'lgan ehtiyoj. Google qidiruv tizimi dunyodagi eng mashhur qidiruv tizimlaridan biri ekanligi barchaga ma'lum. Bu esa o'z foydalanuvchilari uchun Google jamoasi tomonidan yanada ko'proq imkoniyatlar va qulayliklar yaratishni taqozo qiladi. Foydalanuvchilari soni kundan kunga ortib borayotgani hamda axborotga va axborotni qayta ishlashga bo'lgan ehtiyoj borgan sari ortayotganini inobatga olgan holda Google jamoasi ushbu qidiruv tizimining rivojlanib borishi va undan foydalanuvchilarga yanada ko'proq qulaylik va imkoniyat yaratish maqsadida ko'plab yangi loyihalarga ehtiyoj sezdi.

Shunday loyihalardan biri BERT modelining yaratilganligi bo'ldi. BERT modeli 2018-yilda Google tadqiqotchilari jamoasi tomonidan ishlab chiqilgan til modellari oilasi hisoblanadi. BERT (inglizcha Bidirectional Encoder Representations from Transformers – transformatorlardan ikki tomonlama kodlovchi tasvirlar) – bu transformatorlar arxitekturasiga asoslangan til modeli bo'lib, keyinchalik ularni [tabiiy tilni qayta ishlashning](#) keng ko'lamli vazifalarida qo'llash maqsadida til ko'rinishlarini oldindan tayyorlash uchun mo'ljallangan. Masalan, BERT modeli yordamida matnlarni tasniflash, his-tuyg'ularni tahlil qilish, matn yaratish, tillarni aniqlash kabi ko'plab vazifalarini amalga oshirish mumkin. Matnlarni tasniflash tizimini ishlab chiqishda, yuqorida ta'kidlanganidek, eng samarali usul BERT modelidan foydalanishdir. Chunki BERT modelining boshqa til modellaridan farqi shundaki, bu model qidiruvda kontekstdagi so'zlarning butun to'plamidan foydalanadi. Ya'ni “Tilni taqdim etishning so'nggi modellaridan farqli

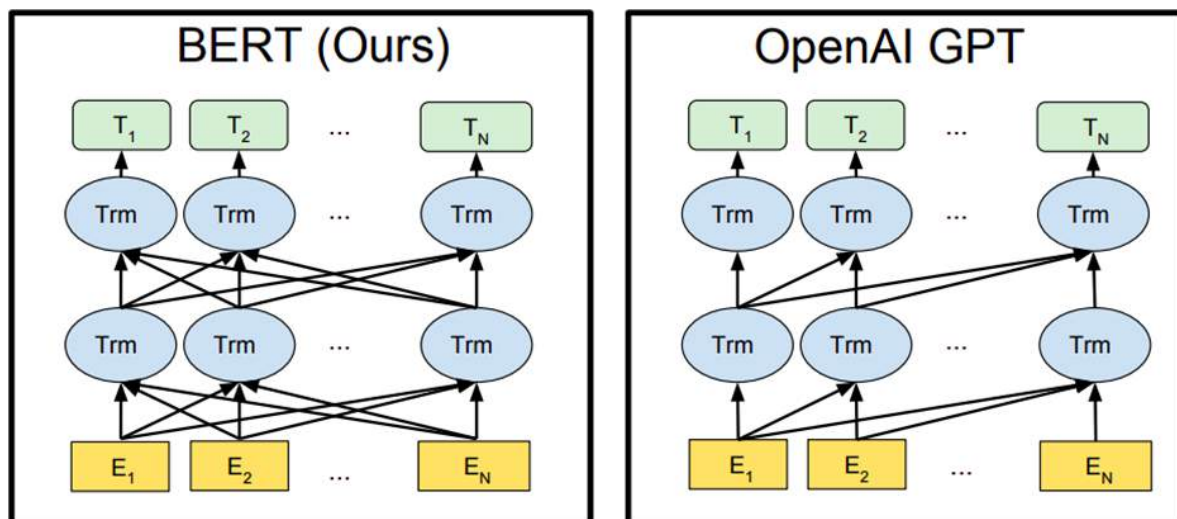
o'laroq, BERT barcha qatlamlarda chap va o'ng kontekstni birgalikda shartlash orqali yorliqsiz matndan chuqur ikki tomonlama tasvirlarni oldindan tayyorlash uchun mo'ljallangan". Shunday ekan, bu model orqali matnni tasniflash, his-tuyg'ularni tahlil qilish, spamni aniqlash, savollarga javob berish kabi vazifalarni yaratish uchun BERT modeli arxitekturasini ba'zi o'zgarishlar bilan osongina sozlab olish mumkin.

BERT modeli orqali tabiiy tilni qayta ishlashda, jumladan, matnni tasniflashda ikki tomonlama transformator kodlovchisidan foydalanadi. Bundan maqsad esa gaplardagi har bir so'zning unga bog'liq bo'lgan va u bilan birga keladigan barcha so'zlarni hisobga oladi. Masalan, BERT [qo'y] so'zi hayvon ekanligini tushunish uchun [qo'y og'zidan cho'p olmagan] jumlasidagi [qo'y] so'zi bilan bog'liq bo'lgan matndagi barcha so'zlarni tekshirib chiqadi. (1-rasmga qarang)



1-rasm. Qo'y so'zi bilan bog'liq bo'lgan so'zlar to'plami

BERTning bunday xususiyati kiritilgan matnni juda yaxshi tushunish va matn bilan ishlashda yuqori darajadagi aniqlikka erishish imkonini beradi. Qiyoslash uchun BERT modeli va GPT modellarini keltirish mumkin. BERT modeli matn bo'ylab yo'nalishsiz harakat qiladi. Bu esa matnni yaxshiroq tushunishga va aniqroq natijani olishga yordam beradi. GPT esa bir tomonlama chapdan o'ngga tomon harakat qiladi. Bu esa kerakli natijani ishlab chiqishda biroz noaniqlikni yuzaga keltirishi mumkin. (2-rasmga qarang) Ammo, shuni ham nazarda tutish kerakki, tabiiy tilni qayta ishlashda eng yaxshi va yuqori natijaga erishish uchun ma'lumotlar to'plamining hajmi qay darajada ko'p ekanligiga bog'liq bo'ladi.



2-rasm. BERT va GPTning ishlash sxemasi

BERT modeli oldindan tayyorlanadigan modeldir. Bu tabiiy tilni qayta ishlash uchun sozlanishdan oldin turli xil katta hajmdagi matnlarning o'qitilishini anglatadi. Uni ma'lum bir tilga o'qitgach, qandaydir aniq vazifa, masalan, his-tuyg'ular tahlili bo'yicha moslashtirish mumkin. Bu esa BERT modeliga unga berilgan vazifaning aniqligini oshirish imkonini beradi.

Matnlarni tasniflash jarayonini, birinchi navbatda, ma'lum bir dasturiy ta'minot kutubxonasini Python ish muhitiga import qilib olishdan boshlash kerak. Bunday dasturiy ta'minot kutubxonalarining bir qancha turlari mavjud. Dasturiy ta'minot kutubxonalarini mashinani o'qitish (ML), neyron tarmoq yaratish kabi muammolarni hal qilish uchun zarur bo'lgan barcha vositalarni taqdim etadi. Shunday kutubxonalardan biri TensorFlowdir (TensorFlow 2015-yilda Google Brain jamoasi tomonidan ishlab chiqilgan). U quyidagicha sozlanadi:

```
Pip install tensorflow
```

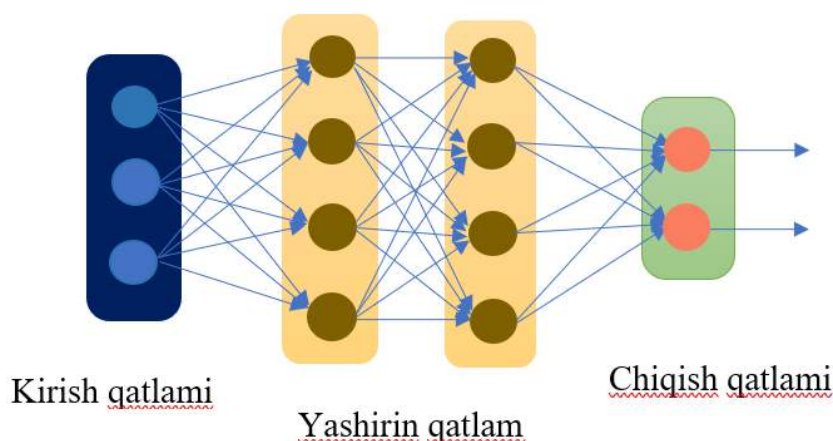
Bu buyruq orqali TensorFlow kutubxonasi o'rnatiladi.

```
Import tensorflow as tf
```

Bu buyruq orqali esa TensorFlow kutubxonasi ishga tushiriladi.

Yuqoridagi amallar bajarib bo'lingandan so'ng esa TensorFlowga tegishli barcha funksiyalarni va modullarni *tf* nomi orqali chaqirish va kerakli amallarni bajarish mumkin.

TensorFlow mashinani o'rganishning kichik to'plami bo'lmish sun'iy intellektga asoslangan chuqur o'rganish (Deep learning)dir. Chuqur o'rganishning asosiy qismlaridan esa biri neyron tarmoqdir. Neyron tarmoq quyidagi sxemaga o'xshaydi:



3-rasm. Neyron tarmoqning ishlash sxemasi

Bu yerda kirish qatlam neyron tarmog'ini qurish uchun katta hajmdagi turli xil ma'lumotlarni qabul qiladi. Bu ma'lumotlar matn, rasm, ovozi fayllar, video fayllar kabi turli xil ma'lumotlar bo'lishi mumkin:

– yashirin qatlamda kiritilgan ma'lumotlar qayta ishlanadi, ma'lumotlarni o'qish, ya'ni tanib olish jarayoni amalga oshadi va raqamlashtiriladi. O'quv jarayoni tugagandan so'ng ma'lumotlar chiqish qatlamiga o'tkaziladi.

– chiqish qatlami tegishli faollashtirish funksiyalarini qo'llash orqali bashorat qilingan chiqishni hosil qiladi. Masalan, matnni tasniflovchi ilovada ma'lum bir matnning qaysi guruhga tegishli ekanligi, tillarni aniqlovchi ilovada esa matnning qaysi tilga tegishli ekanligi haqidagi natijani beradi.(3-rasmga qarang)

Xulosa

Xulosa qilib aytadigan bo'lsak, avvalo, BERT modelining yaratilishi tabiiy tilni qayta ishlashda katta inqilob bo'ldi. Shuningdek, BERT modelining osongina sozlanishi istalgan vazifalarni bajarish uchun imkon beradi. TensorFlow kutubxonasidan foydalanish ham ma'lumotlarni qayta ishlashda eng samarali natijalarni olish imkonini beradi. Jumladan, matnlarni tasniflashda, his-tuyg'ularni tahlil qilishda va boshqa tabiiy tilni qayta ishlash bilan bog'liq jarayonlarda BERT modeli va TensorFlow dasturiy ta'minot kutubxonasidan foydalanish eng samarali natijalarni olishga yordam beradi.



Foydalanilgan adabiyotlar:

1. Baz M., Sami A. , Prottasha N., Kowsher Md, Murad S., Bairagi A., Masud M. (2022). Transfer Learning for Sentiment Analysis Using BERT Based Supervised Fine-Tuning. Sensors. 22. 10.3390/s22114157.
2. Kuriyozov E., Salaev U., Matlatipov S., Matlatipov G'. (2023). Text classification dataset and analysis for Uzbek language. 10.48550/arXiv.2302.14494.
3. Мотовских Л.В. Автоматизированная классификация новостных
4. интернет-текстов (на основе метода латентно-семантического анализа LSA) / Л.В. Мотовских. – Москва: Московский государственный лингвистический университет, 2022.
5. [https://en.wikipedia.org/wiki/BERT_\(language_model\)](https://en.wikipedia.org/wiki/BERT_(language_model))
6. <https://arxiv.org/abs/1810.04805>
7. <https://www.simplilearn.com/tutorials/deep-learning-tutorial/what-is-tensorflow>

NLPDA KOREFERENSIYA MASALASI

Jumaqulova Nasiba Qurbonovna

nasiba.qurbonovna@gmail.com

ToshDO‘TAU magistranti

Annotatsiya. Ushbu tezisda NLPda koreferensiya muammosini hal qilishning ahamiyati va qoidaga asoslangan usullar, Hobbs algoritmi hamda koreferensiya yechimidagi qiyinchiliklar haqida fikr yuritiladi.

Abstract. In this thesis, the importance of resolving coreference issues in NLP and the methods based on rules, the Hobbs algorithm, and the difficulties in coreference resolution are discussed.

Аннотация. В данной диссертации обсуждается важность разрешения кореференции в NLP, методы, основанные на правилах, алгоритм Хоббса и трудности в разрешении кореференции.

Kalit so‘zlar: *NLP, koreferensiya, qoidaga asoslangan usullar, Hobbs algoritmi.*

Tabiiy tilni qayta ishlash (NLP) – tilshunoslik va sun’iy intellekt kesishmasida joylashgan soha bo‘lib, u tabiiy tilning rasmiy axborot uslubini yaratish bilan shug‘ullanadi. NLPdagi asosiy muammolardan biri bu koreferensiya muammosini hal qilishdir, bu jarayon matndagi turli xil so‘z yoki ifodalar bir xil obyekt yoki tushunchaga ishora qilishini avtomatik aniq aniqlashni o‘z ichiga oladi. Bu vazifa matnni umumlashtirish va savollarga javob berishdan tortib chatbotlar va ma’lumot qidirish tizimlarigacha bo‘lgan turli ilovalar uchun juda muhimdir.

Azim issiq yoz kuni ko‘cha chetida qovun sotardi, u har o‘tgan piyodalarga uning mazzasini ta’riflashdan charchamasdi.

Azim → u

Qovun → uning

Koreferensiya muammosining ahamiyati

Matn ma’lumotlarining tuzilishi va semantikasini tushunishda asosiy rol o‘ynaydi. Bu olmoshlar, otlar va boshqa ishora qiluvchi ifodalarni ajratishda yordam beradi va natijada matnni umumiy tushunishni yaxshilaydi. Koreferensiya muammosini hal qilish muhim bo‘lgan bir nechta asosiy sabablar mavjud [1]:

1. **Matndagi koreferensiya:** Koreferensiya yechimini hal qilish matnning uyg‘unligi va o‘qilishini oshiradi. Bu o‘quvchilar yoki NLP tizimlari matnning turli qismlari orasidagi mos yozuvlar aloqalarini chalkashmasdan kuzatishini ta’minlaydi.

2. **Ma’lumot olish:** Ma’lumot olish vazifalarida, masalan, nomli obyektни tanib olish yoki hodisani ajratib olishda, koreferensiya yechimi matnning turli qismlaridan ma’lum bir obyekt yoki hodisa haqidagi ma’lumotlarni jamlashda yordam beradi.

3. **Matnni umumlashtirish:** Avtomatik matnni umumlashtirish uchun matndagi qaysi ifodalar yoki atoqli otlar obyekt bir obyektga tegishli ekanligini tushunish juda muhimdir. Bu qisqa, izchil xulosalar yaratish imkonini beradi.

4. **Savol-javob:** Savollarga javob berish tizimlari uchun koreferensiya yechimi zarur. Savol berilganda, asosiy ma'lumotlarni hal qilish savolga javob berish uchun zarur bo'lgan tegishli ma'lumotlarni aniqlashi mumkin.

5. **Chatbotlar va virtual yordamchilar:** Virtual yordamchilar, chatbotlar va mijozlarni qo'llab-quvvatlash tizimlari foydalanuvchilar bilan tabiiy tilda suhbatlar o'tkazyotganda koreferensiya yechimidan foyda ko'radi. Bu ularga kontekstni saqlab qolish va aniqroq javob berish imkonini beradi.

Koreferensiya muammosini hal qilish usullari

Koreferensiya muammosini hal qilishga turli usullar orqali erishish mumkin, ularning har biri o'zining afzalliklari va cheklovlariga ega:

1. **Qoidalarga asoslangan usullar** [2]: Qoidalarga asoslangan tizimlar asosiy murojaatlarni aniqlash uchun oldindan belgilangan lingvistik qoidalaridan foydalanadi. Ushbu qoidalar sintaktik tuzilma, shaxs-son moslashuvi va jins kabi omillarni hisobga oladi. Qoidaga asoslangan usullar koreferensiyaning hal etishni to'liq qamrab olmaydi. Tabiiy til ijtimoiylik xususiyatiga ega murakkab hodisa bo'lganligi uchun, qoidalarga bo'ysinmaydigan holatlar ham uchraydi.

Ushbu modellar qaysi eslatmalar bir obyektga tegishli ekanligini aniqlash uchun oldindan belgilangan qoidalar va usullar to'plamiga tayanadi. Qoidalarga asoslangan keng tarqalgan modellardan biri, Hobbsning olmoshlarni aniqlash algoritmidir. Biroq, boshqa qoidalarga asoslangan modellar mavjud bo'lib, ular ot ifodalari, tegishli otlar va boshqa turdagi ifodalar o'rtasidagi bog'liqlikni hal qilishga qaratilgan. Ushbu qoidaga asoslangan modellar odatda matndagi eslatmalarning sintaktik va semantik xususiyatlarini tahlil qilib, ularning asosiy munosabatlarini aniqlaydi. Ular asosiy qarorlarni qabul qilish uchun grammatik shaxs-son kelishuvi, miqdor, jins kelishuvi va semantik muvofiqlik kabi omillarni hisobga oladi.

Qoidalarga asoslangan modellar ma'lum kontekstlarda samarali bo'lgan bo'lsa-da, ular ham cheklovlarga ega. Ular noaniq yoki murakkab lingvistik konstruktsiyalarga duch kelishi va koreferensiya muammosini hal qilishga ta'sir qiluvchi lingvistik hodisalarni to'liqligicha qamrab ololmasligi mumkin. Garchi, qoidaga asoslangan modellar usulining cheklovlari mavjud bo'lsa-da, ular tabiiy tilni qayta ishlashda asosiy koreferensiya yechimlarini aniqlash usullarining muhim tarkibiy qismi bo'lib qolmoqda.

Hobbs algoritmi [3] gapdagi olmoshning ishorasini aniqlash uchun tabiiy tilni qayta ishlashda qo'llaniladigan usuldir. Bu olmosh va uning atrofidagi so'zlarning kontekstini tahlil qilish uchun bir qator qoidalar va metodlardan foydalanadi. Ushbu algoritmi, ayniqsa, so'zlar va ularning referentlari o'rtasidagi munosabatlarni tushunish va to'g'ri talqin qilish uchun juda muhim bo'lgan mashina tarjimai, ma'lumotni qidirish va matnni umumlashtirish kabi vazifalarda foydalidir.

Algoritm 1978-yilda Wiilliam A.Woods tomonidan taklif qilingan va keyinchalik 1979-yilda Jerry R.Hobbs tomonidan takomillashtirilgan. O'shandan beri u tabiiy tilni qayta ishlashda asosiy texnikaga aylandi va turli ilovalarda keng qo'llanildi.

Xobbs algoritmining asosiy xususiyatlaridan biri uning olmoshlarni yechishda ham sintaktik, ham semantik axborotni hisobga olish qobiliyatidir. Unda potentsial oldingi gapning grammatik roli, olmoshga yaqinligi, ikkalasi o'rtasidagi semantik moslik kabi omillar ko'rib chiqiladi.

Algoritmni qo'llash uchun odatda jumlaning sintaktik tuzilishini tahlil qilish va olmoshning potentsial oldingilarini aniqlash uchun tahlilchi ishlatiladi. Keyin algoritm potentsial oldingilarni referent bo'lish ehtimoli asosida tartiblash uchun o'z qoidalari va evristikasini qo'llaydi.

Xobbsning algoritmi tabiiy tilni qayta ishlash vazifalarida olmoshlarning aniqligini oshirishda samarali ekanligi isbotlangan. Biroq, bu cheklovlardan holi emas va tadqiqotchilar murakkab lingvistik kontekstlarda olmoshlarning yechimini yanada yaxshilash uchun yangi yondashuvlarni o'rganishda davom etmoqdalar.

Umuman olganda, Xobbsning olmoshlarni aniqlash algoritmi tabiiy tilni qayta ishlashda muhim vosita bo'lib, matndagi olmoshlarning murojaatlarini aniqlashning tizimli va ishonchli usulini ta'minlaydi. Uning doimiy rivojlanishi va takomillashtirilishi turli NLP ilovalaridagi yutuqlarga hissa qo'shishi mumkin.

2. Mashinali o'rganish yondashuvlari: Nazorat qilinadigan mashinani o'rganish modellari asosiy aloqalarni bashorat qilish uchun etiketli ma'lumotlarga o'rgatilgan bo'lishi mumkin. Ushbu modellar haqida ma'lumot berish uchun sintaktik va semantik o'xshashlik, so'zlarni joylashtirish va jins kelishuvi kabi xususiyatlardan foydalaniladi. Mashinali o'rganish algoritmlariga misollar orasida qarorlar to'plami, qo'llab-quvvatlovchi vektor mashinalari va logistik regressiya kiradi.

3. Chuqur o'rganish modellari: Chuqur o'rganish, xususan, neyron tarmoqlar, koreferensiya yechim rezolyutsiyasida inqilob qildi. Neyron tarmoqlar, shu jumladan takroriy neyron tarmoqlari (RNNs) va transformatorga asoslangan modellar matndagi murakkab naqshlar va bog'liqliklarni olishlari mumkin. BERT va uning variantlari kabi zamonaviy modellar koreferensiya yechimida ajoyib natijalarga erishdi.

4. Gibrid yondashuvlar: Ba'zi tizimlar ikkalasining kuchli tomonlaridan foydalanish uchun qoidalarga asoslangan va mashina o'rganish yoki chuqur o'rganish usullarini birlashtiradi. Misol uchun, qoidaga asoslangan usullardan dastlabki yadroviy klasterlarni yaratish uchun foydalanish mumkin, ular keyinchalik mashinani o'rganish yoki chuqur o'rganish modellari yordamida takomillashtiriladi.

Koreferensiya yechimidagi qiyinchiliklar

Koreferensiya yechimi quyidagi omillar tufayli qiyin NLP vazifasidir[1]:

1. Noaniqlik: Til tabiatan noaniq va bir nechta obyektlar yoki tushunchalar o'xshash atributlarga ega bo'lsa, asosiy ma'lumotlarni hal qilish qiyin bo'ladi.

2. Ichki ma'lumotnomalar: Matn ko'pincha ichki o'rnatilgan ko'rsatmalarni o'z ichiga oladi, bu yerda bir obyekt boshqa obyektga ishora qiladi, bu esa o'z navbatida boshqa obyektga ishora qiladi. Bunday ichki o'zaro munosabatlarni boshqarish ilg'or texnikani talab qiladi.

3. Uzoq masofaga bog'liqliklar: Bir nechta jumalarni yoki hatto paragraflarni qamrab oluvchi asosiy ma'lumotlarni hal qilish murakkab vazifa bo'lib, kengaytirilgan nutq orqali kontekstni tushunishni talab qiladi.

4. Jahon bilimlari: Ma'lumotnomalarni tushunish uchun ko'pincha dunyo bilimlari talab qilinadi, ular trening ma'lumotlarida mavjud bo'lmasligi mumkin. Masalan, “president” ma'lum bir shaxsga tegishli ekanligini bilish tashqi bilimlarni talab qiladi.

5. Pronominal va Null Anaphora: Ba'zi tillarda olmoshlar o'tkazib yuborilgan qo'zg'atuvchi hodisalar mavjud va koreferensiya yechimi aniq bo'lmasligi mumkin. Null anaforani aniqlash ayniqsa qiyin.

Xulosa

Xulosa qilib aytadigan bo'lsak, koreferensiya yechimi tabiiy tilni qayta ishlashning muhim tarkibiy qismi bo'lib, mashinalarga inson tilini yanada samaraliroq tushunish va talqin qilish imkonini beradi. U matnning uyg'unligini osonlashtiradi, ma'lumot olishda yordam beradi va NLP vazifalarining keng doirasi uchun ilovalarga ega. NLP tadqiqotlari va texnologiyasi rivojlanishda davom etar ekan, koreferensiya yechim o'lchamlari mashinalarga inson tilini aniqroq va ravonroq tushunish va yaratishga imkon beradigan hal qiluvchi jihat bo'lib qolmoqda.

Foydalanilgan adabiyotlar:

1. <https://medium.com/aimonks/coreference-resolution-in-natural-language-processing-nlp-5ba4f570bffe>
2. <https://medium.com/analytics-vidhya/hobbs-algorithm-pronoun-resolution-7620aa1af538>
3. <https://galhever.medium.com/a-review-to-co-reference-resolution-models-f44b4360a00>

KORPUS LINGVISTIKASI

O‘ZBEK TILIDAGI SO‘ZLARNI MORFOLOGIK TAHLIL QILUVCHI DASTUR

Qarshiyev Abduvali Berkinovich

Fizika-matematika fanlari nomzodi

TATU Samarqand filiali professori

Karimov Suyun Amirovich,

Filologiya fanlari doktori

SamDU professori

Tursunov Muhammadsolih Sa’din o‘g‘li,

muhammadsolih927@gmail.com

TATU Samarqand filiali katta o‘qituvchisi

Annotatsiya. Ushbu maqolada o‘zbek tilidagi so‘zlarni morfologik tahlil qiluvchi dasturdan foydalanish ko‘rsatib o‘tilgan. Matnlarni tahlil qiluvchi avtomatik analizatorni yaratishda dastlabki bosqich sifatida o‘zbek tilining elektron morfologik lug‘atini yaratish masalasi qo‘yiladi. Unga ko‘ra, har bir so‘zning qaysi so‘z turkumiga xosligining aniq statistikasi ko‘rsatiladi. Har bir leksemaning muayyan morfologik xususiyatlari jadvallarda aks etadi. Undan keyin leksemalar so‘z turkumlari bo‘yicha to‘plamlarga birlashadi. To‘plamlarga ajratishda leksemalarning umumiy semantik, qolaversa, formal jihatlarini inobatga olinadi. To‘plamlarga ajratilgan har bir so‘zga qo‘shilishi mumkin bo‘lgan qo‘shimchalar kombinatsiyasi aniqlanadi. Undan oldin barcha qo‘shimchalar funksional belgilariga ko‘ra guruhlariga bo‘linib, ularning o‘zbek tilidagi barcha mumkin bo‘lgan mantiqiy bazasi shakllantiriladi.

Abstract. This article shows the use of the program for the morphological analysis of words in the Uzbek language. The creation of an electronic morphological dictionary of the Uzbek language is the first step towards creating an automatic text analyzer. It displays precise statistics on which word group each word belongs to. The tables reflect individual morphological features of each lexeme. After this, lexemes are combined into sets of word groups. When dividing into sets, the general semantic and formal aspects of lexemes are taken into account. A combination of suffixes is determined that can be added to each word divided into sets. Previously, all suffixes are divided into groups according to their functional characteristics and all possible logical bases for them in the Uzbek language are formed.

Аннотация. В данной статье показано использование программы для морфологического анализа слов узбекского языка. Создание электронного морфологического словаря узбекского языка является первым шагом на пути создания автоматического анализатора текста. В соответствии с ним отображается точная статистика того, к какой группе слов принадлежит

каждое слово. В таблицах отражены отдельные морфологические особенности каждой лексемы. После этого лексемы объединяются в наборы по группам слов. При разделении на наборы учитываются общие семантические и формальные аспекты лексем. Определяется комбинация суффиксов, которую можно добавить к каждому слову, разделенному на множества. Предварительно все суффиксы разделяются на группы по их функциональным признакам и формируются все возможные логические основы их в узбекском языке.

Kalit so‘zlar: *o‘zbek tili, so‘z turkumi, morfologik tahlil, so‘zning asosi, ma’lumotlar bazasi, lug‘at, affikslar, dasturiy ta’minot.*

So‘z turkumini aniqlovchi morfoanalizator uchun dastlabki algoritmlar ikki tarkibli tuzilishga asoslangan. Birinchi bosqich har bir so‘zning turkumini aniqlash bo‘lsa, ikkinchi bosqich ajratilgan har bir so‘zning grammatik qoidalariga tuzatishlar kiritish hisoblangan [Mahlov, Piotrowski, 2009]. O‘zbek tili uchun ma’lumotlar bazasi juda muhim sanaladi. Shuni unutmash kerakki, morfoanaliz faqat shakl bilan cheklanib qolmasligi kerak. Kompyuter lingvistikasida chuqur tahlil qilish metodi semantik tahlil deb ataladi. Morfoanalizator quyidagi ikki bosqichdan iborat bo‘ladi:

- so‘zlarning asosini aniqlash;
- so‘z turkumini aniqlash;
- asosga ko‘ra har bir affiksni tahlil qilish. Bu jarayon chapdan o‘ngga yoki aksincha bo‘lishi mumkin;
- har bir qo‘shimchani turini belgilash.

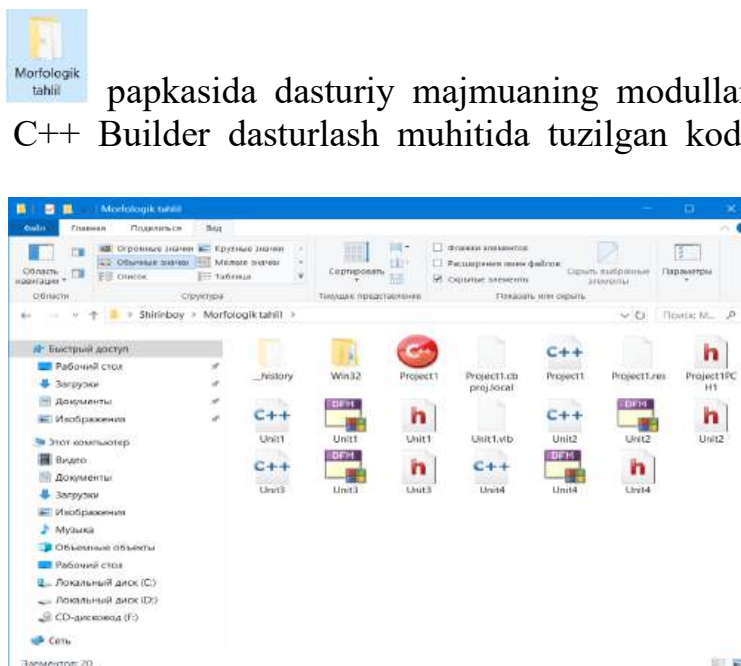
So‘zlarni ma’lumotlar bazasidan o‘qib olish uchun ularning simvollar ketma-ketligiga amal qilinadi. Belgilash o‘sha so‘z yaxlit birlik sifatida aniqlanmagunicha bu holat takrorlanaveradi. Morfologik tahlil uchun komponent tahlil metodidan ham foydalanish mumkin. Unga ko‘ra so‘zning uzunligi aniqlanadi. Shu tarkibdagi simvollar soniga qarab ma’lumotlar bazasidan qidiriladi. O‘sha so‘z bazadan aniqlanmagunicha tekshirish davom etaveradi.

O‘zbek tilida grammatik qo‘shimchalarning ketma-ketligi aniq belgilab qo‘yilgan, biroq qoidadan tashqari holatlar ham uchraydi. Chunonchi, ot va fe‘l so‘z turkumlarida quyidagicha:

OT	bola+jon+lar+im+dagi+lar+niki+mas+mi
FE’L	yoz+dir+tir+il+ma+yotgan+lar+ga+mi yoz+dir+tir+il+ma+di

Morfotaktik qoidalar morfemalarning o‘zaro kombinatsiyalarini boshqaradi. Shu jihatdan, matnlarni tahlil qilganda so‘zlarning asos va affikslarni aniqlashda morfotaktik tamoyillarga tayaniladi [Baqayev, 2021].

Ish stolning papkasida dasturiy majmuaning modullari va ularning Embarcadero XE3 C++ Builder dasturlash muhitida tuzilgan kodlari joylangan (1 – rasm):



1 – rasm. Dasturni ishga tushirish

1 – rasmdagi papkaning ichida **Win32** va **Debug** papkalariga ketma-ket kiriladi va dasturiy modulning **Project1** fayli ishga tushiriladi. Natijada ekranda quyidagi bosh oyna hosil bo'ladi (2 – rasm).



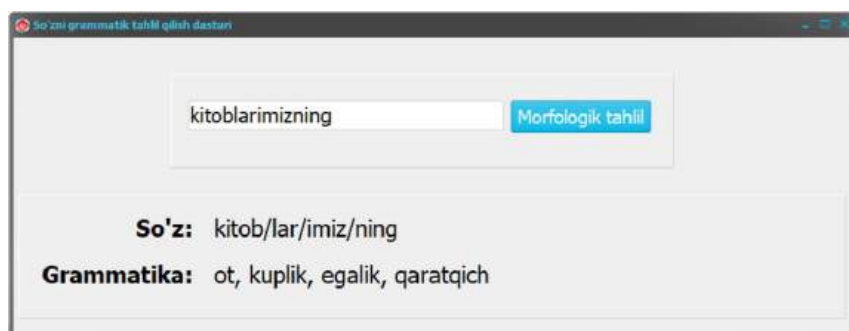
2 – rasm. Dasturning asosiy oynasi

2 – rasmdagi oynada dastur ikkita rejimda ishlashi ko'rinib turibdi. Bitta so'zni grammatik tahlil qilish va matnni grammatik tahlil qilish dasturlaridan iborat. "So'zni Grammatik tahlil qilish" tugmasi bosilsa 3 – oyna hosil bo'ladi. So'z kiritish maydoniga biror so'z kiritilsa, masalan "kitoblarimizning" so'zi kiritilsa, so'zdan birinchi navbatda asosni topish zarur bo'ladi. Bunda so'zdan asos sifatida "kitob" so'zini ajratib oladi. Sababi dasturning asoslar lug'atida kitob so'zi mavjud. Undan so'ng so'zning asosidan keyingi qo'shimchalarni aniqlaydi va qo'shimchalarga mos ravishda uning ma'nolarini foydalanuvchiga taqdim etadi (3-rasm).

Dastur "kitoblarimizning" so'zini quyidagicha tahlil qiladi:

Asos: kitob – ot;

Qo'shimcha: -lar – ko'plik, -imiz – egalik, -ning – qaratqich.



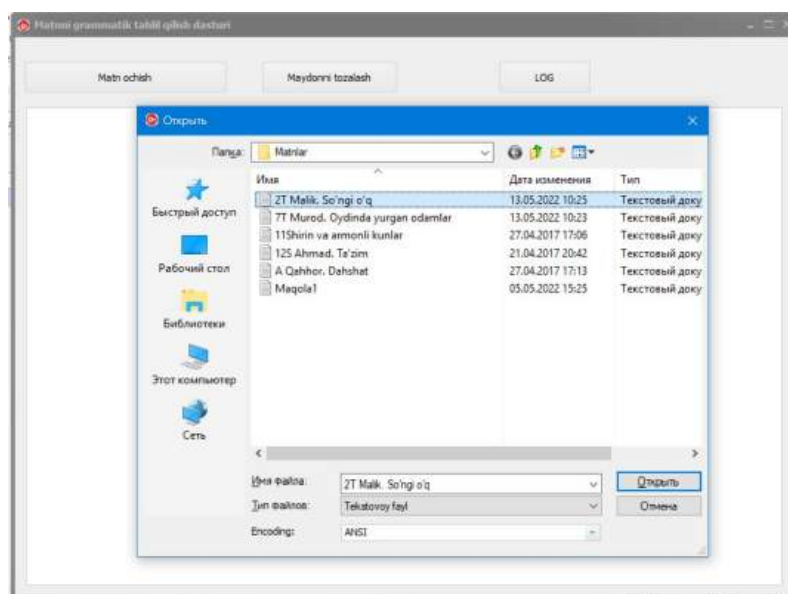
3 – rasm. So'zni grammatik tahlil qilish dasturi oynasi

Endi 3-rasmdagi dastur oynasiga “universitetning” degan so'zni kiritib, uni morfologik tahlil qilinsin. Bunda dastur asoslar lug'atida yuqoridagi so'zning asosi bo'lmaganligi uchun berilgan so'zni morfologik tahlil qilib bo'lmadi (4-rasm).



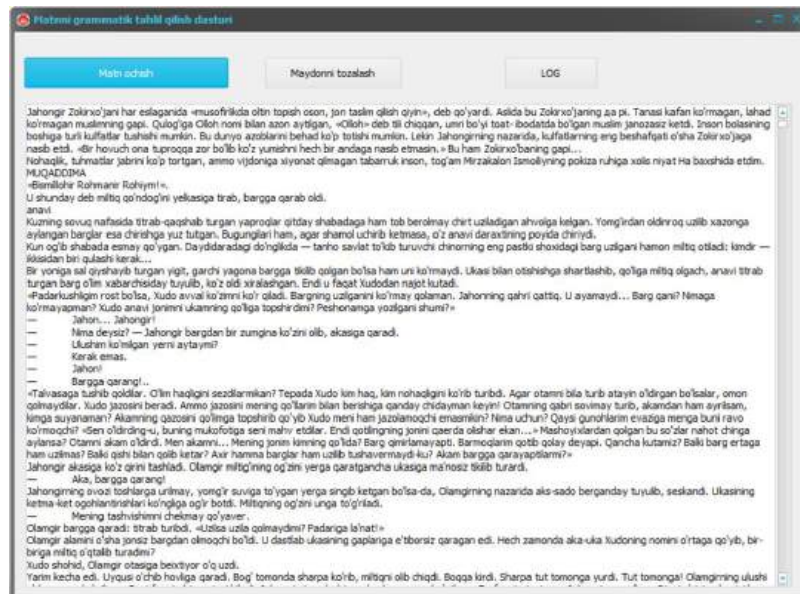
4 – rasm. So'zni Grammatik tahlil qilishda xatolik

Matnni grammatik tahlil qilish uchun 2-rasmdagi “Matnni grammatik tahlil qilish” tugmasi bosiladi. Keyingi qadamda “matnni ochish” tugmasi bosilib, dasturga matn yuklanadi (5-rasm).



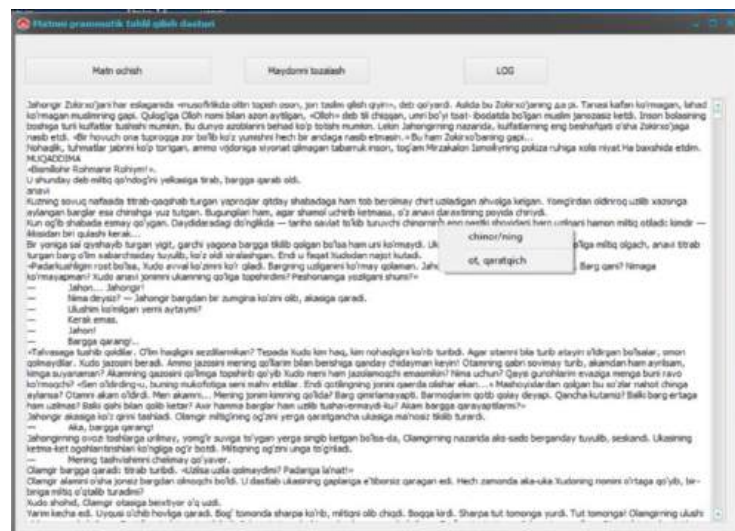
5 – rasm. Matnni dasturga yuklash

Matnni dasturga yuklagandan so‘ng matn dastur ishchi oynasida hosil bo‘ladi (6-rasm).

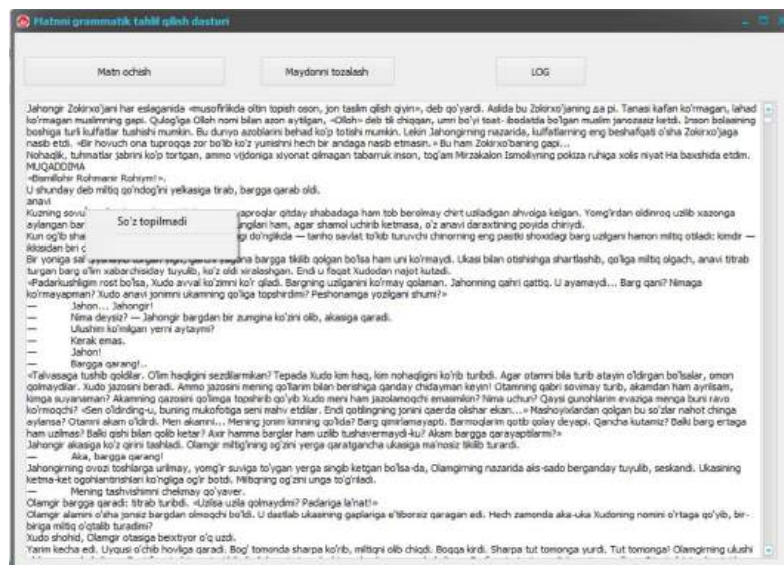


6 – rasm. Matnni dasturga yuklangan holati

Dasturga matn yuklangandan so‘ng, matnning ichida ixtiyoriy so‘zning, masalan “chinorning” so‘zini ustidan sichqonchaning o‘ng tugmasi bosilsa, tanlangan so‘zning grammatik xarakteristikalarini aniqlab beradi (7-rasm). Agar tanlangan so‘z bosilsa va bu so‘zning asosi lug‘atda bo‘lmasa, u holda “So‘z topilmadi” degan xabar paydo bo‘ladi. Masalan “sovuq” so‘zini morfologik tahlil qilinsin (8-rasm).



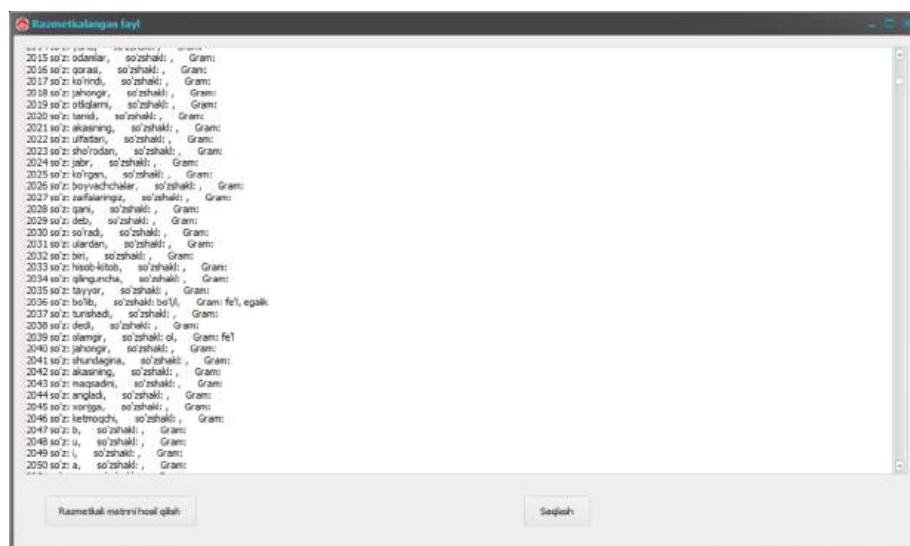
7 – rasm. Matndagi so‘zning morfologik tahlili



8 – rasm. Matndagi so‘zni morfologik tahlil qilib bo‘lmagan holat

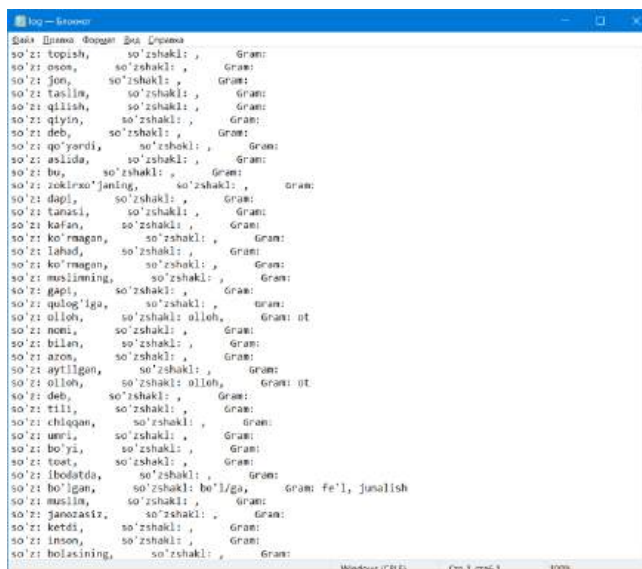
Dasturga yuklangan matnni to‘liq tahlil tahlil qilib, kompyuter xotirasiga saqlash lozim bo‘lsin. Bunda “LOG” tugmasi bosiladi va 9-rasm ochiladi.

9-rasmda “Razmetkali matnni hosil qilish” tugmasi bosiladi va tanlangan matndagi hamma so‘zlar razmetkalanib, foydalanuvchiga taqdim etiladi (10-rasm).



10 – rasm. Matnni razmetkalash

Matnni to‘liq razmetkalanganda ishchi oynaga tartib raqam, so‘z, so‘z shakl va grammatikasi chiqadi. So‘z shaklida so‘zning asosi va qo‘shimchalari ajratib ko‘rsatiladi. Hosil bo‘lgan razmetkali matnni kompyuter xotirasida saqlash uchun “saqlash” tugmasi bosiladi va “LOG.txt” nomli faylga saqlanadi.



11 – rasm. Faylda saqlangan razmetkali matnning ko‘rinishi

Matnlarni tahlil qiluvchi avtomatik analizatorni yaratishda dastlabki bosqich sifatida o‘zbek tilining elektron morfologik lug‘atini yaratish masalasi qo‘yiladi. Unga ko‘ra, har bir so‘zning qaysi so‘z turkumiga xosligining aniq statistikasi ko‘rsatiladi. Har bir leksemaning muayyan morfologik xususiyatlari jadvallarda aks etadi. Undan keyin leksemalar so‘z turkumlari bo‘yicha to‘plamlarga birlashadi. To‘plamlarga ajratishda leksemalarning umumiy semantik, qolaversa, formal jihatlarini inobatga olinadi. [Tursunov, Qarshiyev, 2019] To‘plamlarga ajratilgan har bir so‘zga qo‘shilishi mumkin bo‘lgan qo‘shimchalar kombinatsiyasi aniqlanadi. Undan oldin barcha qo‘shimchalar funksional belgilariga ko‘ra guruhlariga bo‘linib, ularning o‘zbek tilidagi barcha mumkin bo‘lgan mantiqiy bazasi shakllantiriladi.

Foydalanilgan adabiyotlar:

1. Mahlov C., Piotrowski M., JSLIM – “Computational Morphology in the Framework of the SLIM, Theory Language” / State of Art in Computational Morphology, Zurich, 2009, 15-p.
2. Baqayev I.I., “O‘zbek tili so‘z shakllarini morfologik tahlil qilish modellari va algoritmlari” / avtoreferat, Toshkent: TATU, 2021, 11-bet.
3. Tursunov M.S., Qarshiyev A.B., “O‘zbekcha matnlarni morfologik va leksik tahlil qilish algoritmlari va dasturlari” / International scientific conference “Actual problems of applied mathematics and information technologies”. – Toshkent, 14-15.11.2019, 290-291 betlar.

O'ZBEK TILI KORPUSINI SINTAKTIK TEGGLASH MASALASI

Elov Botir Boltayevich

Texnika fanlari bo'yicha falsafa doktori PhD, dotsent

elov@navoiy-uni.uz

ToshDO'TAU Kompyuter lingvistikasi va raqamli
texnologiyalar kafedrasini mudiri

Abdullayeva Oqila Xolmo'minovna

abdullayeva.oqila@navoiy-uni.uz

f.f.f.d., PhD, ToshDO'TAU doktoranti

Annotatsiya. Bugungi kunda zamonaviy NLP modellarini yaratish lingvistik bilimlarning akustika va prosodika, fonetika, orfografiya, morfologiya, leksikologiya, sintaksis, semantika, pragmatika va diskurs kabi spetsifikatsiyasini talab qiladi. NLP bilan bog'liq lingvistik bilimlar leksik, sintaktik, semantik va pragmatik xususiyatlarni o'z ichiga oladi. Ushbu maqolada sintaktik teglashning asosiy amallari haqida fikrlar keltiriladi. Shuningdek, sintaktik teglash (parsing)da qo'llanadigan shajara daraxti xususiyatlari tavsiflangan.

Abstract. Today, the creation of modern NLP models requires the specification of linguistic knowledge such as acoustics and prosody, phonetics, orthography, morphology, lexicology, syntax, semantics, pragmatics and discourse. Linguistic knowledge related to NLP includes lexical, syntactic, semantic and pragmatic features. This article provides an overview of the basic operations of syntactic tagging. Also, the features of the family tree used in syntactic tagging (parsing) are described.

Аннотация. Сегодня создание современных моделей НЛП требует уточнения таких лингвистических знаний, как акустика и просодия, фонетика, орфография, морфология, лексикология, синтаксис, семантика, прагматика и дискурс. Лингвистические знания, связанные с НЛП, включают лексические, синтаксические, семантические и прагматические особенности. В этой статье представлен обзор основных операций синтаксической разметки. Также описаны особенности генеалогического древа, используемые при синтаксической разметке (парсинге).

Kalit so'zlar: *til korpusi, lingvistik teglash, morfologik tahlil, tokenlash, lemmalash, POS teglash, sintaktik tahlil.*

Kirish

Korpus matnlarini lingvistik teglash dastlab lingvistik nazariyalarni yoki bugungi kunda ma'lum bo'lganidek, til korpuslarni ishlab chiqish va tahlil qilish uchun ma'lumot berish uchun amalga oshirildi. Matnlarni “qo'lda” teglash uchun ko'p vaqt va kuch talab qilinadi. Biroq so'nggi o'ttiz yil ichida hisoblash texnikalarning quvvati va katta hajmdagu ma'lumotlarni saqlash imkoniyati hosil

bo'lgach, katta hajmdagi til kopruslarini ishlab chiqish va avtomatik teglash sohasidagi bir qator yutuqlarga erishildi. Matnlardan iborat katta hajmli til korpuslari ustida turli lingvistik va amaliy tadqiqotlar olib borildi. Tabiiy tilni yangi texnologiyalar asosda tadqiq qilish va, eng muhimi, ishonchli statistik modellarni ishlab chiqish uchun lingvistik teglangan matn va til korpusiga asoslangan tabiiy tilni qayta ishlash (NLP) sohasiga muhim xizmat qiladi.

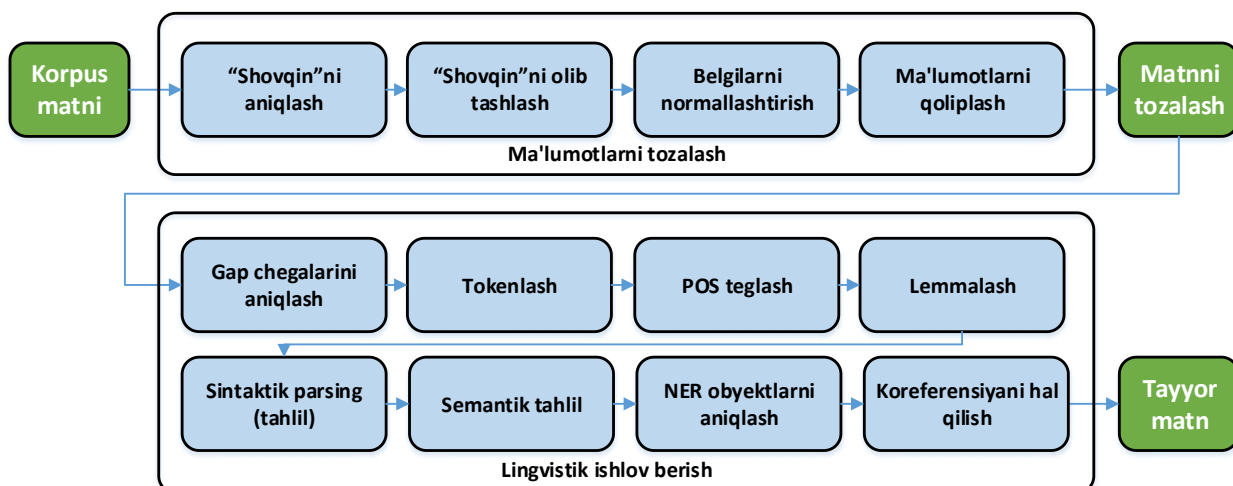
Bugungi kunga qadar o'zbek tili korpusi matnlarini teglash uchun standartlashtirilgan tokenizatsiya amallari, lemmalash metodlari, POS teglashdagi noaniqliklarni bartaraf qilish usullari ishlab chiqilmagan. Bugungi kunda korpus matnlarini lingvistik teglashni qo'lda yoki yarim avtomatik tarzda amalga oshirish mumkin.

NLP bilan bog'liq lingvistik bilimlarning jihatlari haqida turli xil qarashlar mavjud. Umuman olganda, NLP tadqiqotchilarining aksariyati NLP bilan bog'liq lingvistik bilimlar, hech bo'lmaganda, **leksik, sintaktik, semantik** va **pragmatik** xususiyatlarni o'z ichiga olishi kerak, deb hisoblashadi. Har bir xususiyat lingvistik ma'lumotni farqli yo'llar bilan ko'rsatadi. Masalan, sintaktik xususiyat ma'lum bir tildagi so'z yoki so'z birikmalarining gap hosil qilishini o'z ichiga oladi.

Bugungi kunda tilshunoslikda matnni teglashning zamonaviy usullari, shuningdek, turli xil teglash vazifalarini quyidagi – 1-jadval va 1-rasmda ko'rsatilganidek, o'xshash qatlamlar to'plamiga joylashtirish mumkin bo'lgan turli bosqichlarga ajratish mumkin.

1-jadval. Lingvistik teglash bosqichlari

№	Teglash usuli	Tavsif
1.	Gap chegalarini aniqlash (sentence boundaries)	matnni gaplarga ajratish
2.	Tokenlash (tokenization)	matnni so'zlarga ajratish
3.	Lemmalash	so'zshaklni lemma (lug'atdagi shakli)ga keltirish
4.	POS teglash (part-of-speech tagging)	so'zlarning turkumlari bilan aniqlash va belgilash
5.	Sintaktik parsing (tahlil)	gapdagi tarkibiy so'z birikmalarini tahlil qilish
6.	Semantik parsing (tahlil)	predikat-argument munosabatlarini belgilash (labeling predicate-argument relations)
7.	NER obyektlarini aniqlash (named entity recognition)	atoqli otlarni aniqlash va belgilash
8.	Koreferensiyani hal qilish (coreference resolution)	matndagi bir xil obyektlarga havolalarni bog'lash



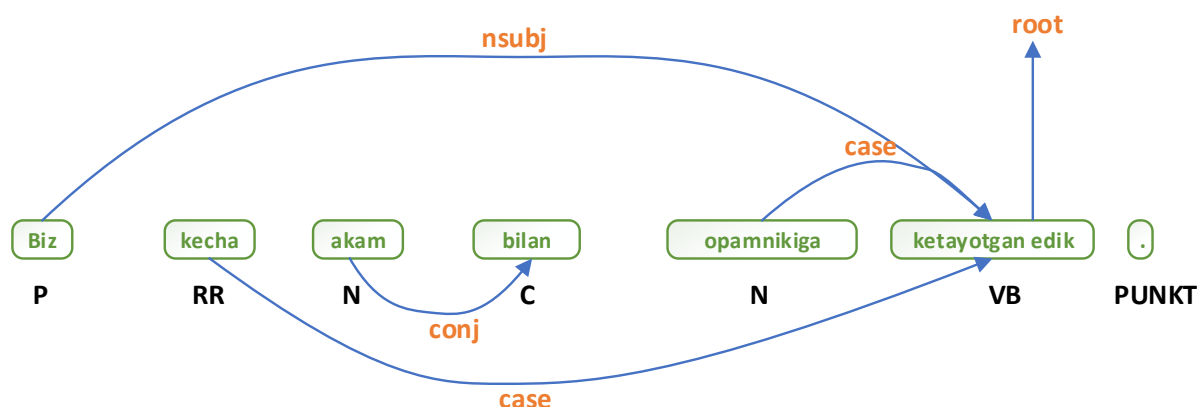
1-rasm. Korpus matnlarini teglashning zamonaviy usullari

Sintaktik parsing (tahlil)

Sintaktik tahlilning vazifasi gap qurilmalarini tahlil qilishdir. Ma'lum qurilmaning turi ishlatiladigan grammatik nazariyaga bog'liq. Masalan, **bog'liqlik grammatikasi (dependency grammars)** so'zlar orasidagi bog'liqlik munosabatlarini aniqlash uchun grafik qurilmalardan foydalansa, **so'z birikmalari qurilmasi grammatikasi (phrase structure grammars)** esa so'z birikmalari orasidagi tobelik munosabatini ko'rsatish uchun daraxt qurilmalaridan foydalanadi.

Gap yoki boshqa matn ma'lumotlari berilganda sintaktik tahlil qilish va ma'lumotlardan ajratish daraxtini yaratish vazifasi **parsing** qilish deyiladi. Bu daraxt grammatikada rasmiy qoidalarga asoslangan gapning sintaktik tuzilishini ajratib ko'rsatadi. So'z birikmasi grammatikasi asosida sintaktik tahlil qilishda gapdagi so'zlar o'zaro birlashtiriladi; uzunroq so'z birikmalari hamda gapning tarkibiy qismlarini tashkil qiladi. Sintaktik tahlil qilish uchun eng ko'p ishlatiladigan ikkita – **tobe munosabatlar tahlili (Dependency Parsing)** va **tarkibiy qismlar tahlili (Constituency Parsing)** qo'llanadi.

Turli xil so'z turkumlariga asoslangan so'z birikmalari turlicha bo'lib, otli birikma (NP) otga, fe'lli so'z birikma (VP) fe'lga asoslanadi. So'z birikmasiga asos bo'lgan so'z birikmaning **bosh so'zi** deyiladi. Dependency Parsing jarayonida “bosh so'z” va boshqa barcha so'zlar o'rtasidagi munosabatlarni aniq belgilaydigan **tobelik daraxti (Dependency Tree)** deb nomlanuvchi grammatik struktura shakllantiriladi. Ushbu strukturani graf ko'rinishida hosil qilganimizda so'zlar *tugunlar (nodes)*, ular orasidagi bog'liqliklar *qirralardir (edges)*ga ajratiladi. Ba'zi ilmiy tadqiqotlarda ushbu graf **tobelik strukturasi (Dependency Structure)** deb yuritiladi. Ushbu strukturada muayyan ildiz so'z mavjud. Matn ma'lumotlarining boshlang'ich nuqtasi sifatida ushbu so'z orqali boshqa barcha so'zlarga erishish mumkin. Bu so'z gapning markazidir.



2-rasm. “Dependency Tree”ga namuna

Shuningdek, gapga mos tobelik daraxtida munosabatlar deb ataladigan yorliqlarni⁸ shakllantirish mumkin. Ushbu munosabatlar tobelik turi haqida batafsil ma'lumot beradi.

Yuqoridagi 2-rasmdagi $h \rightarrow d$ munosabatda, h – bosh va d – tobe so'z hisoblanadi⁹. **Bosh qism (head)** so'z birikmasining eng muhim tugunidir, **ildiz (root)** esa gapning eng muhim tugunidir: u bevosita yoki bilvosita boshqa har bir tugunning boshidir.

O'zbek tilidagi POS teglarning to'liq ro'yxatida ot so'z turkumining turli pastki turkumlarini ajratib turadi: *birlik shakldagi ot (NN)*, *ko'plik shakldagi ot (NNS)*, ... (*NNP*), ... (*NNPS*) va boshqalar (2-jadval). Otli birikma ot so'z turkumidagi *NN*, *NNS*, *NNP*, *NNPS* kabi kategoriyalarga asoslangan bo'lishi kerak. Xuddi shunday, fe'lli birikma ham fe'lni har xil kichik toifalariga asoslanishi mumkin: *VBP*, *VBZ*, *VBD*, *VCN*, *VBG* va boshqalar.

2-jadval. O'zbek tilidagi sintaktik teglar ro'yxati

№	Teg	Tavsif	Namuna
1.	NP	Otli birikma	Akamning moshinasi
2.	ADVP	Ravishli birikma	Juda ko'p
3.	INTJ	Undov so'zlar	Oh-ho
4.	S	Gap (darak gap)	Men keldim.
5.	SINV	Inversiyaga uchragan gap	Keldi bu bahorlar.
6.	WHADJP	So'roq so'zli birikma	Qayerda?
7.	SBAR	Murakkab gapdagi tobe qism	Bahor kelsa, gullar ochiladi.
8.	NX	Otli birikmaning boshi	Mening ukam
9.	RRC	Qisqartirilgan nisbiy bo'laklar	Kelayotgan odam

⁸ <https://universaldependencies.org/u/dep/index.html>

⁹ <https://universaldependencies.org/u/pos/all.html>

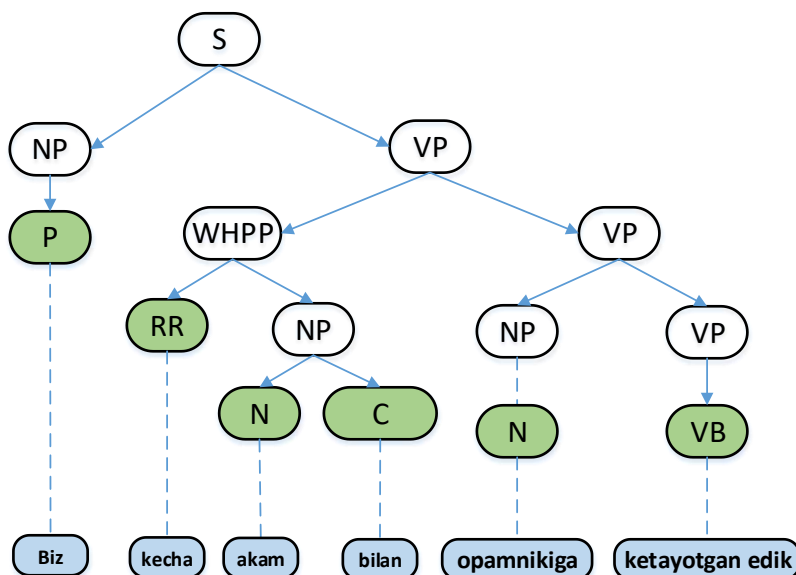
10.	CONJP	Bog'lovchili birikma	Va, bilan, ba'zan.... ba'zan, ..
11.	VP	Fe'lli birikma	Men <i>ertaga maktabga boraman</i> .
12.	FRAG	Gaplardagi bo'lak qismlar	<i>Yoqimli shabada!</i>
13.	QP	Miqdor ko'rsatgichli birikma (otli birikmada)	<i>Ko'p bolalar</i>
14.	PRT	Yuklama	<i>Faqat, axir, -mi?, -chi?</i>
15.	ADJP	Sifatlovchi ibora	<i>Qora sochlar, eng uzun bino</i>
16.	UCP	Gapdagi bo'lak	<i>U ustozlari faxrlanadigan talaba bo'lgan.</i>
17.	TOP	Daraxtning eng yuqori tuguni (root)	Men (men - root) bordim (bor-root)
18.	PRN	Qavs ichidagi izohlar	Alisher Navoiy (1441-1501) buyuk shoir, mutafakkir.
19.	SBARQ	So'roq so'zli birikmalar	nima?, nima uchun?
20.	WHADVP	Holli birikmalar	Qachon? Kecha
21.	WHPP	To'ldiruvchi	Akam bilan, sen uchun
22.	WHNP	So'roq so'z (otli birikmada)	Kim?
23.	NML	Aniqlovchi (Otli birikmada)	O'zbek olimlari, rasmiy kitob

Tobe munosabatlar tahlili (Dependency Parsing) tobelik grammatikasiga asoslangan bo'lsa, *Tarkibiy qismlar tahlili (Constituency Parsing)* **kontekstga bog'liq bo'lmagan grammatikaga (context-free grammar)** asoslanadi. Sintaktik tahlilning bu turi matn ma'lumotlaridagi so'z birikmalari ustida amallar bajaradi. *Tarkibiy qismlar tahlili* jarayonida matnni grammatik toifaga asoslanib, quyi so'z birikmalari, tarkibiy qismlarga ajratiladi. Ushbu qismlar asosan grammatik birliklardir. Grammatik birliklariga **otli birikma (NP)**, **fe'lli birikma (VP)** yoki **ba'zan old so'z birikmasi (PP)** misol bo'ladi.

Sintaktik daraxtlar tabiiy tilda gap yoki so'z birtikmasining sintaktik tuzilishi grafik tasvirini ifodalaydi. Bu daraxtlar gapdagi so'z va birikmalar qanday iyerarxik tarzda tuzilganligi, ular orasidagi grammatik munosabatlarni ko'rsatadi. Sintaktik daraxt lingvistik tahlilning asosiy tushunchasi bo'lib, turli xil tabiiy tillarni qayta ishlash vazifalarida muhim rol o'ynaydi. Sintaktik daraxtlarining asosiy elementlari:

Sintaktik daraxt lingvistik tahlil uchun muhim ahamiyat kasb etadi va turli NLP ilovalarida, masalan, *tahlil qilish, imloni tekshirish* va *mashina tarjimasida* qo'llaniladi. U gapning grammatik tuzilishini vizual tarzda taqdim etadi, bu inson va kompyuter uchun tilni tushunish va boshqarishni osonlashtiradi.

Soʻz birikmasi tuzilishi qoidalari bosh soʻzning qanday qilib boshqa soʻzlar bilan birikma hosil qilishi mumkinligini belgilaydi. Berilgan gapdan soʻz birikmalarini hosil qilish usuli koʻpincha quyidagi 3-rasmdagi kabi tahlil daraxti diagrammasi orqali koʻrsatiladi.



3-rasm. Tahlil daraxti diagrammasi

Baʼzi ilmiy tadqiqotlarda sintaktik tahlil qilingan gapga mos koʻrinishlardan bir qavs ichidagi gapdir. Yuqoridagi 3-rasmga mos ifoda quyidagi koʻrinishga ega:

(S (NP (P Biz)) (VP (WHPP (RR kecha) (NP (N akam) (C bilan))) (VP (NP (N opamnikiga)) (VP (VB ketayotgan edik)))))

Matnni sintaktik tglash uchun turli xil soʻz birikmalari teglari (NP, VP, PP, S va boshqalar) **sintaktik tarkibiy teglar** deb ataladigan teglar toʻplamidir.

Ushbu maqolada matnni sintaktik tglash va uning usullari haqida maʼlumot keltirildi va yechimlari oʻzbek tili leksik birliklari misolida keltirildi.

Foydalanilgan adabiyotlar:

1. Maverick, G. v. (1969). Computational Analysis of Present-Day American English. Henry Kučera, W. Nelson Francis. International Journal of American Linguistics, 35(1). <https://doi.org/10.1086/465045>
2. Demirşahin, I., & Zeyrek, D. (2017). Pair Annotation as a Novel Annotation Procedure: The Case of Turkish Discourse Bank. In Handbook of Linguistic Annotation. https://doi.org/10.1007/978-94-024-0881-2_46
3. Dickinson, M., & Tufiş, D. (2017). Iterative Enhancement. In Handbook of Linguistic Annotation. https://doi.org/10.1007/978-94-024-0881-2_9

4. Core, M., Ishizaki, M., Moore, J., & Nakatani, C. H. (1999). The report of the Third Workshop of the Discourse Resource Initiative. Chiba University and Kazusa Academia Hall.
5. Cunningham, H. (2002). GATE, a general architecture for text engineering. *Computers and the Humanities*, 36(2). <https://doi.org/10.1023/A:1014348124664>
6. Day, D., Aberdeen, J., Hirschman, L., Kozierok, R., Robinson, P., & Vilain, M. (1997). Mixed-initiative development of language processing systems. 5th Conference on Applied Natural Language Processing, ANLP 1997 - Proceedings. <https://doi.org/10.3115/974557.974608>
7. Bird, S., Day, D., Garofolo, J., Henderson, J., Laprun, C., & Liberman, M. (2000). ATLAS: A flexible and extensible architecture for linguistic annotation. 2nd International Conference on Language Resources and Evaluation, LREC 2000.
8. Elov B., Hamroyeva Sh., Xudayberganov N., Yodgorov U., Yuldashev A. Pos tagging of Uzbek texts using hidden Markov models (HMM) and Viterbi algorithm. *O'zMU xabarлари. Mirzo Ulug'bek nomidagi O'zbekiston Milliy universiteti ilmiy jurnali*. 2023 yil Maxsys son.
9. Elov, B.B., Hamroyeva, Sh.M., Abdullayeva, O.X., Husainova, Z.Y., Xudoyberganov, N.U. 2023. “Agglutinatив tillar uchun pos teglash va stemming masalasi (turk, uyg'ur, o'zbek tillari misolida)”. *O'zbekiston: til va madaniyat* 2: 6-39.
10. Ramatova M. O'zbek tilidagi sodda gaplarning sintaktik teglangan bazasi orqali tahlil daraxtini qurish // *Educational Research in Universal Sciences*. VOLUME 2 | ISSUE 12 | 2023. // <https://zenodo.org/records/10463799>.

FRAZEOLOGIK POLISEMIYA VA UNING KORPUS TILSHUNOSLIGIDA MODELLASHTIRISH MASALALARI

Gulyamova Shahnoza Qahromonovna

Filologiya fanlari doktori. DSc

bunnycham@mail.ru

Turkiston yangi innovatsiyalar
universiteti dotsenti,

Nurboyeva Maftuna Vahobjon qizi

ToshDO'TAU 2-bosqich tayanch doktoranti,

maftunanurboyeva@gmail.com

Annotatsiya. Maqolada frazeologik polisemiyaning o'zbek tilshunosligida o'rganilishi, lug'atlarda berilishi va polisemantik iboralarning korpus tilshunosligi tizimiga lingvistik model asosida kiritish hamda iboralarning ko'p ma'noliligi tufayli yuzaga kelishi mumkin bo'lgan noaniqliklarni bartaraf etish masalalari haqida so'z boradi.

Abstract. The article deals with the study of phraseological polysemy in Uzbek linguistics, its inclusion in dictionaries and inclusion of polysemantic expressions into the system of corpus linguistics based on the linguistic model, and the elimination of ambiguities that may arise due to the polysemy of expressions.

Аннотация. В статье речь идет об изучении фразеологической многозначности в узбекском языкознании, ее включении в словари и включении многозначных выражений в систему корпусной лингвистики на основе лингвистической модели, а также устранении неясностей, которые могут возникнуть из-за многозначности выражений.

Kalit so'zlar: *frazeologiya, frazeologik polisemiya, frazeologik omonimiya, lingvistik model, variatsiya, morfo-semantik kollokatsiya, idioma, erkin birikma.*

Tilshunosligimizda nisbatan yangi tarmoq bo'lgan frazeologiya sohasiga oid ilk tadqiqotlar o'tgan asrning 50-yillariga borib taqaladi. Dastavval, o'zbek tilshunos olimlaridan Shavkat Rahmatullayev iboralar va ularning ta'rif va tasniflarini o'z tadqiqotlarida yoritgan bo'lsa, bugungi kungacha bir qancha olimlar bu soha bo'yicha chuqur tadqiqotlar olib bordilar va shunga ko'ra frazeologik qarashlar xilma-xillik kasb etdi.

An'anaviy va substansial tilshunosligimizda frazeologiya sohasi bir qator tadqiqotlarga obyekt bo'lib xizmat qilgan bo'lsa, uning korpus tizimiga tatbiq etilishi, ayniqsa, ko'p ma'noli va omonimik xarakterli iboralarni modellash masalasi nafaqat o'zbek tilshunosligida, balki jahon tilshunosligida ham yechimini kutayotgan mavzulardan biri hisoblanadi. Shuningdek, frazeologik birliklar va ularning polisemantik imkoniyatlari masalasida ham shu vaqtga qadar qilingan tadqiqotlar va lug'atlarni ko'zdan kechirganimizda bir qator xilma-xilliklar borligini kuzatdik.



Frazeologik birliklarning xuddi leksema kabi shakl va ma'no munosabatiga ko'ra turlari mavjudligi bo'yicha ilk izlanishlar Sh.Rahmatullayev tomonidan tasnif qilingan bo'lib, olim fikricha, frazeologik polisemiya leksik polisemiya kabi keng qamrovli bo'lmasligi, leksema kabi iboralarning 10-15 talab ma'nosi bo'lmasligi, ma'no imkoniyati 7-8 tadan oshmasligi va buning sababi esa iboraning bir nechta so'zdan tarkib topishi hamda bu holat yangi qo'shimcha ma'no hosil bo'lishini birmuncha qiyinlashtiradi [Rahmatullayev, 1966: 26]. Ammo olimning nomzodlik dissertatsiyasida iboralar tizimida biroz nomutanosiblik borligi kuzatiladi, ya'ni unda *oxiratda*, *qiyomatda*, *latta*, *muzlamoq* kabi bosh ma'nosidan uzoqlashib, o'z ma'nosida qo'llanish imkoniyatini deyarli yo'qotgan so'zlarni ham frazeologik (idiomatik) so'zlar qatorida sanab o'tilgan [Rahmatullayev, 1952: 17]. Shuningdek, olim o'zining idiomatik so'zlar haqidagi fikrlariga yana to'xtalib, nomzodlik ishining keyingi o'rinlarida *jonkuyar*, *jonkuyarlik*, *tilyog'lama*, *tilyog'lamachilik*, *bichib-to'qimoq*, *manmanlik*, *sansalorlik*, *o'zboshimchalik* kabi birliklarni etimologik jihatdan qo'shma so'zdan o'sib chiqqan idiomatik so'zlar deya baholaydi. Vaholanki, olimning yana bir asarida iboralarni grammatik shakllangan birdan ortiq mustaqil leksik negizdan tashkil topuvchi tarkibli til birligi ekanligi ta'kidlangan [Rahmatullayev, 1970: 4].

M.Husaynov va A.Shomaqsudov ishlarida ham, umumxalq frazeologik birliklari bilan shaxsiy (ommalashmagan, muayyan shaxs nutqidagina ishlatilgan) frazemalar o'rtasidagi farq aniq sezilmasligini ko'rish mumkin va tadqiqotlarda sof lingvistik qarashlar bilan badiiy-uslubiy tahlil aralash kelganligiga guvoh bo'lamiz [Husaynov, 1959: 18]. Bu murakkabliklar U.Tursunov va A.Koklyanova ishlarida alohida qayd qilinib, frazeologik sathni aniq chegaralash va tizimli tahlil qilish kerakligi aytiladi [Tursunov, 1961: 106; Koklyanova, 1961: 99-101].

Ammo Sh.Rahmatullayevning keyingi tadqiqot ishlarida frazeologik sath anchayin tizimlashgan holda tadqiq etilganligini alohida ta'kidlash joiz. Olimning nomzodlik ishi orqali o'zbek tilidagi fe'l frazemalarining grammatik xususiyatini tahlil qilgan bo'lsa, 1966-yilda himoya qilingan doktorlik dissertatsiyasi va shu asosda nashr qilingan “O'zbek frazeologiyasining ba'zi masalalari” nomli monografiyasida frazeologik sath birliklarini lug'aviy birlik sifatida baholab, ularda ko'p ma'nolilik, sinonimlik, omonimlik, antonimlik, paronimlik, variatsiya kabi til hodisalarini o'ttiz mingdan ortiq kartochkalar orqali tahlil qilib berdi. Monografiyaning e'tiborga molik jihati shundaki, unda frazeologik omonimlar, paronimlar va frazeologik paraformalarni o'zaro farqlab chiqilgan, ya'ni olim fikricha, bu uch hodisa bir-biriga shu qadar o'xshash bo'ladiki, ularni xato qo'llab yuborish mumkin: agar iboralar o'z tarkibidagi biror so'z-komponenti bilangina farq qilsa, ularni frazeologik paronim deya baholash mumkin, misol tariqasida esa quyidagi iboralar keltiriladi: *ko'ngl(i) tog'day ko'tarilmoq*, *ko'ngl(i)ni tog'day ko'tarmoq*. Frazeologik paraforma qatoriga esa ba'zi biri ikkinchisida yo'q grammatik shakllanish sistemasiga ega bo'lgan turli xil shakllarda qo'llaniluvchi iboralarni kiritadi, masalan, *jon(i) kirdi* – *jon(i)ni kirgizmoq* kabi. Ammo frazeologik



paronimga keltirilgan misollar ham aynan grammatik jihatdan bir-biridan farq qilib turgani va shu jihatiga ko'ra, aniqroq aytadigan bo'lsak, shaxsga xoslanishiga polisemantik iboralar qatoriga kiritilgan “ona(m/si)ni Uchqo'rg'ondan ko'r(sat)moq” iborasi bilan solishtirilganda tasnifda biroz tushunmovchilik borligini payqash mumkin, chunki olimning “O'zbek tilida fe'l frazemalari bog'lashuvi” kitobida U nomardga *onasini Uchqo'rg'ondan ko'rsataman* – Ha, birovga og'iz ocha ko'rmalaring, komandir *onamni Uchqo'rg'ondan ko'rsatadi-ya*” misollari orqali ushbu ibora polisemik xarakterli iboralar qatoriga kiritilgan [Rahmatullayev, 1992: 104].

Shuningdek, frazeologik polisemiya, umuman frazeologik birliklar bilan bog'liq yana bir muammo borki, u frazeologik polisemiyaning chegaralab olishda qiyinchilik tug'diradi, bu ham bo'lsa, qo'shma so'zlar. Chunki ko'pgina tadqiqotlarda frazeologik birliklar qo'shma so'zlar bilan aralashtirib yuborilgani va ularni o'zaro farqlash murakkabligini ko'rish mumkin. Xususan, U.Tursunov va J.Muxtorovlarning “Hozirgi zamon o'zbek tili” nomli qo'llanmasida *javob bermoq, ko'ngil qo'yimoq, xalal bermoq, quloq solmoq, nazar tashlamoq* kabi birikmalar qo'shma fe'llar qatoriga kiritilganini ko'rishimiz mumkin, vaholanki, bu birikmalarning hammasi frazeologik lug'atdan o'rin olgan bo'lib, ularning bir nechtasi polisemik xarakterga ham ega. Shu bois, biz tahlilga tortayotgan birikmalar aslida qo'shma so'zmi yoki frazeologik birlikmi, bu xususda bir to'xtamga kelish lozim bo'ladi, aslida ism-fe'l birikuvidagi barcha qo'shma fe'llarga diqqat qaratsak, ularning barchasi ma'no ko'chishi tufayli hosil bo'lganini aniqlash mumkin, balki, bu tipdagi barcha fe'llarni iboralar qatoriga kiritish kerakdir. Olim A.Hojiyevning “Fe'l” nomli darsligida qo'shma fe'l va iboraning farqli tomonlari alohida aytib o'tiladi, ya'ni qo'shma fe'l ma'no va forma butunligi bo'ladi, iborada esa ma'no va qisman forma butunligi mavjuddir, aniqroq aytadigan bo'lsak, qo'shma fe'lni tashkil qilayotgan so'zlar qanday bo'lsa, shundayligicha matnga olib kiriladi, shaxsga moslashmaydi, iborada ham qisman butunlik mavjudligi uni tashkil qilgan so'zlar orasiga boshqa gap bo'laklari qo'yib bo'lmasligida namoyon bo'ladi, ammo ularni gap shaxsiga moslash, birliklarning o'rnini o'zgartirib qo'llash mumkin, degan fikrni ilgari suradi [Hojiyev, 1973: 4-7]. Biroq olim bu birliklar xususida mutlaq to'xtamga kelish lozimligini ta'kidlab o'tganidan payqash mumkinki, hali bu masala to'laqonli yechimini topgan emas.

Shuningdek, tadqiqotimiz obyekti bo'lgan frazeologik lug'atlarda ham o'zaro nomutanosiblikni uchratish mumkin, ya'ni ba'zi polisemantik shakldagi iboralar bir lug'atda bo'lib, boshqasida bir ma'noli shaklda ham uchramasligi, polisemantik deya ko'rsatilgan iboralar aslida so'z birikmasi bilan teng kelib qolgan holatlar bo'lishi yoki polisemiya va omonimiya holatlarining noaniq bo'lib qolishi, frazeologik birliklardan tashqari paremiologik birliklar qo'shib ketganini ko'ramiz. Yuqoridagi holatlarga o'rnak sifatida *dars olmoq* iborasini kiritish mumkin: Sh.Rahmatullayevning frazeologik izohli lug'atida bir ma'noli ibora sifatida ham uchramagani holda, yangi izohli lug'atda polisemik ibora sifatida uchraydi: dars

olmoq kim? kimdan yoki nimadan? – 1. O'rganmoq. - *Bo'taboy... o'qishga juda astoydil bo'lib, ba'zi kechalari O'rmonjonning uyiga borib, qo'shimcha dars ola boshladi. A.Qahhor. Qo'shchinor chiroqlari.* 2. Saboq olmoq, kamchiliklarini tuzatish uchun tajriba hosil qilmoq - *Har bir qoqilgan o'rnim, olgan zarbalarimdan o'zim uchun hayotiy dars oldim.* Gazetadan.

Shuningdek, diqqatimizni tortgan frazeologik birliklardan yana biri *orqada qolmoq* iborasidir. Mazkur ibora Sh.Rahmatullayevning frazeologik lug'atida mavjud emas, O'zbek tilining izohli lug'atida esa bir ma'noli ibora sifatida quyidagicha izohlanadi: **orqada qolmoq** – qoloqlik qilmoq, biror vazifani o'z vaqtida bajarmaslik. *Shalola qurilishi hozir eng orqada qoldi.* S.Ahmad. Ufq. Yangi frazeologik izohli lug'atda esa mazkur ibora 3 xil ma'no kasb etishi ta'kidlangan, ammo 3-ma'noning polisemik yoki omonimik xarakterda ekanligi izohlanmagan. Ya'ni mazkur iboraning dastlabki ikki ma'no imkoniyati o'zaro omonimlik hosil qiladi, deb baholangan: **Orqada qolmoq** - I. Kim? Kimdan? Nimada? Biror faoliyatda boshqalar erishgan darajaga yetisha olmaslik. Varianti: ortda qolmoq. *Ular eng ibtidoiy tilga – harakatlari, imolar bilan fikr uqtirishga murojaat qildilar. Bu ishda Petrov mahorat ko'rsatdi. Yo'lchi ham undan ortda qolmaslikka tirishdi.* Oybek. Qutlug'qon. II. Nima? Taraqqiyotning keyingi bosqichida barham topmoq, xoli bo'lmoq. *Vaqtiki kelib, insoniyat balog'atga yetadi, uning salbiylik fe'l-atvorlari orqada qoladi.* A.Qahhor. Sarob. Frazemaning uchinchi ma'nosi esa aytib o'tganimizdek, ochiqdanmagan: **orqa(sida) qolmoq** – kim? kimning? Bevaqt o'lmay, kattalar vafotidan keyin yashamoq. *Ona, bolam orqamda qoladi, chirog'imni yoqadi, deb niyat qilardi.* S.Ahmad, Ufq. Fikrimizcha, mazkur iboraning omonim sifatida ko'rsatilgan dastlabki ikki ma'nosi omonimik emas, polisemik xarakter kasb etmoqda, mansubligi izohlanmagan uchinchi ma'nosi esa yuqoridagi ikki ma'noga nisbatan omonimlik hosil qilmoqda, chunki semantik imkoniyati uyg'unlik kasb etmagan. Bu kabi holatlar bizga polisemantik iboralarni lingvistik modellashirishda ularni aniq chegaralab olishda biroz muammo tug'dirishi mumkin.

Yuqorida semantik jihatlari tavsiflangan ikki iboraning korpus tizimidagi lingvistik modelini yaratishga harakat qilamiz, dastavval iboralarning morfo-semantik kollokatsiyasini tuzamiz, negaki iboralarning polisemantik yoki omonimik holatlarini korpus tizimida bartaraf etish uchun uning har bir semasidagi boshqa leksema va grammatik shakllar bilan birikuvchanlik holatlarini farqlab olish va shu orqali ularning har birini tizimga “tanitish” mumkin bo'ladi. Dastlab, *dars olmoq* iborasining morfo-semantik kollokatsiyasiga diqqat qaratamiz:

T/r	Ibora nomi	Ma'nosi	Birikuvi	Namunalar
1.	Dars olmoq	O'rganmoq	Shaxs oti + ch.k. ~ dars olmoq	<i>Bo'taboy... o'qishga juda astoydil bo'lib, ba'zi</i>

				<i>kechalari</i> <i>O'rmonjonning</i> <i>uyiga borib,</i> <i>qo'shimcha dars ola</i> <i>boshladi. A.Qahhor.</i> <i>Qo'shchinor</i> <i>chiroqlari.</i>
2.	Dars olmoq	Saboq olmoq, kamchiliklarni tuzatish uchun tajriba hosil qilmoq	Mavhum ot+ch.k. ~ dars olmoq	<i>Har bir qoqilgan</i> <i>o'rnim, olgan</i> <i>zarbalarimdan</i> <i>o'zim uchun hayotiy</i> <i>dars oldim.</i> <i>Gazetadan</i>

Shu asosda, iboraning boshqa soʻz va grammatik shakllar bilan birika olish imkoniyatidan kelib chiqib, uning har bir semasi boʻyicha lingvistik model yaratishga harakat qilamiz. Demak, mazkur iboraning 1-maʼnosi korpus tizimida quyidagicha model asosida oʻrin olishi mumkin:

1. PN + Abl_CS_aff + P_{(v)phr}

Bunda PN – Personal noun – shaxs oti, Abl_CS_aff – chiqish kelishi, P_{(v)phr} – feʼl turkumiga oid polisemantik ibora. Yaʼni biz dasturga grammatik sxema tanish boʻlishi uchun modellarda ingizcha atamalarining qisqartma shaklidan foydalandik. Iboraning 2-maʼnosini esa tizimga quyidagi model asosida biriktirish mumkin:

2. PN + Abl_CS_aff + P_{(v)phr} - bunda AN (abstract noun) – mavhum ot hisoblanib, unga chiqish kelishi qoʻshilishi va undan keyin iboraning joy olishi orqali iboraning maʼnosi ochiqatlanadi.

Yuqorida *orqada qolmoq* iborasining ham semantikasiga diqqat qaratgan edik, uning lingvistik modelini tuzish uchun ham avvaliga morfo-semantik valentligini koʻzdan kechiramiz:

T/r	Ibora nomi	Maʼnosi	Birikuvi	Namunalar
1.	Orqada qolmoq	Biror faaoliyatda boshqalar erishgan darajaga yetisha olmaslik	Shaxs nomi/oti + ch.k./ oʻ.k. ~ orqada qolmoq	<i>Xotinlar</i> <i>ko'ksilarini silkitib,</i> <i>qattiq kulishdi.</i> <i>Anna ham</i> <i>Mariyadan orqada</i> <i>qolishni istamadi.</i> <i>Oybek. Qutlugʻqon</i>
2.	Orqada qolmoq	Taraqqiyotning keyingi bosqichida	Mavhum ot+ b.k. ~ orqada qolmoq	<i>Vaqtiki kelib,</i> <i>insoniyat</i> <i>balogʻatga yetadi,</i>

		barham topmoq, xoli bo'lmoq.		<i>uning salbiylik fe'l- atvorlari orqada qoladi. A.Qahhor. Sarob.</i>
I.	Orqa(-m/- ng/-si)da qolmoq	Bevaqt o'lmay, kattalar vafotidan keyin yashamoq.	Shaxs oti/Kishilik olmoshi + b.k./q.k. ~ orqa(-m/-ng/- si)da qolmoq	<i>Ona, bolam orqamda qoladi, chirog'imni yoqadi, deb niyat qilardi. S.Ahmad. Ufq.</i>

Ko'rinadiki, ibora ham polisemantik, ham omonimik xarakterga ega. Lingvistik modellashtirish uchun iboraning polisemantik yoki omonimik xususiyatga ega ekanligi katta ahamiyatga ega emas, negaki har bir ibora va uning ma'no imkoniyati uning morfo-semantik birikuv usuliga ko'ra amalga oshiriladi. Demak, mazkur iboramizning har bir ma'nosi uchun alohida model tuzish lozim:

2. PN + Abl_CS_aff/Loc_CS_aff + P_{(v)phr}
3. AN + Nom_CS_aff + P_{(v)phr}
4. PN/PP + Nom_CS_aff/Gen_CS_aff + P_{(v)phr}

Bu modellar korpus tizimiga kiritilgandan so'ng mazkur iboralar jumla tarkibida kelgan vaqtda korpus tizimi iboraning grammatik birikuvidan kelib chiqib, tegishli ma'noni chiqarib beradi. Shu o'rinda savol tug'ilishi mumkin: iboralarning so'z birikmasiga teng kelib qolish holatlari ham mavjud. Bu holatda yuqoridagi modellar ibora bilan uning erkin so'z birikmasi shaklini qanday farqlay oladi? Aytish mumkinki, so'z birikmasi erkin birikuvga ega bo'lgani uchun ham yuqoridagi modellarga mos tushish holatlari juda kam bo'ladi, shuningdek, uning yaxlit so'z emas, erkin birikma ekanligi nutq vaziyatidan ham anglashilib turadi. Negaki, biz iboralar necha so'zdan iborat bo'lishidan qat'iy nazar model tarkibida uni bitta belgi ostida berdik. Erkin birikma shakli tarkibidagi har bir so'z esa alohida belgi ostida modellashtirilishi shubhasiz.

Frazeologik birliklar va ularning nutqdagi nazarda tutilgan ma'nosini anglash barcha tillarda doim murakkablik tug'dirgan. Ayniqsa, mashina tarjimonligining eng zaif nuqtasi ham aynan shu. Chunki iboralarning bosh ma'nosini anglash alohida jarayon bo'lsa, uning hamroh - qo'shimcha ma'nolarini ajrata olish yana bir murakkab jarayondir. Polisemantik iboralarni modellashtirish esa bu holatni yengillashtiribgina qolmay, matnni to'g'ri anglashga ham yordam beradi.

Foydalanilgan adabiyotlar:

1. Раҳматуллаев Ш. Ўзбек фразеологиясининг баъзи масалалари. – Тошкент: Фан, 1966.
2. Раҳматуллаев Ш. Ҳозирги замон ўзбек тилидаги образли феъл фразеологик бирликларнинг асосий грамматик хусусиятлари. Номзодлик диссертацияси. – Москва, 1952. – Б. 52.
3. Раҳматуллаев Ш. Нутқимиз кўрки. –Тошкент, 1970.– 56 б. 4-б
4. Хусайнов М. Фразеология прозы писательницы Айдын. АКД. – Самарканд, 1959. – 18 с.
5. Шамаксудов А.Ш. Язык сатиры Муками (лексика и фразеология). АКД. – Ташкент, 1956. – 18 с.
6. Турсунов У.Т. Состояние и задачи изучения фразеологии узбекского языка // Вопросы фразеологии, Труды СамГУ, новая серия, выпуск 106. - Самарканд, 1961. – С. 8-9.
6. Коклянова А.А. О границах фразеологии тюркских языков // Вопросы фразеологии, Труды СамГУ, новая серия, выпуск 106. - Самарканд, 1961. – С. 99-101
7. Хожиев. А. Феъл. - Т.: 1973. 191.б
8. Rahmatullayev Sh., Mahmudov N., Xolmonova Z., O‘razova I., Rixsiyeva K. O‘zbek tili frazeologik lug‘ati. “G‘afur G‘ulom”. -T.: 2022.

TIL KORPUSLARINI LINGVISTIK TEGDLASH BOSQICHLARI

Elov Botir Boltayevich

Texnika fanlari bo'yicha falsafa doktori PhD, dotsent

elov@navoiy-uni.uz

ToshDO'TAU Kompyuter lingvistikasi va raqamli
texnologiyalar kafedrası mudiri

Xusainova Zilola Yuldashevna

xusainovazilola@navoiy-uni.uz

ToshDO'TAU tayanch doktoranti

Annotatsiya. Til korpusi elektron shaklda taqdim etiladigan *katta hajmli* va *strukturlangan matnlar to'plami* sifatida qaraladi. Til korpusi *yozma* yoki *og'zaki* materialni ifodalab, NLP tizimini mavjud resurslarni o'rganishi uchun lingvistik tahlilni amalga oshirish lozim. Bugungi kunda zamonaviy NLP modellarini yaratish lingvistik bilimlarning akustika va prosodika, fonetika, orfografiya, morfologiya, leksikologiya, sintaksis, semantika, pragmatika va diskurs kabi spetsifikatsiyasini talab qiladi. NLP bilan bog'liq lingvistik bilimlar leksik, sintaktik, semantik va pragmatik xususiyatlarni o'z ichiga oladi. Ushbu maqolada lingvistik tegdlashning asosiy amallari haqida fikrlar keltiriladi. Maqolada orfografiya bilan bog'liq tegdlash – tokenizatsiya va gap chegaralarini aniqlash, token yoki tokenlar to'plamining so'z turkumini belgilash (PoS tegdlash), sintaktik, semantik tegdlash, korpus matnlaridagi koreferensiyani aniqlash masalalari va ularni amalga oshirish usullari haqida umumiy ma'lumot berilgan.

Abstract. A linguistic corpus is considered a large and structured collection of texts presented in electronic form. In order for the language corpus to represent written or spoken material, it is necessary to perform a linguistic analysis in order for the NLP system to learn the available resources. Today, the creation of modern NLP models requires the specification of linguistic knowledge such as acoustics and prosody, phonetics, orthography, morphology, lexicology, syntax, semantics, pragmatics and discourse. Linguistic knowledge related to NLP includes lexical, syntactic, semantic and pragmatic features. This article provides an overview of the basic operations of linguistic tagging. The article deals with orthography-related tagging – tokenization and sentence boundary determination, token or set of tokens tagging (PoS tagging), syntactic, semantic tagging, coreference detection in corpus texts, and general information about their implementation methods is given.

Аннотация. Лингвистический корпус – это большая и структурированная коллекция текстов, представленная в электронном виде. Чтобы языковой корпус представлял письменный или устный материал, необходимо провести лингвистический анализ, чтобы система НЛП могла изучить доступные ресурсы. Сегодня создание современных моделей НЛП требует уточнения таких лингвистических знаний, как акустика и просодия,

фонетика, орфография, морфология, лексикология, синтаксис, семантика, прагматика и дискурс. Лингвистические знания, связанные с НЛП, включают лексические, синтаксические, семантические и прагматические особенности. В этой статье представлен обзор основных операций лингвистического тегирования. В статье представлен обзор тегирования, связанного с орфографией – токенизация и определение границ предложения, токенизация или набор токенов (PoS-тегирование), синтаксическое, семантическое тегирование, обнаружение кореференции в корпусных текстах и методы их реализации.

Kalit soʻzlar: *til korpusi, lingvistik teglash, morfologik tahlil, tokenlash, lemmalash, POS teglash, sintaktik tahlil, semantik tahlil, NER, teglangan korpuslar.*

Kirish

Tabiiy tilni kompyuter yordamida tadqiq qilish va qayta ishlash, odatda, toʻrt jihatni: formallashtirish, algoritmlash, dasturlash va amaliyotga qoʻllashni oʻz ichiga oladi.

Birinchidan, tilshunoslik nuqtayi nazaridan oʻrganiladigan muammolarni **formallashtirish**, tilning maʼlum bir matematik shaklda qatʼiy va muntazam ravishda ifodalanishi uchun formal modelni ishlab chiqish kerak.

Ikkinchidan, ushbu qatʼiy va muntazam matematik shaklni algoritm shaklida ifodalash talab etiladi: bu jarayon **algoritmlash** deb ataladi.

Uchinchidan, NLPning turli amaliy tizimlarini shakllantirish uchun algoritmgaga asoslangan kompyuter dasturini yozish, uni kompyuterda amalga oshirish kerak: bu jarayonni **dasturlash** deb atash mumkin.

Toʻrtinchidan, oʻrnatilgan NLP tizimini foydalanuvchining ehtiyojini qondirish, sifat va ish faoliyatini doimiy ravishda yaxshilab borish uchun baholash lozim; bu jarayonni **amaliylashtirish** deb atash mumkin.

Korpus matnlarini lingvistik teglash dastlab lingvistik nazariyalarni yoki bugungi kunda maʼlum boʻlganidek, til korpuslarni ishlab chiqish va tahlil qilish uchun maʼlumot berish uchun amalga oshirildi [Hovy, Lavid, 2010; Arista, 2022]. Matnlarni “qoʻlda” teglash uchun koʻp vaqt va kuch talab qilinadi. Biroq soʻnggi oʻttiz yil ichida hisoblash texnikalarning quvvati va katta hajmdagi maʼlumotlarni saqlash imkoniyati hosil boʻlgach, katta hajmdagi til korpuslarini ishlab chiqish va avtomatik teglash sohasidagi bir qator yutuqlarga erishildi [Xusainova, 2022; Bi, 2018]. Matnlardan iborat katta hajmli til korpuslari ustida turli lingvistik va amaliy tadqiqotlar olib borildi [Xusainova, 2023]. Tabiiy tilni yangi texnologiyalar asosda tadqiq qilish va, eng muhimi, ishonchli statistik modellarni ishlab chiqish uchun lingvistik teglangan matn va til korpusiga asoslangan tabiiy tilni qayta ishlash (NLP) sohasiga muhim xizmat qiladi.

Lingvistik teglash tavsiflovchi yoki analitik belgilarni til maʼlumotlari bilan bogʻlashni oʻz ichiga oladi. Strukturlanmagan (qayta ishlanmagan) maʼlumotlar matnli yoki har qanday manba yoki janrdan olingan yoki vaqt funksiyalari (audio,



video va/yoki fiziologik yozuvlar) shaklida bo'lishi mumkin. Teglar barcha turdagi transkripsiyalarni (fonetik xususiyatdan nutqiy qurilmagacha), POS teglash, ma'no belgilari, sintaktik tahlil, NER obyektlar, semantik rol belgilari, vaqt va hodisalarni aniqlash, so'zlarning sintaktik zanjirlarini, nutq darajasini o'z ichiga olishi mumkin [Xusainova, 2022; Bi, 2018; Elov, Hamroyeva, Axmedova, 2023]. Lingvistik teglashning eng muhim komponenti bu tegishli teg birligi (masalan, tovush, token yoki so'z, birikma, fragment, hujjat) bilan bog'lanishi kerak bo'lgan teglar va tegishli xususiyatlarni belgilaydigan teglash sxemasi hisoblanadi.

Bugungi kunda korpus matnlarini lingvistik teglashni qo'lda yoki yarim avtomatik tarzda amalga oshirish mumkin. NLP bilan bog'liq lingvistik bilimlarning jihatlarini haqida turli xil qarashlar mavjud. Umuman olganda, NLP tadqiqotchilarining aksariyati NLP bilan bog'liq lingvistik bilimlar, hech bo'lmaganda, **leksik, sintaktik, semantik** va **pragmatik** xususiyatlarni o'z ichiga olishi kerak, deb hisoblashadi.

Har bir xususiyat lingvistik ma'lumotni farqli yo'llar bilan uzatadi. Masalan, leksik xususiyat so'z darajasidagi asosiy tarkibiy qismlar (masalan, morfema) va uning flektiv shakllari haqidagi bilimlarni qamrab olishi mumkin; sintaktik xususiyat ma'lum bir tildagi so'z yoki so'z birikmalarining gap hosil qilishini o'z ichiga oladi; semantika ma'lum so'zlar yoki gaplarga qanday ma'no berishni; pragmatika suhbatda nutq markazidagi o'zgarishlarni, berilgan kontekstdagi gap ma'nosini izohlash haqida bilimlarni o'z ichiga oladi.

Yuqorida keltirilgan to'qqizta xususiyat, asosan, lingvistik bilimlarni o'z ichiga oladi, chunki NLP, asosan, lingvistik muammo deb hisoblanadi. Biroq lingvistik bilimga qo'shimcha ravishda NLP *axborot texnologiyalari, matematika, psixologiya, falsafa, statistika* va *biologiya* sohalaridagi boshqa bilimlarni ham o'z ichiga olishi mumkin.

Tilshunoslikda lisoniy hodisalarni tahlil qilish va tavsiflash, odatda, bir necha alohida bosqichlarda amalga oshiriladi. Tilning turli tovushlari **fonetikada** tavsiflansa, yozuv tizimi **orfografiyada** o'rganiladi. **Morfologiya** so'zning shakllanishi, o'zgarishini, **sintaksis** so'zlarning joylashuvi, ularning so'z birikmasi va gapga aylanishini aniqlaydi. **Semantika** so'z (*leksik semantika*), birikma hamda gapning ma'nosini (*kompozitsion semantika*) tahlil qiladi. So'z va so'z birikmalarining turli sharoit va vaziyatda o'ziga xos ma'no anglatishini **pragmatik tahlil** natijasida o'rganish mumkin. Shaxs va narsalar qanday qilib mavzu sifatida kiritilishi va keyinchalik qanday tilga olinishi **diskurs tahlil**ning predmeti sanaladi.

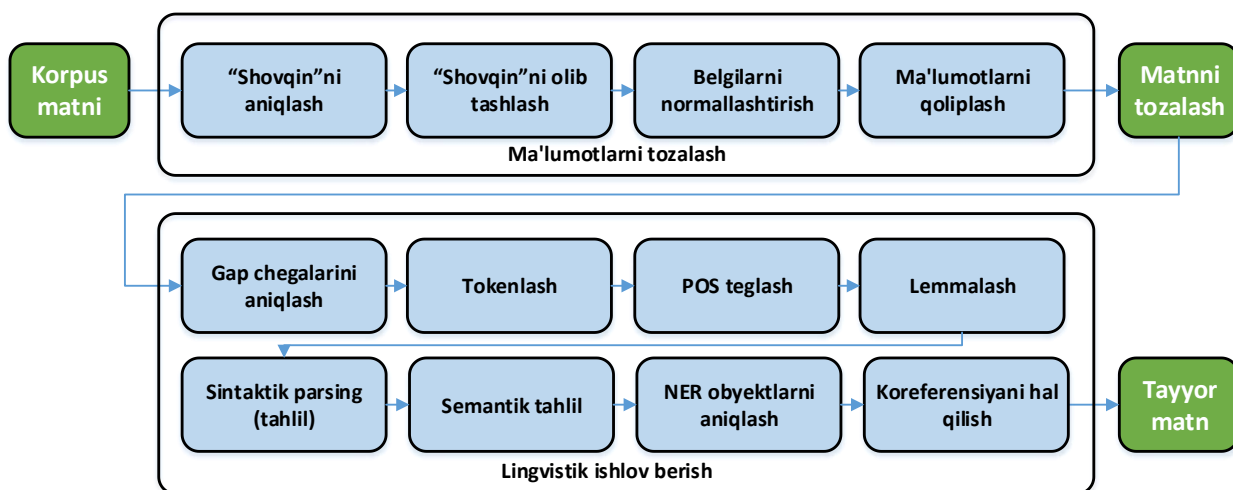
Fonetika va **orfografiya** eng kichik birliklar (alohida tovush va harf) bilan shug'ullanadi. **Morfologiya, sintaksis** va **semantika** o'rta hajmdagi birlik(so'z, so'z birikmasi va gap)larni o'rganadi. **Diskurs** va **pragmatikaning** predmeti yuqorida keltirilganidek eng katta birliklar (*paragraf* va *dialog*)dir.

Bugungi kunda tilshunoslikda matnni teglashning zamonaviy usullari, shuningdek, turli xil teglash vazifalarini quyidagi – 1-jadval va 1-rasmda

ko'rsatilganidek, o'xshash qatlamlar to'plamiga joylashtirish mumkin bo'lgan turli bosqichlarga ajratish mumkin.

2-jadval. Lingvistik teglash bosqichlari

№	Teglash usuli	Tavsif
1.	Gap chegaralarini aniqlash (sentence boundaries)	matnni gaplarga ajratish
2.	Tokenlash (tokenization)	matnni so'zlarga ajratish
3.	Lemmalash	so'zshaklni lemma (lug'atdagi shakli)ga keltirish
4.	POS teglash (part-of-speech tagging)	so'zlarning turkumlari bilan aniqlash va belgilash
5.	Sintaktik parsing (tahlil)	gapdagi tarkibiy so'z birikmalarini tahlil qilish
6.	Semantik parsing (tahlil)	predikat-argument munosabatlarini belgilash (labeling predicate-argument relations)
7.	NER obyektlarini aniqlash (named entity recognition)	atoqli otlarni aniqlash va belgilash
8.	Koreferensiyani hal qilish (coreference resolution)	matndagi bir xil obyektlarga havolalarni bog'lash



1-rasm. Korpus matnlarini teglashning zamonaviy usullari

Biroq til korpuslarini teglashda bajariladigan vazifalar bilan tilshunoslik nazariyasidagi tavsif bosqichlari o'rtasida faqat umumiy muvofiqlik mavjud. Ushbu maqolada matnlarni teglashning zamonaviy yondashuvlari ko'rib chiqiladi.

Orfografiya bilan bog'liq teglash vazifalari **tokenlash** va gap chegaralarini aniqlashdir. Ushbu vazifalar matnni alohida **so'z (belgi)lar** va alohida **gaplarga**

ajratishda qo'llanadi. Odatda, bu ikki vazifaning qaysi biri oldin bajarilishi muhim emas, lekin ikkala vazifa ham keyingi bosqichdagi vazifalarni bajarishdan oldin bajarilishi lozim.

Matn alohida tokenlarga ajratilgandan so'ng har bir token yoki tokenlar to'plamini **so'z turkumi** (ot, fe'l, olmosh va boshqalar) bilan belgilash mumkin. Bu vazifa NLPda **so'z turkumini aniqlash va teglash** yoki **PoS teglash**dir. PoS teglash morfologik tahlil sanaladi, chunki o'zakka qo'shilgan grammatik shakl orqali so'z turkumini aniqlashga yordam beradi. Misol uchun, morfologik teglash jarayoni - *gan*, *-di* qo'shimchalarining fe'llar bilan qo'llaniladigan morfologik birlik ekanligini anglatadi (istisnolar ham mavjud).

Matn gaplarga bo'lingandan so'ng har bir gapdan *otli birikma* va *fe'lli birikma* kabi tarkibiy so'z birikmalari aniqlanadi. Bu tahlil sintaktik tahlil bo'lib, **lingvistik daraja (sintaksis)** va **annotatsiya vazifasi (sintaktik parsing)** o'rtasida aniq muvofiqlik mavjud. Bugungi kunda lingvistik tahlil bosqichi bo'yicha ko'plab lingvistik nazariyalar mavjud bo'lib, o'zbek tili uchun hozirda keng qo'llaniladigan **so'z birikmalari modeli** yondashuvidan foydalanish maqsadga muvofiq.

Lingvistik tavsifning yuqori darajalari hisoblangan **semantika**, **pragmatika** va **diskursda** turli nazariyalar mavjud. Bugungi kunda o'zbek tili korpusini ushbu darajalar asosida lingvistik teglashdan NLPning amaliy vazifalarida foydalanish uchun aniq konsensus topish qiyin. Shu sababli til korpusining semantik yoki pragmatik tahlili haqida o'zbek tilidagi ilmiy tadqiqotlar deyarli amalga oshirilmagan. Lingvistik tavsifning diskurs darajasiga mos keladigan lingvistik teglash vazifalari haqida ayrim ishlar mavjud. Til korpusi matnlaridan **NER (Named Entity Recognition) obyektlarini aniqlash** – bu matndagi atoqli otlarni aniqlash va ularni teglash vazifasi (shaxs, tashkilot, joy nomlari va boshqalar) hisoblanadi. **Koreferensiyani hal qilish** – matndagi aynan bir obyektни atovchi yoki ularga bir xil obyektlarga ishora qilishini, tegishli ekanligini aniqlash vazifasi.

Xulosa

NLPga asoslangan ilovalar(axborot tizimlari)ni ishlab chiqish uchun NLP tizimni mavjud ma'lumotlarni o'rganishini ta'minlash kerak. Ushbu amalni til korpusi vositasida amalga oshirish mumkin. Til korpusi elektron shaklda taqdim etiladigan *katta hajmli* va *strukturlangan matnlar to'plami* sifatida qaraladi. Til korpusi *yozma* yoki *og'zaki* materialni ifodalab, NLP tizimini mavjud resurslarni o'rganishi uchun lingvistik tahlilni amalga oshirish lozim. Korpus NLP tizimlarining asosidir. Ular AI va mashinali o'rgatish tizimlarini o'rgatish uchun ishlatiladi. Ular axborot hayotiy davrlarini modellashtirish, bashorat qilish uchun keng va xilma-xil ma'lumotlar to'plamini taqdim etadi. Korpus NLP tizimini tabiiy tilni boshqarish va sharhlash uchun tayyorlaydi, bu esa odamlar bilan tabiiy tilda oson muloqot qilish imkonini beradi.

Til korpusi – bu tilning haqiqiy foydalanuvchilari tomonidan ishlab chiqarilgan, so'z, ibora va, umuman, til qanday ishlatilishini tahlil qilish uchun ishlatiladigan juda katta matnlar to'plami. U tilshunos, leksikograf, ijtimoiy olim,

gumanitar fanlar, tabiiy tillarni qayta ishlash bo'yicha mutaxassislar va boshqa ko'plab sohalarda qo'llaniladi. Korpus, shuningdek, dasturiy ta'minotni ishlab chiqishda ishlatiladigan turli til ma'lumotlar bazalarini yaratish uchun ishlatiladi, masalan, *bashoratli klaviatura, imloni tekshirish va tuzatish, matn/nutqni tushunish tizimlari, matndan nutqqa modullar, mashina tarjima tizimlari* va boshqalar.

Til korpusi foydalanuvchilar uchun to'liq foydali bo'lishi uchun uni teglash kerak. Bugungi kunda dunyodagi mashhur korpuslar turli mezonlar bo'yicha teglangan.

Lingvistik teglash – bu qaror qabul qilish maqsadida kompyuterda o'qiladigan ma'lumotlarni uning ma'nosiga bog'lash jarayoni. Texnik jihatdan, u tildagi murakkab naqshlarni aniqlash uchun hissiyotlarni tahlil qilish yoki NLP ilovalari tomonidan ishlatilishi mumkin bo'lgan lingvistik metama'lumotlarga ega matnga izoh berishni o'z ichiga oladi.

So'nggi bir necha yil ichida lingvistik tahlillar onlayn korpuslar orqali amalga oshirilmoqda. Masalan, Sketch Engine 38 milliard so'zni o'z ichiga olgan English Web 2020 (enTenTen20) kabi eng mashhur ochiq korpuslar xizmatini taqdim etadi. Shunisi e'tiborga loyiqki, ijtimoiy tarmoqlar va axborot tizimlaridagi katta hajmdagi strukturlanmagan ma'lumotlar faqat matn emas, balki turli formatlarda (audio, video va boshqalar) ham bo'lishi mumkin. Natijada, lingvistik teglash matn tahlili bilan chegaralanib qolmaydi. Teglar turli formatlarga qo'llanilishi mumkin: *transkripsiya, vaqt belgisi, nutq sharhi, ma'no teglari* va boshqalar.

Tabiiy tilni tushunishga (natural language understanding, NLU) asoslangan ko'plab NLP ilovalarini ishlab chiqish uchun, til korpuslarini shakllantirish va ularni teglash lozim. Shu sababli mualliflar tomondan o'zbek tili korpusni lingvistik teglash turlari hisoblangan, *fonetik, morfologik, sintaktik* va *semantik teglarning* o'zbek tiliga mos to'plami shakllantirildi hamda korpusga qo'llandi.

Strukturlanmagan matnlarni tozalash NLP vazifasi orqali matn tahlil qilish sifati va aniqligini oshirish uchun muhim qadamdir. Matnni tozlash jarayoni: imlo va formatlashdagi nomuvofiqliklarni bartaraf etish orqali matnni kichik harflarga aylantirish bilan birga maxsus belgi, raqam va nomuhim so'zlar kabi ahamiyatsiz (yoki ortiqcha) ma'lumotlarni olib tashlashni o'z ichiga oladi. Matnni tozalash, shuningdek, imlo xatolarni qayta ishlash, so'zlarni o'zak shakliga keltirish (lemmatizatsiya) va matnni kodlash muammolarini hal qilish yechimlarini taklif qiladi. Ijtimoiy tarmoqlar, onlayn servislarda hosil qilinadigan katta hajmdagi ma'lumotlar, matn terishdagi xatolik, nomuvofiq so'zlar, sheva unsurlarini misol sifatida keltirish mumkin. Tozalanmagan matnlar NLP modeli ish jarayoniga salbiy ta'sir o'tkazadi.

Ushbu maqolda matnni teglashning zamonaviy usullari hisoblangan: gap chegaralarni aniqlash, tokenlash, lemmalash, POS teglash, sintaktik tahlil, semantik tahlil, NER obyektlarini aniqlash va koreferensiyani hal qilish kabi NLP vazifalari yechimlari o'zbek tili leksik birliklari misolida keltirildi. Maqolada keltirilgan teglash usullaridan turli NLP ilovalarida, masalan, *matnni tushunish, ma'lumot*

olish, hissiyotlarni tahlil qilish, imloni tekshirish, hujjatlarni umumlashtirish va mashina tarjimasida foydalanish mumkin.

Foydalanilgan adabiyotlar:

1. Hovy, E., & Lavid, J. (2010). Towards a ‘Science’ of Corpus Annotation: A New Methodological Challenge for Corpus Linguistics. *International Journal of Translation*, 22(1).
2. Arista, J. M. (2022). Toward the morpho-syntactic annotation of an old english corpus with universal dependencies. *Revista de Linguistica y Lenguas Aplicadas*, 17. <https://doi.org/10.4995/rlyla.2022.16787>
3. Xusainova Z.Y. NLP: tokenizatsiya, stemming, lemmatizatsiya va nutq qismlarini teglash // “O‘zbek amaliy filologiyasi istiqbollari” mavzusidagi respublika ilmiy-amaliy konferensiyasi – Toshkent, 2022. №.1. – B.154-163.
4. Bi, P. (2018). Handbook of Linguistic Annotation. *Journal of Quantitative Linguistics*. <https://doi.org/10.1080/09296174.2018.1424495>
5. Xusainova Z.Y. BPE algoritmi asosida tokenizatsiya jarayonini amalga oshirish // O‘zbekiston milliy universiteti xabarleri jurnali. Toshkent, 2023. №1/3/1. – B.296-298.
6. Boltayevich, E. B., Mirdjonovna, H. S., & Ilxomovna, A. X. (2023). Methods for Creating a Morphological Analyzer. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 13741 LNCS. https://doi.org/10.1007/978-3-031-27199-1_4
7. Elov B., Hamroyeva Sh., Xudayberganov N., Yodgorov U., Yuldashev A. Pos tagging of Uzbek texts using hidden Markov models (HMM) and Viterbi algorithm. O‘zMU xabarleri. Mirzo Ulug‘bek nomidagi O‘zbekiston Milliy universiteti ilmiy jurnali. 2023 yil Maxsys son.
8. Gries, S. Th., & Berez, A. L. (2017). Linguistic Annotation in/for Corpus Linguistics. In *Handbook of Linguistic Annotation*. https://doi.org/10.1007/978-94-024-0881-2_15
9. Finlayson, M. A., & Erjavec, T. (2017). Overview of Annotation Creation: Processes and Tools. In *Handbook of Linguistic Annotation*. https://doi.org/10.1007/978-94-024-0881-2_5
10. Elov, B.B., Hamroyeva, Sh.M., Abdullayeva, O.X., Husainova, Z.Y., Xudoyberganov, N.U. 2023. “Agglutinatив tillar uchun pos teglash va stemming masalasi (turk, uyg‘ur, o‘zbek tillari misolida)”. O‘zbekiston: til va madaniyat 2: 6-39
11. Arista, J. M. (2022). Toward the morpho-syntactic annotation of an old english corpus with universal dependencies. *Revista de linguistica y Lenguas Aplicadas*, 17. <https://doi.org/10.4995/rlyla.2022.16787>

QUR'ONI KARIM MA'NOLARI TARJIMASI VA TAFSIRINING TIL KORPUSI – DASTURIY TA'MINOTI INTERFEYSI VA QIDIRUV TIZIMIDAN FOYDALANISH

Xolmuxamedov Baxtiyor Farxodovich

bxolmuxamedov@mail.ru

Filologiya fanlari bo'yicha falsafa doktori PhD

Sharof Rashidov nomidagi Samarqand davlat universiteti doktoranti

Annotatsiya. O'zbek tilini o'rganishga bo'lgan qiziqish kundan-kunga oshib borayotgan bir vaqtda uni kompyuter tiliga aylantirish bugungi kunning dolzarb vazifalaridan biri bo'lib qolmoqda. Buning uchun o'zbek tilining milliy korpusini yaratish va uning ustida tadqiqotlar olib borish zarurati mavjudligini ko'rsatadi.

Ushbu maqola o'zbek tilshunosligining yangi yo'nalishlaridan biri – korpus lingvistikasiga mo'ljallangan Qur'oni Karimning Shayx Usmonxon Temurxon Samarqandiy tomonidan qilingan ma'nolar tarjimasi va tafsirining lingvistik tadqiqi va tafsiri-irfon.uz platformasiga bag'ishlanadi. Ushbu platforma o'zbek tilini o'rganuvchilar, o'zbek tili ustida ilmiy tadqiqot olib boruvchilar va butun ommaga mo'ljallangan bepul platforma hisoblanadi.

Abstract. Interest in learning the Uzbek language is growing day by day; turning it into a computer language remains one of the pressing tasks of our time. To do this, it is necessary to create a national corpus of the Uzbek language and conduct research on it.

This article is devoted to one of the new areas of Uzbek linguistics - the linguistic study of the translation and interpretation of the meanings of the Holy Quran by Sheikh Osmankhan Temurkhan Samarkandi, intended for corpus linguistics, and the tavsiri-irfon.uz platform. This platform is a free platform for students and researchers of the Uzbek language as well as for the general public.

Аннотация. Интерес к изучению узбекского языка возрастает с каждым днем, превращение его в компьютерный язык остается одной из актуальных задач современности. Для этого необходимо создать национальный корпус узбекского языка и провести по нему исследования.

Данная статья посвящена одному из новых направлений узбекского языкознания – лингвистическому исследованию перевода и толкования смыслов Священного Корана шейха Османхана Темурхана Самарканди, предназначенного для корпусной лингвистики, и платформы tavsiri-irfon.uz. Эта платформа является бесплатной платформой для изучающих и исследователей узбекского языка а так же для широкой общественности.

Kalit so'zlar: *tilshunoslik, korpus lingvistikasi, milliy til, milliy korpus, kompyuter lingvistikasi, mashina tarjimasi, elektron ma'lumotlar.*

Shiddat bilan rivojlanib borayotgan bugungi zamonda barcha sohalarda bo'lgani kabi, o'zbek tilshunosligida ham ulkan yutuqlarga erishilmoqda. Tilshunoslik yo'nalishlaridan biri bo'lgan amaliy tilshunoslik sohasida ham bir qancha yangi terminlar, yangi yo'nalishlar vujudga kelmoqda. Mana shunday rivojlanib borayotgan yangi yo'nalishlardan biri – korpus lingvistikasi hisoblanadi.

Jahon tilshunosligida korpus lingvistikasi o'tgan asrning 60-yillarida o'rganish obyektiga aylangan. “Ishonarli lingvistik ma'lumotlar katta massivli matnlar majmuasidagina olinishi mumkin” degan qarash o'tgan asrning 60-yillarida R.G.Piatrovskiy tomonidan aytilgan bo'lsa-da, korpus sohasidagi maqsadli tadqiqotlar 40-yillarda Blumfeld, Frays va Bondjerslar tomonidan boshlangan [Blumfeld, 1968: 608]. Braun korpusi (1961-1964) tuzuvchilari Nilson Frensis va Genri Kuchera ilk marta korpus tuzish prinsiplarini ishlab chiqqan. Bu borada ingliz tili banki (1980) loyihasi muallifi Jon Sinkler ishlari ham e'tiborga sazovor [Frensis, Kuchera, 1967]. Rus tilshunosligida V.P.Zaxarov, A.B.Kutuzov, Ye.V.Nedoshivina, V.V.Rikov, V.Plungyanlar korpus, uning turlari, o'ziga xos xususiyati, korpusning ijtimoiy ahamiyati, korpus tuzish tamoyillari borasida tadqiqot olib borishgan [Nedoshivina, 2006:26]. Mualliflik korpuslari O.V.Kukushkina, A.A.Polikarpov, Ye.V.Surovtsevalar tomonidan tadqiq etilgan [Kukushkina, Polikarpov, Surovseva, 2011: 416]. O'zbek tilshunosligida kompyuter lingvistikasi, matnga leksikografik ishlov berish va lingvostatistik tahlil etish borasida birmuncha tadqiqotlar amalga oshirilgan. A.Po'latov, H.Orziqulov, S.Muhamedov, M.Ayimbetov, S.Muhamedova, S.Karimov, G.Jumanazarova, A.Babanarov, D.O'rinboyeva, N.Abdurahmonova, A.Norov va boshqalarning kuzatishlarini ana shunday ishlar sifatida qayd etamiz [Muhammedov, 1980: 24]. Bu tadqiqotlar matnni innovatsion yondashuv – kompyuter lingvistikasi yutuqlari yordamida leksikografik va lingvostatistik tadqiq etishning zamonaviy usullarini tavsiya etganligi bilan dolzarblik kasb etgan bo'lsa-da, ularning hech birida o'zbek tili korpuslarini yaratish masalasi kun tartibiga qo'yilgan emas. Ammo ular milliy tilimiz korpusini yaratish yo'lida qo'yilgan tamal toshlari bo'lganligini ta'kidlash joiz.

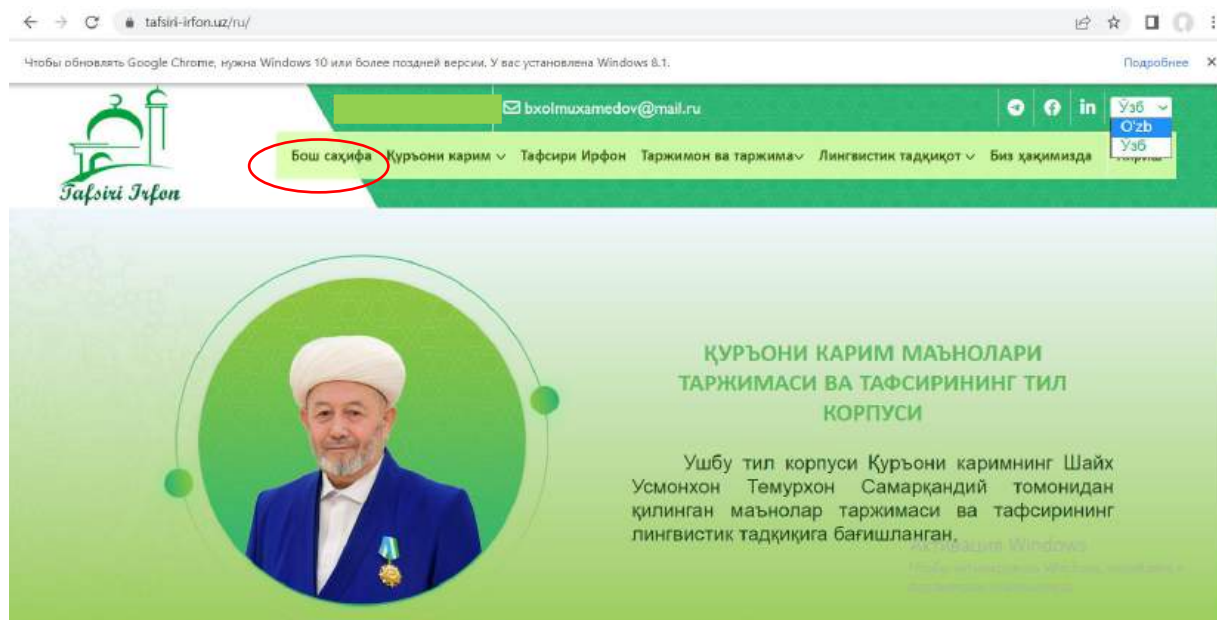
B.Mengliyev, S.Bobojonov, Sh.Hamroyevalar milliy korpus haqida e'lon qilingan dastlabki maqolalarida korpusga online yoki offline tizimda ishlaydigan vosita sifatida qarab, mazkur fikrni bildirishadi: “Korpus – til birliklarining xususiyatlarini aniqlash maqsadida qidiruv dasturiga bo'ysundirilgan matnlar majmuyi, tabiiy tildagi elektron shaklda saqlanadigan yozma yoki og'zaki, kompyuterlashtirilgan qidiruv tizimiga ishlaydigan matnlar jamlanmasi” [Mengliyev, Bobojonov, Hamroyeva, 2018].

Demak, til korpusi – matnlar to'plami, u turli lingvistik so'rovga javob beruvchi qidiruv tizimi. Lingvistik korpus til o'rganish, ma'lum bir tilning xususiyatlarini kuzatish va boshqa lingvistik maqsadlarda qo'llanadi, degan xulosaga kelish mumkin. Bundan tashqari til korpuslari lug'at tuzishda ham asosiy manba bo'lib, u yirik turdagi lug'atlarni dunyoga kelishiga ham xizmat qiladi. Oxirgi

vaqtda yaratilayotgan lug‘atlar asosan korpusga bog‘langan. Buning yagona sabablaridan u yerdagi ma’lumotlarning haqqoniyligi va asoslanganligi bilan birga tilning tabiiyligini o‘zida mujassam etishida. Hukmingizga havola etadigan platforma ham shular jumlasidan.

Platformaning asosiy interfeysi:

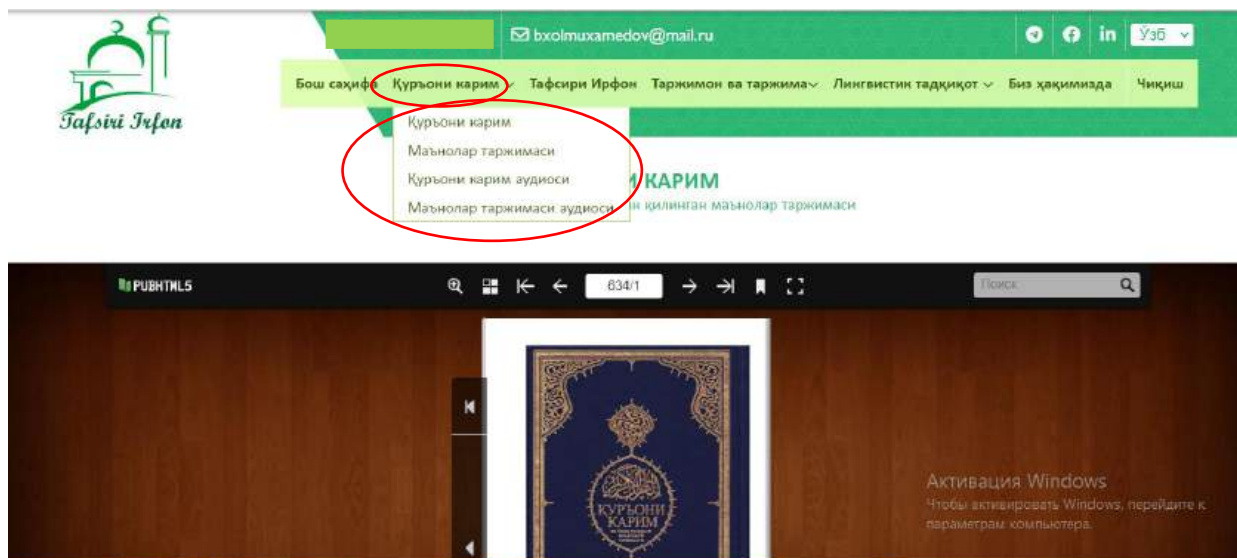
Qur’oni Karimning Shayx Usmonxon Temurxon Samarqandiy tomonidan qilingan ma’nolar tarjimasi va tafsirining lingvistik tadqiqiga bag‘ishlangan platformaga kirish uchun brauzer yordamida <https://tafsiri-irfon.uz/> saytiga kiriladi. Saytga kirgan vaqtda 1-rasmdagi interfeys chiqadi. Bu interfeysning tepa qismi menyular panelidan iborat bo‘lib, “Bosh sahifa”, “Qur’oni karim”, “Tafsiri irfon”, “Tarjimon va tarjima”, “Lingvistik tadqiqot”, “Biz haqimizda” va “Kirish” kabi menyulardan iborat. (1-rasm)



1-rasm.

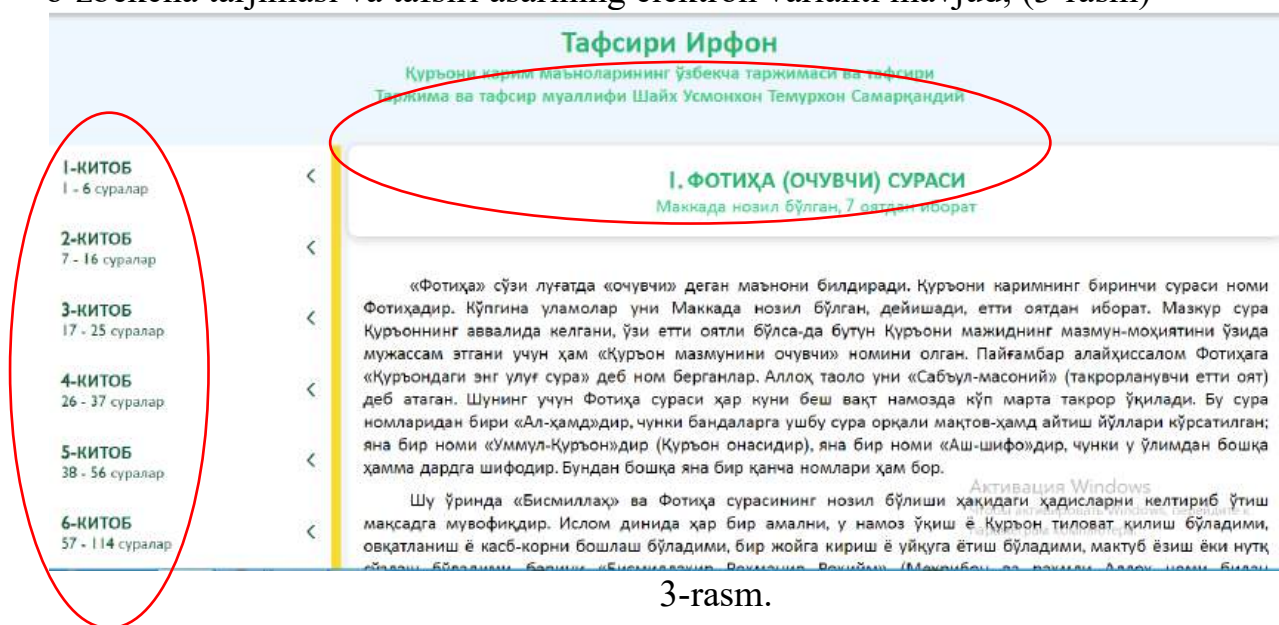
Bosh sahifa – menyusida platformaning umumiy ko‘rinishi, undagi mavjud menyular va ushbu korpusning qaysi asarning lingvistik tadqiqiga bag‘ishlanganligini ko‘rish mumkin;

Qur’oni karim – menyusida Qur’oni Karimning asl nusxasi, Shayx Usmonxon Temurxon Samarqandiy tomonidan tarjima qilingan Qur’oni Karimning o‘zbek tilidagi ma’nolar tarjimasi, Qur’oni Karimning arab tilidagi audiosi va Qur’oni Karimning o‘zbek tilidagi ma’nolar tarjimasi audiosi mavjud; (2-rasm)



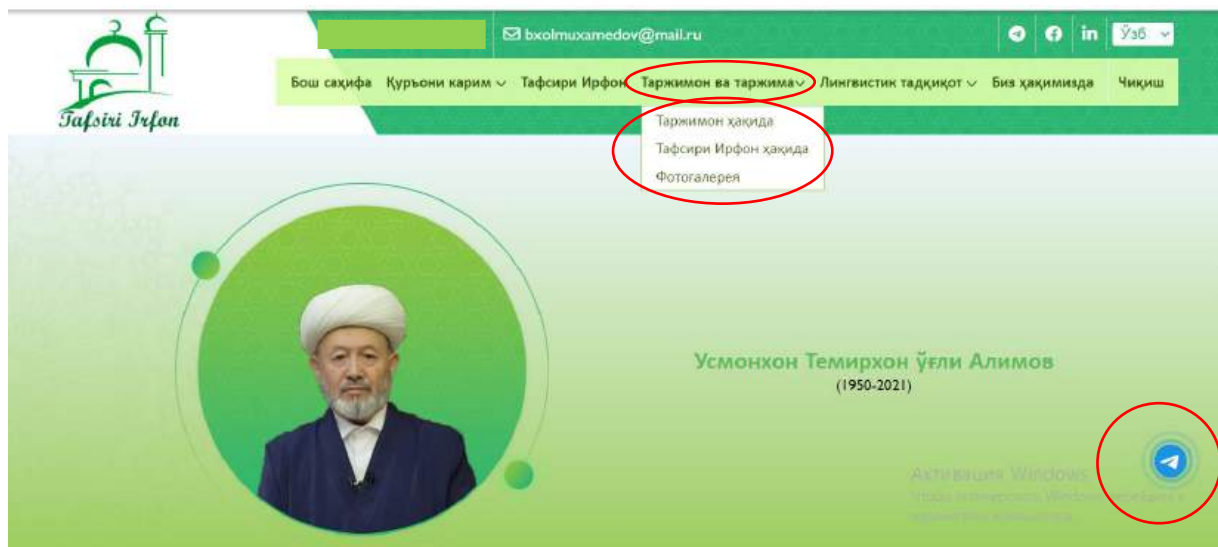
2-rasm.

Tafsiri Irfon – menyusida 6 tomlik Shayx Usmonxon Temurxon Samarqandiy qalamiga mansub “Tafsiri Irfon” – Qur’oni Karim ma’nolarining o‘zbekcha tarjimasi va tafsiri asarining elektron varianti mavjud; (3-rasm)



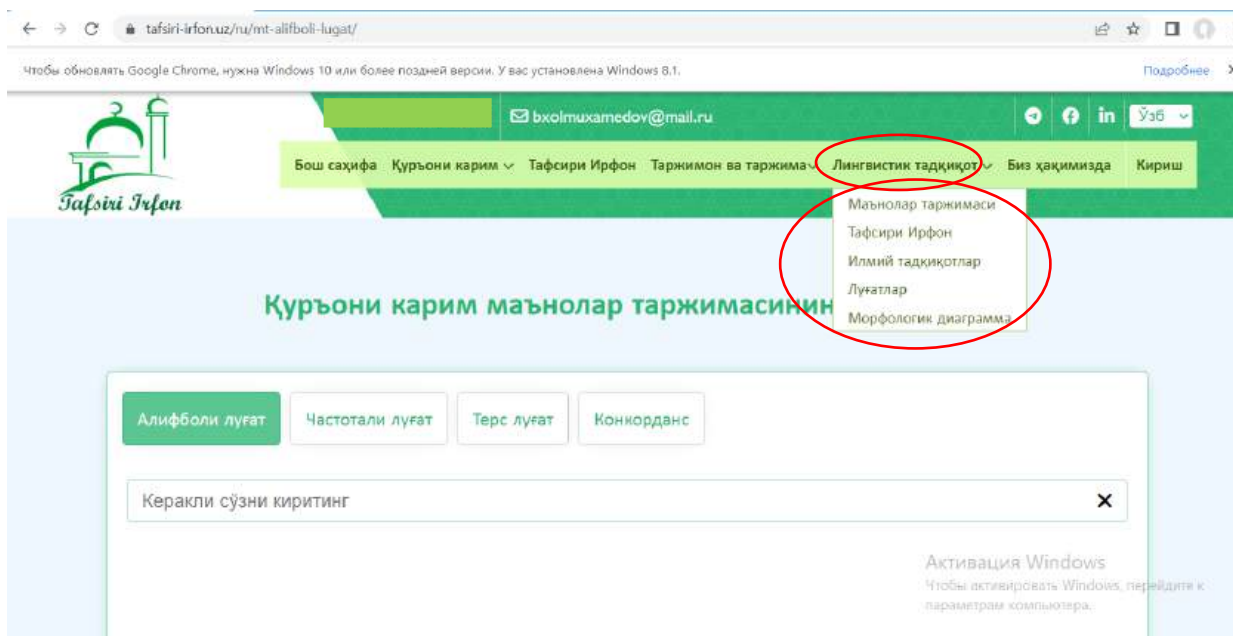
3-rasm.

Tarjimon va tarjima – menyusida Usmonxon Temurxon o‘g‘li Alimov hazratlarining hayot yo‘llari, <https://t.me/ShayxUsmonxonAlimov> telegram havolasi orqali hazratning ilmiy-adabiy meroslari bilan tanishish, “Tafsiri Irfon” asarining yaralish tarixi, hazratning hayot davrlaridagi fotojamlanmalari va <https://t.me/TafsiriIrfonUz> telegram havolasi orqali “Tafsiri Irfon” asarining suxandon Shuhrat Qayumov tomonidan o‘qilgan audiosini tinglash mumkin; 4-rasm



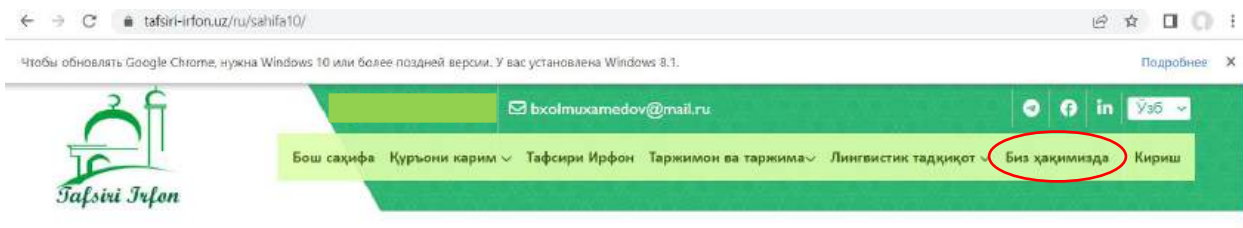
4-rasm

Lingvistik tadqiqot – menyusida Qur’oni Karimning o‘zbek tilidagi ma’nolar tarjimasining alifboli, chastotali, ters lug‘atlari va konkardansi, 6 tomlik “Tafsiri Irfon” asarining alifboli, chastotali, ters lug‘atlari va konkardansi, ushbu korpusni yaratsihda chop etilgan ilmiy maqolalar, Qur’oni Karimning o‘zbek tilidagi ma’nolar tarjimasi va “Tafsiri Irfon” asarining alifboli, chastotali, ters lug‘atlarining elektron varianti mavjud; 5-rasm



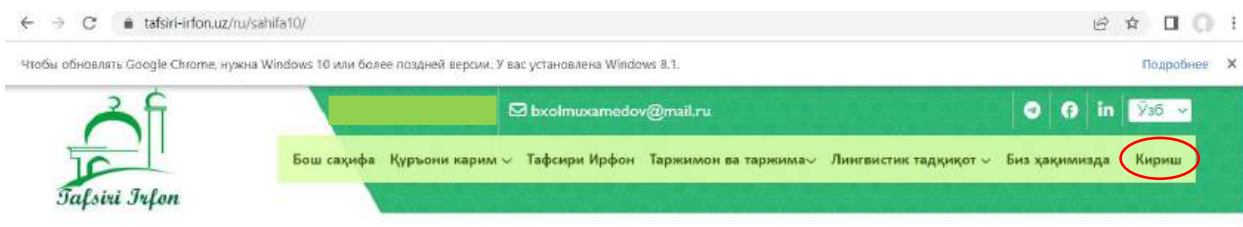
5-rasm.

Biz haqimizda – menyusida ushbu platformaning muallifi haqida qisqacha ma’lumot berilgan; 6-rasm

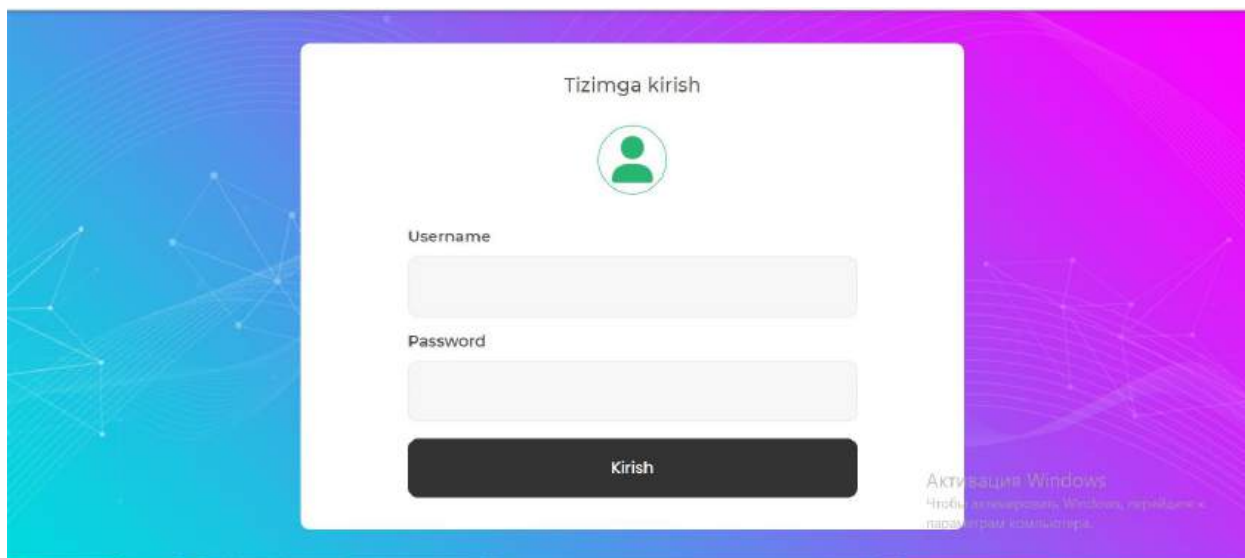


6-rasm.

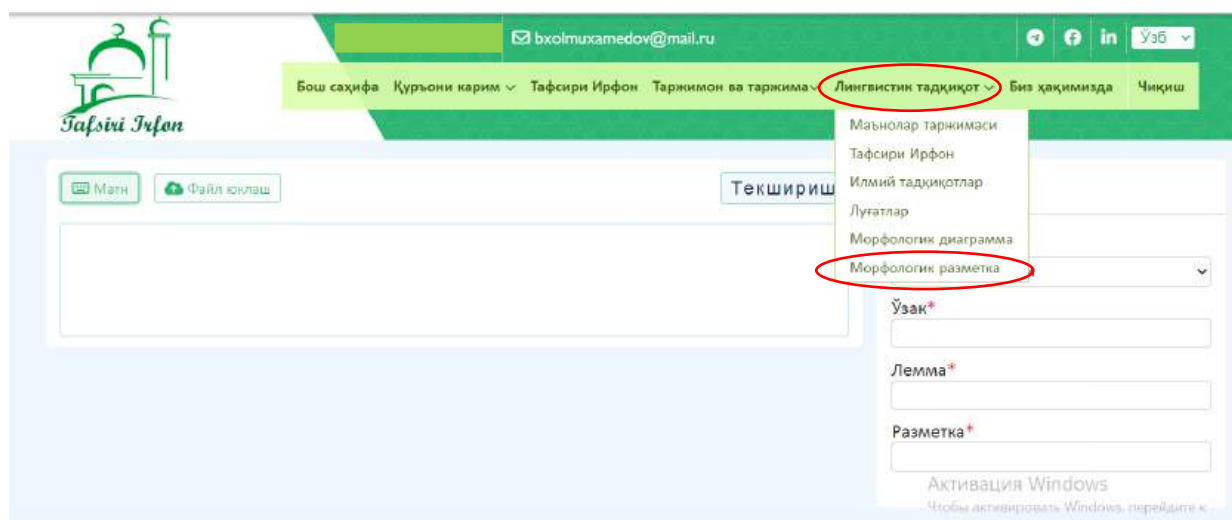
Kirish – menyusi (7-rasm) login, parol yordamida tizim administratori ruxsati bilan ma'lumotlar bazasiga o'tish mumkin (7.1-rasm). Bunda ma'lumotlar bazasini boshqarish (ma'lumlarni qo'shish, tahrirlash, o'chirish) imkoniyatlari mavjud. Tizimga login, parol orqali kirganda *Lingvistik tadqiqot* menyusida yangi *Morfologik razmetka* oynasi (7.2-rasm) ochiladi va shu oynada matnlarni yuklab, ularning ustida amallar bajarish imkoniyati paydo bo'ladi.



7-rasm.



7.1-rasm.



7.2-rasm.

Xulosa o'rnida shuni aytish mumkinki til korpuslari vaqtlar o'tgan sari tilshunoslik borasidagi tadqiqotlar uchun ma'lumotlar bazasiga aylanadi. Korpusga kiritilgan ma'lumotlar ma'lum dasturlar asosida tezlik bilan qayta ishlanish imkoniyatiga ega, shu bilan birga leksikografik manbalarni boyitib boraveradi. Bu jarayonda tilning turli davrlardagi holatini aniq ko'rishimiz, so'zning qo'llanish davrini, chastotasini aniqlashimiz oson kechadi. Bizning Qur'oni Karim ma'nolari tarjimasi va tafsirining til korpusi (tafsiri-irfon.uz) platformasi ham xuddi shunday muhim manba bo'lib xizmat qiladi.

Foydalanilgan adabiyotlar:

1. Блумфилд Л. Язык. – М.: Прогресс, 1968. – 608 с.
2. Френсис Н., Кучера Г. Вычислительный анализ современного американского варианта английского языка. – М., 1967.
3. Кутузов А.Б. Корпусная лингвистика. (Электрон ресурс): Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс) - //lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf.
4. Кукушкина О.В., Поликарпов А.А., Суrowцева Е.В. (Под ред. В.В.Дубичинский) Электронный корпус текстов художественных произведений А.П.Чехова: принципы организации и возможности лексикографического использования// Слово и словарь. Vocabulum et vocabularium. Сборник научных трудов по лексикографии. Вып. 12. – Харьков-Клагенфурт, 2011. – 416 с.
5. Менглиев Б., Бобожонов С., Хамроева Ш. Ўзбек тили миллий корпуси. 2018 йил, 26 апрель, <http://marifat.uz/marifat/ruknlar/fan/1241.htm>
6. Муҳаммедов С.А. Статистический анализ лексико-морфологической структуры узбекских газетных текстов: Автореф. диссер. ... канд. филол. наук. – Ташкент, 1980. – 24 с.



7. Недошивина Е.В. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. Учебно-методическое пособие. – Санкт-Петербург. 2006. – 26 с.

8. [Xolmuxamedov B.F.](#) Korpusga kiritilgan kavsar surasi tadqiqi// “O‘zbek tilining milliy korpusi: muammo va vazifalar” mavzusidagi xalqaro ilmiy-amaliy konferensiya. – Toshkent, 31.05.2022. –B.229-234.

9. [Xolmuxamedov B.F.](#) Milliy korpusga bo‘lgan ehtiyoj// BuxDU ilmiy axborotnomasi. – Buxoro, 2024. – № 1. –B.202-207.

10. <http://uzbekcorpora.uz/>

11. <https://tafsiri-irfon.uz/>

MUALLIFLIK KORPUSI KENG QAMROVLI QIDIRUV TIZIMI

Akramova Shohista Islom qizi

akramovashohista88@gmail.com

ToshDO‘TAU doktoranti

Annotatsiya. Ushbu maqolada mualliflik korpuslari muallif asarlari tilini to‘liq va aniq ko‘rasata olishi bilan boshqa axborot banklaridan ajralib turishi haqida ma’lumotlar keltirilgan. Korpusning dunyodagi ko‘plab sohalarda ma’lumotlarning keng va to‘liq manbayi vazifasini o‘tay olishi bilan amaliy ahamiyati oshib borayotgani va korpuslarsiz raqamli texnologiyalarga asoslangan axborot asrida ta’lim jarayonini tasavvur qilish juda qiyinligi izohlangan.

Abstract. This article provides information on how author corpora differ from other information banks by being able to fully and clearly display the language of author's works. It is explained that the practical importance of the corpus is increasing as it can serve as a wide and complete source of information in many areas of the world, and it is very difficult to imagine the educational process in the information age based on digital technologies without the corpus

Аннотация. В данной статье представлена информация о том, чем авторские корпуса отличаются от других информационных банков возможностью полного и четкого отображения языка авторских произведений. Поясняется, что практическая значимость корпуса возрастает, поскольку он может служить широким и полным источником информации во многих регионах мира, а образовательный процесс в информационную эпоху, основанный на цифровых технологиях, очень сложно представить без корпус

Kalit so‘zlar: *korpus lingvistikasi, mualliflik korpusi, ta’limiy ahamiyat, milliy korpus.*

Axborot asriga kelib, korpus lingvistikasi jahon tilshunosligining shiddat bilan rivojlanayotgan sohasiga, korpusning o‘zi esa tilshunoslikning ish quroliga aylandi. Tilning milliy korpusi axborot manbayi sifatida quyidagi qulayliklarni yaratdi:

1) korpuslashgan tilda yaratilgan og‘zaki, yozma yodgorliklar, milliy, madaniy meros namunalari elektron shaklda internet tarmog‘idan joy oladi;

2) korpus tabiiy tilning elektron shakldagi, qidiruv dasturiga joylashtirilgan matnlar yig‘indisi bo‘lganligidan bir marta tuzilib, mukammal razmetkalangan korpus lingvistik tadqiqotlar samaradorligini ta’minlashda baza vazifasini bajaradi;

3) korpus elektron kutubxona, lug‘at, grammatikalar yaratish uchun asos bo‘ladi. U keng ko‘lamli bo‘lganligi uchun ma’lumotning o‘ziga xosligini kafolatlab, til hodisalarining barcha qirralarini to‘liq namoyish etishni kafolatlaydi;

4) turli ma’lumotlar til korpusida o‘zining tabiiy kontekstual shaklida joylashadi, bu esa ularni har tomonlama, chuqur, obyektiv o‘rganishga asos bo‘ladi;

5) korpus tilni tadqiq etish (soʻzning oʻzgarishi, istorizm, neologizm vujudga kelishi, maʼno torayishi, maʼno kengayishi; yangi frazeologizmlarning paydo boʻlishini kuzatish), til oʻrganish, lugʻat tuzishda eng zamonaviy, keng imkoniyatli dasturlashtirilgan tizimdir

Korpus lingvistikasi XX asrning 60-yillarida kompyuter lingvistikasining boʻlimi sifatida shakllangan soha boʻlib, kompyuter texnologiyalaridan foydalangan holda lingvistik korpuslarni (matnlar korpuslarini) qurish va ulardan foydalanishning umumiy tamoyillarini ishlab chiqish bilan shugʻullanadi. Lingvistik (leksik) korpus yoxud til (matnlar) korpusi muayyan lingvistik muammolarni hal qilishga moʻljallangan, maxsus saralangan va teglangan elektron shakldagi matnlar majmuasi hisoblanadi va pragmatik ahamiyatga ega yirik elektron baza sanaladi.[Abduraxmonova, 2021]

Oʻtgan asrning 40-yillarida Blumfeld, Fritz va Bonderslar tomonidan korpus lingvistikasi boʻyicha tadqiqotlar boshlangan boʻlib, Nils Frensis va Genri Kuchera birinchi boʻlib korpus yaratish tamoyillarini ishlab chiqdi. Korpus lingvistikasi termini birinchi marta 1963-yilda ingliz tilidagi matnlar asosida yaratilgan ilk korpus – Braun korpusining yaratilishi bilan qoʻllanila boshlandi. Rossiyada bu termin 1996-yilda Sidney Grinbaumning maʼruzalari orqali maʼlum boʻldi. XX asrning 90-yillariga kelib korpus lingvistikasiga oid ilmiy tushunchalar mukammalroq shakllandi, bu yangi soha haqida manbalarda bir-biriga yaqin taʼriflar berila boshlandi. V.Zaxarov til korpusiga elektron shakldagi katta hajmda tizimlashtirilgan (muayyan strukturaga ega), lingvistik izohlar – maxsus elektron belgilar qoʻyilgan, filologik jihatdan keng qamrovli va muayyan lingvistik masalalarni hal etishga moʻljallangan til maʼlumotlari majmuyidir, deya taʼrif beradi. [Ataboyev, 2019]

N.Kozlova esa korpus oddiy matnlardan raqamlashtirilganligi bilan farqlanib, yozma va ogʻzaki shakldagi matnlar yigʻindisi hisoblanishi hamda maxsus lingvistik izohlar qoʻyilgan holatda kompyuterga joylashtirilgan matnlar yigʻindisi ekanligi hamda maʼlumotlarni tartibga soladigan annotatsiyalardan iboratligini aytadi. Korpus kiritilgan matnlarning oʻziga xos jihatlari toʻgʻrisidagi muhim qoʻshimcha maʼlumotlarni annotatsiya yoki lingvistik izoh koʻrinishida saqlaydi.

Lingvistik izoh (rus tilida pazметка, ingliz tilida tag) – korpusning asosiy bosh maʼlumotnomasi boʻlib, u korpusni zamonaviy internet tarmogʻida mavjud boʻlgan oddiy matnlar toʻplami (yoki kutubxonalar)dan ajratib turadigan, masalan, Ziyonet kutubxonasi, Hilol virtual kutubxonasi, Maksim Morshkov kutubxonasi yoki rus virtual kutubxonasi kabilardan farqlanadigan matnlardir. Korpuslar umumiylikiga koʻra umumiy va yakka muallifli boʻladi. Umumiy korpuslar muayyan tilning bir yoki bir necha masalalari boʻyicha yaratiladi. Korpuslar umumiylikiga koʻra umumiy va yakka muallifli boʻladi. Umumiy korpuslar muayyan tilning bir yoki bir necha masalalari boʻyicha yaratiladi. Unda matnlarning grammatik yoxud semantik teglari berilishi, parallel tarjimai taqdim etilishi, matnlar manbasi haqida metamaʼlumotlar koʻrsatilishi mumkin. Yakka muallifli korpuslar muayyan



ijodkorga mansub materiallarni qamrab olishi hamda jahon ilmida korpusshunosligida yaratilgan mualliflik korpuslari o'rganildi. Mualliflik korpusi muayyan bir muallif qalamiga mansub matnlar majmuasidan iborat tizim bo'lib, asosan, yozuvchining yozish uslubi, asarlaridagi til, matn xususiyatlarini axborot texnologiyalari yordamida semantik tasniflab, tadqiq etishga asoslangan elektron shakldagi ma'lumotlar bazasi sanaladi. Mualliflik korpusining boshqa korpuslardan farqi uning bazasida muayyan bir muallif asarlari jamlanadi, matnlar qayta ishlanadi, grammatik va semantik teglanadi, qidiruv tizimi orqali muallifga tegishli materiallardan zaruriy ma'lumotlar, lingvistik ifodalar aniqlanadi, matnlar asosida statistik ma'lumotlar taqdim etiladi, metama'lumotlar aniq beriladi. Shunday qilib mualliflik korpusi keng imkoniyatli qidiruv tizimiga ega bo'lib, muallif ijod qilgan barcha tur va janrdagi asarlar qamrab olinadigan, maxsus parametrlar asosida ham qidiruv imkoniyatiga ega, hajmi chegaralanmagan, muallif va uning ijodi bilan bog'liq manbalarni qulay va tezkor olish mumkin bo'lgan elektron ma'lumotlar bazasi hisoblanadi. Jahon tilshunosligida hozirgi kunga qadar mualliflik manbalarini yaratish borasidagi tadqiqotlar amalga oshirilmoqda. Xususan, 1956-1961-yillarda Pushkin tilining lug'ati, shuningdek, alohida asarlarning lug'atlari, masalan, 1978-yilda Lev Tolstoyning “Urush va tinchlik” romanining chastotali lug'ati nashr etildi. 1981-yilda R.Qo'ng'urov va S.Karimovlar tomonidan “Zulfiya poeziyasining lug'ati. Konkordans” tuzilgan. 1986-yilda esa S.Rizayev va N.Bo'ronovlar tomonidan “A.Qahhor “Sinchalak” povesti tilining chastotali lug'ati” yaratilgan. Bunday tadqiqotlar, shubhasiz, dolzarb hisoblanib, hozirgi kunda yozuvchilarning zamonaviy mualliflik korpuslarini yaratish borasidagi izlanishlarning shakllanishiga zamin bo'lgan.

Umuman, mualliflik korpuslarini yaratish asosi mualliflik lug'atlarining tuzilishi bilan bog'liq va MK yaratish ishlari shakllanishi o'z tendensiyasiga ega. Buni quyidagicha aks ettirish mumkin:

Rus adiblarining mualliflik korpuslari bo'yicha izlanishlar nafaqat rus tilshunosligida, balki jahon korpusshunosligida shiddat bilan rivojlanib, uning taraqqiyoti natijasida juda ko'plab ijodkorlarning elektron lug'at va korpuslari yaratildi. Mualliflik korpuslarining yaratilishi bir qancha yozuvchi, shoir tili konkordansini tuzish, milliy korpusda ularning asarlari bazasini yaratish imkonini berdi. Bu borada rus korpus lingvistikasida amalga oshirilgan ishlar va erishilgan yutuqlar diqqatga sazovor.

Mualliflik korpusshunosligining asosiy vazifasi va mualliflik korpuslaridan foydalanish sabablari quyidagicha izohlandi:

- muayyan ijodkor shaxsini o'rganish;
- adib uslubini tadqiq etish;
- muayyan shoir ijodini lingvopoetik tahlil qilish;
- ijodkorning so'z qo'llash imkoniyati va mahoratini tadqiq etish;
- mualliflik lug'atlarini yaratish;
- mualliflik iboralarini jamlash;

– MKdagi ijodkor shaxsi va uslubini namoyon etuvchi parametrlar orqali anonim asarlar muallifini topish;

– mualliflik parafraza, parema, hikmatlarini jamlash; tasviriy ifodalarning qo‘llanilish ko‘lamini ijodkor kontekstidan aniqlash mumkin bo‘ladi. Ma‘lum bir mualliflik korpusining tarixiy, lingvistik, ta‘limiy, ijtimoiy, tarbiyaviy ahamiyatlari yoritiladi. Mualliflik korpusida muallif tilini to‘liq, batafsil, obyektiv ko‘rsatish imkoniyati bor, shuning uchun bunday korpuslar boshqa axborot bankidan afzalligi bilan ajralib turadi. Korpus turli tadqiqotlar uchun asos, manba, vosita vazifasini o‘tay oladi. Bunday korpuslarning afzalligi yana shundaki, ular yordamida nafaqat biron bir so‘z yoki jumlaning, balki butun boshli asar tillarini bilish imkoniyati mavjud. Mualliflik korpuslaridagi ma‘lumotlarning ilmiy manbalar asosida tahrir qilinganligi, unda berilgan ma‘lumotlarning aniqliligi va ishonchliligini kafolatlaydi hamda lingvistik hodisalarning butun spektrini to‘liq aks ettirishi bilan ularni har tomonlama va obyektiv o‘rganishga imkon beradi. MK ta‘lim jarayonida muallif asarlarining ta‘limiy, ijtimoiy, tarixiy, lingvodidaktik, ahamiyatini yoritib berishda, muallif asarlarida qo‘llanilgan so‘zlarning dialektal xususiyatlarini tasniflashda, Milliy tilning semantik xususiyatlarini o‘rganishda muhim sanaladi. XV asr til uslubini bugungi tilga qiyosan o‘quvchilarga o‘rgatish mumkin. Badiiy matnlarning statistik tahlili natijasida matnda tez-tez qo‘llaniluvchi til birliklari (otlar, sifatlar, kalit so‘zlar, fe‘llar, grammatik shakllar, jumla qurilishi, bir so‘z bilan aytganda, yozuvchining uslubini ko‘rsatuvchi vositalar) lingvostatistik tahlil yordamida aniqlanadi. Turli matnlardan olingan dalillarning qiyosiy tahlili bizga o‘sha matnning mazmun-mundarijasini, matn tuzilgan davrni, dalillarni argumentlash xarakterini va hatto mualliflikni aniqlashga imkon beradi. Mualliflik korpuslari o‘quvchilarning izlanuvchanlik salohiyatini o‘stirib, kichik tadqiqotlarni ishonchli dalillar asosida bajarishga imkon beradi. Ta‘kidlash joizki, mualliflik korpuslarining ijtimoiy ahamiyati – keng qamrovli hisoblanib, undan lingvistik, etnopsixolingvistik tadqiqotlarda, ona tili, adabiyot, xorijiy til ta‘limida, matnga avtomatik ishlov berish, tarjima dasturlarini tuzishda foydalanish mumkin. Mualliflik korpusi orqali Ma‘naviy merosimizni asrab-avaylash, ajdodlar ruhini ulug‘lash ko‘nikmalarini shakllantirish, mutolaa madaniyatini oshirish natijasida teran fikrga ega bo‘lish, tafakkurni kengaytirishga erishiladi.

Mualliflik korpuslari ta‘lim jarayonida muallif asarlarining ta‘limiy, ijtimoiy, tarixiy, lingvodidaktik, ahamiyatini yoritib berishda, muallif asarlarida qo‘llanilgan so‘zlarning dialektal xususiyatlarini tasniflashda, milliy tilning semantik xususiyatlarini o‘rganishda muhim sanaladi. Mualliflik korpusi materialidan ta‘limiy maqsadda foydalanish o‘ta ahamiyatli masala hisoblanadi. Darhaqiqat, til korpuslari ta‘lim jarayonida eng muhim o‘quv quroli. Sababi mavzulashtirilgan matn ta‘lim jarayonida o‘quv materialini tayyorlashda qulay axborot manbayi vazifasini bajaradi. Lingvodidaktikada mualliflik korpusining ahamiyati shundan iboratki, o‘qituvchi uchun korpus tengsiz xazina, undan o‘quv mashg‘uloti uchun serqirra, mazmunli o‘quv materialini tayyorlashi mumkin.

Mualliflik korpuslari o'quvchilarning izlanuvchanlik salohiyatini o'stirib, kichik tadqiqotlarni ishonchli dalillar asosida bajarishga imkoniyat beradi. Ta'kidlash joizki, mualliflik korpuslarining ijtimoiy ahamiyati – keng qamrovli hisoblanib, undan lingvistik, etnopsixolingvistik tadqiqotlarda, ona tili, adabiyot, xorijiy til ta'limida, matnga avtomatik ishlov berish, tarjima dasturlarini tuzishda foydalanish mumkin. Mualliflik korpuslaridan – “ish quroli – so‘z” bo‘lgan barcha soha vakillari: tilshunos, tarjimon, o‘qituvchi, dasturchi, jurnalist, muharrir kundalik faoliyati jarayonida keng foydalanishlari mumkin[Hamroyeva, 2020].

Foydalanilgan adabiyotlar:

1. Abduraxmonova N. O‘zbek tili elektron korpusining kompyuter modellari. Monografiya/ Abduraxmonova N. –Toshkent, 2021.
2. Ataboyev N.B. Korpus lingvistikasining asosiy xususiyatlari // Ўзбекистонда хорижий тиллар илмий-методик электрон журнал journal.fledu.uz №2/2019
3. Hamroyeva Sh.M. O‘zbek tili mualliflik korpusini tuzishning lingvistik asoslari. Monografiya. – Toshkent, 2020.

TIL KORPUSI MATNLARI CHASTOTASI, KONKORDANS VA KWIC

Yuldashev Aziz

yuldashevaziz@navoiy-uni.uz

ToshDO'TAU o'qituvchisi

Annotatsiya. NLPga asoslangan ilovalar(axborot tizimlari)ni ishlab chiqish uchun NLP tizimni mavjud ma'lumotlarni o'rganishini ta'minlash kerak. Ushbu amalni til korpusi vositasida amalga oshirish mumkin. Korpus matnlarini annotatsiyalash yoki teglash orqali so'z, ibora, so'z birikmasi shu kabi leksik birliklarni izohlash kabi matnga teg (izoh) qo'shish va ularni belgilashdan iborat. Til korpusi foydalanuvchilar uchun to'liq foydali bo'lishi uchun uni teglash kerak. Bugungi kunda dunyodagi mashur korpuslar turli mezonlar bo'yicha teglangan. Lingvistik teglash – bu qaror qabul qilish maqsadida kompyuterda o'qiladigan ma'lumotlarni uning ma'nosiga bog'lash jarayoni. Texnik jihatdan, u tildagi murakkab naqshlarni aniqlash uchun hissiyotlarni tahlil qilish yoki NLP ilovalari tomonidan ishlatilishi mumkin bo'lgan lingvistik metama'lumotlarga ega matnga izoh berishni o'z ichiga oladi.

Abstract. To develop applications (information systems) based on NLP, it is necessary to ensure that the NLP system learns the available data. This can be done using a language corpus. Annotating or tagging corpus texts consists of adding a tag (explanation) to the text and defining them, such as explaining words, phrases, phrases, and similar lexical units. In order for a language corpus to be fully useful to users, it needs to be tagged. Today, the world's most famous corpora are tagged according to various criteria. Linguistic tagging is the process of linking computer-readable information to its meaning for decision-making purposes. Technically, it involves annotating text with linguistic metadata that can be used by sentiment analysis or NLP applications to identify complex patterns in language.

Аннотация. Для разработки приложений (информационных систем) на основе НЛП необходимо обеспечить, чтобы система НЛП обучалась имеющимся данным. Это можно сделать с помощью языкового корпуса. Аннотирование или разметка корпусных текстов заключается в добавлении к тексту тега (пояснения) и его определении, например пояснении слов, словосочетаний, словосочетаний и подобных лексических единиц. Чтобы языковой корпус был полностью полезен пользователям, его необходимо пометить. Сегодня самые известные корпорации в мире маркируются по различным критериям. Лингвистическая разметка — это процесс связывания читаемой компьютером информации с ее значением для целей принятия решений. Технически это включает в себя аннотирование текста лингвистическими метаданными, которые могут использоваться анализом

настроений или приложениями НЛП для выявления сложных языковых шаблонов.

Kalit so‘zlar: *Til korpusi, matnni qayta ishlash, lingvistik tahlil, mashinali o‘qitish modellari, korpus turlari, teglangan korpuslar, POS teg.*

Kirish

So‘nggi bir necha yil ichida lingvistik tahlillar onlayn korpuslar orqali amalga oshirilmoqda. Masalan, Sketch Engine 38 milliard so‘zni o‘z ichiga olgan English Web 2020 (enTenTen20) kabi eng mashhur ochiq korpuslar xizmatini taqdim etadi. Shunisi e‘tiborga loyiqki, ijtimoiy tarmoqlar va axborot tizimlaridagi katta hajmdagi strukturlanmagan ma‘lumotlar faqat matn emas, balki turli formatlarda (audio, video va boshqalar) ham bo‘lishi mumkin. Natijada, lingvistik teglash matn tahlili bilan chegaralanib qolmaydi. Teglar turli formatlarga qo‘llanilishi mumkin: *transkripsiya, vaqt belgisi, nutq sharhi, ma‘no teglari* va boshqalar [Bi, 2018].

Tabiiy tilni tushunishga (natural language understanding, NLU) asoslangan ko‘plab NLP ilovalarini ishlab chiqish uchun, til korpuslarini shakllantirish va ularni teglash lozim [Bender, Lascarides, 2019]. Shu sababli mualliflar tomonidan o‘zbek tili korpusini lingvistik teglash turlari hisoblangan, *fonetik, morfologik, sintaktik va semantik teglarning* o‘zbek tiliga mos to‘plami shakllantirildi hamda korpusda qo‘llandi.

Strukturlanmagan matnlarni tozalash NLP vazifasi orqali matn tahlil qilish sifati va aniqligini oshirish uchun muhim qadamdir. Matnni tozlash jarayoni: imlo va formatlashdagi nomuvofiqliklarni bartaraf etish orqali matnni kichik harflarga aylantirish bilan birga maxsus belgi, raqam va nomuhim so‘zlar kabi ahamiyatsiz (yoki ortiqcha) ma‘lumotlarni olib tashlashni o‘z ichiga oladi. Matnni tozalash, shuningdek, imlo xatolarni qayta ishlash, so‘zlarni o‘zak shakliga keltirish (lemmatizatsiya) va matnni kodlash muammolarini hal qilish yechimlarini taklif qiladi. Ijtimoiy tarmoqlar, onlayn servislarda hosil qilinadigan katta hajmdagi ma‘lumotlar, matn terishdagi xatolik, nomuvofiq so‘zlar, sheva unsurlarini misol sifatida keltirish mumkin. Tozalanmagan matnlar NLP modeli ish jarayoniga salbiy ta‘sir o‘tkazadi.

Korpus chastotasi

Tilshunoslik – bu statistika va matematika sohasi bilan uzviy bog‘liq bo‘lgan fan. Matematik tilshunoslik, kompyuter lingvistikasi, korpus lingvistikasi, amaliy tilshunoslik, sud lingvistikasi, stilometriya va boshqalar tabiiy til korpusidan olingan turli statistik natijalarni talab qiladi. Tilning turli xossalari haqidagi statistik ma‘lumotlarni yetarli darajada bilmaslik natijasida lingvistik ma‘lumotlar bilan ishlashda ham, kuzatishda ham xatoga yo‘l qo‘yilishi mumkin.

Til korpusi ham miqdoriy, ham sifat jihatidan tahlil qilinishi mumkin.

Miqdoriy tahlilda tilning turli lingvistik xususiyatlari tasniflanadi va murakkabroq statistik modellar ishlab chiqiladi. Statistik modellar asosida qaysi

hodisalar til yoki turli xil xatti-harakatlarning haqiqiy aks etishi mumkinligini va qaysilari shunchaki tasodifiy hodisalar ekanligini aniqlash imkonini beradi.

Sifat tahlili miqdoriy tahlil orqali korpusda kuzatilgan hodisalarning to'liq va batafsil ta'rif berishga qaratilgan. Korpusning sifat tahlili aniq xulosalarni chiqarishga imkon beradi. Chunki bu holda ma'lumotlarni cheklangan tasniflar to'plamiga kiritish shart emas. Ham miqdoriy, ham sifat tahlillarini amalga oshirish natijasida korpusni tahlil qilish samaradorligi oshadi. Biroq miqdoriy tahlil evaziga olingan natijalar sifat tahlilidan olingan natijalarga qaraganda samaradorligi pastroq hisoblanadi.

Bugungi kunda korpus ustida miqdoriy tahlilni amal oshirishning quyida keltirilgan statistik yondashuvlari mavjud:

- *oddiy tavsiflovchi statistik yondashuv*;
- *inferensial statistik yondashuv*;
- *baholovchi statistik yondashuv*.

Oddiy tavsiflovchi statistik yondashuv kuzatilgan ma'lumotlarning eng muhim xususiyatlarini umumlashtirish imkonini beradi. *Inferensial statistik yondashuv* savollarga javob berish yoki gipotezani shakllantirish uchun tavsiflovchi statistik yondashuvdan olingan ma'lumotlardan foydalanadi. *Baholovchi statistik yondashuv* gipoteza ma'lumotlardagi dalillar bilan tasdiqlanganligini va matematik model yoki ma'lumotlarning nazariy taqsimoti haqiqatga qanday bog'liqligini tekshirishda ahamiyatli.

Korpus ustida amalga oshirilgan miqdoriy tahlil natijalarini taqqoslash va yashirin shablonlarni aniqlash uchun quyida keltirilgan ko'p o'lchovli statistik usullardan foydalaniladi:

- *omillar tahlili (Factor Analysis)*;
- *ko'p o'lchovli masshtablash (Multidimensional Scaling)*;
- *klaster tahlili (Cluster Analysis)*;
- *log-linear modellari (Log-linear Models)*.

So'zlarni tartiblash/saralash

Korpusda saqlangan so'zlarni ikki usulda tartiblash mumkin. **Raqamli saralash** jarayoni miqdoriy ma'lumotlar bilan ishlashning eng oddiy usuli hisoblanadi. Ushbu usul orqali saralashda obyektlar ma'lum bir sxema bo'yicha tasniflanadi hamda sxemadagi har bir sinfga tegishli bo'lgan *matnlar ichidagi elementlarning soni bo'yicha arifmetik hisoblash* amalga oshiriladi. Oddiy chastotalarni hisoblashda mavjud bo'lgan ma'lumotlar alifbo tartibida yoki raqamli tartibda foydalanuvchiga taqdim etiladi. Hosil qilingan ro'yxat foydalanuvchi talabiga ko'ra *o'sish* yoki *kamayish* tartibida joylashtirilishi mumkin. Matn tahlili jarayonida har xil element (so'z, token, so'z birikmasi, lemma)ning necha martadan uchrashi aniqlanadi.

Tahlil natijasida **so'zlarning chastota ro'yxati** hosil qilinadi. Ro'yxat o'rganib chiqish matn tuzilishi haqida tasavvur hosil qiladi; shunga mos ravishda tahlil rejalashtiriladi. Shuningdek, alifbo tartibida tartiblangan so'zlar ro'yxati oddiy umumiy havolalar uchun ishlatilishi mumkin. Alfavit tartibidagi chastotalar ro'yxati o'rganish obyekti sifatida ahamiyatli, chunki u ko'pincha tekshirilishi kerak bo'lgan gipotezalarni shakllantirish va qaror qabul qilishga yordam beradi.

O'zbek tili korpusida chastotani hisoblashdan oldin korpusda ishlatiladigan *belgi, so'z, idioma, ibora, fraza va gaplar* bilan ishlash jarayoni haqida mulohaza yuritish kerak. Ushbu amal bizni o'zbek tilining turli lingvistik xususiyatlari haqida noto'g'ri kuzatish va xulosalardan saqlaydi.

Konkordans

Konkordans tuzish jarayoni korpusda ishlatiladigan so'zlarni indekslashni anglatadi. Ushbu jarayonda har bir so'z matn(lar)da kelgan joyiga qarab indekslanadi. Konkordans tuzish jarayoni til modellarini qurishda muhim ahamiyat kasd etadi va tabiiy tildagi qoliplarni aniqlash imkonini beradi.

Bugungi kunda korpusni tahlil qilish uchun konkordans dasturlari ishlab chiqilgan. Masalan, korpusni tartiblash va chastotasini aniqlash uchun **MonoConc**, **Conc**; matnlarni parallel qayta ishlash uchun **ParaConc**, matnlarni qayta ishlash va saralash uchun **FreeText** kabi dasturiy ta'minotlar mavjud. Konkordans funksiyasi, asosan, leksikografik tadqiqotlar va til o'rgatishda qo'llanadi. Shuningdek, undan bir va ko'p so'zli satrlar, so'z, ibora, idioma, maqol va boshqa leksik birliklarni qidirishda foydalanish mumkin.

Konkordans jarayoni **leksik, semantik, sintaktik, matn** modellarini tadqiq qilish va matnlarning uslubiy qonuniyatlarini o'rganishda ishlatiladi. Bu ko'p ma'noli, polifunksional va polisemantik so'zlar va morfemalarni tekshirishda muhim vosita bo'lib xizmat qiladi.

Leksik kollokatsiya (Erkin birikma)

So'zlarni qo'shish usuli matndagi so'zlarning o'rni va qurshovini tushunishga yordam beradi. Korpusdan aniqlangan so'z birikmalari tilda mavjud ba'zi an'anaviy lingvistik tavsif va farazlardan farq qiladi. Leksikografiya bo'yicha amalga oshirilgan tadqiqotlar shuni ko'rsatadiki, ko'p ishlatiladigan faol so'zlar ma'nosi bizning ongimizga birinchi kelgan va lug'atlarda keltirilgan ma'no bilan aynan emas. Demak, korpusdan aniqlangan so'z birikmalari o'rtasida *leksik kollokatsiyani* aniqlashga yordam beradi. Bunda ikki so'zning yonma-yon qo'llanish ehtimoli aniqlanadi.

Har bir so'z juftligiga ball beriladi. Ball qancha yuqori qimmatli bo'lsa, so'zlarning birikuvchanligi shunchalik yuqori sanaladi. Bu amal tilshunoslikdagi *leksikografiya* va *texnik tarjimada* foydalanish uchun korpusdan so'z birliklarini ajratib olish imkonini beradi. Shuningdek, mazkur amal orqali ma'no o'zgarishini aniqlash maqsadida o'xshash so'zlarni aniqlash mumkin. Masalan, **daryo qirg'og'i = landshaft, ustoz piri komil = muallim**.

Bu esa ma'nosi o'xshash so'zlar o'rtasidagi foydalanishdagi farqlarni ajratish imkonini yaratadi. Til ta'limida bunday ma'lumotlar ikki o'xshash so'z o'rtasidagi tafovutlar haqidagi bilimlarni olish, shu bilan birga, tilni yaxshiroq o'rganishda muhim rol o'ynaydi. Turli lingvistik elementlarning (masalan, so'z, morf, idioma va boshqalar) birikmalari haqidagi ma'lumotlar til ta'limidan tashqari *lug'atlarni ishlab chiqish, NLP vazifalari va mashina tarjimasi* uchun muhim. Biroq qaysi so'z birikmasi muhim ekanligini aniqlash oson jarayon emas.

Kontekstdagi kalit so'z (Key-Word-In-Context, KWIC)

KWIC funksiyasi korpusni qayta ishlashda mayyan so'zning hujjatlarda mavjudligini qidirishda keng qo'llaniladi. Tekshirilayotgan so'z har bir satrning o'rtasida paydo bo'lib, ikkala tomonda qo'shimcha qismlardan tashkil topadi. NLP vazifasiga ko'ra, kontekstning uzunligi turli xil bo'lishi mumkin. U markazda so'zning har ikki tomonida ikki, uch yoki to'rtta so'zdan iborat qism (n-gram)ni ko'rsatadi. So'z, ibora va gaplarni tahlil qilishda yaxshiroq tushunish uchun qo'shimcha kontekstga ehtijoz seziladi. KWICni o'z-o'zidan matn sifatida talqin qilish va **markaziy so'z** qurshovidagi so'zlarning chastotasini o'rganish ahamiyatli.

Til korpusi KWIC vosidasida tahlil qilganidan so'ng lingvistik tavsifda turli vazifalarni hal qilish maqsadida turli amallarni ishlab chiqish mumkin.

KWIC kontekstning ahamiyati, assotsiativ so'zlarning roli, kontekstdagi so'zlarning xarakteristikasi, qurshovdagi so'zlar haqidagi ma'lumotlar va so'zdan foydalanishda kontekstli cheklovlar mavjudligini tushunishga yordam beradi.

So'zlarni guruhlash (Local Word Grouping, LWG)

LWG – bu matnlarni tahlil qilish yoki korpusni qayta ishlashning yana bir turi bo'lib, u matnlardagi so'zlar guruhidan foydalanish shablonlarini aniqlab beradi. Bu usul gapning ma'nosini aniqlashda so'z tartibi muhim bo'lgan NLP vazifalarini hal qilishda foydalaniladi. LWG usuli so'z birikmalari va gaplar darajasida tahlil qilish vaqtida tarkibiy qismlarning funksional xatti-harakatlari haqida ma'lumot beradi. LWG usuli unchalik keng tarqalgan bo'lmasa-da, refleksiv shakllarga o'ziga xos yaqinlikka ega bo'lgan fe'l shakllarining (masalan, o'yin-kulgi, o'z-o'zini xursand qilish, qarzga berish, eslatish va h.) taqsimlanishini belgilaydi.

Bunday namuna(shablon)larni bilish til o'rganuvchilarni o'rta darajadan yuqori malaka darajasiga o'tkazishda muhim ahamiyat kasb etadi.

LWG usuli o'zbek tili korpusi matnlaridagi ma'lumotlar asosida fe'lli birikma va otli birikmalarini (yoki guruh) tahlil qilishga yordam beradi. LWG ma'lumotlari turli leksik elementlarning assotsiatsiyasidan kelib chiqadigan leksik noaniqlikni bartaraf etishga yordam beradi.

O'zbek tili korpusi ustida olib borilgan tajribalar shuni ko'rsatadiki, omonim so'zlarning ma'nosini aniqlashda LWG usulidan foydalanish yuqori samaradorlikni ta'minladi. Ko'pgina qo'shma fe'l, sifat va otlar uchun ma'lum bir so'z birikmasi bilan ifodalangan ma'noni alohida so'zlarning ma'nolaridan olish mumkin emas.

Xulosa

Bugun axborot asrida tabiiy tillarning tuzilishini o'rganuvchi NLP yaqin kelajakda fizika, biologiya kabi juda muhim sohaga aylanishi kerak, chunki axborot tuzilmasi, asosan, tabiiy tillar birliklaridan iborat. Fizika dunyodagi jismoniy harakatlar qonunlarini o'rgansa, NLP tabiiy tillar qoidalarini, axborot tashuvchilarni, axborot tarmoqlari dunyosini o'rganadi. Tabiiy tilni kompyuter yordamida tadqiq qilish va qayta ishlash, odatda, to'rt jihatni: formallashtirish, algoritmlash, dasturlash va amaliyotga qo'llashni o'z ichiga oladi. Korpus matnlarini lingvistik teglash dastlab lingvistik nazariyalarni yoki bugungi kunda ma'lum bo'lganidek, til korpuslarni ishlab chiqish va tahlil qilish uchun ma'lumot berish uchun amalga oshirildi. Ushbu maqolada til korpusi matnlari chastotasi, konkordans va KWIC texnologiyasi haqida fikr mulohaza yuritildi.

Foydalanilgan adabiyotlar:

1. Bi, P. (2018). Handbook of Linguistic Annotation. *Journal of Quantitative Linguistics*. <https://doi.org/10.1080/09296174.2018.1424495>
2. Bender, E. M., & Lascarides, A. (2019). Linguistic Fundamentals for Natural Language Processing II: 100 Essentials from Semantics and Pragmatics. In *Synthesis Lectures on Human Language Technologies* (Vol. 12, Issue 3). <https://doi.org/10.2200/S00935ED1V02Y201907HLT043>
3. Boltayevich, E. B., Mirdjonovna, H. S., & Ilxomovna, A. X. (2023). Methods for Creating a Morphological Analyzer. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 13741 LNCS. https://doi.org/10.1007/978-3-031-27199-1_4
4. Elov B.B., Hamroyeva Sh.M., Xusainova Z.Y. The pipeline processing of NLP // E3S Web of Conferences INTERAGROMASH. Rossiya, Rostov, 2023. <https://doi.org/10.1051/e3sconf/202341303011>
5. Xusainova Z.Y. NLP: tokenizatsiya, stemming, lemmatizatsiya va nutq qismlarini teglash // “O'zbek amaliy filologiyasi istiqbollari” mavzusidagi respublika ilmiy-amaliy konferensiyasi – Toshkent, 2022. №.1. – B.154-163.
6. Garside, R. G., Leech, G., & Mcenery, A. M. (2020). Introducing corpus annotation. In *Corpus Annotation*. <https://doi.org/10.4324/9781315841366-7>
7. Hovy, E., & Lavid, J. (2010). Towards a ‘ Science ’ of Corpus Annotation : A New Methodological Challenge for Corpus Linguistics. *International Journal of Translation*, 22(1).
8. Pustejovsky, J., & Stubbs, a. (2013). Natural language annotation for machine learning. In *Vasa*.
9. Sheng, D. (2023). Statistics in corpus linguistics: a practical guide. *Social Semiotics*, 33(4). <https://doi.org/10.1080/10350330.2021.1969215>

KORPUS MATNLARINI TEGGLASH MUAMMOLARI

Yodgorov Umid Saydilla o'g'li

yodgorov@navoiy-uni.uz

ToshDO'TAU o'qituvchisi

Annotatsiya. Korpus matnlarini teglash chatbotlar, virtual yordamchilar, qidiruv tizimlari va mashina tarjimasi kabi tabiiy tilni qayta ishlash (NLP) ilovalarida asosiy amal hisoblanib, u sun'iy intellekt modellariga inson tili va muloqotni yanada samaraliroq tushunishga yordam beradi. NLP ilovalaridagi o'quv ma'lumotlarini teglashdagi noaniqliklar AI modellarini real jahonda mavjud ilovalarida ishlamay qolishiga olib kelishi mumkin. Gartner hisobotiga ko'ra, 8% noto'g'ri ta'lim ma'lumotlarini kiritish sun'iy intellektning aniqligini 75%ga kamaytirishi mumkin. Korpus matnlarini annotatsiyalash yoki teglash orqali so'z, ibora, so'z birikmasi shu kabi leksik birliklarni izohlash kabi matnga teg (izoh) qo'shish va ularni belgilashdan iborat. Ushbu maqolada til korpusi matnlarini teglash muammolari ko'rib chiqiladi.

Abstract. Corpus text tagging is a key practice in natural language processing (NLP) applications such as chatbots, virtual assistants, search engines, and machine translation, where it helps AI models understand human language and communication more effectively. Inaccuracies in the labeling of training data in NLP applications can cause AI models to fail in their existing real-time applications. According to a Gartner report, entering 8% of the wrong training data can reduce the accuracy of artificial intelligence by 75%. Annotating or tagging corpus texts consists of adding a tag (explanation) to the text and defining them, such as explaining words, phrases, phrases, and similar lexical units. This article examines the problems of tagging language corpus texts.

Аннотация. Разметка корпуса текста - это ключевая практика в приложениях обработки естественного языка (NLP), таких как чат-боты, виртуальные помощники, поисковые системы и машинный перевод, где они помогают моделям ИИ более эффективно понимать человеческий язык и общение. Неточности в маркировке обучающих данных в приложениях НЛП могут привести к сбою моделей ИИ в существующих приложениях реального времени. Согласно отчету Gartner, ввод 8% неправильных данных обучения может снизить точность искусственного интеллекта на 75%. Аннотирование или маркировка корпусных текстов заключается в добавлении к тексту тега (пояснения) и его определении, например, пояснении слов, словосочетаний, словосочетаний и подобных лексических единиц. В данной статье рассматриваются проблемы разметки текстов языкового корпуса.

Kalit so'zlar: *Til korpusi, matnni teglash, mashinali o'qitish modellari, korpus turlari, parallel korpus, teglangan korpuslar, POS teg.*

Kirish

Korpus shakllantirilganidan so'ng undan NLP vazifalarini bajarishda foydalanish uchun qayta ishlash lozim. Til korpusidan tegishli ma'lumotlarni olish uchun *tizim, vosita, metod* va *dasturlarni* ishlab chiqish talab etiladi. Korpusni qayta ishlash nafaqat asosiy lingvistik tadqiqot va ishlanmalar uchun, balki til texnologiyasi ishlarida ham ajralmas jarayon hisoblanadi.

Bugungi kunda korpusni qayta ishlashda *statistik tahlil, konkordans, leksik kollokatsiya, kalit so'zlarni qidirish, so'zlarni lokal guruhlash, lemmalash, morfologik ishlov berish va generatsiya, qismlarga ajratish, so'zlarni qayta ishlash, so'z turkumlarini belgilash, izohlash, tahlil qilish* kabi amallar bajariladi [Garside, Leech, Mcenery, 2020; Hovy, Lavid, 2010]. Ba'zi hollarda korpusni qayta ishlash evaziga olingan natijalar til va uning xususiyatlari haqidagi xulosalarga zid kelishi kuzatiladi.

Jahonda keng foydalaniladigan ingliz, fransuz, nemis va shunga o'xshash tillar uchun korpuslarni qayta ishlovchi ko'plab dasturlar mavjud. O'zbek tili korpusi uchun yuqorida keltirilgan amallarning ba'zilar amaliyotga joriy etilgan bo'lsa-da, samaradorligi qoniqarli emas. O'zbek tilining tabiatini (o'ziga xosligini) hisobga olgan holda o'zbek tili korpusini qayta ishlash vositalarini loyihalash dolzarb. Shu sababli ingliz va jahonning boshqa tillarida faol qo'llaniladigan ba'zi taniqli korpuslarni qayta ishlash usullari va vositalarini tahlil qilishga ehtiyoj tug'iladi.

Korpus matnlarini teglash

Korpus matnlarini teglash – gaplarning xususiyatlarini aniqlash uchun matnga yoki uning turli elementlariga **teg(lar)** biriktirish (yozis/belgilash) jarayoni [Gries, Berez, 2017]. Bu jarayonda matn yoki audio yozuvlardagi *so'z turkumlari (parts of speech, POS), grammatik kategoriya va grammatik ma'no, sintaktik xususiyatlar, kalit so'zlar, iboralar, kinoya* va boshqa leksik birliklar ajratib olinadi hamda lisoniy ma'lumotlar qo'shimcha yoki metama'lumotlar bilan belgilanadi [Elov va boshqalar, 2022; Elov va boshqalar, 2023].

Matnni teglash/annotatsiyalashda matn ma'lumotlarini turli xil tabiiy tillarni qayta ishlash vazifalari uchun tushunarli va foydalanishga imkon beradigan **yorliq(metka)lar** yoki **toifalarni belgilash** jarayonidir.

Shuni ta'kidlash kerakki, “matn annotatsiyasi”, “matnni teglash” atamallari ko'pincha bir-birining o'rnida ishlatiladi, chunki ular AI/ML modellarini o'rgatish uchun matn ma'lumotlariga ma'lumot qo'shishni o'z ichiga oladi. Biroq matn annotatsiyasi – kengroq tushuncha, matnni teglash esa matn izohidagi o'ziga xos kichik vazifadir (1-jadval).

1-jadval. Matn annotatsiyasi va matn teglash o'rtasidagi farq

	Matn annotatsiyasi	matn teglash
Ta'rif	Matnga qo'shimcha ma'lumot, <i>metama'lumotlar</i> yoki kontekst qo'shishni o'z ichiga olgan kengroq atama.	Matn ma'lumotlarini turli NLP vazifalari uchun tushunarli va foydalanishga yaroqli/qulay qilish uchun <i>teg</i> yoki <i>toifalarni belgilash</i> jarayoni.
Maqsad	AI/ML modellarini matnning turli qismlari orasidagi <i>kontekst</i> , <i>hissiyot</i> va <i>munosabatlarni</i> tushunish, o'rgatishda yordam beradi.	Modellarni o'rgatish va baholash uchun etiketli ma'lumotlar talab qilinadigan nazorat ostidagi mashinali o'rgatish vazifalarida ishlatiladi, bu esa ularga teglangan ma'lumotlar asosida bashorat qilish, ma'lumot olish yoki boshqa NLP vazifalarini bajarish imkonini beradi.
Qo'llanilishi	Aniqlik yoki qo'shimcha kontekstni ta'minlash uchun matnga <i>sharh</i> yoki <i>izoh</i> qo'shish, <i>rasm</i> , <i>video</i> yoki <i>matn fragmentlari</i> bilan <i>havolalar</i> kabi multimedia elementlarini bog'lash.	Matnni tasniflash, NERlarni aniqlash, kalit so'zlarni ajratib olish va hissiyotlarni tahlil qilish.
Metodologiya	Muayyan vazifaga qarab qo'lda yoki avtomatlashtirilgan vositalar yordamida amalga oshirilishi mumkin.	Qo'lda yoki avtomatlashtirilgan vositalar yordamida amalga oshirilishi mumkin, ammo bu matn izohidagi ma'lum bir kichik vazifa.

Matn annotatsiyasi xatolarni aniqlash va qidiruv natijalarini yaxshilash uchun turli sohalarda qo'llaniladi [Elov va boshqalar, 2023; Xusainova, 2023]. Quyida turli sohalarda yaxshiroq natijalarga erishish uchun matn annotatsiyasidan qanday foydalanish keltirilgan:

1. Sog'liqni saqlash. Matn annotatsiyasi elektron sog'liqni saqlash yozuvlari (electronic health records, EHR) va tibbiy adabiyotlarni tahlil qilishda ishlatiladi. Bu sog'liqni saqlash xodimlariga bemor ma'lumotlaridagi *shablon*, *tendentsiya* va *korrelyatsiyalarni* aniqlashga yordam beradi. Tahlillar shuni ko'rsatdiki, EHRlarni tahlil qilish uchun tabiiy tilni qayta ishlashdan foydalanish qayta davolashni 71,2% aniqlik bilan bashorat qilishi mumkin.

2. Chakana savdo. Matn annotatsiyasi mijozlarning *fikr-mulohazalari* va *ijtimoiy tarmoq sharhlarini* tahlil qilishda ishlatiladi. Bu esa chakana savdo sotuvchilarga tendensiya, imtiyoz va sohalar samaradorligini aniqlashga imkon beradi. IBM kompaniyasi tomonidan olib borilgan tadqiqot shuni ko'rsatdiki, chakana savdo sotuvchilarning 62 foizi tahlil va ma'lumotlar tomonidan taqdim etilgan tushunchalar ularga raqobatdosh ustunlik berganini ta'kidlagan.

3. Moliya. Moliya institutlari *firibgarlikni aniqlash, mijozlarga xizmat ko'rsatishni yaxshilash* va *biznes operatsiyalarni soddalashtirish* uchun mijoz so'rovlari, tranzaksiyalari va boshqa matnli ma'lumotlarni tahlil qilishda matnli annotatsiyadan foydalanadi. Misol uchun, Bank of America Erica deb nomlangan virtual yordamchini yaratish uchun tabiiy tilni qayta ishlash va bashoratli tahlildan foydalanadi. Bu esa mijozlarga yaqinlashib kelayotgan veksellar yoki tranzaksiyalar tarixi haqidagi ma'lumotlarni ko'rishga yordam beradi.

4. Mijozlarni qo'llab-quvvatlash. Matn annotatsiyasi umumiy muammo, tendensiya va soha samaradorligini yaxshilashda *chat transkriptlari* va *elektron pochta xabarlari* kabi mijozlarni qo'llab-quvvatlash xizmatlari faoliyatini tahlil qilish uchun ishlatiladi. Korxonalar mijozlarning his-tuyg'u va ehtiyojlarini tushunib, o'z qo'llab-quvvatlash xizmatlarini yaxshilashi va mijozlar talabiga javob berishi mumkin.

5. Inson resurslari (Human Resources, HR). Matn annotatsiyasi *ishga qabul qilish haqidagi ariza, rezyume* va *xodimlarning fikr-mulohazalarini* tahlil qilishda qo'llanadi. Bu HR mutaxassislariga shablon, tendensiya va potentsial muammolarni aniqlashga yordam beradi. Misol uchun, matn annotatsiyasidan ish joyidagi madaniyat yoki boshqaruv muammolari kabi sohalarni aniqlash va ularni hal qilishda tegishli choralarni ko'rish uchun xodimlarning fikr-mulohazalarini tahlil qilishda amaliy ahamiyat kasb etdi.

Matn annotatsiyasidan foydalangan holda, yuqorida keltirilgan sohalardagi strukturlanmagan ma'lumotlarni qimmatli/zarur ma'lumotga aylantirish mumkin. Bu esa sohada muhim qarorlarni qabul qilish sifatini yaxshilash, mijozlar tajribasi va umumiy biznes natijalar samadorligining oshishiga olib keladi.

Matnni teglash(annotatsiyalash) muammolari

AI va ML kompaniyalari matnni annotatsiyalash jarayonlarida bir nechta qiyinchiliklarga duch kelishadi. Bularga ma'lumotlar *sifatini ta'minlash, katta hajmdagi ma'lumotlar to'plamini samarali boshqarish, konfedensial ma'lumotlarni himoyalash* va *ma'lumotlar o'sishi bilan masshtabni kengaytirish* kiradi [Cassidy, Schmidt, 2017; Bi, 2018; Arista, 2022]. Matnni annotatsiyalashga asoslangan NLP ilovalarining samaradorligini oshirish uchun ushbu muammolarini hal qilish juda muhimdir.

1. Noaniqlik. Bu so'z, so'z birikmasi yoki gap bir nechta talqinga (ma'noga) ega bo'lganda sodir bo'ladi. Bu esa matnni annotatsiyalash jarayonidagi nomuvofiqliklar va xatolarga olib kelishi mumkin.

For example, the phrase “I saw the man with the telescope” can be interpreted in two ways: either the speaker saw a man who had a telescope, or the speaker used a telescope to see the man.

2. Subyektivlik. Bu matnda shaxsiy fikrlar, his-tuyg'ular yoki baholashlarning mavjudligini anglatadi. Subyektiv tilni izohlash qiyin bo'lishi mumkin, chunki turli izohlovchilar bir xil matnni turlicha talqin qilishlari mumkin, bu esa annotatsiyalarda nomuvofiqliklarga olib keladi. Masalan, mijozlar sharhlarida fikr-mulohazalarni izohlash turli sharhlovchilarning matnni tushunishi va idrokiga asoslanib, bir xil sharhni *ijobiy*, *neytral* yoki *salbiy* deb belgilashiga olib kelishi mumkin.

3. Kontekstni tushunish. Matndagi so'zlar va so'z birikmalarning ma'nosi ko'pincha ular ishlatilgan kontekstga bog'liq. Annotatorlar ma'lumotlarni aniq belgilash va tasniflash uchun matnni va parchaning umumiy ma'nosini hisobga olishlari kerak. For example, the word “bank” can refer to a financial institution or the side of a river, depending on the context. Failing to consider the context may lead to incorrect annotations and negatively impact the performance of machine learning models.

4. Tillarning xilma-xilligi. Tabiiy tillarning xilma-xilligi matn annotatsiyasida qiyinchilik tug'diradi. Chunki annotatorlar turli tillardagi ma'lumotlarni to'g'ri belgilash va turkumlashtirish uchun bir nechta tillarni bilishi kerak. Bu, ayniqsa, kam resursli tillar yoki lahjalar bilan ishlashda qiyin bo'lishi mumkin, chunki kerakli tajribaga ega izohlovchilarni topish qiyin bo'lishi mumkin. Shuningdek, tillarning xilma-xilligi annotatsiyalash jarayonida nomuvofiqliklarga olib kelishi mumkin, chunki turli annotatorlar ma'lum bir tilda turli darajadagi malakaga ega bo'lishi mumkin.

5. Masshtablilik. Katta hajmdagi ma'lumotlarni annotatsiyalash jarayoni ko'p vaqt va resurs talab qiladi. Bu esa matn annotatsiyasida jiddiy muammo tug'diradi. Tadqiqotlar shuni ko'rsatadiki, AI loyihasi vaqtining 80% dan ortig'i ma'lumotlarni boshqarishga, jumladan, ularni *yig'ish*, *umumlashtirish*, *tozalash* va *teglash* uchun sarflanadi. Ma'lumotlar hajmi oshgani sayin, ma'lumotlar annotatorlariga bo'lgan talab ham oshib boradi, bu esa kompaniyalar uchun o'zlarining izohlash ishlarini samarali ravishda kengaytirishni qiyinlashtiradi.

6. Narx. Bu kompaniyalar uchun katta to'siqdir, chunki annotatorlarni yollash va o'qitish, shuningdek, teglash vositalari va texnologiyalariga sarmoya kiritish qimmatga tushishi mumkin. 2030 yilga kelib ma'lumotlarni teglash bo'yicha global bozor qiymati 13 milliard dollarga yetishi kutilmoqda. Bu esa yuqori sifatli ma'lumotlarni annotatsiyalash uchun zarur bo'lgan katta sarmoyani talab qiladi. Annotatsiya xizmatlarining narxi bilan aniq va izchil izohlarga bo'lgan ehtiyojni muvozanatlash sun'iy intellekt va mashinali o'qitish bo'yicha yechimlarni joriy qilmoqchi bo'lgan kompaniyalar uchun muhim muammo hisoblanadi.

Xulosa

Tabiiy tilni kompyuter yordamida tadqiq qilish va qayta ishlashga asoslangan dasturiy ilovalar(axborot tizimlari)ni ishlab chiqish uchun NLP tizimini mavjud ma'lumotlarni o'rganishini ta'minlash kerak. Ushbu amalni til korpusi orqali amalga oshirish lozim. Til korpusi elektron shaklda taqdim etiladigan katta hajmli va strukturlangan matnlar to'plami sifatida qaraladi. Til korpusi tabiiy tilni qayta ishlash tizimining asosidir. NLP bilan bog'liq lingvistik bilimlar leksik, sintaktik, semantik va pragmatik xususiyatlarni o'z ichiga oladi. Ushbu korpus matnlarini teglash usullari va muammolari haqida keltirildi. Korpus matnlariga ishlov berishdan oldin unga dastlabki ishlov berish matn tozalash tizimiga oid vazifa sanalib, u *nomuhim so'zlar, maxsus belgi, kulgich, raqam, tinish belgi, ortiqcha bo'shliq, elektron pochta manzili, HTML teglari hamda URL* kabi elementlarni olib tashlash orqali matn ma'lumotlarining sifatini oshirishga xizmat qiladi.

Foydalanilgan adabiyotlar:

1. Garside, R. G., Leech, G., & Mcenery, A. M. (2020). Introducing corpus annotation. In *Corpus Annotation*. <https://doi.org/10.4324/9781315841366-7>
2. Hovy, E., & Lavid, J. (2010). Towards a 'Science' of Corpus Annotation: A New Methodological Challenge for Corpus Linguistics. *International Journal of Translation*, 22(1).
3. Gries, S. Th., & Berez, A. L. (2017). Linguistic Annotation in/for Corpus Linguistics. In *Handbook of Linguistic Annotation*. https://doi.org/10.1007/978-94-024-0881-2_15
4. Elov B., Hamroyeva Sh., Abdullayeva O., Uzakova M. O'zbek tilida POS tagging masalasi: muammo va takliflar / Uzbekistan: Language and Culture. *Applied philology*. 2022/2(5). – 51-68-b.
5. Elov B., Hamroyeva Sh., Xudayberganov N., Yodgorov U., Yuldashev A. Pos tagging of Uzbek texts using hidden Markov models (HMM) and Viterbi algorithm. O'zMU xabarlar. Mirzo Ulug'bek nomidagi O'zbekiston Milliy universiteti ilmiy jurnali. 2023 yil Maxsys son.
6. Elov, B.B., Hamroyeva, Sh.M., Abdullayeva, O.X., Husainova, Z.Y., Xudoyberganov, N.U. 2023. "Agglutinatив tillar uchun pos teglash va stemming masalasi (turk, uyg'ur, o'zbek tillari misolida)". O'zbekiston: til va madaniyat 2: 6-39.
7. Xusainova Z.Y. O'zbek tili milliy korpusi qidiruv tizimini optimallashtirishda lemmatizatsiyadan foydalanish / O'zbekiston: Til va Madaniyat (Kompyuter lingvistikasi). Toshkent, 2023, №2. – B.20-37.
8. Cassidy, S., & Schmidt, T. (2017). Tools for Multimodal Annotation. In *Handbook of Linguistic Annotation*. https://doi.org/10.1007/978-94-024-0881-2_7
9. Bi, P. (2018). *Handbook of Linguistic Annotation*. *Journal of Quantitative Linguistics*. <https://doi.org/10.1080/09296174.2018.1424495>



10. Arista, J. M. (2022). TOWARD THE MORPHO-SYNTACTIC ANNOTATION OF AN OLD ENGLISH CORPUS WITH UNIVERSAL DEPENDENCIES. Revista de Linguistica y Lenguas Aplicadas, 17. <https://doi.org/10.4995/rlyla.2022.16787>

JADID DAVRI ADABIY MUHITIGA DOIR ASARLARDAN KORPUSDA FOYDALANISH

Nuritdinov Abrorbek Sayfiddin o'g'li

nuriddinovabrorbek@gmail.com

University of business and science
nodavlat oliy ta'lim muassasasi o'qituvchisi

Annotatsiya. O'zbek tili korpusida matn bilan ishlash, korpusga kiritilayotgan manbaning tadqiq imkoniyatlari va unda tug'iluvchi muammolar haqida so'z yuritildi.

Abstract. The Uzbek language consisted of working with the text, the research capabilities of the source entered into the corps, and the problems that arise there.

Аннотация. В корпусе узбекского языка обсуждались возможности работы с текстом, исследовательские возможности источника, входящего в корпус, и проблемы, которые в нем возникают.

Kalit so'zlar: *qidiruv tizimi, korpus, korpus interfeysi, korpus annotatsiyasi, O'zbek tili korpusi, token, lemma, konkordans.*

Korpus turli xildagi tadqiqot uchun asos, manba, vosita vazifasini o'tay olganligi uchun ham dunyodagi juda ko'p sohalar korpusga tayanib ish ko'radi [Hamroyeva, 2020:16]. Aksariyat qidiruv tizimlari uchun eng murakkab muammolaridan biri – bu matnli qidirish. Ko'pgina qidiruv tizimlari uncha katta bo'lmagan massivlarda ishlaydi, lekin korpus hajmi ortgan sayin ishlash tezligi va qidirish sifati kamayadi. Bunda qidiruv tezligi nafaqat korpus hajmi, balki talabning murakkabligiga, so'zlar soniga uzviy bog'liq.

O'zbek tili korpusining qidiruv tizimiga bo'lgan asosiy talablarni ifodalaymiz:

- 1) so'z va so'z birikmasini ularning belgilari (grammatik, semantik va b.)ga qarab qidirish;
- 2) matn (nutq yoki asarning tugal fikr anglatuvchi parchasi) va so'zlar orasidagi masofani hisobga olish;
- 3) metamatnli ma'lumotni qidirish;
- 4) taraqqiy etgan til talablariga xos, o'z ichiga mantiqiy bog'lamalar, qavs va matn operatorlarni qamrab olish;
- 5) indeksatsiyalash samaradorligi;
- 6) yetarli darajadagi murakkab savolga yetarli darajadagi tezlikda javob topish;
- 7) keng ko'lamlilik, yetarli darajadagi katta hajmgacha (yuz millionlab so'zlarni ishlatish).

Korpus bilan ishlashda yuqori samaradorlikka erishish uchun, to'liq matnli qidiruv tizimi zarur. Mazkur tizim ulkan matnlar massivida matn qidiruv vositasi uchun maxsus optimallashtirilgan. Bunday tizimlar o'rtasida eng mashhurlari – Yandex-server va Google internet qidiruvi, shuningdek, indeksatsiya va internet-manbalar qidiruvi kabi tizimlar ham. Tabiiy ravishda o'zbek tili korpusi uchun Yandex-server tizimi juda mos. Yandex-server tizimi juda yuqori samaradorlik va keng ko'lamlilik, matn qidiruvni to'liq amalga oshirish, qudratli til so'rovnomasiga ega bo'lish, matnlarning nostandart turlari uchun egiluvchan indeksator sozlagichini o'rnatish, o'zbek tilini to'liq ko'tarish kabi imkoniyatlariga ega. Yandex-server qisqa fursat ichida ulkan matnlar massivida murakkab so'rovlarni amalga oshirish imkoniyatiga ega bo'lib, bunda qidiruv tezligi korpus hajmiga umuman bog'liq emas.

Bilamizki, matnlarni qayta ishlashda matndan foydalanuvchi qaysi maqsad bilan yondashganiga qarab matnlarni avtomatik qayta ishlovchi dasturlar mavjud. Bu dasturlar turlicha bo'lgani uchun ularni topishimiz va ulardan qanday foydalanishimiz haqida ma'lumotlarga ega bo'lishimiz kerak. Bu esa bizdan ko'proq vaqt sarflashimizni talab qiladi. Bularning barchasi korpusda jamlangan holda o'z yechimini topmoqda. Korpus bilan qanday ishlashni esa korpus interfeysi ko'rsatib beradi. Interfeys asli inglizcha so'z bo'lib, “tashqi ko'rinish” degan ma'noni anglatadi. Asosan kompyuter texnologiyasida qo'llanadi. Korpus interfeysi korpusning dastlabki kirganimizdagi ko'rinishi hisoblanadi. Bu vizual interfeys hisoblanadi. O'zbek tili korpusining interfeysini ko'rishimiz uchun internet tarmog'idagi <https://uzbekcorpus.uz> havolasiga kiramiz. Hozirda O'zbek tili korpusi online holda ishlaydi. O'zbek tili korpusi interfeysida foydalanuvchiga qulay bo'lgan til (o'zbek, rus, ingliz) va korpus annotatsiyasi (unda: bosh sahifa, loyiha a'zolari, izlash, morfoanalizator, parallel korpus, ta'limiy korpus, mualliflik korpusi, lingvistik resurslar, tezaurus, foydalanuvchi yo'riqnomasi, korpus haqida, ilmiy nashrlar)dan iborat. Interfeys orqali berilgan buyruqlardan ichki interfeyslar paydo bo'ladi.

Korpusdan izlash to'rt xil usul bilan (token, lemma, birikma bo'yicha va konkordans) amalga oshiriladi. Qidiruv davomida qaysi usulda qidirish buyrug'i bilan birgalikda uslub va davrni ham tanlash zarur. Token bo'yicha qidiruv so'z shakllarini topib beradi. Token bo'yicha qidirishda izlashga so'zning leksema shaklini kiritish zarur. Chunki Abdurauf Fitrat asarlarida, xususan, jadid davri adabiyotida grammatik qo'shimchalarning hozirgidan ancha farqi bo'lgan. Masalan, token bo'yicha qidirishda *ot* so'zini izlamoqchimiz. Foydalanuvchi biror manbada jadid davrida bu so'zning *otg'a* shaklini ko'rib, shunday shaklda qidirishi bu davr adabiyoti uchun natija doirasini toraytiradi. Bu so'zni u davrda *otqa* kabi shakllarda ham qo'llanganligini uchratish mumkin (manba kirillda): *Қараш белгиси бир **отқа** қўшилгач, ундан кейин келган бир **отға** учинчи турли “олмош излари” қўшиладур* [Abdurauf Fitrat, 2006:171]. Yoki *to'ldiruvchi* terminini hozirgi shu nom bilan izlashi mumkin. Lekin Abdurauf Fitrat asarlarida *to'ldirg'ich* shaklida

berilgan: *Бориш тўлдиригичи гапнинг шундай бир бўлагидирким, гапнинг бошқа бир бўлмидан ўзига томон бир ҳаракат, бир узаниш борлигини билдирадир. Бунинг белгиси “га — қа — га — ка” дир. Бориш тўлдиригичидан белгисининг тушириб қолгани ҳам бўладир: Бозор бордим — бозорга бордим. Бориш тўлдиригичига “га — га, давур — довур” қўшилганда ҳаракат ва узанишнинг сўнг ўзи бўладир. Бузоқнинг югургани оғилгача. Бозоргача бордим. Бу йил қишқача далада қолдим. Бу иш қиёматқа довур бўлмайдир. Шаҳарга давур айланиб келдим* [Abdurauf Fitrat, 2006:182].

Bunday hozirgi kundagi farqli shakllarni izlashda ham munosib yechim topish zarur. Dastlabki barchaga ma'lum usul so'zshaklni yanada qisqartirish orqali izlash, bunda *to'ldir* shakli kiritiladi va kerakli natijaga erishiladi. Agar asos tarkibida farq bo'lsa-chi, masalan, *sifat* so'zi *sufat* shaklida uchraydi: Сўз: *от, суфат, феъл, олмош, кўмакчи сўзлари каби қисимларга ажраладир* [Abdurauf Fitrat, 2006:145]. Bu holat tezausda o'z yechimini topadi. Tezaus shakllantirilguncha qidirilishi lozim bo'lgan so'zlarning hozirgi va jadid davri mos ravishda variantlashtirilib sinonimlar qatori ishlab chiqiladi. Bu orqali so'zning hozirgi davridagi shakliga jadid davridagi shakli sinonim qilib so'z haqidagi metama'lumotga kiritiladi. Jadid davridagi shaklini ko'rgandan so'ng shu shaklda qidiriladi.

Lemma bo'yicha izlashda so'zlarning lug'atdagi shakli topiladi. Birikma bo'yicha izlashda birikma shaklida kelgan shakllar qidiriladi. Bu izlash jarayoni Abdurauf Fitratning asarlarida birikma holda kelgan tilshunoslik terminlarini topishda yordam beradi. Konkordansda esa so'zning boshqa so'zlar bilan qo'llanish qurshovini belgilab beradi.

Abdurauf Fitrat asarlari O'zbek tili korpusiga mualliflik korpusi sifatida joylandi. Asli tanlab olingan matnlar chegaralanganligi uchun ular maxsus matnlar korpus bazasi hisoblanadi. Maxsus matnlar korpusi – balansga keltirilgan korpus, u hajman u qadar katta bo'lmagan, alohida tadqiqiy topshiriqqa xizmat qiluvchi, ma'lum maqsadni ko'zlagan, tuzuvchi rejasiga muvofiq tayyorlanadi [Hamroyeva, 2020:55]. Bular sirasiga mualliflik va ta'limiy korpuslarni kiritish mumkin. Bu korpuslarning milliy korpuslardan farqli jihatlari ma'lum bir davrga yoki muallifga tegishli o'ziga xos jihatlari, so'zning zamonga, yozuvchiga mos ma'nosi, ajratilgan manbani til xususiyatlari, so'zlari orqali nimarni nazarda tutganligi chuqurroq tahlili namoyon bo'ladi. Abdurauf Fitratning asarlarini mualliflik korpusi qatorida tadqiq etish davr nuqtayi nazaridan ham, til nuqtayi nazaridan ham, mazmun nuqtayi nazaridan ham ahamiyatlidir. Avvalo, yuqorida ko'rib o'tganimizdek, qidiruv tizimida chigalliklarni keltirib chiqaradi. Jadid davri va hozirgi davr tili to'liq mos emasligi foydalanuvchiga kerakli natijani bermaydi. Mualliflik korpusi sifatida foydalanganda esa aynan jadid davri tili nuqtayi nazaridan yondashadi. So'zlarni shu davrga mos ravishda izlaydi. Mazmun jihatidan Abdurauf Fitratning asarlarini o'zini ham ajratgan holda tadqiq etish zarur. Asarlar badiiy adabiyotga, publitsistikaga, tarixga, adabiyotshunoslikka, tilshunoslikka oid ekanligini ham hisobga olinadi.

Masalan, tilshunoslikka oid asarlarini tahlil qilishda ko'proq tilshunoslik terminlariga e'tiborni qaratish mumkin. Terminlarning alohida lug'atini shakllantirib foydalanuvchiga havola etiladi. Korpusdan foydalanuvchi jadid davri so'zlarining shaklini bilishi unga korpusdan foydalanish imkoniyatini kengaytirib beradi. Shuning uchun morfologik teglash jarayonida so'zning jadid davridagi har bir shaklini ko'rsatish zarur. Agar korpus foydalanuvchi qidirgan so'zni topa olmasa, mos variantlarni taqdim etadi. Shu joyga foydalanuvchi so'zni hozirgi shakl bilan qidirsam, korpus jadid davridagi variantini taqdim etishi lozimligini hisobga olish zarur. Shularni hisobga olib Abdurauf Fitratning asarlarini mualliflik korpusi sifatida tadqiq etdik va O'zbek tili korpusining mualliflik korpusi bazasiga joyladik. Tajribalar asosida keyinchalik jadid tilshunoslari asarlarining, balki barcha jadid ma'rifatparvarlarining asarlari asosida korpus yaratishimiz mumkin. Bunda bizga hozirgi tajribalarimiz dasturi amal vazifasini o'taydi.

Shunday qilib, jadidlar korpusini quyidagicha interfeysda shakllantirish mumkin: (1-rasm)



1-rasm. Asosiy oyna

Jadid korpusini yaratishda e'tibor beriladigan jihatlar quyidagilardan iborat:

- korpus nomini interfeysning yuqori qismiga joylanadi;
- korpusning annotatsiya qismida mundarija sifatida korpusga joylangan jadid ma'rifatparvarlarining ro'yxati turadi. Bu korpusdan qanday ma'lumotlar topishimiz mumkinligini ko'rsatadi;



— mundarijadagi har bir nomning ustiga kelib bosganimizda qo'shimcha ravishda yana uchta tugma (avtobiografiya, asarlari va tadqiqotlar) hosil bo'ladi. Bu tugmalarni bosish orqali birinchisidan muallif avtobiografiyasi, ikkinchisidan asarlar ro'yxati hamda uchinchisidan muallif va asarlari ustida olib borilgan keyingi tadqiqotlar ro'yxati haqida ma'lumotlar chiqadi;

— annotatsiya qismida jadidlar nomi tugagandan so'ng barcha korpuslar kabi korpusdan foydalanish, korpus haqida, loyiha a'zolari va boshqa qismlar joylanadi;

— asosiy oynada dastlab foydalanuvchiga qulay bo'lgan til (o'zbek, ingliz, rus), ikki qidiruv tizimi joylashadi;

— bu ikki qidiruv tizimi umumiy va maxsus qidiruvlardan iborat bo'ladi;

— umumiy qidiruv tizimi korpusning ma'lumotlar bazasidagi barcha manba bilan izlashga yordam beradi, aynan ma'lum manbani chegaralab olmaydi;

— maxsus qidiruv tizimida esa manbani chegaralab olishimiz mumkin bo'ladi. U uchta izlash qismi (so'zni kiriting, muallifni kiriting va asar nomini kiriting) orqali qidiradi. Bulardan so'z va muallif bo'yicha izlash majburiy, asar nomi bo'yicha qidirish esa ixtiyoriy, ya'ni bu qismga asar nomi kiritilmagan taqdirda ham topib bera olishi kerak. Chunki foydalanuvchi ba'zan muallifning barcha asari bo'yicha, ba'zan ma'lum asar bo'yicha izlashi mumkin;

— umumiy va maxsus qidiruv tizimida ham korpusda bashorat qila olish imkoniyati bo'lishi zarur. Qidiruv qismiga so'zni kiritish davomida uning bazadagi shakllarini (*ot*→*oti*, *otqa*, *ota*, *otadi*), bu holat mos ravishda lotin grafikasida kiritilsa, kirillda yoki aksincha (*o'xsha*→*yxuamulgan*, *yxuamuu*, *yxuauu*, *yxuaagan*) kabi ko'rsatishi zarur;

— qidiruvni lemma, token, birikma va konkordans bo'yicha ajratish natija chiqqandan so'ng qilinsa, qulaylik tug'diradi. Bunda dastlabki qidiruv token bo'yicha bo'ladi va natijalarning yuqori qismida yonma-yon token, lemma, birikma va konkordans tugmalari joylashadi. Ularni tanlash orqali moslanadi. Bu ko'rinish barcha qidiruv shakllarining natijalarini bir joyga jamlashga yordam beradi. Qidiruv natijalarining yuqori qismida joylashgan qidiruv shakllari o'zi bilan birga hujjatlar statistikasini ko'rsatib beradi: (2-rasm)

Jadid korpusi



[O'zbek](#) [English](#) [Русский](#)

Token 17 ta hujjatda 28 ta so'z	Lemma	Birikma	Konkordanser
Abdurauf Fitrat. Tanlangan asarlar. IV jild. Oila yoki oila boshqarish tartiblari.			Zikr qilingan uchta tarbiyaning eng muhimi axloqiy tarbiya hisoblanadi.
Abdurauf Fitrat. Tanlangan asarlar. IV jild. Oila yoki oila boshqarish tartiblari.			O'zingiz xohlaganingizcha bolangizning aqliy va badan tarbiyasi bilan mashg'ul bo'lishingiz mumkin va qo'lingizdan ko'proq...
Abdurauf Fitrat. Tanlangan asarlar. IV jild. Oila yoki oila boshqarish tartiblari.			Binobarin, bolaning aqliy va jismoniy tarbiyasi samarasiz qolmasligi uchun farzandning axloqiy tarbiyasi jiddiy shug'ullanib ko'proq...
Abdurauf Fitrat. Tanlangan asarlar. IV jild. Oila yoki oila boshqarish tartiblari.			Axloqiy tarbiya — odam axloqini kamolga yetkazish demakdir, ya'ni odamni shunday tarbiya qilish kerakki, fe'li va amali o'ziga ham, ko'proq...

2-rasm. Qidiruv natijasining ko'rinishi

Shunday shaklda ko'rinish foydalanuvchiga qulay bo'ladi va osonlik bilan bir qidiruv shaklidan ikkinchi qidiruv shakliga o'tkaza oladi.

Bugungi kunda manbalarini elektron resursga aylanishi bilan birga uning so'zma-so'z qidiruv imkoniyatiga egaligi, bu jarayon tizimli ravishda bo'lishi korpusda matnlardan foydalanish imkoniyatining kengligini ko'rsatadi. Xususan, tilshunoslikda axborot tizimlarini yaratish ehtiyoji tadqiqotchi uchun kerakli bo'lgan ma'lumotlardan foydalana olish, o'z ishida unumliroq harakatlarni amalga oshirish va axborot oqimlarini samarali boshqarish imkonini beradi.

O'zbek tili korpusi qidiruv tizimi Abdurauf Fitratning asarlarini lemma, token, birikma va konkordans bo'yicha izlash imkoniyatini beradi. Bu orqali foydalanuvchiga qulay bo'lgan qidiruv shaklini tanlay oladi.

Korpusda kengaytirish va ixchamlik talabi mavjud. Kengaytirish talabi orqali ma'lumotlar bazasiga kirgan manbalarni yangi ma'lumotlar orqali boyitib boriladi. Bundan tashqari matn ustidagi yangi tahlillarni ham qo'shiladi. Ixchamlik talabi orqali korpus foydalanuvchiga zarur ma'lumotlarni o'zinigina taqdim etish, ortiqcha ma'lumotlarni saralay olish imkoniyatini ham yaratadi.



Foydalanilgan adabiyotlar:

1. Abduraxmonova N. Kompyuter lingvistikasi. – Toshkent, 2020.
2. Абдурауф Фитрат. Танланган асарлар. IV жилд. – Тошкент: Маънавият, 2006.
3. Ҳамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари. Монография. – Тошкент, 2020.
4. Ҳамроева Ш. Корпус лингвистикаси атамаларининг қисқача изоҳли луғати. – Бухоро, 2018.
5. Xolmonova Z. Kompyuter lingvistikasi. – Toshkent, 2019.
6. <https://uzbekcorpus.uz> — O‘zbek tili korpusi.
7. <http://uzschoolcorpara.uz> — O‘zbek tilining ta’limiy korpusi.

O'ZBEK TILI BOLALAR NUTQI AUDIO KORPUSINI YARATISH

Masharipova Baxtigul Rajabboy qizi

baxtigulmasharipova94@gmail.com

UrDU magistranti

Qutlimuratova Barno Xursandbekovna

kutlimuratovab0712@urdu.uz

UrDU o'qituvchisi

Annotatsiya. O'zbek tili uchun bolalar nutqi korpusini yaratish O'zbekistonda tilni qayta ishlash texnologiyalari va ta'lim resurslarini takomillashtirish yo'lidagi muhim qadamdir. Ushbu loyiha keng qamrovli va vakillik korpusini yaratish uchun turli yosh guruhlari, dialektlari va mintaqalaridagi bolalarning nutq ma'lumotlarini to'plash, tahlil qilish va tartibga solishga qaratilgan. Bolalar so'zlashadigan o'zbek tilining o'ziga xos fonetik, leksik va sintaktik xususiyatlariga moslashtirilgan nutqni aniqlash va sintez qilish tizimlari, ta'lim vositalari va lingvistik tadqiqotlarni rivojlantirishni qo'llab-quvvatlash asosiy vazifalardan iborat.

Abstract. Creating a children's speech corpus for the Uzbek language is a pivotal step toward enhancing language processing technologies and educational resources in Uzbekistan. This project aims to collect, analyze, and organize speech data from children across various age groups, dialects, and regions to create a comprehensive and representative corpus. The primary objectives are to support the development of speech recognition and synthesis systems, educational tools, and linguistic research, specifically tailored to the unique phonetic, lexical, and syntactic characteristics of the Uzbek language as spoken by children.

Аннотация. Создание детского речевого корпуса на узбекском языке является важнейшим шагом на пути совершенствования технологий обработки языка и образовательных ресурсов в Узбекистане. Целью этого проекта является сбор, анализ и систематизация речевых данных детей разных возрастных групп, диалектов и регионов для создания комплексного и репрезентативного корпуса. Основными задачами являются поддержка разработки систем распознавания и синтеза речи, образовательных инструментов и лингвистических исследований, специально адаптированных к уникальным фонетическим, лексическим и синтаксическим характеристикам узбекского языка, на котором говорят дети.

Kalit so'zlar: *NLP, Nutq korpusi, bolalar nutqi, O'zbek tili, korpus tahlili.*

Kirish.

Til texnologiyalarini rivojlantirish, ayniqsa kam vakil tillar uchun, til xilma-xilligini saqlash va inklyuziv ta'lim va muloqotni rivojlantirish uchun muhim ahamiyatga ega. Markaziy Osiyoda millionlab odamlar so'zlashadigan o'zbek tili

turli yosh guruhlari va mintaqalarda farq qiluvchi o'ziga xos xususiyatlarga ega. Bolalar nutqi, ayniqsa, kattalar nutqidan sezilarli darajada farq qiladigan aniq fonetik, leksik va sintaktik xususiyatlarni namoyon qiladi. Buni anglagan holda, o'zbek tili uchun bolalar nutqi korpusini yaratish hal qiluvchi ahamiyatga ega bo'ladi.

O'zbek tili uchun maxsus tayyorlangan bolalar nutqi korpusining joriy etilishi lingvistik resurslar va texnologiyalarni rivojlantirishdagi jiddiy bo'shliqni bartaraf etadi. Hozirgi til texnologiyalari asosan yaxshi hujjatlashtirilgan tillarga xizmat qiladi va o'zbek tili kabi tillar uchun resurslarda bo'sh joy qoldiradi. Ushbu korpus nafaqat o'zbek tilida so'zlashuvchilarga mo'ljallangan ilg'or ta'lim vositalari va ilovalarni ishlab chiqishga yordam beradi, balki nutqni aniqlash va sintez qilish tizimlarini ham yaxshilaydi va ularni bolalar uchun yanada qulayroq va samaraliroq qiladi.

Bundan tashqari, bu tashabbus faqat texnologiyani rivojlantirish bilan bog'liq emas; madaniy o'ziga xoslik va merosni tarbiyalash haqida. Bolalarning o'zbek tilida so'zlashuv usullarini, jumladan, mintaqaviy shevalar va so'zlashuv iboralarini hujjatlashtirish orqali korpus tilshunoslik tadqiqotlari va saqlash ishlari uchun qimmatli manba bo'lib xizmat qiladi. Bu tilshunoslar va pedagoglarga yosh so'zlovchilar orasida til o'zlashtirish va evolyutsiyani tushunishga yordam beradi, o'zbek tilining lingvistik kelajagi haqida tushuncha beradi.

Shu nuqtayi nazardan, o'zbek tili uchun bolalar nutqi korpusini joriy etish texnologik tafovutni bartaraf etish, lingvistik tadqiqotlarni boyitish, o'zbek tilida so'zlashuvchi yoshlarning ta'lim va kommunikativ ehtiyojlarini qo'llab-quvvatlashni va'da qiladigan istiqbolli ishdir. Ushbu loyiha orqali biz til texnologiyalari inklyuziv, qulay va O'zbekiston aholisining til xilma-xilligini ifodalovchi kelajak uchun zamin yaratishni maqsad qilganmiz.

Nutq korpusini yig'ish

Bolalar nutq korpusini yig'ish jarayoni ma'lumotlarning har tomonlama, vakolatli va sifatli bo'lishini ta'minlash uchun bir necha tizimli bosqichlarni o'z ichiga oladi. Umuman olganda, bu bosqichlar rejalashtirish, ma'lumotlarni yig'ish, qayta ishlash va tekshirishni o'z ichiga oladi. Bizning yondashuvimizda biz yig'ish jarayonini soddalashtirish va yaxshilash uchun texnologiyani integratsiyalash orqali, xususan, bolalar nutqi ma'lumotlarini to'plashni osonlashtirish uchun mo'ljallangan Telegram botini yaratish orqali innovatsiyalar qildik.

Rejalashtirish bosqichi

Dastlabki bosqich bolalarning yosh doirasi, qamrab olinadigan lahjalar va nutqiy o'zaro ta'sir turlari (masalan, suhbat, hikoya yoki ko'rsatma nutqi) kabi korpusning ko'lami va maqsadlarini aniqlashni o'z ichiga oladi. Shuningdek, u axloqiy me'yorlarni belgilash va vasiylardan zarur roziliklarni olish, ishtirokchilarning shaxsiy hayoti va himoyasini ta'minlashni o'z ichiga oladi.

Ma'lumotlar yig'ish

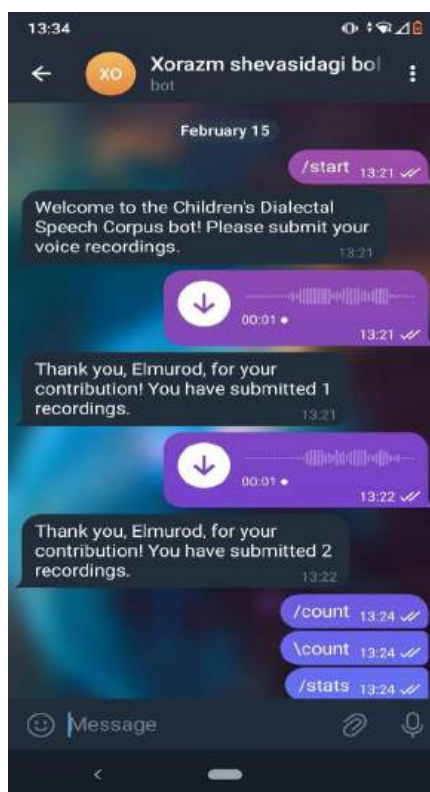
An'anaga ko'ra, ma'lumotlarni yig'ish maktablar yoki tadqiqot markazlari kabi boshqariladigan muhitda sessiyalar tashkil qilishni o'z ichiga oladi. Biroq, kengroq demografiyaga erishish va jarayonni soddalashtirish uchun biz Telegram botidan foydalangan holda yangi yondashuvni qo'lladik. Bu bot ota-onalar va vasiylar bilan muloqot qilish uchun mo'ljallangan bo'lib, ular o'z farzandlarining nutqini bevosita platforma orqali yozib olishlari va yuklashlari mumkin. Bu usul tanish muhitda tabiiy nutqni to'plash, bolalarning haqiqiyliги va qulayligini oshirish imkonini beradi.



1-rasm. Nutq korpusini yig'ish platformasi QR kodi

Telegram botdan foydalanish

Telegram boti barcha ishtirokchilarning to'liq xabardor bo'lishini va loyiha shartlariga rozi bo'lishini ta'minlab, rozilik olish jarayonida foydalanuvchilarni yo'naltirish uchun dasturlashtirilgan. Rozilik olingandan so'ng, bot nutq namunalarini yozib olish va yuklash bo'yicha ko'rsatmalar beradi. U to'plangan ma'lumotlarning xilma-xilligini ta'minlash uchun muayyan turdagi nutq yoki mavzularni taklif qilishi mumkin. Masalan, u hikoyalar, kundalik faoliyatning tavsifi yoki turli til elementlarini keltirib chiqaradigan savollar va takliflarga javoblarni so'rashi mumkin. Quyida Telegram botining ko'rinishi



2-rasm. Nutq korpusini yig'ish uchun yaratilgan Telegram bot interfeysi

Birlashtirish va saqlash

Barcha yig'ilgan nutq ma'lumotlari avtomatik ravishda markazlashtirilgan serverga yuboriladi, u yerda saqlanadi va tizimli ravishda tashkil etiladi. Bu server boshqa omillar qatori yosh, dialekt va nutq turiga qarab ma'lumotlarni toifalarga ajratuvchi ombor vazifasini bajaradi. Ushbu tizimli yig'ish va saqlash keyingi bosqichlarda ma'lumotlarni samarali qayta ishlash va tahlil qilishni osonlashtiradi.

Sifat nazorati va axloqiy mulohazalar

To'plash jarayonida yuqori ma'lumotlar sifati va axloqiy me'yorlarni saqlash muhim ahamiyatga ega. Tizim yozuvlarning ravshanligi va foydalanish qulayligini tekshirish mexanizmlarini o'z ichiga oladi va ishtirokchilarning maxfiyligini himoya qilish uchun barcha ma'lumotlarning anonim va xavfsiz saqlanishini ta'minlaydi.

Bolalar nutqi korpusini to'plash uchun Telegram botidan foydalanish orqali biz O'zbekiston bo'ylab turli demografik guruhlardan keng ko'lamli nutq namunalarini samarali to'plashimiz mumkin. Bunday yondashuv nafaqat to'plash jarayonini soddalashtiribgina qolmay, balki turli sheva va nutq turlarining ifodalanishini ta'minlab, pirovardida o'zbek tili uchun boy va qimmatli lingvistik manba bo'lishiga hissa qo'shadi.

Xulosa

O'zbek tili bo'yicha bolalar nutqi korpusini yaratish tashabbusi lingvistik resurslarni ko'paytirish va tilni qayta ishlashda texnologik yutuqlarni qo'llab-quvvatlash yo'lidagi muhim qadamdir. Bu sa'y-harakatlar nafaqat yosh

so'zlovchilar uchun moslashtirilgan o'quv va texnologik vositalarni ishlab chiqishda, balki o'zbek tilining lingvistik rang-barangligini saqlashda ham muhim rol o'ynaydi.

Kelajakdagi ish sifatida loyiha O'zbekistonning turli mintaqalari va shevalari bo'yicha yanada xilma-xil nutq namunalarini to'plash orqali o'z ma'lumotlar bazasini kengaytirishni maqsad qilgan. Ushbu kengayish yanada mustahkam mashina o'rganish va sun'iy intellekt modellarini yaratish imkonini beradi, nutqni aniqlash va bolalar nutqi uchun sintez texnologiyalarini yaxshilaydi. Yakuniy maqsad bu yutuqlarni ta'lim platformalari va interfaol ilovalarga integratsiya qilish, shu tariqa O'zbekistondagi bolalar uchun ta'lim tajribasini va texnologik foydalanish imkoniyatlarini boyitishdir.

Foydalanilgan adabiyotlar:

1. Salaev, U. I., Kuriyozov, E. R., & Matlatipov, G. R. (2023, November). Design and Implementation of a Tool for Extracting Uzbek Syllables. In *2023 IEEE XVI International Scientific and Technical Conference Actual Problems of Electronic Instrument Engineering (APEIE)* (pp. 1750-1755). IEEE.
2. Madatov, K. (2019). A prolog format of uzbek WordNet's entries. *Human Language Technology as a Challenge for Computer Science and Linguistics*, 316-320.
3. Allaberdiev, B., Matlatipov, G., Kuriyozov, E., & Rakhmonov, Z. (2024). Parallel texts dataset for Uzbek-Kazakh machine translation. *Data in Brief*, 110194.
4. Kutlimuratova, B., Kuriyozov, E., & Tillaeva, M. (2022). Teaching english as a foreign language for primary school children: literature reveiw. *Foreign language teaching and applied linguistics*, 161-171.
5. Sharipov, M., Salaev, U., & Matlatipov, G. (2021). O'zbek tili fe'l so'z turkumi uchun chekli avtomatlar asosida stemming algoritmini yaratish. *Computer linguistics: problems, solutions, prospects*, 1(1).
6. Kutlimuratova, B. (2021). Uzbek students learning English as a foreign language: Error analysis using corpora.
7. Sharipov, M., & Salaev, U. (2022). Uzbek affix finite state machine for stemming. *arXiv preprint arXiv:2205.10078*.
8. Madatov, K., Bekchanov, S., & Vičič, J. (2023). Uzbek text summarization based on TF-IDF. *arXiv preprint arXiv:2303.00461*.

O‘ZBEK TILI KORPUSIDA TEGGLASH METODLARI VA USULLARI

O‘dayeva Shahnoza Shuxratovna

BuxDU mustaqil izlanuvchisi

Annotatsiya. Ushbu maqolada o‘zbek tili korpusida teglash metodlari va usullari xususida so‘z yuritiladi. O‘zbek milliy korpusida teglashdan foydalanishning o‘ziga xos jihatlari mavjud. Qayd etish lozimki, teglash jarayoni ma’lum qoidalarga asoslanib, undan foydalanish orqali tilda mavjud elementlarni modellashtirish mumkin.

Abstract. This article talks about tagging methods and techniques in the Uzbek language corpus. The use of tagging in the Uzbek national corpus has its own characteristics. It should be noted that the tagging process is based on certain rules, and by using it, it is possible to model the elements present in the language.

Аннотация. В данной статье рассказывается о методах и приемах разметки в корпусе узбекского языка. Использование тегов в узбекском национальном корпусе имеет свои особенности. Следует отметить, что процесс разметки основан на определенных правилах, и с их помощью можно моделировать элементы, присутствующие в языке.

Kalit so‘zlar: *korpus lingvistikasi, teglash, modellashtirish, qolip, asos, til birliklari.*

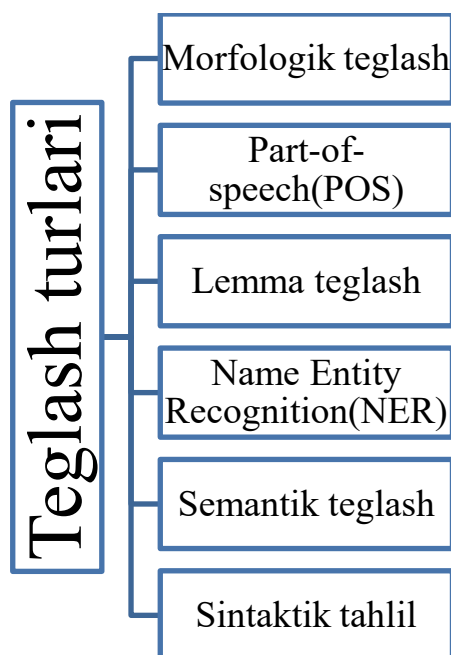
Ta’lim tizimi taraqqiy etgani sari uning tarkibida yangidan yangi sohalar ajralib chiqmoqda. Xususan, tilshunoslik sohasini oladigan bo‘lsak, ushbu fanning yangi integratsion ko‘rinishlari vujudga kelmoqda, jumladan, psixolingvistika, lingvokulturologiya, lingvoma’naviyatshunoslik kabilar shular jumlasidandir. Alqissa, kompyuter lingvistikasi tarmog‘ida shakllangan korpus lingvistikasi ham shular jumlasidandir. Korpus lingvistikasi til tarkibidagi lug‘at bazasini korpuslashtirish masalasi bilan shug‘ullanuvchi soha sanalib, bizning ushbu maqolamiz ham aynan bu yo‘nalishni qamrab oladi.

Ma’lumki, bugungi kunda tilshunoslikda korpus sohasi, ayniqsa, taraqqiy etmoqda. Korpus lingvistikasida teglash matn korpusidagi so‘zlarga grammatik, morfologik, sintaktik yoki semantik ma’lumotlarni qo‘shish jarayonini bildiradi. Bu teglash lingvistik tahlil uchun korpusdan foydalanish imkoniyatini oshiradi. Korpus lingvistikasida teglardan foydalanish vaqt o‘tishi bilan rivojlanib, texnologiya, lingvistik nazariyalar va tadqiqot metodologiyasidagi yutuqlarni aks ettirdi. Korpus lingvistikasida teglar tarixini hisoblash tilshunosligining dastlabki rivojlanishi va matn korpusida tizimli va tuzilgan lingvistik annotatsiyaga bo‘lgan ehtiyoj bilan bog‘lash mumkin [Charniak, 1997:89]. Korpus lingvistikasining dastlabki bosqichlarida tadqiqotchilar nutq qismi teglari, lemma shakllari va grammatik kategoriyalar kabi lingvistik ma’lumotlarga ega bo‘lgan matn korpusiga qo‘lda izoh

berishgan. Ushbu qo'lda izohlash jarayoni ko'p vaqt talab qildi va keng miqyosda cheklangan edi, ammo avtomatlashtirilgan teglashning keyingi ishlanmalari uchun asos yaratdi. Davrlar o'tishi bilan olib borilgan tadqiqotlar tufayli teglash jarayonining yanada taraqqiy etgan ko'rinishi yuzaga keldi.

Teglar turli tillarda amalga oshirilgan va ishlatiladigan teglar to'plami tilga qarab katta farq qiladi. Teglar, odatda, ochiq morfologik farqlarni o'z ichiga olishi uchun mo'ljallangan, ammo bu ingliz tilidagi otlar uchun emas, balki olmoshlar uchun harf belgilari kabi nomuvofiqliklarga va tillararo kattaroq farqlarga olib keladi. Yunon va lotin kabi murakkab fleksiyaga ega tillar uchun teglar miqdori juda katta bo'lishi mumkin; Inuit tillari kabi agglutinativ tillardagi so'zlarni belgilash deyarli imkonsiz. Boshqa tomondan, Petrov va boshqalar 12 toifadan iborat (masalan, otlar, fe'llar, tinish belgilari va boshqalarning pastki turlari yo'q) “universal” teglar to'plamini taklif qildilar [Charniak, 1997:56]. Juda keng teglarning juda kichik to'plami yoki aniqroq bo'lganlarning kattaroq to'plami afzalroq bo'ladimi, qo'llanilayotgan maqsadga bog'liq. Avtomatik teglash kichikroq teglar to'plamidan foydalanish jarayonini osonlashtiradi.

Korpus lingvistikasida qo'llaniladigan teglashning xilma-xil turlari mavjud. Ularni quyidagicha guruhlash mumkin [1]:



1. Part-of-speech (POS) teglash. POS teglash korpusdagi har bir so'zga grammatik kategoriya belgilashni o'z ichiga oladi. Ya'ni so'zning morfologik xususiyatlarini aniqlashda ushbu teglashdan foydalanish samarali hisoblanadi. Ushbu teglashga otlar, fe'llar, sifatlar, ergash gaplar, olmoshlar, bog'lovchilar, yuklamalar va boshqalarni teglashni kiritish mumkin. POS belgilari jumla tuzilmalarini tahlil qilish va lingvistik tadqiqotlar o'tkazishda yordam beradi.

Gap tuzilmasi	Men	Musiq	tingladim
POS teglash modeli	Pronoun(Olmosh)	Noun(Ot)	Fe'l(Verb)

Demak, keltirilgan misoldan ko‘rinib turibdiki, teglashning ushbu guruhida umumiy tarzda (morfologik va sintaktik) xususiyatlar aniqlashtiriladi va uning quyidagi teglari mavjud:

JJ – sifat (adjective)

JJR – sifatning qiyosiy darajasi (adjective, comparative)

JJS – sifatning ortirma darajasi (adjective, superlative)

MD – modal so‘zlar

NN – ot (noun)

NNS – ko‘plikdagi ot va hokazo [Wikipedia.org].

Yuqorida nomlari keltirilgan teglash usullari miqdoran ko‘p bo‘lib, teglash jarayonida so‘zning qaysi grammatik xususiyatini aks ettirishiga binoan tanlanadi.

2. Lemma teglash: Lemma teglash so‘zning asosiy yoki lug‘at shaklini aniqlashni o‘z ichiga oladi. Masalan, “yugurish” lemmasi “yugurmoq”dir. Ushbu teglash turli xil qo‘shimchalarni o‘zida jamlagan so‘zlarni tahlil qilishni soddalashtirishga yordam beradi va tadqiqotchilarga so‘zning asosiy ma‘nosiga e‘tibor qaratish imkonini beradi. Lemma teglashni quyidagi jadval asosida ko‘rib chiqamiz:

№	So‘zning turli shakllari	Lemma
1	Yugurish, yuguryapti, yugurajak, yugursin, yuguringlar	Yugurmoq
2	Daftari, daftarga, daftardan, daftarsiz, daftarni	Daftar
3	Qizilroq, qip-qizil, qizg‘ish, eng qizil, sal qizil	Qizil
4	Yozish, yozilajak, yozib, yozdir, yozildi, yozishmoqda	Yozmoq

3. Named Entity Recognition (NER) teglash: NER teglashi korpus ichidagi shaxslar, tashkilotlar, manzillar, sanalar va hokazo nomlari kabi nomli obyektlarni aniqlash va tasniflashni o‘z ichiga oladi. Bu teglash ma‘lumot olish, matnni umumlashtirish va obyektlarni bog‘lash kabi vazifalar uchun foydalidir.

Kamil A‘zamovichning faoliyati Turkiyada 12- may kundan boshlanadi.	Shaxs(person)	Kamil A‘zamovich
	Joylashuv(Location)	Turkiya
	Sana(Date)	12-may

Yoki:

25-mayda “Tilshunoslar” guruhining tashkil topganligiga 8 yil to‘ladi	Tashkilot nomi(Organization)	“Tilshunoslar”
	Sana(Date)	25-may
	Qo‘shimcha ma’lumot(extra information)	8 yil

4. Sintaktik tahlil: Sintaktik tahlil korpusdagi gaplarning grammatik tuzilishini tahlil qilishni o‘z ichiga oladi. Bu so‘zlar va birikmalar o‘rtasidagi munosabatlarni aniqlashga yordam beradi, shu bilan birga, lingvistik tuzilmalarni sintaktik tahlil qilish imkonini beradi. Ushbu tahlil turi “Part-of-speech” tahliliy guruhiga kiritilib, uning alohida ko‘rinishi sanaladi.


Bugun havo yaxshi [Hans, 2001:69].

Keltirilgan misolda ko‘rsatilgan teglar orqali ifodalangan gapning bog‘lanish tuzilmasi aks ettirilmoqda.

5. Semantik teglash. Semantik teglash jumladagi so‘zlarning semantik rollarini aniqlashni o‘z ichiga oladi. Bu teglash jumladagi yoki so‘zlar orasidagi ma’no va munosabatlarni tushunishga yordam beradi. Ushbu teglash turi o‘z tarkibida qator guruhlarni mujassam etadi, xususan: anaforik, sifat teglash, konsept teglash, mantiqiy teglash, diskurs teglash, aspektga oid teglash. Bunda so‘zning ma’noviy xususiyati e’tiborga olinadi:

Sifat teglash	COL	<i>Colour/rang</i>	<i>Qizil, yashil, sariq, qora</i>
	QUC	<i>Quantity/miqdor</i>	<i>Ikki, yetti, o‘n besh, to‘qqiz</i>
	IST	<i>Intersective/xususiyat</i>	<i>Samimiy, yaxshi, ochiqko‘ngil</i>

6. Morfologik teglash: Morfologik teglash korpus tarkibidagi so‘zlarning morfologik belgilarini, masalan, old qo‘shimchalar, qo‘shimchalar, zamon belgilari, ko‘plik belgilari va boshqalarni tahlil qilishni o‘z ichiga oladi. Bu teglash so‘z shakllari, qo‘shimchalar va hosilalarni tushunishga yordam beradi.

Korpus tilshunosligida teglash xususida so‘z yuritar ekanmiz, uning ham o‘ziga yarasha kamchiliklari mavjud ekanligini qayd etishimiz lozim. To‘g‘ri, teglash so‘zlarni modellashtirish jarayonini osonlashtirish, biroq so‘zlarni teglash ayrim murakkablik tug‘diradi. Quyidagi misolga e’tibor bering:

Oting nima?

Qo'lingdagi to'pni bizga ot.

Birinchi gapda keltirilgan ot so'zi ot so'z turkumiga taalluqli, ikkinchi birlik esa fe'l so'z turkumining buyruq maylida qo'llanmoqda. Mana shu kabi bir xil shaklli so'zlar(omonimlar)ni korpus lingvistikasida shakllantirish teglash jarayonida birmuncha qiyinchilikka yuz tutadi. Ushbu muammoning yechimi sifatida so'zlar bilan bog'lanib kelgan birliklarni aniqlik bilan teglash masalasini keltirish mumkin. Sababi, birikmalarning aniq teganishi kompyuter dasturida modellashtirish jarayonining osonlik bilan kechishiga yordam beradi.

Xulosa sifatida aytish mumkinki, ushbu teglash usullarini korpus lingvistikasida qo'llash orqali tadqiqotchilar lingvistik tahlilning chuqurligi va ko'lamini oshirishlari, qimmatli tushunchalarni olishlari va til ma'lumotlar to'plamidagi tuzilmalarni aniqlashlari mumkin.

Foydalanilgan adabiyotlar:

1. Charniak, Eugene. 1997. "Statistical Techniques for Natural Language Parsing". AI Magazine 18(4):33–44.
2. Hans van Halteren, Jakub Zavrel, Walter Daelemans. 2001. Improving Accuracy in NLP Through Combination of Machine Learning Systems. Computational Linguistics. 27(2): 199–229. PDF
3. DeRose, Steven J. 1990. "Stochastic Methods for Resolution of Grammatical Category Ambiguity in Inflected and Uninflected Languages." Ph.D. Dissertation. Providence, RI: Brown University Department of Cognitive and Linguistic Sciences. Electronic Edition available at 2.
4. D.Q. Nguyen, D.Q. Nguyen, D.D. Pham and S.B. Pham (2016). "A Robust Transformation-Based Learning Approach Using Ripple Down Rules for Part-Of-Speech Tagging." AI Communications, vol. 29, no. 3, pages 409–422.

ABDURAUF FITRAT ASARLARI KORPUSI MATNLARINI TO‘PLASH VA ULARGA DASTLABKI ISHLOV BERISH MUAMMOLARI

Alimbekova Mavjuda Xalimjon qizi

alimbekovamavjuda18@gmail.com

ToshDO‘TAU tayanch doktoranti

Annotatsiya. Ushbu maqolada Abdurauf Fitrat asarlari korpusi matnlarini to‘plash, yig‘ilgan asarlarni tekst holiga keltirish, matnga dastlabki ishlov berish masalalari yoritilgan. Matnga dastlabki ishlov berish jarayonida yuzaga kelgan qiyinchiliklar va unga berilgan yechimlar haqida fikr yuritilgan. Shuningdek, muallifning korpusga kiritish imkoni bo‘lmagan asarlari va buning sababi keltirilgan.

Abstract. This article covers the issues of collecting the texts of Abdurauf Fitrat authorship corpus, turning the collected works into text, and initial processing of the text. The difficulties that appeared during the initial processing of the text and solutions given to it are discussed. Also, the works of the author that can not included in the corpus and the reason for this are given.

Аннотация. В данной статье освещены вопросы сбора текстов корпуса Абдурауфа Фитрата, приведения в порядок собрания сочинений в текстовую форму и предварительной обработки текста. Обсуждаются проблемы, возникшие при первичной обработке текста, и пути их решения, а также приводятся авторские произведения и причины, из за которых не удалось включить их в корпус .

Kalit so‘zlar: *mualliflik korpusi, Fitrat asarlari matni, txt, matnlarni to‘plash, matnga dastlabki ishlov berish.*

Kirish

“O‘z mohiyat e‘tiboriga ko‘ra noyob ijtimoiy-siyosiy fenomen bo‘lgan jadidchilik g‘oyasining shakllanishi va taraqqiyotiga doir dunyoning bir qator mamlakatlarida ko‘plab tadqiqotlar yaratilgan bo‘lsa-da, ushbu harakat namoyandalarining Markaziy Osiyo hududida milliy davlatchilik va mintaqaviy o‘ziga xoslikni, fuqarolik jamiyatining rivojlanishiga qo‘shgan ulkan hissasini konseptual va tizimli asosda atroflicha o‘rganish dolzarb masala bo‘lib qolmoqda” [Mirziyoyev, 2023]. Mamlakatimiz prezidenti Shavkat Mirziyoyev 2023-yilning 11-12-dekabr kunlari Toshkentda bo‘lib o‘tgan “Jadidlar: milliy o‘zlik, istiqlol va davlatchilik g‘oyalari” mavzusidagi xalqaro konferensiya qatnashchilariga aytilgan tabrik nutqida jadidlarning hayoti va jasorati yoshlarni Vatanni ko‘z qorachig‘iday asrashga da‘vat ekanini, ularning ilmiy me‘rosini o‘rganish, tadqiq qilish millat kelajagi uchun ahamiyatli ekanini ta’kidladi.

Abdurauf Fitrat ijodi adabiyotshunoslik, tilshunoslik, tarix, pedagogika, siyosat, din, falsafa, musiqa sohaları uchun bugungi kunda ham ahamiyatli. Fitratning ijodiy merosi o'zbek millati uchun qanchalik ahamiyatli bo'lsa, fors-tojik adabiyotshunos, tilshunoslari uchun ham birday qadrlidir.

Asosiy qism

Shoir Rauf Parfi Fitratning 4 ta: “Sayha”, “Uchqun”, “She’rlar va doston”, “Mening tunim” she’riy to’plami mavjud ekanini yozgan [Parfi]. “Sayha” she’riy to’plami vatanparvarlik ruhida yozilgan she’rlardan iborat ekani, o’z davrida xalq orasida katta shov-shuvlarga sabab bo’lganini Fayzulla Xo’jayev ta’kidlab o’tgan, ammo “Uchqun”, “She’rlar va doston”, “Mening tunim” nomli to’plamlari va u haqidagi ma’lumotlarni boshqa fitratshunoslar tadqiqotlarida uchratmadik. 1922-yilda Fitrat boshchiligida “O’zbek yosh shoirlari” nomli she’riy to’plam yaratilgan bo’lib, unga Fitratning she’rlari bilan birga Cho’lpon, Elbek va Botuning ba’zi she’rlari jamlangan. Ushbu to’plamga Fitratning o’n uchta she’ri kiritilgan bo’lib, ular 1919-1921-yillarda yozilgan she’rlar sanaladi va quyidagilarni o’z ichiga oladi: “Kim deyay seni?”, “Biroz kul”, “Achchig’lanma degan eding”, “Mirrix yulduziga”, “Ovunchoq”, “Ishqimning tarixi”, “Yana yondim”, “Nega bo’yla?”, “Shoir”, “Behbudiyni sag’anasini izladim”, “Parcha”, “O’gut”, “Sharq”. Fitrat 1917-yilda “Hurriyat” gazetasida “Yurt qayg’usi” nomli she’riy turkumini e’lon qilgan. Ulardan bittasi aruz vaznida yozilgan, “Yurt qayg’usi (sochma)”, “Yurt qayg’usi (bir o’zbek tilidan)”, “Yurt qayg’usi (Temur oldinda)” nomli she’rlari sochma she’r hisoblanadi. Fitrat “Adabiyot qoidalari” kitobida bunday she’rlarni “mansura” deb atagan. Shuningdek, shoirning “O’qitg’uvchilar yurtiga”, “Qor”, “Mening kecham”, “Go’zalim, bevafo gulistonim” kabi she’rlari ham mavjud. Ular to’plamga kirmagan she’rlaridir. Hamidulla Boltaboyev boshchiligida chop etilgan Fitratning besh jildlik “Tanlangan asarlar”ining I jildiga Fitratning yuqorida sanab o’tilgan 21 ta she’ri kiritilgan. Korpusga kiritish uchun ushbu she’rlar olindi.

Uning dramatik asarlaridan “Begijon”, “Qon”, “Abo Muslim”, “O’g’izxon” pyesalari eski o’zbek yozuvidan lotin alifbosiga o’girilmagan.

“Temur sag’anasi” dramasi bizgacha yetib kelmagan. Hamid Olimjon Fitrat faoliyatini tanqid qilib maqola yozgani ma’lum, shunga tayanib H. Boltaboyev H.Olimjonda ushbu asarning biror bir nusxasi bo’lishi mumkinligini yozadi. Ushbu dramadan quyidagi parcha yetib kelgan, xolos: “Men sizlarga ko’p narsalar yetkardim. Ne bo’ldi, bir zamonlar sharaflı va jasur bir millatning zulm ostida qolmish. Kimlar bog’laringdan qushlarimni quvmoqda!... Otalar merosidan nimani oldingiz! Sizdan talab etaman, qalqingiz!... O’lkani tuzatingiz, avlodlarimning hur yashamoqlarini ta’min etingiz. Agar bunday qilmasangiz, o’lka bir buyuk mozorlik holiga kelur”. Ushbu parcha “Yosh leninchi” jurnalida 1935-yilda e’lon qilingan. [Fitrat, 2000: 18]

1934-yilda yozilgan “To’lqin” opera librettosi bizga qadar yetib kelmagan. Lotin alofbosida chop etilgan “Abulfayzxon”, “Vose’ qo’zg’oloni”, “Chin sevish”,

“Arslon”, “Hind ixtilochilari”, “Shaytonning tangriga isyoni” dramalari korpus matni uchun olindi.

“Sarfi zaboni tojik”, “Cho‘zg‘ilar”, “Bitim yo‘llari”, “Amir Olimxonning hukmronlik davri”, “Turk tili”, “Sharq shahmati” asarlari va Qayum Ramazon, Shokirjon Rahimiylar bilan hammualliflikda chop etilgan “Ona tili” darsligi lotin alifbosida chop etilmagani bois ushbu asarlar ham Abdurauf Fitrat asarlari korpusi matni uchun olinmadi.

Fitratshunos olim H.Boltaboyev boshchiligida chop etilgan Fitratning V tomlik tanlangan asarlariga kiritilmagan “Me’roj”, “Zayd va zaynab”, “Oqmozor”, “Zahroning iymoni” nomli hikoyalari Sh.Turdiyev nashrga tayyorlagan “Hiylaiy shar’iy. Kichik hajviy asarlar” to‘plamidan olindi. Ushbu to‘plam 1973-yili “Adabiyot va san’at nashriyoti” tomonidan O‘zSSR FA akademigi, muxbir a’zosi I.Sultonov tahriri ostida nashr qilingan.

Korpus matnlari uchun yig‘ilgan asarlar va ularning janrlari quyidagi jadvalda berilgan:

T/r	Abdurauf Fitrat ijodi janrlari	Soni	Asar nomlari
1.	She’r	17	“Achchig‘lanma degan eding”, “Behbudiyning sag‘anasin izladim”, “Bir oz kul”, “Go‘zalim, bevafo gulistonim”, “Ishqimning tarixi”, “Kim deyay seni”, “Mening kecham”, “Mirrix yulduziga”, “Nega bo‘yla”, “O‘gut”, “O‘qitg‘uvchilar yurtiga”, “Ovunchoq”, “Parcha”, “Qor”, “Sharq”, “Shoir”, “Yana yondim”
2.	G‘azal	1	“Yurt qayg‘usi”
3.	Mansura	3	“Yurt qayg‘usi (bir o‘zbek tilidan)”, “Yurt qayg‘usi (Sochma)”, “Yurt qayg‘usi” (Temur oldinda)
4.	Qissa	1	“Hind sayyohi bayonoti”
5.	Hikoya	5	“Qiyomat”, “Me’roj”, “Zayd va Zaynab”, “Oq mozor”, “Zahroning iymoni”
6.	Drama	6	“Abulfayzxon”, “Arslon”, “Shaytonning tangriga isyoni”, “Hind ixtilochilari”, “Chin sevish”, “Vose’ qo‘zg‘oloni”
7.	Ocherk	1	“Bedil”



8.	Maktub	2	“Buxoro vaziri Nasrullohbey parvonachi afandi hazratlarina ochiq maktub”, “Yopishmagan gajaklar”
9.	Maqola	11	“Angliz va Turkiston”, “Buxoro ulamosi”, “Buxoroda inqilob”, “Buxoroning holi”, “Johilona taassubg‘a misol”, “Musulmonlar, g‘ofil qolmang!”, “Muxtoriyat”, “Sharq siyosati”, “Sho‘royi Islomiyanning xatosi”, “Siyosiy hollar”, “Turkistonda ruslar”.
	Ilmiy	18	“Ahmad Yassaviy”, “Chig‘atoy adabiyoti”, “Farhod va Shirin dostoni to‘g‘risida”, “Fors shoiri Umar Hayyom”, “Hibat ul-haqoyiq”, “Mashrab”, “Muhammad Solih”, “Muqaddimat ul-adab”, “Navoiyning forsiy shoirlik‘i ham uning forsiy devoni xususida”, “O‘zbek shoiri Turdi”, “Qutadg‘u bilig”, “She‘r va shoirliq (Adabiy mubohasa)”, “Tilimiz I”, “Tilimiz II”, “Tilimizninng abadiyligi”, “XVI asrdan so‘ngra o‘zbek adabiyotig‘a umumiy bir qarash”, “Yassaviy maktabi shoirlari to‘g‘rusida”
10.	Munozara	1	“Hindistonda bir farangi ila Buxoroli mudarrisning jadid maktablari xususida qilg‘an munozarasi”
11.	Qo‘llanma	3	“O‘zbek tili qoidalari to‘g‘risida bir tajriba. Birinchi kitob: Sarf”, “O‘zbek tili qoidalari to‘g‘risida bir tajriba. Ikkinchi kitob: Nahv”, “Adabiyot qoidalari”.
12.	Darslik	1	“O‘quv”
13.	Majmua	1	“Eng eski turk adabiyoti namunalari”
14.	Risola	4	“Aruz haqida”, “Rahbari najot (“Najot yo‘li”)", “Oila yoki oilani boshqarish tartiblari”, “Muxtasar islom tarixi”

Korpus matnlariga dastlabki ishlov berish muammolari. Pdf formatidagi matnlarni txt formatiga o‘tkazish jarayonida qadimgi turkiy til tadqiqiga bag‘ishlangan asari “Eng eski turk adabiyoti namunalari” asarida arab alifbosida

berilgan ifodalar mavjud. Bunda muallif “Devoni lug‘at”dagi harf va cho‘zg‘ilarning eski o‘zbek alifbosida qanday fodalanishini keltirib o‘tgan. “Sublime text” dasturida matnni “text” formatiga keltirayotganimizda ushbu o‘rinda qiyinchiliklar yuzaga keldi. Bunday arabiy ifodalar, harf va cho‘zg‘ilarning o‘zbek tilidagi nomini yozish bilan cheklanishni ma’qul topdik.

“Adabiyot qoidalar” asarida tarix, muammo janrlari tavsifida arab harflari, “Abjad hisobi” yozilgan o‘rinlar mavjud:

Dedi ta’rixini duo birla,

Komil xoksor besomon.

Avj izzat uza muqim o‘lsun

“نور عين سيد محمد خان”

(Nuri ayni Said Muhammadxon)

Ushbu ta’rixda “نور عين سيد محمد خان” misrasi ta’rix janrining asosi hisoblanadi. Ushbu imlo bilan yozilganda to‘rtinchi misraning soni abjad hisobiga ko‘ra 1203 bo‘ladi. Korpusda bunay o‘rinlarda kelgan arab alifbosida yozilgan jumla ahamiyatli bo‘lgani bois txt formatida ham aynan shunday holatda kiritishni o‘rinli deb hisobladik. Qidiruv tizimida ushbu ta’rix yaxlit holda chiqishi uchun quyidagi tartibda bir satrga keltirildi:

[*Dedi ta’rixini duo birla,*][*Komil xoksor besomon.*][*Avj izzat uza muqim o‘lsun*][*“نور عين سيد محمد خان”*](Nuri ayni Said Muhammadxon)]

Shuningdek, “Aruz haqida” risolasida aruz qoidalariga doir chizma, taqte’lar, jadval ko‘rinishida — ustun shaklda berilgan ma’lumotlar ham ko‘p uchraydi. Misol sifatida quyidagi parchani keltirishimiz mumkin:

Qo‘lg‘a olib dombiram, chertar uni tirnog‘im,

Bordi-keldi maqomim kelishtirg‘an barmog‘im.

Tilim zikrin so‘ylayib, ko‘nglim fikrin o‘ylayib,

So‘ylaganda ko‘runar ko‘rguligi (ko‘rgulik?) oq tomog‘im.

-V|V-|-V-|--|VV|-V|-

-V|-V|VV|-|V-|--|-V|-

V|--|-V-|--|--|-V|-

-V|-V|VV|-|-V|--|VV|-

Asardagi tahlil qilinib, taqte’si chizilgan bunday she’riy parchalarning har bir misrasini chizmalari bilan birgalikda ifodalash uchun bir satrga yaxlit holda quyidagicha kiritishni lozim topdik:

[*Qo‘lg‘a olib dombiram, chertar uni tirnog‘im,*][*-V|V-|-V-|--|VV|-V|-*][*Bordi-keldi maqomim kelishtirg‘an barmog‘im.*][*-V|-V|VV|-|V-|--|-V|-*][*Tilim zikrin so‘ylayib, ko‘nglim fikrin o‘ylayib,*][*V|--|-V-|--|--|-V|-*][*So‘ylaganda ko‘runar ko‘rguligi (ko‘rgulik?) oq tomog‘im.*][*-V|-V|VV|-|-V|--|VV|-*]

Xulosa

Abdurauf Fitrat mualliflik korpusini yaratish uchun lotin va kiril alifbolarida nashr etilgan barcha asarlari “txt” holatiga keltirildi. Yuqorida sanab o‘tilgan - eski

o'zbek yozuvidan lotin alifbosiga yoki kirill alifbosiga tabdil qilinmagan, qayta nashr etilmagan yoki qo'lyozma holida saqlanib kelinayotgan asarlari kiritilmadi. Jadid namoyondalari asarlarini o'rganish, tadqiq etishga e'tibor kun sayin ortib bormoqda. O'rganilishni, tadqiq etilishini kutayotgan yozma yodgorliklar talaygina. Ushbu soha egalari tomonidan Fitratning bu turdagi asarlari lotin, kiril alifbolariga tabdil qilinib, nashr etilsa, korpus matnlarini yanada kengaytirishimizga imkon yaratilgan bo'ladi.

Foydalanilgan adabiyotlar:

1. Fitrat A. Tanlangan asarlar. I jild. – T.: Ma'naviyat, 2000. – 256 b. – B. 18.
2. Mirziyoyev Sh. “Jadidlar g'oyasi —Yangi O'zbekiston strategiyasi bilan har tomonlama uyg'un va hamohang”. — URL: <https://daryo.uz/2023/12/11/prezidentning-jadidlar-milliy-ozlik-istiqlol-va-davlatchilik-goyalari-mavzusidagi-xalqaro-konferensiya-ishtirokchilariga-tabrigi-elon-qilindi>
3. Parfi R. Fitrat she'riyati. — URL: <https://kh-davron.uz/kutubxona/uzbek/rauf-parfi-fitrat-va-cholpon-sheriyati-haqida.html>

O‘ZBEK TILIDAGI OMONIM SO‘ZLAR KORPUSINI YARATISH

Qurbondurdiyeva Laylo Madiyor qizi

layloqurbondurdiyeva40@gmail.com

UrDU magistranti

Rahimbayeva Muhabbat Dushambayevna

mrd_aspi@mail.ru

UrDU Kompyuter ilmlari kafedrası dotsenti

Kuriyozov Elmurod Rajabboy o‘g‘li

elmurod1202@urdu.uz

UrDU Kompyuter ilmlari kafedrası o‘qituvchisi

Annotatsiya. O‘zbek tilidagi omonim so‘zlar bazasini mobil ilova ko‘rinishida yaratilishi lingvistik resurslarga va til o‘rganish vositalariga bo‘lgan ehtiyojini hal etib, sohadagi mavjud bo‘shliqni bartaraf etadi. Ushbu maqola o‘zbek tili omonimlarining keng qamrovli so‘zlar bazasini tuzish, ularni o‘rgatish va tushunishni osonlashtirish uchun interaktiv mobil ilovani ishlab chiqishga qaratilgan. Ishlab chiqilgan lug‘at foydalanuvchiga til o‘rganishga yordam beradi va omonim so‘zlardagi lingvistik aniqlikni oshiradi va o‘zbek tilining boy lingvistik merosini saqlashni qo‘llab-quvvatlashga qaratiladi.

Abstract. The creation of a database of Uzbek homonyms in the form of a mobile application addresses the existing void in this area, addressing the need for linguistic resources and language learning tools. The purpose of this article is to develop an interactive mobile application to create an extensive vocabulary database of Uzbek homonyms to facilitate their study and understanding. The developed dictionary helps the user learn the language by increasing the linguistic accuracy of homonyms and is aimed at promoting the preservation of the rich linguistic heritage of the Uzbek language.

Аннотация. Создание базы данных узбекских омонимов в виде мобильного приложения устраняет существующий пустоту в этой области, решая потребность в лингвистических ресурсах и инструментах изучения языка. Целью данной статьи является разработка интерактивного мобильного приложения для создания обширной словарной базы узбекских омонимов, чтобы облегчить их изучение и понимание. Разработанный словарь помогает пользователю изучить язык повышая лингвистическую точность омонимов и направлен на содействие сохранению богатого лингвистического наследия узбекского языка.

Kalit so‘zlar: *Omonimlar, shakl, ma’no, lug‘at, tasnif, tilshunoslik, leksika, tahlil, o‘zbek tili*

Kirish.

Omonimlar, ya’ni o‘xshash, ammo turli ma’nolarga ega so‘zlar o‘rganuvchiga qiyinchiliklar tug‘diradi. O‘zbek tilida o‘zining boy fonetik rang-barangligi va

tarixiy ta'siri bilan omonimlar, ayniqsa keng tarqalgan va foydalanuvchiga qiziqish uyg'otadi. Maqolaning maqsadi - bu omonimlarni tizimli ravishda to'plash va turkumlash, lingvistik tadqiqotlar, ta'lim va til texnologiyasini rivojlantirish uchun qimmatli manba yaratish. Keyingi mobil ilova o'zbek tilini o'rganuvchilar va so'zlashuvchilar uchun omonimlarning nozik jihatlarini o'rganish va tushunish uchun qulay platforma bo'lib xizmat qiladi, shu bilan ularning til bilimini boyitadi.

Lug'atlarda omonimlar haqida aniq tavsif va ma'lumotlar yo'q. Hayotning barcha sohalarini, ayniqsa, og'zaki nutqni qamrab oluvchi eng ko'p omonim turkumli lug'atlar tuzish o'zbek tilining dolzarb muammosidir. Bu tadqiqotchilarga barcha ekzistensial shakllarga tayanishga imkon beradi: badiiy til, shevalar, professional lug'at, jargon, jargon, vahshiylilik, qo'pollik va ekzotik lug'at. O'rganishlar shuni ko'rsatdiki, "O'zbek tilining omonimlari lug'ati" ("O'zbek tilining omonimlari lug'ati") lug'atida genetik jihatdan bir-biriga bog'liq bo'lmagan omonimlar qator a'zolarining birdan beshgacha bo'lgan sonlari berilgan. Demak, zil (o'zbek tilida) so'zining omonimlari – uch qator. Har bir qator bitta a'zodan iborat bo'lib, asl tilga ko'ra alohida guruhlariga kiradi. Birinchi ma'noda zil so'zi nemis tilidan olingan bo'lib, "og'ir" (og'ir) ma'nosida qo'llaniladi. Ikkinchi omonim turkiy tildan o'zlashtirilgan bo'lib, "musiqqa asbobi" (musiqqa asbobi) ma'nosini bildiradi. Uchinchi omonim zil arabchadan olingan bo'lib, "kamsitish" (xo'rlash) ma'nosini bildiradi.

Ma'lumotlar yig'ish

Ma'lumotlar yig'ish jarayoni o'zbek tilidagi omonimlarni aniqlash va tekshirish uchun keng qamrovli lingvistik tadqiqotlarni o'z ichiga oladi. Ushbu vazifa ma'lumotlar to'plamining aniqligi va to'liqligini ta'minlash uchun tilshunoslar, ona tilida so'zlashuvchilar va til o'qituvchilari bilan hamkorlik qilishni talab qiladi. Manbalarga lingvistik ma'lumotlar bazalari, ilmiy tadqiqotlar, an'anaviy adabiyotlar va zamonaviy ommaviy axborot vositalari kiradi. Ma'lumotlar to'plamidagi har bir yozuv omonimlarni yaxlit tushunishni ta'minlash uchun fonetik transkripsiyalar, ma'nolar, foydalanish misollari va etimologik ma'lumotlar bilan sinchkovlik bilan izohlanadi.

Biz o'zbek tilining izohli lug'atidan foydalangan holda omonim so'zlarni quyidagicha tasnif qildik.

"id" : 32,

"leksema" : "ALLA",

"miqdor" : "I",

"til" : "",

"turkum" : "ot",

"mano" : " ",

"misol" : "Beshikda qulog'ingga kirgan mungli va ayni chog'da nurli allami? O.Sharafiddinov, "O'lsam, ayrilmasman quchoqlaringdan""

```
"visits" : 0,  
"muallif_id" : 2,  
"sana" : "2022-11-11 12:40:31",  
"uslub" : "Go‘dakni uxlatishda yakkaxon usulida aytiladigan qo‘shiq."  
},  
{  
  "id" : 33,  
  "leksema" : "ALLA",  
  "miqdor" : "II",  
  "til" : "",  
  "turkum" : "fe'l",  
  "mano" : " ",  
  "misol" : "U, (Elmurod) kiyimlarini qoqar ekan, yigitlar atrofini  
o‘rashib, shovqin, kulgi bilan hazil qilishdi: -Bay-bay, zo‘r ekansiz-a! Polvonni  
alla qilib qo‘ydingiz-a! P.Tursun, O‘qituvchi; o‘zimizam alla qilib qo‘lga  
tushiruvdik, — dedi Soli sovuq. M.Ismoil, “Farg‘ona tong otguncha””,  
  "visits" : 0,  
  "muallif_id" : 2,  
  "sana" : "2022-11-11 12:40:31",  
  "uslub" : "1) charxpalak bo‘lmoq, chirpirak bo‘lib, umbaloq oshib  
yiqilmoq, qulamoq. Qoqilib ketib, alla bo‘ldi. Alla qilmoq; 2) Ko‘chma. boplab  
aldamoq, firib bermoq."  
},
```

Natijalar

Ushbu loyihaning asosiy natijasi mobil ilova uchun asos bo‘lib xizmat qiluvchi o‘zbek omonimlarining mustahkam ma’lumotlar to‘plamidir. Foydalanish uchun qulay interfeyslar va interaktiv o‘quv modullari bilan yaratilgan ilova foydalanuvchilarga o‘yinlar, viktorinalar va kontekstli mashqlar orqali omonimlarni o‘rganish imkonini beradi. Shuningdek, u omonimlarni talaffuz qilish va eshitishda farqlashda yordam berish uchun nutqni aniqlash va sintez kabi xususiyatlarni o‘z ichiga oladi. Foydalanuvchi testidan olingan dastlabki fikr-mulohazalar til o‘rganishga ijobiy ta’sir ko‘rsatadi, foydalanuvchilar kundalik muloqotda omonimlarni tushunish va ulardan foydalanish yaxshilangani haqida xabar berishadi. O‘quvchi omonim so‘zlar boyligini oshirish uchun biz ishlab chiqqan lug‘atda viktorina bo‘limiga kirib ishlay oladi. Talaffuzda farqlanmaydigan so‘zlarni chuqurroq o‘rganish uchun ishlab chiqilgan mobil lug‘atda mikrofon tugmachasini bosish orqali omonim so‘zlarning ma’nolarini o‘rganadi.

Xulosa

O‘zbek omonimlari uchun ma’lumotlar to‘plami va mobil ilovani yaratish til o‘rganish, texnologiyani rivojlantirish va madaniyatni saqlashga katta ta’sir

ko'rsatadi. Ushbu loyiha nafaqat o'quvchilarning tilshunoslik atamalarini chuqurroq tushinishiga yordam beradi, balki o'zbek tilini raqamlashtirish va asrab-avaylash kabi kengroq maqsadni amalga oshirishga ham hissa qo'shadi. Kelgusi ishlar o'zbek tili foydalanuvchilari uchun o'rganish tajribasini yanada qo'llab-quvvatlash va boyitish uchun ma'lumotlar to'plamini kengaytirish, ilova funksiyalarini yaxshilash va sun'iy intellekt asosidagi tilni qayta ishlashning ilg'or vositalarini integratsiyalashga qaratilgan.

Foydalanilgan adabiyotlar:

1. Salaev, U. I., Kuriyozov, E. R., & Matlatipov, G. R. (2023, November). Design and Implementation of a Tool for Extracting Uzbek Syllables. In 2023 IEEE XVI International Scientific and Technical Conference Actual Problems of Electronic Instrument Engineering (APEIE) (pp. 1750-1755). IEEE.
2. Madatov, K. (2019). A prolog format of uzbek WordNet's entries. Human Language Technology as a Challenge for Computer Science and Linguistics, 316-320.
3. Allaberdiev, B., Matlatipov, G., Kuriyozov, E., & Rakhmonov, Z. (2024). Parallel texts dataset for Uzbek-Kazakh machine translation. Data in Brief, 110194.
4. Kutlimuratova, B., Kuriyozov, E., & Tillaeva, M. (2022). Teaching english as a foreign language for primary school children: literature reveiw. Foreign language teaching and applied linguistics, 161-171.
5. Sharipov M., Salayev U., Matlatipov G. (2021). O'zbek tili fe'l so'z turkumi uchun chekli avtomatlar asosida stemming algoritmini yaratish. Computer linguistics: problems, solutions, prospects, 1(1).
6. Kutlimuratova B. (2021). Uzbek students learning English as a foreign language: Error analysis using corpora.
7. Sharipov M., Salaev U. (2022). Uzbek affix finite state machine for stemming. arXiv preprint arXiv:2205.10078.
8. Madatov, K., Bekchanov, S., Vičič, J. (2023). Uzbek text summarization based on TF-IDF. arXiv preprint arXiv:2303.00461.
9. Abdullaev, A., Yusupova, N. (2020). "Homonyms and Polysemous Words" textbook.
10. Rahmatillayev Sh. (1984). "Omonimlarning izohli lug'ati".
11. Smirnickij, A.I. (1956). Leksikologija anglijskogo jazyka. (260-bet). - Moskva.

MUALLIFLIK KORPUSLARINI YARATISH BOSQICHLARI (Usmon Azim dramalarining til korpusi dasturining yaratilishi misolida)

Maxammadiyeva Diyora Raxmatovna

diyoramaxammadiyeva@gmail.com

O'zbekiston-Finlandiya pedagogika instituti xodimi

Annotatsiya: Ushbu maqolada O'zbekiston xalq shoiri Usmon Azimning “Adibning umri” dramatik asarlar to'plamiga kiritilgan dramalari uchun mualliflik korpusining xususiyatlari va undan foydalanishning ahamiyati yoritilgan.

Abstract: This article describes the characteristics of the author's corpus and the importance of its use for the dramas of the national poet of Uzbekistan Usman Azim included in the collection of dramatic works "The Life of an Author"

Аннотация: В данной статье описаны особенности авторского корпуса и значение его использования для драм народного поэта Узбекистана Усмана Азима, включенных в сборник драматических произведений «Жизнь автора».

Kalit so'zlar: *Til korpusi, ta'limiy korpus, dramatik korpus, korpusdan foydalanish, konkordans, korpusning ahamiyati.*

Korpus lingvistikasi kompyuter lingvistikasining yangi tarmog'i sifatida so'nggi yillarda shakllandi va bu sohada bir qancha ilmiy, amaliy loyihalar amalga oshirildi. Har bir korpus, albatta, lingvistik tahrirdan, ishlov berishdan o'tgan va dasturiy ta'minot bilan ta'minlangan bo'lishi lozim. Bugungi kunda korpuslar leksikografiya va grammatika kabi tilshunoslik sohalarining ajralmas qismiga aylandi. Korpus paydo bo'lganidan so'ng tilshunoslik fanlari yangi umr yo'lini boshlashdi, aytish mumkinki, butun tilshunoslikni korpus tilshunosligi materiali sifatida qamrab olish mumkin.

O'rganilgan adabiyotlar natijasidan ma'lum bo'ldiki, korpusni loyihalashtirish jarayonida material (matn) tanlash, saralash, uni texnik jihatdan korpusga moslashtirish eng asosiy bosqichdir [Хамроева, 2018]. Usmon Azim dramalari til korpusi dasturini yaratish ishlarida lingvistik resurs sifatida ijodkorning “Adibning umri” [Usmon Azim, 2015] to'plami tanlandi va u quyidagi bosqichlarda amalga oshirildi:

1. Matnning belgilangan manbaga muvofiq holda korpusga kirishi ta'minlandi.

2. Matn avtomatik o'qilish shakliga keltirildi, ya'ni kiritiladigan matnning elektron shakldagi na'munasi Usmon Azimning “Adibning umri” dramatik asarlar to'plamidan skanerlash yo'li orqali docx formatda hosil qilindi.

3. Tahlil, matnga dastlabki ishlov berish bosqichida skaner qilingan matnlar dastur yordamida kril alifbosi shaklidan lotin alifbosi shakliga o'tkazildi. Shundan so'ng filologik tekshiruv, tahrirdan o'tkazildi. Bunda hosil qilingan matn kitobdagi

namunasi bilan solishtirildi, texnik jarayondagi imloviy, ishoraviy xatolar bartaraf etildi

4. Konversiyalash, grafematik tahlil. Bunda matnlar qayta kodlashtirish jarayoniga qadar mashina ishlovidan o'tkazildi, nomatniy qismlar (rasm, jadval) o'chirildi. Matndagi bo'g'in ko'chirish, chegaralar (MS-DOS matnlarida) bekor qilindi, tire, boshqa belgilar bir xilligiga erishildi. Grafematik tahlil korpusga kiruvchi matnni qismga (so'z, bog'lovchi) ajratish, nomatniy elementni o'chirish kabi amallarni bajarishdan iborat.

5. Mashina ishlovining so'nggi bosqichida nostandart (noleksik) elementni belgilash, rasmiylashtirish, maxsusmatniy elementni (qisqartma asosida yozilgan nom (ism, familiya), boshqa alifboda yozilgan o'zlashma leksema, rasmga berilgan nom, izoh, zarvara,q, adabiyotlar ro'yxati va b.) bir xil mezon asosida qayta ko'rib chiqish ishlari bajarildi. Albatta, bu amallar avtomatik ravishda matn muharriri tomonidan bajariladi.

Endi Usmon Azim dramalari til korpusining dasturiy ta'minot strukturasi haqida to'xtalamiz. Dasturiy ta'minot ikki qismdan iborat bo'ladi (2.1-rasm) [Qarshiyev, Tursunov, 2022]:

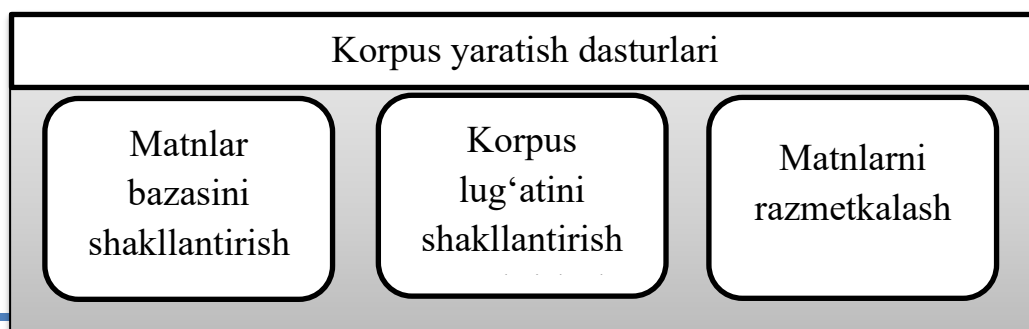
- 1) korpusni yaratishga mo'ljallangan dasturlar;
- 2) korpusdan foydalanishga xizmat qiladigan dasturlar.



2.1-rasm. Dasturiy ta'minot tarkibi

Korpus yaratishga mo'ljallangan dasturlar

Korpusni yaratishga mo'ljallangan dasturlar matnlar bazasini shakllantirish, korpus lug'atini shakllantirish va tahrirlash, matnlarni razmetkalash kabi vazifalarni bajarishi kerak (2.2-rasm) [Qarshiyev, Tursunov, 2022].



2.2-rasm. Korpus yaratish dasturlari tarkibiy qismlari

Matnlar bazasini shakllantirish dasturlari

Shuni eslatib o'tish kerakki, bizning ushbu tadqiqotimiz doirasida yaratilgan mualliflik korpusi ma'lum bir vazifalarni bajaradigan, foydalanuvchilarning ma'lum bir doirasiga yo'naltirilgan maxsus kichik korpus sanaladi.

Usmon Azim dramalari til korpusi dasturining matnlar bazasini shakllantirishda quyidagi ishlar bajarildi:

- matnlarni raqamlashtirish, ularni tahrirlash;
- matn bo'yicha ma'lumotlarni faylga yozish;
- matnni bazaga kiritish.

Matnlarni raqamlashtirishda qog'oz manbadagi matn bo'lganligi uchun dastlab uni skanerlab oldik. Bunda skanerlash apparati va skanerlash dasturlari ishlatildi. Hosil bo'lgan elektron matn *MS Word* dasturi yordamida *.docx formatga o'girildi, so'ngra asosiy manba bilan solishtirilib, sinchiklab tekshirildi, xatolar to'g'rilandi va saqlandi.

Matnlar bazasini yaratishga doir ishning keyingi bosqichlarini *Dast_MtnBaza* nomli dastur amalga oshiradi. Bu dastur saqlangan fayllar asosida korpusning matnlar bazasini shakllantiradi. Matnlar bazasi kompyuter xotirasidagi bitta *Matn_Baza* nomli papka va bitta *MeteRazm* nomli metarazmetka ma'lumotlari faylidan iborat bo'ladi. *Matn_Baza* papkada korpus tarkibiga kiritishga qaror qilingan matnlarni o'z ichiga oluvchi *.docx fayllar saqlanadi. *MeteRazm* faylda *Matn_Baza* papkaga kiritilgan har bir faylda mavjud matn haqida metarazmetka ma'lumotlari saqlanadi. Metarazmetka ma'lumotlari matn bo'yicha umumiy ma'lumotlardan iborat bo'lib, unda matn nomi, matn muallifiga oid ma'lumotlar saqlangan.

Gram_Lugat nomli dastur yordamida korpus lug'ati shakllantirildi. Lug'at korpusning matnlar bazasiga kiritilgan matnlar asosida shakllanadi. *Gram_Lugat* dastur korpusning *Matn_Baza* matnlar bazasidagi har bir faylni navbat bilan ishlovdan o'tkazadi. Har bir fayl uchun *Gram_Lugat* dastur quyidagi ishlarni amalga oshiradi [Qarshiyev, Tursunov, 2022]:

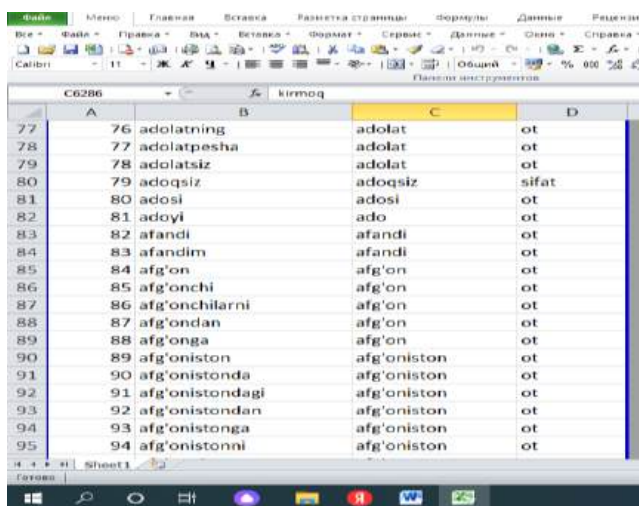
- matnni tokenlash;
- matndagi so'zlarning turli ro'yxatlarini (alfavit-chastotali tartibdagi, chastota-alfavit tartibdagi va ters ro'yxat) tuzish;
- so'zlarni grammatik lug'atga kiritish;

Matnni tokenlashda, matnni avtomatik qayta ishlashda, birinchi navbatda, undagi so'zlarni ajratib olish yoki boshqacha aytganda, matnni birliklarga bo'laklash masalasi yuzaga keladi. Buning uchun ajratuvchi belgilarni (probel, tinish belgilari va h.k.) o'z ichiga olmagan hamma qisman satrlar matndan ajratilib

olinishi lozim. Bu esa tokenlar to'plami bo'ladi. Matnlarni avtomatik qayta ishlashning fundamental algoritmlaridan biri, berilgan matnni tokenlarga bo'lib tashlashdan iborat. Algoritm kirishiga matn berilib, chiqishida matndagi tokenlar ro'yxati olinadi. Bu algoritmnii amalga oshiruvchi dasturni tokenayzer deb atashadi. Odatda, tokenlar so'z shakllari bilan bir xil ma'noni beradi. Lekin, leksik birliklarni ifodalash uchun “so'z” emas, balki “token” termini ishlatiladi. Bunga sabab, ba'zi hollarda token sifatida so'zdan kichikroq birliklar (alohida morfema) yoki so'zdan kattaroq birliklar (so'z birikmalari) ishlatilishi mumkin.

Tokenayzer matnni, dastlab, so'zlar orasidagi probellar (bo'shliq belgilari) asosida bo'laklaydi, so'ngra so'zlardan tinish belgilari olib tashlanadi. Qisqartmalar (masalan, TATU, BMT, MDH, h.k.) va sana yozuvi (masalan, 09.04.2018) ham token sifatida olinadi.

Tokenayzer dramalar matnini tokenlarga bo'laklab berdi va dastur yordamida so'zshakllardan alfavitli-chastotali lug'at hosil qilinib, quyidagi rasmda keltirilganidek morfologik razmetka ishlari amalga oshirildi (2.3-rasm).



	A	B	C	D
77	76	adolatning	adolat	ot
78	77	adolatpesha	adolat	ot
79	78	adolatsiz	adolat	ot
80	79	adoqsiz	adoqsiz	sifat
81	80	adosi	adosi	ot
82	81	adoyi	ado	ot
83	82	afandi	afandi	ot
84	83	afandim	afandi	ot
85	84	afg'on	afg'on	ot
86	85	afg'onchi	afg'on	ot
87	86	afg'onchilarni	afg'on	ot
88	87	afg'ondan	afg'on	ot
89	88	afg'onga	afg'on	ot
90	89	afg'oniston	afg'oniston	ot
91	90	afg'onistonda	afg'oniston	ot
92	91	afg'onistondagi	afg'oniston	ot
93	92	afg'onistondan	afg'oniston	ot
94	93	afg'onistonga	afg'oniston	ot
95	94	afg'onistonni	afg'oniston	ot

2.3-rasm. Morfologik annotatsiyalash.

Bu dasturlar quyidagi sifatlarga ega bo'lishi kerak:

- foydalanuvchi so'zlar yoki gaplarning grammatik yoki sintaksis xarakteristikalarini yetarli darajada to'liq aniqlash va o'zgartirish imkoniyatiga ega bo'lsin;
- tahlil maksimal darajada qulay va sodda bo'lib, tez bajarilsin va bu tahlilning to'liqligini kamaytirmasin.

Korpus yaratishda quyidagi savollarga asoslangan holda asosiy talablarni belgilab olish muhim [Qarshiyev, Tursunov, 2022]:

1. Korpus foydalanuvchisi kim? (Individual, guruh, lingvistik jamiyat).
2. Korpusning mantiqiy g'oyasi nimada?
3. Korpusni qurishda qancha hajmdagi ma'lumotlar bilan ishlamoqchisiz?

4. Nima ishlatiladi: matn parchalari, to'liq matnlar yoki ikkalasi ham?
5. Korpusdagi matnlarni tanlash tartibi qanday?

Usmon Azim dramalari til korpusi lingvistik korpuslar qo'llanilishiga qo'yilgan quyidagi talablar bajaradi:

1. Korpus o'quv va ma'lumot qo'llanmalar yozishda, lug'atlar tuzishda hamda lingvistik ekspertizalar va tahlillarni ishlab chiqishda, til ma'lumotlari sifatida ishonchli manba hisoblanadi;
2. Korpus turli xil lingvistik fanlarni o'qitishda xuddi tilning ko'rgazmali (illyustratsiya) manbasi, hamda tahlillarni taqdimotlilik ko'rinishi sifatida ishlatish mumkin;
3. Korpus ilmiy tadqiqot ishlarida.
4. Korpus til faoliyatining ma'lumotlar qo'llanmasi sifatida bo'lishi mumkin.

Xulosa qilib aytganda, razmetkaning yuqori aniqlikda bo'lishini faqat filolog mutaxassis ta'minlashi mumkin. Filolog mutaxassis har bir matnning razmetkalanishini so'zma-so'z sinchiklab tekshirishi, avtomatik razmetka xatolarini tuzatishi va omonim holatlarni bartaraf qilishi kerak. Ishning asosiy qismi qo'l mehnati orqali bajarilishi mumkin. Shuning uchun filolog mutaxassisning kompyuterdagi ishiga mo'ljallangan maxsus razmetkalash dasturlari yaratilishi zarur.

Foydalanilgan adabiyotlar:

1. Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари. filol.fan.bo'y. PhD diss. –Qarshi, 2018.
2. Usmon Azim. Dramatik asarlar. - Toshkent, 2015.
3. A.Qarshiyev, M.Tursunov. O'zbek tili milliy korpusini loyihalash. "O'zbek milliy va ta'limiy korpuslarini yaratishning nazariy hamda amaliy masalalari", 2022.
4. A.Qarshiyev, M.Tursunov. O'zbek tili milliy korpusining dasturiy ta'minot strukturasi va vazifalari, 2022.

CORPUS PROCESSING TOOLS IN BUILDING MULTILINGUAL CORPORA

Amirqulov Ma'rufjon Aliqul o'g'li

amirkulov.edu01@gmail.com

ToshDO'TAU magistranti

Abstract. Corpora annotation constitutes a pivotal phase in the development of multilingual corpora, encompassing various processes like lemmatization, tagging, parsing, and more. These procedures involve the addition of meta-information into the corpus, enabling the organization of textual elements such as words into specific categories like part-of-speech, thereby facilitating comprehensive analysis of their relationships and structures. Without such meticulous sorting and understanding of interrelations, the corpus data cannot be fully leveraged to its potential. Thus, this article discusses aforementioned matters briefly.

Annotatsiya. Korpus annotatsiyasi lemmatizatsiya, teglash, tahlil qilish va boshqalar kabi turli jarayonlarni o'z ichiga olgan ko'p tilli korpusni rivojlantirishning asosiy bosqichini tashkil qiladi. Ushbu protseduralar korpusga metama'lumotni qo'shishni o'z ichiga oladi, bu so'zlar kabi matn elementlarini nutqning bir qismi kabi muayyan toifalarga ajratish imkonini beradi va shu bilan ularning munosabatlari va tuzilmalarini har tomonlama tahlil qilishni osonlashtiradi. Bunday sinchkovlik bilan tartiblash va o'zaro munosabatlarni to'liq tushunib yetmaslik korpus ma'lumotlarining to'liq imkoniyatlaridan foydalanishga to'siq bo'ladi. Ushbu maqola yuqorida ta'kidlab o'tilgan masalalarni qisqacha muhokama qiladi.

Аннотация. Аннотации Корпуса представляют собой ключевой этап в разработке многоязычных корпусов, охватывающий различные процессы, такие как лемматизация, тегирование, синтаксический анализ и многое другое. Эти процедуры включают добавление метаинформации в корпус, что позволяет организовать текстовые элементы, такие как слова, в определенные категории, такие как части речи, тем самым облегчая всесторонний анализ их отношений и структур. Без такой тщательной сортировки и понимания взаимосвязей корпусные данные не могут быть полностью использованы в полной мере. Таким образом, в данной статье кратко обсуждаются вышеупомянутые вопросы.

Keywords: *corpora, concordance, aligner, metadata, lemmatizator.*

While manual annotation of metadata is theoretically accurate, it is exceedingly laborious and time-intensive. Hence, there is a prevalent preference for employing software tools to automate these tasks. Despite the potential for lower accuracy levels compared to manual methods, the practicality and efficiency offered

by automated annotation software make it the predominant choice in contemporary corpus development practices. This shift is primarily driven by the sheer volume of textual information present in corpora, especially those of considerable size, rendering manual processing unfeasible. Therefore, specialized software tools play a critical role in expediting the annotation process, thereby saving both time and resources.

In scenarios where mid-sized corpora or high precision outputs are required, semi-automatic annotation tools serve as a viable alternative. These tools automate a portion of the annotation process, which is then refined and enhanced through manual editing and further software training. Conversely, for smaller and specialized corpora with limited sample sizes, manual metadata addition proves advantageous, allowing for meticulous attention to detail and precision.

The thesis at hand exclusively focuses on automatic annotation tools, showcasing a variety of software options available for corpus processing. Examples include Wordfisher, Corpus Builder, BNC, Gestrolex, ANTConc, Sketch Engine, and Shoebox, each offering unique functionalities tailored to specific research needs. While some tools provide specialized features, others are comprehensive and intricate, catering to diverse requirements within the field of corpus linguistics.

Historically, early iterations of corpus processing software were notorious for their lack of user-friendliness and steep learning curves, often necessitating substantial computer literacy. However, modern software trends prioritize accessibility and ease of use, making these tools increasingly approachable for both linguists and non-experts alike. As a result, the current landscape emphasizes user-friendly interfaces and intuitive design, ensuring broader accessibility and usability across the corpus linguistics community.

The subsequent section of the thesis delves into fundamental tools designed to enhance the overall reliability of corpus data. These tools are geared towards differentiating between grammatical or syntactic categories, resolving language ambiguities, and refining structural analysis, rather than directly searching the corpus or providing finite data outputs.

LEMMAZIZER

A lemmatizer plays a crucial role in linguistic analysis by assigning a lemma—a collection of lexical forms with a shared stem and belonging to the same major word class—to each word entry in a corpus [Francis & Kučera, 1982:1]. This process simplifies the corpus by reducing words to their respective lexemes, enabling researchers to explore all variations of a specific lexeme without manually inputting every potential variant. Additionally, it facilitates the extraction of frequency and distribution data for each lexeme, as highlighted by Baker, Hardie, and McEnery [Baker, Hardie & McEnery, 2006:104-105].

Given the unique characteristics of each language, lemmatization tools must be developed separately for each linguistic system. However, linguists may

sometimes prioritize broader language features, such as grammatical categories and syntactic relationships, over specific lexical compositions. In such cases, it becomes essential to distinguish these features within the corpus texts. This is achieved through the use of linguistic markers known as tags. These tags help delineate the grammatical and syntactic structures present in the texts, allowing researchers to analyze language features beyond individual word forms.

TAGGER

Tagging, as described by Baker, Hardie, and McEnery [Baker, Hardie & McEnery, 2006:154] involves applying additional layers of annotation to corpus data, specifically adding codes to each word to indicate its part of speech [Hounston, 2002]. Initially, during the early stages of corpus development, tagging was a manual process. However, advancements in technology have led to the creation of software tools for automatic tagging.

Despite these advancements, fully automated tagging remains elusive. Instead, a manual tagging process is typically employed initially, using a sufficiently large text as a "tutorial" for the tagging software. Through a process of learning based on verisimilitude or predetermined linguistic rules, the tagger then assigns grammatical category markers (tags) to selected texts. According to Baker, Hardie, and McEnery [2006:154], a tag can be attached to various linguistic units, including phonemes, morphemes, words, phrases, or longer stretches of text.

Although human correction and post-editing are still necessary, the accuracy of automated tagging has surpassed 90%. This high level of accuracy allows tagging software tools to be utilized in much larger projects, streamlining the tagging process and enhancing efficiency in corpus annotation endeavors.

PARSER

Parsing is another essential procedure in corpus annotation, as it involves marking texts not only grammatically but also syntactically. Unlike tagging, which focuses on grammatical categories, parsing delves deeper into the syntactic structure of texts.

To enable parsing, a small number of manually processed texts are typically required to instruct parsers on recognizing and marking syntactical categories within a selected text in a corpus. Parsing tags may include delineating the start and end points of units such as noun phrases, verb phrases, and clauses. Additionally, parsing may provide information about how these syntactic units relate to one another [Baker, Hardie & McEnery, 2006:127]. However, parsing is less commonly used compared to tagging, primarily due to its complexity. Parsing a corpus is far more intricate and resource-intensive than tagging it. Moreover, there are limited parsing tools available to individual researchers or students, further contributing to its lower adoption rate in corpus annotation endeavors.

ALIGNER

In parallel corpus projects, a vital tool is employed to facilitate the alignment of text segments in one language with their counterparts in another language. This tool serves as a cornerstone for linguists, enabling them to synchronize specific parts of a text across different languages seamlessly. Often, this alignment process relies on the principle of verisimilitude, ensuring accurate correspondence between the texts. Once texts undergo modification through this software, researchers can proceed to delve into the corpus for their analyses. However, unlike the annotation tools previously discussed, the software tools described below offer tangible data outputs. These outputs provide researchers with concrete information that can be utilized in various ways. They can be treated as "finite," allowing for the formulation of generalizations and insights based on the extracted data. Alternatively, researchers may subject the data to further scrutiny and analysis, exploring nuances and patterns hidden within the corpus. Thus, these software tools serve as invaluable assets in the extraction and interpretation of linguistic data from parallel corpora.

WORD LISTER

One of the fundamental tools utilized in corpus processing is the word lister, which serves as a cornerstone for basic statistical analyses within a corpus. Often referred to simply as a word lister, this tool enables researchers to perform straightforward calculations and analyses on textual data. As described by Lynne Bowker, a word lister can calculate the total number of "tokens" and tally the frequency of each individual word form within the corpus [Bowker & Pearson, 2002:13]. These word lists can be sorted in various ways, such as alphabetically or by frequency, facilitating easy access to information. Furthermore, certain word listers possess the capability to generate lists of keywords, offering additional utility beyond basic word frequency analysis. These keyword lists can serve as valuable resources for activities like terminology mining and further web searches. For example, word lists derived from specialized corpora can be compared with keyword lists from other corpora, allowing users to discern differences and similarities in lexical usage across different fields of study or domains. This comparative analysis provides valuable insights into the language specific to various contexts, enriching linguistic research and facilitating cross-disciplinary investigations.

CONCORDANCER

A concordancer is a fundamental tool shared by all corpus-processing software, providing users with invaluable insights into the occurrences of specific search terms within a corpus. As defined by Baker, Hardie, and McEnery [Baker, Hardie & McEnery, 2006:42-43], a concordance is essentially a comprehensive list of all instances where a particular search term appears in the corpus, presented alongside the surrounding context in which each occurrence occurs. According to Lynne Bowker and Jennifer Pearson, a concordancer offers users the ability to

observe every instance of a particular word within its immediate contexts. This information is typically presented using a format known as keyword in context (KWIC). In a KWIC display, all occurrences of the search term are aligned at the center of the screen, with a specified amount of context displayed on either side. Similar to word lists, concordances can be sorted to facilitate the identification of patterns, enabling users to glean deeper insights into the usage and distribution of specific terms within the corpus.

CONCLUSION

The realm of corpus processing tools is vast, with a plethora of options available to researchers. In fact, the sheer variety of tools is such that an entire thesis could be dedicated solely to their description. Many software packages encompass a range of tools, some of which may be exclusive to a particular software (for instance, Word Sketches are unique to the Sketch Engine). These tools offer users a diverse array of functionalities, including the ability to access original files, search for collocations, save search results, visualize the clustering of search terms (such as in the AntConc Concordance Plot), utilize the corpus as a thesaurus, and consult additional corpora (as seen in the Sketch Engine). Despite the extensive range of tools available, certain basic tools are indispensable for anyone involved in corpus compilation. These tools, discussed earlier, form the foundation of corpus processing and are typically included in most comprehensive software packages. They serve as essential components of the corpus compilation process, facilitating various analyses and inquiries into linguistic data.

References:

1. Baker, P., Hardie, A., & McEnery, T. (2006). *A Glossary of Corpus Linguistics*. Edinburgh: Edinburgh University Press. (Baker, Hardie & McEnery, 2006)
 2. Bowker, L., & Pearson, L. (2002). *Working with Specialized Language: A Practical Guide to Using Corpora*. New York: Routledge. (Bowker & Pearson, 2002)
 3. Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press. (Hunston, 2002)
- Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin. (Francis & Kučera, 1982)

LINGVISTIK ANNOTATSIYA TURLARI VA UNING TIL KORPUSLARIDAGI AHAMIYATI

Xudayarova Sabura Shuxrat qizi

xudayorovasabura@navoiy-uni.uz

ToshDO‘TAU magistranti

Annotatsiya. Til va kompyuter, ushbu tushunchalar bugungi kunda tilshunoslikning zamonaviy yo‘nalishlarini shakllantirishi bilan birga sohada ko‘plab amaliy natijalarga erishmoqda. Ko‘plab til korpuslarining yaratilishi hamda til birliklarining korpuslarda annotatsiyalanishi erishilgan yutuqlar natijasidir. Ushbu maqolada lingvistik annotatsiya va uning turlari hamda bugungi kunga qadar yaratilgan til korpuslari, turli til korpuslarida til birliklarining annotatsiyalanishi haqida fikr yuritilgan.

Abstract. Language and computers, these concepts shape the modern directions of linguistics today and achieve many practical results in the field. The creation of many language corpora and the annotation of language units in corpora are the result of the achievements. This article discusses linguistic annotation and its types, language corpora created to date, annotation of language units in different language corpora.

Аннотация. Язык и компьютеры — эти концепции сегодня формируют современные направления лингвистики и достигают многих практических результатов в этой области. Результатом достижений является создание множества языковых корпусов и аннотирование языковых единиц в корпусах. В данной статье рассматриваются лингвистическая аннотация и ее виды, созданные к настоящему времени языковые корпуса, аннотация языковых единиц в разных языковых корпусах.

Kalit so‘zlar. *Obyekt, modellashtirish, sun’iy intellekt, xalqaro korpuslar, NLP, algoritm, matn, token, lemma, annotatsiyalash.*

Shaxsning hujjat va matn ma’lumotlarini ifodalovchi belgilar bilan o‘zaro aloqasi Tabiiy tilni qayta ishlash (NLP) sun’iy intellektidagi eng yirik ishlanmalardan biridir. Chatbotlar, nutqni avtomatik aniqlash va his-tuyg‘ularni tahlil qilish dasturlari kabi ko‘plab NLP yechimlari butun dunyo bo‘ylab son-sanoqsiz biznes samaradorligini va mahsuldorligini oshirmoqda. NLP sohasidagi so‘nggi yutuqlar, hatto nutqida nuqsoni bo‘lgan odamlarga avtomatik nutqni aniqlash qurilmalari va ularning atrofidagi odamlar bilan erkin muloqot qilish imkoniyatini ko‘rsatdi. Biroq, bu ajoyib texnologiyalarning hech biri matn annotatsiyasiz amalga oshirilmaydi.

NLP algoritmlari uchun izohli matn ma’lumotlarining katta to‘plami talab qilinadi. Matn izohlarining beshta keng tarqalgan turiga qisqacha to‘xtalamiz:

1. Entity Annotation Entity annotatsiyasi chatbot trening ma'lumotlar to'plami va boshqa NLP o'quv ma'lumotlarini yaratishda eng muhim jarayonlardan biridir. Bu matndagi obyektlarni aniqlash, ajratib olish va etiketlash harakati.

Obyekt izohi turlariga quyidagilar kiradi:

- Nomlangan obyektни tanib olish (NER): obyektlarning tegishli nomlari bilan izohlash.

- Kalit iboralarni belgilash: matn ma'lumotlaridagi kalit so'zlar yoki kalit iboralarning joylashuvi va teglari.

- Nutq qismini (POS) teglash: tanib olish va izohlash [Core, Ishizaki, Moore, Nakatani, Reithinger, Traum, Tutiya, 1998: 65]. Nutqning funksional elementlari (sifatlar, otlar, qo'shimchalar, fe'llar va boshqalar). Shaxsiy izohlar NLP modelini matndagi nutq qismlarini, nomlangan shaxslarni va asosiy iboralarni aniqlashni o'rgatadi. Ushbu vazifada izohlovchilar matnni diqqat bilan o'qiydilar, maqsadli obyektlarni topadilar, ularni annotatsiya platformasida ajratib ko'rsatishadi va oldindan belgilangan ro'yxatdan teglarni tanlaydilar. NLP modellariga nom berilgan obyektlar haqida ko'proq ma'lumot olishga yordam berish uchun obyekt izohi ko'pincha obyektни bog'lash bilan birlashtiriladi.

2. Obyektни bog'lash. Obyekt izohi matndagi muayyan obyektlarni tartibga solish va izohlash bo'lsa, obyektни bog'lash bu obyektlarni ular haqidagi ma'lumotlarning kattaroq omborlariga ulash jarayonidir.

3. **Matnni tasniflash**, shuningdek, matnni turkumlashtirish yoki hujjat tasnifi sifatida ham tanilgan, matnni tasniflash matnni yoki matnning qisqa satrlarini o'qish bilan izohlovchi vazifalarni bajaradi. Annotatorlar tarkibni tahlil qilishlari, undagi mavzu, niyat va his-tuyg'ularni aniqlashlari va oldindan belgilangan toifalar ro'yxati asosida tasniflashlari kerak. Obyekt annotatsiyasi alohida so'zlar yoki iboralarni belgilash bo'lsa, matn tasnifi butun matn yoki matn qatoriga bitta yorliq bilan izoh berish jarayonidir.

4. **Hissiy intellekt** - bu mashinani o'rganishning eng qiyin sohalaridan biri. Ba'zan hatto odamlar uchun matnli xabar yoki elektron pochta ortidagi haqiqiy tuyg'uni taxmin qilish qiyin. Mashina uchun istehzoli, aqlli yoki boshqa tasodifiy aloqa shakllaridan foydalanadigan matnlarda yashirin konnotatsiyalarni aniqlash eksponent jihatdan qiyinroq. Mashinani o'rganish modellariga matn ichidagi his-tuyg'ularni tushunishga yordam berish uchun modellar his-tuyg'ularga izohli matn ma'lumotlari bilan o'rgatiladi. Kengroq ma'noda his-tuyg'ularni tahlil qilish yoki fikrni tahlil qilish deb ataladi, hissiyot izohi - bu tuyg'u, fikr yoki his-tuyg'ularning yorlig'i. Annotatorlarga tahlil qilish uchun matnlar beriladi va matn ichidagi his-tuyg'u yoki fikrni qaysi belgi eng yaxshi ifodalashini tanlashi kerak.

5. **Lingvistik annotatsiya**, shuningdek, korpus annotatsiyasi deb ataladi, lingvistik annotatsiya shunchaki matn yoki audio yozuvlardagi til ma'lumotlarini belgilash jarayonini tasvirlaydi. Lingvistik annotatsiya bilan annotatorlarga matn yoki audio ma'lumotlardagi grammatik, semantik yoki fonetik elementlarni aniqlash

va belgilash vazifasi yuklanadi. Lingvistik annotatsiyaning turlariga quyidagilar kiradi [Bunt H, 2010. 29–46]:

- **Diskurs annotatsiyasi:** anafor va kataforalarning o'z oldingi yoki postsedentlari bilan bog'lanishi. Nutqning bir qismi (POS) teglari: matndagi turli funksiyali so'zlarga izoh beradi.

- **Fonetik annotatsiya:** nutqdagi intonatsiya, urg'u va tabiiy pauzalarni belgilaydi.

- **Semantik annotatsiya:** ta'riflarga izoh beradi. Lingvistik annotatsiya chatbotlar, virtual yordamchilar, qidiruv tizimlari, mashina tarjimasini va boshqalar kabi turli NLP yechimlari uchun sun'iy intellekt bo'yicha o'quv ma'lumotlar to'plamini yaratishda qo'llaniladi. Bular bugungi kunda mashinani o'rganishda keng qo'llaniladigan beshta turdagi matn izohlaridir. Korpusdagi til namunasining xususiyatlari. Bunday izoh tokenizatsiya deb ataladigan dastlabki segmentatsiya jarayonini talab qiladi, bu esa korpusdagi birliklarni belgilaydi – so'zlar, raqamlar, tinish belgilari va boshqalar. Ba'zi hollarda bu qo'shimcha qadamni o'z ichiga oladi.

Obyektni tan olish, bu korpusdagi tegishli birliklarni aniqlashga xizmat qiladi. Lemmalar annotatsiyasi lemmatizatsiyaning eng asosiy turlaridan biri, korpusdagi har bir so'zni asos (iqtibos yoki lug'at) shakli bilan belgilash. Lemmatizatsiya mavjud shakl-lemma ma'lumotlar bazasi asosida amalga oshirilishi mumkin, (yarim) avtomatik yondashuv bo'lib, unda so'z shakllari kesiladi. Lemmaning umumiy ifodasiga kelish uchun belgilarni kesib tashlash yoki ushbu ikki strategiyaning ba'zi gibridd yondashuvlari, ular morfologik birlikni ham o'z ichiga olishi mumkin. Sintaktik va morfologik jihatdan Til birliklarini teglash ham shular jumlasidandir. Til birliklarini teglash annotatsiyaning eng tez-tez uchraydigan va eng ko'p ishlatiladigan turlaridan biridir. Annotatsiya ko'plab korpus-lingvistik tadqiqotlarga tegishli semantik annotatsiya. Bu har bir tokenlashtirilgan so'zga yorliq belgilashni o'z ichiga oladi. Ba'zi grammatik kategoriya ma'lumotlari so'zning nutq qismini minimal darajada aniqlaydi,

Annotatsiyaning keng tarqalgan turi - nutqning bir qismini teglash korpusda juda ko'p turli xil annotatsiya turlari va formatlari qo'llanilishi ajablanarli emas. Tilshunoslikda korpus lingvistikasining ko'p qismida hali ham nisbatan kam sonli dastur turlari hukmronlik qilmoqda.

Lingvistik annotatsiya til korpuslarining ajralmas qisimidir. Korpusdagi har bir til birliklari lingvistik annotatsiyalanganagina til korpuslarini tashkil etadi. Lingvistik annotatsiyaning ahamiyati til korpuslarida namoyan bo'ladi. Shunday ekan til korpuslari va uning turlari haqida fikr yuritish o'rinli bo'ladi. Korpus lingvistikasi amaliy tilshunoslikning lingvistik korpus (matnlar korpusi) qurish va ulardan foydalanishning umumiy tamoyillarini ishlab chiqish bilan shug'ullanuvchi bo'limidir. Turli xil matnlarni tahlil qilish asosida tadqiqotchini qiziqtiradigan lingvistik hodisa, masalan, grammatik tuzilmaning xatti-harakati, tilda ekspressiv vositalardan foydalanish va boshqalar haqida xulosa chiqarish mumkin. Kompyuter texnologiyalarining rivojlanishi ko'plab matnlarning elektron shaklda paydo

bo'lishiga yordam berdi. Bunday hajmdagi matnlar bilan ishlash, ulardan kerakli ma'lumotlarni olish uchun butun dunyoda lingvistik korpuslar yaratila boshlandi, ya'ni, maxsus tanlangan, turli lingvistik parametrlar bo'yicha belgilangan va qidiruv tizimi bilan ta'minlangan matnlar to'plami. Tadqiqot materiali hajmining ortishi lingvistik ma'lumotlarni tahlil qilishning yangi usullarini, jumladan, ularni statistik qayta ishlashni qo'llashni taqozo etdi. Demak, korpus lingvistikasi ikki jihatni o'z ichiga oladi: birinchidan, matn korpusini yaratish va belgilash (annotatsiya) va qidiruv vositalarini ishlab chiqish. ular uchun, ikkinchidan, lingvistikaning o'zi - korpuslarga asoslangan eksperimental tadqiqotlar. Bu nisbatan yosh va faol rivojlanayotgan, hisoblash tilshunosligi bilan chambarchas bog'liq va miqdoriy usullardan keng foydalaniladi. Elektron resurslardan foydalanish imkoniyati tilshunoslik tadqiqotlarida material to'plash jarayonini ancha osonlashtirdi. Biroq, lingvistik ma'lumotlarning bunday mavjudligi lingvistik tadqiqotlarning dalillar bazasiga qo'yiladigan talablarni tubdan o'zgartirdi. Korpus lingvistikasi tilshunoslikning matn korpusini ishlab chiqish, yaratish va ishlatish bilan shug'ullanuvchi bo'limidir. Bu atama 1960-yillarda korpus yaratish amaliyotining rivojlanishi bilan bog'liq bo'lib, 1980-yillardan boshlab kompyuter texnologiyalarining rivojlanishi bilan bog'liq holda qo'llanila boshlandi. Korpus lingvistikasi tilni o'rganadi, chunki bu til uning matn korpusida (ko'plikda), uning “haqiqiy dunyo” matn korpusida ifodalanadi. Korpus tilshunosligi shuni ko'rsatadiki, tilni ishonchli tahlil qilish eng kam tajriba aralashuvi bilan ushbu tilning tabiiy konteksti sohasida to'plangan korpuslardan foydalanish mumkin. Katta matn to'plamlari tilshunoslarga lingvistik tushunchalarning miqdoriy tahlilini o'tkazishga imkon beradi. Lingvistik yoki til, matnlar korpusi - bu katta, taqdim etilgan. Mashinada o'qilishi mumkin bo'lgan formatda, birlashtirilgan, tuzilgan, belgilangan, aniq lingvistik muammolarni hal qilish uchun mo'ljallangan filologik jihatdan malakali til ma'lumotlari to'plami. Zamonaviy korpusning asosiy xususiyatlari - mashinada o'qiladigan format, reprezentativlik va metalingvistik ma'lumotlarning mavjudligi. Reprezentativlikka matnlarni tanlashning maxsus tartibi yordamida erishiladi. Lingvistik korpus – ma'lum tamoyillarga muvofiq to'plangan, ma'lum standart bo'yicha belgilangan va ixtisoslashtirilgan qidiruv tizimi bilan ta'minlangan. Ba'zan korpus (“birinchi tartibli korpus”) oddiygina umumiy xususiyat (til, janr, muallif, matnlarni yaratish davri) bilan birlashtirilgan har qanday matnlar to'plami deb ataladi. Matn korpusini yaratishning maqsadga muvofiqligi quyidagilar bilan izohlanadi: taqdimot, real kontekstdagi lingvistik ma'lumotlar; ma'lumotlarning yetarlicha katta vakili (korpusning katta hajmi bilan); turli lingvistik muammolarni hal qilish uchun bir marta yaratilgan korpusdan qayta foydalanish imkoniyati, masalan, grafik va leksikani amalga oshirish.

Matnning grammatik tahlili. Birinchi yirik kompyuter korpusi 1960-yillarda Braun universitetida yaratilgan va har biri 2 ming so'zdan iborat 500 ta matn bo'laklarini o'z ichiga olgan, 1961-yilda AQShda ingliz tilida nashr etilgan Brown



Corpus (BC) hisoblanadi. Natijada, u boshqa tillarda vakillik korpusini yaratish uchun 1 million hodisa standartini o'rnatdi.

1970-yillarda Zazorina rus tilining chastotali lug'ati yaratilgan bo'lib, u matnlar korpusi asosida, shuningdek, 1 million so'z hajmida va taxminan teng nisbatda ijtimoiy-siyosiy matnlarni, badiiy adabiyotlarni o'z ichiga olgan, turli sohalaridagi ilmiy va ilmiy-ommabop matnlar va drama. 1980-yillarda Shvetsiyaning Upsala universitetida yaratilgan rus Korpusi ham xuddi shunday model bo'yicha qurilgan. Bir million so'z hajmi faqat eng tez-tez uchraydigan so'zlarning leksik-orfografik tavsifi uchun yetarli, chunki o'rtacha chastotali so'zlar va grammatik tuzilmalar har million so'zga bir necha marta uchraydi (statistik nuqtayi nazardan, til noyob hodisalarning katta to'plamidir, shuningdek, katta hajmdagi matnlar bilan ishlashga qodir kompyuter quvvatining o'sishi tufayli 1980-yillarda butun dunyo bo'ylab kattaroq korpuslarni yaratishga bir necha bor urinishlar qilingan. Buyuk Britaniyada bunday loyihalar Birmingem universitetidagi ingliz banki va Britaniya milliy korpusi (BNC) edi. Minglab matnlarni to'plash, mualliflik huquqi muammolarini bartaraf etish, barcha matnlarni yagona shaklga keltirish, korpusni mavzular va janrlar bo'yicha muvozanatlash juda ko'p vaqtni oladi.

Nemis, polyak, chex, sloven, fin, zamonaviy yunon, arman, xitoy, yapon, bolgar va boshqa tillar uchun vakillik korpusi mavjud (yoki ishlab chiqilmoqda). Rossiya Fanlar akademiyasida yaratilgan rus tilining milliy korpusi; Hozirda 500 milliondan ortiq so'z qo'llanishlarini o'z ichiga oladi [Core, Ishizaki, Moore, Nakatani, Reithinger, Traum, Tutiya, 1998:79]. Keng janrlar va funksional uslublarni qamrab oluvchi vakillik korpusi bilan bir qatorda lingvistik tadqiqotlar ko'pincha opportunistik matnlar to'plamidan, masalan, gazetalardan (ko'pincha The Wall Street Journal va The New) foydalanadi. Ishning hajmi ham, tuzilishi ham reprezentativlik uchun muhimdir. Vakillik hajmi muammoga bog'liq, chunki u o'rganilayotgan hodisalar uchun qancha misol topish mumkinligi bilan belgilanadi.

Matn korpusi usuli har qanday tabiiy tilda yozilgan matnlardan ushbu tilni boshqaradigan mavhum qoidalar to'plamini olish uchun foydalanadi. Ushbu natijalar ushbu maqsadli til va shunga o'xshash tahlildan o'tgan boshqa tillar o'rtasidagi munosabatni o'rganish uchun ishlatilishi mumkin. Birinchi bunday korpuslar manba matnlari asosida qo'lda yaratilgan, ammo hozir bu ish avtomatlashtirilgan. Korpus nafaqat lingvistik tadqiqotlar uchun, balki lug'atlarni [Carletta, 1996: 249–254] ham qayta tahrir qilish mumkin.

Tilshunoslikda va tabiiy tilni qayta ishlashda korpus yoki matn korpusi raqamli va raqamlashtirilgan, izohli yoki izohsiz til resurslaridan iborat ma'lumotlar to'plamidir. Ular korpus lingvistikasida statistik gipotezalarni tekshirish, hodisalarni tekshirish yoki ma'lum bir til hududida lingvistik qoidalarni tasdiqlash uchun ishlatilgan. Qidiruv texnologiyasida korpus bu qidirilayotgan hujjatlar to'plamidir. Odatda, u manba nomi va u olingan butun matnga giperhavola bilan ta'minlanadi. Korpus menejeri korpus bo'yicha har xil turdagi statistik ma'lumotlar haqida hisobot berishi mumkin. Masalan, ma'lum bir matn birligidan

foydalanish chastotasi lug‘atini tuzing. Bu ma’lumotlarning barchasi tilshunos tomonidan tilni bir butun sifatida tavsiflashda yoki alohida hodisani o‘rganishda qo‘llaniladi.

Bizning fikrimizcha, korpusning vakili bo‘lishi zaruriy shartdir, bu sizga turli xil matnlar to‘plamiga korpus maqomini berishga imkon beradi. Lingvistik tahlil qilish imkonini beruvchi matnlar ma’lum bir korpusning so‘zsiz vakilidir.

Xulosa qilib aytganda, ko‘plab til korpuslari yaratilgan bo‘lib, har bir tilning ichki imkoniyatlari o‘ziga xos xususiyatlarini aniq belgilab olgan holda uning muayyan bir shaklini ishlab chiqish bilan bir qatorda har bir til birligi lingvistik annotatsiyalanishi til korpusida amalga oshiriladi.

Foydalanilgan adabiyotlar:

1. Artstein, R., Poesio, M.: Inter-coder agreement for computational linguistics. Computational Linguistics, - 2008. – c 555–596
2. Bernsen, N.O., Dybkjær, L., Kolodnytsky, M.: The NITE Workbench. A Tool for Annotation of Natural Interactivity and Multimodal Data. In: Proceedings of the Third International Conference on Language Resources and Evaluation, - 2002. – C 214
3. Bird, S., Day, D., Garofolo, J., Henderson, J., Laprun, C., Liberman, M.: Atlas: A flexible and extensible architecture for linguistic annotation. In: Proceedings of the Second International Conference. – 2000
4. Bunt, H.: A methodology for designing semantic annotation languages exploiting semanticsyntactic isomorphisms. In: Proceedings of the Second International Conference on Global Interoperability for Language Resources, – 2010. – C 29-46.
5. Carletta, J.: Assessing agreement on classification tasks: The kappa statistic. Computational Linguistics, - 1996, - pp 249–254
6. Church, K.W.: A stochastic parts programm and noun phrase parser for unrestricted text. In: Proceedings of the Second Conference on Applied Natural Language Processing, ANLC '88, pp.
7. Clear, J.H.: The british national corpus. In: G.P. Landow, P. Delany (eds.) The Digital Word, - 1993. pp. 163–187.
8. Core, M., Ishizaki, M., Moore, J., Nakatani, C., Reithinger, N., Traum, D., Tutiya, S.: The report of the third workshop of the discourse resource initiative. Tech. rep., Chiba University and Kazusa Academia Hall, - 1998.

PARALLEL KORPUS TUZISHDA STEMMING VA LEMMATIZATSIYANING AHAMIYATI

Xolmonova Iqbola Alisher qizi

iqbolabintualisher@gmail.com

ToshDO‘TAU 2-kurs magistranti

Annotatsiya. Ushbu maqolada korpus tuzish uchun stemming va lemmatizatsiyaning ahamiyati yoritilgan. Stemming va lemmatizatsiyaning bir biridan farqi ochib berilgan va misollar bilan isbotlangan. O‘zbek tilidagi so‘zlar lemmasini aniqlash uchun lemmatizatsiya algoritmi izohlangan. Shu bilan birga, mazkur maqolada lemmatizatsiyaning o‘zbek-turk parallel korpusini yaratishdagi ahamiyati ham ochib berilgan. O‘zbek va turk tilidagi lemmani aniqlash uchun modellar taklif qilingan.

Abstract. This article highlights the importance of stemming and lemmatization for corpus building. The difference between stemming and lemmatization is explained and illustrated with examples. The lemmatization algorithm for determining the lemma of words in the Uzbek language is explained. At the same time, this article also reveals the importance of lemmatization in creating the Uzbek-Turkish parallel corpus. Models for determining the lemma in Uzbek and Turkish languages are proposed.

Аннотация. В этой статье подчеркивается важность стемминга и лемматизации для построения корпуса. Объясняется и иллюстрируется примерами разница между стеммингом и лемматизацией. Объясняется алгоритм лемматизации для определения леммы слов узбекского языка. В то же время в данной статье также раскрывается значение лемматизации при создании узбекско-турецкого параллельного корпуса. Предложены модели определения леммы на узбекском и турецком языках.

Kalit so‘zlar: *lemma, lemmatizatsiya, stem, stemming, model, UzLemmatizator, token, tokenizatsiya*

Kirish

Lemmatizatsiya so‘zning lemma (leksema) shaklini aniqlash va uni flektiv/hosila shakllari o‘rnida ishlatishni o‘z ichiga oladi. Lemmatizatsiya jarayoni orqali til korpusi matnlari mazmunini tahlil qilish aniqligini oshirish mumkin. Shuningdek, lemmatizatsiya jarayoni orqali qidiruv tizimlari algoritmlari korpus matnlari mazmunini tushunishi, uni tartiblash jarayonini amalga oshirishga yordam beradi [Jabeen, 2014: 264].

Lemmatizatsiya odatda quyidagicha ishlaydi:

➤ Tokenizatsiya: Birinchi qadam matnni alohida so‘zlarga yoki tokenlarga bo‘lishdir. Buni turli usullar yordamida amalga oshirish mumkin, masalan, matnni bo‘shliqlar asosida ajratish.

➤ POS yorlig'i: Nutq qismlarini teglash har bir tokenga grammatik toifani (masalan, ot, fe'l, sifat va boshqalar) belgilashni o'z ichiga oladi. Lemmatizatsiya ko'pincha bu ma'lumotlarga tayanadi, chunki so'zning asosiy shakli uning gapdagi grammatik rolga bog'liq bo'lishi mumkin.

➤ Lemmatizatsiya: Har bir so'z tokenizatsiya qilingan va nutq qismi tegi tayinlangandan so'ng, lemmatizatsiya algoritmi har bir so'zning lemmasini aniqlash uchun leksika yoki til qoidalaridan foydalanadi. Lemma so'zning asosiy shakli bo'lib, u so'zning ildizi bilan bir xil bo'lishi shart emas.

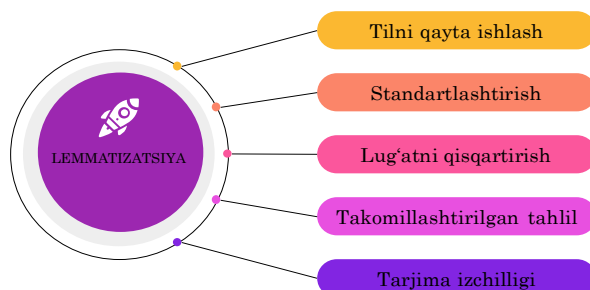
➤ Qoidalarni qo'llash: Lemmatizatsiya algoritmlari ko'pincha lingvistik qoidalarga tayanadi.

➤ Natija: Lemmatizatsiya natijasi so'zlarning asosiy yoki lug'at shaklidagi to'plami bo'lib, matnning asosiy ma'nosini tahlil qilish va tushunishni osonlashtiradi [Saumyab, 2024: 14].

Korpus yaratish uchun lemmatizatsiya jarayoni muhim hisoblanadi. Chunki lemmatizatsiya so'zlarni asosiy shaklga standartlashtirishga yordam beradi, bu so'zlarning turli xil flektiv shakllarini birlashtirib, korpusning murakkabligini kamaytiradi. Ushbu standartlashtirish matnlarni tahlil qilish va solishtirishni osonlashtiradi. So'zlarni asosiy shakliga qisqartirish orqali lemmatizatsiya hissiyotlarni tahlil qilish, mavzuni modellashtirish va ma'lumotlarni qidirish kabi matn tahlili vazifalarining aniqligini oshirishga yordam beradi. Bu bir xil so'zning o'zgarishlari yagona obyekt sifatida ko'rib chiqilishini ta'minlaydi. Lemmatizatsiya so'zlarning flektiv shakllarini umumiy tayanch shaklga jamlash orqali umumiy lug'at hajmini kamaytiradi. Tabiiy tilni qayta ishlash (NLP) ilovalarida lemmatizatsiya ko'pincha matn ma'lumotlarini keyingi qayta ishlashga tayyorlash uchun ishlatiladi, masalan, xususiyatlarni ajratib olish, mashinani o'rganish va tilni modellashtirish. Umuman olganda, samarali matn tahlili va NLP ilovalari uchun yanada standartlashtirilgan, izchil va boshqariladigan korpus yaratish uchun lemmatizatsiya zarur.

Lemmatizatsiya jarayoni parallel korpus yaratish uchun ham muhim ahamiyat kasb etadi. Jumladan, turli tillardagi tarjima matnlaridan iborat parallel korpusda lemmatizatsiya manba va maqsadli tillar o'rtasida mos keladigan so'zlarni yanada aniqroq moslashtirishga yordam beradi. Lemmatizatsiya so'zlarni asosiy shakllariga qisqartirish orqali tillardagi ekvivalent so'zlar va iboralarni yaxshiroq moslashtirishga yordam beradi. Lemmatizatsiya manba tildagi turli flektiv shakllardagi bir xil so'zning maqsadli tildagi bir xil tayanch shaklga o'tishini ta'minlaydi. Bu tarjimalarda izchillikni saqlashga va parallel korpus sifatini yaxshilashga yordam beradi. So'zlarni asosiy shakllariga standartlashtirishda lemmatizatsiya noaniqlikni kamaytirish va ekvivalent so'z va iboralarning tillar o'rtasida to'g'ri mos kelishini ta'minlash orqali tarjima sifatini yaxshilashga yordam beradi. Bundan tashqari, lemmatizatsiya mashina tarjimasi, tillararo ma'lumotni qidirish va ko'p tilli tabiiy tilni qayta ishlash kabi tahlil vazifalarining aniqligini oshirishi mumkin. Umuman olganda, lemmatizatsiya turli xil ko'p tilli matnlarni

qayta ishlash va tarjima dasturlari uchun zarur bo'lgan yuqori sifatli, izchil va moslashtirilgan parallel korpusni yaratishda hal qiluvchi rol o'ynaydi.



1-rasm. Lemmatizatsiyaning parallel korpus uchun ahamiyati

Asosiy qism. Lemmatizatsiya jarayoni korpus funktsionalligini va tahlil samaradorligini oshirish uchun foydali jarayon hisoblanadi [Xusainova, 2023: 2].

Lemmatizatsiyani korpusda to'g'ri qo'llash uchun uni stemmingdan farqlash lozim. Lemmatizatsiya va stemming – tabiiy tilni qayta ishlash (NLP) da tahlil murakkabligini kamaytirish maqsadida ishlatiladigan ikkita usul. Bu ikkala usul ham so'zlarni asosiy shakllariga qisqartirishdan iborat, ammo ular bu amalni bajarish maqsadi va unga erishish usullari bilan farqlanadi [Elov, 2023: 46].

Stemming asosan so'zlardan affikslarni ajratib, faqat asosiy shaklni qoldiradi [Manning, 1999: 134]. Stemming va lemmatizatsiya ikkala usul ham so'zlarni asosiy yoki ildiz shakliga qisqartirish uchun tabiiy tilni qayta ishlashda qo'llaniladi. Biroq, ular o'zlarining yondashuvlari va ular ta'minlaydigan normalizatsiya darajasida farqlanadi. Quyidagi jadvallarda stemming va lemmatizatsiya farqlarini ko'rib chiqamiz:

STEMMING	LEMMATIZATSIYA
Stemming - bu qo'shimcha va old qo'shimchalarni olib tashlash orqali so'zlarni ildiz shakliga keltirish jarayonidir. Bu oddiy qoidalar va evristika yordamida, so'zning kontekstini hisobga olmagan holda amalga oshiriladi	Lemmatizatsiya, aksincha, so'z kontekstini ko'rib chiqishda so'zlarni ularning asosiga yoki lug'at shakliga (lemma) qisqartirishni o'z ichiga oladi.
Stemmingda so'zlar har doim ham haqiqiy so'zlar bo'lavermaydi, chunki jarayon so'z o'zagini olish uchun affikslarni olib tashlashga qaratilgan.	Lemmatizatsiya so'zning asos yoki lug'at shaklini to'g'ri qaytarish uchun lug'at tahlili va morfologik tahlildan foydalanadi.

Agglyutativ tillarda ko‘pincha ildizga qo‘shila oladigan qo‘shimcha va old qo‘shimchalar ko‘p bo‘lganligi sababli, stemming o‘zaklarning haqiqiy so‘z bo‘lmasligi va so‘zning asosiy ma’nosini to‘liq anglatmasligi mumkin.	Agglyutativ tillarda lemmatizatsiya so‘zning asos shaklini to‘g‘ri qaytarish uchun turli affikslar va ularning ma’nolarini hisobga oladi. Bu lemmalashtirilgan shaklning haqiqiy so‘z bo‘lishini va mo‘ljallangan ma’noni saqlab qolishini ta’minlaydi.

1-jadval. Stemming va lemmatizatsiyaning farqlari

STEMMING	LEMMATIZATSIYA
taroq → taro	taroq → tara
maktabim → maktab	maktabim → maktab
mening → me	mening → men
shunday → shun	shunday → shu
burni → bur	burni → burn
oltovlon → olt	oltovlon → olti

2-jadval. Stemming va lemmatizatsiyaga misollar

Umuman olganda, stemming oddiy va tezkor usul bo‘lib, har doim ham haqiqiy so‘zlarni keltirib chiqarmasligi mumkin, lemmatizatsiya esa so‘z kontekstini hisobga olgan holda haqiqiy so‘zlarni qaytaradigan murakkabroq jarayondir. Agglyutativ tillarda ya’ni so‘zlar ildizga affikslar qo‘shilishi bilan hosil bo‘lgan tillarda o‘zak va lemmatizatsiya o‘rtasidagi farqlar alohida ahamiyatga ega bo‘ladi. Bunda affikslarning ko‘pligi sababli so‘z shakllarining to‘liq murakkabligini anglash qiyin bo‘lishi mumkin, lemmatizatsiya esa so‘z morfologiyasining nozik tomonlarini ko‘rib chiqish va kontekstda to‘g‘ri asos shakllarini ta’minlash uchun ko‘proq mos keladi.

So‘zlarning tarkibiy qismlari bilan shug‘ullanadigan morfologik tahlil NLPning asosiy yo‘nalishlaridan biri hisoblanadi. Bugungi kunda morfologik tahlilni yanada samaraliroq amalga oshirish uchun turli usullar va algortimlar ishlab chiqilgan hamda joriy qilingan [Sharipov, 2022: 4]. O‘zbek tilidagi so‘zlar lemmasini aniqlash uchun turli xil algortimlar taklif qilingan [Elov, 2022: 16].

Quyida Maksud Sharipov va Og‘abek Sobirov tomonidan ishlab chiqilgan o‘zbek tilidagi so‘zning lemmasini aniqlash uchun lemmatizatsiya algoritmi taqdim etilgan (1-chizma).

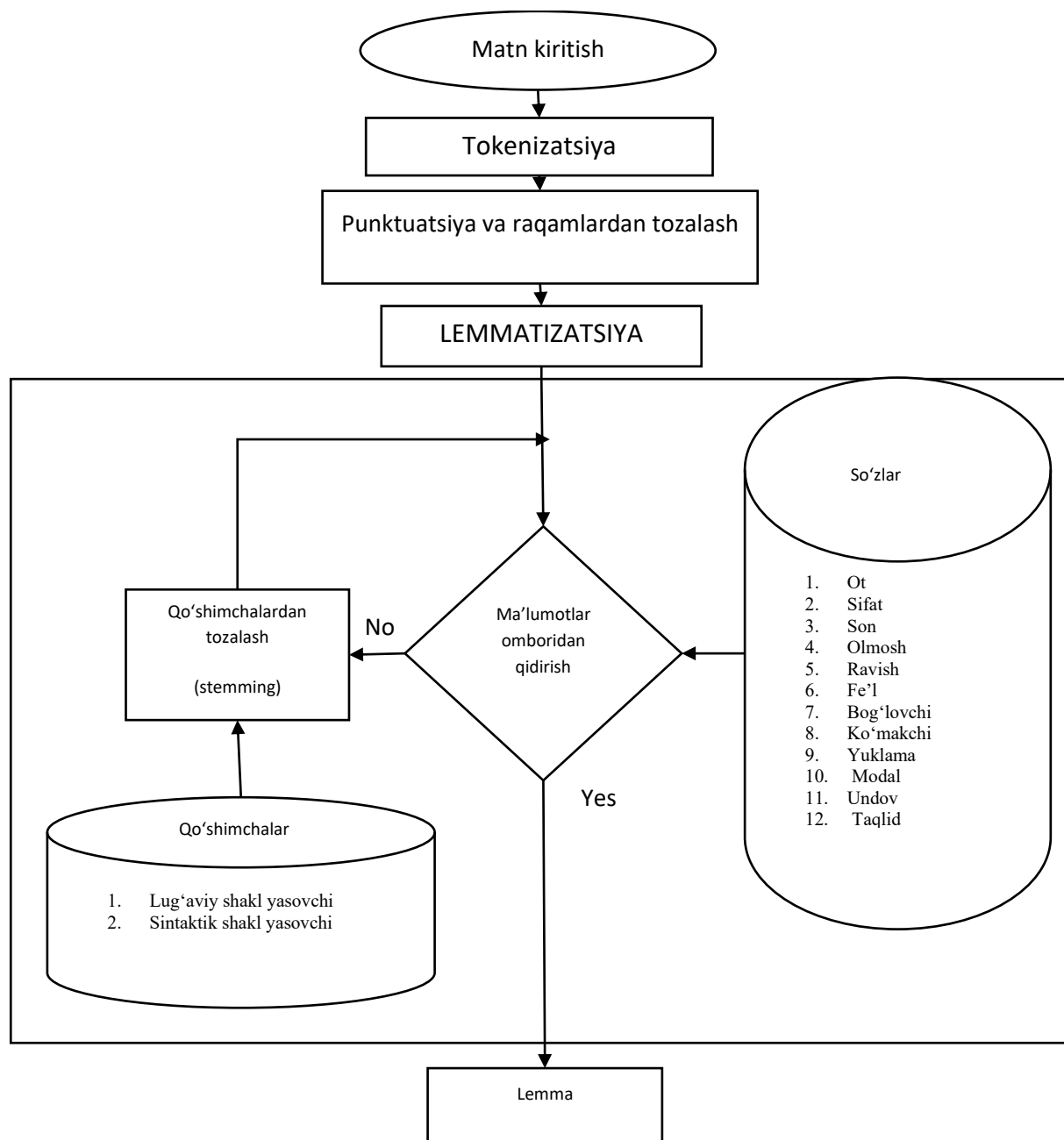
Lemmatizatsiya algoritmini amalda qo‘llash uchun qadamlar ketma-ketligi:

- 1-qadam. Dasturga matn kiritiladi;
- 2-qadam. Matn tokenizatsiya yordamida tokenlarga ajratiladi;
- 3-qadam. Agar matnda tinish belgisi yoki raqamlar bo‘lsa, u olib tashlanadi.
- 4-qadam. Token “so‘zlar” ma’lumotlar bazasidan qidiriladi, agar bazada mavjud bo‘lsa o‘tkazib yuboriladi. Mavjud bo‘lmasa 5-qadam bajariladi.



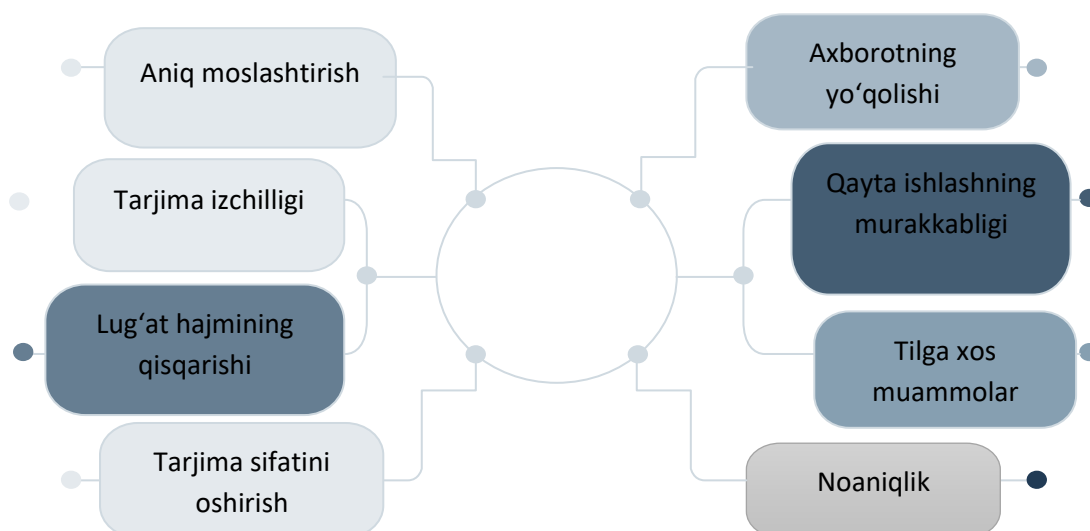
5-qadam. Token qo'shimchalardan tozalanadi, ya'ni stemming amalga oshiriladi. Bunda qo'shimchalar bazasidan foydalaniladi.

Parallel korpuslar yaratish uchun stemming va lemmatizatsiya bir necha sabablarga ko'ra muhimdir: stemming va lemmatizatsiya so'zlarni parallel korpusda turli tillarda ifodalashda izchillikka erishishga yordam beradi. So'zlarni ildiz shakllariga qisqartirish orqali, stemming va lemmatizatsiya mashina tarjimai tizimlarining aniqligini oshirishi mumkin. So'zlar o'zlarining asosiy shakllariga tushirilganda, bu boshqa tildagi mos keladigan so'zlarni aniqroq topishga yordam beradi. Bu tarjima modellarini yaratish va mashinali tarjima tizimlarini o'qitish uchun parallel korpuslarni moslashtirish uchun zarurdir. Umuman olganda, stemming va lemmatizatsiya parallel korpuslarni mashina tarjimai va boshqa tabiiy tilni qayta ishlash vazifalariga tayyorlashda hal qiluvchi rol o'ynaydi.



1-chizma. O'zbek tilidagi so'zlar lemmasini aniqlash uchun
lemmatizatsiya algoritmi

Quyidagi sxemada parallel korpus uchun lemmatizatsiyaning afzalliklari va kamchiliklarini ko'rib chiqish mumkin:



2-chizma. Lemmatizatsiyaning afzalliklari va kamchiliklari

Mazkur 2-chizma quyidagi 3-jadvalda izohlangan:

Parallel korpus uchun lemmatizatsiyaning afzalliklari	Parallel korpus uchun lemmatizatsiyaning kamchiliklari
Lemmatizatsiya manba va maqsadli tillar o'rtasida mos keladigan so'zlarni yanada aniqroq moslashtirishga yordam beradi, bu esa yanada sifatli parallel korpusga olib keladi.	Lemmatizatsiya fleksiya shakllarda mavjud bo'lgan o'ziga xos morfologik yoki sintaktik ma'lumotlarning yo'qolishiga olib kelishi mumkin, bu tarjima matnlarining boyligi va nuanslariga ta'sir qiladi.
Lemmatizatsiya manba tildagi turli flektiv shakllardagi bir xil so'zning tarjimalarda izchillikni saqlab, maqsadli tildagi bir xil tayanch shaklga o'tishini ta'minlaydi.	Lemmatizatsiya jarayoni, ayniqsa, keng ko'lamli parallel korpus qurish va texnik xizmat ko'rsatish uchun hisoblash murakkabligi va vaqtni oshishiga sabab bo'ladi.
Lemmatizatsiya so'zlarning flektiv shakllarini umumiy tayanch shaklga birlashtirib, umumiy lug'at hajmini qisqartiradi, korpus bilan ishlash va tahlil qilishni osonlashtiradi.	Lemmatizatsiya algoritmlari barcha tillar uchun bir xil darajada samarali bo'lmashligi mumkin, chunki ular flektiv yoki agglutinativ tillar bilan kurash olib borishi mumkin, bu esa moslashtirish va tarjimada noaniqliklarga olib keladi.
So'zlarni asosiy shakllariga standartlashtirish noaniqlikni kamaytiradi va tillar o'rtasida to'g'ri	Lemmatizatsiya ba'zi so'zlar yoki iboralardagi noaniqliklarni to'liq bartaraf etmasligi mumkin, bu parallel

moslashishni ta'minlaydi va tarjima sifatini yaxshilashga yordam beradi.	korpusda moslashtirish va tarjimaning to'g'riligiga ta'sir qilishi mumkin.

3-jadval. Lemmatizatsiyaning afzalliklari va kamchiliklari

Umuman olganda, lemmatizatsiya parallel korpus qurish uchun muhim afzalliklarni taqdim etsa-da, u potensial ma'lumot yo'qotilishi, hisoblash murakkabligi, tilga xos muammolar va noaniqliklarni hal qilish bilan bog'liq muammolarni ham keltirib chiqaradi. Parallel korpuslarni yaratish va tahlil qilishda lemmatizatsiyani qo'llashda ushbu omillarni diqqat bilan ko'rib chiqish kerak.

Bizning ishimizda asosiy maqsad o'zbek-turk parallel korpusini yaratishdir. Shuning uchun, o'zbek va turk tilidagi so'zlarning lemmasini parallel ravishda aniqlash tadqiqotning asosiy maqsadidir. Quydagi jadvalda o'zbek va turk tilidagi so'zlarning lemmasi parallel ravishda aniqlanganligiga misollarni ko'rishingiz mumkin:

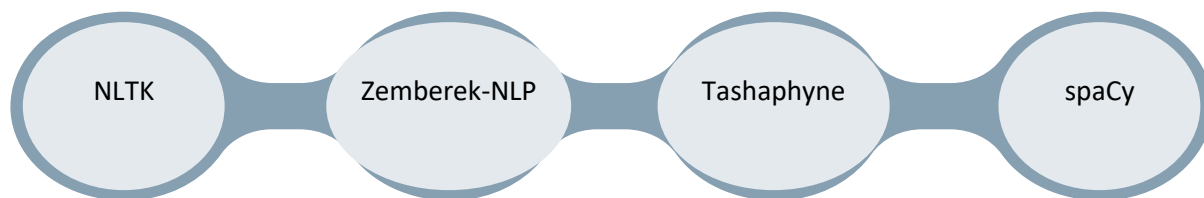
O'zbekcha so'zlar	Turkcha so'zlar	O'zbekcha va Turkcha so'zlar lemmasi
derazadan → deraza o'yladim → o'yla nafasimni → nafas issiqligidan → issiq ko'ryapman → ko'r meniki → men	pencereden → pencere düşündüm → düşün nefesimi → nefes sıcaklığından → sıcak görüyorum → gör benim → ben	deraza – pencere o'ylamoq – düşünmek nafas – nefes issiq – sıcak ko'rmoq – görmek men – ben

4-jadval. O'zbek va turk tilidagi so'zlarning lemmasi parallel ravishda aniqlanganligiga misollar

O'zbek va turk tillarining parallel korpusini yaratish uchun bir nechta lemmatizatsiya modellari variantlari ishlab chiqilgan. Masalan, o'zbek va turk tillari uchun oldindan tayyorlangan lemmatizatsiya modellari mavjud. Ushbu modellar katta korpuslarda o'qitilgan va so'zlarni o'z tillarida aniq lemmatizatsiya qila oladi. O'zbek tili uchun o'zbekcha matn bo'yicha o'rgatilgan modellarni, turkcha uchun esa turkcha lemmatizatsiya uchun maxsus ishlab chiqilgan modellarni topish mumkin. Jumladan:

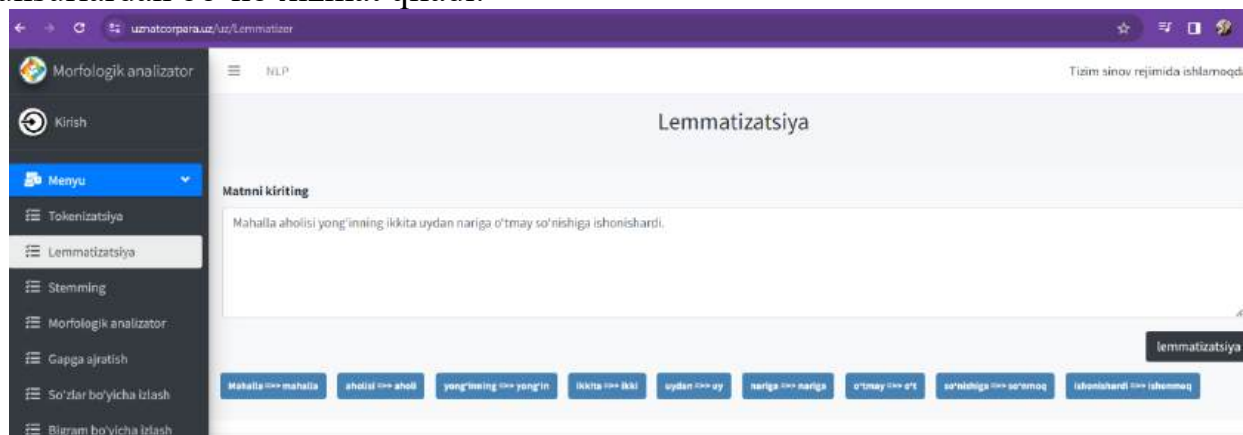
- NLTK (Natural Language Toolkit): NLTK o'zining nltk.stem.snowball moduli orqali turk tilini lemmatizatsiya qilishni ta'minlaydi.
- spaCy: spaCy – Turk va o'zbek tillari uchun lemmatizatsiyani qo'llab-quvvatlaydigan mashhur tabiiy tillarni qayta ishlash kutubxonasi.
- Tashaphyne: Tashaphyne turkiy tillarda, jumladan turk va o'zbek tillarida tabiiy tillarni qayta ishlash vazifalari uchun maxsus ishlab chiqilgan Python kutubxonasidir. Bu tillar uchun lemmatizatsiya funksiyasini taqdim etadi.

➤ Zemberek-NLP: Zemberek-NLP Java-ga asoslangan tabiiy tillarni qayta ishlash kutubxonasi bo'lib, turk tili uchun lemmatizatsiyani qo'llab-quvvatlaydi.



3-chizma. Lemmatizatsiya modellari

Bundan tashqari, o'zbek tili uchun O'zbek tili morfologik analizatoridagi UzLemmatizatoridan foydalanish mumkin. Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti kompyuter lingvistikasi va raqamli texnologiyalar kafedrasining tashabbuskorligi natijasida yaratilgan *uznatcorpara.uz*¹⁰ sayti orqali lemmatizatsiya amallarini bajarish imkoni yaratilgan bo'lib, bu o'zbek-turk tillari parallel korpusini yaratishda muhim va kerakli unsurlardan bo'lib xizmat qiladi.



2-rasm. UzLemmatizatoridan qidiruv natijasi

Oldindan tayyorlangan modellarning mavjudligi va sifati har xil bo'lishi mumkinligini yodda tutgan holda, ularni maxsus talablar va foydalanish holati asosida baholash muhimdir. Agar oldindan o'rgatilgan modellar mavjud bo'lmasa yoki qoniqarli ishlashni ta'minlamasa, maxsus lemmatizatsiya modellari mashinani o'rganish texnikasi yordamida ishlab chiqilishi mumkin. Bu har ikki tildagi so'zlarning asosiy shakllarini to'g'ri bashorat qilish uchun o'zbek va turk matnlarning katta korpusida lemmatizatsiya modelini o'rgatishdan iborat. O'zbek va turk tillarining parallel korpusini yaratishda har bir til uchun maxsus ishlab chiqilgan yoki bu tillarda yaxshi ishlashi ko'rsatilgan lemmatizatsiya modellarini tanlash

¹⁰ <https://uznatcorpara.uz/uz/Lemmatizer>

muhim ahamiyatga ega. Bundan tashqari, ikkala tildagi ma'lumotlarning kichik namunasi bo'yicha lemmatizatsiya modellarining to'g'riligini baholash parallel korpus sifatini ta'minlash uchun juda muhimdir.

Xulosa qilib shuni aytish mumkinki, lemmatizatsiya ayniqsa turkiy tillar kabi agglyutinatív tillardagi parallel korpuslar uchun bu tillarning murakkab morfologik tuzilishi tufayli muhim ahamiyatga ega. Agglyutinatív tillarda so'zlar o'zak yoki o'zakga old qo'shimcha, qo'shimcha va infiks qo'shish orqali hosil bo'ladi, natijada har bir so'z uchun turli xil flektiv shakllar paydo bo'ladi. Umuman olganda, lemmatizatsiya parallel korpusdagi agglyutinatív tillarning boy morfologiyasini boshqarishda hal qiluvchi rol o'ynaydi, bu aniqroq moslashtirish, tarjima sifatini yaxshilash, so'z boyligini oshirish, hisoblash samaradorligi va lingvistik tahlilni yaxshilash imkonini beradi.

Foydalanilgan adabiyotlar:

1. Jabeen, Balakrishnan, Ethel, Khyani. Stemming and Lemmatization: A Comparison of Retrieval Performances. January 2014. Lecture Notes on Software Engineering 2(3):262-267 DOI:[10.7763/LNSE.2014.V2.134](https://doi.org/10.7763/LNSE.2014.V2.134)
2. Saumyab. Stemming vs Lemmatization in NLP: Must-Know Differences. 15 Jan, 2024
3. Xusainova Z. O'zbek tili milliy korpusi qidiruv tizimini optimallashtirishda lemmatizatsiyadan foydalanish. 2023 Vol. 2 (6) compling.tsuull.uz
4. Elov B.B, Hamroyeva sh.M., Abdullayeva O.X., Xusainova Z.Y., Xudayberganov N.U., 2023: 46
5. Christopher Manning, Hinrich Schütze. Foundations of Statistical Natural Language Processing, MIT Press, 1999
6. Sharipov, Sobirov. Development of a Rule-Based Lemmatization Algorithm Through Finite State Machine for Uzbek Language Urgench State University, Department of Information Technologies, 14, Kh.Alimdjon str, Urgench city, 220100, Uzbekistan
7. Elov B. Hamroyeva Sh. Axmedova X. Methods for creating a morphological analyzer// 14th International Conference on Intelligent human Computer Interaction, IhCI Tashkent.2022, 19-23 October. https://link.springer.com/chapter/10.1007/978-3-031-27199-1_4

O'LCHOV BIRLIKLARINING MILLIY KORPUSDA BERILISHI

Achilova Noila Hamroqulovna

Qorako'l tumani ixtisoslashtirilgan maktab internati
ona tili va adabiyot fani o'qituvchisi

Annotatsiya. Ushbu maqolada bugungi kunda tilshunoslik bilan integratsion fan sifatida yuzaga kelgan korpus lingvistikasi va unda o'lchov so'zlarning aks ettirilishi xususida fikrlar tahlilga tortiladi. O'lchov birliklaridan milliy korpusda foydalanish ushbu so'zlarning kompyuter tilida foydalanilishiga keng imkoniyat yaratadi hamda tilning tadqiq manbasiga oid modellashtirish jarayonini osonlashtiradi.

Abstract. In this article, corpus linguistics, which has emerged today as an interactive science with linguistics, and the reflection of measurement words in it, are analyzed. The use of units of measurement in the national corpus provides a wide opportunity for the use of these words in a computer language and facilitates the modeling process of language research resources.

Аннотация. В данной статье анализируется корпусная лингвистика, возникшая сегодня как интерактивная наука с лингвистикой, и отражение в ней измерительных слов. Использование единиц измерения в национальном корпусе дает широкие возможности для использования этих слов в компьютерном языке и облегчает процесс моделирования языковых исследовательских ресурсов.

Kalit so'zlar: *o'lchov birliklari, milliy korpus, tadqiqot manbai, modellashtirish, tahlil qilish.*

Korpus lingvistikasi so'nggi o'n yillikda amaliy tilshunoslikning muhim sohasi sifatida maydonga chiqdi. Korpus tilshunosligining asosiy yo'nalishi grammatikalar va lug'atlar yozishda, tillar o'rtasidagi farqlarni o'rganishda va til o'qitish resurslarida korpusdan foydalanish bo'ldi. Korpus lingvistikasidan turli xil aniq maqsadlarda foydalanilgan: nutq belgilarining roli, darslik tilining tabiati, tildagi kuch va maqomning roli, matnlarda leksikani o'rganish, o'quv materiallarida matn segmentlarini tanlash va baholash vazifalari. Bu faoliyatlar katta korpuslar yoki muayyan maqsadlar uchun yig'ilgan kichikroq korpuslardan foydalanish orqali amalga oshirilishi mumkin. Korpus lingvistikasi atamasi, umuman, korpusga asoslangan lingvistik tadqiqotlarni anglatadi. Arxetip korpus ishi zamonaviy raqamli davrdan ancha oldin mavjud bo'lgan. Biroq, korpus lingvistikasining akademik fan sifatida paydo bo'lishi kompyuter texnologiyalarining jadal rivojlanishi, shuningdek, o'tgan asrning ikkinchi yarmidan boshlab raqamli kontentning mavjudligi bilan bog'liq bo'lgan korpuslarni yozib olish, saqlash va tekshirish uchun raqamli vositalar mavjudligi bilan chambarchas bog'liq. Hozirgi vaqtda korpus



tilshunosligida hisoblash va statistik vositalardan foydalanish hal qiluvchi rol o'ynaydi.

Korpus lingvistikasi doirasida ko'plab tadqiqotlar amalga oshirilgan bo'lib, bizning ushbu mavzuyimiz ham bevosita milliy korpus tarkibida tahlil qilinuvchi hisob so'zlar sanaladi. Korpus lingvistikasi kontekstida numerativ so'zlar matn korpusidagi son ma'lumotlarini yoki miqdorlarni ifodalovchi so'zlarni anglatadi. Ushbu so'zlar turli xil lingvistik tahlillarda, jumladan, chastotalarni hisoblashda, miqdoriy tadqiqotlarda va statistik modellashtirishda hal qiluvchi rol o'ynaydi. Numerativ so'zlar korpus ichidagi raqamli ma'lumotlarning tarqalishi, ma'no va munosabatlari haqida tushuncha beradi. Korpus tilshunosligida numerativ so'zlar bilan bog'liq ba'zi asosiy jihatlar mavjud bo'lib, ularni quyidagicha izohlash mumkin. Numerativ so'zlar son, miqdor, o'lchov, nisbat va boshqa son qiymatlarni bildiruvchi lug'aviy birikmalarning keng doirasini o'z ichiga oladi. Ushbu leksik birliklarni milliy korpus tarkibida o'rganish so'z modellashtirish jarayoniga yordam beradi.

Numerativ so'zlar korpus tilshunosligida chastota tahlilini o'tkazish uchun zarurdir. Tadqiqotchilar korpus ichida muayyan atamalar, iboralar yoki lingvistik xususiyatlarning mavjudligini miqdoriy aniqlash uchun numerativ so'zlardan foydalanadilar. Numerativ so'zlarning chastotasi lug'atdan foydalanish, leksik xilma-xillik va til ma'nolari haqida tushuncha beradi [James, 1975:69].

Hisob so'zlar tadqiqotchilar tomonidan korpus lingvistikasida lingvistik ma'lumotlarni raqamli o'lchash va tahlil qilish imkonini beruvchi miqdoriy tadqiqotlarni amalga oshirishlariga yordam beradi. Miqdoriy tadqiqotlar so'z chastotalarini hisoblash, raqamli taqsimotlarni solishtirish, matn o'zgaruvchanligini baholash yoki lingvistik elementlar o'rtasidagi raqamli munosabatlarni o'rganishni o'z ichiga olishi mumkin. Numerativ so'zlar asosida statistik modellashtirish va til hodisalarini miqdoriy tahlil qilishda muhim rol o'ynaydi. Tadqiqotchilar statistik modellarni yaratish, gipoteza sinovini o'tkazish, lingvistik ma'nolarni ochib berish va tildan foydalanish haqida tushuncha olish uchun korpusdan olingan raqamli ma'lumotlardan, jumladan, numerativ so'zlardan foydalanadilar. Matn mazmuniga qo'shimcha ravishda numerativ so'zlar korpus lingvistikasida matnlarning raqamli atributlari haqida qo'shimcha kontekstni ta'minlash uchun metama'lumotlar sifatida ishlatilishi mumkin. Raqamli metama'lumotlar so'zlar soni, jumlar uzunligi, hujjat chastotasi, leksik zichlik yoki tahlilga tegishli har qanday boshqa miqdoriy xususiyatlar kabi ma'lumotlarni o'z ichiga olishi mumkin. Ko'p tilli yoki xilma-xil korpuslar bilan ishlashda bir xil tahlil qilish uchun tillar yoki dialektlar bo'ylab sonli so'zlarni standartlashtirish juda muhimdir. Normalizatsiya tartib-qoidalari raqamli ko'rinishlardagi izchillikni ta'minlaydi va turli til turlari bo'yicha qiyosiy tadqiqotlarni osonlashtiradi [James, 1975:156].

Yuqorida qayd etib o'tganimizdek, hisob so'zlar korpus lingvistikasida sonlarning atributiv belgilari sifatida namoyon bo'ladi. Ushbu ifodalarni milliy korpusda aks ettirish orqali tilimizdagi so'z boyligini kompyuter dasturlariga

kiritish, uning soʻz modellash formalarini yaratishda qoʻllash mumkin. Avvalo, ularning milliy korpus bazasida qanday teg asosida aks ettirilishini tahlil qilsak.

Hisob soʻzlarni milliy korpusga kiritishni quyidagi ketma-ketlik asosida amalga oshirish lozim:

1. Milliy korpus tarkibiga hisob soʻzlarga oid belgi kiritish. Masalan, oʻzbek milliy korpusida uni HS-hisob soʻzlar tarzida (ingliz tilida Numerals) keltirish mumkin. Bu til modelini tahlil qilish jarayonida soʻzning qaysi guruhga oid ekanligini ochiqlashda yordam beradi.

2. Hisob soʻzlarga oid guruhni milliy korpusda aks ettirish. Ularni kichik guruhlariga boʻlish hamda ushbu guruhga kiruvchi leksik ifodalarni ham modulda aks ettirish. Uni quyidagi jadval asosida koʻrib oʻtamiz [Achilova, 2023:5]

<i>Hisob soʻz(HS)</i>	<i>Ogʻirlik oʻlchovini ifodalovchi hisob soʻzlar</i>	<i>Gramm</i>
		<i>Kilogramm</i>
		<i>Botmon</i>
		<i>Misqol</i>
		<i>Tonna</i>
		<i>Sentner</i>

Hisob soʻzlarni milliy korpusda modellashtirish uchun korpus bazasiga yuqoridagi jadvalda keltirilgan namuna asosida (guruh ichida guruh) keltirish ushbu turkum soʻzlarning yanada kengroq aniqlashtirilishiga, soʻzlarning semantik, grammatik(morfologik va sintaktik), lemma tahliliga oid jarayonlarda ushbu guruh soʻzlarning yanada aks ettirilishiga xizmat qiladi [Crespo, 2023:78].

Xulosa sifatida aytish mumkinki, korpus lingvistikasida numerativ soʻzlardan foydalanish orqali tadqiqotchilar chuqur miqdoriy tahlillar oʻtkazishlari, tildan foydalanishning raqamli jihatlarini oʻrganishlari va matn maʼlumotlaridagi til tuzilishi, oʻzgaruvchanligi va aloqa shakllari haqida mazmunli tushunchalar olishlari mumkin.

Foydalanilgan adabiyotlar:

1. Achilova N. Oʻlchov soʻzlar pragmatikasi//Oʻzbekistonda milliy tadqiqotlar:davriy anjumanlar [7-qism], 2022. – B. 19-21.

2. Achilova N. Pragmatic features of numerical words expressing the meaning of money//Web if scientist: International scientific research journal, 2022. – B. 524-530.

3. Achilova N. Pul qiymatini ifodalovchi numerativlarning pragmatik xususiyatlari//Taʼlimda filologiyani rivojlantirishning global masalalari, 2022. – B. 140-143.



4. Achilova N. Suyuqlik miqdorini anglatuvchi lisoniy birliklarning pragmatik xususiyatlari//So‘z san’ati xalqaro jurnali, 2022. – B. 74-80.
5. Crespo Cantalapiedra, I. (2023). La diversidad en las lenguas: los numerales. Online book (in Spanish).
6. James R. Hurford (2010) [1975]. The Linguistic Theory of Numerals. Cambridge University Press. ISBN 978-0-521-13368-5.
7. Qilichev E. Hozirgi o‘zbek adabiy tili. –Buxoro: Buxoro universiteti, 1999.
8. Rahimov S., Umurqulov B. Hozirgi o‘zbek adabiy tili. –Toshkent: O‘qituvchi, 2003. 220 b.
9. Rahmatullayev Sh. Hozirgi o‘zbek adabiy tili. –Toshkent: Universitet, 2006.

PANXRONIK KORPUS TAVSIFI, IMKONIYAT VA AFZALLIKLARI (RUS TILI PANXRONIK KORPUSI MISOLIDA)

Bekmuradova Iroda Zokir qizi

irodabekmuradova9@gmail.com

ToshDO‘TAU magistranti

Annotatsiya. Ushbu maqolada “panxron” atamasi ta’rifi, panxronik korpus xususiyatlari tavsifi yoritilgan bo‘lib, jahon tilshunoslik ilmiga panxronik metodning kirib kelishi tarixi va zarurati haqida nazariy qarashlar keltirilgan. Ushbu maqolada panxronik korpus haqida umumiy ma’lumot beriladi, uning tuzilishi, mazmuni va lingvistik tahlilda qo‘llanilishi mumkin bo‘lgan ahamiyati muhokama qilinadi. Maqolada, shuningdek, til birliklarining o‘zgarishini o‘rganish, tillararo aloqalarning mazkur o‘zgarishlarga ta’sirini baholash uchun panxronik korpusdan foydalanishning afzalliklari haqida so‘z yuritiladi.

Abstract. In this article, the definition of the term “panchronic”, the description of the characteristics of the panchronic corpus, and the theoretical views on the history and necessity of the introduction of the panchronic method to world linguistics are presented. This article provides an overview of the panchronic corpus, discussing its structure, content, and potential applications in linguistic analysis. The article also discusses the advantages of using a panchronic corpus to study the change of language units and to assess the impact of interlinguistic contacts on these changes.

Аннотация. В данной статье представлены определение термина «панхронный», описание характеристик панхронного корпуса, а также теоретические взгляды на историю и необходимость внедрения панхронического метода в мировое языкознание. В этой статье представлен обзор панхронического корпуса, обсуждаются его структура, содержание и потенциальные применения в лингвистическом анализе. В статье также рассматриваются преимущества использования панхронного корпуса для изучения смены языковых единиц и оценки влияния межъязыковых контактов на эти изменения.

Kalit so‘zlar: *panxronistika, panxronik va axronik metod, panxronik korpus.*

Kirish

Milliy korpuslar orasida rus tili milliy korpusi o‘zining ichki funksiyalari, qamrab olgan baza hajmi hamda korpus turlari xilma-xilligi bo‘yicha alohida ajralib turadi. Rus tili milliy korpusi – ommabop rus matnlari korpusi. U 2004-yilda tashkil etilgan, 2007-yilda 140 million so‘z qo‘llash holatini qamrab olgan yozma, og‘zaki, badiiy, dialektal matnlar yig‘indisini tashkil etgan. 2013-yilda hajmi 500 million so‘z bo‘lgan. Mazkur korpus doimiy ravishda yangilanib to‘ldirilib boriladi. Uning yana bir xususiyati matnning janriy xilma-xilligi, rus tilida qo‘llanadigan barcha

soha uslub va janrlarni qamrab olganidir. [Hamroyeva, 2020:52] Rus milliy korpusi asosiy korpusdan tashqari gazeta matnlari korpusi, ijtimoiy tarmoqlar korpusi, ogʻzaki, multimediali, dialektal, taʼlim, poetik, tarixiy, panxronik korpus kabilarni oʻz ichiga olgan. Shular orasida eng eʼtiborlisi rus panxronik korpusidir.

“Panxron” soʻzi atama sifatida tilshunoslikda turli darajadagi umumlashtirish maʼnosida keng qoʻllanadi. Bu tushunchaning asosiy taʼrifi quyidagicha: “Til hodisalarini universal, zamonaviy va umuman, inson tiliga xos boʻlgan xususiyatlar asosida koʻrib chiqish.” Sodda tarzda aytganda “panxron” atamasi badiiy matnlardagi interpretatsiya jarayonini ifodalaydi. Yaʼni, bunda adabiy matnlar tarkibidagi vaqtinchalik-lingvokulturologik belgilari bilan birga tahlilga olinadi hamda boshqa xronologik davrlarda yaratilgan matnlar bilan taqqoslanadi. Tilshunos olim V.I. Postovalova V.Gumboldt qarashlarini tahlil qilar ekan “panxronizm da doimiy va azaliy tushunchalar yordamida tahlil qilinayotgan protsedura uchun vaqt dolzarbligi muammolari dialektik harakatda oʻrganiladi”, deydi. Yaʼni mazkur tahlilda obyekt tarixiy nuqtayi nazardan klassik dialektiv metod asosida tekshiriladi hamda bugungi xususiyatlari ham qoʻshib tahlil qilinadi, muayyan xronologik davrdagi oʻzgarishlari aniqlanadi. [Постовалова, 2003:606]

Panxronik korpus xususiyatlarining umumiy tavsifini berishda obyekt sifatida rus tili panxronik korpusi olingan. Tahlildan asosiy maqsad milliy korpus funksiyalarini kengaytirishda panxronik korpus imkoniyatlaridan foydalanish hamda kelgusida mazkur korpusni tilimiz uchun qoʻllashga zamin hozirlashdan iborat.

Asosiy qism

Fanda panxronistikaga oid eng dastlabki dalillar oʻtgan asrning 40-yillaridan boshlab berila boshlangan. Soʻzning atama sifatida kelib chiqishi F.de Sossyurga borib taqaladi. Olim bu atamani eng umumiy tamoyillarga murojaat qilish uchun ishlatadi. Panxron usulning maqsadi har qanday muayyan til birliklarining u yoki bu xususiyati boʻyicha barcha xulosa va tahlillar umumlashmalarini shakllantirishdir.

Tilshunos olim L. Yelmslev universal lingvistik nazariyani oʻrganishga doir tadqiqotlarida tilshunos til elementlari oʻrtasidagi oʻzaro munosabatlarni emas, balki til strukturalari oʻrtasidagi munosabatni oʻrganishi kerak deb hisoblaydi. Tildagi ayrim elementlar funksiyalarning bir butogʻi boʻlsa, til funksiyalar zanjiridan iboratdir. Glossematiklar tilni oʻrganishdagi F. de Sossyur tomonidan ilgari surilgan sinxronik va diaxronik metod oʻrniga panxronik (uch zamonni oʻz ichiga oladigan) va axronik (zamondan tashqari) metodni tavsiya etadi. Panxronik va axronik metod tarixiylikka qarshi ketma-ketlik metodidir. Bu metod orqali tilning haqiqiy mohiyati emas, balki til muammolarini oʻrganish mumkin boʻladi. Ular tilni oʻrganishda falsafiy empirizmdan oʻtib, mantiqiy pozitivizmga tomon bordilar. [Oʻrinboyev, Musayev, 2021:98]

Panxronik korpus – tadqiqotchilarga vaqt oʻtishi bilan tildagi u yoki bu soʻzlarning yozilishi yoxud muayyan qonuniyatlar oʻzgarishi xususiyatlarini oʻrganish imkonini beradigan, bir necha vaqt oraliqini qamrab oluvchi korpus



turidir. Ushbu turdagi korpus tillarning qanday rivojlanishi va yangi ijtimoiy, madaniy o'zgarishlarga moslashishini kuzatmoqchi bo'lgan tilshunoslar va til tadqiqotchilari uchun qimmatlidir. Panxronik korpus lingvistik tadqiqotlar uchun qimmatli manba hisoblanadi, chunki u turli davrlar va geografik joylashuvlarga oid keng doiradagi matnlarni o'z ichiga oladi.

Panxronik korpus, odatda, qadimgi zamonlardan tortib to hozirgi davrgacha bo'lgan turli tarixiy davrlardagi matnlarni o'z ichiga oladi. Bu matnlar har xil manbalardan, jumladan, badiiy asarlar, tarixiy hujjatlar, diniy matnlar va kundalik yozuvlar bo'lishi mumkin. Tadqiqotchilar ushbu matnlarda til xususiyatlarini tahlil qilib, ularning vaqt o'tishi bilan qanday o'zgargani, shuningdek, migratsiya, savdo va madaniy almashinuv kabi tashqi omillar qanday ta'sir qilgani haqida tushunchaga ega bo'lishlari mumkin.

Panxronik korpusning asosiy afzalliklaridan biri shundaki, tadqiqotchilarga muayyan tilga xos lingvistik xususiyatlarning uzoq vaqt davomida rivojlanishini kuzatish imkonini beradi. Misol uchun, tilning lug'at boyligi, grammatika va talaffuzdagi o'zgarishlar, shuningdek, yozish uslublari va qoidalaridagi yangilanishlarni kuzatish mumkin. Mazkur korpus nafaqat, ma'lum bir vaqt oralig'ida sodir bo'lgan til o'zgarishini tushunish uchun ham muhim, balki turli mintaqalar yoki ijtimoiy guruhlar matnlarini solishtirish, mintaqaviy dialektlar hamda sotsiolektlarning til xilma-xilligi va boshqa shu kabi xususiyatlarini aniqlashda ahamiyatlidir. Bu esa turli jamoalar tomonidan tildan qanday foydalanilishi va idrok etilishi hamda ularning o'ziga xosliklari, tajribalarini yoritishga yordam beradi.

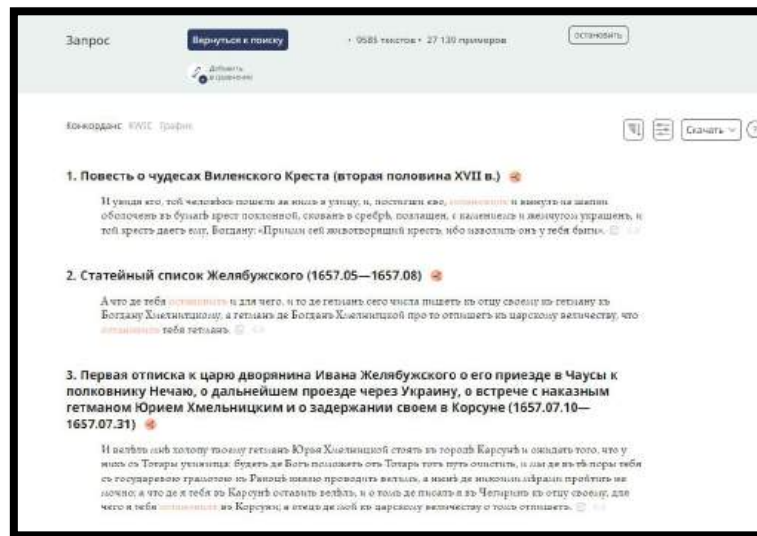
Rus panxronik korpus uchta tarixiy korpus – “drevnerusskiy”, “starorusskiy” va “berestyannix gramot” korpuslarini o'z ichiga olgan bo'lib, eng qadimgi matnlari XVII-XVIII asrlarning boshlariga to'g'ri keladi. Unda rus tili asosiy korpusidagi 300 milliondan ortiq matnlar ustunlik qilsa-da xronologik diapazonning 70 foizi tarixiy korpusga to'g'ri keladi. Mazkur korpus ustida I. I. Vinogradova, S. A. Gladilin, D. A. Morozov, V. G. Sizov, O. N. Lyashevskaya, D V. Sichinava, T. A. Arxangelskiy, A. N. Dyushkant kabi olimlar ishlagan.

Panxronik korpusda lemma, so'zshakl, so'zning grammatik belgilari, asosiy semantikasi va boshqa qo'shimcha belgilari asosida qidiruv so'rovlarni amalga oshirish mumkin. Bunda natijalar rus tilining bir necha asrlik tarixi davomida uchragan o'zgarishlari bilan o'sha davr matnlari misolida taqdim etiladi.

Panxronik korpusni yaratish zarurati, birinchi navbatda, turli xil tarixiy va zamonaviy korpuslarda lemmalar, imlo va grammatik qoidalarni o'zgarishini tahlil qilish bilan bog'liq. Panxronik korpus mavjud alohida tarixiy va asosiy korpuslarni almashtirmaydi. Grammatik, metamatn va boshqa tegishli imlo me'yorlar to'liq saqlab qolinadi. Mazkur mezonlar muntazam ravishda to'ldirib boriladi, yangilanadi va panxronik korpus ushbu to'ldirishlarni hisobga olgan holda natijalarni sinxronlashtiradi. Panxronik qidiruvda lemmaga oid qidiruv natijalari

soʻzning qadimgi tarixiy, yangi tarixiy va zamonaviy rus tilidagi qoʻllanilishi meʼyorlashtirilgan holda taqdim etiladi. Masalan, “*сѣлати – слати – слать*”.

Tarixiy korpusdan olingan matnlarda, barcha leksemalar hozirda qoʻllanilsa ham, barcha lugʻat shakllari koʻrsatiladi. Misol uchun, “berestyannix gramot” korpusida garchi oʻrta asrlarga kelib yoʻqolib ketgan boʻlsa-da, “*продажъникъ*” soʻzining keyingi shakllari “*продажники*” va “*продажник*”, “*крити*” soʻzining yangi shartli shakli “*крить*” mavjud. Soʻzning avvalgi shakllari faqatgina tarixiy davr matnlarida qoʻllanilgan holatdagina taqdim etiladi. Masalan, “*компьютер*” soʻzida bu imkoniyat mavjud emas, chunki ushbu soʻz yangi davr matnlaridan boshlab qoʻllanilib kelinadi.



1-rasm. Korpusda qidiruv natijalarini “konkordans” shaklda taqdim etish tizimi

Rus panxronik korpusidagi grammatik belgilar majmuasi alohida korpus belgilaridan farq qiladi. Masalan, panxronik korpusga kiritilgan matnlarda grammatik omonimiya olib tashlangan. Faqat “drevnerusskiy” va “berestyannix gramot” korpusi va zamonaviy matnlarning olti million subkorpusida bunday birliklar qoʻlda kiritib chiqilgan. Qadimgi rus korpusida va asosiy korpusdagi aksariyat matnlarda lemmalar va grammatik xususiyatlarni belgilash neyron tarmoq mexanizmlari yordamida avtomatik ravishda amalga oshiriladi. Bu degani belgilashda muayyan xatolar uchrashi mumkin demakdir.

Tarixiy matnlarda zamonaviy rus tilidagi tegishli soʻzlarning semantik xususiyatlari (etimologik kognatlarning) semantik sinflariga muvofiq belgilanadi. Leksik semantika tarixiy oʻzgarishlarga uchraganligi va bir qator soʻzlar zamonaviy tilda umuman yoʻqolib ketganligi hamda zamonaviy semantik lugʻatda yoʻqligi sababli tarixiy matnlarning semantik belgilanishida noaniqliklar va toʻliqsizliklar uchraydi. Shunga qaramay, tarixiy matnlarning ushbu yondashuv bilan olingan semantik belgilar bilan yuqori qamrab olinishi va aksariyat leksemalar semantik sinflarining barqarorligi bunday belgilanishning kamchiliklarini toʻldiradi.

tomonlama tadqiqini ta'minlashda mazkur korpus hajman katta va boy ma'lumotlar manbasini taqdim etadi. Texnologiya taraqqiyotda davom etar ekan, panxronik korpusni yaratish va tahlil qilish yanada qulayroq bo'lib, muayyan tilning murakkab va dinamik tabiatini tushunish uchun yangi imkoniyatlarni taqdim etadi.

Yuqoridagi tahlil va tafsilotlar kelgusida mazkur korpusni o'zbek tiliga oid manbalar asosida ham ishlab chiqish dolzarbligini ko'rsatadi. O'zbek panxronik korpusi hozirgi o'zbek tilining shakllanish bosqichlari va bu davrlarga oid qadimgi va eski o'zbek tili manbalarining til xususiyatlarini o'rganish, tilimizning bugungi holati shakllanishidagi o'rnini belgilash, til o'zgarishlarini tadqiq qilishda nihoyatda ahamiyatli bo'ladi.

Foydalanilgan adabiyotlar:

1. Hamroyeva Sh. Korpus lingvistikasi atamalarining qisqacha izohli lug'ati. Globe edit, 2020, 52-b.
2. O'rinboyev B., Musayev A. Umumiy tilshunoslik. -Jizzax, 2021, 98-b.
3. Постовалова В. И. Язык как деятельность. Опыт интерпретации концепции В. Гумбольта. -М., 2003, С.606.
4. <https://ruscorpora.ru>

ZAHIRIDDIN MUHAMMAD BOBUR MUALLIFLIK KORPUSINING TUZILISHI VA TARKIBI

Muradova Nilufar Baxodir qizi

nilufarmurodova818@gmail.com

ToshDO‘TAU 1-kurs magistranti

Annotatsiya. Bugungi kungacha dunyo bo'yicha uch yarim mingdan ortiq korpuslar. Ayni damda zamonaviy korpuslar ustida ishlar olib borilmoqda. Xususan, mualliflik korpuslariga ham bo'lgan talab ortib bormoqda. Ushbu maqolada Zahiriddin Muhammad Bobur mualliflik korpusining tuzilishi va tarkibi borasidagi fikrlar bayon etilgan. Muallif ijodiy merosi va asarlarining janriy xususiyatlari o'rganilgan.

Abstract. To date, more than three and a half thousand cases around the world. Modern hulls are currently being worked on. In particular, the demand for author corpora is also increasing. This article describes the views on the structure and composition of Zahiruddin Muhammad Babur's authorship. The author's creative heritage and genre features of his works are studied.

Аннотация. На сегодняшний день более трех с половиной тысяч случаев по всему миру. В настоящее время ведутся работы над современными корпусами. суд, требование или контроль над корпусом авторских прав. В трех статьях описан авторский корпус Захируддина Мухамада Бабура по его структуре и составу. Изучено творческое наследие автора и жанровые особенности его произведений.

Kalit so'zlar: *korpus, mualliflik korpusi, korpus tarkibi, qidiruv menejeri.*

Zahiriddin Muhammad Bobur va uning adabiy merosi nafaqat o'zbek xalqining, balki dunyo xalqlari, xususan, turkiy xalqlar adabiyotlarining ham xazinasini sifatida e'tiborga molik. Dunyo adabiyotida Bobur ijodidek ko'p o'rganilgan va tadqiq qilingan shaxslar kam. Hozirgi kunda Bobur shaxsi birlamchi bo'lib yozilgan asarlar soni 12 mingga yaqinni tashkil etadi. Bundan ko'rinadiki, Boburshunoslik sohasidagi ishlar ko'lami keng va miqyosi ham baland. Bobur turkiy she'riyat an'analarini Hindiston adabiy muhitiga olib kirgan va hind shoirlarining ham turk tilida ijod qilishiga rahnamolik qilgan shaxs. U o'zbek va fors tillarida ijod qilgan zullisonayn shoir.

Bobur muallifli korpusini ishlab chiqish hozirgi kundagi dolzarb vazifalardan biri sanaladi. Korpusdagi eng muhim tarkibiy qismidan biri korpus interfeysidir. Mualliflik korpusi interfeysining aniqligi juda muhim. Sababi foydalanuvchi tizimga kirganda unga tushunarli holatda bo'lishi, erkin holda foydalana olishi kerak. Interfeysda ma'lumotlar tartibli va tizimli shaklda berilsa, bu foydalanuvchiga yanada qulaylik yaratadi. Mualliflik korpusini ishlab chiqishda, har bir yozuvchi uchun alohida yondashuv zarur. Avvalo, yozuvchi ijodining o'ziga xosligini

o'rganish, ijodiy merosi bilan batafsil tanishish darkor. Negaki, mualliflik korpusining tuzilishini, tarkibini ishlab chiqishda yozuvchi ijodi xususiyatidan kelib chiqqan holda yondashiladi. Masalan, U.Shekspir mualliflik korpusi 6 bo'limdan iborat. Unda yozuvchining poemalari, sonetlari va dramatik asarlari alohida bo'lim holida joylashtirilgan. Boshqa ijodkor asarlari uchun bu bo'limlar mos kelmasligi aniq. Sababi, aynan U.Shekspir ijodi yuqoridagi janrlardan iborat. Shu sababi ham, har bir mualliflik korpusi uchun individual yondashuv zarur.

Zahiriddin Muhammad Bobur mualliflik korpusi 7 bo'limdan iborat bo'ladi.

1. Muallif hayoti va ijodiy faoliyati
2. “Boburnoma” asari
3. Bobur devoni (116 g'azal, 196 ruboiy, 8 masnaviy, 52 muammo, 19 qit'a).
4. Mubayyin
5. Muhtasar (4 jild)
6. Korpus bo'yicha qidiruv
7. Zahiriddin Muhammad Bobur mualliflik korpusi haqida

Muallif hayoti va ijodiy faoliyati

Zahiriddin Muhammad Bobur muallif korpusi tarkibini tuzishda, avvalo, yozuvchining ijodiy merosini o'rganish, yozuvchi ijodi o'ziga xos xususiyatlarini aniqlash talab etiladi. Korpusni birinchi bo'limi muallif hayoti va ijodiy faoliyati deb nomlanadi. Bu bo'limda muallif hayoti bilan va ijodiy faoliyati bog'liq bo'lgan eng muhim ma'lumotlar beriladi. Bobur hayotining ko'p qismini jang-u jadallarda o'tkazgan bo'lsa-da, u yaratgan asarlar ilm-fanning bir qancha yo'nalishlarida, xususan, adabiyot, tilshunoslik, tarix, falsafa, fiqh, tabiatshunoslik, madaniyat va boshqa sohalarga oid ijodiy meros qoldirgan. Zahiriddin Muhammad Bobur daho ijodkorlardan biridir. Bobur ijodiy merosi hazin ohanglarga yo'g'rilgan, ohu armon to'la gamgin misralar, hasbi hol xarakteriga ega bo'lgan she'riyat, “Boburnoma” dek shoh memuar, aruz ilmidan saboq beruvchi “Muhtasar”, soliq ishlariga oid, o'g'li Humoyunga atab yozilgan “Mubayyin” hamda “Volidiya” tarjima asarlaridan iborat. Harbiy tizimga oid “Harb ishi”, musiqa borasidagi “Musiqa ilmi” asarlari ham Bobur yaratgan tengsiz xazinani boyitishi shubhasiz [Xolmonova, 2021:42]. Bundan ko'rinadiki, Bobur nafaqat davlat rahbari, siyosatchi, balki o'zidan boy, adabiy meros qoldirgan yetuk ijodkor hamdir.

Mualliflik korpusining tuzilishini belgilab olish uchun, avvalo ijodkor asarlarning umumiy hajmi, janrlar bo'yicha tasnifini o'rganish talab etiladi. Ma'lumki, Bobur ijodi, asosan lirik asarlardan iborat. Manbalarda keltirilishicha, shoir she'rlarining umumiy hajmi 400 dan ortishi va sakkizta she'riy janrda ijod qilganligi aytiladi [Захириддин Мухаммад Бобур энциклопедияси. -Тошкент, 2014:122]. Shundan g'azallar – 126 ta (3 tasi tugallanmagan), masnaviy – 12 ta, qit'a – 23 ta, tuyuq – 20 ta, fard – 117 ta, muammo – 54 ta, ruboiy – 230 ta va bir nechta masnu she'rlardan iborat. Bobur she'riyatining ushbu ko'rsatgichlari ko'pgina adabiyotlarda 119 ta g'azal, 321 ta ruboiy tarzida uchraydi [Захириддин Мухаммад Бобур, 2007:4]. Bobur ijodiy merosi turli manbalarda turlicha

ko'rsatilgan. Bobur shaxsi va ijodi bo'yicha ilmiy izlanishlar olib borgan xorijiy olimlarning tadqiqotlarida bu ko'rsatgichlar ham farqli. Bunga sabablardan biri shoir asarlari qo'lyozma nusxalarining ko'pligidir. Xususan, birgina “Bobur devoni”ning hozirgi kunda 10 ga yaqin qo'lyozma nusxasi mavjud.

“Boburnoma” asari

Mualliflik korpusining ikkinchi bo'limi “Boburnoma” asari deb nomlanadi. Bobur ijodining eng nodir adabiy manbasi hisoblanadi. “Boburnoma”da O'rta Osiyo, Afg'oniston, Hindiston xalqlarining tarixiy, ijtimoiy-siyosiy hayoti, madaniyati, turmush tarzi, geografiyasi, etnografiyasi, o'simliklar va hayvonot dunyosi to'g'risida ko'plab ma'lumotlar keltirilgan. Asar o'z xarakteri bilan nafaqat turkiy tildagi, balki Sharq madaniyati tarixida ham ajoyib yangilik va nodir adabiy manba bo'lib xizmat qilmoqda. XVI asrdayoq bir necha marta fors tiliga tarjima qilindi, keyinchalik Yevropa tillariga ham tarjima qilinib, batafsil o'rganildi.

“Boburnoma” nafaqat tarixiy asar, balki unda o'sha davrdagi bir qator fanlar-jug'rofiya, nabotot, etnografiya, harbiy usul va boshqalar haqidagi bilimlar ham o'z ifodasini topgan. Bobur tog' va dalalar, ko'l-daryolar, xilma-xil o'simliklar, har xil hayvonlar, turli hudud va mintaqalarning yer osti va yer usti boyliklari, u yerda yashaydigan xalqlarlarning urf-odati, tili, adabiyoti, san'ati, ayrim so'zlarning kelib chiqishi kabilarga ham batafsil to'xtaladi.

“Boburnoma” matnini aniqlash va tadqiq etishda R.M.Kaldekot, F.G.Talbot, A.Denison-Ross, A.S.Beverij, A.Shimmel, N.N.Pantusov, V.V.Vyatkin, N.I.Veselovskiy, V.V.Bartold, A.N.Samoylovich, I.V.Stebleva, G.F.Blagova, S.Azimjonova, V.Zohidov, A.Qayumov, N.Otajonov, A.Abdug'affurov, B.Qosimov, H.Boltaboyev, S.Shukrullayeva, H.Hasanov singari tarixchi, faylasuf, filolog va tabiatshunos olimlarning xizmati katta [G'aybulloh, 1996:5]. “Boburnoma” badiiyati, uslubiy xususiyatlariga doir tadqiqotlar olib borildi, maqolalar e'lon qilindi [Nazarova, 1971:39]. “Boburnoma” adabiy manbashunoslik va matnshunoslik nuqtayi nazaridan tadqiq etildi. Asarning o'zbek tilidagi nashrlari, ingliz tilidagi tarjimalari, tanqidiy matnning yapon nashri qo'lyozma manbalar bilan qiyoslandi, “Boburnoma”ning ilmiy-tanqidiy matnini yaratishdagi nazariy konsepsiyalar belgilandi [Hoshimova, 2006:22]. A.Qayumovning shoir lirikasiga bag'ishlangan tadqiqotlari, S.Azimjonovanning tarixiy yo'nalishdagi “Umar Shayx va Boburning Farg'ona ulusi” mavzusidagi nomzodlik (1949), “Boburning Kobul va Hindistondagi davlati (1969) mavzusidagi doktorlik dissertatsiyalari ham mavjud. “Boburnoma”ning leksik xususiyatlariga bag'ishlangan qator tadqiqotlar bo'lib, jumladan H.Dadaboyevning “Boburnoma”da qo'llangan ijtimoiy-siyosiy, harbiy terminlarni tadqiq qildi. N.Jiyanova numerativlar haqidagi [Jiyanova, 2001], A.Ibrohimov asardagi hindcha so'zlar tahliliga doir [Ibrahimov, 2000] nomzodlik, Bobur asarlaridagi o'zlashma leksika tadqiqiga bag'ishlangan doktorlik dissertatsiyasini himoya qildi.

A.Ibrohimov “Boburnoma”da takrorlarsiz 32 100 dan ortiq so'zshakl va birikma (Boburning jami asarlarida 46 200 dan ortiq so'zshakl va birikma)

ishlatilganligi haqida ma'lumot bergan. “Boburnoma” leksikasida aks etgan nomlar statistikasi haqida ham xilma-xil ma'lumotlar mavjud. Jumladan, M.Hoshim “Boburnoma”da 1605 nafar shaxs, 1151 ta hayvon va o'simliklar, 40 ta kitob va risola nomi, 18 ta maqol va hikmatli so'z hamda 94 ta she'riy parcha keltirilganligini yozadi [Hoshim, 2003:24]. “Boburnoma” da jami 14 728 ta so'z qo'llangan. Shundan 4007 tasi (asar leksikasining 27,2%) ni nomlar tashkil qiladi, jumladan: 1507 ta antroponim, 1253 ta toponim, 248 ta etnonim, 175 ta zoonim, 247 ta fitonim, 37 ta kitob va risola, 4 ta kema nomi. 1925 tasi (13,1%) arab fors-tojik, hind pushtu tiliga faol qo'llangan leksik birliklardir. 700 tasi (4,75%) arxaik birliklar, 195 tasi (1,31%) okkazional so'zlardan iborat [Xolmonova, 2021:74].

	Asardagi nomlar	Soni
1	Antroponim	1507
2	Toponim	1253
3	Etnonim	248
4	Zoonim	175
5	Fitonim	247
6	Kitob va risola	37
7	Kema	4

Zahiriddin Muhammad Boburning “Boburnoma” asari tilini o'rganish borasida jahon miqyosida bir qator izlanishlar amalga oshirilgan. G'.Salomov, N. Otajonov, S.Hasanov, M.Xolbekov, L.Xo'jayeva, G'.Xo'jayev, S.Shukurullayeva, F.Salimova, A.Ibrohimov, R.Karimov, M.Sobirov, Z.Xolmonova, R.Shirnova, Sh.Xayitov, Z. Teshaboyeva kabi olimlarining tadqiqotlarini qayd etish lozim.

Bobur devoni

Boburshunoslik tarixida shu kunga qadar “Bobur devoni” qo'lyozmalari asosida bir qator noshirlik ishlari bajarilgan bo'lib, ular D.Ross (1910), F.Ko'prulu (1915), A.Samaylovich (1917), S.Azimjonova va A.Qayumov (1965), S.Hasanov (1982), Sh.Yorqin (1983, 2008), A.Abdug'afforov (1994), B.Yujel (1995), H. Muhammadxo'jayev (1996), E.Ochilov (2007), V.Rahmonov (2008), R.Kurinbayev (2011) A.A'zamov (2014) tomonidan amalga oshirilgan. Shubhasiz, ushbu nashrlar orasida turkcha nashr o'zining bir qancha yutuqlari bilan e'tiborga loyiqdir. Turk olimi Yujel Bilol tomonidan mumtoz devonchilik tarzida tuzilgan “Bobur devoni” nashri ham mavjud.

“Bobur devoni” ning bir qancha nusxalari mavjud. Xususan, Turkiya nusxasi Istanbul universiteti kutubxonasida 3743 raqami ostida saqlanayotgan nusxadir. Qo'lyozma “Devoni Bobur” Chig'atoy” so'zlari bilan boshlanadi. Devonning birinchi bo'limiga “Risolayi volidiya” tarjimasi joylashtirilgan. Ikkinchi qismida 119 ta g'azal, 8 ta masnaviy, 3 ta ruboiy, bir musarra' bayt, 202 ruboiy, 53 ta

muammo (6 ruboiy, 6 qit'a, 41 musarra' bayt), 18 ta qit'a, 15 ta tuyuq, 2 ta musarra' bayt, 79 matla', 7 ta masnu she'r (2 g'azal, 5 ruboiy), 2 musarra' bayt, 4 ta fard, 1 ta qit'a, 3 musarra bayt, 1 masnaviy, 8 ta tugallanmagan g'azal, 1 nazm, 1 mansur parcha, bulardan tashqari fors tilida 1 mansur parcha 2 g'azal, 17 matla', 11 ruboiy, 7 qit'a va aruzga doir ikki sahifalik ilmiy parcha [Shimmel, 1996:514] dan iborat. Alisher Navoiy nomidagi Adabiyot muzeyi fondida 67 raqam ostida saqlanayotgan ushbu qo'lyozma To'xtamurod Usmon o'g'li Zafariy tomonidan Parij, Rampur qo'lyozma nusxalari va F. Ko'prulu nashri asosida ko'chirilgan. Ushbu nusxada 116 g'azal, 196 ruboiy, 8 masnaviy, 52 muammo, 19 qit'a keltirilgan.

Bobur devoni nusxalari bo'yicha janrlar tasnifi	Turkiya nuxasi	R nuxasi	Topqapu nuxasi	Muallim Javdat nuxasi	Rampur nuxasi	Bobur devoni tarjimasi
Risolayi volidiya tarjimasi						
G'azallar soni	121 ta	87 ta	113 ta	113 ta	1 ta	121 ta
Masnaviylar soni	18 ta	2 ta	12 ta	12 ta	2 ta	19 ta
Ruboiylar soni	216 ta	65 ta	139 ta	138 ta	24 ta	222 ta
Muammolar soni	53 ta	28 ta	57 ta	57 ta	-	57 ta
Qit'alar soni	26 ta	1 ta	14 ta	18 ta	3 ta	27 ta
Tuyuqlar soni	15 ta	9 ta	15 ta	13 ta	-	15 ta
Matla'lar soni	87 ta	15 ta	78 ta	78 ta	-	96 ta
Masnu she'r soni	7 ta	3 ta	7 ta	7 ta	-	7 ta
Tugallanmagan g'azallar soni	12 ta	2 ta	8 ta	8 ta	1 ta	16 ta
Nazmlar soni	2 ta	1 ta	1 ta	1 ta	1 ta	3 ta
Musarra baytlar soni	14 ta	5 ta	8 ta	8 ta	-	16 ta
Mufridlar soni	5 ta	1 ta	4 ta	4 ta	7 ta	5 ta
Mansur parcha	3 ta	-	3 ta	3 ta	-	5 ta
Nusxadagi umumiy she'rlar	593 ta	219 ta	454 ta	431		614 ta

Bobur devonining qo'lyozma nusxalaridagi she'rlar soni va janrini ko'rsatib beruvchi ushbu jadval Y.Bilol ma'lumotlari asosida tuzilgan. Unda Boburning o'zbekcha, fors-tojikcha she'rlari soni umumiy holda ko'rsatilgan.

Mubayyin

“Mubayyin” – Zahiriddin Muhammad Boburning islom dini qonun-qoidalarini haqida batafsil yoritilgan muhim manba. Zahiriddin Muhammad Bobur nafaqat shoh va shoir, balki shariat bilimdoni bo'lganini o'zining “Mubayyin” asari orqali namoyon etgan. Bobur Mirzo “Mubayyin” da eng muhim islomiy ilmlarni saralab,

o'zining din olimi, faqih ekanini ko'rsatgan. Sovetlar davrida “Mubayyin” asari xalqdan yashirildi, faqat istiqlool yillari izohlar bilan nashr etildi. “Mubayyin” Hindiston yurishlari davrida, hijriy 928(mil.avv.1521) yili she'riy yo'lda tasnif etilgan. Asarda ahli sunnat val jamoat aqoidi ravshan bayon qilingan [<https://kh-davron.uz/kutubxona/uzbek/zahiriddin-muhammad-bobur-mubayyin-mirzo-kenjabek-bobur-mirzoning-mubayyin-asari.html>].

Bobur bu asarni o'g'illari Humoyun Mirzo va Kamron Mirzolarga bag'ishlab yozgan. “Mubayyin” asari “Hamd” – Alloh taologa hamdu sano, “Na't” – Payg'ambarimizning (sollolohu alayhi va sallam) ta'riflari, va quyidagi 5 qismdan iborat. Bular: “E'tiqodiyya”, “Kitob us-salot”, “Kitob uz-zakot”, “Kitob us-savm”, “Kitob ul-haj”.

Muhtasar (4 jild)

“Muxtasar” – Zahiriddin Muhammad Boburning aruz vazni nazariyasi haqidagi asari (taxminan 1525-1523). Manbalarda “Mufasssal”, “Risolayi aruz” deb ham yuritiladi. Unda aruz vaznidagi barcha yutuq va kamchiliklar ko'rsatilgan. Bobur aruzning ba'zi noaniq masalalariga aniqlik kiritdi va uni yangi fikrlar bilan boyitdi, o'zbek, arab, fors aruzlarini bir-biri bilan qiyoslab, aruz vaznining 248 turini aniqladi.

Bobur aruzda zihoflar sonini 46 taga yetkazdi. U taqte' bobida eski o'zbek tili fonetik me'yorlari asosida Navoiyning “Mezon ul-avzon”ida ishlab chiqilgan aruzning qonun-qoidalarini tizimga solib, ba'zi bahrlar orasida o'xshashliklar topdi. Asarda qo'shiqlarning tarxoniy, turkiy, o'lang, shuningdek, tuyuq va uning rang-barang turlari haqida yangi fikrlar aytilgan. Bobur bu asari bilan o'zbek she'riyatida qo'llanadigan aruz bahrlarining ko'p yangi vazn, zihof va far'larini aniqladi. “Mezon ul-avzon”da ruknlar to'liq berilgan. Ularning sonini aniqlashda talaffuz hisobga olingan. “Mezon ul-avzon” vaznlarning ko'pligi va ularga berilgan izohlarning aniqligi bilan boshqa asarlardan ajralib turadi. Risolaning kotib Hoji Muhammad Samarqandiy ko'chirgan (1533—1534) yagona qo'lyozmasi Parij Milliy kutubxonasida saqlanmoqda [<https://uz.wikipedia.org/wiki/Muxtasar>]. “Muxtasar” asari yirik 4 jildidan iborat. Ushbu asar aruz ilmi bo'yicha muhim manbalar sirasiga kiradi. Unda muallif aruz ilmi qonun-qoidalarini batafsil yoritgan.

Korpus bo'yicha qidiruv

Korpusdagi eng muhim xususiyatlardan biri qidiruv tizimi hisoblanadi. Shu sababli qidiruv tizimi qanchalik mukammal tuzilgan bo'lsa, korpus ham shunchalik sifatli bo'ladi. Korpusda qidiruv tizimining bir necha shakllari mavjud. Birinchisi so'zshaklida, ikkinchisi lemma shaklidagi qidiruv tizimi hisoblanadi. Ushbu mualliflik korpusida esa qidiruv so'zshaklida amalga oshiriladi. Shu bilan birga mualliflik korpusi xususiyatlaridan kelib chiqib, korpus qidiruvida janr bo'yicha ham qidiruvni amalga oshirish mumkin. Bu yanada muallif asarlarida qidiruvni amalga oshirishda qulaylik yaratadi. Korpus menejerining mukammal ishlab chiqilganligi korpus qidiruv tizimi imkoniyatini yanada oshirishga xizmat qiladi.

Zahiriddin Muhammad Bobur mualliflik korpusi haqida

Mualliflik korpusi ma'lum bir shoir va yozuvchining ijodiy merosi elektron shaklda jamlangan axborot kutubxonasi. Korpusning nafaqat, ta'lim tizimida balki, ko'plab sohalarda ham ahamiyati ortib bormoqda. Hozirgi raqamli texnologiyalar asrida hayotimizni korpuslarsiz tasavvur qilish qiyin. Mualliflik korpusiga kiritilgan asarlar uzoq yillar davomida saqlanishiga yordam beradi. Ilmiy tadqiqotlarda ham korpusdagi ma'lumotlarning aniqligi va reprezentativlik xususiyati tadqiqotchining vaqtini tejaydi va unga amaliy yordamchi vazifasini bajaradi. Ma'lum bir asar bo'yicha statistik ma'lumotlarga ham oson ega bo'ladi. Shuningdek, muallif korpusi yordamida shoir ijodi chastotasini bilish mumkin.

Masalan, U. Shekspir muallif korpusi yordamida “war” (urush) so'zida qidiruv amalga oshirilganda quyidagicha natija olindi: 42 ta asardagi 356 ta nutqda 261 marta uchraydi. Bir vaqtning o'zida berilgan barcha misollarni ko'rish mumkin. Bu ham tadqiqotchi uchun tayyor statistik ma'lumotlardan foydalanish imkoniyatini yaratadi. Korpus yordamida lug'at tarkibini ham kuzatish, yozuvchi qo'llagan so'zning hozirgi kundagi muqobil varianti bilan qiyoslash imkoniyatini beradi.

O'zbek korpus lingvistikasini targ'ib etish hamda rivojlantirishga, tadqiqot ishlarini yuksaltirishga, barcha yoshdagi mutaxassislarining ta'limiy ehtiyojini qondirishga, so'z boyligi hamda savodxonlikni oshirishga qaratilgan zamonaviy mualliflik korpusi yaratilish zarur. Hozirgi kunda jahon kompyuter lingvistikasida ko'plab korpuslar, xususan mualliflik korpuslari tizimli shaklda yaratilgan. O'zbek tilshunosligida bugungi kunga qadar A. Navoiy va A. Qahhor mualliflik korpusining lingvistik bazasi yaratilgan bo'lsa-da, dasturiy ta'minoti ishlab chiqilmagan. Shu sababli ham mualliflik korpusini ham lingvistik, ham dasturiy ta'minotini ishlab chiqish dolzarb masalalardan biri hisoblanadi. Ma'lum yozuvchi asarlari uchun mualliflik korpusini tuzishda muallif ijodiy merosini mukammal o'rganish zarur. Negaki, korpus tarkibi, bo'limlar soni muallif ijodining o'ziga xos xususiyatidan kelib chiqqan holda tuziladi.

Foydalanilgan adabiyotlar:

1. G'aybulloh A., Ne'matulloh O., Jahongashta “Boburnoma”. – Toshkent: Xalq merosi, 1996.
2. Hoshim M. Besh yuz yillik qomus // Tafakkur. 2003.
3. Hoshimova D. “Boburnoma” asarining tanqidiy matnini yaratish muammolari. Filol. Fan. Nomz... diss. – Toshkent, 2006.
4. Ibrohimov A. “Boburnoma”dagi hindcha so'zlar tadqiqi. Filol.fan.nomz... diss. – Toshkent, 2000.
5. Jiyanova N. “Boburnoma”dagi numerativlarning ma'no xususiyatlari. Filol. fan.nomz...diss. – Toshkent, 2001.
6. Nazarova X. Bobur va o'zbek adabiy tili. – Toshkent, 1971.
7. Xolmonova Z. “Boburnoma”- til qomusi. – Toshkent: Akademnashr, 2021.



8. А.Шиммелнинг «Бобуршоҳ ва оилавий истеъдод ҳақида». Инглизчадан Н.Отажонов таржимаси // Адабиёт қўзғуси, – Тошкент, 1996.
9. Бу ҳақда қаранг: Заҳириддин Муҳаммад Бобур энциклопедияси. Тошкент, 2014.
10. <https://uz.wikipedia.org/wiki/Muxtasar>
11. <https://kh-davron.uz/kutubxona/uzbek/zahiriddin-muhammad-bobur-mubayyin-mirzo-kenjabek-bobur-mirzoning-mubayyin-asari.html>.

DIALEKTAL KORPUSNING ARXITEKTURASI VA MA'LUMOTLAR BAZASI

Muftillayeva Ruxsora Toshmuhammad qizi

ruxsoramuftillayeva851@gmail.com

ToshDO'TAU magistranti

Annotatsiya. Dialektal korpus bu tilning turli mintaqaviy yoki ijtimoiy turlarini ifodalovchi matnlar va og'zaki til namunalari to'plami. U lingvistik tadqiqotlar, til hujjatlari va madaniy merosni saqlash uchun muhim manba hisoblanadi. Dialektal korpusni boshqarish va tahlil qilish uchun yaxshi ishlangan arxitektura, ma'lumotlar bazasi tizimi zarur. Dialektal yozuvlari to'plami keng ko'lamli fonologik, morfologik, leksik va sintaktik tadqiqotlar uchun ham qimmatli ma'lumotlar manbayi hisoblanadi. Ushbu maqolada biz dialektal korpus uchun arxitektura va ma'lumotlar bazasi talablarini o'rganamiz. Bu xususiyatlar quyidagilarni o'z ichiga oladi: ma'lumot to'plash, ma'lumotlarni qayta ishlash, saqlash, qidiruv tizimi, xavfsizlik va maxfiylik, ma'lumot beruvchi tafsilotlarini ko'rsatish (so'zlar soni va axborot beruvchilar haqidagi boshqa ma'lumotlar).

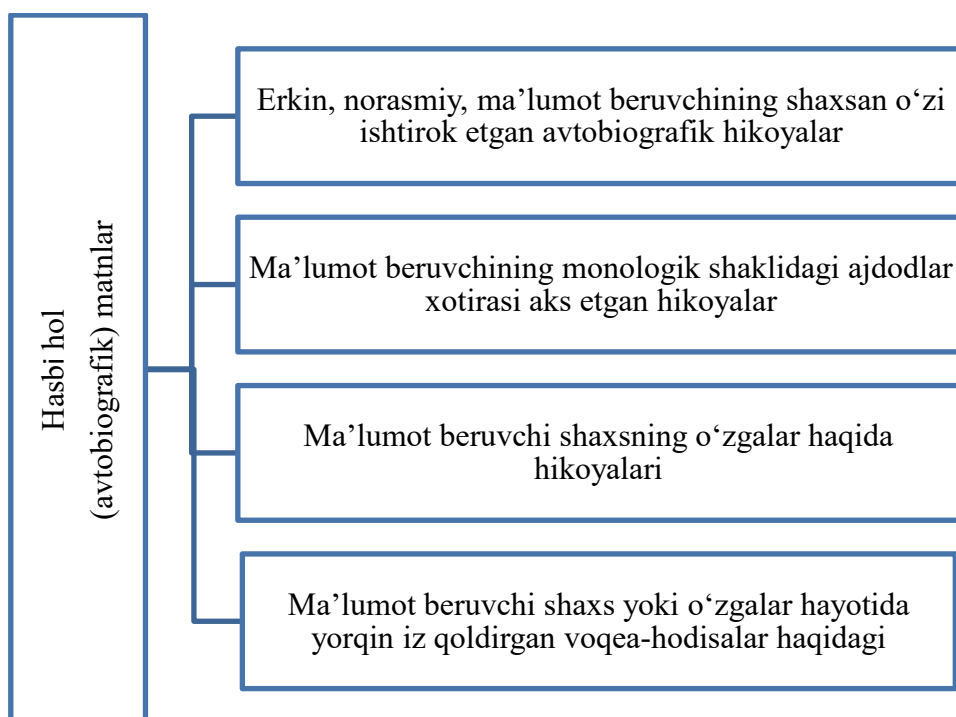
Abstract. A dialectal corpus is a collection of texts and spoken language samples representing different regional or social varieties of a language. It is an important resource for linguistic research, language documentation and cultural heritage preservation. A well-designed architecture and database system is necessary for the management and analysis of the dialectal corpus. The collection of dialect records is also a valuable source of information for large-scale phonological, morphological, lexical and syntactic studies. In this article, we explore the architecture and database requirements for a dialectal corpus. These features include: data collection, data processing, storage, search engine, security and privacy, disclosure of informant details (word count and other information about informants 'data).

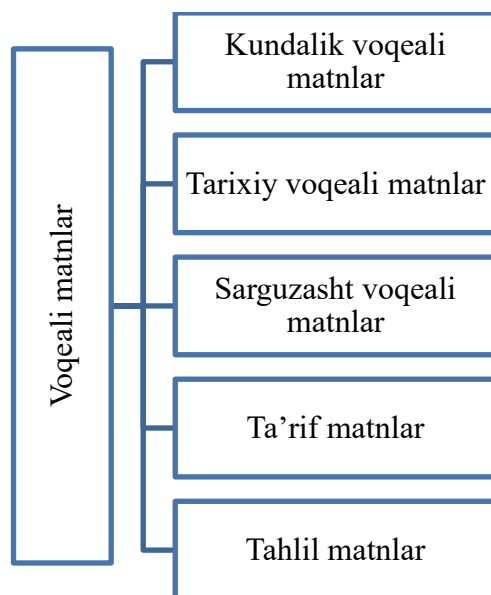
Аннотация. Диалектный корпус — это совокупность текстов и образцов разговорного языка, представляющих различные региональные или социальные разновидности языка. Это важный ресурс для лингвистических исследований, языковой документации и сохранения культурного наследия. Для управления и анализа диалектного корпуса необходимы хорошо продуманная архитектура и система баз данных. Коллекция диалектных записей является также ценным источником информации для масштабных фонологических, морфологических, лексических и синтаксических исследований. В этой статье мы исследуем требования к архитектуре и базе данных для диалектного корпуса. Эти функции включают в себя: сбор данных, обработку данных, хранение, поисковую систему, безопасность и конфиденциальность, раскрытие данных информанта (количество слов и другая информация о данных информаторов).

Kalit soʻzlar: *dialektal korpus, transkripsiya, dialekt, integratsiya, qidiruv tizimi.*

Dialektal korpus tizimining arxitekturasini dialektal ma'lumotlarni saqlash, olish, tahlil qilish uchun birgalikda ishlaydigan turli komponentlar va qatlamlarni o'z ichiga oladi. Shevalar korpusining texnik tomoniga ishlov berish uchun juda katta mablag' va ilmiy salohiyatga ega guruh shakllantirish eng asosiy vazifa. Dialektal korpus yaratish jarayoni uzoq va qiyin mehnatni talab etishini hisobga olgan holda korpusning dasturiy ta'minotini ishlab chiquvchi dasturchi texniklar guruhini shakllantirish lozim bo'ladi. Birinchi navbatda korpusdan foydalanish qulay va ommaviy bo'lishini ta'minlash kerak. Dialektal korpus arxitekturasining ba'zi asosiy komponentlari:

Ma'lumot to'plash. Dialektal korpusni qurishda birinchi qadam turli dialekt mintaqalari yoki hududlardan ma'lumotlarni yig'ishdir. Bunda asosan intervyu manbalardan foydalaniladi. Qishloq aholisi o'rtasidagi muloqatlar yozib olinadi. Intervyular dialog yoki polilog shaklida bo'ladi. Respondentlarning asosiy qismi 55 yoshdan oshgan odamlar. Tadqiqot uchun ataylab o'rta ma'lumotga ega respondentlar tanlanadi, chunki aynan ularning nutqida dialektal xususiyatlar yaxshiroq saqlangan bo'ladi. Har bir matn uchun “pasport” ma'lumotlari yozuvning yili, joyini va respondent haqida ma'lumotlarni o'z ichiga oladi. Yozilgan dialekt ma'lumotlarining mavjudligi lingvistik va tarixiy tadqiqotlar uchun bebaho manba bo'lib xizmat qiladi. Ma'lumot to'plash suhbat shaklida o'tkaziladi.





1.1-jadval. Sheva suhbat janrining ichki qismi

Dialektal korpus tuzishda ishtirok etgan respondentlar haqida ma'lumotlar beriladi.

T /p	So'rovnoma	Javoblar
1	Yozib olingan joy (viloyat, tuman)	Yangihayot mahallasi, Nurota tumani, Navoiy viloyati
2	Yozib olingan vaqt (yil, kun, oy)	2023- yil 23- avgust
3	Axborot beruvchi (familiya, ism, otasining ismi)	Muftillayeva Ruxsora Toshmuhammad qizi
4	Axborot beruvchining jinsi (ayol, erkak)	Ayol
5	Axborot beruvchining tug'ilgan joyi va yili	Nurota tumani, Yangihayot mahallasi
6	Axborot beruvchining ma'lumoti (o'rta, oliy)	O'rta ma'lumotli
7	Axborot beruvchining kasbi	Uy bekasi
8	Ma'lumot yig'uvchi (familiya, ism, otasining ismi)	Muftillayeva Ruxsora Toshmuhammad qizi
9	Ish olib borilayotgan muassasa	Toshkent davlat o'zbek tili va adabiyoti unversiteti

10	Matn shakli	Og‘zaki
11	Janr	Suhbat
12	Yozuv turi (qo‘lyozma, audio, kamera, diktafon, fotoapparat)	Diktafon, kamera
13	Mavzu	Dehqonchilik
14	Voqealik davri Mustaqillikka qadar Mustaqillikdan keyin	Mustaqillikdan keyin

1.2-jadval. Shevaning pasport metarazmetkasi

Suhbat matnining transkripsiyasi. Yuqoridagi so‘rovlar asosida qidiruv o‘tkazilsa, quyidagicha natijaga erishiladi: Dehqonchilik bilan bog‘liq suhbat. Matn haqidagi ma‘lumotlar aks etgan matn metarazmetkasi.

Bu yozuv (diktafon, kamera) 2023-yil 23-avgustda Yangihayot mahallasida dehqonchilik bilan shug‘ullangan Muftillayeva Ruxsora Toshmuhammad qizidan yozib olindi; nutq maqsadi: o‘tgan kunlarni eslash, kasbiy bilimlar berish orqali shevaga xos so‘zlarni aniqlash va tadqiq etish; nutq jarayonida tilga olinadigan shaxslar; turmush o‘rtog‘i, qaynota, onasi, dadasi; voqea-hodisaning mustaqillikdan oldin yoki keyin ro‘y berganligi; voqea-hodisaning mustaqillikdan oldin 1954-yillarda, kelin bo‘lganidan keyin dehqonchilik bilan bog‘liq voqealar. [Xolova, 2022:25]

Ma‘lumotlarni oldindan qayta ishlash. Ma‘lumotlar to‘planganidan keyin, izchillik va sifatni ta‘minlash uchun uni oldindan qayta ishlash kerak. Bu dialektal matnlar yoki nutq namunalarini transkripsiya qilish, normallashtirish, izohlash va moslashtirishni o‘z ichiga olishi mumkin. Ma‘lumotlarni oldindan qayta ishlashda qo‘llaniladigan ba‘zi umumiy usullar mavjud:

Ma‘lumotni tozalash: to‘plangan ma‘lumotlardagi takroriy yozuvlar, chetga chiqish kabi xatolardan tozalash yoki tuzatish.

Ma‘lumotlarni o‘zgartirish: ma‘lumotlardan erkin foydalanish va tahlil qilish uchun ularni mos formatga aylantirish.

Ma‘lumotlar hajmini qisqartirish: O‘lchamlarni kichraytirish yoki ma‘lumotlarning muhim xususiyatlarini saqlab qolgan holda ma‘lumotlar hajmini kamaytirish.

Ma‘lumotlar integratsiyasi: bir nechta manbalar va usullar orqali olingan ma‘lumotlarni tahlil qilish uchun ularni yagona ma‘lumotlar to‘plamiga birlashtirish. Ma‘lumotlarni oldindan qayta ishlashni amalga oshirish orqali tahlilchilar va ma‘lumot oluvchilar ma‘lumotlarning toza, izchil, tahlilga tayyor bo‘lishini ta‘minlashi mumkin, bu esa kompyuter, raqamli texnologiyalarga asoslangan ilovalarda aniq va ishonchli natijalarga olib keladi.

Saqlash. Ma'lumotlar kelajakda foydalanish uchun raqamli texnologiyada saqlash tushuniladi. Dialektal ma'lumotlarning katta hajmini saqlash uchun ishonchli va kengaytiriladigan saqlash tizimi talab qilinadi. Barcha natijalar saqlanishi yoki yuklab olinishi mumkin.¹¹ Saqlash orqali natijalarni qayta ko'rish imkoniyatiga ega bo'lamiz. Boshqa tomondan yuklab olingan natijalar fayl yoki matn ko'rinishida import qilinishi mumkin. Bundan tashqari saqlash tizimi kompyuter tizimida hal qiluvchi rol o'ynaydi, chunki u turli maqsadlar uchun ma'lumotlarni saqlash va olish imkonini beradi, masalan, dastur ma'lumotlari, fayllar, ma'lumotlar bazalari va tizimning zaxira nusxalari. Saqlashni samarli boshqarish ma'lumotlar mavjudligi, ishonchliligi va xavfsizligini ta'minlash uchun zarur.

Qidirish tizimi. Dialektal korpus kontekstida qidirish va izlash ma'lum bir tadqiqot savoliga yoki tahliliga tegishli bo'lgan ma'lum lingvistik ma'lumotlarni yoki korpus ichidagi ma'lumotlarni topish va ularga kirish jarayonini anglatadi. Dialektal korpus bu dialektlardagi yoki mintaqaviy tillardan foydalanishdagi o'zgarishlarni aks ettiruvchi til ma'lumotlari to'plamidir¹². Tadqiqotchilar ko'pincha o'z tadqiqotlari uchun ushbu korpus ichidagi o'ziga xos lingvistik xususiyatlarni izlashlari va olishlari kerak. Tadqiqotchilar dialektal korpus ichida o'rganishga qiziqqan muayyan lingvistik xususiyatlar, atamalar yoki shevalar asosida qidiruv so'rovlarini tuzadi. Ushbu so'rovlar kerakli til hodisalarni qamrab oluvchi kalit so'zlar, iboralar yoki muntazam iboralar yordamida tuzilishi mumkin. Tadqiqotchilar dialektal korpusga kirish va o'z so'rovlarini bajarish uchun qidiruv vositalari yoki dasturiy ilovalardan foydalanadilar. Qidiruv vositasi korpus ma'lumotlarni skanerlaydi va belgilangan mezonlarga mos keladigan tegishli lingvistik misollarni oladi. Tadqiqotchilar olingan lingvistik ma'lumotlarni to'g'riligini baholash uchun qidiruv natijalarini ko'rib chiqadi. Bu ma'lumotlar matndan parchalar, audio yozuvlar, transkripsiyalar yoki dialektal o'zgarishlarni aks ettiruvchi boshqa lingvistik izohlarni o'z ichiga olishi mumkin. Tadqiqotchilar olingan lingvistik ma'lumotlarni tahlil qilib, korpus ichidagi dialektal o'zgarishlar yoki tildan foydalanish bilan bog'liq xususiyatlarni, tushunchalarni aniqlashadi. Ular ma'lumotlarni sharhlash va tadqiqot maqsadlari uchun xulosalar chiqarish sifat yoki miqdoriy usullardan foydalanishlari mumkin. Foydalanuvchilar uchun qulay qidiruv interfeysi tadqiqotchilarning tadqiqot qiziqishlari asosida aniq dialektal ma'lumotlarga kirishlari uchun juda muhim. Korpus tarixchilar, etnologlar, filolog olimlar uchun emas balki oddiy odamlar uchun ham ochiq bo'ladi, shu sababli, matnlar fonetik alifbolarni yoki mahalliy dialektlarning barcha o'ziga xos xususiyatlarini yaxshi bilmaydiganlar uchun ham tushunarli bo'lishi kerak. So'rovnoman quyidagicha amalga oshirish maqsadga muvofiqdir. Korpus

¹¹ <https://www.helsinki.fi/varieng/CoRD/corpora/Dialects/fieldwork.html>

¹² <http://www.tekstlab.uio.no/scandiasyn/index.html>

kengaytirilgan qidiruv tizimi, natijalarni o'chirish, izohlash va saqlash funksiyasiga ega bo'ladi.

- 1) aniq dialektal so'z kiritish orqali;
- 2) gap oxirida, boshida va istalgan qismida kelgan dialektal so'zni kiritish orqali amalga oshiriladi.

Bir so'zni turli xil yo'l bilan qidirish bazadan keng foydalanish imkoniyatini beradi va xulosaning aniqligi uchun xizmat qiladi. Bazaga kiritilgan shevaga oid ma'lumotlar bazasidagi lingvistik tahlillar bir necha bosqich va prinsiplarga asoslangan holda tuziladi. Dastlabki shevalar bazasi quyidagi talablar asosida quriladi.

Qidiruv tizimi parametrlari	Misol
1. Shevaga oid so'z	Jang'i
2. Adabiy til	Yangi
3. Izoh	Ilgari bo'lmagan, endi yuzaga kelgan
4. Turkum	Sifat/ravish
5. Omonim	Yangi - ravish (yaqindagina, hozirgina) Yangi - sifat
6. Sinonim	Mavjud emas
7. Antonim	Eski

1.3-jadval. Korpusning qidiruv turlari

Bazada qidiruv tizimi ayni shu 7 maydonda amalga oshiriladi. Shevalar bazasida so'zlarning turkumlar bo'yicha statistikasi beriladi. Bunda bazaga kiritilgan jami kiritilgan so'zlarning tashqi va ichki statistikasi keltiriladi. Tashqi statistikada jami so'zlar soni, ya'ni jami mustaqil so'z turkumi, yordamchi so'z turkumi va alohida olingan so'zlar soni berilishi. Ichki statistikada turkumlarning xususiyatiga e'tibor beriladi. Unga ko'ra jami shevaga oid so'zlarning turkumga oidligi qayd etiladi.

T/R	Turkum	So'zlar soni
1.	Ot	323
2.	Sifat	48
3.	Son	12
4.	Fe'l	101
5.	Olmosh	4
6.	Ravish	6
7.	Ko'makchi	6

8.	Bog'lovchi	7
9.	Yuklama	9
10.	Modal	2
11.	Undov	8
12.	Taqlid	9

1.4- jadval. Turkumlar statistikasi

Moslashuvchanlik. Moslashuvchan dialektal korpus tadqiqotchilarga o'zlarining tahlil usullarini, tadqiqot savollarini va ma'lumotlarni o'rganish usullarini o'zlarining aniq tadqiqot maqsadlari va qiziqishlariga mos ravishda sozlash va moslashtirish qobiliyatini taklif qiladi. Moslashuvchan dialektal korpus turli dialektal o'zgarishlarni, mintaqaviy til xususiyatlarini va tildan foydalanish kontekstlarini ifodalovchi keng ko'lamli lingvistik ma'lumotlarni o'z ichiga olishi kerak. Korpus tadqiqotchilarga til o'zgarishi va turli til o'lchovlari bo'yicha o'zgarishlarni o'rganishga imkon berish uchun bir nechta dialektlardan, sotsiolektlardan, tarixiy davrlardan yoki janrlardan olingan ma'lumotlarni o'z ichiga olishi kerak. [Sirazitdinov, Buskunbayeva, Karimova, 2008: 212] Ma'lumotlar bazasi har xil turdagi ma'lumotlar jumladan, matn, audio, video tasvirlar va izohlar bilan ishlash uchun yetarlicha moslashuvchan bo'lishi kerak. Korpus bir nechta ma'lumotlar formatlarini qo'llab-quvvatlashi va boshqa vositalar yoki tizimlar bilan oson integratsiyalashuviga imkon berishi kerak.

Dialektal korpus vaqt o'tishi bilan sezilarli darajada o'sishi mumkin, shuning uchun ma'lumotlar bazasi katta hajmdagi ma'lumotlarni samarali boshqarish uchun kengaytirilishi kerak.

Xavfsizlik va maxfiylik. Dialektal korpusda shaxslar yoki jamoalar haqida nozik ma'lumotlar bo'lishi mumkin. Korpusga kirishni boshqarish mexanizmlari va shifrlashni himoya qilish uchun mustahkam xavfsizlik choralariga ega bo'lishi kerak.

Dialektal korpus ko'pincha lingvistik tahlil uchun maxsus vositalarni talab qiladi. Ushbu vositalar asosiy statistik ma'lumotdan tortib tabiiy tilni qayta ishlashning ilg'or usullarigacha bo'lishi mumkin. Bunday vositalarni korpus arxitekturasiga integratsiyalashuvi dialektal ma'lumotlarni samarali tahlil qilishga yordam beradi. Dialektal korpus tadqiqotida ishlatiladigan ba'zi umumiy tahlil vositalari quyidagilardan iborat: Konkordans tadqiqotchilarga ma'lum kalit so'zlar, iboralar asosida korpus ichida lingvistik ma'lumotlarni qidirish va tahlil qilish imkonini beruvchi dasturiy vositalar. Muvofiqlashtiruvchilar kontekstda qidiruv so'zlari misollarni aks ettiruvchi muvofiqlik qatorlarini taqdim etadi, bu esa tadqiqotchilarga turli dialektal kontekstlarda o'ziga xos lingvistik xususiyatlar qanday ishlatilishini tekshirishga imkon beradi.

Xulosa qilib aytganda, yaxshi ishlab chiqilgan arxitektura va ma'lumotlar bazasi tizimi dialektal korpusni samarali boshqarish va tahlil qilish uchun muhimdir. Arxitektura ma'lumotlarni yig'ish, oldindan qayta ishlash, saqlash, metama'lumotlarni boshqarish, qidirish, tahlil qilish vositalarini o'z ichiga olishi kerak. Ma'lumotlar bazasi moslashuvchan, kengaytirilgan, ma'lumotlar xavfsizligi va maxfiyligini ta'minlashi kerak.

Foydalanilgan adabiyotlar:

1. Холова М. ўзбек миллий шевалари корпусини тузишнинг лингвистик асослари (Бойсун тумани “ж”ловчи шевалари мисолида): Фил. фан. бўйича фалсафа доктори (PhD)...дисс. Автореф. – Термиз, 2022. -12 Б
2. <https://www.helsinki.fi/varieng/CoRD/corpora/Dialects/fieldwork.html>
3. <http://www.tekstlab.uio.no/scandiasyn/index.html>
4. Sirazitdinov Z.A., Buskunbaeva L.A., Karimova R.N. Dialectological Base of the Machine Fund of the Bashkir Language as a Tool for Research Bashkir Dialect Continuum.\\ Russian Federation Bulletin of the Kalmyk Institute for Humanities of the Russian Academy of Sciences Has been 149 issued since. 2008. Vol. №23, Is. №1. –P. 212–219. DOI: 10.22162/2075- 7794-2016-23-1-212-219

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

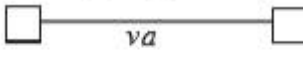
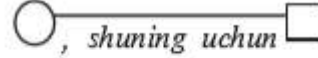
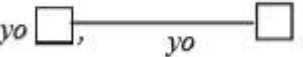
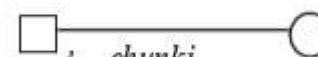
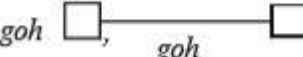
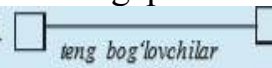
Har qanday gapning tuzilish jihatidan turlarini uning grammatik asosi belgilab beradi. Gapning grammatik asosini bosh bo'laklar: ega va kesim tashkil etadi. Gap tarkibidagi grammatik asos (ega va kesim) ning ishtirokiga ko'ra gaplar tuzilish jihatidan ikki turga bo'linadi: sodda gaplar va qo'shma gaplar. 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.

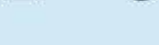
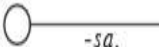
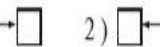
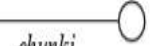
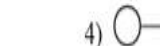
Gapning tuzilishiga ko'ra turlari jadvali:

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)

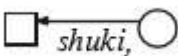
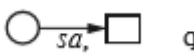
Sodda va qo'shma gap strukturalari orasidagi farqlar va ularni aniqlash uchun qoidalar

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 ifodalangan 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 <i>qo‘shma gap qismlari vergul bilan ajratiladi.</i>
Bir so‘z bilan ifodalangan 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 ifodalangan 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, iʼqibati 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 modellashtirish 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. **lda2vec:** 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-atcilikni-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

MATNGA QAYTA ISHLOV BERISH TIZIMLARI

O‘ZBEK TILI SO‘ZLARINING STEMINI TOPISH UCHUN LINGVISTIK MANBA YARATISH

Sharipov Maqsud Siddiqovich

Texnika fanlari nomzodi, dotsent,

m.sharipov@urdu.uz

UrDU, Kompyuter ilmlari kafedrası dotsenti

Amanbayeva Raxima Xushnud qizi,

rahimaamanbayeva@gmail.com,

UrDU, Kompyuter ilmlari kafedrası magistranti

Jumanazarova Onajon Matnazarovna,

anaxon54@gmail.com,

UrDU, Kompyuter ilmlari kafedrası katta o'qituvchisi

Annotatsiya. Stemlarni (soʻzning asosiy qismi yoki oʻzak soʻz) aniqlash uchun lingvistik manbalar, soʻzning tarkibiy qismi, maʼnosi, va qoʻllanilishining tushunchalarini taʼkidlaydigan, bu soʻzlarga oid maʼlumotlarni taqdim etadigan manbalardir. Bu maqola oʻzbek tili soʻzlarining tarkibini va ularning stemini aniqlash boʻyicha lingvistik manbaning yaratilish jarayonini taʼriflaydi. Maqolada tahlil va yigʻilish bosqichlari, koʻplab manbalar bilan ishlash, lugʻatlar va tahlil kitoblardan foydalanish, izlanadigan standartlarga rioya qilish, tahlil va natijalarni koʻrib chiqish, natijalarni yigʻib chiqish va yuklash bosqichlari koʻrsatilgan. Oʻzbek tili soʻzlarining stemini aniqlash uchun lingvistik manba yaratishning metodologiyasini va jarayonning muhimligini belgilaydi.

Abstract. Linguistic resources for identifying stems (the main part of a word or root word) are resources that provide information about these words, emphasizing concepts of word structure, meaning, and usage. This article describes the process of creating a linguistic resource for determining the composition of Uzbek words and their stem. The article illustrates the steps of analysis and assembly, working with multiple sources, using dictionaries and analysis books, following the standards sought, reviewing the analysis and results, collating and uploading the results. Summarizes the methodology and importance of the process of creating a linguistic resource for determining the stem of Uzbek words.

Аннотация. Лингвистические ресурсы для определения основы (основной части слова или корневого слова) — это ресурсы, которые предоставляют информацию об этих словах, подчеркивая концепции структуры, значения и использования слова. В данной статье описан процесс создания лингвистического ресурса для определения состава узбекских слов и их основы. В статье показаны этапы анализа и сборки, работа с несколькими источниками, использование словарей и книг по анализу, следование искомым стандартам, просмотр анализа и результатов, сопоставление и загрузка



результатов. Обобщены методология и значение процесса создания лингвистического ресурса для определения основы узбекских слов.

Kalit so‘zlar. *Lingvistik manba, o‘zak, stem, stemming, lug‘at, qo‘shimcha, so‘z, o‘zbek tili.*

Lingvistik manba, tilning tarkibiy elementlarini, ularning o‘zaro aloqalarini va qonuniyatlarni izlash, aniqlash va tushuntirish uchun ishlatiladigan muhim ma’lumot manbasidir. Lingvistik manbalar tilining lug‘atlarini, tarixiy manbalari, akademik tadqiqotlar va maktab darsliklarini o‘z ichiga oladi. Lingvistik manbalar, tilning aloqador so‘zlari, izohlar va ta’riflari orqali tilning o‘zini tushunishda muhim vazifalarni o‘taydi. Stem – so‘zning qo‘shimchalarsiz qismi. Masalan, “*kitob*” so‘zi esa “*kitoblar*”, “*kitobni*”, “*kitobdan*” so‘zlarining stemidir. Stemlarni (so‘zning asosiy qismi yoki o‘zak so‘z) aniqlash uchun lingvistik manbalar, so‘zning tarkibiy qismi, ma’nosi, va qo‘llanilishining tushunchalarini ta’kidlaydigan, bu so‘zlarga oid ma’lumotlarni taqdim etadigan manbalardir. Ushbu maqola o‘zbek tili so‘zlarining tarkibini, strukturasi va tarixiy rivojlanishini tahlil qilish va ularning stemini aniqlash uchun lingvistik manba yaratish jarayonini ta’kidlaydi.

Stemming uchun lingvistik manba yaratish jarayonini amalga oshirishdan oldin, avvalgi ma’lumotlar yoki oldingi bajarilgan ishlarni ko‘rib chiqish, ushbu faoliyatni boshlash va maqsadlarni belgilash uchun muhim nazariy va amaliy yordam beradi. Ushbu maqolada qo‘shimchalarni guruhlariga ajratgan holda o‘zbek tili uchun qoidaga asoslangan stemming algoritmini chekli holatlar mashinasi (ChHM) yordamida ishlab chiqish haqida so‘z boradi [Sharipov, Yuldashov, 2022]. Ushbu maqolada tabiiy tilni qayta ishlash (NLP)ning muhim yo‘nalishlaridan biri hisoblangan stemming masalasi o‘zbek tilidagi fe’l so‘z turkumi uchun qo‘llanilishi ko‘rib chiqilgan bo‘lib, o‘zbek tilidagi fe’l so‘z turkumiga mansub so‘zlarni chekli avtomatlar asosida qo‘shimchalarini qirqib olish algoritmini yaratish masalasi tushuntirilgan [Sharipov, Salayev, Matlatipov, 2021]. So‘z turkumlarini belgilash (teglash) vositasi va belgilangan ma’lumotlar bazasi haqida ushbu maqolada so‘z borgan bo‘lib, bunda ular o‘zbek tilidagi barcha 12 ta so‘z turkumlariga belgi kiritib belgilab chiqishgan va 20 ta sohaga tegishli balanslangan matnlarda o‘zlarining bazasidan va teglash usulidan foydalanib testlab, muhim natijalar olishgan va ularni jadvallar shaklida e’lon qilishgan [Sharipov, Kuriyozov, Yuldashev, Sobirov, 2023]. O‘zbek tilidagi so‘zlarni lemmasini topish uchun lemmatizatsiya algoritmi ishlab chiqish, bunda qoidaga asoslangan usuldan foydalanilgan [Sharipov, Sobirov, 2022]. Ushbu maqolada matnlarni tahlil qilishda mazmunini o‘zgartirmagan holda matndan nomuhim so‘zlarni olib tashlash masalasi katta ahamiyatga ega ekanligi va o‘zbek tilida yozilgan matn uchun avtomatik aniqlash xususiyatiga ega bo‘lgan nomuhim so‘zlarni olib tashlash, kerak bo‘lganda uni asl matnga qaytarib qo‘yish masalasi ko‘rib chiqilgan [Madatov, Sharipov, Bekchanov, 2021]. Bu maqolada amaliy loyiha doirasida yaratilgan ta’limiy korpus haqida so‘z borgan bo‘lib, ushbu korpusni yaratish tamoyillari, omillari, ma’lumotlar bazasi, modellari, axborot

tizimi, arxitekturasi va imkoniyatlari haqida ham soʻz borgan va barcha jarayonli muhim chizmalar va jadvallar orqali keltirilgan [Elov, Raupova, Abjalova, 2021]. Ushbu maqolada matnlarni belgilash zarurati, muhimligi haqida soʻz borgan boʻlib, ushbu jarayonni amalga oshirish uchun Yashirin Markov Modeli yordamida soʻzlarni teglash haqida soʻz yuritilgan, bundan tashqari jarayon Atoqli otlarni ham tanish xususiyatiga ega, bu jarayonlarni amalga oshirishda esa oʻzbek tili korpusidan foydalanilgan [Elov, Sirojiddinov, Hamroyeva, Adali, Xusainova, 2023]. Ushbu maqolada soʻzlarni ikkita asosiy qism nominal va verbal kombinatsiyasining lingvistik modeli asosida oʻzbek tili korpusini sintaktik teglash haqida soʻz boradi, modelda soʻzlar bir-biriga uch xil holatda bogʻlangan boʻlib, maqolada yana quyidagilar haqida soʻz yuritilgan, sintaktik munosabatlarni tahlil qilish va NLP uchun meta til modelini qurish [Abduraxmanova, 2020] Bundan tashqari ushbu maqola ilmiy mazmun-mohiyati ostida yotgan lingvistik manba yaratish ushuni tadqiqot obyekti sifatida bir nechta manbalardan foydalanilgan [Tojiyev, Nazarova, Tojiyeva, 2012].

Lingvistik soʻzlar, ularning tarixi va tarkibiy qonuniyatlari, uslubi va tarixi haqida maʼlumotlarni oʻrganishda asosiy maʼlumot manbalarining roli juda katta. Bu maqolada, tilshunoslik sohasidagi muhim konseptlarni va lingvistik manbalar haqida tafsilotlar keltirilib, ularning qanday qilib ilmiy tadqiqotlarda va boshqa sohalarida qoʻllanilishini tahlil qilganmiz. Lingvistik manba tildagi maʼlumotlarni izlash va tushunish uchun foydalaniladigan resurslar va manbalarni anglatadi. Quyidagi koʻrinishdagi manbalardan lingvistik manbalar hosil qilishda foydalanish mumkin:

- **tilshunoslik manbalari:** “Oʻzbek tilining tuzilishi”, “Oʻzbek tilining morfologiyasi”, “Oʻzbek tilida sintaksis”; “Oʻzbek tilining frazeologiyasi”, “Oʻzbek tilining qoʻshma soʻzlar toʻplami”;

- **lugʻatlar;**
- **internet manbalari;**
- **elektron kutubxonalar;**
- **ilmiy maqolalar;**
- **korpuslar.**

Lugʻatlar. Tildagi soʻzlar va ularning maʼnolari, koʻrinishi va tarixi haqida maʼlumotlar taqdim etuvchi kitoblar. Bu lugʻatlar odatda oʻzbek tilining tarixi va zamonaviy lugʻatlari, etimologik lugʻatlar, sinonimlar va antonimlar lugʻatlari, shuningdek, atamalardan foydalanilgan jargon va soʻzlar uchun mahsulotlar boʻlishi mumkin, masalan: “Oʻzbek tilining izohli lugʻati”, “Oʻzbek tilining imlo lugʻati” va b.

Tilshunoslik manbalari. Tilshunoslik sohasidagi tadqiqotlarda foydalanuvchilarga tilning strukturasini, oʻzgarishlarini va rivojlanishini tushunishga yordam beradi. Bu manbalar, oʻzlarining turli xil xususiyatlari va

foydalanish sohalariga ega bo'lganlik tufayli, tilshunoslik sohasidagi tadqiqotchilar uchun juda muhimdir.

Internet manbalari. Bu kabi manbalar internetdagi bloglar, web-saytlar, akademik platformalar va elektron kutubxonalardan foydalanishni o'z ichiga oladi.

Ilmiy maqolalar. Tilshunoslik sohasidagi so'nggi ilmiy o'zgarishlar, nazariy va amaliyotdagi yangiliklar, ko'nikmalar va teoriyalar haqida ma'lumot taqdim etuvchi ilmiy jurnallar va kitoblar. Ular odatda lug'atlar va tadqiqotlarda ko'rsatilgan ma'lumotlarni amaliyotda sinovdan o'tkazish uchun foydalaniladi.

Stemming - tilshunoslik (lingvistika) sohasida amal qiladigan bir tahlil metodidir. Ushbu metod ma'lum bir so'zni o'zining asosiy qismini (stemini) aniqlash va uni ko'p sonli variantlardan, ya'ni qo'shimchalar (masalan, ko'plik, zamon va hokazo) qo'shilgan holatidan qo'shimchalarni olib tashlashni amalga oshiradi. So'zning stemini topish uchun lingvistik manba yaratish o'zgaruvchan so'zlar, ularning tarkibiy qonuniyatlari va tarixi bo'yicha ma'lumotlar taqdim etadi. Bu tilshunoslik sohasidagi ilmiy tadqiqotlarda foydalanish uchun muhimdir. Lingvistikada stemming uchun manba yaratishda avvalo, so'zlar tarkibining analizini amalga oshirish kerak. Bu muhim yoki kerakli so'zlar ro'yxatini yaratish va ularning tarkibiy elementlarini aniqlashga yordam beradi. Misol uchun, “*ishchi*” so'zda “*ish*” so'zining stemi, “*chi*” esa suffiksdir. Tilning ma'nolari va tarkibiy elementlari haqida ma'lumotlarni o'rganishda lug'atlar va leksikonlardan foydalanish muhimdir. Ular odatda so'zlarning stemini belgilovchi to'plamlar, so'zlar tarkibini tahlil qiladigan muhim manbalar sifatida xizmat qiladi. Tildagi so'zlar va ularning o'zgartirishlarini tahlil qilish uchun korpuslar va tilning monologlaridan foydalanish ham muhimdir. Bu amaliy ma'lumotlarga ega bo'lish va so'zning har xil kontekstlaridagi o'zgarishlarni tushunishga yordam beradi. Stemming uchun lingvistik manba yaratish jarayonida, muhim bo'lgan so'zlarni, ularning tarkibiy qonuniyatlari va tarixi bo'yicha tahlil qilish uchun tajribali tadqiqotchilarning yordamiga ega bo'lish ham juda muhimdir.

Ushbu maqola o'zbek tili so'zlarining stemini aniqlash uchun lingvistik manba yaratishning metodologiyasini va jarayonning muhimligini belgilaydi. Bundan tashqari ushbu maqola o'zbek tilining lingvistik tahlilida va izlanishida muhim ma'lumotlar bilan ta'minlangan va tilshunoslik sohasidagi ilmiy tadqiqotchilar uchun muhimdir.

Foydalanilgan adabiyotlar:

1. Sharipov M, Yuldashov O. UzbekStemmer: Development of a Rule-Based Stemming Algorithm for Uzbek Language. CEUR Workshop Proceedings, CEUR-WS 2022, 137 – 144.
2. Sharipov M, Salayev U, Matlatipov G. O‘zbek tili fe’l so‘z turkumi uchun chekli avtomatlar asosidastemming algoritmini yaratish. Computer Linguistics: Problems, Solutions, Prospects 2021; 1.
3. Sharipov M, Kuriyozov E, Yuldashev O, Sobirov O. UzbekTagger: The rule-based POS tagger for Uzbek language. arxiv preprint arXiv:230112711 2023;
4. Sharipov M, Sobirov O. Development of a Rule-Based Lemmatization Algorithm Through Finite State Machine for Uzbek Language. CEUR Workshop Proceedings, CEUR-WS 2022, 154 – 159.
5. Madatov X, Sharipov M, Bekchanov S. O‘zbek tili matnlaridagi nomuhim so‘zlar. Computer linguistics: problems, solutions, prospect 2021; 1: 51–55.
6. Elov B, Raupova L, Abjalova M. O‘zbek tilining ta’limiy korpusi va uning imkoniyatlari. 2021; 4.
7. Boltayevich EB, Samariddinovich SS, Mirdjonovna KS, Adali E, Yuldashevna XZ. Pos Taging of Uzbek Text Using Hidden Markov Model. UBMK 2023 - Proceedings: 8th International Conference on Computer Science and Engineering, 2023, 63 – 68.
8. Zaynobiddin AN. Syntactic Models For Parsing Of Uzbek Corpus. Вестник науки и образования 2020; 15–19.
9. Tojiyev Y, Nazarova N, Tojiyeva G. O‘zbek tilidagi ergash morfemalarning semantik-stilistik xususiyatlari (affiksial ma’nodoshlik). - Toshkent, 2012.

ALGORITHMS BEHIND UZBEK CORPUS PARAPHRASING: A CLOSER LOOK AT LINGUISTIC SUPPORT AND METHODOLOGIES

Khayatova Zarnigor Marufovna

khayatovazarnigor@gmail.com

Doctor of Philosophy in Philological Sciences, PhD

Post-doctorate student at Tashkent State University of Uzbek

Abstract. This article explores the fascinating world behind AI-powered paraphrasing tools for the Uzbek language. We will crack open the code and see how these tools use clever algorithms and the power of the Uzbek Corpus to rephrase text in an accurate and meaningful way.

Get ready to learn about Machine Learning and Natural Language Processing (NLP) – the secret weapons in these tools' arsenal. We'll also explore how the Uzbek Corpus acts as a giant language teacher, providing the tools with the knowledge they need to understand and rephrase Uzbek text effectively.

But it is not just brute force – we will also delve into the linguistic support systems that work behind the scenes. Imagine the tool as a writer with a built-in grammar checker and thesaurus – that's the power of these linguistic systems!

By the end of this article, you will have a clear understanding of how these paraphrasing tools work, making you a more informed user and perhaps even inspiring you to explore the world of AI and language further.

Annotasiya. Ushbu maqolada o'zbek tili uchun sun'iy intellekt bilan ishlaydigan parafrazlash vositalari ortidagi maftunkor dunyoni ko'rib chiqilgan. Biz sirlarni ochamiz va ushbu vositalar aqlli algoritmlardan qanday foydalanishni va o'zbek korpusining matnni to'g'ri va mazmunli tarzda qayta yozish imkoniyatlarini ko'rib chiqamiz.

Mashinaviy o'rganish va tabiiy tilni qayta ishlash (NLP) – ushbu vositalar arsenalidagi maxfiy qurollar haqida bilib olishga tayyor bo'ling. Shuningdek, biz O'zbek korpusining ulkan til o'qituvchisi sifatida qanday ishlashini, o'zbek matnini samarali tushunish va qayta ifodalash uchun zarur bo'lgan bilimlarni taqdim etishini o'rganamiz.

Ammo bu shunchaki vosita emas-biz sahna ortida ishlaydigan lingvistik qo'llab-quvvatlash tizimlarini ham ko'rib chiqamiz. Asbobni o'rnatilgan grammatik tekshirgich va tezaurusga ega yozuvchi sifatida tasavvur qilib lingvistik tizimlarning kuchini sinab ko'ramiz.

Ushbu maqolaning oxiriga kelib, biz ushbu parafrazlash vositalarining qanday ishlashini tushunish, bu bizni yanada ma'lumotli foydalanuvchiga aylantiradi va ehtimol bizni AI va til dunyosini yanada ko'proq o'rganishga ilhomlantiradi.

Аннотация. В этой статье исследуется увлекательный мир инструментов перефразирования на базе искусственного интеллекта для узбекского языка. Мы взломаем код и увидим, как эти инструменты

используют умные алгоритмы и возможности корпуса узбекского языка для точной и осмысленной перефразировки текста.

Приготовьтесь узнать о машинном обучении и обработке естественного языка (NLP) – секретном оружии в арсенале этих инструментов. Мы также рассмотрим, как узбекский корпус выступает в роли гигантского преподавателя языка, предоставляя инструментам знания, необходимые для эффективного понимания и перефразирования узбекского текста.

Но это не просто грубая сила – мы также углубимся в системы лингвистической поддержки, которые работают за кулисами. Представьте себе инструмент в виде программы для написания текстов со встроенной системой проверки грамматики и тезаурусом - вот в чем сила этих лингвистических систем!

К концу этой статьи у вас будет четкое представление о том, как работают эти инструменты перефразирования, что сделает вас более информированным пользователем и, возможно, даже вдохновит вас на дальнейшее изучение мира искусственного интеллекта и языка.

Keywords: *Artificial Intelligence (AI), Paraphrasing, Uzbek Language, Uzbek Corpus, Machine Learning, Natural Language Processing (NLP), Linguistic Support Systems, Part-of-Speech Tagging, Named Entity Recognition, Academic Writing.*

The ability to paraphrase effectively is a cornerstone of academic writing, allowing for clarity, conciseness, and the avoidance of plagiarism. However, for Uzbek speakers, this task can be particularly challenging, especially when dealing with complex concepts and nuanced sentence structures. Thankfully, the emergence of Artificial Intelligence (AI)-powered paraphrasing tools offers a revolutionary solution. These tools leverage sophisticated algorithms and the rich resources of the Uzbek Corpus to generate accurate and meaningful paraphrases. But how exactly do they achieve this seemingly magical feat? Let's embark on a journey to unveil the intricate mechanisms powering AI paraphrasing for the Uzbek language.

1. The Engine Room: Unveiling the Algorithmic Heart

At the heart of AI paraphrasing tools lies a complex interplay of algorithms, each playing a crucial role in the transformation process. Machine Learning (ML) forms the core functionality, acting as a tireless student eager to learn. Drawing inspiration from vast amounts of text data within the Uzbek Corpus, ML algorithms like Support Vector Machines (SVMs) or Recurrent Neural Networks (RNNs) [Batyrkhanov, Yusupova, Bolatov, 2020] identify patterns and relationships between words and phrases. These patterns, represented mathematically as feature vectors (X) in Equation 1, encode information about word order, syntactic structure, and semantic relationships:

$$X = [x_1, x_2, \dots, x_n]$$

where each element x_i represents a specific feature extracted from the text data.

The learned patterns are then used to predict new paraphrases for unseen input sentences. This prediction process often involves a scoring function (f) that evaluates the quality of candidate paraphrases. One popular approach utilizes a combination of language models and attention mechanisms [Bird, Klein, Loper, 2006]. Language models estimate the probability of a word sequence occurring in a grammatically correct and fluent paraphrase. Attention mechanisms, on the other hand, focus the model's attention on specific parts of the original sentence that are most relevant for generating an accurate paraphrase.

Complementing the power of ML is Natural Language Processing (NLP). This field empowers AI paraphrasing tools with the ability to understand the nuances of the Uzbek language. Techniques like part-of-speech tagging (POS tagging) and named entity recognition (NER) play a critical role in this process. POS tagging assigns each word in the sentence a grammatical category (e.g., noun, verb, adjective) as shown in Equation 2:

$$S = [w_1(\text{POS}_1), w_2(\text{POS}_2), \dots, w_n(\text{POS}_n)]$$

where S represents the tagged sentence, w_i represents the i -th word, and POS_i denotes its corresponding part-of-speech tag. This information is crucial for maintaining grammatical accuracy during paraphrasing. Similarly, NER identifies and classifies named entities like people, places, and organizations, ensuring their preservation and accurate rephrasing.

Together, ML algorithms and NLP techniques form a powerful engine that drives the AI paraphrasing process. By analyzing the intricate relationships within the Uzbek language as captured in the Corpus data, these algorithms can generate paraphrases that not only preserve meaning but also maintain grammatical correctness and stylistic flow.

At the heart of AI paraphrasing tools lie sophisticated algorithms. Two key players are:

- **Machine Learning (ML):** Imagine a student learning from examples. Similarly, ML algorithms learn from vast amounts of text data to identify patterns and relationships between words and phrases [Carroll, O'Neill, 2017]. These patterns are then used to generate paraphrases for new text.
- **Natural Language Processing (NLP):** This field empowers machines to understand the nuances of human language. NLP techniques like part-of-speech tagging (identifying nouns, verbs, etc.) and named entity recognition (recognizing people, places, etc.) allow the AI to grasp the deeper meaning of the text and generate contextually appropriate paraphrases [Jurafsky, Martin, 2020].

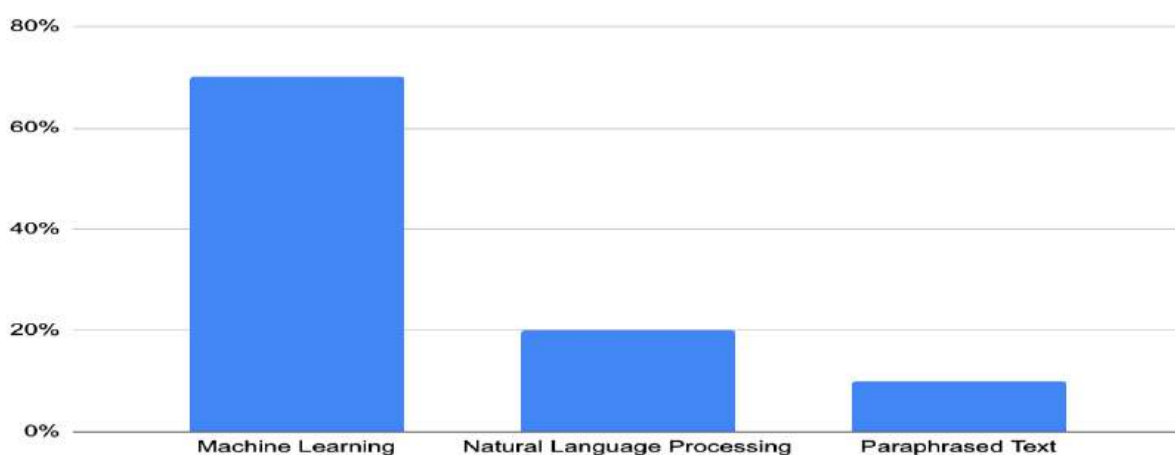


Figure 1: Unveiling AI Paraphrasing - A Model Overview

The percentages in the diagram represent a rough estimate of the relative importance of each component. Machine Learning forms the core functionality (70%), as it's responsible for identifying patterns and generating paraphrases. NLP plays a crucial role in understanding meaning and context (20%), while the final output, the paraphrased text, represents the user-facing outcome (10%).

2. The Knowledge Bank: The Uzbek Corpus as a Training Ground

The Uzbek Corpus plays a vital role in shaping the capabilities of AI paraphrasing tools. The sheer volume and diversity of text data it provides allow the algorithms to learn the intricacies of the Uzbek language and generate paraphrases that are not only accurate but also natural-sounding.

The ability to paraphrase effectively is a valuable skill, especially in academic writing. AI-powered tools are transforming the way Uzbek speakers can rephrase text while maintaining meaning and clarity. But how do these tools achieve this seemingly magical feat? Let's delve into the three key components that power these paraphrasing wonders: algorithms, the Uzbek Corpus, and linguistic support systems. The Uzbek Corpus serves as a treasure trove of Uzbek language data. Think of it as a massive library of Uzbek text, containing books, articles, and other resources. This data is crucial for training the AI paraphrasing tools. Here's how:

- **Data Pre-processing:** The vast amount of text in the Corpus needs to be cleaned and organized for the AI to understand it. This involves removing irrelevant information like punctuation and formatting errors.
- **Feature Extraction:** The AI focuses on specific elements within the text, such as word order, sentence structure, and word usage patterns. These elements are extracted and analyzed to create a "fingerprint" of the original text.
- **Model Training:** the extracted features and paraphrased examples from the Corpus are used to train the Machine Learning algorithms. This training allows the AI to learn how to identify patterns and generate new paraphrases based on the provided data.

Process	Description
Data Collection	Gathering a vast amount of Uzbek text from various sources.
Data Cleaning	Removing irrelevant information and formatting errors from the text.
Feature Extraction	Identifying key elements like word order and usage patterns.
Model Training	Training the AI algorithms using features and paraphrased examples.

Figure 2: The Uzbek Corpus - Fueling AI Paraphrasing

The world of language is a vast and intricate one, filled with unique structures and cultural nuances. Natural Language Processing (NLP), a branch of Artificial Intelligence, strives to bridge the gap between human and machine understanding of this complex system. One crucial aspect of NLP is paraphrase generation, the ability to rephrase a sentence while preserving its core meaning. This becomes particularly challenging when dealing with languages like Uzbek, with its rich grammar and distinct cultural expressions. Here, we delve into the exciting potential of Machine Learning (ML) in tackling the challenge of Uzbek text paraphrasing.

At the heart of successful paraphrase generation lies NLP, a subfield dedicated to enabling computers to process and understand human language. In the context of Uzbek paraphrasing, several NLP techniques play crucial roles. Part-of-speech (POS) tagging acts as the foundation, categorizing each word in a sentence by its grammatical function, such as noun, verb, or adjective. This ensures that the paraphrased sentence maintains a coherent structure. Imagine the sentence "كِتابنى o'qish kerak" (kitobni - the book, o'qish - reading, kerak - needs). POS tagging helps the ML model understand that "kitobni" is a noun and "o'qish" is a verb, allowing it to identify potential substitutions while guaranteeing a grammatically correct outcome in the paraphrase.

Another critical NLP technique is Named Entity Recognition (NER). This technique identifies and classifies named entities within a sentence, such as people, locations, and organizations. Names hold significant meaning and should be preserved accurately during paraphrasing. NER allows the ML model to recognize these entities and potentially find suitable synonyms while maintaining the context and information conveyed by the original sentence. For instance, if the sentence mentions a specific historical figure like "Amir Temur" (a historical ruler), NER ensures this name remains unchanged in the paraphrase.

Beyond individual words, dependency parsing dives deeper, revealing the relationships between words within a sentence. This information is crucial for understanding the overall structure and meaning of a sentence. By analyzing these dependencies, the ML model can identify potential areas for paraphrasing while

ensuring the new sentence retains the intended meaning and relationships between words. Imagine the sentence "Madrasa yopiq bo'ldi" (madrasa - school, yopiq - closed, bo'ldi - became). Dependency parsing helps the model understand the relationship between "madrasa" and "yopiq," allowing it to paraphrase the sentence accurately while preserving the connection between the school and its closure.

Machine learning algorithms offer a powerful foundation for paraphrase generation. However, to effectively rephrase Uzbek text, the model needs to focus on the most relevant parts of the source sentence. This is where attention mechanisms come into play. These are specific architectures within the machine learning model that enable it to selectively focus on crucial elements of the input sentence. By attending to specific words or phrases, the model can understand the core meaning and context of the sentence, ultimately leading to more accurate paraphrases. For example, the model might pay closer attention to the verb and its surrounding words ("o'qish kerak" - needs reading) while paraphrasing, as opposed to focusing on less critical articles ("kitobni" - the book).

Despite the promising prospects, challenges remain associated with developing ML-based Uzbek text paraphrasing. One significant hurdle is the limited availability of large, high-quality Uzbek text datasets for training machine learning models. The more data the model has to learn from, the better it can understand the nuances of the language and generate accurate paraphrases.

Additionally, ensuring cultural nuances and idiomatic expressions are preserved during paraphrasing requires further exploration and adaptation of existing ML techniques.

Reference:

1. Batyrkhanov, M., Yusupova, A., & Bolatov, Z. (2020). Morphological Analysis of Kazakh and Uzbek Languages Using Conditional Random Fields. <https://aclanthology.org/W04-3230>
2. Bird, S., Klein, E., & Loper, E. (2006). Natural language processing with Python. O'Reilly Media, Inc.
3. Carroll, J., & O'Neill, S. (2017). Machine learning in language processing. Springer.
4. Jurafsky, D., & Martin, J. H. (2000). Speech and language processing. Prentice Hall pp.193-199)
5. Manning, C. D., Surdeanu, M., João Silva, J., Greenstein, R., Sagot, S. (2015). The Stanford CoreNLP natural language processing toolkit. In Proceedings of the 52nd Annual Meeting of the Association.

UNDERSTANDING THE DISTINCTION: WORD AND PHRASE ALIGNMENT

Matyakubova Noila

matyakubovanoila@navoiy-uni.uz

ToshDO‘TAU tayanch doktoranti

Abstract. Word and phrase alignment in natural language processing (NLP) plays a crucial role in various tasks such as machine translation, text summarization, information retrieval, and sentiment analysis. Parallel texts of parallel corpus, where content exists in multiple languages, offer a treasure trove for language processing tasks like translation and dictionary development. But extracting meaning from these texts involves aligning corresponding units across languages. While both aim to bridge the language gap, they differ in their approach and scope.

Key words. *Word alignment, phrase alignment, alignment algorithms, source language, target language.*

Introduction

In the realm of linguistics and natural language processing (NLP), the alignment of words and phrases plays a crucial role in tasks like machine translation, sentiment analysis, and text summarization. However, it's essential to grasp the nuanced differences between word and phrase alignment to effectively harness their power in various applications. But both word and phrase aligners can be useful for the applications like:

- **Machine Translation:** In machine translation, aligning words and phrases between source and target languages is fundamental. It helps in identifying corresponding words and phrases, which is essential for generating accurate translations.
- **Text Summarization:** When generating summaries, aligning important words and phrases between the original text and the summary helps in preserving the meaning and coherence of the content. It ensures that the summary reflects the key points of the original document.
- **Information Retrieval:** In information retrieval systems, aligning query terms with relevant documents is crucial for retrieving accurate and relevant results. Word and phrase alignment help in matching the user's query with the content of documents.
- **Sentiment Analysis:** Aligning sentiment-bearing words and phrases in opinionated texts is important for sentiment analysis tasks. Identifying the sentiment-bearing units accurately helps in determining the overall sentiment polarity of the text.
- **Text Alignment:** In tasks like plagiarism detection and paraphrase identification, aligning words and phrases between two texts helps in detecting

similarities and differences. This alignment is essential for determining the degree of similarity between texts.

- **Named Entity Recognition (NER):** Aligning named entities across different documents or languages is important in tasks like cross-lingual entity linking or entity alignment. It helps in identifying the same entities mentioned in different contexts.

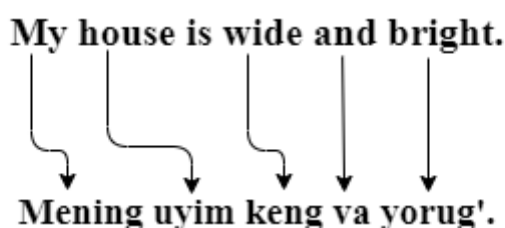
- **Coreference Resolution:** Aligning pronouns with their corresponding antecedents is essential for coreference resolution tasks. This ensures that pronouns are correctly resolved to their referents, leading to a better understanding of the text.

- **Language Model Training:** Word and phrase alignment can be crucial in the training of language models, especially for tasks like machine translation where parallel corpora are used. Aligning sentences, words, and phrases between source and target languages helps in training robust translation models.

Word alignment

Word alignment primarily deals with aligning individual words between two or more languages or sentences. This process is fundamental in machine translation systems, where the goal is to accurately map words from a source language to a target language. Word alignment involves establishing correspondences between words in different languages, forming pairs known as alignments.

For instance, if we take into consideration the sentence pair “My house is wide and bright” (English) and “Mening uyim keng va yorug’” (Uzbek), alignment process looks like this.



Picture 1. Word alignment process.

This alignment enables the translation system to understand the relationship between words in the source and target languages, facilitating accurate translation.

Table 1. Advantages and disadvantages of word alignment process.

Advantage	Disadvantage
Provides a foundation for more complex alignment tasks like phrase alignment.	Fails to capture the nuances of language, where meaning is often conveyed through multi-word expressions.

Useful for tasks like bilingual dictionary creation, where individual word equivalences are crucial.

Simpler and computationally efficient, making it suitable for large datasets.

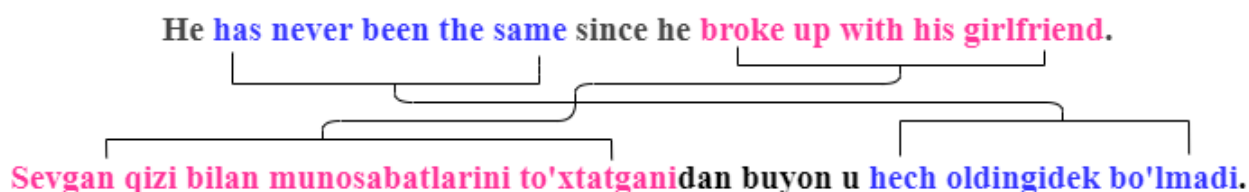
Can be misled by synonyms, homonyms, and different sentence structures across languages.

Word alignment algorithms often rely on statistical methods or neural network architectures to learn alignments from parallel corpora, which are collections of translated texts in multiple languages. These algorithms aim to maximize the alignment's likelihood based on observed patterns in the data.

Phrase alignment

While word alignment focuses on individual words, *phrase alignment* operates at a higher linguistic level by aligning sequences of words, or phrases, between languages or sentences. Phrases are groups of words that function as a single unit in a sentence and convey a specific meaning or grammatical function.

Phrase alignment recognizes that translation equivalence often extends beyond individual words, capturing multi-word expressions, idiomatic phrases, and syntactic structures that may not align perfectly on a word-by-word basis. By aligning phrases, translation systems can achieve more accurate and fluent translations, preserving the original sentence's structure and semantics.



Picture 2. Phrase alignment process.

Phrase alignment algorithms build upon word alignment techniques but extend them to handle larger linguistic units. These algorithms may incorporate syntactic information, linguistic rules, or machine learning models to identify meaningful phrase correspondences across languages.

Table 2. Advantages and disadvantages of phrase alignment process.

Advantage	Disadvantage
Captures the essence of meaning, leading to more accurate translation and analysis.	Computationally more complex compared to word alignment.

Accounts for word order variations and idiomatic expressions that can challenge word alignment.	Requires sophisticated algorithms incorporating syntactic and semantic knowledge.
Enables applications like sentiment analysis, where understanding the overall tone of a phrase is crucial.	May face ambiguity when phrases can be interpreted in multiple ways.

Several word and phrase aligners are used in natural language processing tasks nowadays. These aligners vary in their approaches, complexity, and the tasks they are designed for. Here are some common ones:

- **GIZA++ and MGIZA++:** These are widely used tools for word alignment in machine translation. They are based on IBM Models and are often used to align words in parallel corpora[1,2].
- **FastAlign:** FastAlign is a fast and memory-efficient word alignment tool based on an unsupervised approach using Expectation-Maximization algorithm. It's commonly used in machine translation and other tasks requiring word alignment[3].
- **Berkeley Aligner:** The Berkeley Aligner is a phrase-based alignment tool designed for statistical machine translation. It aligns phrases across parallel sentences in a parallel corpus.
- **Bitextor:** Bitextor is a tool for bilingual text alignment, commonly used for building parallel corpora. It aligns texts at various granularities, including words, phrases, and sentences[4].
- **OpenNMT:** OpenNMT is an open-source neural machine translation framework that includes modules for word alignment. It allows for end-to-end training of neural machine translation models, which implicitly learn alignments.
- **Gentle Align:** Gentle Align is a library for aligning phonetic transcriptions to audio files. It's often used in speech processing tasks, such as aligning transcripts to audio recordings for speech recognition.
- **Stanford CoreNLP:** Stanford CoreNLP toolkit includes modules for various NLP tasks, including sentence alignment, coreference resolution, and named entity recognition, which involve aligning words and phrases across documents or sentences[5].
- **FastText:** FastText, a library developed by Facebook, provides tools for word alignment and similarity computation. While primarily known for its word embedding capabilities, it can also be used for aligning words based on vector representations.

These aligners serve different purposes and are utilized in various NLP applications, ranging from machine translation to speech processing. Researchers and practitioners choose aligners based on factors such as the task requirements, dataset characteristics, and computational resources available.

Key Differences

- **Granularity:** Word alignment operates at the level of individual words, whereas phrase alignment considers sequences of words or phrases.
- **Complexity:** Phrase alignment is inherently more complex than word alignment due to the larger search space and the need to capture linguistic units beyond single words.
- **Translation Quality:** Phrase alignment often leads to higher translation quality by preserving syntactic structures and capturing multi-word expressions, enhancing fluency and accuracy.
- **Application:** Word alignment is foundational in machine translation and related tasks, while phrase alignment further refines translation quality and is crucial for advanced translation systems.

The choice between word and phrase alignment depends on the specific task and the desired level of detail. Word alignment is suitable for tasks requiring basic word-level mapping and large data analysis.

Phrase alignment offers a deeper understanding for tasks where capturing the complete meaning of an expression is critical.

Conclusion.

In conclusion, while both word and phrase alignment serve the overarching goal of facilitating accurate translation between languages, they operate at different linguistic levels and offer distinct advantages. Understanding these differences is essential for developing robust and effective translation systems in the field of natural language processing. The future of alignment is likely to see a hybrid approach, combining the strengths of both methods. Advances in machine learning and natural language processing promise even more sophisticated techniques that can seamlessly switch between word and phrase alignment depending on the context.

References:

1. Qin Gao, Stephan Vogel, “Parallel Implementations of Word Alignment Tool”, Software Engineering, Testing, and Quality Assurance for Natural Language Processing, Columbus, Ohio, USA, June 2008. c 2008 Association for Computational Linguistics.
2. Marcin Junczys-Dowmunt, Arkadiusz Szal, “SyMGiza++: Symmetrized Word Alignment Models for Statistical Machine Translation”, International Joint



Conferences, SIIS 2011, Warsaw, Poland, June 13-14, 2011, Revised Selected Papers.

3. Chenchen Ding, Masao Utiyama, Eiichiro Sumita, "Improving fast align by Reordering", Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal, 17-21 September 2015. c 2015 Association for Computational Linguistics.

4. Miquel Esplà-Gomis, "Combining Content-Based and URL-Based Heuristics to Harvest Aligned Bitexts from Multilingual Sites with Bitextor", Prague Bulletin of Mathematical Linguistics, January 2010.

5. Christofor. D Manning, John Bauer, Jenny Finkel, Mihai Surdeanu, "The Stanford CoreNLP Natural Language Processing Toolkit", Conference: Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System, January 2014.

O'ZBEK TILIDA GAP BO'LAKLARI TARTIBINI AVTOMATIK ANIQLASH (SODDA GAPLAR MISOLIDA)

Abdullayeva Oqila Xolmo'minovna

abdullayeva.oqila@navoiy-uni.uz

f.f.f.d., PhD, ToshDO'TAU doktoranti

Nizomova Fotima Bo'rixo'ja qizi

ToshDO'TAU talabasi

Qodirov Kamoliddin Qodirjon o'g'li

ToshDO'TAU talabasi

Annotatsiya. Ushbu maqolada o'zbek tilida gap bo'laklari tartibini avtomatik aniqlash yuzasidan nazariy va amaliy fikrlar muhokama qilingan. Gap bo'laklarini avtomatik aniqlashda o'zbek tilidagi sodda gaplar olindi va ularning lisoniy sintaktik qolipi berildi. O'zbek va ingliz tillaridagi gap bo'laklari tartibi lingvistik modellash tirish tahlil qilindi. O'zbek tilida gap bo'laklari tartibini avtomatik aniqlash dasturi to'g'risida fikr va xulosalar keltirib o'tildi.

Abstract. In this article, theoretical and practical ideas about the automatic determination of the order of sentence fragments in the Uzbek language are discussed. Simple sentences in the Uzbek language were taken and their linguistic syntactic pattern was given in the automatic detection of parts of speech. Linguistic modeling of the order of sentence fragments in Uzbek and English languages was analyzed. Opinions and conclusions were given about the program for automatic determination of the order of sentence fragments in the Uzbek language.

Абстракт. В данной статье обсуждаются теоретические и практические представления об автоматическом определении порядка фрагментов предложений в узбекском языке. Были взяты простые предложения на узбекском языке и задана их лингвосинтаксическая картина при автоматическом определении частей речи. Проанализировано лингвистическое моделирование порядка фрагментов предложений в узбекском и английском языках. Приведены мнения и выводы о программе автоматического определения порядка фрагментов предложений на узбекском языке.

Kalit so'zlar: *Sodda gap, lisoniy sintaktik qolip, noodatiy tartibdagi gap, to'g'ri tartibdagi gap, lingvistik modellash tirish, sintaktik analiz.*

Kirish. Gaplar strukturasiga ko'ra, to'rt xil bo'ladi: sodda, murakkab, qo'shma gaplar va periodlar.

Sodda gap ma'lum fikr, maqsadni ifodalovchi, grammatik va intonatsion shakllangan sintaktik birlikdir. Sodda gap tuzilishiga ko'ra yig'iq yoki yoyiq bo'lishi mumkin. Bosh bo'laklardan tarkib topgan gaplar — *yig'iq gap*, agar bosh

bo'laklardan tashqari ikkinchi darajali bo'laklar ham ishtirok etsa, *yoyiq gap* sanaladi [Abdullayev, 1976: 95].

Sodda gaplar eganing ishtirok etish yoki etmasligiga ko'ra ikki xil bo'ladi: *egali gaplar* (Biz yangi filmni tomosha qildik) va *ega ishtirok etmagan gaplar* (Yangi filmni tomosha qildik). Sodda gaplar ikkinchi darajali bo'laklarning ishtirok etgani yoki etmaganiga ko'ra: *sodda yig'iq gaplar* (Kamola kirib keldi), *sodda yoyiq gaplar* (Kamola darvozadan shoshilib kirib keldi)ga bo'linadi [Qodirov, 2014: 54].

Gapda so'zlar o'zaro grammatik munosabatga kirishib, gap bo'laklarini hosil qiladi. Gap bo'laklari vazifasini mustaqil so'zlar va gapda mustaqil so'zlar vazifasida kelgan yordamchi so'zlar bajaradi. Har bir gap bo'lagi boshqa gap bo'laklari bilan bo'lgan grammatik munosabatiga ko'ra belgilanadi, ya'ni har bir gap bo'lagi o'zi munosabatga kirishgan so'z bilan ma'lum sintaktik aloqada bo'ladi. Masalan, Dildora kechagi uchrashuv haqida zavqlanib gapirdi gapida quyidagi gap bo'laklari mavjud: Dildora (kim?) - ega, kechagi (qachongi? qaysi?) — aniqlovchi, uchrashuv haqida (nima haqida?) — to'ldiruvchi, zavqlanib (qay tarzda?) — hol, gapirdi (nima qildi?) — kesim. So'zlar quyidagicha sintaktik aloqaga kirishgan:

1) Dildora gapirdi; 2) zavqlanib gapirdi; 3) uchrashuv haqida gapirdi; 4) kechagi uchrashuv. Gap bo'laklari 5 ta: kesim, ega, to'ldiruvchi, aniqlovchi, hol. Kesim gapning asosini tashkil qiluvchi markazdir. Ega kesimga ergashib, kesimda ifodalangan ish-harakatning bajaruvchisini ko'rsatadi.

To'ldiruvchi, aniqlovchi, hol bo'laklari ega va kesimga ergashib kelib, ularni to'ldiradi, aniqlaydi yoki izohlaydi. Shuning uchun ular gapning *ikkinchi darajali bo'laklari* deyiladi [Qodirov, 2014: 56].

Kesim nima qildi? nima qilyapti? nima qiladi? kabi so'roqlarga javob bo'ladi. Kesim qaysi turkum so'zlari bilan ifodalanishiga ko'ra ikki xil: fe'l kesim, ot kesim shaklida, tuzilishiga ko'ra ikki xil: *sodda kesim* va *murakkab kesim* bo'ladi [Qodirov, 2014: 58].

Gap kesimining qo'shimchalaridan anglashilib turgan shaxs-son ma'nosini aniqlashtiruvchi bo'lak egadir. Ega, asosan, kim? nima? qayer? so'roqlariga javob bo'lib keladi. Ega doimo bosh kelishik shaklida bo'ladi, egalik va ko'plik shakllarini qabul qilishi mumkin. Ega quyidagi so'z turkumlari bilan ifodalanadi:

1. Ot bilan: *Aziz* o'sha kitobni o'qidi. Milt-milt yonib turgan *chiroq* birozdan so'ng o'chdi.

2. Olmosh bilan: *Sen* o'zingni maqtama, seni birov maqtasin. (Maqol)

3. Son bilan: *Ikkalasi* ham chiqib ketishdi. (M. Ism.)

4. Sifatdosh bilan: *O'qigan* o'qdan o'zar, *o'qimagan* turtkidan shoshar. (Maqol)

5. Sifat bilan: *Yaxshilar* ko'paysin, *yomon* qolmasin. (Qo'shiqdan)

6. Harakat nomi bilan: *O'qish* jafoli, keti vafoli. (Maqol) Ish ishtaha ochar. (Maqol)

7. Taqlid so'z bilan: *Qars* ikki qo'ldan chiqadi. (Maqol) [Qodirov, 2014: 68].

Gapning kesimi yoki egasiga tobelanib, ularga ergashib keladigan bo‘laklar ikkinchi darajali bo‘laklardir. Ikkinchi darajali bo‘laklar uch turli: bular aniqlovchi, to‘ldiruvchi va hollardir. Ular fe‘l, ot, sifat, son, ravish, taqlid so‘zlar, olmoshlar bilan ifodalanishi mumkin. Aniqlovchi, to‘ldiruvchi, hollar gapni, uning ayrim bo‘laklarini kengaytirishga, aniqlashga xizmat qiladi. Gap tarkibida ikkinchi darajali bo‘laklardan birontasi ishtirok etsa, bunday gap yoyiq gap hisoblanadi [Qodirov, 2014: 80].

Hol ko‘proq gap markazi — kesimga bog‘lanib, undan anglashilgan ish-harakatning o‘rnini, paytini, holatini, bajarilish sababini, maqsadini, darajamiqdorini bildiradigan ikkinchi darajali bo‘lakdir. Hollar qanday? qanday qilib? qay tarzda? qayerda? qayerga? qayerdan? qachon? nega? nima uchun? qancha? kabi so‘roqlarning biriga javob bo‘ladi. Hollar ravish so‘zlar bilan, kelishik shaklidagi va ko‘makchili otlar bilan, sifat, son, taqlid so‘z va undovlar bilan, fe‘lning xoslangan shakllari bilan ifodalanishi mumkin. Hollar sodda (bir so‘zdan iborat) va murakkab (ikki va undan ortiq so‘zdan iborat) tuzilishga ega bo‘ladi [Qodirov, 2014: 82].

Gapning biror bo‘lagiga (ko‘pincha kesimga) boshqaruv yo‘li bilan bog‘lanib, uni to‘ldirib kelgan ikkinchi darajali bo‘lak to‘ldiruvchi deyiladi.

To‘ldiruvchi quyidagicha ifodalanadi:

1. **Ot bilan:** Dilshod maktubni xolasiga berib, ko‘chaga chiqib ketdi. (M.Ism.) Mansurov odamlar bilan xayrlashdi.
2. **Otlashgan sifat bilan:** Yaxshiga qora yuqmas, yomonga el boqmas. (Maqol).
3. **Otlashgan son bilan:** Beshni beshga qo‘shsak, o‘n bo‘ladi.
4. **Sifatdosh bilan:** Bilmagandan bilgan yaxshi, to‘g‘ri ishni qilgan yaxshi. (Maqol)
5. **Harakat nomi bilan:** Uning kulib boqishlarini bir zum ham unutmayman.
6. **Taqlid so‘z bilan:** Taqir-tuquringni yig‘ishtir.
7. **Modal so‘z bilan:** Yo‘qni yo‘q deydi, borni bor deydi.

To‘ldiruvchilar yordamida aytilmoqchi bo‘lgan fikr ravshanlashadi: Oyni (etak bilan) yopib bo‘lmas. (Maqol) Odam (mehnatda) sinaladi. Bu gaplardagi etak bilan, mehnatda bo‘laklari yordamida so‘zlovchining niyati to‘la ochilgan. [Qodirov, 2014: 89-90].

Gapning biror bo‘lagini aniqlab kelgan ikkinchi darajali bo‘lak aniqlovchi deyiladi. Aniqlovchi qanday? qaysi? nechanchi? qancha? nechta? kimning? nimaning? so‘roqlaridan biriga javob bo‘ladi. Aniqlovchi aniqlagan (bog‘langan) bo‘lak aniqlanmish hisoblanadi. Aniqlovchi ega, ot kesim, to‘ldiruvchi va holga bog‘lanadi: Bahaybat samolyot (ega) yerga ohista qo‘ndi. Salimjon — a‘lochi o‘quvchi (ot kesim). Ilg‘or ishchilarga (to‘ldiruvchi) mukofotlar topshirildi. Mehnat faxriylari bilan yangi bog‘da (hol) uchrashdik.

Aniqlovchilarning bir xili narsa-buyumning belgisi (rangi, miqdori, mazasi, hajmi)ni bildirsa, boshqa xili narsa-buyumning kimga yoki nimaga qarashlilikini bildiradi. Shunga ko‘ra, aniqlovchilar ikkiga bo‘linadi: 1) sifatlovchi aniqlovchi; 2)

qaratqichli aniqlovchi. Misol: Yangi (sifatlovchi aniqlovchi) binolar qishloqning (qaratqichli aniqlovchi) husniga husn qo'shmoqda. [Qodirov, 2014: 94-95].

O'zbek tilida so'zlar tartibi, asosan erkin. Gapdagi bo'laklarning o'rnini o'zgartirish ularning grammatik holatlariga ta'sir qilmaydi.

1. Kinoga kecha bordim.

2. Kecha kinoga bordim.

Biroq har bir tartib o'zgarishi mazmunga, uning ottenkasiga, yo'nalishiga ma'lum darajada ta'sir ko'rsatadi. Birinchi misolda vaqt ma'nosi bosh o'rinda, ikkinchisida-o'rin ma'nosi. Demak, bunda so'zlovchining maqsadi, nimaning birinchi o'ringa qo'yilishi katta ahamiyatga ega bo'ladi.

So'zlar tartibi, ba'zan grammatik holatga ham ta'sir etishi mumkin. Bu bilan ma'no ham o'zgaradi. Bu hodisa biror so'zning vazifasi-grammatik funksiyasi morfologik shakllanishdan bilinmay, tartib bilan tayin etilgan vaqtlarda uchraydi. Bu hodisa tartibning grammatik rolidir. Boshqa o'rinlarda tartib stilistik rol o'ynaydi [G'ulomov A., Asqarova M., 1965].

O'zbek tilida gap bo'laklari tartibini avtomatik aniqlashda gap bo'laklarining odatiy va o'zgargan tartibini o'rganib chiqishimiz lozimdir. Bu bizga sodda gap modellarini yaratishga va algoritm ishlab chiqishda kerak bo'ladi. O'zbek tilida gap bo'laklarining joylashish tartibi, asosan, erkin bo'lsa-da, ular gapda ma'lum grammatik qoidalar asosida o'rinlashadi. Odatda, gapda ega va unga bog'langan bo'laklar oldin, kesim va unga bog'langan bo'laklar keyin joylashadi, ba'zan aksinchasi ham bo'ladi: Bu yo'llar menga tanish. — Menga tanish bu yo'llar. Eganing ko'proq gap boshida, kesimning, asosan, gap oxirida (Men keldim), hol (ohista gapirdi) va to'ldiruvchining kesimdan (she'rni yod oldi), aniqlovchining egadan oldin (qizil gul, do'stimning akasi) joylashuvi odatdagi, me'yoriy tartibdir. Lekin nutqda doimo ham gap bo'laklarining odatdagi tartibi saqlanavermaydi, gapdan ko'zda tutilgan umumiy fikrga qo'shimcha ma'no, uslubiy yuk berish uchun tartib o'zgartirilishi mumkin [Qodirov, 2014: 101-102].

Morfologik analizning mantiqiy davomi sintaktik analiz hisoblanadi. Sintaktik analiz morfologik analizga asoslanadi. Bunda sintaktik analiz turli xil modellar asosida amalga oshirilishi mumkin. Formal grammatika asoschisi bo'lgan Noam Xomskiy belgilagan algoritm o'zining optimalligi bilan boshqa sintaktik modellardan ajralib turadi. Matnning avtomatik sintaktik analizi masalasi XX asrning 50-yillarida bir tildan boshqa tilga kompyuter yordamida tarjimani amalga oshiruvchi tizimlarga oid tadqiqotlar bilan uzviylikda o'rganilgan. Zero, matn tarjimasida gaplarning sintaktik strukturasi, sintaktik aloqasi hal qiluvchi ahamiyat kasb etadi [Xolmanova, 2020: 78].

O'zbek tilida gap bo'laklari tartibini aniqlash va noto'g'ri tartibdagi sodda gaplarni to'g'ri tartibga keltirish uchun quyidagi amallarni amalga oshirishimiz lozim.

Gapdagi grammatik tuzilmani tahlil qilish va o'zaro bog'langan so'zlarni, shuningdek, ular o'rtasidagi munosabat turini aniqlash.

O'zbek tilining rasmiy tilida barcha mumkin bo'lgan so'z tartibi qoidalarni yaratish uchun zarur bo'lgan grammatik qoidalarni yig'ish.

Grammatik qoidalardan kelib chiqqan holda sodda gaplarning modellarini tuzish.

Sodda gaplarni avtomatik aniqlash algoritmlarini ishlab chiqish.

Dasturiy ta'minotni yaratish.

O'zbek tili matnlaridan sodda gaplarni avtomatik tahlil qiluvchi ilova yaratish.

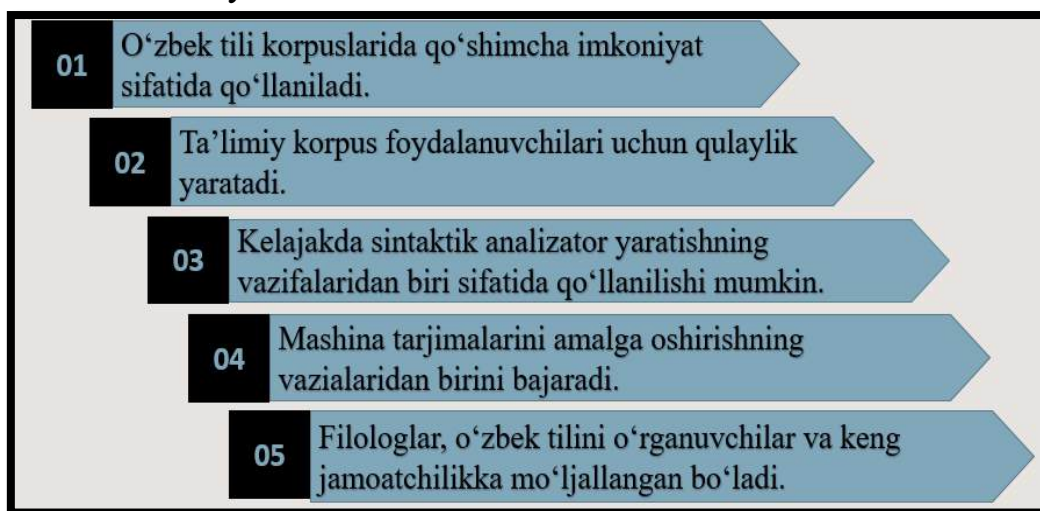
Bu amallarni bajarish uchun biz o'zbek tilidagi sodda gaplarni manba sifatida olishimiz lozim ularni yuqorida keltirib o'tgan nazariy va grammatik qoidalar asosida, odatiy va noodatiy tartibini o'rganib, o'zbek va ingliz tilidagi ma'lum bir belgilar orqali teglab olamiz. Qoidalar va teglar orqali dasturiy ta'minot ishlab chiqamiz.

ID	Teg	O'zbek tilidagi teg	Ma'nosi	O'zbekcha misollar
1	N	N	Ot(<u>N</u>)	olma (<i>apple</i>)
2	V	VB	Fe'l(<u>V</u>)	yugurmoq (<i>run</i>)
3	ADJ	JJ	SIFAT(<u>Adjective</u>)	ko'p (<i>many/much</i>)
4	NUM	NUM	SON(<u>Numeral</u>)	besht (<i>five</i>)
5	ADV	RR	RAVISH(<u>Adv</u>)	tez (<i>fast</i>)
6	PRON	P	OLMOSH(<u>ProN</u>)	bu (<i>this</i>)
7	AUX	II	KO'MAKCHI (<u>Auxiliary</u>)	bilan (<i>with</i>)



No	Sodda gaplar	O'zbek tilidagi POS tegging	Ingliz tilidagi POS tegging	Gapning sintaktik qolipi
1	Poyezd keldi	N+VB di	N+V	E+K
2	Kitobni o'qidi	N ni+VB di	N+V	T +K
3	Dalaga boryapman	N ga+VB yap man	N+V	H+K
4	Men taqdimNga kechikdim	P+N ga+VB di m	PRON+N +V	E+H+K
5	Ali kimyo kitobini o'qiyapti	N+N+N+ni+VB yapti	N+N+N+V	E+A+T+K
6	Ali qo'shiqni kuyladi	N+N ni+VB di	N+N+V	E+T+K
7	Ali bu yomon qo'shiqni kuyladi	N+P+JJ+N ni+ VB di	N+PRON+ADJ+N+V	E+A+A+T+K

Ushbu O'zbek tilida gap bo'laklari tartibini aniqlash va noto'g'ri tartibdagi sodda gaplarni to'g'ri tartibga keltirish uchun ishlab chiqilgan web sayt qayerlarda qo'llaniladi va kimlar foydalanishi mumkin.





Foydalanilgan adabiyotlar:

1. Abdullaev F. O'zbek tili grammatikasi. -Toshkent, 1976.
2. G'ulomov A, Asqarova M. Hozirgi o'zbek adabiy tili, sintaksis.-Toshkent, 1965
3. Ona tili: o'rta maktablarning 8-sinfi uchun darslik / M. Qodirov va boshqalar. – Toshkent, 2014
4. Rahmatullayev Sh. Hozirgi adabiy o'zbek tili. – Toshkent, 2006.
5. Xolmanova Z. Kompyuter lingvistikasi. -Toshkent, 2020.

MATN MA'LUMOTLARIDAN KERAKLI MA'LUMOTLARNI OLISH

Primova Mastura Hakim qizi

primovamastura@navoiy-uni.uz

ToshDO'TAU o'qituvchisi

Annotatsiya. Matnli ma'lumotlarni intellektual tahlil qilish – strukturlanmagan ma'lumotlarni tahlil qilish va qayta ishlashning eng muhim usullaridan biri bo'lib, dunyodagi ma'lumotlarining qariyb 80 foizini tashkil qiladi. Bugungi kunda ko'pchilik tashkilot va muassasalar ma'lumotlar omborlari va bulutli platformalarda katta hajmdagi ma'lumotlarni to'playdi, saqlaydi. Har soniyada bir nechta manbalardan turli xil turdagi yangi ma'lumotlar axborot tizimiga qabul qilinadi va ular eksponent ravishda o'sishda davom etadi. Bugungi kunda katta hajmdagi ma'tnli ma'lumotlarni qayta ishlash NLPning dolzarb vazifasi hisoblanadi. NLP da har bir jumla bilan bog'liq ma'noni ajratib olish, uni kompyuter tushunadigan shaklga aylantirish uchun bir nechta mashina o'rganish algoritmlaridan foydalaniladi. Ushbu maqolada text mining, text analytics va text analysis farqlari haqida keltirilgan.

Abstract. Text mining is one of the most important methods for analyzing and processing unstructured data, accounting for about 80% of the world's data. Today, most organizations and institutions collect and store large amounts of data in data warehouses and cloud platforms. Every second, new data of various types from several sources enters the information system, and it continues to grow exponentially. Today, processing large volumes of text data is a pressing NLP task. NLP uses several machine learning algorithms to extract the meaning associated with each sentence and convert it into a form that a computer can understand. This article discusses the differences between text mining, text analytics, and text analytics.

Аннотация. Интеллектуальный анализ текстовых данных — один из важнейших способов анализа и обработки неструктурированных данных, на долю которого приходится около 80% мировых данных. Сегодня большинство организаций и учреждений собирают и хранят большие объемы данных в хранилищах данных и облачных платформах. Каждую секунду в информационную систему поступают новые данные разного типа из нескольких источников, и они продолжают расти в геометрической прогрессии. Сегодня обработка больших объемов текстовых данных является актуальной задачей НЛП. НЛП использует несколько алгоритмов машинного обучения для извлечения значения, связанного с каждым предложением, и преобразования его в форму, понятную компьютеру. В этой статье обсуждаются различия между интеллектуальным анализом текста, текстовой аналитикой и анализом текста.

Kalit soʻzlar: *Text Mining, Text analytics, text analysis, NLP, Concordance, Word frequency va Collocation.*

Hozirgi kunga matnli maʼlumotlar bizni qamrob olgan. Uygʻonganimizdan koʻp oʻtmay, odatda matnli xabarlar, elektron pochta xabarlar va ijtimoiy media xabarlar koʻrinishidagi katta hajmdagi maʼlumotlarni koʻrishimiz mumkin. Koʻp tashkilot va muassasalar maʼlumotlar omborlari va bulutli platformalarda katta hajmdagi maʼlumotlarni toʻplaydi va saqlaydi. Shuningdek, tashkilotlarda koʻplab manbaalar orqali: *ilovalar, veb-sahifalar, ijtimoiy media, mijozlar sharhlari* va *qoʻngʻiroq transkriptlaridan* toʻplangan maʼlumotlar katta hajmdagi matnli maʼlumotlarni tashkil qiladi. Bunday katta hajmli maʼlumotlardan kerakli maʼlumotlarni topib olish, tahlil qilish va qayta ishlash juda murakkab vazifalardan biri hisoblanadi. Bunday matnli maʼlumotlardan *kerakli* va *dolzarb maʼlumotlarni* olish uchun **text mining** jarayonidan va **text analysis** dasturi yordamida amalga oshiriladi.

Text mining – matnli maʼlumotlarni intellektual tahlili boʻlib, strukturlanmagan katta hajmdagi matn maʼlumotlarni tahlil qilish orqali, yangi maʼlumotlarni olish jarayonidir. Unda matn maʼlumotlardagi *faktlar, tendensiyalar, qonuniyatlar, konsepsiyalar, kalit soʻzlar* va *boshqa elementlarni* aniqlashga yordam beradi. Tashkilotlarda strukturlanmagan maʼlumotlarni strukturlangan maʼlumotlarga aylantirib, matnli hujjatlarning katta toʻplamlarini tahlil qiladi. Bu jarayon **text analysis** deb ham ataladi.

Text mining quyidagi vazifalarini qamrab oladi: matnni tasniflash, matnni klasterlash, taksonomiya yaratish, hujjatlarni umumlashtirish, obyektlarni ajratib olish va hissiyotlarni (sentiment) tahlil qilish.

Text mining bir nechta metodologiyalardan foydalanadi, ulardan biri *tabiiy tilni qayta ishlashdir (NLP)*. **Tabiiy tilni qayta ishlash (NLP)** - bu tabiiy tildan foydalangan holda odamlar va kompyuterlar oʻrtasidagi oʻzaro aloqani taʼminlab beradi. Ular kompyuter fanlari, tilshunoslik, maʼlumotlar fanlari va sunʼiy intellektning bir boʻlimi hisoblanib, kompyuterlarga inson tilining yozma va ogʻzaki shakllarini tushunishga imkon beradi. Olimlarning fikriga koʻra, 2025-yilga kelib NLP dunyo bozorida 43 milliard dollargacha baholanadi deb bashorat qilinadi.

Hozirgi kunda NLP ga asoslangan koʻplab ilovalar mavjud: Siri, Alexa va Google Assistant kabi ovozi yordamchilar soʻrovlarni tushunish va javoblar tayyorlash, Grammarly jummalarning grammatik toʻgʻriligini tekshirish va Google Tarjimon ham NLP orqali amalga oshirilgan.

NLPda har bir jumla bilan bogʻliq maʼnoni ajratib olish, uni kompyuter tushunadigan shaklga aylantirish uchun bir nechta mashina oʻrganish algoritmlaridan foydalaniladi. Bu vazifalarni bajarishda ikkita asosiy usullar bor: *semantik tahlil* va *sintaktik tahlil*.

Semantik tahlil- kompyuterlarga jumalar kontekstini inson tushungandek, tushunish imkonini yaratadi, yaʼni inson tilini tushunish jarayonidir. *Semantik tahlil*

semantikaga asoslanadi - matn orqali berilgan ma'noga asoslanadi. Unda gapning matn elementlarini aniqlab, ularga grammatik va semantik rolini belgilaydi. Natijada, bir nechta so'zlarga ega bo'lgan so'zlarning ma'nosini aniqlash uchun atrofdagi matnning kontekstini tahlil qiladi.

Sintaktik tahlil - tilning grammatik qoidalarga qanchalik rioya qilishini aniqlashda ishlatiladi. U sintaksisga asoslanib, bu tilshunoslikning jumladagi so'zlarni grammatik ma'noga ega bo'lishi uchun tartibga solish qoidalari bilan shug'ullanadi.

NLPda qo'llaniladigan ba'zi sintaktik usullar:

- *Nutqni belgilash qismi (Part-of-speech tagging):* har bir so'z uchun nutq qismini aniqlash
- *Jumlaning uzilishi (Sentence breaking):* katta matn bo'lagidan jumla chegaralarini topish
- *Morfologik segmentatsiya (Morphological segmentation):* so'zlarni individual qismlarga bo'lish
- *So'zlarni segmentatsiyalash (Word segmentation):* uzluksiz matnning katta qismlarini kichikroq alohida bloklarga bo'lish
- *Lemmatizatsiya (Lemmatization):* qulay tahlil qilish uchun so'zning flektiv shakllarini birlik shaklga o'tkazish
- *O'zaklash (Stemming):* flektiv so'zlarni ildiz shakllariga kesish
- *Sintaktik tahlil:* Gapning grammatik tahlilini bajarish.

Text mining nima uchun muhim?

Ba'zi tashkilotlar katta hajmdagi matnli ma'lumotlarni to'plash imkoniyatiga ega. Misol uchun, mijozlarning sharhlari, mahsulot sharhlari va ijtimoiy media xabarlar orqali to'plangan ma'lumotlar katta hajmli ma'lumotlarning bir qismi hisoblanadi. Ushbu ma'lumotlar tashkilotlarga yuqori daromad keltiradi va foydalanuvchilarning talablariga mos tarzda qulayliklar yaratib berishi mumkin. Bunday ma'lumotlarning 80% strukturlanmagan ma'lumotlar hisoblanadi. Shuningdek, tashkilotlarga bir qancha noqulayliklar vujudga keladi. *Text mining* orqali matnlarni tasniflash (klassifikatsiyalash) jarayonini avtomatlashtirishga yordam beradi. Tasniflash bir nechta atributlarga asoslangan bo'lishi mumkin: *mavzu (topic)*, *niyat (intent)*, *hissiyot (sentiment)* va *til (language)*.

Foydalanuvchiga taklif qilinayotgan dasturiy ta'minot haqida qanday munosabatda ekanligini bilish kerak. Foydalanuvchi sharhlarini qo'lda ko'rib chiqish mumkin, ammo minglab sharhlar bo'lsa, *zerikarli jarayon* va *ko'p vaqt* talab qiladi. Bunday muammoni bartaraf etish uchun text mining yordamida bajarish mumkin. Text mining katta va murakkab ma'lumotlar to'plamini tez va oson tahlil qilib, ulardan kerakli ma'lumotlarni olish imkonini beradi. Bunday holda, text mining mahsulotning umumiy hissiyotini aniqlash imkonini beradi. Sharhlarning ijobiy, salbiy yoki neytral ekanligini aniqlash jarayoni *hissiyotlarni tahlil qilish* yoki *fikrni aniqlash* deb ataladi. Text mining orqali foydalanuvchilarga nimani yoqtirishi,

yoqtirmasligini yoki dasturning keyingi yangilanishga nimani kiritilishini xohlashini aniqlash uchun ishlatilishi mumkin. Foydalanuvchilarning muayyan mahsulotlar yoki mavzular bilan bog'liq holda foydalanadigan kalit so'zlarni aniqlashda ishlatiladi.

Shuningdek, tashkilotlar matn ma'lumotlarini chuqurroq o'rganishi, tegishli biznes tushunchalarini aniqlash yoki matnlar ichida qidiruv tizimlarini, an'anaviy ilovalar bilan aniqlanmaydigan o'zaro aloqalarni aniqlash uchun text mining vositalaridan foydalanishlari mumkin.

Text miningdan tashkilotlar foyda olishlari mumkin bo'lgan ba'zi bir aniq usullar:

- Farmatsevtika sanoatida maxfiy ma'lumotlarni ochish orqali, dori vosiolarini yanada rivojlantirish;
- Ishlab chiqaruvchi tashkilotlar real vaqt rejimida mijozlarning fikr-mulohazalarini tahlil qilishlari va darhol e'tibor berishni talab qiladigan mahsulotdagi xatolar yoki kamchiliklarni aniqlashlari;
- Kompaniyalar tuzilgan ma'lumotlarni yaratishi, ularni ma'lumotlar bazalariga integratsiyalashi va undan tavsiflovchi yoki bashoratli tahlil kabi katta ma'lumotlar tahlilining har xil turlari uchun foydalanishi.

Muxtassar qilib aytganda, text mining tashkilotlarga ma'lumotlarni ishga tushirishga, foydalanuvchilarni xursand qiladigan va natijada daromadlilik oshiradigan ma'lumotlarga asoslangan qarorlar qabul qilishga yordam beradi.

Text mining, text analytics va text analysis

Text mining va **text analysis** ko'p hollarda sinonim sifatida ishlatiladi. Biroq, *text analysis* ikkalasidan farq qiladi.

Text analytics foydalanuvchilarga strukturlangan va strukturlanmagan matn ma'lumotlaridan ma'lumot olish imkonini beruvchi text analysis yoki text mining dasturiy ta'minoti sifatida ishlatiladi. Text mining va text analytics bir xil muammoni hal qiladi, ya'ni “xom” matn ma'lumotlarini tahlil qiladi. Ammo ularning natijalari sezilarli darajada farqlanadi. *Text mining* matn ma'lumotlaridan **sifatli natijalar** deb hisoblanishi mumkin bo'lgan kerakli ma'lumotlarni chiqaradi. Boshqa tomondan, *text analytics* **miqdoriy natijalar** sifatida ko'rish mumkin bo'lgan katta hajmdagi matn ma'lumotlaridagi tendensiyalar va qonuniyatlarni aniqlashga qaratilgan.

Boshqacha qilib aytganda; *text analytics* katta hajmdagi matn ma'lumotlarini tahlil qilish orqali grafik va jadvallar kabi vizual hisobotlarni yaratish uchun mo'ljallangan. Shuningdek, *text mining* tahlilni osonlashtirish uchun strukturlanmagan ma'lumotlarni strukturlangan ma'lumotlarga aylantiradi.

Text mining - bu misollardan o'rganish va yangi ma'lumotlardan natijalarni bashorat qilish mumkin bo'lgan modellarni yaratish uchun statistika, tilshunoslik va mashinani o'rganishga tayanadigan ma'lumotlar data mining kichik to'plami. Text

analytics ma'lumotlarni vizuallashtirish uchun text mining modellari tomonidan olingan ma'lumotlardan foydalanadi.

Text mining usullari

Matn ma'lumotlaridan kerakli ma'lumotlarni olish uchun ko'plab text mining usullaridan foydalaniladi. Quyida bir nechta eng mashhur usullar keltirilgan:

- **Konkordans (Concordance)**

Concardas so'z yoki so'zlar to'plamini kontekstni aniqlashda ishlatiladi. Ya'ni “so'z” inson tilida turli xil ma'nolarni anglatishi sababli, so'zlarni moslashtirish tahlili kontekstga qarab so'zning aniq ma'nosini tushunishga yordam beradi.

- **So'z chastotasi (Word frequency)**

So'z chastotasi strukturlanmagan matn ma'lumotlarida so'zning necha marta takrorlanganini aniqlashda ishlatiladi. Misol uchun, mijozlar sharhlarida "bugs", "errors", "failure" kabi so'zlarning mavjudligini tekshiradi va bu atamalar tez- tez uchrab turishi orqali mahsulotlarni yangilab boriladi.

- **Ibora (Collocation)**

Ibora - ko'pincha birga kelgan so'zlar ketma-ketligi. “Decision making”, “time-consuming” va “keep in touch” ba'zi misollarni ko'rish mumkin. Iborani aniqlash matn tafsilotlarini yaxshilashi va matnni yaxshiroq tahlil qilish natijalariga olib keladi.

Foydalanilgan adabiyotlar:

1. Aggarwal, C. C., & Zhai, C. X. (2013). An introduction to text mining. In *Mining Text Data* (Vol. 9781461432234). https://doi.org/10.1007/978-1-4614-3223-4_1
2. Kao, A., & Poteet, S. (2005). Text mining and natural language processing: introduction for the special issue. *SIGKDD Explor Newsl*, 7(1).
3. Salloum, S. A., Al-Emran, M., Monem, A. A., & Shaalan, K. (2017). A survey of text mining in social media: Facebook and Twitter perspectives. *Advances in Science, Technology and Engineering Systems*, 2(1). <https://doi.org/10.25046/aj020115>
4. Zanini, N., & Dhawan, V. (2015). Text Mining: An introduction to theory and some applications. *Research Matters: A Cambridge Assessment Publication*, 19.

MATNLARNING ASPEKTGA ASOSLANGAN HISSIY TAHLILI

Allanazarova Saboxat Yusupboyevna

allanazarova.sabosh@gmail.com

ToshDO‘TAU tayanch doktoranti

Annotatsiya. Hissiyotlarni tahlil qilish biznes va jamiyat uchun muhim ahamiyati tufayli Tabiiy tilni qayta ishlash (NLP) ning eng faol sohalaridan biri bo‘lib kelgan. Bu fikrning (ijobiy, salbiy, betaraf) bahosini aniqlashga qaratilgan tadqiqot sohasi. Hissiy (sentiment) tahlil qilishning turli darajalari bo‘lib, biz ushbu maqolada eng qiyin turi bo‘lgan aspekt darajasidagi tahlilga e’tibor qaratamiz.

Abstract. Sentiment analysis has been one of the most active areas of Natural Language Processing (NLP) due to its importance to business and society. This is a field of research aimed at determining the value of an opinion (positive, negative, neutral). There are different levels of sentiment analysis, and in this article we will focus on the most difficult type, aspect level analysis.

Аннотация. Анализ тональности был одной из наиболее активных областей обработки естественного языка (НЛП) из-за его важности для бизнеса и общества. Это область исследований, направленная на определение ценности мнения (положительного, отрицательного, нейтрального). Существуют разные уровни анализа настроений, и в этой статье мы сосредоточимся на самом сложном типе — анализе на уровне аспектов.

Kalit so‘zlar: *Tabiiy tilni qayta ishlash, ijobiylik, betaraf, salbiylik, hissiy tasnif, fikrlarni o‘rganish.*

Hissiyot tahlili katta hajmdagi umumiy fikrlarni avtomatik saralab, tushunish, tezkor xulosa chiqarish imkonini beradi. Ya’ni qiyinchiliklarni oddiy misollar orqali ko‘rsatish mumkin bo‘ladi. Sentiment (hissiy) tahlil - matndagi emotsional rangdagi so‘z boyligini avtomatlashtirilgan tarzda aniqlash uchun ishlatiladi. Hissiy tahlil quyidagi darajalarda amalga oshiriladi:

1. *Hujjat darajasidagi tahlil.* Matnda ifodalangan butun fikrlarni tasniflashdir. Masalan, biror bir mahsulot yoki xizmat haqida qoldirilgan sharhni tahlil qilish. Biror matnni boshqa matnga qiyoslamasdan to‘liq matn darajasida tahlil qiladi, bitta hissiy baho bitta sharh/matn uchun tegishli bo‘ladi.

Yangi versiyasi juda ham qulay ekan. Mb kam ketishi bilan boshqa applardan farqi bor ekan. Ishingizga omad sizlardan yangidan yangi xabarlar kutamiz. (ijobiy)

2. *Gap darajasidagi tahlil.* Har bitta gapga alohida hissiy baho beriladi. Bu tahlil turi subyektivlik tahlili uchun amalga oshiriladi.

1) *Men telefon sotib oldim.* (obyektiv)

2) *Restoranning ovqatlari ajoyib ekan.* (Subyektiv)

3. *Aspektga darajasidagi tahlil.* Bir xil so‘zning turli gaplardagi ma’nosi qarama-qarshi bo‘lishi mumkin yoki bir xil jumladagi bir xil shaxsning qutbliligi

turli jihatlarda farqlanadi. Shu sababli, matndan ma'lum obyektlar va ularning jihatlarini haqidagi fikrlarni aniqlash uchun aspekt darajasidagi hissiyotlarni tasniflash taklif etiladi.

Mehmonxonaning ovqatlari shirin ammo, narxi qimmat.

Aspekt darajasidagi hissiyotlarni tahlil qilishda muhim muammolardan biri jumlar o'rtasidagi semantik munosabatlarni qanday modellashtirishdir. Muammoni hal qilish uchun ko'plab usullar taklif qilingan [Rahman, 2013: 6]. So'nggi yillardagi tadqiqotlarda aspekt darajasidagi hissiyotlarni tahlil qilishda neyron tarmoqlar (NN) ustunlik qilmoqda. Aspekt darajasidagi hissiyotlarni tasniflash ko'pincha adabiyotlarda tasniflash muammosi sifatida qaraladi. Mashinaviy o'qitish ham hissiyotlarni tasniflashning asosiy usuli hisoblanadi. Bu usullar, asosan, matn ko'rinishini va xususiyatlarni ajratib olishni masalan, so'zlar to'plami modellari va hissiyot leksikasining xususiyatlari, so'ngra qo'llab-quvvatlash vektor mashinasi (SVM) va logistik regressiya (LR) kabi hissiyotlar tasniflagichini o'rgatishni o'z ichiga oladi. Bunda aspekt atamaları va aspekt toifalarini aniqlash, shuningdek, mijozlar sharhlarida atamalar va tomonlar toifalariga nisbatan bildirilgan his-tuyg'ularni aniqlash uchun boshqariladigan mashinaviy o'qitish usullari tavsiflanadi. Biroq, an'anaviy mashinaviy o'rganish usullari odatda ko'plab muhim xususiyatlar kontekstini modellashtira olmaydi.

Hissiyotlarni aspekt bo'yicha tahlil qilish muammosida odatda matn bitta obyektning baholashni o'z ichiga oladi deb taxmin qilinadi, aslida obyektning turli qismlari VA / YOKI atributlari bilan bog'liq holda ko'rib chiqilishi tushuniladi. Aspekt bo'yicha his-tuyg'ularni tahlil qilish uchun odatiy matndagi (bu ba'zi mahsulot yoki xizmatlar haqida foydalanuvchilar qoldirgan sharhlar bo'lishi mumkin) aspektlarni toifalarga guruhlash mumkin. Aspektga asoslangan tahlilda obyektning biror bir jihatiga qaratilgan bo'ladi. Masalan, oshxonalarning xizmat ko'rsatish sifati, yoki ovqatlari tanlab olinishi mumkin. Shuningdek, sharhlarda siz obyektning umumiy bahosini topishingiz ham mumkin: ajoyib restoran. Matndagi biror jihatga ishora qilish uchun ishlatiladigan so'z va iboralar aspekt atamaları deyiladi. Restoranlar uchun bu odatda oshxona, ovqat, likobcha, xizmat ko'rsatish, interyer (mebellar) kabilar.

Tovarlar yoki xizmatlar uchun belgi/xususiyatlarni aniqlashni avtomatlashtirish. Aspekt terminlari sifatida ko'pincha otlar va otli birikmalari ko'rib chiqiladi [Asuncion, 2009: 27]. Ismlar guruhining uzunligi 3-4 so'zdan oshmasligi kerak. Shu bilan birga, faqat individual otlar aspekt sifatida ajratilsa, ular ko'pincha noaniq bo'lishi mumkin, bu mutaxassislar tomonidan xato talqinga olib keladi [Боронцов, 2014: 676]. Matnlardan aspekt atamalarini ajratib olishni avtomatlashtirishning to'rtta asosiy yondashuvi mavjud:

1. Otlar va ot guruhlariga asoslangan yondashuv;
2. Baholovchi iboralar va aspekt terminlari o'rtasidagi munosabatlardan foydalanadigan yondashuv;
3. Nazorat ostida mashinaviy o'qitish yondashuvi;

4. Statistik modellarga asoslangan yondashuv.

Shunday qilib, aspektlar bo'yicha hissiyotlarni tahlil qilish vazifasi quyidagi kichik vazifalarni o'z ichiga oladi:

- Aspektni ifodalovchi atamalarni tanlash;
- Aspekt atamalari orqali toifalarga ajratish;
- Tanlangan toifalarga nisbatan aspektlarni avtomatik aniqlash.

Mavzu sohasidagi aspekt atamalarini bir necha jihatdan tasniflash mumkin. Aspekt atamalarining eng keng tarqalgan turi - obyekt, uning qismlarini yoki xususiyatlarini aniq nomlaydigan, matn muallifi tomonidan baholanadigan, masalan, osh, xizmat, zal, restoran sharhlarida aniq aspekt terminlaridir. Aniq aspekt terminlari ko'pincha otlar yoki ot guruhiga mansub so'zlar bilan ifodalanadi, lekin ba'zi jihatlar fe'llar bilan ham ifodalanishi mumkin, masalan, restoranlarda xizmat ko'rsatish sifatini baholashda kutish, kelmaslik, kuttirish kabi birliklar uchraydi.

Aspekt atamalarini aniqlashning ikkinchi turiga yashirin aspekt atamalari kiradi, ular bir vaqtning o'zida muhokama qilinayotgan jihatni (odatda ancha umumlashgan jihat kategoriyasi) ko'rsatadigan ma'noning aniq baholovchi komponentiga ega bo'lgan so'zlardir, masalan: **mazali** (ijobiy + taom) restoran sharhlarida), **qulay** (avtomobil sharhlarida ijobiy + qulaylik). Boshqa baholovchi so'zlar singari, noaniq jihat atamalar deb ataladigan so'zlar bilan birlashtirilishi mumkin. O'z bahosini o'zgartiradigan yoki kuchaytiradigan baholash modifikatorlari: **juda** mazali emas, **uncha** qulay emas.

Avtomatik his-tuyg'ularni tahlil qilish tizimlarining lug'atlari uchun bunday aspektli atamalarining ahamiyati shundaki, ular tan olinmagan holatlarda aniq atamalar muallifi (matbaa xatolari, yangi lug'at, murakkab havolalar) yashirin aspekt shartlari foydalanuvchining ba'zi jihatlar toifasiga nisbatan pozitsiyasini ajratib olish imkonini beradi.

Ma'lum bir obyektning ba'zi bir xususiyati haqida o'z fikrini ifodalashning uchinchi turi - bu sodir bo'lgan salbiy yoki ijobiy fakt to'g'risidagi xabar, bu bir vaqtning o'zida ham aspekt toifasini, ham uning foydalanuvchi tomonidan baholanishini (baholash faktlari) ko'rsatadi. Baholash faktlarining turlaridan biri bu sharhlarda qayd etilgan texnik muammolardir [Girolami, 2003: 433]. Ingliz tilidagi texnik muammolarga havolalar ko'pincha quyidagilarni o'z ichiga oladi:

- *Biror narsa sodir bo'lganligini bildiruvchi maxsus fe'llar to'plami* (fail-muvaffaqiyatsizlik, halokat-crash, ortiqcha yuk-overload, sayohat-trip, tuzatish-fix, tartibsizlik-mess, buzish-break, ortiqcha zaryadlash-overcharge, vayron qilish-distrupt);

- *Biror narsa sodir bo'lmaganligini bildiruvchi fe'llar to'plami* va ko'pincha bu fe'llar inkorlar bilan, shuningdek, fe'llar bilan (to'xtash, rad etish, to'xtatish va boshqalar) bilan tilga olinadi;

- *shuningdek, ot va iboralar.*

Ma'no jihatdan yaqin bo'lgan baholovchi faktlar matnda turli yo'llar bilan ifodalanishi mumkin, bu esa ularni aniqlashni qiyinlashtiradi. Matnda tez-tez

uchraydigan lekin baholanmagan soʻzlar, ijobiy yoki salbiy konnatativ naʼno ifodalashi mumkin. Ijtimoiy-siyosiy matnlarda salbiy maʼnoga ega boʻlgan bunday soʻzlarga **ishsizlik, inflyatsiya, turgʻunlik** soʻzlari misol boʻla oladi. Restoran sharhlari sohasida **soch, mayonez** soʻzlari salbiy maʼnoga ega, yaʼni matnlarda bunday soʻzlarning paydo boʻlishi matnning ohangi ancha salbiy boʻlishidan dalolat beradi.

Texnik sohada bunday soʻzlar [Lee, 2011: 884] da koʻrsatilganidek, buzilishlarni bildiruvchi soʻzlar (qobiliyatsiz, avariya, ortiqcha yuk, safar, tuzatish, tartibsizlik, tanaffus) hisoblanadi. Shuningdek, “resurs muammosi” deb ataydigan yashirin baholar va jihatlarning yana bir toifasi mavjudligi koʻrsatilgan: **Bu kir yuvish mashinasi juda koʻp suv sarflaydi**. Shunday qilib, suv isteʼmoli bu yerda jihat, suv esa resurs atamasi boʻlib, uni ortiqcha isteʼmol qilish salbiy faktdir.

Baʼzi tadqiqotlarda da resurs atamalarini koʻp sonli miqdor koʻrsatkichlari bilan, shuningdek, isteʼmol feʼllari asosida ajratib olish kerakligi koʻrsatilgan [Wang, 2009: 1982]. Maqolada maʼlum isteʼmol feʼllari, shuningdek, tabiiy resurslar: gaz, suv, elektr, gaz maʼlum sonli algoritm asosida koʻrib chiqiladi.

Mikolov maqolasida restoranlar va avtomobillar haqidagi sharhlar sohasida eng koʻp uchraydigan baholovchi faktlarni koʻrib chiqadi [Mikolov, 2013: 4546]. Restoran sharhlaridagi manbalarga quyidagilar kiradi:

- mehmonlarni qabul qilish vaqti;
- ofitsiantlarning eʼtibori;
- oziq-ovqat bilan bogʻliq boʻlgan uch turdagi resurslar: likobchadagi oziq-ovqat miqdori, menyuda tanlovning kengligi va idishlarning tozaligi;
- restorandagi joy va bepul oʻrindiqlar;
- xizmat koʻrsatish narxi.

Aspekt terminlari sifatida nafaqat otlar, balki otli birikma ham ajratilishi, keyin ot soʻz birikmasining uzunligini toʻgʻri aniqlash uchun qoʻshimcha belgilardan foydalanish kerak.

Xulosa. Matnlarning avtomatik hissiy tahlil qilish vazifasi murakkab muammodir. Hissiyotlarni tahlil qilishning mashhur va ommalashgan turi bu – aspektga asoslangan tahlildir. Matnlarni avtomatik tahlil qilishning dastlabki natijalarini muhandis-linvistik yondashuv asosida ham, mashinani oʻrganish usullari asosida ham tezda olish mumkin.

Foydalanilgan adabiyotlar:

1. Asuncion A., Welling M., Smyth P., Teh Y. W. On smoothing and inference for topic models // Proceedings of the International Conference on Uncertainty in Artificial Intelligence. — 2009. — Pp. 27–34.
2. Girolami M., Kabʼan A. On an equivalence between PLSI and LDA // SIGIRʼ03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval. — 2003. — Pp. 433–434.



3. Lee S. S., Chung T., McLeod D. Dynamic item recommendation by topic modeling for social networks // Information Technology: New Generations (ITNG), 2011 Eighth International Conference on. — IEEE, 2011. — Pp. 884–889.
4. Mikolov T., Sutskever I., Chen K., Corrado G., Dean J. Distributed representations of words and phrases and their compositionality // CoRR. — 2013. — Vol. abs/1310.4546.
5. Rahman M. Search engines going beyond keyword search: A survey // International Journal of Computer Applications. — August 2013. — Vol. 75, no. 17. — Pp. 1–8.
6. Wang C., Blei D. M. Decoupling sparsity and smoothness in the discrete hierarchical Dirichlet process // NIPS. — Curran Associates, Inc., 2009. — Pp. 1982–1989.
7. Воронцов К. В., Потапенко А. А. Регуляризация вероятностных тематических моделей для повышения интерпретируемости и определения числа тем // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 4–8 июня 2014 г.). — Вып. 13 (20). — М: Изд-во РГГУ, 2014. — С. 676–687.

TAHRIRLOVCHI DASTUR (SPELLCHECKER) TEXNOLOGIYASI

Sobirova Nazira G'anijon qizi

nazirasobirova@gmail.com

Alisher Navoiy nomidagi Toshkent davlat o'zbek
tili va adabiyoti universiteti tayanch doktoranti

Annotatsiya. Kompyuter lingvistikasiga doir qator izlanishlar olib borish natijasida mashina tarjimasi, korpus texnologiyasi, kompyuter leksikografiyasi hamda dasturiy ilovalarning lingvistik asoslarini ishlab chiqish, leksikografik matnlarga ishlov berish va lingvostatistik tahlil etish borasida muayyan darajada natijalarga erishib kelinmoqda. Ayniqsa, bugungi shiddatkor davr uchun matnlarni turli jihatdan avtomatik tahrirlash va tahlil qilish, matnlarning avtomatik tarzda morfologik, sintaktik va semantik tahlilini (inglizcha parsing) ta'minlovchi tizimlar yaratish, mashina tarjimasi uchun mo'ljallangan dasturlarni ishlab chiqish muhimlik kasb etadi. Mazkur maqolada avtomatik tahrir hamda matnni avtomatik qayta ishlaydigan dasturlarning avtomatik tarjima sifatini yaxshilashdagi roli, xususan, spellchecker dasturlar, ularning avtomatik tarjimadagi ahamiyati, ushbu dasturlarning ishlab chiqish masalalari muhokama qilingan.

Abstract. As a result of carrying out a number of researches on computer linguistics, some results have been achieved in the field of machine translation, corpus technology, computer lexicography and the development of linguistic bases of software applications, lexicographic text processing and linguostatistical analysis. Especially for today's intense era, it is important to automatically edit and analyze text in various ways, to create systems that provide automatic morphological, syntactic and semantic analysis of text (English parsing), to develop software for machine translation.

This article discusses the role of automatic editing and automatic text processing programs in improving the quality of automatic translation, in particular, spellchecker programs, their importance in automatic translation, and development issues of these programs.

Аннотация. В результате проведения ряда исследований по компьютерной лингвистике достигнуты некоторые результаты в области машинного перевода, корпусных технологий, компьютерной лексикографии и разработки лингвистических основ программных приложений, лексикографической обработки текста и лингвостатистического анализа. Особенно в современную напряженную эпоху важно автоматически редактировать и анализировать текст различными способами, создавать системы, обеспечивающие автоматический морфологический, синтаксический и семантический анализ текста (синтаксический анализ английского языка), разрабатывать программное обеспечение для машинного перевода. В данной статье рассматривается роль программ автоматического

редактирования и автоматической обработки текста в повышении качества автоматического перевода, в частности программ проверки орфографии, их значение в автоматическом переводе, а также вопросы развития этих программ.

Kalit soʻzlar: *avtomatik tahrir, orfografiya, sun'iy intellekt, korrektura, imlo qoidalari, tahlil metodlari, tokenizatsiya, lemmatizatsiya, stemmizatsiya, spellchecker.*

Dunyo tilshunosligida muayyan tillarda yaratilgan ogʻzaki va yozma manbalarni raqamlashtirish hamda ulardan tabiiy tillarni qayta ishlashda foydalanish muhim masalalardan biriga aylandi. Prezidentimiz tomonidan 2020-yil 20-oktabrdagi “Mamlakatimizda oʻzbek tilini yanada rivojlantirish va til siyosatini takomillashtirish chora-tadbirlari toʻgʻrisida”gi Farmoni bilan tasdiqlangan “2020-2030-yillarda oʻzbek tilini rivojlantirish va til siyosatini takomillashtirish Konsepsiyasi”da belgilangan davlat tilining zamonaviy axborot texnologiyalari va kommunikatsiyalariga faol integratsiyalashuvini taʼminlash maqsadida oʻzbek tilini Internet jahon axborot tarmogʻida ommalashtirish, unda munosib oʻrin egallashini taʼminlash, dasturiy mahsulotlarning oʻzbekcha ilovalarini yaratish, oʻzbek tilini oʻrgatuvchi kompyuter dasturlarini keng miqyosda amaliyotga tatbiq qilish hamda oʻzbek tilidagi matnlar tahririga moʻljallangan kompyuter dasturlarini yaratish kabi vazifalarining koʻrsatib berilishi bugungi kun tilshunos-dasturchi mutaxassislariga katta masʼuliyat yukladi. Shu jihatdan, oʻzbek tiliga zamonaviy axborot texnologiyalari vositalarida ishlov berilishiga erishish, buning uchun til korpuslari, elektron tarjimon, tezaurus, orfokorrektor, ularga dastlabki avtomatik ishlov berish vosita (lingvistik analizator)lari hamda ularning lingvistik taʼminotini yaratish singari masalalar dolzarblik kasb etmoqda. [Hamroyeva, Mengliyev, Gʻulomova, 2022].

Avtomatik tahrir - kompyuter lingvistikasining asosiy yoʻnalishlaridan biri boʻlib, u kompyuterdagi matn muharrirlari (Microsoft Word, Exsel, Wordpad, Lexicon) rivojlanishi bilan bogʻliqdir. Avtomatik tahrir deganda, kompyuterga kiritilayotgan matnning mexanik xatolari avtomatik tarzda toʻgʻrilanishi va xato ekanligi haqidagi signallarning foydalanuvchiga taklif etilishi nazarda tutiladi.

Xorijda kompyuter lingvistikasi sohasida matnlarni avtomatik tahlil va tahrir qilish masalasi koʻplab olimlar tomonidan oʻrganilgan [Peterson, James, 1980.] J.Peterson xatolarni tekshiruvchi va ularni tuzatuvchi dasturiy taʼminot (DT)lar borasida soʻz yuritgan. Matnning avtomatik tahlil jarayonini lingvistik aspektda oʻrganish D.Jurafskiy nomi bilan bogʻliq.

Avtomatik tahrirlash yoʻnalishida avtomatik tahrir qiluvchi dasturlar ishlab chiqilgan, ular ruscha, inglizcha, nemischa, fransuzcha, italyancha matnlarni tahrirlashga yoʻnaltirilgan. Hatto tatar, boshqird, qozoq, qirgʻiz, ozarbayjon kabi turkiy tillarga mansub matnlar uchun morfologik va sintaktik analizatorlar yaratilgan, MS WORD sistemasiga avtotahrir paketlari joylashtirilgan. Bunday

tahrirlash dasturlari, asosan, imloni tekshirishga mo'ljallangan. Rus tilini avtomatik tahlil qilish yo'nalishida R.G.Kotov, V.E.Berzon, V.G.Britvin, I.A.Melchuk, A.B.Kuznetsov, L.I.Belyayeva, V.A.Chijakovskiy, G.G.Belonogov, I.S.Duganova, L.L.Iomdin, E.K.Lavoshnikova, A.V.Sokirko, I.M.Nojoy, E.I.Bolshakova, D.V.Klishinskiy, N.N.Leonteva va boshqalar samarali tadqiqot olib borishgan.

Turkiy tillarga mansub matnlarni statistik metodlar asosida lingvistik tahlil qilish sohasida R.G.Piotrovskiy, M.A.Mahmudov, A.A.Piotrovskaya, T.Sadikov ishlari diqqatga sazovor. O'zbek tili materiallari bo'yicha KLga oid dastlabki tadqiqotlarni olib borgan olimlar H.Arziqulov, S.Rizayev va S.Muhamedovlardir. Ular, asosan, KLning statistik tahlil yo'nalishlari bo'yicha nazariy va amaliy ishlarni amalga oshirgan. H.Arziqulov o'zining “Informatika i pererabotka teksta sredstvami vychislitel'noy tekhniki” (“Информатика и переработка текста средствами вычислительной техники”) nomli monografiyasida kompyuter yordamida katta hajmdagi matnlarga ishlov berishning avtomatik tizimi asoslarini tahlil qilgan. S.Muhamedov va R.Piotrovskiy hammuallifligida yozilgan “Injenernaya lingvistika i opit sistemnostatisticheskogo issledovaniya uzbekskix tekstov” (“Инженерная лингвистика и опыт системно-статистического исследования узбекских текстов”) nomli kitob, shuningdek, R.G.Piotrovskiy, T.Sadikovlarning ilmiy izlanishlarida lingvistik modellar, modellash va uning umumiy tamoyillari haqida fikr yuritilgan, o'zbekcha matnlarning kvantitativ modellari keltirilgan. S.Muhamedov va R.Piotrovskiy hammuallifligidagi asarda sun'iy intellekt yaratish va injener lingvistikasi metodlari bilan o'zbekcha nutqni avtomatik qayta ishlash jarayonida foydalanish uchun zarur bo'lgan o'zbekcha matnlarning leksik-morfologik tuzilishi statistik tahlili natijalari keltirilgan. S.Rizayevning “Kibernetika va tilshunoslik” asarida esa tilshunoslikda aniq metodlarning qo'llanilishi, bunda kompyuterdan foydalanish, til va nutq hodisalariga statistik yondashishning sabablari, matnlarni avtomatik qayta ishlash, mashina tarjimai muammolari, shuningdek, harflarning qo'llanilish chastotasini aniqlashda kompyuterning o'rni borasidagi ma'lumotlar keltirilgan. M.K.Ayimbetov statistik metodlardan foydalangan holda turkiy tillar – o'zbek, qozoq va qoraqalpoq tillari morfologiyasini modellash orqali ularning qarindosh tillar ekanligini isbotlab, ushbu tillarning kvantitativ tipologiyasini ishlab chiqqan bo'lsa, X.F.Isxakova turkiy tillarning formal morfologiyasini tadqiq qilish bo'yicha samarali izlanish olib borgan. XXI asr boshiga kelib o'zbek va ingliz tillarida bir qator izlanishlar qilgan, o'zbek KL rivojiga salmoqli ulush qo'shgan A.Po'latov boshchiligidagi o'zbek tilshunosligiga oid masalalarni kompyuter vositasida hal etishga bag'ishlangan bitiruv malakaviy ishlari, magistrlik dissertatsiyalari himoya qilindi, o'nlab ilmiy maqolalar chop etildi, o'quv qo'llanmalar yaratildi va kitoblar nashrdan chiqarildi. Ayniqsa, fe'l turkumini formallashtirish, ularning modelini ishlab chiqish va fe'llar asosida lingvistik ta'min yaratishga bag'ishlangan talay ishlar dunyo yuzini ko'rdi.

So'nggi yillarda M.X.Xakimov rahbarligida “Tarjimon LMX” sistemasi uchun bajarilgan tadqiqot ishlarida o'zbek va rus tillarining leksik-sintaktik tahlili

berilgan. Sh.Xamroyeva, N.Abdurahmonovalar ishlarida aniq metodlarning qo'llanilishi, bunda kompyuterdan foydalanish, til va nutq hodisalariga statistik yondashishning sabablari, matnlarni avtomatik qayta ishlash, mashina tarjimasini muammolari, shuningdek, harflarning qo'llanilish chastotasini aniqlashda kompyuterning o'rni borasidagi ma'lumotlar keltirilgan. M.K.Ayimbetov statistik metodlardan foydalangan holda turkiy tillar – o'zbek, qozoq va qoraqalpoq tillari morfologiyasini modellashtirish orqali ularning qarindosh tillar ekanligini isbotlab, ushbu tillarning kvantitativ tipologiyasini ishlab chiqqan bo'lsa, X.F.Isxakova turkiy tillarning formal morfologiyasini tadqiq qilish bo'yicha samarali izlanish olib borgan. XXI asr boshiga kelib o'zbek va ingliz tillarida bir qator izlanishlar qilgan, o'zbek KL rivojiga salmoqli ulush qo'shgan A.Po'latov boshchiligida o'zbek tilshunosligiga oid masalalarni kompyuter vositasida hal etishga bag'ishlangan bitiruv malakaviy ishlari, magistrlik dissertatsiyalari himoya qilindi, o'nlab ilmiy maqolalar chop etildi, o'quv qo'llanmalar yaratildi va kitoblar nashrdan chiqarildi. Ayniqsa, fe'l turkumini formallashtirish, ularning modelini ishlab chiqish va fe'llar asosida lingvistik ta'min yaratishga bag'ishlangan talay ishlar dunyo yuzini ko'rdi. So'nggi yillarda M.X.Xakimov rahbarligida “Tarjimon LMX” sistemasi uchun bajarilgan tadqiqot ishlarida o'zbek va rus tillarining leksik-sintaktik tahlili berilgan. Sh.Xamroyeva, N.Abdurahmonova va D.O'rinboyevalar izlanishlarida matnlarni tahlil qilish bo'yicha muayyan fikrlar berilgan, bu jihat lingvistik dasturlarni yaratishda lingvistik modullarni ishlab chiqishning muhim ahamiyatga egaligini belgilab bergan Jahon tilshunosligida XX asrning 50-yillaridan til va matn muammolari bilan bog'liq masalalar kompyuter texnologiyalari yordamida hal qilinmoqda. [Abjalova, 2020].

Kompyuter yordamida matnni tahrirlash bosqichlariga quyidagilar kiradi: **fonetik, grafematik, morfologik, sintaktik, semantik**. Bulardan matnni tahrirlashning morfologik bosqichida matn me'yorlashtiriladi. U

- *Tokenizatsiya* (matndagi so'z shakllarini aniqlaydi);
- *Lemmatizatsiya* (so'zlarning lug'atdagi ifodasini aniqlaydi);
- *Stemmizatsiya* (derivativ so'zlarning o'zagini aniqlash) kabi bosqichlarni o'z ichiga oladi.

Morfologik bosqich sintaktik tahlil bosqichi uchun juda muhim sanaladi.

Avtomatik tahrir va tahlil jarayonini quyidagicha tavsiflash mumkin:

Avtomatik tahrir va tahlil modullari	Texnologiya nomi	Dasturiy ta'minot nomi
Orfografik tahrir	Spelling	Spell checker
Grafematik tahrir (tahlil)	Tokenizatsiya/ segmentlash	Tokenizator
Morfologik tahlil	Stemming Lemmatizatsiya	Stemmer morfoanalizator
Sintaktik tahlil	Parsing	Parser

Matnlarni tahrirlash inson tomonidan (***muharrirlik ishi***) yoki kompyuter vositasida (***avtomatik tahrir***) amalga oshirilishi mumkin. Shuni ta'kidlash lozimki, har ikki holatda ham mutaxassisning faoliyati (qobiliyat, temperament, xarakter, irodaviy sifatlar, ijtimoiy holat) muhim ahamiyatga ega. Muharrir matnlarni bevosita tahrir va tahlil qiladi, avtomatik tahlilda esa tilshunos bilvosita qatnashadi, ya'ni dasturning lingvistik bazasini yaratib beradi. [Abjalova, Toshimov, 2011].

Avtomatik tahrirning asosiy uch vazifasi:

* *Orfografiyani* tekshirish – bu vazifa morfologik tahlil yordamida amalga oshiriladi, bunda etalon lug'atdagi baza asosida tekshiriladi.

* *Sintaksisni* tekshirish orfografiyani tekshirishdan ko'ra murakkabroq. Chunki orfografiyani tekshiruvchi dasturlar yaratilgan, lekin norasmiy matnlarning sintaktik jihatdan avtomatik tahriri masalasi haligacha hal etilgani yo'q. Rasmiy matnlarning avtomatik tahriri bo'yicha muhim natijalarga erishilgan.

* *Semantik* tahrirlash inson tafakkurini modellashtirishga asoslangan ***sun'iy intellekt*** tizimi bilan bog'liq ravishda o'rganilmoqda. (Sun'iy intellekt muayyan vazifalarni kompyuter orqali boshqariladigan algoritmdir.)

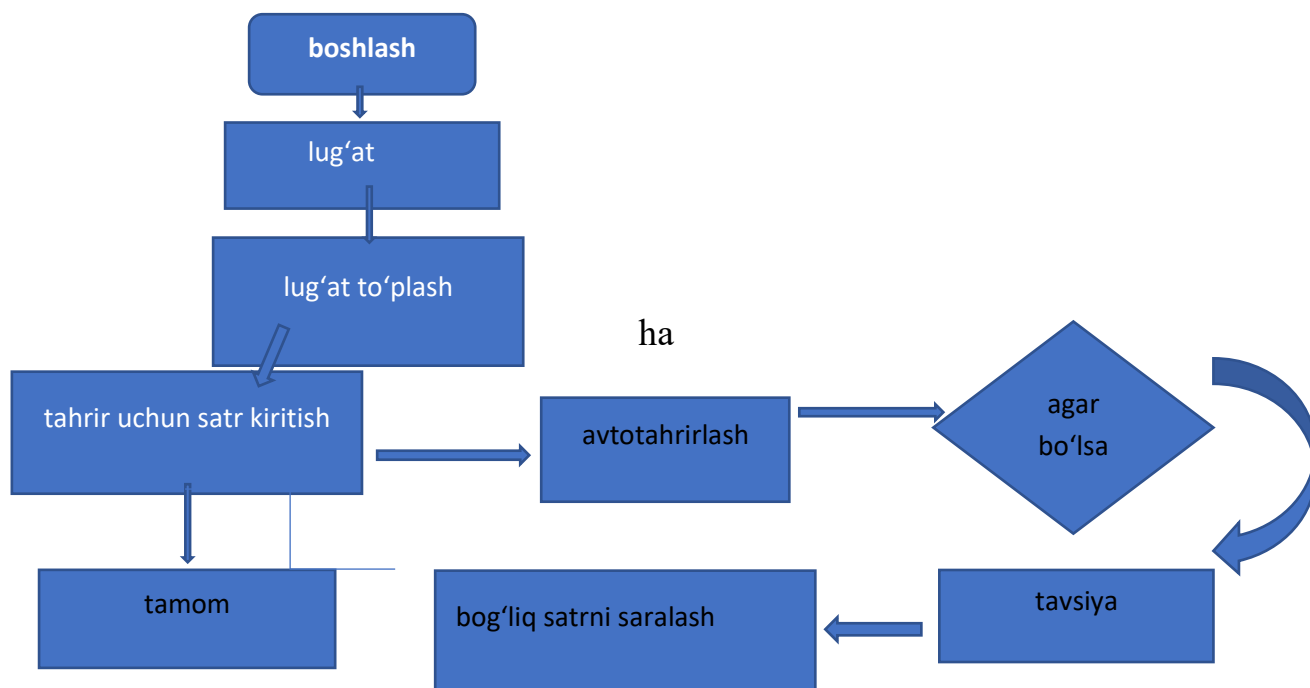
Ixtiyoriy tahrirlovchi dasturning algoritmi quyidagi bosqichlardan iborat bo'ladi:

1. Kompyuter matn tarkibidagi so'zlarni o'qib oladi;
2. Har bir so'z lug'atdagi imlosi bilan solishtiriladi. Leksik va grammatik atributlariga ko'ra so'zlarning lug'ati tuziladi;
3. Ajratilgan segment birliklar morfologik va sintaktik me'yorlarga ko'ra tekshiriladi.

Tahrirlovchi dasturlarning onlayn tizimlari ham yaratilgan, ularning imkoniyatlari bir-biridan farq qiladi, matnni har jihatdan tekshirishga qaratilgan. Hozirda ularning: *Grammarly, Reverso, Ginger software, white Smoke, language Tool, Paper Rater, Pro Writing Aid, Online Correction.com, Hemingway Editor, Spell Check plus, Grammar Slammer, Virtual Writing Tutor, Microsost Word, Google Docs, Slick Write, Grammar Check, WordPerfect Office X8, SentenceChecker.org, After the Deadline, Textly.ai* kabi turlari mavjud.

Matnlarni tahrirlovchi dasturlar matn tarkibida tushirib qoldirilgan yoki ortiqcha qo'llanilgan so'z yoki iborani aniqlash, imloviy va uslubiy xato qo'llanilgan so'zlarni aniqlash, matnning sintaktik strukturasi xatolarni aniqlash, grammatik xato qo'llanilgan shakllarni hamda so'zlarning sinonimlarini taqdim etish orqali semantik imkoniyatini oshirish singari bir qator funksional imkoniyatlarga ega bo'ladi.

Tahrirlovchi dastur algoritmining blok-sxemasini quyidagicha ifodalash mumkin:



Lingvistik tahrirlash (korrektura) – turli ko‘rinishdagi (ilmiy, badiiy, publitsistik va rasmiy uslublardagi) matnlarning orfografik, grammatik, stilistik hamda mantiqiy qurilishdagi xatolarni to‘g‘rilash demakdir.

MC Word dasturida orfografiyani tekshirish imkoniyati cheklanganligi bois hozirgi kunda maxsus *SpellChecker* dasturlar yaratilmoqda. Bunday dasturlar sirasiga quyidagilarni kiritish mumkin: *SpellLink Multilingual Pro* (ingliz, ispan, italyan, nemis, fransuz, portugal, ukrain tillaridagi so‘zlarning to‘g‘ri yozilishini tekshiradi. (*SpellLink -Enlis’h*, *SpellLink French*, *SpellLink Italian*, *SpellLink Spanish*, *Ruta 1.0* rus va ukrain tillaridagi so‘zlarning imlosini tekshirishga yordam beradi), *Asutyp*, *Spellex*, *Spell atche* kabi. Mazkur dasturlar Word dasturiga qo‘shimcha tarzda o‘rnatiladi.

Imloni tekshirish tizimi (spell-cheker ingl. spell checker) – kompyuter dasturi bo‘lib, kiritilgan matnning orfografik tahririni amalga oshiradi. Aniqlangan imlo xatolari maxsus belgilanadi, ya‘ni xato yozilgan leksema tagiga chiziladi. Ko‘p hollarda matn teruvchiga imloviy xatolarga ishora qilishdan tashqari dastur maxsus eslatmasi sifatida so‘zning to‘g‘ri yozilish variantlarini ham taklif qiladi. Shuningdek, matnga qanday tuzatish kiritish mumkinligiga izohlar ham beriladi.

Matnlarni **orfografik tahrirlovchi dastur** hozirda faqat offis dasturlari uchun emas, balki matn bilan bog‘liq boshqa tizimlarda ham keng qo‘llanilmoqda.

Spellcheckerlar odatda quyidagi vazifalarni bajaradi:

Imlo tekshirish: Spellcheckerlar matndagi yozuvning imlo xatolarini aniqlaydi. Bu xatolar o‘rtacha yozuvchilar tomonidan qilinadigan xatolar, harf tartibidagi xatolar, nomlarda yoki kichik harflarda xatolar, yozuv imlo qoidalari va boshqalar bo‘lishi mumkin.

Tavsiyalar berish: Spellcheckerlar imlo xatolarini aniqlagan holda, foydalanuvchiga to‘g‘ri variantlar yoki o‘zgartirishlar taklif etadi. Bu, to‘g‘ri

yozuvni tanlashga yordam beradi va yozuvchilarning yozuvlarini to'g'ri qilishga yordam beradi.

So'zlarni saqlash: Spellcheckerlar odatda o'z ichiga bir nechta so'zlar ketma-ketligini saqlaydi. Bu so'zlarning to'g'ri yozilganligini tekshirishda va foydalanuvchining foydalanishini qo'llab-quvvatlashda yordam beradi.

Foydalanuvchi lug'atini yaratish: Spellcheckerlar foydalanuvchilar o'zlarining so'zlarini qo'shishlari uchun imkon beruvchi o'zlarining lug'atlarini yaratish imkonini beradi. Bu, foydalanuvchi maxsus so'zlari, nomlarni yoki yuridik terminlarni qo'shish imkoniyatini beradi.

Spellcheckerlar uchun qo'llanadigan texnologiyalar va algoritmlar tez-tez rivojlanib borayotgan so'zlar, lug'atlar, imlo qoidalari va texnologiyalarga asoslanadi. Bu, yozuvchilarga yuqori sifatli imlo va yozuvlarni ta'minlashda yordam beradi.

Bugungi kunda jahon bo'ylab spellcheckerlar yuqori darajada ommalashtirilgan vositalar hisoblanadi. Bu vositalar matnlar ustida imlo xatolarini aniqlash va to'g'rilashda yordam beradi. Imlo tekshiruvchilari va spellcheckerlar yozuvning sifatini oshirishga yordam beradigan muhim vositalardir. Ular matnlar ustida yozuv xatolarini aniqlashda va to'g'rilashda yordam beradi va yozuvning o'qilishi va tushunishini osonlashtiradi. Imlo qoidalari va grammatika qoidalari spellcheckerlar va yozuvchilar uchun juda muhim vositalardir. Ular matndagi imlo xatolarini tez fursatda to'g'rilashda ko'maklashadi, shuningdek, yozuvchilar uchun to'g'ri yozuvni ta'minlashda ham yordam beradi.

Matnlarning leksik, morfologik va sintaktik sathlarda tahrir va tahlilini amalga oshiruvchi spell-cheker, orfografik tekshiruv (Speller), orfokorrektorlar, morfoanalizator, parser kabi sistemalarga tildagi matnlarni tahrir va tahlil qilishda asosiy dasturiy vosita sifatida tayanish ish samaradorligini oshiradi.

Hozirgi kunda, spellcheckerlar ko'p platformalarda va dasturlarda o'rtacha va mustahkam bir qismini tashkil etib, matndagi so'zlarning qo'llanishlari hamda imlo xatolarini aniqlash va to'g'rilash usullari eng yaxshi natijalarni olish uchun yanada yaxshilanmoqda.

Foydalanilgan adabiyotlar:

1. Абжалова М.А. Ўзбек тилидаги матнларни компьютер воситасида таҳрир қилиш дастурининг лингвистик таъмини // Магистр илмий даражасини олиш учун ёзилган диссертация// -Тошкент, 2011.
2. Abjalova M.A. Tahrir va tahlil dasturlarining lingvistik modullari /monografiya / Toshkent: Nodirabegim, 2020. – 176 b.
3. Хамроева Шахло Ўзбек тили морфологик анализаторининг лингвистик таъминоти: филол.фан. доктори (DSc)...дис. автореф. – Фаргона, 2021.– 76 б.
4. Rahimov A. Kompyuter lingvistikasi asoslari. -Toshkent, 2011.
5. Po'latov A., Muhamedova S. Kompyuter lingvistikasi.-Toshkent, 2007



6. https://en.wikipedia.org/wiki/Spell_checker
7. https://en.wikipedia.org/wiki/Spell_checker. Peterson, James. Computer Programs for Detecting and Correcting Spelling Errors. 1980.
8. <http://web.stanford.edu/~learnest/les/> Earnest, Les. Visible Legacies for Y3K.
9. <http://en.wikipedia.org>. Computational linguistics.
10. <https://financesonline.com/top-20-grammar-checker-software-solutions/>

MASHINA TARJIMASI UCHUN INGLIZ VA O'ZBEK TILLARIDA GRAMMATIK MA'NO IFODALASHNING O'ZIGA XOSLIKLARI

Qodirova Madinabonu

beginammm@gmail.com

ToshDO'TAU tayanch doktoranti

Annotatsiya. O'zbek tilshunosligida grammatik ma'no, grammatik shakl, grammatik vazifa va grammatik kategoriya tushunchalariga ko'plab murojaat qilingan va ular turli tomondan ko'p bora izohlab berilgan. Ushbu maqolaning maqsadi esa o'zbek va ingliz tillarida Grammatik ma'no hosil qilishning o'ziga xosliklari, o'xshash va farqli jihatlari haqida aniq misollar asosida ma'lumot berish sanaladi.

Abstract. In Uzbek linguistics, the concepts of grammatical meaning, grammatical form, grammatical function, and grammatical category have been referred to many times, and they have been explained many times from different angles. The purpose of this article is to provide information on the peculiarities, similarities and differences of grammatical meaning in Uzbek and English based on specific examples.

Аннотация. В узбекском языкознании понятия грамматического значения, грамматической формы, грамматической функции и грамматической категории неоднократно упоминались и неоднократно объяснялись с разных сторон. Цель данной статьи – на конкретных примерах предоставить информацию об особенностях, сходствах и различиях грамматических значений в узбекском и английском языках.

Kalit so'zlar: *mashina tarjima, grammatik ma'no, grammatik shakl, kategoriya, sintetik va analitik vositalar, suppletivizm, fleksiya, affiksatsiya.*

Matnlarni avtomatik tarjima qilish uchun grammatik shakllarning mos variantini topish muhim sanaladi. Ingliz-o'zbek tarjima yo'nalishida ham bu masala hamon dolzarb bo'lib qolmoqda. Ma'lumki, ingliz va o'zbek tillari morfologik tuzilishi, strukturasi bilan bir-biridan keskin farq qiladi.

Grammatika har bir tilning asosi, mavjud bo'lish va rivojlanishi uchun asosiy tayanch nuqtasi sanaladi. Grammatikaning asosiy tushunchalari esa grammatik ma'no, grammatik shakl va grammatik kategoriya sanaladi.

Grammatik ma'no so'zning tilshunoslik bilan bog'langan, tilning grammatik xususiyatlarini qamrab oluvchi umumiy, mavhum ma'nodir. Grammatik ma'no leksik ma'noga bog'liq bo'lib, obyektiv voqelik bilan bilvosita, leksik ma'no orqali bog'lanadi. U nisbiy bo'lib, so'z shakllari munosabatlarida namoyon bo'ladi: quvnoq - quvnoqroq. Grammatik ma'no barcha birliklarga tegishli va majburiydir [Xolmanova, 2007: 125-127].

Grammatik shakl esa grammatik ma'noni yuzaga chiqaruvchi vositalarni anglatadi. Ularga sintetik va analitik shakllar, melodika, prosodika kabi supersegment birliklar, so'zlarni juftlash va takrorlash orqali hosil qilingan birliklar kiradi.

Grammatik ma'no ifodaning grammatik shakliga ega bo'lishi (masalan, grammatik vositalar, fleksiylar, analitik shakllar, so'z tartibi va boshqalar) yoki ega bo'lmasligi mumkin. Shaklan mavjud bo'lmasa ham til birligi tarkibida Grammatik ma'no saqlanishi mumkin: *kitob - kitoblar*. Ko'plik ma'nosi *-lar* morfemasi bilan ifodalanadi. Birlik ma'nosini ifodalovchi maxsus morfema yo'q. Shuning uchun nol morfema deyiladi.

Til tizimida grammatik elementlar o'rtasida bir vaqtning o'zida o'xshashlik va qarama-qarshilik munosabatlari mavjud bo'ladi. Grammatik shakllar majmuyi paradigmani tashkil qiladi. Bir turkumdagi grammatik shakllarning paradigmatic munosabatlari ular orasida mavjud bo'lgan ziddiyatlar orqali ochib beriladi. Boshqacha qilib aytadigan bo'lsak, grammatik kategoriya til birliklari orasidagi qarama-qarshiliklarining umumiy yig'indisidir [Болкова, 2016: 28].

Grammatik ma'no hosil qiluvchi quyidagi usullarni ajratib ko'rsatish mumkin:

Affiksatsiya. Agglyutativ tillarda eng faol grammatik ma'no hosil qiluvchi usul. So'z asosiga affiks qo'shish orqali grammatik ma'no hosil qilinadi. O'zbek tilida ular ikki turga bo'linadi: lug'aviy shakl yasovchi va sintaktik shakl yasovchi affikslar. Tilimizda bu usul bilan hosil qilingan grammatik ma'nolarga minglab misollarni keltirish mumkin. Masalan, *uy – uylar; kel – kelsang; uzun – uzunroq* kabi. O'zbek tilida grammatik ko'rsatkich so'zning faqat oxiriga qo'shiladi.

Ingliz tilida ham affiksatsiya usuli yordamida grammatik ma'no hosil qilinishi mumkin. [Кожанова, 2021: 194-196]

Affiks	Vazifasi	Misol
-ing	Ingliz tilida davomli zamonlar (Continuous Tenses)ni hosil qiladi.	<i>I am writing a letter – Men xat yozyapman</i>
	Gerundiy shaklini yasaydi. Gerund fe'lining otga xoslangan shakli bo'lib, o'zbek tilidagi harakat nomiga teng keladi. [2]	<i>Reading is a useful for you – O'qish sen uchun foydali.</i>
-(e)s	Oddiy hozirgi zamon (Present Simple) da III shaxs birlik sonni ko'rsatish uchun ishlatiladi. ss, ch, sh, x harflar (sirg'aluvchi tovushlar) bilan tugagan fe'llarga -es shaklida qo'shiladi.	<i>She works at the station – U stansiyada ishlaydi.</i> <i>He teaches 30 students – U 30 talabani o'qitadi.</i>
-(e)s	Otlarga qo'shilib ko'plik ma'nosini	<i>Book – books;</i>

	ifodalaydi.	<i>girl – girls</i>
‘s	Qaratqich kelishigi –ning qo‘shimchasiga teng keladi. Qaratqich kelishigi ko‘plikdagi otlardan keyin faqat apostrof belgisini qo‘yish bilan yasaladi.	<i>Worker’s book –</i> ishchining kitobi <i>the boys’ books –</i> bolalarning kitoblari <i>the workers’ tools –</i> ishchilarning asboblari
-er	Sifatning qiyosiy daraja shakli (the Comparative Degree) ni hosil qiladi.	<i>Long – longer</i> uzun – uzunroq
-est	Sifatning orttirma daraja shakli (the Superlative Degree)ni hosil qiladi.	<i>Long – longest</i> uzun – eng uzun
-th	Tartib sonni hosil qiladi	<i>Fifth</i> - beshinchi <i>Sixth</i> - oltinchi

Intonatsiya. U melodika va urg‘uni o‘z ichiga oladi. Bu vositalar ham tilda grammatik ma‘no ifodalash uchun xizmat qiladi.

Melodika. Melodika deganda, ton balandligining o‘zgarishi va tovushlarning cho‘zilishi tushuniladi. [Anne, 1980:42] Masalan, Xitoy-tibet tillar oilasiga kiruvchi xitoy tilida to‘rtta ton mavjud bo‘lib, so‘zning ana shu tonlardan qaysi biri yordamida talaffuz qilinishiga qarab so‘zning ma‘nosi o‘zgaradi. Masalan, *ma* so‘zi cho‘ziq va baland talaffuz etilishiga qarab *ona*, *kanop*, *chumoli*, *so‘kinmoq* kabi ma‘nolarni ifodalashi mumkin. O‘zbek tilida ham bu usul orqali ma‘no ifodalash holatlari uchrab turadi. Masalan, sifat va ravishlar tarkibidagi unlini cho‘zish orqali orttirma darajasini ifodalash mumkin - *uzu:n*, *bala:nd*.

Urg‘u. O‘zbek va ingliz tillarida urg‘u so‘zlarning leksik va grammatik ma‘nolarini ajratishda ishlatiladi:

do‘stimsi‘z (-siz qo‘shimchasi sifat yasovchi) – *do‘stimsiz* (-siz qo‘shimchasi bog‘lama). *Takror* biror grammatik ma‘noni ifodalash uchun bir so‘zni ikki marta takrorlash bo‘lib, u lingvistikada reduplikatsiya (ikkilantirish) deb yuritiladi. O‘zbek tilida bu usul bilan turli grammatik ma‘nolarni hosil qilish mumkin:

Ichki fleksiya. U turli grammatik ma‘nolarni ifodalash uchun so‘z tarkibidagi tovushlarning o‘zgarishidir. O‘zbek va ingliz tillarida ham bu hodisa uchrab turadi.

Otda	Umumlashtirish	<i>Non-pon, suv-puv</i>
	Sifatlashuv	<i>Yo 'l-yo 'l, katak-katak</i>
Sifatda	Ma'noni kuchaytirish	<i>Baland-baland, uzun-uzun</i>
	Orttirma daraja	<i>Qip-qizil, bo 'm-bo 'sh</i>
Sonda	Taqsim son	<i>Ikkita-ikkita</i>
Olmoshda	Payt	<i>Shu-shu ko'rinmay ketdi.</i>
	Ko'plik	<i>Do'kondan nima-nima olib kelding?</i>
Fe'lda	davomiylik	<i>Yoza-yoza charchadik.</i>

Ingliz tilida [G'aprov, 2010:36]:

Fe'lda	O'tgan zamon (Past Simple) ni yasashda	<i>See - saw</i> <i>Ko'rmoq – ko'rdi</i>
	Sifatdosh (Participle) shaklini hosil qilishda	<i>Do – done</i> <i>Bajarmoq - bajargan</i>
Otda	Ko'plik shaklini hosil qilishda	<i>man — men — erkak</i> <i>woman — women — ayol</i> <i>goose — geese — g'oz</i> <i>foot — feet — oyoq</i> <i>mouse — mice — sichqon</i> <i>ox — oxen — ho'kiz</i> <i>child — children — bola</i>
Sifatda	Daraja shakllarini hosil qilishda	<i>Little – less – least</i> <i>Oz – ozroq – juda oz</i> <i>Good – better – best</i> <i>Yaxshi – yaxshiroq – juda yaxshi</i>
Sonda	Tartib sonni hosil qilishda	<i>First – birinchi</i> <i>Second – ikkinchi</i> <i>Third - uchinchi</i>

Suppletivizm. Suppletivizm deganda, bir so'zning turli grammatik formalarning har xil so'z o'zaklaridan yasalishi tushuniladi.

O'zbek tilida:

Men kishilik olmoshining ko'plik shakli –lar qo'shimchasi bilan emas, “*biz*” so'zi yordamida hosil bo'ladi.

Ingliz tilida ham olmoshlar turlanishida suppletivizmni kuzatish mumkin. Masalan, *I* (men) – *me* (menga).

Analitik vositalar. Ularga tilda mustaqil qo'llana oluvchi, biroq leksik ma'noga ega bo'lmay, faqat grammatik ma'no ifodalaydigan birliklar kiradi.

Bunday soʻzlar boshqa soʻzlarning tarkibiy qismi emasligi bilan affikslardan farq qiladi. Ularga oʻzbek tilidagi quyidagi birliklarni kiritish mumkin:

1. Yordamchi soʻzlar (bogʻlovchi, koʻmakchi, yuklama);
2. Undov va modal, tasdiq va inkor soʻzlar.
3. Koʻmakchi feʼllar (ayta *olmoq*, aytdi-*qoʻydi*, ura *boshlamoq*);
4. Toʻliqsiz feʼllar (*edi*, *ekan*, *emish*, *emas*);
5. Bogʻlama-feʼllar (*boʻlmoq*, *hisoblanmoq*, *sanalmoq*);
6. Funksional soʻzlar. Ular kontekstdan kelib chiqib gapda turli

vazifalarda kelishi mumkin. Masalan, *bir* soʻzi:

Sizni *bir* kishi kutib turibdi – gumon olmoshi

Bir quyosh chiqadi, *bir* yomgʻir yogʻadi – ayiruv bogʻlovchisi.

Bu birliklarning koʻpchiligi alohida olinganda mustaqil soʻz boʻlib, ayrim oʻrinlarda leksik maʼnosini yoʻqotadi.

Ingliz tilida leksik maʼno bildirmaydigan analitik vositalar koʻpchilikni tashkil etadi. Masalan:

1. Yordamchi feʼllar (Auxiliary verbs) - *to be*, *to have*, *to do*, *shall* (*should*), *will* (*would*);
2. Modal feʼllar (Modal verbs) - *can* (*could*), *may* (*might*), *must*, *ought to*, *need*;
3. Artikllar (Articles)– *a*, *an*, *the*
4. Predloglar (Prepositions)– *to*, *in*, *for*, *from*, *of*, *off*;
5. Bogʻlovchilar (Conjunctions) – *and*, *or*, *with*, *although*, *if*
6. Bogʻlovchi soʻz boʻlib kelgan soʻroq olmoshlari: *when*, *what*, *which*, *who*;
7. Bogʻlovchi vazifasida kelgan *that* koʻrsatish olmoshi;
8. *It* soxta ega boʻlib kelganda: *It is winter*. – Qish.
9. Qism-soʻzlar (Particles). Ingliz tilidagi *in*, *up*, *off*, *over* kabi soʻzlar boʻlib, frazemalarni yasash uchun feʼllar bilan ishlatiladi: *put on*, *come up with*.
10. Tasdiq va inkor soʻzlar (Response words) [Mirzayeva, 2023: 43];
11. Undov soʻzlar (Interjunction).

Sintaktik qolip ham soʻz va gaplarning grammatik maʼnosiga taʼsir etadi. Masalan, oʻzbek tilida:

Bu tanti, *mehribon oʻzbek yigiti* – xabar maʼnosi.

Bu oʻzbek yigiti – *tanti*, *mehribon* – taʼkid maʼnosi.

Ingliz tilida:

It is rain – darak gap.

Is it rain? – soʻroq gap.

Shunday qilib, tillardagi grammatik maʼno ifodalash usullari xilma-xil boʻlib, bir tildagi ifoda usullari boshqalaridan farq qiladi. Biroq, shuni taʼkidlash kerakki, tillararo integratsiyalar natijasida bu usullarning koʻrinishlari til xususiyatlaridan qatʼiy nazar istisno tariqasida har qanday tilda uchrab turishi mumkin. Yaʼni agglyutinativ tabiatli tilda faqat affiksatsiya usuli orqali grammatik maʼno hosil

qilinadi, deb xulosa qilib bo'lmaydi. Unda suppletivizm, fleksiya, analitik vositalarni ham uchratish mumkin. Bu esa tillarning murakkab va tasodifiy jarayon ekanligidan dalolat beradi.

Foydalanilgan adabiyotlar:

1. Mirzayeva D.I., Shodiyeva N.G. Ingliz tili morfologiyasida soʻz turkumlarini oʻrganish zaruriyati. // Scientific progress, volume 4 | issue 5, 2023.: <http://sjifactor.com/passport.php?id=22257>
2. Волкова Л.М Лекции по теоритическое грамматике, Киев – 2016, 28 с.
3. Кожанова, Н.В. Грамматика английского языка: учебно-методическое пособие / Н.В. Кожанова, Л.Ф. Дашкевич. – Барнаул: АлтГПУ, 2021. – 194 с.
4. G'apporov M. Ingliz tili grammatikasi. Akademik litseylar va kasb-hunar kollejlari uchun o'quv qo'llanma – T., 2010 – 336 b.
5. Anne C. ERRORS OF STRESS AND INTONATION. Laboratory of Experimental Psychology. University of Sussex. –Brighton, England. 1980.
6. Xolmanova Z.T. TILSHUNOSLIKKA KIRISH. Toshkent -2007, 300-b.

O'ZBEK MATNLARI UCHUN COUNT VECTORIZER IMKONIYATLARI VA KAMCHILIKLARI

Mahmudjonova Gulshahnoz Ulug'bek qizi

gulshahnozmahmudjonova@gmail.com

ToshDO'TAU 2-kurs magistranti

Annotatsiya. Ushbu maqola matn ma'lumotlarini raqamli tasvirlarga aylantirish uchun tabiiy tilni qayta ishlash (NLP)ning asosiy vositasi bo'lgan CountVectorizer haqida umumiy ma'lumot beradi. U mashinani o'rganish vazifalarida qo'llanilishi, soddaligi va samaradorligi kabi afzalliklari, shuningdek, semantik tushunish kabi kamchiliklarni muhokama qiladi.

Abstract. This article provides an overview of CountVectorizer, a key tool in natural language processing (NLP) for converting text data into text representations. It discusses its application in machine learning tasks, advantages such as simplicity and efficiency, as well as limitations such as the lack of semantic understanding.

Аннотация. В данной статье представлен обзор CountVectorizer, ключевого инструмента в обработке естественного языка (Natural Language Processing, NLP), который используется для преобразования текстовых данных в числовые представления. Обсуждаются его применение в задачах машинного обучения, преимущества такие как простота и эффективность, а также ограничения, такие как отсутствие семантического понимания.

Kalit so'zlar. *Matn, hujjat, normallashtirish, scikit-learn, indeks, vektorlar, NLP.*

NLP (natural language processing – tabiiy tilga ishlov berish) – informatikaning inson tilini tahlil qilish, modellashtirish va tushunish usullari bilan shug'ullanadigan sohasi. Tabiiy til (inson tili) ni o'z ichiga olgan har bir intellektual dasturning asosini NLP tashkil etadi. [Elov, Hamroyeva, Xusainova, 2022:20] Insonda rivojlangan har qanday til tabiiy til hisoblanadi. Tabiiy tilni nutq, qo'shiq aytish, yuz ifodasi, belgilar va tana imo-ishoralari kabi turli shakllarda ko'rish mumkin. NLP 1950-yillarda sun'iy intellekt va tilshunoslikning kesishmasi sifatida boshlangan bo'lib, komputer dasturlariga inson tilini tushunish va qayta ishlashga yordam beradi. Bu sohaning ko'plab vazifalari uchun muhim qadamlardan biri matnlarni, ya'ni so'zlarni raqamlashtirishdir. Bu jarayon matnni oldindan qayta ishlash bo'lib, til korpusida toza bo'lmagan (shovqin, turli xil belgilar, URL

manzillarni o'z ichiga olgan) matnni qayta ishlab dastur o'qiy oladigan va ustida turli vazifalarni amalga oshira oladigan holatga olib kelishdir. Til korpusi matniga dastlabki ishlov berish bosqichlari haqida [Elov, Hamroeva, Alayev, Xusainova, Yodgorov, 2023:144] maqolada to'liq ma'lumot olish mumkin.

Count Vectorizer – bu matn ma'lumotlarini mashinani o'rganish algoritmlari tushunadigan formatga aylantirishga yordam beradigan an'anaviy usul. [<https://towardsdatascience.com/basics-of-countvectorizer>]. CountVectorizer – matnli hujjat to'plamini raqamli ifodaga aylantirish uchun NLP vazifalarida keng qo'llaniladi va matn ma'lumotlarini tokenizatsiya qilish va har bir tokenning paydo bo'lishini hisoblash orqali ishlaydi [Neri, 2023]. CountVectorizer – matnli hujjatlar to'plamini raqamli tasvirga aylantirish uchun tabiiy tilni qayta ishlashda keng qo'llanilib, matnni tasniflash, klasterlash va ma'lumotlarni qidirishni o'z ichiga olgan ko'plab NLP vazifalarini qayta ishlash bosqichida asosiy vositadir. CountVectorizer dasturining asosiy g'oyasi hujjat atamasi matritsasini yaratishdan iborat bo'lib, unda har bir satr matn korpusidagi hujjatni, har bir ustun esa korpusda topilgan unikal so'z yoki “termin”ni ifodalaydi. Matritsadaqi qiymatlar har bir hujjatdagi har bir atamaning paydo bo'lish chastotasini ko'rsatadi.

CountVectorizer afzalliklari

Oddiy: CountVectorizerdan foydalanish va tushunish oson. U o'ziga xos parametrlarga ega va matnni oldindan qayta ishlashni boshlash uchun kichik konfiguratsiyani talab qiladi.

Tez va samarador: hisoblash jihatidan samarali va ko'plab hujjatlar bilan katta matnli ma'lumotlar to'plamini boshqara oladi. U kompyuter xotirasini tejaydi va qisqa vaqtda natija qaytaradi, ayniqsa yuqori o'lchamli ma'lumotlar bilan ishlashda. Shuningdek, u nomuhim so'zlarini filtrlaydi va lug'at hajmini nazorat qilish imkoniyatlarini beradi.

CountVectorizerning kamchiliklari

Semantik ma'noni e'tiborsiz qoldiradi: U har bir tokenga alohida obyekt sifatida qaraydi va so'zlar orasidagi semantik munosabatlarni saqlamaydi. U kontekstdagi so'zlar ma'nosini hisobga olmaydi, bu esa so'z semantikasini tushunishni talab qiladigan vazifalarda uning samaradorligini cheklashi mumkin.

Hujjat uzunligini hisobga olmaydi: hajm jihatdan katta bo'lgan hujjatlar qisqaroq hujjatlarga qaraganda ko'proq belgilar soniga ega bo'lsa ham ma'no tarafdin ulardan farq qilishi mumkin. Ushbu normalizatsiya yetishmasligi hujjat uzunligiga tayanadigan maxsus tahlillar va algoritmlarga ta'sir qilishi mumkin.

Cheklangan ma'lumot: faqat hujjatlardagi tokenlarning chastotasini ushlaydi. U his-tuyg'ularni tahlil qilish yoki tilni modellashtirish kabi maxsus matn tahlili vazifalarida tegishli bo'lishi mumkin bo'lgan so'zlarning tartibi yoki ketma-ketligini hisobga olmaydi.

Xato xulosa: Yuqori chastotali so'zlar matn uchun muhimroq ahamiyatga ega ma'lumot beradi degan xato xulosani beradi. “bilan, va, ammo..” kabi nomuhim so'zlarni bunga misol sifatida keltirish mumkin [Elov, Hamroyeva, Xusainova, Xudayberganov, 2023:81].

CountVectorizerdan olingan hujjat muddatli matritsasi talqin qilinadigan natijalarni beradi. Matritsada har bir katak ma'lum bir hujjatdagi tokenning sonini yoki chastotasini ifodalaydi, bu to'g'ridan-to'g'ri tahlil qilish va o'rganish imkonini beradi. CountVectorizer, Pythonning scikit-learn kutubxonasida joylashgan. Ushbu kutubxona ma'lumotlar bilan ishlash, o'rganish, va ma'lumotlar analizi uchun ishlatiladi. CountVectorizer esa matnlar bilan ishlashda o'rtacha bo'lgan bir vositadir, u matn ma'lumotlarini raqamlar bilan ifodalaydi. Batafsilroq tushunish uchun, quyidagi matnlar asosida tahlil qilinadi.

CountVectorizer – bu Pythonning scikit-learn kutubxonasi tomonidan taqdim etilgan. U berilgan matnni butun hujjatda paydo bo'ladigan har bir so'zning chastotasi (soni) asosida vektorga aylantirish uchun ishlatiladi. Bu bir nechta matnlar mavjud bo'lganda va har bir matndagi har bir so'zni vektorlarga aylantirmoqchi bo'linganda Count Vectorizerdan foydali bo'ladi (matn tahlilida foydalanish uchun). Quyida hujjatdagi bir nechta matn misollari orqali ko'rib chiqiladi (har biri (list) ro'yxat elementi sifatida):

hujjat = [“Men dasturlashni yaxshi bilaman.”, “Dasturlash doimiy bilimlar olishni talab qiladi.”, “Men kod yozishni yoqtiraman.”] quyidagi jadvalda hujjat to'plami raqamli ifodaga (matritsa) aylantirilgan bo'lib, ustunlar hujjatlarni, qatorlar esa unikal so'zlarni ifodalovchi matritsa hosil qilgan (2-jadval).

Unikal so'zlar	Hujjat [0]	Hujjat [1]	Hujjat [2]
bilaman	1	0	0
dasturlashni	1	0	0
olishni	0	1	0
men	1	0	1
dasturlash	0	1	0
kod	0	0	1
qiladi	0	1	0
yaxshi	1	0	0
yoqtiraman	0	0	1

bilimlar	0	1	0
doimiy	0	1	0
yo'zishni	0	0	1
talab	0	1	0

1-jadval. Unikal so'zlar

Hujjatda jadval qatorlari sifatida ifodalangan 13 ta unikal so'z mavjud.

1. Hujjatda 3 ta matn namunasi mavjud bo'lib, ularning har biri jadval ustunlari sifatida ko'rsatilgan.
2. Har bir katak o'sha matndagi so'zning sonini ko'rsatadigan raqamni o'z ichiga oladi.
3. Barcha so'zlar kichik harflarga o'tkaziladi.

CountVectorizer ichida bu so'zlar satr, ya'ni so'z sifatida saqlanmaydi. Aksincha, ularga ma'lum bir indeks qiymati beriladi. Bunday holda, “bilaman” indeksi 0, “dasturlashni” indeksi 1, “men” indeksi 3 kabi ifodalanadi. Unikal so'zlar indeksi quyidagi jadvalda ko'rsatilgan (2-jadval).

Indeks	Hujjat [0]	Hujjat [1]	Hujjat [2]
0	1	0	0
1	1	0	0
2	0	1	0
3	1	0	1
4	0	1	0
5	0	0	1
6	0	1	0
7	1	0	0
8	0	0	1
9	0	1	0
10	0	1	0
11	0	0	1
12	0	1	0

2-jadval. Unikal so'zlar indeksi

Python dasturidan foydalanib CountVectorizer ishlatilindi:

```
from sklearn.feature_extraction.text import CountVectorizer
hujjat = [ "Men dasturlashni yaxshi bilaman.",
           "Dasturlash doimiy bilimlar olishni talab qiladi.",
```

“Men kod yozishni yoqtiraman.”

]

CountVectorizer obyektini yaratish

vectorizer = CountVectorizer ()

vectorizer.fit(document)

Belgilangan noyob so'zlarni indeksleri bilan birga chop etish

print (“Vocabulary:”, vectorizer.vocabulary_)

Hujjatni kodlashtirish

vector = vectorizer.transform(document)

Kodlangan matnlarni umumlashtirish

print (“Encoded Document is:”)

print (vector.toarray())

#Natija

Lug‘at: {‘men’: 6, ‘dasturlashni’: 3, ‘yaxshi’: 10, ‘bilaman’: 0, ‘dasturlash’: 2, ‘doimiy’: 4, ‘bilimlar’: 1, ‘olishni’: 7, ‘talab’: 9, ‘qiladi’: 8, ‘kod’: 5, ‘yozishni’: 12, ‘yoqtiraman’: 11}

Kodlangan hujjat:

[[1 0 0 1 0 0 1 0 0 0 1 0 0]

[0 1 1 0 1 0 0 1 1 1 0 0 0]

[0 0 0 0 0 1 1 0 0 0 0 1 1]]

Count Vectorizer o‘ziga xos afzallik va kamchiliklarga ega. Undan foydalanishda matn ustida bajariladigan amalni hisobga olish kerak. Maqolada Count Vectorizerdan sodda dastur uchun sinab ko‘rildi. Keyingi izlanishlarda ixtiyoriy matn tanlanib, unda so‘zlar statistikasini aniqlash kabi amaliyotlar bajariladi.

Foydalanilgan adabiyotlar:

1. Elov B., Hamroyeva Sh., Xusainova Z., “NLP (tabiiy tilga ishlov berish)ning vazifalari va zamonaviy yondashuvlar” Filologik tadqiqotlar: til, adabiyot, ta’lim, 2022. - B. 20
2. Elov, B., Hamroyeva, Sh., Alaev, R., Xusainova, Z., Yodgorov, U., (2023). Methods of processing the Uzbek language corpus texts.
3. <https://towardsdatascience.com/basics-of-countvectorizer>
4. Neri Van Otten., “CountVectorizer Tutorial In Scikit-Learn And Python (NLP) With Advantages, Disadvantages & Alternatives” <https://spotintelligence.com/2023/05/17/countvectorizer/> 2023
5. Elov B, Hamroyeva Sh, Xusainova Z, Xudayberganov N (2023). “O‘zbek tili korpusi matnlarini qayta ishlashda CountVectorizer, TF-IDF hamda Co-occurrence matrix usullarining ahamiyati” Elektron lug‘atlar yaratishning nazariy va amaliy asoslari mavzusidagi xalqaro ilmiy-amaliy anjuman materiallari ., Andijon-2023. – B. 81

N-GRAMMGA ASOSLANGAN STEMMING ALGORITMINI ISHLAB CHIQISH, TATBIQ QILISH VA BAHOLASH

Qosimova Sarvinoz Furqat qizi

sarvinozq54@gmail.com

ToshDO‘TAU magistranti

Annotatsiya. Bugungi kunda tabiiy tilga bog‘liq bo‘lmagan stemmerlarni ishlab chiqish muhim ahamiyat kasb etadi. Tokenizatsiya jarayonining N-gramm usuli ba’zi hollarda stemlarni noto‘g‘ri aniqlashi mumkin. Shu sababli, ushbu N-gramm usuliga biroz o‘zgartirishlar kiritish orqali, yangi stemmer ishlab chiqildi. Ushbu maqola n-grammlardan foydalangan holda innovatsion stemming algoritmini ishlab chiqish, amalga oshirish va baholashni taqdim etadi. Puxta kodlash va sinchkovlik bilan baholash orqali algoritmnining samaradorligi baholanadi va uning tabiiy tilni qayta ishlash vazifalariga qo‘shadigan potensial hissasi haqida qimmatli tushunchalarni taqdim etadi.

Abstract. Today, the development of stemmers that do not depend on natural language is of great importance. The N-gram method of the tokenization process can sometimes incorrectly identify stems. Therefore, by making some modifications to this N-gram method, a new stemmer was developed. This paper presents the development, implementation, and evaluation of an innovative stemming algorithm using n-grams. Through careful coding and careful evaluation, the performance of the algorithm is evaluated and provides valuable insights into its potential contribution to natural language processing tasks.

Аннотация. Сегодня большое значение имеет разработка стеммеров, не зависящих от естественного языка. N-граммный метод процесса токенизации иногда может неправильно идентифицировать основы. Поэтому, внося некоторые изменения в этот метод N-грамм, был разработан новый стеммер. В этой статье представлены разработка, реализация и оценка инновационного алгоритма стемминга с использованием n-грамм. Благодаря тщательному кодированию и тщательной оценке оценивается производительность алгоритма и предоставляется ценная информация о его потенциальном вкладе в задачи обработки естественного языка.

Kalit so‘zlar: *Axborot olish (IR), Stemming, Stemmer, N-gramm, statistik metod, adekvat, tokenizatsiya, algoritm, gipoteza.*

Kirish

Bugungi kunda zamonaviy qidiruv va indeksatsiya tizimlari samaradorligini oshirish uchun **stemming jarayonidan** foydalanadi. Stemming jarayoni lemmatizatsiya jarayoniga o‘xshash bo‘lib, so‘zning morfologik variantlarini bir shaklga keltirish amallarini bajarishga asoslangan. Qidiruv tizimlari hujjatlarni indekslash jarayonida matn so‘zlaridagi qo‘shimchalar (affiks/prefikslar)ni

qisqartirishni bajradi. Natijada hosil qilingan o‘zaklar to‘plami IR tizimida olingan hujjatlar sonini oshiradi. Hozirda mavjud stemmerlar aniqlik ko‘rsatkichi to‘liq natijalarni qaytarmasligi ba’zi olimlar tomonidan qayd etilsa-da, zamonaviy IR tizimlari bugungi kunda ham stemmingdan muhim vosita sifatida foydalanadi.

Stemming jarayoni ustida bir qator ilmiy izlanishlar olib borilgan va muhim natijalar olingan. Jumladan, **qoidalariga asoslangan yondashuvli** stemmerlarni ishlab chiqish uchun til morfologiyasi bo‘yicha zarur bilimlar talab qilinadi. **Til korpusining statistik tahlili yondashuviga** asoslangan stemmerlar muayyan til uchun yaxshi natijalarni qaytaradi. Biroq ushbu lingvistik stemmerlar bitta katta kamchilikka ega, ular ko‘p tilli matnlarni tahlil qila olmaydi. *Belgilar chastotalariga asoslangan statistik algoritmlar* ushbu kamchilikni bartaraf qilishi mumkin.

Statistik metodlarda bizning e’tiborimizni tortgan eng qiziqarli va samarali usul Jeyms Mayfild va Pol MakNaming *Single N-gram Stemming* [Mayfield, 2003]. Mualliflar N belgining bir-biriga o‘xshash ketma-ketliklaridan foydalanib, stemming foydasiga tildan mustaqil ravishda erishish mumkin, deb taxmin qilishdi. Ular so‘zlardan olingan, N-grammlarning ba’zilari (aslida, kam tez-tez) so‘zning morfologik o‘zgarishlarni ko‘rsatmaydigan qismlari uchun umumiy bo‘lishini, ya’ni ular aynan bir xil o‘zak o‘rnini bosuvchi bo‘lib xizmat qilishi mumkinligini ta’kidladilar. Ular so‘zning “*pseudo stem*” (soxta o‘zak) sifatida *Inverse Document Frequency* (IDF) bilan ichki N-gramm so‘zini tanlashni taklif qilishdi. Ular affiksalar bilan namoyon bo‘ladigan ma’lum bir morfologik o‘zgarish ko‘plab turli so‘zlarda takrorlanishi shuning uchun ham past IDF ko‘rsatishi haqida ajoyib asoslar berdi.. Masalan, ular o‘zlarining kontsepsiyalariga ko‘ra, “juggling” so‘zining (CLEF 2002 to‘plamidan) va taklif qilingan psevd-4 va psevd-5 so‘zlarining turli N-grammlarining hujjat chastotalarini jadvalga kiritdilar.

Bizning hozirgi tadqiqotimizning kaliti yuqorida aytib o‘tilgan jadvallarda joylashgan. Currency va warrens so‘zlari uchun psevd-4 o‘zak mos ravishda *rren* va *rens* edi. Juggler, currency va warrens uchun psevd-5 poyasi mos ravishda *ggler*, *rency* va *rrens* edi. Umumiy tushunchaga ko‘ra, o‘zak boshqa joydan emas, balki so‘zning boshlang‘ich belgisidan boshlanishi kerakligini aytadi. Xo‘sh, bu allaqachon oqlangan usulni takomillashtirish imkoniyati bormi?

Yuqoridagi savolga javob berishga harakat qilish bizning maqolamizning mohiyatidir. Biz n-grammlarni yaxshiroq tatbiq qilish uchun faoliyat olib boramiz, ularning chastotalarini hisoblash g‘oyasini saqlab qolamiz. IDFning oddiy kontsepsiyasi o‘rniga, biz o‘zaklarni yaxshiroq taxmin qilish imkonini beradigan yangi algoritmni ishlab chiqamiz.

2. Tavsiya etilgan usul:

Mualliflar [Mayfield, 2003] o‘zlarining eksperimental natijalariga ko‘ra, “Snowball stems [Porter, 2001], psevd stems va 4-gramm so‘zlaridan eng yaxshi natija ko‘rsatish tendensiyasi 4 grammda kuzatilgan. Ular 4 grammlar, har qanday so‘zning boshlang‘ich 4 ta belgisi so‘z o‘zagining eng yaxshi ifodasidir degan xulosaga kelishdi. Bu bizning to‘g‘ri (qat’iy lingvistik nuqtayi nazardan emas) va

ajoyib taxmin qilingan o'zak uchun u boshidan boshlanishi kerakligi haqidagi tushunchamizni mustahkamlaydi. Xo'sh, bir-biriga o'xshash N ta belgining chastotalarini (ya'ni, ichki N-gramm so'zlari) topish o'rniga, 4 gramm, 5 gramm, 6 gramm va hokazo chastotalardan kelib chiqqan taxminiy asosni topsak nima bo'ladi? Istiqbolli yo'nalish ko'rinadi. Bizning fikrimiz, o'zakni topish uchun dastlabki taxminimiz sifatida 4 grammni olishdir. Shubhasiz, bu uzunligi 3 yoki undan kam bo'lgan so'zlarni istisno qiladi. Bu ajoyib afzallik bo'ladi, chunki to'xtash so'zlarining ko'pchiligi 3 yoki undan kam uzunlikda. Endi ma'lum bir so'z uchun bizda 4-gramm, 5-gramm va shunga o'xshash chastotalarni topish va ulardan eng yaxshisini tanlash qoladi. Yondashuvimizni boshlashdan oldin, keling, COCAning turli N-gramm chastotalarini ko'rib chiqaylik [COCA, 2014] Biz xuddi shu jungling so'zidan boshlaymiz.

Jadval 2.1: “Juggling” so'zi uchun COCA dan turli N-gramm chastotalari

<i>N-gram</i>	<i>Chastotasi</i>	<i>N-gram</i>	<i>Chastotasi</i>
jugg*	915	jugglin*	328
juggl*	729	juggling*	328
juggli*	328		

Biz o'zakning adekvat (aynan bir xil) yondashuvi bo'lib xizmat qiladigan so'zning N-grammasi korpus orqali nisbatan yaxshi chastotaga ega bo'lishi kerak degan xulosaga keldik. Bundan kam N-grammlarning barchasi tez-tez bo'lib va undan yuqori chastotalar esa asta-sekin pastroq bo'lib, berilgan so'zning umumiy chastotasi bilan tugaydi (ikki yoki undan ortiq ketma-ket yozuvlar bir xil bo'lishi mumkin).

Yuqoridagi jadval bizning taxminimizga muvaffaqiyatli javob beradi. Shunday qilib, bizda qisman so'z uzunligi (N-gram) bir birlikka ko'payishi bilan qiymatlari pasayishga (yoki doimiy bo'lib qolishga) moyil bo'lgan butun son o'zgaruvchiga mavjud. So'zning (o'zakning) o'zgarmas qismidan qo'shimcha qismiga o'tishda bu pasayish juda keskin bo'lishi ehtimoldan holi emas. Ammo nisbiy maksimal tushishni qanday ushlay olamiz? Buning bir usuli - har ikki ketma-ket yozuvning qaytishini olishdir. Ko'rsatkichni quyidagi tarzda belgilaymiz

(2.1)

$$\lambda_i = \begin{cases} abs(F_i - F_{i-1}) & \text{if } i > 4 \\ 0 & \text{else} \end{cases}$$

Bu yerda F_i - i^{th} chastotasi

N-gram. Keyin biz ikkinchi tartibli og'ishlarni hisoblaymiz.

(2.2)

$$\Delta_i = \lambda_i - \lambda_{i-1}$$

N uzunlikdagi N-grammga mos keladigan o'zgaruvchi bo'lsa. Biz protseduramizni ikki bosqichda aniqlaymiz:

1-bosqich:

1) Initializatsiya:

bu erda M - juda yuqori butun qiymat

$$\psi_N = 4$$

2) Rekursiya:

Calculate λ_i , $4 < i \leq |\text{word}|$

If $\lambda_i > \Gamma$

$$\psi_N = \underset{4 < i \leq |\text{word}|}{\operatorname{argmax}} [F_i, F_{i-1}]$$

Else

$$\psi_N = i$$

3) Tugatish:

a.) If $N=|\text{word}|$, stop. Else, calculate Δ_i .

b.) If $\Delta_i > 0$, stop.

Bu yerda Γ - chegara qiymati bo'lib, u ikki N-gramm chastotali burilishning qanchalik maqbulligini belgilaydi. U nolga yoki minimal qiymatga saqlanadi. Rekursiya bosqichi kiritmoq so'zining ortib borayotgan N-grammlari orqali aylanishni amalga oshiradi. Agar 1-bosqich oxirida $\psi_N = |W|$ bo'lsa, 2-bosqichga o'ting.

2-bosqich:

Agar oxirgi uchta N-grammning chastotalari ψ_N bir xil bo'lsa, kiritilgan so'zning oxirgi uchta belgisini o'chiring, agar ularni o'chirish natijasida hosil bo'lgan asosiy uchdan katta bo'lsa, ya'ni

$$\psi_N = \psi_N - 3 \quad \text{If } \psi_N - 3 > 3$$

Yuqoridagi protsedura oxirida ψ_N ortiqcha talab qilingan o'zakka yaqin bo'lgan N gramm qiymatini o'z ichiga oladi. 2-bosqich umumiy qo'shimchalarni (-ing kabi) ko'rsatadigan so'zlarni boshqaradi, agar ular 1-bosqichda ajratilmagan bo'lsa.

3. Illyustratsiyalar va tajribalar:

Keling, o'z uslubimizni ingliz, ispan va portugal so'zlari klasteriga qo'llaylik. Bizga kerak bo'lgan narsa - bu boy korpusdan N-grammlarning ketma-ket chastotalari. Ingliz tili uchun biz COCA ga tayandik [COCA. 2014]. Ushbu korpusning ikkita afzalligi bor, birinchidan u inglizcha so'zlarning boy to'plamiga ega, ikkinchidan, wild card [*] yordamida N -gram chastotalarini hisoblash oson. Ispan va portugal so'zlari uchun ketma-ket chastotalar mos ravishda CorpusDelEspanol [Corpus Del Español, 2014] va Corpus Do Português [Corpus Do Português, 2014] dan olingan. Ba'zi nomzodlar bo'yicha test natijalari quyida keltirilgan:

3.1-jadval. So‘z turkumlari va o‘zaklari

Klaster raqami.	So‘z	O‘zak	Klaster raqami.	So‘z	O‘zak
1 (Inglizcha)	create	creat	2 (Ispancha)	trabajan	traba
	creates	creat		trabajar	traba
	creating	creat		trabajado	traba
	created	creat		trabajador	traba
	creation	creat	3 (Portugalcha)	dificil	dific
	creative	creat		dificilmente	dific

Keling, bizning stemmerimizni empirik tarzda (tajribaga asoslangan holda) o‘rnatilgan lingvistik stemmer bilan taqqoslaylik, deydi Porterning stemmeri [Porter, 1980]. Sinov tili sifatida biz ingliz tilini tanlaymiz, chunki Porter algoritmining natijalarini Snowball loyihasida osongina olish mumkin [Porter, 2001]. Biz 100 ta tasodifiy inglizcha so‘zlardan namuna olamiz va Porter’s Snowball tizimi va ular ustidagi protseduramizni qo‘llaymiz. Natijalar 3.2-jadvalda keltirilgan. Ikkala algoritm tomonidan yaratilgan bir xil ildizlar qalin qilib ko‘rsatilgan. To‘rtta POS, ot, sifat, fe‘l va qo‘shimchalar hisobga olindi. Biz COCAda gazetalarni qidirish maydonini tanlaymiz.

Snowball (Porter) stemmer va bizning N-gram stemmerning natijalarini solishtirish uchun biz to‘g‘ridan-to‘g‘ri baholash usulidan foydalandik. Chris D. Paice [Paice, 1994] tomonidan amalga oshirilgan taqqoslash kabi, uzunlikni qisqartirishni asosiy chiziqli sifatida ishlatib, biz Levenshteyn masofasidan [Levenshteyn, 1966] asosiy o‘lchov sifatida foydalandik. Masofa - bu manba qatorini maqsadli satrga aylantirish uchun zarur bo‘lgan o‘chirishlar, qo‘shishlar yoki almashtirishlar soni. Shunday qilib, Levenshteyn masofasi so‘zning stemmer tomonidan qancha birlikdan ajratilganligini ko‘rsatadi.

Biz faraz qilamizki, agar bizning stemmerimiz Porterniki kabi samarali bo‘lsa, u holda Levenshteyn masofalarining so‘z va uning ajratilgan o‘zak o‘rtasidagi taqsimoti bir xil bo‘ladi. Berilgan so‘z uchun bizda ikkita to‘g‘rilash usuli bor, Porter algoritmi va N-gramm protseduramiz. Shunday qilib, berilgan so‘z uchun bizda bir juft LD mavjud. Taqqoslash uchun biz H0 nol gipotezasini o‘rnatdik, bu ikkita stemmer o‘rtasida hech qanday farq yo‘q.

Ikki chora-tadbirlar to‘plamini bir xil namunaga bog‘langan ikkita o‘lchov sifatida ko‘rib chiqish mumkinligi sababli, biz juftlashtirilgan namunalar uchun statistik testdan foydalanishga qaror qildik.

Xususan, parametrik bo‘lmagan statistik test, Wilcoxon imzolangan darajali test bizda qo‘llaniladi, chunki bu masofalarning taqsimlanishi haqida hech qanday dalil yo‘q edi.

Wilcoxon testi N kuzatilgan qiymatlarning ikkita juftlashgan seriyasiga (x_i , y_i) asoslanadi, taqqoslanadigan kuzatilgan X, Y o'zgaruvchilarning har biri uchun bir qator. Wilcoxon testi N kuzatilgan qiymatlarning ikkita juftlashgan seriyasiga (x_i , y_i) asoslanadi, taqqoslanadigan X va kuzatilgan Y o'zgaruvchilarning har biri uchun bir qator. Mutloq farqlarning $d_i = \text{abs}(x_i - y_i)$ va ishorasi (d_i) mutlaq farqlari hisoblanishi kerak. Keyin mutlaq qiymatlar nollardan voz kechib tartiblanadi.

Agar namuna hajmi kichraytirilgan bo'lsa, testning yakuniy statistikasi katta N uchun Oddiy o'zgaruvchiga yaqinlashtiriladi.

4. Eksperimental natijalar:

Bizning namunaviy ma'lumotlarimiz bo'yicha Wilcoxon testini o'tkazganimizda, biz p qiymatini olamiz 0,54. $p > 0,05$ sifatida, H_0 nol gipotezasini rad eta olmaymiz. Shunday qilib, bizning N-gramm Stemmerimiz Porter's Stemmerdan kam emasligi haqida kuchli dalillarga egamiz.

Xulosa

Biz statistik stemming texnikasini o'zgartirishga harakat qildik va natijalarni ancha tajribali lingvistik stemmer bilan solishtirish mumkin bo'lgan usulni ishlab chiqdik.

Bizning N-gram stemmerimiz N-gramm korpus chastotalari mavjud bo'lgan har qanday tilda ishlashga qodir.

Tilning neytral yondashuvi har doim lingvistik bilim yoki tahlilning zaruriy sharti bo'lgan texnikadan afzalroq bo'lganligi sababli, bizning natijalarimiz istiqbolli ko'rinadi.

Foydalanilgan adabiyotlar:

1. Elov B., Hamroyeva Sh., Abdullayeva O., Xusainova Z., Xudayberganov N. / O'zbek, turk va uyg'ur tillarida POS teglash va stemming. O'zbekiston: til va madaniyat 2023/1. – P. 40-64
2. Elov B., Alayev R., Xusainova Z. O'zbek tilida stemmingni amalga oshirishning gibrid statistik yondashuvi. 2023
3. B.B.Elov, Sh.M.Khamroeva, Z.Y.Khusainova. Pipeline conveyer of NLP (natural language processing). Descendants of Muhammad al-Khwarazmi. Scientific-practical and information – analytical Journal, 1 (23) / 2023, 181-192 pp.
4. Xusainova Z.Y. Tabiiy tilni qayta ishlash (NLP)da stemming jarayoni tavsifi / BuxDU ilmiy axboroti 3/2023. – B. 113-119
5. Corpus Del Español < <http://www.corpusdelespanol.org> > manzilida, 2014-yil 13-yanvarda tashrif buyurilgan.
6. Corpus Do Português. <<http://www.corpusdoportugues.org/>> manzilida, 2014-yil 13-yanvarda tashrif buyurilgan.
7. Zamonaviy Amerika ingliz tili korpusi (COCA). <<http://corpus.byu.edu/coca/>> manzilida, 2014-yil 13-yanvarda tashrif buyurilgan.

8. Douson Jon (1974). Qo'shimchalarni olib tashlash va so'zlarni birlashtirish. ALLC byulleteni, 2-jild, № 3, 33-46.
9. Funchun Peng, Nawaaz Axmed, Xin Li and Yumao Lu (2007) Veb-qidiruv uchun kontekstga sezgir. ACM SIGIR 30-chi yillik xalqaro konferentsiyasi ma'lumotlarini qidirishda tadqiqot va rivojlantirish bo'yicha ma'ruzalar, 639-646
10. Hafer M. va S. Weiss (1974). Harf o'rinbosar navlari bo'yicha so'zlarni segmentatsiyalash. Ma'lumotni saqlash va qidirish, 10, 371-85.
11. Harman Donna (1987). Onlayn muhitda qo'shimcha qo'shish cheklovlari bo'yicha muvaffaqiyatsizlik tahlili. Axborot qidirishda tadqiqot va ishlanmalar bo'yicha 10-yillik xalqaro ACM SIGIR konferentsiyasi materiallari, 102-107.
12. Harman Donna (1991). Qo'shimchalar qanchalik samarali? Journal of the American Society for Information Science, 42, 7-15.
13. Krovetz Robert (1993). Morfologiyani xulosa chiqarish jarayoni sifatida ko'rib chiqish. Axborot qidirishda tadqiqot va ishlanmalar bo'yicha 16-yillik ACM SIGIR konferentsiyasi materiallari, 191-202.
14. Levenshtein V. I. (1966). O'chirish, kiritish va teskari o'zgarishlarni tuzatishga qodir ikkilik kodlar. Sovet fizikasi- Doklady, 10(8),707710.
15. Lovins, J.B. (1968). Striping algoritmini ishlab chiqish. Mexanik tarjima va hisoblash tilshunosligi, 11, 22-31.
16. Majumder Prasenjit, va boshq. (2007). YASS: Yana bir qo'shimcha striptizator. Axborot tizimlari bo'yicha ACM operatsiyalari. 25(4), №18-modda.
17. Mayfield Jeyms va McNamee Paul (2003). Yagona N-gramm stemming. Axborot qidirishda tadqiqot va ishlanmalar bo'yicha 26-yillik xalqaro ACM SIGIR konferentsiyasi materiallari, 415-416.
18. Melucci Massimo va Orio Nicola (2007). Avtomatik Stemmer yaratish metodologiyasini loyihalash, amalga oshirish va baholash. Amerika axborot fanlari va texnologiyalari jamiyati jurnali, 58(5), 673-686.

TILSHUNOSLIKDA “SLENG” TUSHUNCHASI VA UNGA YONDOSH HODISALAR

Bekmuradova Iroda Zokir qizi

irodabekmuradova9@gmail.com

ToshDO‘TAU magistranti

Annotatsiya. Ushbu maqolada o‘zbek tilshunosligida sleng tushunchasini ajratish tamoyillari, tildagi slenglarga o‘xshash va yondosh leksik hodisalar haqida ma’lumot berilgan. Slenglarning ko‘chma ma’noda qo‘llangan so‘zlardan farqi misollar vositasida tushuntirib berilgan. Slenglarni ajratish omillari sanab o‘tilgan va izohlangan.

Abstract. This article provides information on the principles of distinguishing the concept of slang in Uzbek linguistics, similar and related lexical phenomena to slang in the language. The difference between slang and figurative words is explained by means of examples. Factors that distinguish slang are listed and explained.

Аннотация. В данной статье представлена информация о принципах разграничения понятия сленг в узбекском языкознании, сходных и родственных сленгу лексических явлений в языке. Разница между жаргонными и переносными словами объясняется на примерах. Перечислены и объяснены факторы, отличающие сленг.

Kalit so‘zlar: *sleng, o‘xshatish, ko‘chma ma’no, metafora, metonimiya, sinekdoxa.*

Sotsiolingvistik birliklar ichida ko‘p qo‘llanadigan noadabiy leksik qatlamga oid birliklar doimiy o‘rganishga muhtoj. Yozma nutqda kamdan-kam ishlatilsa ham, og‘zaki so‘zlashuvda faol bo‘lgan jargon, argo va slenglar zamonaviy tilga xos so‘zlar hisoblanadi. Ularning barcha keng aholi uchun birday tushunarli bo‘lolmaydi, chunki, bunday birliklar ko‘p hollarda qisqalik, aytilmoqchi bo‘lgan fikrni boshqalardan yashirish maqsadida ishlatiladi. Zamonaviy so‘zlashuv uslubida tilimizga mansub slenglar ham, o‘zga tillardan o‘zlashgan slenglar ham uchraydi. O‘zbek tilshunosligida slenglar tadqiqotlari hali yetarli darajada amalga oshirilmaganligi sababli, bunday so‘zlar mavzuga doir keng sotsial tadqiqotlar, slenglar lug‘atini yaratish kabilar yangi ilmiy izlanishlarga muhtoj. Ushbu vazifani bajarish uchun slenglar va ularga yondosh boshqa ijtimoiy lug‘aviy birliklar va leksik hodisalarning chegaralarini belgilash hamda ularning xususiyatlarini aniqlash zarur.

Sleng – G‘arb tilshunosligi an‘analariga xos termin sanaladi. Mazkur terminga haligacha aniq ta’rif berilmagan. Turli lug‘atlarda va sotsiolingvistikaga oid adabiyotlarda uning “jargon” terminiga yaqin turishi qayd qilingan. Sleng so‘zi ilk bor yozma ravishda XVIII asrda Angliyada “haqorat” ma’nosini anglatuvchi so‘z

sifatida qo'llangan. Keyinchalik sleng “noqonuniy” oddiy leksika ma'nosini bildiruvchi so'z sifatida keng qo'llanila boshladi. Ayni paytda uning jamiyatning quyi tabaqalari qo'llaydigan slang-lingo va “rang-barang” aholi foydalanadigan argot singari sinonimlari ham paydo bo'ldi

Slengga xos eng umumiy ta'rif – bu muayyan kasb egalari yoki ijtimoiy guruhlarining og'zaki nutqida qo'llaniladigan, emotsional-ekspressiv bo'yoqqa ega bo'lgan so'z va iboralar. Rus tilshunosligida bunday xususiyatga ega bo'lgan birliklarning faqat yoshlar nutqiga xos bo'lganlarigina sleng sifatida olinadi. Ingliz tilshunosligida esa slengning qamrovi biroz kengroq. U jargon so'zi bilan sinonim sifatida yoshi va kasbidan qat'iy nazar har qanday ijtimoiy guruhlar uchun xos bo'lgan “kodlangan leksika”ni bildiradi. O'zbek tilshunosligidagi sotsiolingvistik tadqiqotlarda rus tilshunosligidagi qarashga ergashish holati kuzatiladi. Ya'ni slenglar ko'proq yoshlarning nutqida uchraydigan o'ziga xos yashirin ma'noga ega bo'lgan leksik birliklar sifatida e'tirof etiladi. Masalan, *shkaf* – barvasta, gavdali, norg'ul. Ba'zan beso'naqay qomatga ega odamlarga nisbatan. *Sigir* – qizlarga ko'p pul sarflaydigan yigitlar. *G'ijim* – o'ziga qaramaydigan, o'ziga bee'tibor odamlarga nisbatan va shu kabilar. Turshak – o'ziga qaramaydigan, o'ziga bee'tibor odam.

Sleng – jonli, harakatdagi til shakli bo'lib, u mamlakat va jamiyat hayotidagi har qanday o'zgarishdan ta'sirlanadi. Sleng qo'llanish doirasiga ko'ra: hammaga ma'lum va barcha tomonidan qo'llanuvchi umumiy sleng (general slang) va muayyan guruhlariga ma'lum bo'lgan, tor doirada qo'llanuvchi xususiy sleng (special slang)ga ajraladi. Shuningdek, slengizmlar ko'pincha chet tillardan, ularning dialektlaridan va jargonlaridan o'zlashtirilgan bo'ladi. Masalan, “tusovka”, “telejka”, “tachka”, “stukach” rus tilidan, “yuzer” (user), “chat” (chat), ingliz tilidan o'zlashtirilgan.

Slenglarning (ayniqsa, ularning dastlabki shakllanishi davrida) keng ommaga tushunarsiz bo'ladi. Chunki aksariyat slenglar ko'chma ma'noda ishlatiladi. Shunga ko'ra ular tildagi boshqa leksik hodisalar: metafora, o'xshatish, sinekdoxa kabi hodisalarga o'xshab ketadi. Ammo shunday bo'lsa ham mazkur hodisalar bir-biridan farqli holatda qarashni talab etadigan ayrim xususiyatlarga ega.

So'zlashuv jarayonida so'zlovchi ifodalayotgan fikrini til birliklari vositasida ta'sirchan ifodalash, tasvirning o'ziga xos individual xususiyatlarini ochish uchun so'zlarni o'z ma'nosidan tashqari ko'chma ma'nolarda ham qo'llaydi. Bunda so'zlar o'z lingvistik vazifasi bilan bir qatorda qo'shimcha ma'nolarni ham ifodalaydi. So'zlarning bunday ma'nolarda qo'llanishi nutqqa obrazlilik, ekspressiv bo'yoq beradi, ifodani erkin va to'laqonli ochib berib, tinglovchining estetik zavq olishini ta'minlaydi. Metafora ana shunday hodisadir.

Metafora narsa va hodisalar o'rtasidagi o'xshashlikka asoslangan holda so'z ma'nosining ko'chishi hodisasini ifodalaydi. Metafora fikrni obrazli ifodalash vositasi sifatida og'zaki va yozma nutqda ham, badiiy, publitsistik uslublar bilan bir qatorda oddiy so'zlashuv uslubida ham keng qo'llaniladi. Biroq oddiy so'zlashuvda uslubida qo'llangan metaforalar badiiy metaforalardan farqlanadi. Chunki,



tilshunoslikda metafora hodisasining yuzaga kelish omillari, nutqning u yoki bu ko‘rinishida namoyon bo‘lishi jihatidan lingvistik va badiiy metaforalarga ajratiladi. B.Umurqulovning fikricha, “Til metaforasi badiiy metaforadan o‘xshashlikning formasiga ko‘ra farqlanadi: til metaforasi ikki predmet orasidagi bevosita yoki ochiq o‘xshashlikka asoslangan ko‘chim bo‘lsa, badiiy metaforada predmetlar orasidagi o‘xshashlik asosida badiiy bo‘yoq, obrazlilik yotadi. Demak, badiiy metafora til metaforasidan, dastlab, fikrni ifodalashi bilan ajralib turadi. Badiiy metaforada o‘xshashlik yashiringan bo‘ladi” [Begmatov, Boboyeva, Asomiddinova, Umurqulov, 1988: 80]

Fikrni ifodalashda ma‘no ko‘chish hodisalaridan foydalanishning asosiy vazifasi gapga obrazlilik bag‘ishlash bilan bir qatorda so‘zlovchi tomonidan tasvirlanayotgan obrazning barcha qirralarini tinglovchi tasavvurida aniq gavdalanirish, fikrning qisqa va aniqligini ta‘minlashdan iboratdir. Masalan: *Beodob qizning gapini onasi tekislab turdi*. Ushbu gapda onaning qizining gaplarini to‘g‘rilashga bo‘lgan urinishi biror narsani tekislashga o‘xshatilgan hamda tekislamoq so‘zi bilan ifodalangan metaforani yuzaga keltirgan. Yoki: *Bu voqea yanvarning oyoqlarida ro‘y bergan edi*. Gapdagi oyoq so‘zi organizmning eng pastida, eng oxirida joylashganlik xususiyati bilan oyning so‘ngi ma‘nosini bildirib kelgan.

Bir narsa, belgi, harakatning nomi boshqasiga o‘xshashlik asosida emas, balki o‘zaro bog‘liqlik, aloqadorlik asosida ko‘chirilishiga metonimiya usulida ko‘chirilish deyiladi. Masalan: *Butun qishloq uni olqishladi*. Bu gapdagi qishloq so‘zi joy, manzil ma‘nosida emas, balki shu manzilda yashovchi odamlar ma‘nosida (ko‘chma ma’noda) qo‘llangan. Ma‘no qishloq bilan unda yashovchilar o‘rtasidagi bogliqlik, munosabat asosida kelib chiqqan. [O‘zbek tili praktikumi, 2005: 47] Yoki: *Charm kamzuldan hamma qo‘rqadi* va h.k.

Bir predmetning nomi boshqa bir predmetga qism bilan butun munosabati asosida ko‘chirilishiga sinekdoxa usulida ko‘chirilish deyiladi. Masalan: *Tirnoqqa zor bo‘ldi*. Bu gapda tirnoq so‘zi (qismning nomi) “farzand” ma‘nosida (butunning nomi) kelgan. *Besh qo‘l barobar emas*. Bu gapda qo‘l (butunning nomi) beshta barmoq (qismning nomi) ma‘nosida kelgan. [O‘zbek tili praktikumi, 2005: 50]

O‘xshatish orqali so‘zlarda ifodalangan ikki yoki undan ortiq narsa va hodisa, xususiyatlar o‘rtasida mavjud bo‘lgan o‘xshashlik, sifat va belgidagi umumiylik qiyoslanadi, o‘zaro solishtiriladi, tasvirlanayotgan narsa-hodisaning ayrim xususiyati yorqin, chuqurroq va ta’sirliroq ochib beriladi. Sodda tarzda aytadigan bo‘lsak narsa va hodisalar o‘rtasidagi o‘xshashlikka asoslanib, ularning biri orqali ikkinchisining belgisini, mohiyatini to‘laroq, aniqroq, bo‘rttiribroq ko‘rsatib berish o‘xshatish deyiladi. O‘xshatish eng qadimiy va ko‘p qo‘llanadigan tasviriy vositalardandir. [Qo‘ng‘urov, Begmatov, Tojiyev, 1992: 89] O‘xshatishlarning o‘ziga xos fazilati shundaki, ular ko‘zda tutilgan har qanday fikrning aniqlashishida, shu obrazni to‘liq tasavvur etishda hal qiluvchi rol o‘ynaydi. O‘zbek tilida boshqa tillarda bo‘lgani kabi o‘xshatishlar to‘rt unsuridan tarkib topadi va ularni o‘xshatish

subyekt, o'xshatish etaloni, o'xshatish asosi va o'xshatishning shakliy ko'rsatkichi deb nomlanadi [Mahmudov, O'TA, 2011-yil 3-son: 98] M.Mukarramov o'xshatishlarni yuqorida keltirilgan to'rt unsurning qay darajada ishtirok etishi belgisiga qarab ikki guruhga ajratadi:

1. Grammatik belgisi o'xshatishlar. Bunda obyekt, obraz va belgilar grammatik ko'rsatkichi bilan mavjud bo'ladi. To'rt belgidan birontasi ham tushib qolmaydi: *O'choqdan chiqqan zangori tutun osmonga o'rlab, tevarak-atrofga ko'rpadek yoyilar edi. (G'.G'ulom)*

2. Logik o'xshatishlar. Bunda grammatik belgi bo'lmaydi: *Sizning poyingizda gardman, to'zonman. (A.Oripov)* O'xshatishlarning slenglarga o'xshash nuqtasi aynan logik o'xshatishlarda kuzatiladi. Jumladan, *qovun, g'isht, tapochka* kabi slenglar. Grammatik belgisiz qo'llangan o'xshatishlarga o'xshab ketadi. Bunday holatlarda ularni qanday farqlaymiz degan savol tug'ilishi tabiiy, chunki boshqa slenglarda shunga o'xshash holatni ko'rish mumkin. Masalan, yuqorida misol qilib keltirilgan, *shkaf, turshak, g'ijim* kabi slenglar bir qarashda so'zning ko'chma ma'noda qo'llanishidan boshqa narsa emasday ko'rinadi, ammo ma'no ko'chish hodisalariga xos individuallik xususiyati bularning shunchaki ko'chma ma'no bo'la olmasligini ko'rsatadi. Har bir so'zlovchi fikrini xoh og'zaki, xoh yozma tarzda bo'lsin, bayon qilar ekan, undagi fikr obyektini, yoki umuman u yoki bu fikrni nimagadir o'xshatib yoxud, qaysidir tushunchani boshqasining vastiasida ifodalab tasvirlash uning o'z xohishi. So'zlovchi o'z qarashlari va dunyoqarashidan kelib chiqib, nimanidir boshqa narsaga o'xshatishi mumkin. Ya'ni bunda har bir ma'no ko'chish yoki o'xshatishni so'zlovchining o'zi yaratadi va u tinglovchi tomonidan hech qanday qiyinchiliksiz tushuniladi. Slenglarda bu holat biroz boshqacha. Sleng har bir so'zlovchi tomonidan individual yaratilmaydi, guruh doirasida bo'lmagan odamlarning tushunishi qiyin, ular nutqqa tayyor holda olib kiriladi.

Yuqoridagi xususiyatlarni umumlashtirgan holatda tilimizdagi slenglar va ularga yondosh boshqa hodisalarning o'ziga xos farqli tomonlarini quyidagicha ajratish mumkin:

Birinchidan, sleng va metaforalar tilga rang-baranglik, o'ziga xos jilo va kreativlik qo'shish uchun ishlatiladigan lingvistik vositalar bo'lsa-da, ammo ular qo'llanish maqsadi jihatidan qat'iy farqlanadi. So'zlarni ko'chma ma'noda qo'llash nutqqa obrazlilik va ixchamlilik uchun ishlatilsa, slenglarni qo'llashdan aosiy maqsad guruhdan tashqaridagilarning tushunishlariga yo'l qo'ymaslik, fikr ifodasini "yashirish" dir.

Ikkinchidan, sleng norasmiy leksika birligi. Ko'chma ma'noli so'zlar yoxud o'xshatishlarda uslubiy chegara biroz boshqacha. Ular so'zlashuv nutqida keng ishlatilaversa ham, ko'proq badiiy uslubga xos hisoblanadi.

Uchinchidan, sleng, odatda bir so'z yoki so'z birikmasi shaklida, ba'zan asl shaklidan o'zgartirilgan yoki qisqartirilgan yoki muayyan kontekstda yangi ma'no kasb etgan so'z yoki iboralardan iborat bo'ladi. Ko'p hollarda ular boshqa tillardan

o'zlashtiriladi. Metafora, metonimiya yoxud o'xshatishlar ham bir so'z shaklida ham, so'z birikmasi shaklida ham bo'lishi mumkin, ammo ular boshqa tillardan o'zlashtirilmaydi.

Xulosa qilib aytganda, sleng ham, ko'chma ma'noda qo'llangan so'z ham tilga o'ziga xos ekspressivlik, o'ziga xoslik baxsh etadi. Sleng tushunchasining ko'lamini belgilashda ularning o'ziga xos umumiy va xususiy tomonlarini chuqur tahlil qilgan holda chegara qo'yish lozim.

Foydalanilgan adabiyotlar:

1. Begmatov E., Boboyeva A., Asomiddinova M., Umurqulov B., O'zbek nutq madaniyati ocherklari. -Toshkent, Fan, 1988. – B.80.
2. Mahmudov N. O'xshatishlar – obrazli tafakkur mahsuli. O'TA, 2011-yil 3-son.
3. Partridge E. Slang Today and Yesterday. London: Routledge and Kegan Paul, 1979. – P.3.
4. Qo'ng'urov R., Begmatov E., Tojiyev Y. Nutq madaniyati va uslubiyat asoslari. -T., 1992. – B.89.
5. O'zbek tili praktikumi. I qism. Toshkent, 2005. – B.45-50.

O'ZBEK TILIDA HISLARNI IFODALOVCHI BIRLIKLAR VA ULARNING AHAMIYATI

Mamasoyilova Sitora Mirzoxid qizi

sitoramamsoyilova@gmail.com,

ToshDO'TAU magistranti

Annotatsiya: Ushbu maqolada hislar tadqiqi, o'zbek tilida hislarni ifodalovchi birliklar tushunchasi hamda hislarni ifodalovchi birliklar: so'z, so'z birikmasi, ibora haqida so'z yuritilgan. Ularning nutqdagi ahamiyati, shaxslararo munosabatlarda samarali muloqot qilish va aloqalarni o'rnatishdagi roli ta'kidlangan.

Abstract: This article gives an information about the study of feelings, the concept of units expressing feelings in the Uzbek language, and units expressing sentiments: word, word phrase, idiom. Their importance in speech, their role in effective communication and establishing contacts in interpersonal relations is emphasized.

Аннотация: В данной статье дается информация об изучении чувств, понятии единиц, выражающих чувства в узбекском языке, и единиц, выражающих чувства: слово, словосочетание, идиома. Подчеркивается их значение в речи, роль в эффективной коммуникации и установлении контактов в межличностных отношениях.

Kalit so'zlar: *hislar, sentiment analiz, so'z, so'z birikmasi, ibora, shaxslararo munosabat.*

Kirish

Hislar kishilar muloqotida asosiy rol o'ynaydi, til orqali faqat fikrlar emas, balki shaxs, predmet hamda yuz berayotgan voqea-hodisalarga bo'lgan subyektiv munosabat ham ifodalanadi. Negaki til faqat aloqa vositasi emas, u fikrlarni yaratish va tushuntirish usuli hamdir.

Hislarni o'rganish turli fanlar, jumladan, tilshunoslik, psixologiya va antropologiya uchun muhim tadqiqot obyektlaridan biri bo'lib kelgan. Tilshunoslikda hislar tadqiqi Ch. Darvin va V. Jeyms singari tabiatshunos hamda psixolog olimlarning XIX asr oxirlarida hislar va til orasidagi bog'liqlik borasida bildirgan qarashlari negizida shakllangan. Biroq XX asrning o'rtalariga kelibgina tilshunoslar tomonidan hislarning tildagi rolini tizimli ravishda o'rganish boshlandi. Jumladan, A. Vierzbikka, J. Lakof, S. Ervin-Trip, D. Tannen, A. Duranti kabi lingvistlar hislar va madaniy semantika, til, bilish va hissiyotlar o'rtasidagi munosabat, nutq va nutqning hissiy mazmuni, ijtimoiy va madaniy kontekstlarda his-tuyg'ularning qanday uzatilishi va talqin qilinishi, muloqotda hislarning ahamiyatini o'rganishgan. Bu tadqiqotlarning barchasi hislar va til orasidagi o'zaro murakkab ta'sir haqidagi nazariyalarni boyitishga hissa qo'shgan, hislarning qanday

ifodalanishi, talqin qilinishi va til orqali shakllantirilishidagi qarashlarning chuqur tadqiqiga bag'ishlangan.

Bundan tashqari jahon veb tarmog'i bugungi kunda xarid qilingan oziq-ovqat mahsulotlari, kitoblar, texnik jihozlar; mehmonxonalar, ko'ngilochar maskanlar, umumiy ovqatlanish joylari va ularda ko'rsatilgan xizmat sifati haqida xaridorlar hamda iste'molchilar tomonidan yozib qoldirilgan shaxsiy fikrlar omma bilan bo'lishiladigan keng platformaga aylangan. Hozirda onlayn tarmoq 5.3 milliarddan ortiq faol tarmoq foydalanuvchilarini qamrab olgan [www.statista.com]. Sentiment analiz esa shu tarmoq foydalanuvchilari qoldirgan subyektiv fikr, izohlarda ifodalanuvchi hislarni tahlil qilishi bilan ahamiyatlidir. Negaki hislar tahlili orqali tashkilotlar, xaridorlar yoki sayyohlar mahsulotlar xaridi, maishiy xizmat ko'rsatish sohasidan kelgusida qaysi tanlovni amalga oshirishi kerakligi borasida qaror qabul qila oladi, bu esa o'z navbatida vaqtni tejaydi, iste'mol mahsulotlarining sifatini oshiradi hamda raqobatni ta'minlaydi. Yana bir e'tiborli jihati shundaki, mamlakat ichida sodir bo'lgan va bo'ladigan voqealarga nisbatan odamlarning ularga nisbatan shaxsiy munosabatini tahlil qilish sotsiologiya, statistika kabi sohalarni aniq fakt, shu bilan birgalikda, voqeliklarga nisbatan xolis bildirilgan munosabatlar xulosasi bilan ta'minlaydi.

O'zbek tili millionlab odamlar so'zlashadigan tildir. O'zbek tilining o'ziga xos xususiyatlaridan biri hislarni va fikrga munosabatni bildiruvchi o'ziga xos leksik, sintaktik hissiy birliklarning mavjudligi. Ushbu hissiy birliklar muloqotda hal qiluvchi rol o'ynaydi, suhbat yoki xabar ma'nosiga ta'sir ko'rsatadi.

O'zbek tilshunosligida o'zbek tilida ifodalanuvchi hislarni o'rganish boshqa tillarga nisbatan kam tadqiq etilgandir. Bugungi kungacha ushbu mavzu yuzasidan qilingan ishlar jamiyatdagi shaxslararo munosabatlarda hislar va ularni ifodalashning rolini yoritib berish, amaliy jihatdan qo'llash uchun yetarli emas. O'zbek tilining o'ziga xos lingvistik xususiyatlari, ya'ni leksik, morfologik birliklar, iboralarning hislarni ifodalashda o'rni qay darajada ekanligini hamda ularning tildan foydalanish va fikrni yetkazishdagi ahamiyatini o'rganish uchun qo'shimcha izlanishlar zarur. O'zbek tilida hislar va ularni ifodalovchi leksik birliklarni o'rganish esa masalaga ayni yechim bo'la oladi.

I. Turansky fikricha, “Emotsionallik – bu ayrim obyektiv lisoniy birliklar tomonidan aniqlanishi mumkin bo'lgan ruhiy holat” [Turanskiy, 1990]. Shu qatori “hislar” ham obyektiv lisoniy birliklar qatorida turadi, unga tegishli subyektiv lisoniy birliklarni bilish uchun mavzuga chuqurroq yondashish talab etiladi.

“Ekspressivlik ifodalovchi leksik-frazeologik vositalar sifatida polisemantizm, trop, stilistik chegaralangan leksika, sinonimiya va frazeologiya kabi hodisalarni belgilash mumkin. Sanab o'tilgan til hodisalari ekspressivlik hosil qilishda aktiv va sermahsul vositalardir. Polisemantizmlar bir ma'nosi (to'g'ri, nominativ ma'nolar) bilan neytral, yana bir ma'nosi (figural – ko'chma ma'nosi) bilan esa ekspressiv-emotsional fikr ifodalash uchun xizmat qiladi: *Do'stlik oftobining so'nmas quyoshi* (“O'zbekiston adabiyoti va san'ati”).

Sinonimiya stilistikaning asosi bo'lib, sinonimik birliklar ekspressivliklar yaratishda muhim o'rin egallaydi. Masalan, yuz, bet, aft, bashara, chehra, oraz sinonimik qatoridan aft, bashara, chehra, oraz so'zlari ekspressiv-stilistik bo'yoqlidir” [Abdullayev, 1983:4].

Demak, tildagi so'zlar qaysi turkumga oid bo'lmasin, ularning ko'pchiligi hissiy ma'no tashiy oladi. So'z hislar va fikrlarni shakllantirish imkoniyatiga ega va ular ayni vaqtda hislarni tahlil qilishning muhim tarkibiy qismi hisoblanadi. Yuqoridagi fikrga qo'shimcha tarzda aytish mumkinki, faqatgina ma'no hamda shakl jihatdan sinonimiya qatori emas, balki antonimlar ham so'zlarning hissiy ma'nosini yaqqol namoyon eta oladi. Antonimlarda bu aniq ifoda etiladi, chunki bir vaqtning o'zida bir antonim qo'llansa, uning zidi ham esga olinadi. Masalan, arzon-qimmat, yaxshi-yomon, kam-ko'p. Biroq bu birliklar turli vaziyatlarda, sohalar tarmog'ida turli ma'no kasb etishi mumkin. Misol uchun, “qimmat” so'zi bozordagi xaridorning salbiy munosabatini ifodalasa, mehmonxona va xizmat ko'rsatish sohasidagi xaridor uchun sifat jihatidan yuqori baholangani sababli ijobiy baho bo'la oladi.

“Ekspressivlik ifodalashning sintaktik usuli kommunikatsiyada ham, subyektiv munosabatda ham o'zining asl mohiyatini ko'rsata oladi. V.A.Bogorodiskiy aytganidek, “Morfologiya ayrim kategoriyalarning (so'z kategoriyalarining), ular formalarining inventarlaridan iborat. Sintaksis esa bu so'z formalarining hammasini harakatda va hayotda, nutq sostavida ko'rsatadi” [Abdullayev, 1983:6]. So'z nominativ shaklda turganda unga lisonidagi bor ma'nolar yuklangan holda bo'ladi. U boshqa so'zlar bilan sintaktik munosabatga kirishganda kontekst holatidan kelib chiqib aniq vaziyatdagi ma'nosini namoyon eta oladi. M.N.Goder bu haqda shunday deydi, “So'z birikmasi biror voqea-hodisani, u haqidagi tushunchani ifodalaydi. Lekin bu tushuncha sodda tushuncha emas. U yagona tushuncha bo'lishi bilan birga, tushunchalarning birikuvidan tashkil topgan bo'lib, ularning har biri butun so'z birikmasi orqali ifodalangan tushunchadan farq qiladi...” [Sharipov, 1978:15]

Leksik birlik	Farqi	O'xshashligi	Misol
So'z	Nominativ ma'noda turadi, biroz noaniq.	Hissiy belgi tashiy oladi	Yalmog'iz : 1. Mifologik qahramon. 2. Hamma narsani yeb-yutadigan; yuho.
So'z birikmasi	Aniq ma'no tashiy oladi.	Hissiy belgi tashiy oladi.	1. Yalmog'iz kampir – hissiy belgi tashimasligi aniq. 2. Bu yalmog'iz qashshoqlikning zaharli tomirlari o'z rizqini qayerdan va qanday olishini yaxshi bilib oldi – hissiy belgi tashiydi, ma'no aniq bilinmoqda.

1.1-jadval. So'z va so'z birikmasining hissiy ma'no tashuvi orasidagi farq va o'xshashliklar

Bundan xulosa qilib aytish mumkinki, so‘z birikmalari hissiy ma’no tashishda va hislar tahlilida aniqroq natija bera oladi. Masalan, ijtimoiy tarmoqdan olingan xaridor izohidagi hissiy ma’no tashuvchi so‘z birikmasi tahlilini olaylik:



1.1-rasm. So‘z birikmasining hissiy tahlili uchun olingan xaridor izohi

Bu yerda xaridor tomonidan sotib olingan predmetga nisbatan subyektiv munosabat bildirilmoqda. Subyektiv munosabat predmet haqidagi “juda yoqdi” birikmasi orqali aniq ifoda etilgan, bu esa xaridor munosabatida ijobiy fikr borligini anglatadi. Ta’kidlab o’tish lozimki, “yoqdi” so‘zidan ko‘ra “juda yoqdi” birikmasi munosabatdagi aniqlikni to‘la ochib bera oladi. Bu esa o‘z navbatida boshqa xaridorlarning ham ushbu predmetga nisbatan tasavvur hosil qilishi va qaror qabul qilishida muhim rol o‘ynaydi.

“Frazelogizmlar voqea-hodisa, belgilarning inson ongiga kuchli ta’sir qilishi, shu kuchli ta’sir natijasini nutqda to‘laligicha ifodalash, umuman, fikrning ta’sirchanligiga erishish maqsadi zaminida yaratilgan til hodisasidir” [Abdullayev, 1983:4]. Ularning o‘ziga xos jihati bir nechta so‘zdan tashkil topsa ham, ammo ma’nosi bir so‘z yoki undan ortiq so‘zlarga teng kelishi haqidagi informatsiyani tashiy olishidir. Bu oddiy matn holida o‘quvchiga kontekstga muvofiq aniq bo‘ladi, biroq hissiy tahlil uchun qo‘shimcha izlanishlar talab etiladi. O‘zbek tilida iboralar bo‘yicha amalga oshirilgan leksikografik izlanishlar hissiy tahlil uchun yetarli emas. Chunki ular ma’noviy jihatdan kerakli uch: ijobiy, salbiy, neytral kategoriyaga ajratilmagan. Ularni ushbu ko‘rinishda lug‘at va internet tarmoqlaridan to‘plangan matn bazalari asosida tavsiflab chiqish talab etiladi:

Ibora	Tavsif	Kategoriya
A’zoyi badanidan sovuq ter chiqib ketdi	ruhiyati keskin o‘zgarmoq	salbiy
A’zoyi badanidan o‘tib ketmoq	xafa bo‘lmoq	salbiy
Ahd qilishmoq	ahdlashmoq	neytral
Ahvoli tang	juda yomon	salbiy

Og'zi qulog'ida

xursand

ijobiy

1.2-jadval. Iboralarning hissiy tahlil asosida kategoriyalarga bo'linishi

Iboralar ekspressivlik jihatidan bildirilayotgan predmet, shaxs yoki voqea-hodisaga qaratilgan subyektiv munosabatni yuqori ekspressivlik bilan namoyon etuvchi birlik. Ular nazariy tilshunoslik, leksikografiya tarmoqlari bo'yicha tahlil qilingan bo'lsa, hozirgi vaqtda uning tadqiqi tabiiy tilni qayta ishlashda (NLP) hissiy tahlil uchun ahamiyatli hisoblanadi.

Xulosa

O'zbek tilida hislarni ifodalovchi birliklar, jumladan, so'z, so'z birikmasi, iboralar o'zining umumiy va xos xususiyatlariga ko'ra hissiy ma'no tashiydi va ularni shu xususiyatlaridan kelib chiqib tadqiq etish talab qilinadi. Bundan tashqari hissiy birliklar lingvistik jihatdan shunchaki so'z, so'z birikmasi yoki iboralar emas, balki tildan foydalanuvchi shaxsga o'z hislarini ifoda etish, boshqalar bilan munosabatga kirishishda aloqa o'rnatish, ijobiy, salbiy yoki neytral bahosini berish imkonini yaratuvchi vosita. Shuning uchun shaxslararo muloqotda kishilarning subyektiv munosabatini bildiruvchi hissiy ma'no tashuvchi birliklarni tahlil qilish hissiy tahlil uchun muhimdir.

Qilinadigan ishlar tabiiy tilni qayta ishlashning kichik tarmog'i sanaladigan sentiment analiz uchun zaruriy manba bo'lib xizmat qiladi. Bu orqali internet tarmog'idan foydalanuvchi o'zbek segmentining mamlakat hududida yuz berayotgan voqealarga, ko'rsatilayotgan xizmatlar va foydalanilayotgan mahsulotlar borasida bildirgan shaxsiy fikrlarini real vaqt davomida statistik tahlil qilish imkoniyati tug'iladi.

Foydalanilgan adabiyotlar:

1. Abdullayev A. O'zbek tilida ekspressivlik ifodalashning sintaksik usullari. - Toshkent, 1983.
2. Rahmatullayev Sh. O'zbek tili frazeologik lug'ati. -Toshkent: G'afur G'ulom, 2022.
3. Turansky I. Sementicheskaya kategoriya intensivnosti v angliyskom yazike. - Moskva: Vysshaya shkola, 1990.
4. Sharipov M. Hozirgi o'zbek tilida so'z birikmalari sintaksisi masalasi. - Toshkent: Fan, 1978.
5. www.statista.com

SEMANTIK TIZIMLAR VA LINGVISTIK ONTOLOGIYALAR

O'ZBEK TILIDA FE'L SO'Z TURKUMI UCHUN VERBNET YARATISH

Sharipov Maqsud Siddiqovich

Texnika fanlari nomzodi, dotsent,

m.sharipov@urdu.uz

UrDU, Kompyuter ilmlari kafedrası dotsenti

Mammatova Mahbuba Davronbek qizi

mammatovamahbuba1996@mail.com

UrDU, Kompyuter ilmlari kafedrası magistranti

Annotatsiya. Tabiiy tilni qayta ishlash masalalari orasida fe'l so'z turkimiga oid so'zlarni qayta ishlash masalalarini yechish muhim o'rin egallaydi. O'zbek tilida fe'llarni stemming va lemmatizatsiya qilish hamda kompyuter tarjima dasturlarini yaratish kabi masalalarni yechish uchun fe'llar uchun VerbNet yaratish masalasi dolzab hisoblanadi. Ushbu ishda o'zbek tili uchun VerbNetning tuzilmasini loyihalash va uni yaratish masalasi ko'rib chiqilgan.

Annotation. Among the problems of natural language processing, solving the problems of processing words related to the word group of verbs occupies an important place. The issue of creating a VerbNet for verbs to solve problems such as stemming and lemmatization of verbs in the Uzbek language and creating computer translation programs is considered important. In this work, the issue of designing and creating the structure of VerbNet for the Uzbek language is considered.

Аннотация. Среди задач обработки естественного языка важное место занимает решение задач обработки слов, относящихся к словесной группе глаголов. Важным считается вопрос создания VerbNet глаголов для решения таких задач, как stemming и лемматизация глаголов в узбекском языке и создание программ компьютерного перевода. В данной работе рассматривается вопрос проектирования и создания структуры VerbNet для узбекского языка.

Kalit so'zlar. *Sinonim, giperonim, giponim, antonim, paronim, verbnet, wordnet.*

Tabiiy tillarni qayta ishlash masalalarini yechish uchun tildagi so'zlarni bir-biriga bog'lovchi ma'lumotlar bazasini yaratish muhim o'rin egallaydi. Shu sababli ushbu masalani yechish uchun WordNet yaratilgan. Wordnet bu so'z mazmunlari asosida yaratilgan juda keng qamrovli lug'atdir. Wordnetning asosiy qurilish bloklari sinsetlar bo'lib, ular sinonim sinset a'zolaridan iborat. O'zbek WordNetiga asos bo'lgan Turk WordNeti strukturasi, foydalanish holati Til manbalaridan so'zlar munosabatlari chiqarib olingan. Dastlabki 1310 ta sinsetlar to'plamining

tarjimasidan so'ng, turk jamoasi mashina tomonidan o'qiladigan turk tilining izohli lug'atidan sinonimlar, antonimlar va giponimlarni avtomatik ravishda chiqarib olishga harakat qilishdi [Madatov, 2022].

Fe'l sintezlari ham iyerarxiyalarga joylashtirilgandir. Daraxtning eng tubida joylashgan fe'llar (troponimlar) biror hodisani o'ziga xos uslubda tavsiflab beradi, masalan, suhbat qurmoq – gaplashmoq – pichirlashmoq. Belgilangan uslub semantik maydonga bog'liq bo'ladi; Hajm fe'llarni batafsil ishlab chiqishning bir o'lchovi xolos. Boshqa o'lchovlar tezlik (siljish-yurish-yugurish) yoki tuyg'ular kuchliligi (yoqtirmoq-sevmoq-maftun bo'lmoq) bo'lishi mumkin. Bir-biriga doimiy va bir tomonlama ta'sir ko'rsatadigan, hodisalarni tavsiflovchi fe'llar o'zaro bog'langandir. Masalan, sotmoq – sotib olmoq, ko'rmoq -ko'rsatmoq, erishmoq – harakat qilmoq [Ravshanov, 2021].

WordNet ingliz tilining leksik ma'lumotlar bazasi hisoblanadi. Undagi ot, fe'l, sifat va ravish so'z turkumlari muayyan konseptni ifodalovchi sinonimlar guruhiga (Sinset) birlashgan. Sinsetlar konseptual-semantik va leksik munosabatlarga ko'ra o'zaro bog'langan. WordNetdan foydalanish va uni o'rnatib olish qulay. Uning tuzilishi kompyuter lingvistikasi va tabiiy tilni qayta ishlash uchun foydali. WordNetda fe'l so'z turkumiga kiruvchi troponimlar kiritilgan. Unga ko'ra u yoki bu fe'l boshqa muayyan shakliga ko'ra bog'lanadi. Masalan: vaysamoq, gapirmoq fe'lining troponimi hisoblanadi. Shuningdek fe'llar ham antonimik munosabatda bo'lishi mumkin turmush qurmoq-ajrashmoq [Abduraxmonova, Haydarov].

VerbNet_ - Bu fe'llarni semantik, sintaktik bog'lanish harakati asosida guruhlaydigan leksik manba. Biz 5 tomlik o'zbek tilining izohli lug'ati tarkibida mavjud 8519 ta fe'llar bazasini to'pladik. To'plangan fe'l shakllarini tuzilishi va ma'no guruhi jihatdan belgilab oldik. Ro'yxatdagi yasama fe'l shakllarining asosi qaysi turkumga oidligi, qanday qo'shimcha orqali yasalgani, fellar tarkibidagi mavjud lug'aviy shakl yasovchi qo'shimchalarni (nisbat yasovchi qo'shimchalar) guruhlab tahlil qildik. Bazadagi fe'llarning mavjud sinonim, antonim, paronim, giponim va giperonimlari to'plandi. Biz hozirgacha 447 ta sinonimli fe'l, 91 ta paronimli fe'l, 86 ta antonimli fe'l shakllarini to'pladik. Bundan tashqari to'plangan fe'llarni qanchasi tub (faqat asosdan yoki shakl yasovchi qo'shimchali), qanchasi esa so'z yasovchi qo'shimchalar orqali boshqa turkumlardan yasalganligini aniqladik. Va nechta fe'l shakli qaysi so'z turkumidan yasalganligining statistikasi chiqarildi. Qaysi qo'shimcha qaysi turkumdagi so'zdan fe'l shakli yasashi va nechta o'rinda yasaganini hisobladik. Bu ma'lumotlar 1-jadvalda ko'rsatib o'tiladi:

1-jadval

Jami fe'llar	8518
Tub fe'llar	3461
Otdan yasalgan fe'llar	2447
Sifatdan yasalgan fe'llar	1389

Taqlid soʻzdan yasalgan feʼllar	1143
Ravishdan yasalgan feʼllar	35
Sondan yasalgan feʼllar	17
Olmoshdan yasalgan feʼllar	17
Modal soʻzdan yasalgan feʼllar	8

Yuqoridagi maʼlumotlarga tayangan holda har bir turkumdan qaysi qoʻshimcha orqali nechta feʼl shakli hosil qilinganiga asoslanib eng koʻp feʼl shakli yasagan qoʻshimchalarni jadval orqali berib oʻtamiz.

2-jadval

la	3390ta
illa	414ta
lash	410ta
lan	331ta
a	24ta
ulla	60ta

SINONIM - bir umumiy maʼnoga ega boʻlgan qoʻshimcha maʼnosi uslubiy va boshqa munosabat kabilarni ifodalovchi xususiyatlari bilan oʻzaro farqlanadigan til birliklari. Misol uchun: kulmoq, jilmaymoq, tirjaymoq, ishshaymoq. Oʻzaro sinonim boʻlgan soʻzlar guruhi sinonimik qator deyiladi. Yuqoridagi sinonimik qatordagi soʻzlar bir umumiy “kulish” manosi ega, ammo “kulmoq” soʻzi va “tirjaymoq”, “ishshaymoq” soʻzlari bir-biridan uslubiy jihatdan farqlanadi.

GIPERONIM - jins belgisini bildirgan predmetning nomini ifodalovchi koʻpgina maʼnoni semantik jihatdan umumlashtiruvchi mikrotizimning markaziy leksemasi, dominantasi sifatida namoyon boʻluvchi lugʻaviy birlik. Misol uchun: kulmoq, jilmaymoq, tirjaymoq, ishshaymoq qatorida dominanta soʻz “kulmoq” giperonim.

GIPONIM - maʼlum jins turining nomini hamda oʻzining semantik tarkibida implist tarzida jins maʼnosini ham ifodalovchi, semantik jihatdan giperonimga nisbatan boy boʻlgan lugʻaviy birlik. Misol uchun: kulmoq-giperonim; jilmaymoq, tirjaymoq, ishshaymoq-giponim. Feʼl maʼnoviy guruhlarida ham markaz va qurshov leksemalari farqlanadi. Masalan, maʼnodosh feʼllarda bosh (dominanta) soʻz, uyadoshlik paradigmasida uya (giperonim) markaz leksemasi hisoblansa, boshqa maʼnodosh va uyadosh (giponimlar) qurshov leksemasi deyiladi. [Kulmoq] feʼli dominanta sifatida bunga maʼnodosh boshqa leksemalarni, [buzmoq] esa buzishning turli koʻrinishini ifodalaydigan giponim leksemani uyushtirib turadi. Soʻzlar yirikroq butunliklarga birlashtirilganda, markaz leksemasigina bu butunlikdan joy oladi. Qurshov leksemalari esa uning “soyasi” sifatida quyida qolaveradi. Masalan, [kulmoq], [yigʻlamoq], [xursand boʻlmoq] leksemalari oʻz atrofida koʻplab feʼllarni birlashtirgan. Lekin ular bu uch leksema mansub tizimga kira olmavdi, quyi bosqichda qoladi [Sayfullayeva va b., 2009].

ANTONIM-bir-biriga qarama-qarshi ma'nolarni ifodalovchi so'zlar juftligi. Antonimik juft hosil bo'lishi uchun ikkita mustaqil tushuncha ma'no jihatdan o'zaro qarama-qarshi bo'lishi kerak. Fellardagi bo'lishli-bo'lishsizlik xususiyati antonimlikni vujudga keltirmaydi. Misol uchun: keldi-ketdi, maqtamoq-qoralamoq, kirmoq-chiqmoq.

PARONIM - talaffuzi, eshitalishi va morfem tarkibi o'xshash, leksik ma'nolari boshqa-boshqa yoki qisman yaqin bo'lgan so'zlar. Misol uchun: alishmoq-olishmoq, artmoq-archmoq, chuchimoq-cho'chimoq.

Foydalanilgan adabiyotlar:

1. Abdurakhmonova N., Khaydarov M. (2019). Ўзбек тилида wordnet яратиш масалаларига доир. Foreign Languages in Uzbekistan. 19-27. 10.36078/1570167968.
2. Abjalova M, Rashidova U, Abdumuminov B. Matnlarni qayta ishlash manbalari. Computer linguistics: problems, solutions, prospects 2022;
3. Akhmedovich Khabibulla. (2022). Creating of the Uzbek WordNet based on Turkish WordNet. AIP Conference Proceedings. 2432. 10.1063/5.0089532.
4. Boltayevich E.B, Adali E, Mirdjonovna Sh.X, Xolmo'minovna A.O, Yuldashevna X.Z, Nizomaddin Uktamboi o'g'li X. The Problem of Pos Tagging and Stemming for Agglutinative Languages (Turkish, Uyghur, Uzbek Languages). UBMK 2023 - Proceedings: 8th International Conference on Computer Science and Engineering, 2023, 57 – 62.
5. Ravshanov Shuhratjon. (2021). Wordnet–online leksik ma'lumotlar bazasi. Computer linguistics: problems, solutions, prospect. 1. 149-152. 10.52773/cl:psp/vol_1_issue_1/a30.
6. Sayfullayeva R.R., Mengliyev B.R., Boqiyeva G.H., Qurbonova M.M., Yunusova Z.Q., Abuzalova M.Q. Hozirgi o'zbek adabiy tili. O'quv qo'llanma. - T., 2009. - 416 b.
7. Sharipov M, Kuriyozov E, Yuldashev O, Sobirov O. UzbekTagger: The rule-based POS tagger for Uzbek language. arXiv preprint arXiv:230112711 2023;
8. Sharipov M, Mavlonova M, Sobirov O. O'zbek tilidagi matnlarda ot va olmosh so'z turkumlariga tegishli so'zlarni aniqlovchi algoritm yaratish. Computer Linguistics: Problems, Solutions, Prospects 2023; 1.
9. Sharipov M, Salaev U, Matlatipov G. Implemented stemming algorithms based on finite state machine for uzbek verbs| computer linguistics: problems. Solutions, prospects 2021;
10. Sharipov M, Sobirov O. Development of a Rule-Based Lemmatization Algorithm Through Finite State Machine for Uzbek Language. CEUR Workshop Proceedings, CEUR-WS 2022, 154 – 159.
11. Sharipov M, Yuldashov O. UzbekStemmer: Development of a Rule-Based Stemming Algorithm for Uzbek Language. CEUR Workshop Proceedings, CEUR-WS 2022, 137 – 144.

SUN'IY INTELLEKT TIZIMIDA OMONIMLARNI TUSHUNISH VA TARJIMA QILISH MUAMMOLARI

Abdurahmanova Muqaddas Tursunaliyevna

Mabduraxmanova20@gmail.com

filologiya fanlari nomzodi, dotsent, ToshDO‘TAU doktoranti

Annotatsiya. Kompyuter lingvistikasi taraqqiyoti natijasida sun'iy intellekt qo'llanadigan yo'nalishlar doirasi ham kengaydi. Shu bois kompyuter lingvistikasiga doir adabiyotlarda sun'iy intellekt kompyuter lingvistikasining bir yo'nalishi sifatida talqin qilingan. Maqolada sun'iy intellekt tizimida omonimlarni tushunish va tarjima qilish muammolari tahlil qilingan.

Abstract. As a result of the development of computational linguistics, the range of areas in which artificial intelligence is used has expanded. Therefore, in the literature on computational linguistics, artificial intelligence is treated as a branch of computational linguistics. The article analyzes the problems of understanding and translating homonyms in an artificial intelligence system.

Аннотация. В результате развития компьютерной лингвистики расширился и спектр направлений, в которых используется искусственный интеллект. Поэтому в литературе по компьютерной лингвистике искусственный интеллект трактуется как раздел компьютерной лингвистики. В статье анализируются проблемы понимания и перевода омонимов в системе искусственного интеллекта.

Kalit so'zlar: *Sun'iy intellekt, semantika, modellash, ChatGPT, Prolog tili, omonim.*

Sun'iy intellekt – bu oxirgi o'n yillikda turli fan sohalarda, jumladan, kompyuter lingvistikasida faol qo'llanayotgan tushunchalardan biridir. Sun'iy intellekt fanlararo soha hisoblanadi, keyingi yillarda kompyuter lingvistikasi bilan yonma-yon istifoda etilmoqda. Sun'iy intellekt integrativ xususiyatga ega bo'lib, uning turli ko'rinish va usullaridan bir qator sohalarda foydalaniladi. Kompyuter lingvistikasi nutqni va matnlarni tadqiq qilishning kompyuter usullarini taqdim etar ekan, psixologiya, mantiq, sotsiolingvistika kabi fanlar qatorida sun'iy intellekt ko'rinishlarini ham qamrab oladi.

Virtual qomus hisoblangan Wikipedia sahifalarida sun'iy intellektga (artificial intelligence) shunday ta'rif berilgan: “Sun'iy intellekt – insonning aqliy faoliyati bilan bog'liq bo'lgan masalalarni, xususan, fikrlash, nutq mazmunini anglash, ma'lumotlarni umumlashtirish kabilarni yechishga yo'naltirilgan kompyuter yoki robot imkoniyatidir. Bu termin mazkur xususiyatlarga ega tizimlarni ishlab chiqish bilan bog'liq kompyuter texnologiyalarining bir bo'limiga nisbatan ham

ishlatilmoqda”.¹³ Sun’iy intellekt bo’yicha tadqiqotlar olib borgan olim Jon Makkarti shunday yozadi: “Sun’iy intellekt – intellektual mashinalar, intellektual kompyuter dasturlarini yaratish texnologiyasi va u haqidagi fan. Intellektual sistema strukturalari uch asosiy blokni o’z ichiga oladi: bilimlar bazasi, masalani yechuvchi tizim va intellektual interfeys”¹⁴. Bu ta’riflardan kelib chiqib, sun’iy intellekt fan, soha yoki yo’nalish emas, “kompyuter texnologiyasi tizimi”, “intellektual faoliyat”, “imkoniyat” sifatida baholash to’g’ri bo’ladi.

XX asrning 70-yillariga qadar sun’iy intellekt bo’yicha tadqiqotlar kibernetika hamda informatika doirasida olib borilgan. XX asrning 80-90-yillaridan boshlab sun’iy intellekt juda ko’plab fanlarning o’rganish obyekti bo’ldi. Jumladan, neyrolingvistika, psixologiya, informatika, neyrofiziologiya, epistemologiya (falsafadagi bilish haqidagi ta’limot), kognitologiya, kognitiv lingvistika, kompyuter lingvistikasi kabi fanlar ham o’z doirasida sun’iy intellekt muammosi bilan shug’ullanadi.

Sun’iy intellekt tizimlarini yaratish muammosiga uch xil asosiy yondashuvni ajratib ko’rsatish mumkin [Рассел, 2006]:

- quyi (inglizcha *Top-Down AI*), semiotik yondashuv – ekspert tizimlari, bilimlar bazasi, mantiqiy hukm chiqarish tizimi, yuqori darajadagi psixik jarayonlar (tafakkur, nutq, emotsiya, ijod kabi)ning imitatsiyasi va analogi bo’lgan tizimlarni yaratish;

- yuqori (inglizcha *Bottom-Up AI*), biologik yondashuv – neyron tarmoqlarini o’rganish, biologik unsurlarga asoslangan holda intellektual salohiyatni modellashtirish va neyrokompyuter yoki biokompyuterlarni yaratish;

- gibrid tizimli yondashuv – semiotik va biologik yondashuvning sintezi bo’lgan umumlashtirish. Mazkur yondashuvlardan tashqari, simvolli, mantiqiy, agentga yo’naltirilgan yondashuvlar ham mavjud. Simvolli yondashuv, asosan, simvolik mantiq qoidalarini tizimga tatbiq etish bilan bog’liq bo’lib, bu ko’proq tizim qoidalarini ishlab chiqishda, hisoblashda (hukmning chin yoki yolg’onligini baholashda) samara beradi. Mantiqiy yondashuv predikatlar tilidan foydalanish asosida sun’iy intellekt tizimini bilimlar bazasining mantiqiy modellari bilan ta’minlash uchun ximat qiladi. 1980-yilda Prolog nomli mantiqiy programmashtirish tizimi va tili yaratildi. Prolog tilida yozilgan bilimlar bazasi faktlar naborini hamda mantiqiy xulosa chiqarish qoidalarini belgilaydi. Agentga yo’naltirilgan yondashuv (yoki intellektual, ratsional agentlardan foydalanishga asoslangan yondashuv) 1990-yillarda rivojlandi. Unga ko’ra, intellekt – bu qo’yilgan masalani hisoblash qismi (rejalashtirish qismi), xolos. Bunda intellekt masalani to’laqonli yechish uchun atrofdagi tizimlar munosabatga kirisha oladigan qo’shimcha intellektual (ratsional) agentga ehtiyoj sezadi.

¹³ http://www.wikipedia.org/wiki/artificial_intelligence.html

¹⁴ <http://www-formal.stanford.edu/jmc/whatisai.html>



Sun'iy intellekt tizimi insonning ijodiy tafakkurini modellashtirishni ham o'z ichiga oladi. Shu bilan bog'liq ravishda san'at asarlarini avtomatik tarzda yaratish jarayonini aks ettirgan “mashina ijodi” tushunchasi yuzaga keldi. Bugungi kunda ijodni to'la modellashtirishga erishilgani yo'q, lekin bu sohada sezilarli yutuqlar qo'lga kiritilgan. Jumaladan, kompyuter yordamida musiqalar yaratilishi, treker-dasturlar, sintezatorlar, sekvensorlar yordamida bir turdagi melodiylar va tovushlar generatsiyasi oqibatida elektron musiqalarning yaratilishi, matnning avtomatik generatsiyasi yordamida she'rlar, ertaklar – umuman matnlar yaratilishi, rasmlar generatsiyasi va kompyuter grafikasi yordamida kino va o'yinlarning yaratilayotganligi (ayniqsa, strategiya va harbiy manyovrlarga asoslangan kompyuter o'yinlari) mashina ijodiga misol bo'la oladi.

Sun'iy intellekt tizimi doirasida tatbiq etilgan loyihalar (eksperimentlar) sifatida quyidagilarni ajratib ko'rsatish mumkin:

- Deep **Blue** tizimi – IBM kompaniyasi tomonidan ishlab chiqilgan shaxmat bo'yicha superkompyuter. Deep Blue 1997-yil 11-mayda shaxmat bo'yicha jahon chempioni Garri Kasparov ustidan g'alaba qozondi. “Deep Blue” nomi “Deep Thought” (“Chuqur mulohaza, tafakkur” ma'nosini anglatadi) loyihasiga aloqadordir.

- MYCIN tizimi – bir qator kasalliklarga to'g'ri tashxis qo'yishni ta'minlaydi.

- 20Q – “20 savol” klassik o'yini asosida yaratilgan loyiha bo'lib, u sun'iy intellekt bo'yicha o'tkazilgan dastlabki tajribalardan biridir. Bu o'yin muallifi Robin Byorgenerdir. Bunda foydalanuvchidan biror obyektни toppish so'raladi, so'ng foydalanuvchiga 20 ta savol beriladi, berilgan javoblardan kelib chiqqan holda o'sha narsa topiladi. Ayniqsa, bu tizim internet paydo bolgandan so'ng juda mashhur bo'ldi.

- Via Volce – nutqni tovush orqali tanish tizimi.

- RoboCup – futbol bo'yicha robotlar turniri.

“Sun'iy intellekt tizimini baholash bo'yicha ikki xil nuqtayi nazar mavjud: 1) sun'iy intellektни inson tafakkuridan ko'ra kuchliroq, deb baholash. Ushbu qarash tarafdorlari sun'iy intellekt inson intellektiga qaraganda ko'proq operatsiyalarni qisqa muddatlarda bajarishi mumkin, deb hisoblaydilar; 2) sun'iy intellekt inson tafakkuridan, intellektidan kuchsiz. Bu yo'nalishdagi olimlarning fikricha, sun'iy intellektning imkoniyatlari cheklangan bo'ladi, u faqat dasturiy ta'minot asosida muayyan sohalarida ishlaydi, lekin u bilish qobiliyatining, tafakkurning barcha qirralarini qamrab ololmaydi. Buni Jon Syorl tomonidan o'tkazilgan “Xitoy uyi” nomi bilan mashhur bo'lgan semantik eksperiment ham tasdiqlaydi. Olim bu tajribasini 1980-yilda “*Minds, Brains and Programs*” (“Tafakkur, miya va programmalar”) nomli maqolasida e'lon qildi. J.Syorlning fikricha, *kompyuter sintaktik qoidalarni algoritm nazariyasi orqali formallashtira oladi, lekin semantika, tafakkurni to'la modellashtirishga dasturlar ojizlik qiladi* [Searle, 1980]. Xuddi shu pozitsiyada turgan olimlardan yana biri Rodjer Penrouz o'zining “*Qirolning yangi aqli*” kitobida tafakkurni formal tizimlar asosida mutlaq

modellashtirishning imkoni yo'qligini bayon qilgan [Penrouz, 2005].

Sun'iy intellekt yo'nalishidagi tadqiqotlar sirasiga ekspert tizimlari yoki neyron tarmoqlari bo'yicha olib borilgan amaliy tadqiqotlar ham kiradi. Mazkur ikki yo'nalishga nisbatan “kompyuter intellekti” («computer intelligence») termini ishlatilmoqda, bunda inson tafakkurini global modellashtirish jarayoni hamda murakkab masalalarni operativ tarzda yechishga mo'ljallangan avtomatlarning yaratilishi tushuniladi.

Internet tizimida keng qo'llanayotgan sun'iy intellektlardan biri ChatGPT nomli sun'iy intellekt bo'lib, GPT (Generative Pre-trained Transformer) arxitekturasiga asoslangan OpenAI tomonidan ishlab chiqilgan katta til modelidir. Bu sun'iy intellekt tabiiy til modellari asosida insonning fikrlash tarziga muvofiq javoblar tizimini yaratish uchun mo'ljallangan. Bu sun'iy til insonlar bilan muloqot qilish, suhbatlashish imkonini beradi. ChatGPT internetdagi ingliz tilidagi matnlarning katta hajmdagi ma'lumotlar bazasiga asoslanadi, jarayonning bir necha bor takrorlanishi natijasida tilning xususiyatlari va tuzilmalarini o'rganish, til xususiyatlarini o'zlashtirish va inson bilimi, madaniyatini chuqur tushunish imkonini beradi. ChatGPT:

- 1) AI (sun'iy intellekt) modeli hisoblanadi;
- 2) keng ko'lamli savollarga javob beradi;
- 3) turli mavzularga oid ma'lumot bera oladi;
- 4) inson bilan muloqot qila oladi;
- 5) doimiy ravishda rivojlanib boradi va takomillashtiriladi;
- 6) insonga ma'lumot olishda yordam beradigan, muammolarni hal qilish va o'zaro yanada samarali muloqot qilishda yordam beradigan kuchli vositadir.

Amalda qo'llanish jarayonida ChatGPT foydalanuvchilardan ko'proq ma'lumot va fikr-mulohaza oladi.

ChatGPTga ingliz tilidagi 100ta omonimni topish vazifasi yuklanganda, topshiriq deyarli xatosiz bajariladi. Sun'iy intellekt ingliz tilidagi omonimlarning ma'nosini farqlab, bir-biriga oppozitsiya holatida ajratib beradi:

Here are 100 examples of homonyms in English:

- 1) bank (financial institution) – bank (river bank); 2) bear (animal) – bear (to carry); 3) bow (bend forward) – bow (for arrows); 4) can (able to) – can (container); 5) close (nearby) – close (shut); 6) date (fruit) – date (day/month/year); 7) fair (just) – fair (carnival); 8) file (folder) – file (tool); 9) lead (metal) – lead (to guide); 10) lie (untruth) – lie (to recline).

Turk tilidagi 100ta omonimni ajratish so'ralganda ayrim xatolarga yo'l qo'ygan holda ma'lumot beradi:

- 1) banka (finansal kurum) – banka (nehir kenarı); 2) baş (vücut bölümü) – baş (başlamak); 3) bay (erkek) – bay (koy); 4) beyaz (renk) – beyaz (un); 5) can (yaşam) – can (kutu); 6) dava (mahkeme olayı) – dava (çanta); 7) değişim (farklılaşma) – değişim (para bozdurma); 8) düş (hayal) – düş (düşmek); 9) el (vücut bölümü) – el (ölçü birimi); 10) ev (konut) – ev (işareti).

O'zbek tilidagi 100ta omonimni ajratish so'ralganda butunlay xatoga yo'l qo'yadi: 1) qal'a – fortress (himoya qal'asi), qal'a – palace (qasr); 2) boshi – beginning (nimaningdir boshlanishi), boshi – head (tana a'zosi); 3) oq (rang) – white, oq – learn (o'qimoq); 5) qo'ziq – ear (quloq), qo'ziq – ankle (tovon); 6) ko'z – eye, ko'z – spring (bahor, sakrash); 7) to'q – round (aylana shaklidagi jarayon), to'q – straight (to'g'ri); 8) qulupnay – milkmaid (sutga bog'liq), qulupnay – type of shoe (oyoq kiyim turi); 9) tosh – stone, tosh – weight (og'irlik); 10) qiziqtirg'u – curiosity (любопытство), qiziqtirg'u – type of bird (вид птицы); yurak – heart, yurak – pear (nok).

ChatGPTning o'zbek tilidagi omonimlarni ajratishda unlilarni farqlamasligi kuzatiladi: *qo'sh* so'zini *qush* bilan almashtiradi: qo'shquloq – type of bird. Ko'p ma'noli so'zlarni ham omonim sifatida ajratish hollari uchraydi: boshi – beginning (nimaningdir boshlanishi), boshi – head (tana a'zosi).

O'zbek tilida omonim sifatida ajratilgan 100ta shakldosh juftlikdan bor-yo'g'i uchtagina ChatGPT tomonidan to'g'ri tanlangan: ko'k – blue, ko'k – sky (osmon); qo'y – sheep; qo'y – put; qo'y – sheep, qo'y – fire.

Sun'iy intellektda o'zbek tilidagi muloqot kam bo'lgani uchun, o'zbek tilidagi ma'lumotlar bazasi kam o'zlashtirilgani bois ChatGPT o'zbek tilidagi omonimlarni farqlay olmadi.

Avtomatik tarjimada samaradorlikning yuqori bo'lishiga tabiiy tillarning ayrim jihatlari – polisemiya (ko'p ma'nolilik), omonimlik, frazeologizmlar, idiomalar, paremlar, troplar to'sqinlik qiladi. Shularni istisno qilganda, hozirda avtomatik tarjima yo'nalishida yuqori samaradorlikka erishilgan. Hatto hozirda faqat avtomatik tarjima yo'nalishiga bag'ishlangan saytlar ham yaratilgan. Bu avtomatik tarjimaning keng qamrovda taraqqiy etayotgan yo'nalishlardan ekanligini ko'rsatib turibdi [Xolmanova, 2020].

Foydalanilgan adabiyotlar:

1. Рассел С. Норвиг П. Искусственный интеллект: современный подход / Artificial Intelligence: a Modern Approach /Пер. с англ. и ред. К.А.Птицына. 2-е изд. - М.: Вильямс, 2006. – С.8.
2. Po'latov A., Muhamedova S. Kompyuter lingvistikasi. –T., 2007. Rahimov A. Kompyuter lingvistikasi asoslari. –T., 2011.
3. Xolmanova Z. Kompyuter lingvistikasi. –Toshkent, 2020. –B.140.
4. Searle J.R. Minds, brains, and programs // Behavioral and Brain Sciences. 1980. - P.417- 437
5. Пенроуз Р. Новый ум короля. О компьютерах, мышлении и законах физики. - М.: УРСС, 2005. - P.14.
6. http://www.wikipedia.org/wiki/artificial_intelligence.html
7. <http://www-formal.stanford.edu/jmc/whatisai.html>
8. http://www.wikipedia.org/wiki/artificial_intelligence.html

MATN BIRLIKLARINI SEMANTIK TEGGLASHNING AHAMIYATI

Hamroyeva Shahlo Mirdjonovna

shaxlo.xamrayeva@navoiy-uni.uz

filologiya fanlari doktori, ToshDO‘TAU dotsenti

Hamroqulova Marjona Nabijon qizi

hamroqulovamarjona@mail.ru

ToshDO‘TAU tayanch doktoranti

Annotatsiya. Mazkur maqolada ko‘p vazifali o‘qitish (Multi-task learning) turkumiga kiruvchi semantik teglashdan foydalanish usullari tahlil qilingan. Tabiiy tilni qayta ishlashning uchta turli vazifasi, jumladan, so‘z turkumini teglash, universal bog‘liqlikni tahlil qilish va semantik teglashdan yordamchi vazifa sifatida foydalanish tavsiflangan. To‘liq neyron tarmoq, qisman neyron tarmoqda o‘qitish va vazifalar o‘rtasida salbiy o‘tkazish ehtimoli kamroq bo‘lgan tizimlarni o‘qitish usullari taqqoslangan.

Abstract. We investigate the effects of multi-task learning using the recently introduced task of semantic tagging. We employ semantic tagging as an auxiliary task for three different NLP tasks: part-of-speech tagging, Universal Dependency parsing, and Natural Language Inference. We compare full neural network sharing, partial neural network sharing, and what we term the learning what to share setting where negative transfer between tasks is less likely.

Аннотация. В данной статье анализируются методы использования семантической разметки, которая относится к категории многозадачного обучения. Описаны три различные задачи обработки естественного языка, включая маркировку фразы, анализ универсальной релевантности и использование семантической маркировки в качестве вспомогательной задачи. Сравняются полное обучение нейронной сети, частичное обучение нейронной сети и системы обучения с меньшей вероятностью отрицательного перехода между задачами.

Kalit so‘zlar: *semantik teglash, ko‘p vazifali o‘qitish, to‘liq taqsimlangan tarmoq, qisman taqsimlangan tarmoq, so‘z turkumini teglash.*

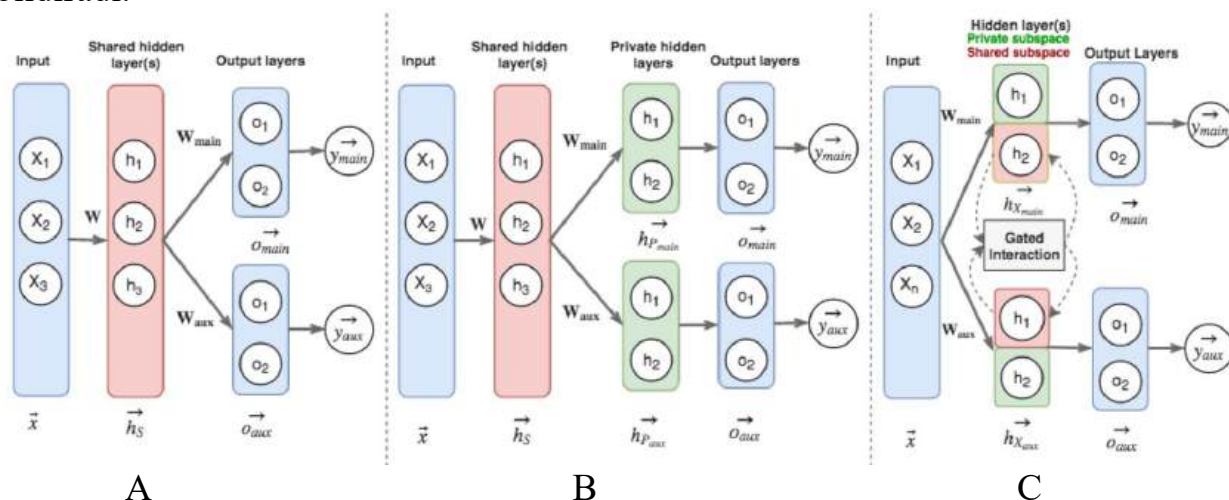
Semantik razmetka (rus., разметка) – til korpuslaridagi so‘z yoki so‘z birikmasining ma’lum bir semantik kategoriya yoki undan ham kichikroq semantik guruh (LSG, semantik maydon va to‘da)ga mansubligini bildiruvchi, ma’noni maxsuslashtiruvchi belgi, izohlar majmuyi [Захаров, Богданова, 2011:48].

Zamonaviy avtomatik-semantik razmetkalash tizimlari lug‘atlardan farqli o‘laroq, semantik tarmoq yoki leksik tasniflarga bog‘lanadi. Ular orasida Oxford Advanced Learners Dictionary (OALD)ga asosan ma’noni parchalab qo‘llovchi WordNet [Кустова, Падучева, 1994] mashhur. Ushbu avtomatik-semantik razmetkalash dasturi asosida ilk marta Braun korpusi razmetkalan. Dastur

234136 so‘zshakl razmetkasini o‘z ichiga olgan. Ushbu so‘zshakllarining 186575 tasi ko‘p ma’noli.

Ko‘p vazifali o‘qitish bir vaqtning o‘zida bir nechta vazifalar o‘qitiladigan mashinali o‘qitishga nisbatan yaqinda qayta tiklangan yondashuv hisoblanadi. Bir vaqtning o‘zida bir-biriga bog‘liq bo‘lgan bir nechta funksiyalarni optimallashtirish orqali ko‘p vazifali o‘qitish modellari bitta vazifaga o‘qitilgan modellarga qaraganda yuqori natijalarga erishishi mumkin. Bu bo‘yicha asosiy tamoyil Karuana [Rich Caruana, 1998:95] tomonidan shunday xulosalangan: “Ko‘p vazifali o‘qitish tegishli vazifalarning o‘qitish signallarida mavjud bo‘lgan domenga xos axborotlardan foydalanish orqali umumlashtirishni yaxshilaydi”. Neyro ko‘p vazifali o‘qitish tabiiy tilni qayta ishlash vazifalari o‘rtasidagi o‘xshashliklardan foydalangan holda borgan sari muvaffaqiyatli yondashuvga aylanmoqda [Anders Sogaard, Yoav Goldberg, 2016:231].

Semantik teglash, odatda, mavjud bo‘lgan Penn Treebank teglari kabi sintaksisga yo‘naltirilgan so‘z turkumlarining teglar to‘plamidagi semantik ma’lumotlarining yetishmasligini bartaraf etishga mo‘ljallangan. Bunday teglar to‘plami muhim semantik farqlarni, jumladan, inkor va modal shakllar, miqdorni aniqlash turlari, atoqli ot turlari va fe’llarning zamon, harakat yoki holat ma’nolarini bildiradi.



1-rasm. Uchta ko‘p vazifali o‘qitish tizimlari: (A) to‘liq taqsimlangan tarmoqlar, (B) qisman taqsimlangan tarmoqlar va (C) nimani taqsimlashni o‘qitish. Qatlamlar matematik jihatdan vektorlar bilan; ular orasidagi o‘qlar bilan ifodalangan bog‘lanishlar matematik jihatdan og‘irliklar matritsalarini belgilanadi. S taqsimlangan qatlamni, P xususiy qatlamni, X taqsimlangan va shaxsiy maydonlarga ega qatlamni bildiradi.

Semantik teglar to‘plami tilga nisbatan neytral hisoblanib, so‘z turkumi va atoqli ot sinflari bo‘yicha mavhum bo‘ladi, bir qancha semantik ma’lumotlarni o‘z ichiga oladi. Teglar to‘plami 13 ta yirik sinfga birlashtirilgan 80 nafar semantik tegdan iborat. Teglar to‘plami kompozitsion semantikaga va semantik tasvirlashning tillararo proyeksiyasida qo‘llangan Parallel Ma’no Banki loyihasida [Abzianidze va

boshq., 2017:242] kelib chiqqan. Oxirgi ishlar vektor tasvirlari tomonidan olingan semantikani baholash uchun o'tkazgich sifatida teglar to'plamining qulayligini alohida ko'rsatadi yoki biz mazkur maqolada keltirganimiz kabi uni yordamchi teglash vazifasida qo'llaydi [Bjerva va boshq., 2016:3531].

Asosiy vazifa va yordamchi vazifa o'rtasida taqsimlangan $\vec{h}_X$ qatlami ikkita kichik maydonga bo'linadi: taqsimlangan $\vec{h}_{X_S}$ maydon hamda xususiy $\vec{h}_{X_P}$ maydon. Taqsimlangan maydonlar o'rtasidagi o'zaro ta'sir o'rganilgan og'irliklar to'plamiga qo'llaniladigan sigmasimon yo'l birligi orqali modulyatsiya qilinadi, (1)

(2) tenglamalar $\vec{h}_{X_{S(main)}}$ va $\vec{h}_{X_{S(aux)}}$ da ko'rsatilganidek, asosiy va yordamchi vazifalarning umumiy qatlamlari, $W_{a \rightarrow m}$ va $W_{m \rightarrow a}$ o'rganilgan og'irliklar va σ sigmasimon funksiya hisoblanadi.

$$\vec{h}_{X_{S(main)}} = \vec{h}_{X_{S(main)}} \sigma(\vec{h}_{X_{S(aux)}} W_{a \rightarrow m}) \quad (1)$$

$$\vec{h}_{X_{S(aux)}} = \vec{h}_{X_{S(aux)}} \sigma(\vec{h}_{X_{S(main)}} W_{m \rightarrow a}) \quad (2)$$

Liu va boshqalarning “Taqsimlangan qatlam arxitekturası” (2016)dan farqli o'laroq, bu qatlamda har bir vazifa bitta odatiy taqsimlangan qatlamga emas, balki o'zining umumiy maydoniga ega. Bu Liu va boshqalarda bo'lgani kabi “barcha vazifalar uchun taqsimlangan ma'lumotni olish yo'lida bitta taqsimlangan qatlam”ga ega bo'lish o'rniga, har bir vazifaga boshqasidan nimani o'qitishni tanlash imkonini bergan holda har bir yo'nalishda turli parametrlarni taqsimlash imkonini beradi (ya'ni asosiydan yordamchi vazifaga va yordamchidan asosiy vazifaga).

1-rasmda ko'rsatilgan uchta neyron ko'p vazifali tizimlarini ko'rib chiqamiz. Ular tarmoq parametrlarining vazifalar o'rtasida taqsimlanishiga ko'ra farqlanadi:

- **To'liq taqsimlangan tarmoq.** Barcha yashirin qatlamlar butunlay vazifalar orasida taqsimlanadi, har bir vazifa alohida chiqish qatlamiga ega. Kirish

$$\vec{h}_S = \sigma(\vec{x}W) \quad (3)$$

vektori $\vec{x}$ ni taqsimlangan yashirin qatlam $\vec{h}_S$ ga aylantirish (3) tenglama bilan tavsiflanadi:

- **Qisman taqsimlangan tarmoq.** Yashirin qatlamlar to'plami vazifalar orasida taqsimlanadi; har bir vazifada kamida bitta xususiy yashirin qatlam va alohida chiqish qatlami mavjud. Taqsimlangan yashirin qatlam $\vec{h}_S$ ning $\vec{h}_{P(main)}$ va $\vec{h}_{P(aux)}$ bilan belgilangan xususiy yashirin qatlamlarga o'zgarishi (4) va (5) tenglamalarda tasvirlangan.

$$\vec{h}_{P_{(main)}} = \sigma(\vec{h}_S W_{(main)}) \quad (4)$$

$$\vec{h}_{P_{(aux)}} = \sigma(\vec{h}_S W_{(aux)}) \quad (5)$$

• **Nimani taqsimlashni o'qitish.** Har bir vazifada yashirin qatlamlarning maxsus to'plami mavjud. Taqsimlash uchun foydalaniladigan yashirin qatlam taqsimlangan va xususiy maydonlarga bo'linadi. Yo'l birligi (1) va (2) tenglamalarda ko'rsatilganidek, taqsimlangan maydonlar o'rtasida ma'lumot uzatishni modulyatsiya qiladi.

Barcha vazifalar bo'yicha natijalar quyidagi jadvalda keltirilgan:

Model	Stemford tabiiy til xulosasi	SICK-E	Universal nutq qismlarini teglash	Universal bog'liqlik
Bir vazifali modellar	87.01	81.30	92.12	80.24/84.87
To'liq taqsimlangan tarmoq	84.96	56.69	92.95	81.03/85.54
Qisman taqsimlangan tarmoq	87.08	77.92	92.34	80.92/85.81
Nimani taqsimlashni o'qitish	87.51	84.57	95.54	81.39/86.00

Bir vazifali modellar, to'liq taqsimlangan tarmoqlar, qisman taqsimlangan tarmoqlar va nimani taqsimlashni o'qitish bo'yicha natijalar. LAS/UAS F1 baholash bo'yicha, universal bog'liqlikdan tashqari barcha ballar aniq hisoblangan holda taqdim etilgan.

Bjerva va boshqa olimlar xulosalariga ko'ra, to'liq taqsimlangan tarmoq tizimlari so'z turkumlarini universal teglashni yaxshilaydi. So'z turkumlarini teglash, ketma-ket teglash vazifasi semantik teglash bilan eng ko'p bog'liq bo'lgan vazifa sifatida qaralishi mumkin, shuning uchun salbiy uzatish minimallashtirish hamda parametrlarni to'liq taqsimlash foydaga ishlaydi. Ajablanarlisi shundaki, to'liq taqsimlovchi tarmoq tizimlari universal bog'liqlikning samarali ishlashini ta'minlaydi. Darhaqiqat, universal bog'liqlik uchun barcha ko'p vazifali o'qitish modellari chegaralarni oshirish orqali bir vazifali modelidan ustun turadi. Biroq tabiiy til xulosasi vazifalarini amalga oshirishda aniq pasayish yuzaga keladi. Qisman taqsimlangan tarmoq tizimlari aralash natijalarni ko'rsatadi hamda so'z turkumlarini universal teglash va universal bog'liqlik uchun to'liq taqsimlangan tarmoqqa nisbatan aniq ustunlikni ko'rsatmaydi. Bu shuni ko'rsatadiki, to'liq taqsimlangandan keyin maxsus vazifali qatlamlarini qo'shish har doim ham vazifalarni yetarli darajada ixtisoslashtirishga imkon bermaydi. Biroq tabiiy til xulosasi vazifalari uchun qisman taqsimlangan tarmoq, ayniqsa, to'liq taqsimlangan tarmoq modeli yetarli darajada o'qita olmaydigan kichik o'lchamli SICK-E ma'lumotlar to'plami uchun, to'liq taqsimlangan tarmoqqa nisbatan afzal.

Tajribalar natijalari shuni ko'rsatadiki, semantik teglashni yordamchi vazifa sifatida qo'llash barcha MTL tizimlarida UPOS teglash va UD DEP ish faoliyatini yaxshilashga olib keladi. Bundan tashqari, semantik teglashni yordamchi vazifa sifatida qo'llash barcha ko'p vazifali o'qitish tizimlarida so'z turkumlarini teglash va universal bog'liqlik ish faoliyatini yaxshilashga olib keladi. Ibora semantikasini tushunishni talab qiluvchi Stenford tabiiy til xulosasi vazifalari uchun biz “nimani taqsimlashni o'qitish” deb nomlaydigan selektiv taqsimlash tizimlari aniq afzalliklarga ega ekanligi ma'lum bo'ldi.

Foydalanilgan adabiyotlar:

1. Abzianidze, Johannes Bjerva, Kilian Evang, Hessel Haagsma, Rik van Noord, Pierre Ludmann, Duc-Duy Nguyen and Johan Bos. 2017. The parallel meaning bank: Towards a multilingual corpus of translations annotated with compositional meaning representations. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2.
2. Anders Sogaard and Yoav Goldberg. 2016. Deep multi-task learning with low level tasks supervised at lower layers. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), volume 2.
3. Johannes Bjerva, Barbara Plank, and Johan Bos. 2016. Semantic tagging with deep residual networks. In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers.
4. Захаров В.П., Богданова С.Ю. Корпусная лингвистика: учебник для студентов гуманитарных вузов. – Иркутск: ИГЛУ, 2011. – 161 с.
5. Кустова Г.И., Падучева Е.В. Словарь как лексическая база данных // ВЯ, 1994. – № 4.
6. Rich Caruana. 1998. Multitask learning. In Learning to learn, Springer.

ISSUES OF IMPLEMENTING MACHINE INTERPRETING INTO THE SPHERE OF MEDICINE

Xabibullayeva Dilnoza Qobiljon qizi

dilnozaxabibullayeva982698@gmail.com

ToshDO‘TAU tayanch doktoranti

Annotatsiya. Ushbu maqolada tibbiyot sohasiga hozirda rivojlanib borayotgan mashina tarjimasini tatbiq etishda yuzaga keluvchi muammolar ko‘rib chiqiladi va ularga muvofiq yechimlar ilgari suriladi. Shuningdek, mashina tarjimasining yutuqli jihatlari va kamchiliklari tadqiq etilib, inson tomonidan amalga oshirilgan tarjima hamda mashina tarjimasini natijalari qiyosiy o‘rganilib, tegishli xulosalar chiqariladi. Mashina tarjimasining tibbiyot sohasida sinxron tarjimani amalga oshirishdagi o‘rni borasida tadqiq etiladi.

Abstract. This article reflects the problems in the implementation of currently developed method - machine interpreting to the sphere of medicine and provides effective solutions to the issue. Furthermore, benefits and disadvantages of machine interpreting are investigated and comparative study about the productivity of machine interpreting and interpreting process by specialists is carried out. Discussing the role of machine interpreting and translation in the training of medical interpreting experts, the article reflects precise conclusion in the field.

Аннотация. В данной статье отражены проблемы внедрения разрабатываемого в настоящее время метода машинного перевода в сферу медицины и предложены эффективные решения данного вопроса. Кроме того, исследуются преимущества и недостатки машинного перевода и проводится сравнительное исследование производительности машинного перевода и процесса устного перевода специалистами. Обсуждая роль машинного устного и письменного перевода в подготовке специалистов по медицинскому устному переводу, статья отражает четкие выводы в этой области.

Kalit so‘zlar: *Mashina tarjimasi, Lexifone, ovozni aniqlash.*

Today, computer science is being developed in an accelerating pace. In many spheres of life, people began to use this advantage instead of using human work power. One of these disciplines is called medical interpreting domain. When there is a language barrier between medical specialists and the patient, the healthcare setting tries to solve this problem by hiring professional medical interpreters in order to render accurate information about patients and treat sick people in a quality manner. The basic issue about the case is that healthcare setting cannot always find professional experts in this field. Because medical interpreting is such a complicated sphere that not all interpreters involve in this discipline. Medical interpreting requires deep medical knowledge as well as foreign language acquisition. So the burden of the career is more than other branches of interpreting. So, in rare cases

people choose to become medical interpreting expert. The lack of specialists in the field has become a trigger to investigate computer science to solve the problem in an effective way.

Machine interpreting (Automated interpreting) is a new term which can prove a fairly new technology. Machine interpreting can be used to convert spoken words into different languages. In the domain of computer science, *Speech – to –Speech translation* specifically refers to the translation of spoken content from one language into spoken content in a different language. Another explanation of the term is that “Machine interpreting is the transmission of a spoken message in one language into a spoken message in a different language using artificial intelligence (AI), without the input of a human interpreter”. It is worthy to mention that the first attempt to build a speech – to- speech translation system was carried out in 2016 by a group of French researchers. Using only a small corpus, the team tried to obtain effective results from their MI engine. Their conclusion was that in order to implement smooth interpreting process they should train future MI engines on more varied datasets. Actually, machine interpreting combines two forms of technology: **voice recognition software** and **machine translation**. To identify source language, translation technology uses Automatic Speech Recognition (ASR) and to speak the message in the target language, they use a synthetic voice. This process consists of three basic stages:

- *Voice recognition software picks up a speaker's voice and renders it into text;*
- *The text is run through a machine translation program;*
- *The translated text is automated into speech in the target language.*

Since medical interpreting experts are not always available for interpreting services in healthcare settings, medical specialists tend to use machine translation technology for interpreting medical content. They consider this method as an effective way to supplement their communication. Initially, machine translation was developed to facilitate simple text-to-text translations. However in recent years it produces more complex translations by incorporating more sophisticated algorithmic approaches. As a result, machine interpreting has begun to be considered as one of the modes of interpreting sphere.

Nowadays, there are a number of programs which can render medical content from one language to another. In healthcare settings, specialists may use a number of translation programs including: *Google translate, Microsoft Bing Translator, DeepL Translator, SYSTRAN Translate, Amazon Translate and etc.* There is also machine interpreting service such as *Lexifone* which has been established by Israeli startup company [Sanita, 2022: 76]. They can be used as an interpreting aid during the process but they cannot provide complete interpreting output due to number of reasons.

Despite having several beneficial sides (e.g. Speech - to – Speech translation can also function live, or in real – time or streaming mode, to use more technical

terminology) machine interpreting cannot always provide promising results in the sphere of medicine due to some reasonable issues. The core problem with machine interpreting is its basis in machine and automated translation. Mostly machine translation can provide the gist of the statement but it is not acceptable in situations which require high accuracy. Sometimes, voice recognition software cannot provide perfect product. In terms of medical cases, accurate interpreting output is so crucial that it has a high influence on the patient's health. As well as having numerous beneficial sides, machine interpreting cannot provide confidential results all the time. In the sphere of simultaneous interpreting there are a lot of interpreting strategies and techniques which can be used effectively only by medical interpreting experts. Considering the particular case, experts choose effective interpreting techniques and solve problems related to the language. The following issues should be considered in machine interpreting:

- Adapting the mouth movements of the speaker to the translation (or vice – versa) and create almost perfect synchronicity.
- Considering context when interpreting tricky language units like homonyms. Medical interpreting involves a thorough grasp of medical terms in both source and target languages.
- There is no possibility to edit the result: the users will consume the raw translation.
- The system must construct a logical and articulate flow of text in the target language, starting from partial segments of speech.
- One of the most effective techniques of interpreting - Anticipation cannot be used in machine interpreting.
- In automated interpreting, there is a risk of using medically offensive terms. (Checking any term with knowledge of the clinical field is essential) [Amrita, 2021; 91].

Google translate in medical communication. Google translate uses statistical matching to translate rather than depending on a dictionary or grammar rules approach. Unfortunately, the method “just Google it” cannot be used in life saving communications between doctors and patients. According to recent statistics, Google Translate has only 57.7 % accuracy when used for medical phrase translation. It means that medical interpreting process cannot be carried out based on translation tools. This also proves that implementing machine interpreting in the sphere of medicine is a very critical issue [Douglas, 2022: 101]. In the investigation, the following errors have been identified and analyzed:

Example 1:

*The **bundle of His** is a group of fibers that carry electrical impulses through the center of the heart. If these signals are blocked, you will have problems with your heartbeat.*

Google translate: *Uning to'plami yurakning markazi orqali elektr impulslarini o'tkazadigan tolalar guruhidir. Agar bu signallar bloklangan bo'lsa, siz yurak urishi bilan bog'liq muammolarga duch kelasiz.*

Correct version : *Gis to'plami yurakning markazi orqali elektr impulslarini o'tkazadigan tolalar guruhidir. Agar bu signallar bloklangan bo'lsa, siz yurak urishi bilan bog'liq muammolarga duch kelasiz.*

When using google translate tool, the term “His bundle” was translated as “Uning to'plami”. But when we checked the medical context in the Uzbek language, we discovered that this kind of term is not used in the Uzbek medical language.

Solution: Discussing with the professional medical expert. When we discussed with the specialist firstly he confirmed that they use this term in the sphere and defined the above mentioned term: His bundle or the bundle of His is a group of fibers that carry electrical impulses through the center of the heart (Gis tutami - yurakning markazi orqali elektr impulslarini o'tkazadigan tolalar guruhidir).

Example 2

*In addition to active ingredients, most **tablets** contain additives that hold the pill together and improve the taste, texture, or appearance.*

Google translate: *Faol moddalarga qo'shimcha ravishda, **planshetlarning** ko'pchiligi tabletkani bir-biriga bog'lab turadigan va ta'mi, tuzilishi yoki ko'rinishini yaxshilaydigan qo'shimchalarni o'z ichiga oladi.*

Correct version: *Faol moddalarga qo'shimcha ravishda, **tabletkalarning** ko'pchiligi ularni bir-biriga bog'lab turadigan hamda ta'mi, tuzilishi yoki ko'rinishini yaxshilaydigan qo'shimchalarni o'z ichiga oladi.*

Rendering homonyms in machine interpreting. One of the most problematic issues of machine interpreting is rendering homonyms in the process. Medical interpreting experts have a chance to identify the term by considering the context or asking explanation from the speaker while the process is going on. But in machine interpreting, if the term is not mentioned in the database there is a high risk to render it in a wrong way. The term “**BS**” has three meanings:

- *Blood sugar, Russian. уровень сахара в крови. Uzbek. qondagi shakar;*
- *Bowel sounds, Russian. звуки кишечника, Uzbek. ichak tovushlari;*
- *Breath sounds, Russian. звуки дыхания, Uzbek. nafas tovushlari.*

The solution for this problem can be including all meanings of homonyms when forming database for machine interpreting. They should be mentioned in both Uzbek and English languages.

In November 2016, Google informed that it changed its way of working manner. It began to use neural machine translation which is a deep learning tool that regularly compares whole sentences at a time from many linguistic sources to translate language units. Despite this advancement, Patil and Davies found serious errors in translations of common medical terms [Flores, 2005: 2010]. This means that Google Translate does not always reflect accurate rendering of medical content.



Cultural awareness. In the healthcare and medical fields, professional interpreters are required to be aware of cultural nuances, which might range from standard medical procedures to idiomatic expressions that may be misunderstood. In machine interpreting, this may be a problematic issue. Machine translation technology may not be aware of medically offensive terminology. For example, “harelip” is a term used to describe an individual with a congenital split in their lip or palate. “Cleft lip” and “cleft palate” are accepted in modern usage instead of “Harelip”. When using Google Translate, it has rendered the terms as “harelip”. Through ages, harelip was associated with negative connotations. Young people with *cleft lip* have several functional and aesthetic problems including trouble clothing their mouth, feeding difficulties, hearing difficulties and speech difficulties. Using the offensive term for them due to the misinterpreting may cause dangerous consequences such as anxiety, depression and even suicide [Hsieh, 2006: 55]. In a word, machine interpreting is to be highly developed to the level of differentiating special linguistic units before its introduction to the medical field.

Using medical interpreting expert during the medical interpreting process enhances the possibility that the interpreter would interpret literally and objectively with high-quality language skills. Effective medical communication between these two languages requires investigation in the translation techniques which are expected to use to make medical interpreting process safe and high quality. In order to improve the quality of machine interpreting, the following changes should be implemented: To have a large database of Uzbek and English medical terminology; to reflect specific features of medical language; to be aware of medical terms' structure; to be aware of confusing medical terms; to analyze medical terms in these two languages. To conclude, machine interpreting should be developed as a perfect interpreting tool since any single minor mistake may cost patient's valuable life and this mistake can never be forgiven.

References:

1. Amrita Naomi DSouza. Improving communication on medical ward rounds with patients who speak limited English with implementation of medical communication charts. BMJ Open Quality, 2021 – P 87
2. Barik. H. Simultaneous interpretation: temporal and quantitative data. Chapel Hill, NC. University of North Carolina. - P 420
3. Douglas Ondara Orang'i . Translating taboo words in health care texts from English into Swahili in Kenya., Mycotaxon, 2022. – P 108
4. Flores, G. The impact of medical interpreter services on the quality of Healthcare: A systematic review, Medical care research and review. 2005. - P 240
5. Hsieh E. Conflicts in how interpreters manage their roles in provider-patient interactions. Soc Sci Med. 2006. – P 156



6. Hsieh E. I am not a robot!” Interpreters’ views of their roles in health care settings. Qual Health Res. 2008. – P 46

7. Sanita Kaur Sandhu. Lost in translation: a caution against using automated and commercial translation services for medical communications. et al., Global health, 2022. – P 108

INTERTEKSTUAL TEZAVRUSLARNING MENTAL LUG‘AT SIFATIDAGI AHAMIYATI

O‘rinova Zarifa Jumanazar qizi

zarifaorinova@gmail.com

ToshDO‘TAU tayanch doktoranti

Annotatsiya. Ushbu maqolada hozirgi kunda elektron lug‘atlarning o‘rni, o‘zbek tili leksikografiyasida haligacha, yetarli darajada elektron lug‘atlar yaratilmaganligi haqida so‘z boradi. Jahon tilshunosligida yaratilgan ilk tezavrus lug‘atlar va ularning o‘ziga xos xususiyatlari, bu turdagi lug‘atlar uchun ma’lumotlar bazasini shakllantirish talablariga ham to‘xtalib o‘tilgan. O‘zbek intertekstual birliklar tezavrusini yaratish asoslari, intertekst haqidagi ma’lumotlarni berish tartibi xususida aniq tavsiyalar ishlab chiqilgan. Nazariy ma’lumotlar misollar asosida dalillangan.

Abstract. This article talks about the role of electronic dictionaries today, the fact that there are still not enough electronic dictionaries in the lexicography of the Uzbek language. The first thesaurus dictionaries created in world linguistics and their characteristics, as well as the requirements for the formation of a database for this type of dictionaries, are also discussed. Specific recommendations have been developed regarding the basics of creating the thesaurus of Uzbek intertextual units, the procedure for providing information about intertext. Theoretical information is proved on the basis of examples.

Аннотация. В данной статье говорится о роли электронных словарей сегодня, о том, что в лексикографии узбекского языка по-прежнему недостаточно электронных словарей. Также обсуждаются первые созданные в мировой лингвистике словари-тезаурусы и их характеристики, а также требования к формированию базы данных для словарей этого типа. Разработаны конкретные рекомендации относительно основ создания тезауруса узбекских интертекстуальных единиц, порядка предоставления информации об интертексте. Теоретическая информация подтверждается на основе примеров.

Kalit so‘zlar: *Leksikografiya, kompyuter lingvistikasi, tezavrus, intertekstual birlik, pretsedent birlik.*

Ma’lumki, XXI asrni kompyuter texnologiyalari asri sifatida qayd etilmoqda. Hozirgi kunda hech qaysi soha yo‘qki, elektrotexnika, kompyuter texnologiyalari kirib bormagan. Zamonaviy tilshunoslikni-da, biz kompyuter texnologiyalarisiz tasavvur qila olmaymiz. Ayniqsa, jamiyatimizda zamonaviy elektron lug‘atlarga ehtiyojning borgan sari ortib borishi, o‘zbek leksikografiyasi rivojiga katta ta’sir o‘tkazdi. Tilshunoslar tomonidan har tomonlama qulayliklarga ega tezavrus lug‘at, wordnet lug‘atlar yaratish zarurati vujudga keldi. Bu kabi mashina tarjimasi tayanch

bazasi vazifasini o'tovchi lug'atlar esa g'arb tilshunosligida allaqachon yaratilgan. Xususan, tezaurus lug'atlarga to'xtalsak, tezaurus umumiy ma'noda maxsus terminologiya, ya'ni lug'at, ma'lumotlar yig'indisi, ma'lum sohadagi bilimlar yoki faoliyat sohasidagi tushunchalar, ta'riflar va terminlarni to'liq o'z ichiga olish xususiyatiga ega bo'lgan elektron lug'atdir. Birinchi zamonaviy ingliz tezaurusi Piter Mark Roger tomonidan 1805-yilda yaratilgan bo'lib, bu lug'at 1852-yilda Angliyada nashr etilgan. Ushbu tezaurus lug'at taxminan 15 mingta tushunchani o'z ichiga oladi va hozirgacha foydalanib kelinadi. Rus tilshunosligida ham o'tgan asrning oxirida mashxur tezaurus lug'at “PyTez” yaratildi. Biroq o'zbek tilida hanuz bu kabi tezaurus lug'at yaratilmagan. Ushbu maqolada biz interkestual tezaurus yaratish va uning ahamiyatini ko'rib chiqamiz.

Hozirgi kunda aksariyat hollarda zamonaviy o'zbek tilida yaratilgan badiiy matnlarni to'laqonli tushunish uchun faqat o'zbek tilini bilish, unda erkin gapira olishning o'zi yetarli emas. Buning sabablaridan biri esa badiiy kontekstda intertekstual birliklarning yozuvchi va shoirlar tomonidan keng qo'llanilayotganligidir. Lisoniy shaxs (til sohibi) badiiy matnni aniq tushuna olishi va tahlil qila olishi uchun esa, albatta, intertekstual birlikni tanishi, u haqida yetarli bilimga ega bo'lishi lozim.

Tilshunoslar O.V.Fokina va E.Shmelova ham zamonaviy rus matnlarini to'liq tushunish uchun “Intertekstual kompetensiya”dan xabardor bo'lishi lozim deya ta'kidlaydi. Sitata, iqtibos, allyuziv nomlar, reminatsiya, pretsedent nomlar v.b.ni birgalikda intertekstual birlik deb atash mumkin. [Фокина, Шмелева, 2012:49]. Rus tilshunosligida chuqur izlanishlar natijasida 2007-yilda birinchi “Zamonaviy rus tili idiomalari tezaurus lug'ati”¹⁵ yaratildi. Bu lug'at tarkibiga rus intertekstual birliklari ham kiritilgan.

Ko'rinib turibdiki, badiiy matnni to'liq tushunish, tarjima jarayonlarini yanada qulaylashtirish maqsadida o'zbek tilidagi matnlarda ko'p qo'llanadigan intertekstual birliklarning ishlatish o'rnini batafsil tavsiflovchi va tushuntiruvchi “O'zbek intertekstual birliklar tezaurusi”ni yaratish bugungi kun o'zbek leksikografiyasi uchun juda zarur. Lug'at uchun o'zbek milliy adabiyoti, ommaviy axborot vositalari va internet matnlarini tahlil qilish orqali ma'lumotlar bazasi to'planishi ko'zda tutiladi. Tezaurus lug'atlarda ma'lumotlar bazasi juda katta bo'lganligi bois kitob shaklida emas, multimedia, ya'ni elektron shaklida yaratish tavsiya etiladi. Tezaurus, ayniqsa, elektron formatdagi tezaurus har bir sohaga oid so'zlarni to'liq ochib beradigan lug'atlardan biri sanaladi. Oddiy lug'atlardan farqli o'laroq, har bir so'z mohiyatining to'liq yoritib berilishini ta'minlaydi. Bu lug'atning elektron formatdagi ko'rinishi insonning vaqtini tejash, bir so'z haqida to'liq ma'lumot olish imkonini beradi. [Xolmanova, 2019] Demak, foydalanuvchilar

¹⁵ Словарь-тезаурус современной русской идиоматики 2007]; [Ашукин, Ашукина 1987]; [Берков, Мокиенко, Шулежкова 2005]; [Душенко 2006]]

kam vaqt ichida ko‘p ma’lumotga ega bo‘lishi, qulaylik his qilishi uchun tezaurus lug‘atlarning elektron shaklda bo‘lishi afzalroq.

Intertekstual birliklar tezaurus lug‘atining o‘ziga xos jihatlaridan biri shuki, bunda intertekstual birlikning asl manbasi, uning paydo bo‘lish vaqti va mualliflik ko‘rsatkichi, agar birlik chet el adabiyoti matnlariga, chet el filmlariga, aforizmlariga ishora qiluvchi tarjima qilingan matn yoki iqtibos ko‘rinishida bo‘lsa, asl tilida ham ko‘rsatilishi kerak. Ba’zan esa rasmlar, plakatlar, intertekstual birlik film va multfilm qahramoni tomonidan qo‘llangan bo‘lsa, aynan shu obrazning fotosurati beriladi. Lug‘at elektron shaklda bo‘lganligi bois, turli xil ma’lumotlar matn, grafika, animatsiya, audio, foto va video va boshqalar ko‘rinishida taqdim etilish imkonini mavjud.

O‘zbek intertekstual tezaurusi bazasini yaratishda boshqa tillarda yaratilgan ushbu tipdagi lug‘atlarni umumlashtirib, ulardan foydalangan holda ma’lumot to‘plash kerak. Quyida biz xalqimiz orasida juda mashhur bo‘lgan “siz o‘shami?” pretsedent birligining tezurus lug‘atda qanday berilishini ko‘rib chiqamiz. Baza taxminan shunday tuzilishga ega bo‘ladi:

Intertekstual birlik	“Siz o‘shami” (1926-yildan beri xalq orasida mashhur. Eng mashhur bo‘lgan davr: 2001-2017-yillar oralig‘i.
Intertekstual birlik turi	Pretsedent birlik
Asl manba	A.Qodiriy. “O‘tkan kunlar” romani (1920-yilda yozilgan, 1926-yilda kitob holida nashr etilgan. Romanning “Kutilmagan baxt” bo‘limidan
Asl matn	Shundan keyin bir necha qadam bosib Otabekning pinjiga yaqin keldi va esankiragan, hayajonlang‘an bir tovush bilan so‘radi: — Siz o‘shami? — Men o‘sha, — dedi bek. Ikkisi ham bir-birisiga beixtiyor termulishib qoldilar. 
Izoh	Kutilmagan vaqtda, kutmagan insonni ko‘rish, o‘zini xursand va baxtli his qilish
Intertekstual birlik sifatida	Xurshid Davron “Siz o‘shami?” she‘ri. Yulduzim deb izlaganim, Armonlarda bo‘zlaganim,

*Kecha-kunduz so'zlaganim –
Siz o'shamu, siz o'shamu? (Shuhrat Qayumov ijrosida
qo'shiq bo'lgan va mashhurlikka erishgan)*



Shu tarzda o'zbek tilida mavjud bo'lgan intertekstlarni har biriga to'xtalib o'tiladi. Qo'shimcha tarzda “siz o'shami” pretsedent birligini bazasini shakllantirishda “O'tgan kunlar” romani asosida ishlangan badiiy filmdan aynan shu intertekst qo'llangan parcha va film epizodining videosi va qo'shiq ijrosidan namuna, ya'ni qo'shiq matni va ovozi ham joylash mumkin.

Intertekstual birliklar tezaurusini tuzishda ma'lumotlarni, avvalo, tematik, ya'ni interteks turi nuqtayi nazaridan berib, an'anaviy alifbo tartibida berilishi tavsiya etilmaydi. Chunki ba'zi intertekstual birliklar so'z ko'rinishida, ba'zilar birikma, ba'zilar gap, ba'zilar esa matn ko'rinishida bo'lishi mumkin.

Tezurus tarkibidagi so'zlar alfavit bo'yicha emas, balki mavzu va mavzular majmuyi sifatida shakllangan lingvistik lug'at hisoblanadi. Tezaurus – so'zlar va so'z birikmalari o'rtasidagi turli bog'lanishlar (munosabatlar) haqidagi axborotlarni saqlashga xizmat qiluvchi ma'lumotlar bazasidir. [Xolmanova, 2019:102] Ko'rinish turibdiki, tezaurus lug'atlar boshqa turdagi lug'atlardan farqli tuzilishga ega bo'ladi va metainformatsiyani (axborot haqidagi ma'lumotlar)ni qamrab oladi.

Intertekstual lug'at tuzishda olimlar turlicha takliflar bildirishgan. Bu takliflarning barchasining maqsadi, intertekstual lug'atlarni foydalanuvchilar uchun qulay va foydali bo'lishga qaratilgan.

Foydalanuvchiga kerakli ma'lumotlarni tezda o'qishga imkon beradigan intertekstual birlik yozuvining turli zonalariga giperhavolalar qo'yilishi kerak. Ma'no zonasidan tematik lug'atga o'tish va boshqa intertekstual birliklar shunga o'xshash ma'noni ifodalashi haqida ma'lumot olish mumkin bo'lishi, shuningdek, manba zonasidan manbalar ro'yxatiga kirish va kelib chiqishi bir xil bo'lgan boshqa intertekstlarni ko'rish imkoni mavjud bo'lmog'i lozim. [Вознесенская, Шмелева, 2019:8]. Ushbu fikrlardan anglashinadiki, o'zbek intertekstual birliklar tezaurusini yaratish uchun tilshunoslar va dasturchilar birgalikda maslahatlashgan holda ish olib borishi lozim. O'zbek tili intertekstual birliklar tezaurusini to'liq yaratish bitta tadqiqotchi yoki ilmiy ish doirasida hal qilish qiyin bo'lgan vazifadir. Sababi intertekstual tezaurus bazasini yaratish jarayonida matnlar badiiy asarlardan, kinofilmlardan, xalq og'zaki ijodi namunalaridan, jonli nutqdan, teleko'rsatuvlar, reklamalar va ilmiy-badiiy maqolalardan terib olinishi talab etiladi.



Xullas, intertekstual birliklar tezaurus lug‘ati tarjimonlarning sifatli faoliyat olib borishi uchun, lisoniy shaxs (til egasi)lari hamda ushbu tilni o‘rganayotgan chet elliklar uchun juda zarurdir.

Foydalanilgan adabiyotlar:

1. Вознесенская М. М., Шмелева Е. Я. О проекте словаря «интертекстуальный тезаурус современного русского языка»: книжный vs. Мультимедийный. Компьютерная лингвистика и интеллектуальные технологии: по материалам международной конференции «Диалог 2019». Москва: 2019. С. 4-10.
2. Fokina O. V., Shmeleva E. Ya. (2012) Intertextual competence and text understanding [Intertekstual'naja kompetetsiya i ponimanie teksta] Linguistic training of students at school and out of school [Lingvisticheskaja rabota so shkol'nikami na urokah raznyh predmetov i vo vneurochnom obrazovatel'nom prostranstve], MIOO, Moscow, p. 49–59.
3. Xolmanova Z. Kompyuter lingvistikasi. – Toshkent, 2019. B. 102.
4. <https://tafakkur.net/otkan-kunlar/kutilmagan-baxt.uzb>

O'ZBEK TILI TEZAVURUSI UCHUN SIFAT LEKSEMALARNING SEMANTIK MUNOSABATLARI

Po'latova Gulhayo G'ofurjon qizi

gpulatova2409@gmail.com

Namangan davlat universiteti mustaqil tadqiqotchisi

Annotatsiya. Ushbu maqolada sifat leksemalarining tildagi o'rni, semantik maydoni haqidagi fikr bildirilgan. Semantik maydon kompyuter leksikografiyasi taraqqiyotida o'ziga xos o'ringa ega. Terminlar tezavrusida semantik maydon nazariyasiga asoslanish sohaga oid tushunchalarni tez va oson o'zlashtirish imkonini berishi yoritilgan. “Achchiq” sifat leksemasining semantik maydoni va tezavrusi haqida mulohazalar keltirilgan.

Abstract. This article expresses an opinion about the place of quality lexemes in the language, the semantic field. The semantic field has a special place in the development of computer lexicography. In the thesaurus of terms, it is explained that basing on the theory of the semantic field allows to quickly and easily master the concepts of the field. Comments are made on the semantic field and thesaurus of the adjective lexeme “achchiq”.

Аннотация. В данной статье высказывается мнение о месте качественных лексем в языке, семантическом поле. Семантическое поле занимает особое место в развитии компьютерной лексикографии. В тезаурусе терминов поясняется, что опирание на теорию семантического поля позволяет быстро и легко освоить понятия поля. Даются комментарии к семантическому полю и тезаурусу лексики «achchiq».

Kalit so'zlar: *sifat leksemalari, sifatlarning semantik-uslubiy xususiyatlari, semantika, semantik maydon, tezavrus, “achchiq” leksemasi tezavrusi.*

Kirish

O'tgan davr mobaynida yurtimizda milliy ma'naviyatimiz, o'zligimizni ifodalovchi tilimizga, uni ilmiy tadqiq qiluvchi tilshunosligimizga katta e'tibor qaratildi va mazkur jarayon izchil davom etmoqda. O'zbek tilshunosligida fan taraqqiyoti sezilarli darajada yuksaldi. Bunda jahon tilshunosligi yutuqlaridan o'rinli foydalanish, ularni amalda qo'llash o'z ijobiy natijalarini berdi. Natijada erishilayotgan yutuqlarimiz nafaqat mamlakatimiz tilshunosligida, balki dunyo tilshunosligida ham ma'lum bo'lib bormoqda. Erishilayotgan muvaffaqiyatlar bilan bir qatorda qilinishi lozim bo'lgan tadqiqotlar ham mavjud bo'lib, ularning sirasiga o'zbek tilida tezavrus lug'atlarni ishlab chiqish, o'zbek tilini o'rganuvchi xorijliklar uchun elektron lug'atlarni yaratish bugungi kunning dolzarb vazifalaridandir.

Jamiyat rivojlanishining qizg'in pallasiga kirgani sari tilimizda o'ziga xos taraqqiyot yuz berishi tabiiydir. Binobarin, tildagi so'zlarning shakl va ma'no jihatdan taraqqiy etishi ulardagi ushbu munosabatlarni kengroq tadqiq etishni taqozo

etadi. Soʻzlarning oʻzaro shakl va maʼno munosabatlari turlicha. Ayrim soʻzlar shaklan har xil, lekin maʼno jihatidan bir-biriga yaqin boʻlsa, boshqalari shaklan bir xil, ammo maʼnolari turli xil, baʼzilari esa bir-biriga qarama-qarshi soʻzlar ham boʻlishi mumkin. Bunday hodisa tilshunoslikda semantik munosabat deyiladi. Sintagmatikada lisoniy birliklar orasidagi munosabatlar aniqligi va izchilligi bilan ajralib turadi. Qiyoslang: Ekilgan koʻchatlar viqor bilan koʻkka boʻy choʻza boshladi. Koʻrinib turibdiki, nutq faoliyatida muayyan “yotiq chiziq”qa joylashib, muayyan tartibda, mantiqiy ketma-ketlikda kelgan birliklar orasidagi munosabatlar sintagmatik munosabatni hosil qiladi. Sintagmatik munosabatlar doimo yaqqol taʼsir qilishi insonda lisoniy sezgi uygʻotishi bilan ham oʻziga xosdir [Israilova, 2023].

Asosiy qism

Oʻzbek tilshunosligida kompyuter lingvistikasiga doir qator izlanishlar olib borilmoqda, jumladan, mashina tarjimasini, korpus texnologiyasi, kompyuter leksikografiyasi hamda dasturiy ilovalarning lingvistik asoslarini ishlab chiqish, leksikografik matnlarga ishlov berish va lingvostatistik tahlil etish borasida muayyan darajada natijalarga erishib kelinmoqda. N.Abduraxmonova oʻzining tadqiqotlarida kompyuter leksikografiyasi, ontologiya tezaurus, WordNet, korpusga asoslangan taʼlim texnologiyasiga bagʻishlangan bir nechta ilmiy ishlarni eʼlon qilgan.

Tezaurus tushunchasi bugungi kunda keng maʼno ottenkasiga ega tushuncha. Keng isteʼmolda konseptologik tamoyillarga asoslangan ideografik lugʻat sifatida tushuniladi. Keyingi vaqtlarda yaratilayotgan tezaurus va ontologiyalar esa funksional jihatdan semantik lugʻatlar tarkibiga kiradi. Shuningdek, bunday lugʻatlar ideografik lugʻatlar tipiga ham mansub, chunki bu lugʻatlarni tuzishda semantikaga alohida eʼtibor beriladi va mavzu maydoni muhim ahamiyatga ega hisoblanadi. Tezauruslar ideografik lugʻatning bir turi sifatida iyerarxik munosabatga asoslanadi va unda semantik modelning “jins-tur”, “qism-butun”, “sinonim” munosabatlari ifodalanadi. Tezaurus lugʻatlarda maʼlum atama uchun aloqador boʻlgan boshqa tushunchalar aks ettiriladi [Abduraxmonova, 2023].

Oʻzbek tili morfologiyasida, boshqa soʻz turkumlari qatori sifat ham katta oʻrinni egallaydi va oʻzining leksik-grammatik va semantik xususiyatlariga koʻra mustaqil soʻz turkumi sifatida ajratilgan. Tilshunoslikda sifat soʻz turkumining oʻrganilish tarixi juda qadim davrlarga borib taqaladi. Mahmud Koshgʻariyning “Devon-u lugʻatit turk”, Mahmud Zamaxshariyning “Muqaddimat-ul adab”, “Asos-ul balogʻa”, Alisher Navoiyning “Muhokamat-ul lugʻatayn” kabi asarlaridayoq sifatning belgi ifodalash xususiyati, ularning turli xil tashbehtar va badiiy sanʼatlarni yuzaga keltirishdagi imkoniyatlari haqida muayyan fikr-mulohazalar mavjud edi.

60-yillarga kelib oʻzbek tilshunosligida sifat turkumiga oid soʻzlarni oʻrganishga tilshunos olim M.Sodiqova katta hissa qoʻshdi. Olima 1963-yilda “Oʻzbek tilida rang-tus bildiruvchi sifatlar” mavzusida nomzodlik ishini himoya

qildi. Ikki tomlik “Hozirgi o‘zbek adabiy tili” (1966) va “O‘zbek tili grammatikasi” (1975) nomli tadqiqotlarning “Sifat” mavzularini yozdi, 1974-yilda olimaning “Hozirgi o‘zbek tilida sifat” nomli monografiyasi e‘lon qilindi.

O‘tgan asrning 70-90-yillarida o‘zbek tilshunosligida sifat turkumini talqin etish yanada chuqurlashdi va ilmiy tus oldi. Jumladan, keyingi o‘ttiz-o‘ttiz besh yil davomida o‘zbek tilshunosligida tilning grammatik qurilishini, lug‘aviy tarkibini o‘rganish sohasida sezilarli yutuqlar qo‘lga kiritildi. Pardayev Z.ning “O‘zbek tilida sifatlarning semantik-uslubiy xususiyatlari” nomzodlik dissertatsiyalarining bir bobi “Rang bildiruvchi sifat leksemalarning semantik-uslubiy xususiyatlari” deb nomlangan. Mazkur bobda “qora”, “oq”, “qizil”, “yashil” kabi tilimizda keng tarqalgan rang bildiruvchi leksemalarning semantik-uslubiy xususiyatlari tahlil qilingan [Pardayev, 2004]. N.Jumatova “O‘zbek xalq maqollaridagi rang bilan bog‘liq ramziy obrazlar” maqolasida “oq” va “qora” ranglari qatnashgan maqollarni atroflicha tahlil qilgan. Ming yillardan beri shakllanib kelayotgan maqol-u matallardagi ranglar jilosiga ahamiyat qaratgan [Isroilova, 2020].

O‘zbek tilida sifat so‘z turkumi bo‘yicha olib borilgan tadqiqotlarning ayrimlarini yuqorida keltirib o‘tdik. Hozirgi kunga qadar sifat leksik, semantik, uslubiy, funksional jihatdan o‘rganilgan bo‘lsa-da, bu turkum uchun o‘zbek tilida hali tezaurus lug‘ati yaratilmagan. Bu yo‘nalishda olimalarimiz N.Abduraxmonova, M.Abjalova, Sh.Hamroyeva o‘z munosabatlarini bildirib o‘tishgan. Bu esa, o‘zbek tili tezaurusi uchun sifat leksemalarning semantik munosabatlarini tadqiq etish hali tilshunoslar oldida turgan vazifalardan ekanligini bildiradi.

Tezaurus lug‘atlar bir so‘z semantikasi bilan bog‘liq barcha belgilarni qamrab oladigan, matnda qo‘llanish xususiyatlarini to‘liq aks ettiradigan kompyuter lug‘atidir. Tezaurus giperhavola vositasida boyitilgan katta hajmdagi ma‘lumotlar bazasidir. Xususan, “achchiq” leksemasining barcha ma‘naviy semalarini yig‘ib, quyidagicha tezaurusni ishlab chiqdik:

№	Tezaurus qolipi	So‘zning semantik pasporti
1.	So‘zning yoki til birligining morfologiya va sintaksisga aloqadorligi	O‘zbek tilida maza-ta‘m ma‘nosidagi achchiq so‘zi tezaurusini ko‘rib chiqamiz. achchiq – maza-ta‘m bildiruvchi sifat. Achchiq – ta‘mi og‘iz-tilni achitadigan, lovullatadigan: achchiq garmdori.
2.	Talaffuz qilinishi	achchiq
3.	Semantik tarkibi	<i>so‘zning asl ma‘nosi:</i> Achchiq – nordon, tishni qamashtiradigan meva-sabzavot navi: achchiq olma, achchiq karam. B) <i>ko‘chma ma‘noda:</i> Umuman, salbiy belgini bildiradi: 1.Achchiq – jazillatadigan, jonni achitadigan: achchiq shamol.

		2.Achchiq – kishiga qattiq tegadigan, noxush, ko'ngilga botadigan: achchiq gap. 3.Achchiq – g'am-iztirobli, alamli: achchiq damlar, achchiq xotira. 4.Achchiq – qahr, jahl, zarda: achchig'i burnini uchida. 5.Achchiq – qiyinchilik, mashaqqat, g'am-g'ussa, azob-uqubat kabi ma'nolarni ifodalaydi (O'TIL, II, 183) C) sinonimi: nordon, tilni lovullatadigan D) omonimi: – E) antonimi: mazali, totli, shakarli, yoqimli, chuchuk, lazzatli F) paronimi: – G) giperonimi: maza-ta'm
4.	Yaqin ma'nodagi so'zlar	achchiq so'z, achchiq hayot, achchiq taqdir, achchiq choy, achchiq shamol
5.	Etimologiyasi	Qadimgi turkiy tildagi a:chi-fe'lidan –q qo'shimchasi bilan yasalgan.
6.	Frazeologizmlarda qo'llanilishi	<u>Achchiq</u> <u>choy</u> <u>damlanib</u>, <u>shopirib-shopirib</u> <u>qaytarilib</u>, <u>xo'rillatib</u> <u>ichilgan</u> <u>damlarda</u> <u>kimlarning</u> <u>taqdiri</u> <u>hal</u> bo'lajak. Tohir Malik “Shaytanat 1-kitob”, 1989 Mahmudxon <u>achchiq</u> <u>ustida</u> <u>biror</u> <u>tarablarga</u> <u>ketub</u> <u>qolish</u> <u>ehtimoli</u> <u>ham</u> <u>bor</u>. Abdulla Qodiriy “Baxtsiz kuyov”, 1915 Faqat sizning shirin turmushingizga achchiq oralamasin, deyman-da....<i>R.Tagor “Aka va uka”</i> Uning oxirgi saltanati xonlik istibdodining achchiq zulm va taaddilarining jonso'z qora sahifalarini tashkil etadilar. Abdulla Qodiriy “Mehrobdan chayon”
	Birikmalar tarkibida	Yaxshidir achchiq haqiqat, Lek shirin yolg'on yomon. <i>E.Vohidov</i>
7.	Boshqa tillarda qo'llanilishi	- inglizcha: <u>bitter</u> - ruscha: горький

Tezaurus “achchiq” leksemasidagi quyidagi semantikani yoritish imkonini beradi: “achchiq” sifat leksemasida quyidagi semalar mavjud:

- 1) nordon; 2) tilni achitadigan; 3) zarda; 4) taxir; 5) jonni achitadigan

№	So‘z birikmalari	nordon	Tilni achitadigan	zarda	jonni achitadigan
1.	achchiq xotira				+
2.	achchiq choy		+		
3.	achchiq haqiqat				+
4.	achchiq gap			+	+

“Achchiq” leksemasining birikma hosil qilishida o‘ziga xos jihatlar ko‘zga tashlanadi. “Achchiq” leksemasi taom, suv, kulgu, hayot, xayol leksemasi bilan semantik maydon hosil qiladi, lekin “nordon” leksemasi achchiq leksemasida bog‘lanuvchi so‘zlar bilan semantik maydon hosil qilmaydi. Yuqorida keltirilgan jadvaldan shuni xulosa qilish mumkinki, “achchiq” leksemasi o‘z ma’no ottenkasi bilan bir qatorda ko‘chma ma’noda ko‘proq qo‘llanilar ekan.

Xulosa

Sifat leksemalarining semantik maydoni va tezaurusini yaratishda grammatik hamda sintaktik yaqinlikni hisobga olish lozim. Shundagina tezaurus lug‘atlarni yaratishda leksemalarning kontekst va semantik maydonini yarata olamiz. Demak, so‘zning semantik maydonga mansubligi sinxron xarakterga ega bo‘lib, davrlar osha o‘zgarishlarga uchrashi mumkin. Shuningdek, semantik maydonda so‘zlar bir-biri bilan tutashmasligi qayd etiladi. So‘z ma’nolarining bir-biriga tutashmasligi nisbiy tushuncha bo‘lib, etimologik tahlillar turli maydonga mansub leksemalarning semantik jihatdan bog‘lanishi mumkinligini ko‘rsatadi.

Foydalanilgan adabiyotlar:

1. Abduraxmonova N., Qodirova Z. Kompyuter leksikografiyasida ontologiya (monografiya) Toshkent – 2023. – B.71.
2. Israilova S. “Ko‘k” leksemasi semantikasi va tezaurusi. Monografiya. Toshkent – 2023. –B.91.
3. Pardayev Z. O‘zbek tilida sifatlarning semantik-uslubiy xususiyatlari. Samarqand – 2004. – B.7.
4. Isroilova S. Sifat so‘z turkumining o‘rganilishi “Filologik ta’limni takomillashtirish muammolari” Respublika ilmiy-amaliy anjumani materiallari. - Toshkent, 2020.

ZAMONAVIY TILSHUNOSLIK TERMINOLOGIYASI MUAMMOLARI

Axmedova Aynura

aynuraaxmedova345@gmail.com

ToshDO‘TAU talabasi

Annotatsiya. Ushbu maqolada tilshunoslikda antroposentrik paradigmaning kirib kelishi, uning leksikologiya va terminologiyaga ta'siri haqida so'z yuritilgan. Shu vaqtgacha nashr etilgan tilshunoslik terminlariga oid lug'atlar haqida ma'lumot berilgan. Shuningdek, zamonaviy lingvistik sohalar terminologiyasida lug'atshunoslik muammolari, xususan, tilshunoslikka oid terminlarning elektron va bosma nashrlardagi izohli lug'atini tayyorlash bo'yicha fikrlar bildirilgan.

Abstract: This article talks about the introduction of the antropocentric paradigm in linguistics, its influence on lexicology and terminology. Information about dictionaries of linguistics terms published so far is given. Also, the problems of lexicography in the terminology of modern linguistic fields, in particular, opinions on the preparation of an explanatory dictionary of terms related to linguistics in the electronic and printed publications are expressed.

Аннотация. В статье рассматривается внедрение антропоцентрической парадигмы в языкознание, ее влияние на лексикологию и терминологию. Приведены сведения об опубликованных к настоящему времени словарях лингвистических терминов. Также высказываются проблемы лексикографии в терминологии современных лингвистических направлений, в частности, мнения о составлении толкового словаря терминов, относящихся к языкознанию, в электронных и печатных изданиях.

Kalit so'zlar: *antroposentrik paradigma, antroposentrik sohalar, termin, terminologiya, leksikografiya, terminologik elektron lug'at.*

Dunyo tilshunosligida o'tgan asrlarda shakllangan antroposentrik paradigma o'zbek tilshunosligiga ham kirib keldi va jadallik bilan rivojlandi. Ma'lumki, til tizimiga dastlab semantik va sintaktik jihatdan yondashilgan bo'lsa, XX asr kashfiyoti – antroposentrik paradigmada tilga antroposentrik yondashish, tahlil qilish kuchaydi. Bunda har qanday matnga faqat lisondan ma'lum me'yorlar asosida nutqqa ko'chuvchi jarayon natijasi sifatidagina emas, balki ijtimoiy va boshqa xususiyatlarga ega bo'lgan muloqot, o'zida muayyan til egasining bilimlarini, lisoniy ko'nikmalarini, lisoniy tafakkurini jamlagan qurilma sifatida ham qarala boshlandi, ya'ni til egasi – shaxs birinchi planga ko'chdi. Natijada tilshunoslik o'ziga yondosh bir nechta sohalar bilan integratsiyaga kirishdi va tilshunoslikning yangi sohaları paydo bo'ldi. Sotsiolingvistika, psixolingvistika, neyrolingvistika, kognitiv tilshunoslik, kompyuter lingvistikasi, pragmalingvistika, lingvokulturologiya, aksiolingvistika, etnolingvistika kabi sohalarning paydo bo'lishi tilshunoslikning deyarli barcha ijtimoiy va tabiiy fanlar bilan o'zaro



bog'liqligini ko'rsatdi. Va albatta, ushbu yangi lingvistik sohalarning tadqiqot obyekti markazini shaxs omili tashkil etadi.

Tilshunoslikning yangi yo'nalishlari orasida kompyuter lingvistikasi bugungi kunimizda tez taraqqiy etmoqda. Chunki hozirgi axborot asrimizda ma'lumotni asl holicha qabul qilish, ishonchli saqlash va eng qulay tarzda boshqalarga yetkazib berish bevosita til bilan bog'liq. Tarjima qilish, til o'qitish jarayonini avtomatlashtirish dasturlariga bo'lgan ehtiyoj kompyuter lingvistikasi fanining shakllanishiga zamin yaratdi.[Xolmanova, 2019:7] Kompyuter lingvistikasining asosini mashina tushuna oladigan, qayta ishlash mumkin bo'lgan til modelini yaratish hisoblanadi. Kompyuter lingvistikasi rivojlanishi bilan bugungi kunda uning vazifalari kengaydi. Til korpusini yaratish, elektron lug'atlar, tezaurus lug'atlar, axborotni saqlovchi, yetkazib beruvchi har xil dasturlar ishlab chiqish kabi amaliy ishlar bugungi kunda kompyuter lingvistikasining oldiga qo'ygan vazifalarini tashkil etadi.

O'zbek tilshunosligiga yangi sohalar kirib kelishi bilan yangi o'zlashma terminlar ham kirib keldi. Sotsiolingvistika, psixolingvistika, lingvokulturologiya va boshqa sohalarga oid neologizmlar asl holida to'g'ridan to'g'ri o'zlashdi. Sotsiolekt, ideolekt, diglossiya, geshtalt, freym, skript kabi terminlar o'zbek terminologiyasini, leksikologiyasini boyitdi. Bunday terminlar ko'plab topiladi, hatto ayrim zamonaviy tilshunoslik sohalarida terminlar shunchalik ko'p o'zlashdiki, ular o'zlarining alohida terminologik tizimini shakllantirdi. Tilshunoslikka oid yangi terminlarning o'zbek leksikologik tizimiga kirib kelishi leksikologik, terminologik izohli lug'atlarni qaytadan tayyorlashni, yangilashni talab qilmoqda. Tilshunosligimizda lingvistik terminlarni umumleksikadan farqli ravishda alohida to'plagan, lug'at holiga keltirgan dastlabki olim akademik Azim Hojiyevdir. Muallifning 1985-yilda “O'qituvchi” nashriyoti tomonidan “Lingvistik terminlarning izohli lug'ati” chop etildi. Ushbu nashrning qayta to'ldirilgan shakli 2002-yilda “Tilshunoslik terminlarining izohli lug'ati” nomi bilan ommaga taqdim etildi. Bunda muallif yuqorida ko'rsatilgan lingvistik terminlar izohli lug'atidan farqli ravishda uni qayta ishlab to'ldirgan, nomaqbul terminlarni chiqarib tashlagan, imkon darajasida ruscha, ruscha-baynalmilal terminlarning o'rniga o'zbek tilidagi variantini, muqobilini ishlatishga harakat qilgan. Muallifning o'zi ma'lumot berishicha, lug'atda o'sha davrdan faol qo'llanilib kelayotgan ruscha-baynalmilal terminlarning to'liq uchdan bir qismi o'zbekchallashtirilgan. [Hojiyev, 2002: 3] Lug'atda 1700ga terminlar alifbo tartibida berilgan. Lug'atdagi har bir maqolada dastlab termin nomi bosh harflar bilan, keyin uning o'zlashgan yoki o'zlashmaganligi, o'zlashgan bo'lsa, qaysi tildan o'zlashganligi va terminning izohi keltirilgan. Keyingi tilshunoslikka oid lug'at Ne'mat Mahkamov va Ixtiyor Ermatov tomonidan 2013-yilda nashr etildi. Ushbu “Tilshunoslik terminlarining izohli lug'ati”da 1500ga yaqin termin keltirilgan. Lug'atning yuqoridagi lug'atlardan farqi uning, asosan, umumta'lim maktablari, akademik-litsey hamda kasb-hunar kollejlari o'quvchilari uchun yaratilganligi, terminga berilgan izohlar umumta'lim maktablari, akademik-litsey hamda kasb-



hunar kollejlari o'quvchilari uchun yaratilgan darsliklar asosida izohlanganligidir. Lug'atda termin nomi bosh so'z sifatida va agar o'zbek tilidan boshqa chet tillardan o'zlashgan bo'lsa, qaysi tildan o'zlashganligi, izohi, ko'p ma'noli bo'lsa har bitta ma'nosi alohida berilgan, izohni tasdiqlovchi misol, agar termin boshqa terminga aloqador bo'lsa, aloqadorligi maxsus belgilar bilan ko'rsatilgan (*ayn, qiyos, zid, qarang*). Bundan keyin tilshunoslik terminlarining umumiy izohli lug'ati yaratilmadi.

XXI asr boshlarida kirib kelgan tilga bo'lgan antroposentrik yondashuv jahon tilshunosligidan o'zbek tilshunosligiga o'zining nazariyalari bilan birgalikda terminologiyasini ham olib kirdi. Natijada bu holat hammaga tanish bo'lavermagan yangi terminlarga ixcham tarzda izoh berilgan lug'atlar yaratishni taqozo etdi. Nizomiddin Mahmudov va Durdona Xudoyberganovalar hammuallifligida yaratilgan “O'zbek tili o'xshatishlarining izohli lug'ati” (Toshkent, 2013) ana shunday lug'atlardan. Lug'at o'zbek tili egalarining obrazli tafakkuriga xos 530dan ortiq turg'un o'xshatishlarni o'z ichiga olgan va izohlar o'zbek tilidagi matnlar asosida misollar bilan dalillangan. Ushbu lug'atni lingvokulturologik xarakterdagi lug'at sifatida ham baholash mumkin. Chunki lug'atda izohlangan o'xshatish etalonlari xalqning milliy tafakkuri, madaniyati va obrazli fikrlashi mahsulidir. [Normamatov, 2020: 5] Shuningdek, lingvokulturologiyaning tilshunosligimizda boshqa sohalarga qaraganda qisqa fursatda jadal sur'atda rivojlanishi natijasida 2015-yilda alohida kitob holida D.Xudoyberganovanning “Lingvokulturologiya terminlarining qisqacha izohli lug'ati” nashrdan chiqdi. Kitob “So'z boshi”, “Lug'atning tuzilishi haqida” va lug'at maqolalaridan iborat. Lug'atda terminlar ko'p emas, faqatgina lingvokulturologiyada foydalaniladigan asosiy terminlar keltirilgan. Keyingi qisqacha izohli lug'at Guli Toirovaning “Pragmalingvistika” o'quv lug'atidir. Ushbu lug'at hajmi ham unchalik katta emas, pragmalingvistikadan tashqari boshqa lingvistik sohalarga oid ayrim tushunchalar ham izohlangan. Lug'at alifbo tartibida, taxminan 278ga yaqin so'zni qamragan.

Xulosa qilib aytganda, til rivojlanar, amaliyotda foydalanilar ekan, uning leksikasi, terminologiyasi ham kengayib boradi. Nafaqat amaliy, balki tilning nazariy jihatdan ham taraqqiy etishi muhim. Ana shu ehtiyojlar natijasida umumadabiy leksikadan leksikadan farqli ravishda o'ziga xos ilmiy leksika (terminologiya) shakllanadi. Albatta, leksika kabi chetdan o'zlashayotgan terminlarni ham tartibga solib turish muhim. Bunda bizga terminologik lug'atlar, ayniqsa, terminologik izohli lug'atlar yordam beradi. Bunday lug'atlarda har bir termin o'zi anglatayotgan tushunchani, mohiyatini to'liq va aniq ifodalashi zarur. Lug'atlar yaratilmaguncha terminologik tizimdagi har xil salbiy hodisalar, masalan, ayni bir tushunchani ifodalash uchun birdan ortiq terminning yoki aksincha bir nechta boshqa-boshqa hodisalarni ifodalash uchun bir terminning qo'llanilishi, hodisaning mohiyatini to'liq, aniq ifodalay olmaydigan terminlarning vujudga kelishi va boshqa shu kabi salbiy oqibatlarni keltirib chiqarishi mumkin. Yuqoridagi ma'lumotlarni kuzatar ekanmiz, “Tilshunoslik terminlarining izohli lug'ati” 2013-

yildan buyon yangilanmaganligiga guvoh bo'lamiz. O'tgan deyarli 10 yil ichida tilshunoslikka oid yangi terminlar ko'plab o'zlashdi. Tilshunoslik sohasining nazariy jihatdan rivojlanishi, yangi tadqiqotlarning paydo bo'lishi bunga keng yo'l ochdi. Tilshunoslikning ko'plab zamonaviy sohalari bo'yicha o'zbek tilshunosligiga dunyo tilshunosligining nazariyalari asos bo'ldi. Shu sababli xalqaro terminlar, inglizcha, ruscha-baynalmilal atamalar soni ancha ko'paydi. Bularni lug'atlarda jamlash, ularni tartibga solib borish muhim vazifalardandir. Shuningdek, tilshunoslikka oid terminlarning elektron shaklini tayyorlash, alohida dastur ishlab chiqish ham muhimdir. XXI asrda zamonaviy lingvistik sohalarda kompyuter lingvistikasi jadal tezlikda rivojlandi, rivojlanmoqda. Bugungi kunda tilni dasturlashtirish bo'yicha amaliy va nazariy ishlar olib borilmoqda. Adabiy tilni jahon miqyosiga olib chiqishda, uni xalqaro foydalanuvchi tillar qatoriga qo'shishda bu juda muhim. Til korpusi, murakkab tezaurus lug'atlar, elektron lug'atlar yaratish ustida ko'plab ishlar amalga oshirilmoqda. Xususan, tilshunoslikka oid jami terminlarni qamrab olgan murakkab, sifatli izohli lug'atlarning elektron va bosma nashrlarini yaratish lingvistik terminografiyaning kechiktirib bo'lmas vazifalaridandir.

Foydalanilgan adabiyotlar:

1. Dadaboyev H. O'zbek terminologiyasi. – Toshkent, 2019.
2. Dadaboyev H. Zamonaviy o'zbek leksikografiyasi va terminologiyasi. – Toshkent, 2016.
3. Hojiyev A. Tilshunoslik terminlarining izohli lug'ati. – Toshkent: O'zbekiston milliy ensiklopediyasi, 2002. – B. 4.
4. Mahkamov N, Ermatov I. Tilshunoslik terminlarining izohli lug'ati. – Toshkent: Fan, 2013.
5. Normamatov S. Leksikografiya asoslari. – Toshkent, 2020. – B. 84.
6. Toirova G. Pragmalingvistika. – Toshkent: Akademnashr, 2016.
7. Xolmanova Z. Kompyuter lingvistikasi. – Toshkent, 2019. – B. 14.
8. Xudayberganova D. Matnning antroposentrik tadqiqi. – Toshkent: Fan, 2013.
9. Xudoyberganova D. Lingvokulturologiya terminlarining qisqacha izohli lug'ati. – Toshkent: Turon zamin ziyo, 2015.

KOMPYUTER LEKSIKOGRAFIYASI VA LINGVODIDAKTIKASI

TURK TILI VA XORAZM SHEVALARI O'XSHASH LEKSEMALARINING QIYOSIY-TARIXIY VA ETIMOLOGIK LUG'ATINI ELEKTRON SHAKLDA YARATISH

Rahimbayeva Muhabbat Dushambayevna

mrd_aspi@mail.ru

UrDU Kompyuter ilmlari kafedrası dotsenti

Rajapboyeva Yulduzxon Ixomboy qizi

yulduzrajapboyeva@gmail.com

UrDU magistranti

Annotatsiya. Ushbu maqolada ko'pgina ilmiy tadqiqotlar qilishga yordam beradigan, turk tili o'rganuvchilari yoki shu tilga qiziquvchilar uchun foydali manba hisoblanadigan qiyosiy-tarixiy va etimologik lug'at yaratishga e'tibor qaratilgan. Unda turkiy tilidagi leksemalar bilan Xorazm shevalari o'rtasidagi o'xshashlik va bog'lanishlar o'rganiladi. Tabiiy tilni qayta ishlash (NLP) va qiyosiy o'rganish usullarini qo'llash orqali ushbu urinish ushbu tillarning tarixiy taraqqiyoti va umumiy ildizlarini yoritishga qaratilgan.

Abstract. This article focuses on creating a comparative historical and etymological dictionary that will help in many scientific studies and become a useful resource for Turkish language learners or those interested in the language. The similarities and connections between Turkic lexemes and Khorezm dialects are explored. Using natural language processing (NLP) and comparative learning, this attempt aims to shed light on the historical development and common roots of these languages.

Аннотация. Цель этой статьи создать сравнительно-исторический и этимологический словарь, который облегчит многие академические исследования и станет полезным ресурсом для изучающих турецкий язык или тех, кто интересуется этим языком. Изучены сходства и связи тюркских лексем с хорезмийскими диалектами. Используя обработку естественного языка (НЛП) и сравнительное обучение, эта попытка направлена на то, чтобы пролить свет на историческое развитие и общие корни этих языков.

Kalit so'zlar: *Etimologik lug'at, lahja, mushtaraklik, ma'noviy va shakliy mushtarak leksemalar, ma'nosi o'zgarishga uchragan leksemalar, omonim so'zlar, mobil ilova.*

Kirish

Qardosh tillarni qiyosiy-tarixiy va etimologik jihatdan tahlil qilish va mana shu nuqtayi nazardan o'rganish, ularning qardoshligini belgilash, tillarning paydo bo'lishidagi umumiy jihatlarni aniqlashda katta rol o'ynaydi. O'zbek tili boyligining asosi shevalarimizda hisoblanadi. Shunday ekan, shevalarimizni saqlab qolish

kelgusi avlodga yetkazish masalasi turibdi. Xususan, iste'moldan chiqib ketayotgan Xorazm shevalarini yosh avlodga o'rgatadigan o'quv manbalarini ishlab chiqish va bu sohani tadqiq qilish hozirgi dolzarb muammolardan biridir. Ushbu maqolada turk tili va Xorazm shevalaridagi o'xshash leksemalarning qiyosiy-tarixiy va etimologik lug'atini ishlab chiqish masalasi o'rganiladi. Etimologik lug'at tildagi muayyan leksemaning kelib chiqishi, qanday shakl va ma'no paydo bo'lgani haqidagi lug'at hisoblanadi. Etimologik tahlilga qiyosiy-tarixiy usul asos bo'ladi.

Xorazm viloyati o'g'uz va qipchoq lahjalari amal qiladigan O'zbekistondagi yagona hudud hisoblanadi. Bir necha shevalar jamlanmasi lahja hisoblanadi. Qiyosiy-tarixiy va etimologik lug'atni yaratish jarayonida turk tili va Xorazm shevalari obyekt qilib olindi. Mazkur muammoni yechish uchun nega aynan turk tili va Xorazm shevalari obyekt qilib olindi? Turk tili N.A.Baskakov [Abdurahmonov, 1982] tasnifiga ko'ra o'g'uz-saljuq guruhiga kiradi. Xorazm shevalari ham shu nuqtayi nazardan turk tiliga o'xshash deb qaraladi. Lug'atni ishlab chiqishda turk tili va Xorazm shevalaridagi o'xshash leksemalar qiyosiy-tarixiy va etimologik elektron lug'atini yaratish uchun qo'yilgan maqsad:

1) turk tili va Xorazm shevalariga xos o'xshashliklarni hamda bu elementlarning qay darajada umumturkiylikka borib taqalishini tahlil qilish;

2) turk tili va Xorazm shevalaridagi o'xshash leksemalarining qiyosiy-tarixiy va etimologik lug'atini mobil ilova shaklida ishlab chiqish.

Belgilangan maqsadga erishish uchun quyidagi vazifalarni bajarish lozim bo'ladi:

1) turk tili va Xorazm shevalarining o'zaro bog'liqligini tahlillar asosida o'rganish;

2) turk tili va Xorazm shevalaridagi o'xshashliklar va turkiy so'zlarning etimologiyasi xususida qilingan ishlarni, manbalarni tahlil qilish;

3) turk tili va Xorazm shevalaridagi o'xshash leksemalarni to'plash;

4) yig'ilgan ma'lumotlarni tartibga solib, misollar bilan izohlash;

5) to'plangan ma'lumotlar uchun mobil ilova strukturasini ishlab chiqish;

6) mobil ilova menyularini belgilash;

7) mobil ilovani Google Playga joylashtirish.

Turk tili va Xorazm shevalaridagi o'xshash leksemalarni ikkiga:

1) shakliy va ma'noviy mushtarak leksemalar;

2) ma'nosi o'zgarishga uchragan leksemalarga bo'linadi.

Mushtaraklik tushunchasi umumiy, birlashgan, o'rtadagi, ko'pchilikka taalluqli ma'nolarni bildiradi. Leksemalarning shakliy va ma'noviy mushtarakligi deganda esa biz shakl jihatdan ham, ma'noviy jihatdan ham bir-biriga o'xshashligini tushunamiz. Shakl jihatdan fonetik o'zgarishlarga uchragan, ammo ma'noviy jihatdan bir xil bo'lgan leksemalarga ham shakl va ma'noviy mushtarak leksemalar desak xato bo'lmaydi. Biz nazarda tutayotgan turk va o'zbek tilining Xorazm shevalarida shakl va ma'noviy jihatdan mushtarak leksemalar anchagina. Bunday so'zlar turk tilida va Xorazm shevalarida aynan bir xil tushuncha ifodalashga xizmat

qiladi. Quyidagi to'plangan so'zlarimiz hozirgi o'zbek adabiy tilida mavjud bo'lmagan, ammo turk tili va Xorazm shevasida ishlatiladigan, shakliy va ma'noviy jihatdan mushtaraklik kasb etuvchi so'zlar hisoblanadi: *alan, ari, ayaklanmak, aldanmak, baldız, bağırmaq, bebek, bal, barışmaq, danışmaq, dayı, dağılmaq, dede, diri, diz, donmaq, dönmek, düdük, düşmek, eşelemek, fişkırmak, fincan, gıdıklamak, gizlemek, hazzetmek, ısı, ikiz, ilişki* kabi leksemalar misol bo'ladi.

Turk tilida maydon, yer ma'nolarini anglatuvchi *alan* leksemasi Xorazm shevasida ham shu tarzda aytilib, *maydon, bo'sh yotgan yer* ma'nolarida ishlatiladi. F.Abdullayevning “O'zbek tilining Xorazm shevalari” lug'atida [Ahmedova, 2019] ham keltirilgan: *alan-bekor yotgan yer, bo'z yer; suv oqmaydigan eski ariq; kanaldan qazib olingan tuproq.*

Sakız leksemasi tarjimasi “saqich”. Turk tilida o'zbek tilidagi ikki qo'sh undoshli so'zlarda bitta undosh bo'lishi mumkin. Bu so'z ham shunga misol bo'ladi. Qadimgi turkiy tilda *sağız* shaklida qo'llangan. [Madrhimov, 1999]

Turk tili va Xorazm shevalarida shunday so'zlar mavjudki, shaklan bir-biriga o'xshash, ammo turk tilida ifodalaydigan ma'nosi Xorazm shevasida ifodalaydigan ma'nosi bilab o'zaro farqlanadi. Odatda, umumiy ma'no unsurlariga ega bo'lmagan, tasavvuriy bog'lanmagan, lekin bir xil yozilib, bir xil aytiladigan so'zlar omonim so'zlar deb talqin qilinadi. Omonim so'zlar orasida semantik aloqa bo'lmaydi. Ammo biz tahlil qilayotgan turk tili va Xorazm shevalaridagi o'zaro o'xshash leksemalarning orasida shunday so'zlar borki, ularning shakllari bir xil, ma'nolari boshqa-boshqa bo'lishidan qat'iy nazar qadimgi turkiy tildagi tub ma'nosi mushtarak hisoblanadi. Deyarli barcha so'zlarda umumturkiylik unsurlari mavjud, lekin yillar davomida ma'no torayishi, kengayishi yoki umuman boshqa ma'noli so'zga aylanib ketishi yuz bergan. Quyidagi so'zlar fikrlarimizning isboti bo'la oladi: *ata, azmak, beklemek, beğenmek, çıkışmaq, çöp, dağlamak, darılmak, dolanmaq, doğmak, eğlenmek, ek, ekmek, gövşemek hala, is, kapılmak, karma* kabi leksemalar misol bo'la oladi.

Turk tilidagi *kapılmak* leksemasi “berilib ketmoq, berilmoq” [Dadaboyev, 2015] ma'nolarini anglatadi. Xorazm shevasi o'g'uz lahjasida *qapılmaq* so'zi “jim bo'lmoq” ma'nosini anglatadi. Bundan tashqari “indamay o'tirmoq, og'zini yummoq, ovoz chiqarmay o'tirmoq, yo'qolmoq, tugamoq” ma'nolarida ham uchraydi. *M., Maña kegändä qapildi.* Bu misol qipchoq lahjasida “tugamoq” ma'nosida qo'llaniladigan *qapılmaq* so'ziga tegishlidir.

Mazkur to'plangan so'zlar lug'atimiz mobil ilova shaklida o'rganuvchilarga taqdim etiladi. Mobil ilova qandaydir bir qurilmalarda, smartfonlar, planshetlarda ishlashga mo'ljallangan dastur. Etimologik lug'at uchun mobil ilovani yaratish foydalanuvchi tajribasi, funksionallik va ma'lumotlar bazasini boshqarishni diqqat bilan ko'rib chiqishni o'z ichiga oladi. Quyida bunday mobil ilova uchun taklif qilingan tuzilma keltirilgan:

Bosh sahifa: Xush kelibsiz xabar va ilovaga qisqacha kirish; ilovaning turli bo‘limlariga kirish uchun navigatsiya menyusi; tezkor so‘zlarni qidirish uchun qidiruv paneli ko‘zga tashlanadi.

Qidiruv funksiyasi: Foydalanuvchilar turk tili yoki Xorazm shevalarida so‘z kiritishlari mumkin. Foydalanuvchilar so‘z tanlashda yordam berish uchun yozganda avtomatik to‘ldirish takliflari; natijalarni til, nutq qismi yoki etimologik ildizlar bo‘yicha filtrlash uchun kengaytirilgan qidiruv imkoniyatlari.

So‘z tafsilotlari sahifasi: Tanlangan so‘z haqida batafsil ma’lumot, jumladan: Etimologik kelib chiqishi va tarixiy konteksti.

So‘zning ishlatilishini ko‘rsatadigan misol jumalari.

Aloqador so‘zlar va turkum so‘zlar.

Masalan, turkcha: *uşak* “xizmatkor, erkak xizmatchi, bola” ma’nolarida uchraydi [Berdak, 1993]. Urganch, Xiva, Xonqa, Xazorasp hududlarida: *uşaq* “mayda, kichkina” ma’nosida uchraydi. Qipchoq lahjasida asosan, adabiy tilimizdagi ma’nosi “nonning mayda bo‘lakchalari” hisoblanadi. Yana bir ma’nosi “mayda pul” ham faol uchraydi. Qadimgi turkiy tilda “kichik bo‘laklarga ajra” ma’nosini anglatgan *uv* fe’lining “takror” ma’nosini ifodalagan *-sha* qo‘shimchasini olgan shaklidan *-q* qo‘shimchasini qo‘shish orqali yasalgan. O’sha davrdayoq so‘z tarkibidagi *v* undoshi talaffuz qilinmay qo‘ygan [Rahmatullayev, 2000].

Sevimlilar va tarix: Foydalanuvchilar sevimli so‘zlarni xatcho‘pga qo‘yishlari yoki tezkor ma’lumot olish uchun oxirgi qidiruvlarni saqlashlari mumkin; oldin qidirilgan so‘zlarni qayta ko‘rib chiqish uchun tarix jurnali; o‘rganishni mustahkamlash uchun interaktiv viktorinalar yoki o‘yinlar.

Sozlamalar: Til afzalliklari va interfeysni sozlash imkoniyatlari; yangilanishlar yoki yangi xususiyatlar uchun bildirishnoma sozlamalari.

Mazkur elektron lug‘atimiz turk tili va Xorazm shevalariga asoslangan, solishtirib qiyoslanib tahlil qilingan etimologik lug‘at hisoblanadi.

Xulosa

Etimologik lug‘atning mobil ilovasi yaratilishi til tarixini birlashtirish va madaniyatlararo rivojlanishga qaratilgan innovatsion urinishdir. Raqamli texnologiya kuchidan foydalangan holda ushbu innovatsion vosita turkiy tilshunoslik sohasida tadqiqot, o‘rganish va ta’lim olish uchun yangi imkoniyatlar ochadi. Til rivojlanishda davom etar ekan, bunday tashabbuslar til boyligini asrab-avaylash, aloqalarni mustahkamlashda muhim rol o‘ynaydi. Shuni ham alohida ta’kidlash kerakki, turkiy tillarni, xususan, turk tilini o‘rganish imkoniyatlarini kengaytirish va shevashunoslik sohasini rivojlantirib, ilmiy izlanishlar olib borish o‘zbek tilshunosligiga qo‘shilgan eng katta samara va foyda deyishimiz mumkin.



Foydalangan adabiyotlar:

1. Abdurahmonov G'., Rustamov A. Qadimgi turkiy til. – T., 1982.
2. Ahmedova N. O'zbek tili o'g'uz lahjasi va turk tilidagi bir xil qo'llanuvchi so'zlar tahlili//O'zbek tili taraqqiyoti va xalqaro hamkorlik masalalari: xalqaro ilmiy-amaliy anjuman materiallari. –T.: Navoiy universiteti nashriyot matbaa uyi, 2019.
3. Dadaboyev H. Turkiy tillarning qiyosiy-tarixiy grammatikasi. – T.: Tafakkur bo'stoni, 2015.
4. Hasan Eren. Türk Dilinin Etimolojik Sözlüğü. – Ankara, 1999.
5. Абдуллаев Ф.А. Ўзбек тилининг Хоразм шеваси. – Т.,1961.
6. Баскаков Н.А. Историко-типологическая фонология тюркских языков. – М., Наука. 1988. – С. 111-112.
7. Мадрахимов О. Ўзбек тили ўғуз лаҳжасининг Хива шеваси. – Урганч, 1999.
8. Раҳматуллаев Ш. Ўзбек тилининг этимологик луғати.–Т.: Университет, 2000.
9. Юсуф Бердак. Туркча-ўзбекча ва ўзбекча-туркча луғат.–Т.: Ўзбекистон, 1993.

ГУМАНИТАРЛЫҚ БІЛІМ БЕРУДЕГІ АҚПАРАТТЫҚ ТЕХНОЛОГИЯЛАР

Мұратбек Бағила Құрманбекқызы
Ф.ғ.к., доцент, Қ.Жұбанов атындағы АӨУ
Айтбенбетова Айгул Кадирановна
Ф.ғ.к., доцент, Қ.Жұбанов атындағы АӨУ

Аннотация. В данной статье речь пойдет о том, что развитие гуманитарного образования в условиях информационного общества в современном мире направлено на его интеграцию с естественнонаучным образованием. Это объединение обеспечивает строгость и точность технических и методологических аспектов гуманитарного знания, что определяет его эффективность и объективность.

Abstract. This article will focus on the fact that the development of humanitarian education in the conditions of the information society in the modern world is aimed at its integration with natural science education. This association ensures the rigor and accuracy of the technical and methodological aspects of humanitarian knowledge, which determines its effectiveness and objectivity.

Ключевые слова: методика обучения, интеграция, информационные технологии, дистанционное обучение

XXI ғасырда гуманитарлық білім берудегі ақпараттық-коммуникациялық технологиялардың рөлі артып келеді. Тарихи тұрғыдан алғанда, нақты пәндер әрдайым ақпараттық технологияларға жақын болды, ал гуманитарлық білім оларға қарамастан жеке дамыды. Гуманитарлық ғылымдар оқытушыларын бағдарламалауды үйрету арқылы ақпараттық технологиялармен таныстыру әрекеттері жасалды, бірақ бұл сабақтар сәтті болмады. Ауысым есептеу техникасы мен бағдарламалық жасақтаманың белсенді дамуы болған кезде пайда болды (жоғары деңгейлі бағдарламалау тілдері түрінде), содан кейін ақпараттық технологиялар тек гуманитарлық зерттеулерде ғана емес, сонымен қатар гуманитарлық білім беруде де қолданыла бастады. Осы кезеңде ақпараттық технологиялар гуманитарлық бағыттағы оқытушылар мен студенттерді негізгі үш міндетті шешуге мүмкіндік беретін деңгейде жалпыға бірдей компьютерлендіруді қамтамасыз етеді:

- интернетке шығуды қамтамасыз ету;
- бірыңғай ақпараттық кеңістікті дамыту;
- басқарылатын ақпараттық білім беру ресурстарын құру, дамыту және пайдалану.

«Білім беру ақпараттық технологиялары білім беру ортасында жүзеге асырылады, онда келесі жұмыс компоненттерін бөліп көрсету керек:

- техникалық (байланыс құралдары және компьютерлік техника);
- бағдарламалық-техникалық (іске асырылатын оқыту технологиясын бағдарламалық қолдау құралдары);
- ұйымдастырушылық-әдістемелік (барлық оқу процесін ұйымдастыру және студенттер мен оқытушыларға арналған нұсқаулықтар).»

Дәстүр бойынша гуманитарлық білім дүниетанымдық міндеттерді шешу, қазіргі әлеуметтік-мәдени жағдайда бағдарлау үшін іргелі негіздерді қалыптастыруға ұсынылды. Қазіргі жағдайда ақпараттық қоғам білім берудің гуманитарлық дайындығының міндеттері мен бағдарларының өзгеруімен байланысты.

Ақпараттық қоғамдағы гуманитарлық білім пәнаралық сипатқа ие болады, оның негізінде гуманитарлық білім беруде ақпараттық технологияларды қолдану әдісі жатыр. Пәнаралық білім беру объектісі қажетті деректерді тиімді өңдеуге, сақтауға және іздеуге бағытталған ақпарат пен ақпараттық технологиялар болып табылады. Маманданудың пайда болуы, білім берудің іргелес салаларының пайда болуы гуманитарлық ғылымды дамытады.

Қазіргі әлемде ақпараттық қоғам жағдайында гуманитарлық білім беруді дамыту оның жаратылыстану білімімен интеграциялануына бағытталған. Бұл біріктіру гуманитарлық білімнің техникалық және әдістемелік жақтарының қатандығы мен дәлдігін қамтамасыз етеді, бұл оның тиімділігі мен объективтілігін анықтайды. Ақпараттық технологиялар гуманитарлық білімнің өзгеруі мен дамуын анықтайды. Гуманитарлық білім беру құрылымына ғылыми білімнің жаңа салаларының, гуманитарлық бағыттағы маманданулардың пайда болуына әкелетін Ақпараттық технологиялар енгізілуде. Математикалық аппаратты және ақпараттық технологиялар құралдарын пайдалану есебінен үлкен көлемдегі ақпаратты өңдеу арқылы гуманитарлық білім берудің тиімділігі артады.

Қазіргі заманғы адам білімге ие болуы, білуі (ақпаратты іздеуі, табуы, туындаған мәселелерді шешуі) әртүрлі ақпарат көздерін пайдалануы, ақпараттық технологиялар саласында жаңа білім алуы (оларды оқу мен зерттеуде пайдалану үшін) керек. Осы міндеттердің барлығын шешу үшін гуманитарлық білім берудің бірыңғай білім беру моделі пайда болады. Ол заманауи ақпараттық және педагогикалық технологияларға, ақпараттық қоғамның әдіснамалық принциптеріне негізделген.

Сонымен, білім беру моделі - бұл біздің алдағы іс-әрекеттерімізді жоспарлауда маңызды рөл атқаратын және осындай әрекеттерде бұрын алынған білімді пайдалануға мүмкіндік беретін білім құрылымы. Ақпараттандыру гуманитарлық және жаратылыстану ғылымдарының синтезіне серпін береді. Пәнаралық - қазіргі білім берудің басты белгісі. Нәтижесінде бірқатар ғылыми гуманитарлық салаларда (экономикалық,

элеуметтік, тарихи, құқықтық, археологиялық, психологиялық, лингвистикалық және т.б.) «салалық» информатиканың пайда болуы. Сондай-ақ, информатика ғылымының өзіне және ақпараттандырудың қоғамның дамуына әсері туралы көзқарас өзгерді, ал гуманитарлық пәндерді оқытуды бүгінде ақпараттық технологияларды қолданбай елестету мүмкін емес. Қазіргі заманғы инновациялық технологиялардың бірі-қашықтықтан білім беру. Бұл күндері Интернет желісінің арқасында мүмкін болды. Интернет гипертехнология ретінде анықталады, оған басқалардың барлығы кіреді және барлығына бәрін береді. Сырттай білім беру қашықтықтан білім беруге батыл және сенімді түрде ауыстырылады.

Қашықтықтан білім беруді дамытушылар келесідей оң жақтарын көрсетеді:

- икемділік (студент сабақтың орнын, уақытын және ұзақтығын өзі жоспарлай алады);
- модульділік (материал модульдермен қамтамасыз етіледі, студент өзіне не ұнайтынын өзі таңдайды);
- қол жетімділік (географиялық және уақыт шеңбері студент үшін маңызды емес);
- рентабельділік (оқу орындарын ұстауға, оқу және әдістемелік материалдарды басып шығаруға жұмсалатын шығындар азаяды);
- ұтқырлық (оқытушы мен студент арасындағы кері байланысты тиімді жүзеге асыру);
- көптеген студенттерді қамту (әр түрлі ақпарат көздеріне (деректер банктері, сараптамалық базалар, электрондық кітапханалар, білім базалары және т. б.) параллель және бір мезгілде жүгіну);
- технологиялық (білім беру процесінде телекоммуникациялық және ақпараттық технологиялардың соңғы жетістіктері қолданылады).

Жоғарыда айтылғандардың барлығы қашықтықтан оқытуда тұлғаға бағытталған оқыту әдісінің белгілері айқын көрінеді деп есептей отырып, білім беру мінез-құлқын дараландыруды нақтылайды. Компьютер ақпаратты берудің барлық түрлерін қамтыды, ол оқытушыны студенттермен жедел және объективті кері байланыспен қамтамасыз етті. Сондай-ақ қашықтықтан оқытудың сырттай оқытудан маңызды айырмашылығы оқытуды дараландыруды жүзеге асыру болып табылады (дәптер мен қалам компьютерді, поштаны – электронды пошта, оқулықты - электрондық оқулықты және т.б. ауыстырады).

Электрондық кітапханаға қызмет көрсетудің кең жүйесі пайда болды, оған мыналар кіреді:

- электрондық оқу залдары арқылы жеке жұмыс орындарынан қолжетімділігі бар ақпараттық қоймалар, электрондық басылымдар желісі;
- кітапханалар бойынша ақпараттық-іздеу жүйелерінің желісі;

- электрондық ақпаратты дайындау және шығару базасы (электрондық басылымдар);

- бірыңғай жүйесі бар электрондық каталогтар.

Бұл жүйе Интернет желісінде ашық қолжетімділігі бар жұмыс орындарының кешендерін құруға ықпал етті. Электрондық кітапханалар гуманитарлық мамандарға ғана емес, сонымен қатар әртүрлі білім салаларының мамандарына да қызмет көрсетеді.

Жоғарыда айтылғандарды қорытындылай келе, ақпараттық және компьютерлік технологиялар гуманитарлық білім беру үшін жақсы алаң деп айтуға болады.. Гуманитарлық пәндер оқытушылары мен студенттерге қажетті траекторияда белсенді дамуға көмектесетін жаңа бағдарламалық өнімдер үнемі жасалып, білім беру процесіне ақпараттық-коммуникациялық технологияларды енгізу және қолдану неғұрлым кең болса, соғұрлым тиімді болады.

Әдебиеттер:

1. Титова С.В. Информационно-коммуникационные технологии в гуманитарном образовании: теория и практика. Пособие для студентов и аспирантов языковых факультетов университетов и вузов. М., П-Центр, 2009. 240 с.

2. Применение ИКТ в высшем образовании стран СНГ и Балтии: текущее состояние, проблемы и перспективы развития. Аналитический обзор / – СПб.: ГУАП, 2009. – 160 с.: ил.

ҚАЗАҚ ЖӘНЕ ТҮРІК ТІЛІНДЕГІ ОРТАҚ ЛЕКСИКА СИПАТЫ

Бисенгали Іңкәр Бақытжанқызы

Қ.Жұбанов атындағы Ақтөбе өңірлік
университетінің білім алушысы

Ғылыми жетекші: филология ғылымдарының кандидаты, доцент
(қауымдастырылған профессор) Мұратбек Бағыла Құрманбекқызы

Аннотация. Түркі тілдері - Қиыр Шығыстан Жерорта теңізіне дейінгі кең байтақ жерді алып жатырған, өзара туыстас халық тілі. Түркі тіл семьясына жататын тілдер ішінара грамматикалық, фонетикалық, лексикалық және синтаксистік құрылымы жағынан ұқсас болып келеді. Бұл жұмыста қазақ және түрік тілінің құрылымдық ұқсастықтарын талдай отыра, ортағасырлық еңбектер арқылы нақты дәлелдер келтірілетін болады. Ортағасырлық Махмұд Қашқаридің «Диуани лұғат-ит-түрк» еңбегіндегі сөздердің қазақ және түрік тіліне ортақтығы сөз етіледі. Сол уақыттардағы еңбектер араб тілінде жазылу себебі мен араб тілінің түрік және қазақ тілінде көрініс табатын лексикалық құрылымның ұқсастықтары көрсетіледі. 8-ғасырға тән «Күлтегін ескерткіші» Ғұбайдолла Айдаровтың зерттеунамасы негізінде бүкіл түркі дүниесіне ортақ ескерткіштің тілдік құбылыстарына кеңінен тоқтала отыра, қазіргі тілдік бірлікпен салыстырылады. «Орхон-Енисей» жазба ескерткіші түркі тілдес халықтардың ортақ ескерткіші ретіндегі маңызын көрсету арқылы, тарих қойнауына шолу жасалады. Синтаксистік құрылым бойынша қазақ және түрік тілдерінің орыс және ағылшын тілдерінен айырмашылығы көрсетіледі. Түркологияда түркітілдес халықтардың тілінің тарихи тамырын бір арнадан іздеу ежелден бар үрдіс. Осы үрдіс негізінде өткен мен бүгінді талдап, бір арнаға тоғыстыратын боламыз.

Abstract. The Turkic languages are mutually related vernacular languages that occupy a vast territory from the Far East to the Mediterranean Sea. The languages belonging to the Turkic language family are partially similar in grammatical, phonetic, lexical and syntactic structure. Analyzing the structural similarities of the Kazakh and Turkish languages, this work will provide convincing evidence of medieval writings. The medieval work of Mahmoud Kashgari "Divani lugat-it-Turk" expresses the commonality of words in the Kazakh and Turkish languages. The works of that time reflected the similarity between the Arabic language and the lexical structure of the Arabic language, reflected in Turkish and Kazakh. The "Monument to Kultegin", characteristic of the 8th century, is compared with modern linguistic unity, broadly touching upon the linguistic phenomena of the monument, common to the entire Turkic world. The written monument "Orkhon-Yensei", reflecting its importance as a common monument to Turkic-speaking peoples, provides an overview of the depths of history. The syntactic structure indicates the difference between Kazakh and

Turkish languages from Russian and English. In Turkology, the search for the historical roots of the language of the Turkic-speaking peoples through one channel is a long-standing trend. Based on this trend, we will analyze the past and the present and combine them into one channel.

Аннотация. Тюркские языки-это взаимно родственные народные языки, занимающие обширную территорию от Дальнего Востока до Средиземного моря. Языки, относящиеся к тюркской языковой семье, частично схожи по грамматической, фонетической, лексической и синтаксической структуре. Анализируя структурные сходства казахского и турецкого языков, в данной работе будут приводиться убедительные доказательства средневековых трудов. В средневековом произведении Махмуда Кашгари «Дивани лугат-ит-Турк» выражена общность слов на казахском и турецком языках. В произведениях того времени отражено сходство между арабским языком и лексической структурой арабского языка, отраженной в турецком и казахском языках. «Памятник Культегину», характерный для 8-го века, сравнивается с современным языковым единством, широко затрагивая языковые явления памятника, общего для всего тюркского мира. Письменный памятник «Орхон-Енсей», отражая его значение как общего памятника тюркоязычным народам, дает обзор глубин истории. По синтаксической структуре указывается отличие казахского и турецкого языков от русского и английского. В тюркологии поиск исторических корней языка тюркоязычных народов по одному каналу-давняя тенденция. На основе этой тенденции мы будем анализировать прошлое и настоящее и объединять их в одно русло.

Кілт сөздер: синтаксистік құрылым, фонетикалық құрылым, түркі тіл семьясы, лексикалық құрылым, грамматикалық құрылым

Әдеби тіл тарихы - сан ғасырлық тарихи кезеңдердің негізінде қалыптасып, сұрыпталып, өңделіп қазіргі уақытта тарихи категорияға ие болып отыр. "Әдеби тілдің тарихы - бұл жазба ескерткіштердің тарихы" деген пікірді академик Р.Сыздықова "Қазақ әдеби тілінің тарихы" атты еңбегінде арнайы келтіреді [1,5]. Түркі тілдері- Қиыр Шығыстан Жерорта теңізіне дейінгі кең байтақ жерді алып жатырған, өзара туыстас халық тілі. Түркі тіл семьясына жататын тілдер өзара грамматикалық, фонетикалық, синтаксистік және лексикалық құрылымы жағынан ұқсас болып келеді. Сол түркі тілдері тобына жататын қазақ пен түрік тіліндегі құрылымдық ұқсастықтарын көрсету үшін ортағасырлық бірқатар еңбектерге шолу жасайтын боламыз.

Ортағасырлық Махмуд Қашқаридің “Диуани лұғат-ит-түрк” (Түркі тілдері сөздігі) еңбегі түркі дүниесіне ортақ мұра және ежелгі тіл жәдігері. Бұл еңбек араб тілінде жазылған түркі сөзінің аудармалары. Бұл еңбектен қазақ және түркі тіліне ортақ лексиканы көре аламыз. Алдымен, бұл еңбек не себепті араб тілінде жазылды деген сауалға тоқталайық. Тарих қойнауына көз

жүгіртетін болсақ 751 жылы Атлах шайқасы түркі әлеміне үлкен өзгеріс әкелді. Біріккен араб пен түрік әскері Тан империясымен соғысының нәтижесі жеңіспен аяқталды. Соғыстың тағы бір нәтижесі күллі түркі жеріне араб мәдениеті мен ғылым-білімі таралды. Әсіресе, 8-10-ғасырларда Ұлы Жібек жолы бойына арабша сауат ашатын медреселер салынып, көне түркі жазуының орнын араб жазуы басты. Араб тілі түркі халықтарында әдеби тіл қызметін атқарып, түркі әлеміне құнды еңбектер араб тілінде жарық көрді. Дегенмен, біздің түркі халықтарының тілі жойылып кетпеді, тек араб қаріпіне көшті.

Талдау:

- « اِنْكَ » іңек деп оқылатын бұл сөздің мағынасы сиыр деген ұғымды білдіреді. Түрік тілінде «сиыр» сөзі дәл осылай, яғни «inek» деп қолданылады. Бір қарағанда араб тілінде жазылғанымен бұның түрік сөзі екендігін аңғаруымызға болады. Себебі араб тілінде сиыр сөзі «البقرة» «бақара» деп қолданылады.

- « اُنْتُى » сөзінің оқылуы унутты. Қазақ тілінде «ұмытты» деген ұғымды білдіретін бұл сөз түрік тілінде «unuttu» деп қолданылып, дәл осы мағынаны білдіреді. Араб тілінде «نسييت» «ұмытты» сөзінің жазылуы әрі айтылуы басқа фонетикалық құрылым екенін негізінде құрылған [2,252].

- « بُشْنى » бос мағынасында жазылған бұл сөздің екі тілде де беретін мағынасы «бос» (қазақ тілінде) «boş» (түрік тілінде), ал негізгі араб тілінде «فارغة» «фарифа» [2,386].

- « قشْنى » қыш деп оқылатын бұл сөз қазақ және түрік тілінде де бір мағынада «қыс» мезгілін білдіру үшін жұмсалады [2,389].

- « اَرْ » ер мағынасында қолданылатын сөз. Қазақ тілінде де түрік тілінде де «ер, ер адам» мағынасында лексикалық қорымызда бар.

- « اَوْتْ » EVET EBET: «дұрыс, мақұл, әлбетте» деген мағынадағы сөз.

Бұл сөз қазақ тілінде қолданыста жоқ, ал түрік тілінде осы күйінде қолданылады.

Диуани лұғат-ит-түрк” еңбегіндегі «Алып ер Тұңғаны жоқтау» жырының тілдік қолданысындағы шылаулар:

- Алп ер Тұңға өлдүмү
Әсіз аздұн қалдымү
Өзлэк өдшін алдымү
Өмді йүрэк йүртілүр [3,63]

Аударма:

Алып ер Тұңға өлгені ме,
Өзіз дүние қалғаны ма.
Тағдыр өшін алғаны ма
Енді жүрек жыртылды

«өлдүмү»; «қалдымү»; «алдымү» - бұл жерде тіркесіп тұрған «мү» қазіргі қазақ тіліміздегі «ма, ме, ба, бе, па, пе» сұраулық мағынадағы шылаулар. Қазақ тілінде дыбыстық өзгеріске ұшырған.

Түрік тіліндегі нұсқа бойынша, «мү» - «tu, mi, mü, mı» сұраулық шылау. Бұл жерде фонетикалық құрылым сақталған.

Осы мысалдар арқылы Махмуд Қашқаридің “Диуани лұғат-ит-түрк” (Түркі тілдері сөздігі) еңбегі араб тілінде жазылды дегенімізбен, сөздік ішінде сөз қолданыстары көне түркі тілін құрайды. Сонымен қатар, қазіргі лексикалық қорымызда қолданыста жүрген бұл сөздер 11-ғасырлық еңбекке тән екендігін көре аламыз.

Араб тілі түркі тілдер семьясына жатпағанымен, қазақ және түркі тіліне ұқсайтын лексикалық қоры да жоқ емес. Сөздік ретінде жинастыратын болсақ, мына бір мысалдарды келтіруімізге болады:

Араб – түрік – қазақ
دفتر – defter – дәптер
كتاب – kitap – кітап
دكان – dükkân – дүкен
دنیا – dünya – дүние
فيل – fil – піл
خبر – haber – хабар
حقیقة – hakikat – ақиқат
بركة – bereket – берекет
هواء – hava – ауа.

Түрік халықтарында араб тілі әдеби тіл қызметін атқарды десек, осы ретте әдеби тілдің қалыптасу жолдарын ашып көрсетейік. Әдеби тілдің негізгі үш қалыптасу жолы бар.

1. Бөтен халықтың тілінен қалыптасады;
2. Ең күшті диалектіден қалыптасады;
- Бұл түрколог ғалым Сәрсен Аманжоловтың тұжырымдамасы бойынша
3. Жалпыхалықтық тілдің негізінде қалыптасады;

Демек, араб тілінің түркі тіліне әсерін әдеби тілдің қалыптасуының бірінші жолы ретінде қарастырамыз. Түркі тілдерінің лексикалық қорын араб тілі ғылым-білім, дін негізінде құрады. Бұл жоғары дамыған мәдениет пен өркениет, әрі тарихи оқиғалардың көрінісі.

Келесі, талдайтын еңбегіміз 8 ғасырға тән «Күлтегін ескерткіші» Ғұбайдолла Айдаровтың зерттеунамасы. 1995 жылы жарық көрген еңбекте түркі тілдерінің ортақ тілдік құбылыстарыды салыстырмалы лексикология негізінде қарастырамыз. «Күлтегін ескерткіші» 682 жылы Шығыс түрік қағанатының негізін қалаған Елтеріс қағанның ұлы. Орхон Енисей өзені бойынан табылған ескерткіште ержүрек қолбасшы Күлтегіннің ерлігі баяндалған. Түркі әлеміне ортақ мұра көне түркі тілінде жазылған.

И.Т.Страленберг швед офицери, Енисей жазуын тауып, тарихта «руна» деген атау берді. Көне түркі жазуы қазіргі түрік және қазақ тілінің лексикалық құрылымында қолданылуына тоқталайық.

Көне түркі тілі – қазақ тілі – түрік тілі

АТЫН – оның атын – onun adı [4,19]

- АТ – адамның есімін атағанда қолданылатын сөз. Бұл сөздің қолданылуы қазіргі қазақ және түрік тіліне де ортақтығын көреміз. Мысалда тәуелдік жалғауы арқылы берілген ұқсастық.

ЕБ – үй – ev [4,20].

- ЕБ - бұл сөздің қазіргі уақыттағы қолданылу аясы тек түрік тіліне ғана тән екендігін көреміз. Көне түркі тілінің ескерткішіндегі сөз бүгінгі күні түрік әдеби тілінің негізінде қарастырылып отыр.

ОТУЗ – отыз – otuz [4,22].

ИІГІРМІ – жиырма – yirmi [4,40].

ЙЕТІ – жеті – yedi [4,36]

- Сандық мағынаны білдіретін сөздерде фонетикалық айырмашылық бар. «Иігірмі» және «Йеті» сөздерінде қазақ тіліндегі «-ж» дауыссызы түрік тілінде «-у» дыбысына ауысып, фонетикалық өзгеріске ұшырап отыр.

ЙЫМЧАҚ – жұмсақ – yumuşak [4,37]

ЙАШЫЛ – жасыл – yaşıl [4,38]

ЙЫЛ - жыл - yıl [4,38]

ЙОҚ – жоқ – yok [4,75]

ЙОЛ – жол – yol [4,36]

- Мына бір мысалдардан да қазақ тіліндегі «-ж» түрік тілінде «-у» дыбысына ауысып, фонетикалық өзгеріске ұшырағандығын көреміз.

ТАШ – тас – taş [4,32]

ҚАШ – қас – kaç [4,76]

- Түрік тіліндегі «-ш» дыбысы қазақ тілінде «-с» дыбысына ауысып қолданылған.

Филология ғылымдарының докторы Ғұбайдолла Айдаровтың «Күлтегін ескерткіші» зерттеунамасы бойынша көне түркі тілі қазақ тіліне аударылған. Саралау жүйесі негізінде қазақ тіліндегі бүгінгі сөздік қорымыздағы сөздер түрік тілінде де қолданыста екендігін анықталды. Қазіргі қазақ тілі мен түрік тіліне де ортақ лексика 8-ғасырға тән ескерткіштен де аңғаруымызға болады.

Академик, профессор, филология ғылымдарының докторы Рабиға Сыздықованың «Қазақ тіліндегі ескіліктер мен жаңалықтар» еңбегін негізге ала отыра, ортағасырлық Қадырғали Жалайыри «Жами'ат-тауарих» атты жылнамалық еңбегінен қазақ тіліне қатысы бар сөздердің қолданысы арқылы түрік тіліне ұқсас тілдік ерекшеліктерді байқауымызға болады.

«САНДАЛ» - қазақ тілінде «орындық, тақ» деген мағынаны білдіреді. Алайда, қазіргі тілдік қолданыста бұл сөз қазақ тілінде жоқ болғанымен, түрік тілінде «sandalye», яғни орындық мағынасында қолданылып жүр. Еңбектен

үзінді: «Мубарак айтунлық сандалны кетуруб, мәсжид ішінде құрдылар»[5,157].

«ТАМАМ БОЛДЫ» - қазақ тілінде «аяқталды, жарайды болды, тәмәмдалды» мағынасында қолданылады. Түрік тілінде «tamam» сөзі дәл осы мағыналық қолданыста жұмсалады [5,158].

«ШАҺАР» - қазақ тілінде көне қолданыстағы «үлкен қала» сөзі қазіргі түрік тілінде «үлкен қала», яғни «şehir» мағынасында қолданылады[5,160].

«ЙАРАШМАҚ» - қазақ тіліндегі мағынасы «жарасу». Түрік тілінде «уараşmak» жарасу, лайық болу мағынасында қолданыста жүр. Еңбектен үзінді: «Йарашмақлықны охшатыб елчі йіберді» [5,153].

«СОҢРАТЫН» - «кейін, кейіннен, соңынан» мағынасын білдіріп, түрік тілінде «sonra» кейін мағынасында жұмсалады. Еңбектен үзінді: «Моғоллар уа түрклер соңратын бу сөзні суралар» [5,158].

Осы тұста түрік тілі морфологиясына тоқталсақ. Қазақ тілінде көптік жалғау « -лар , -лер, -дар, дер, -тар -тер» болып, алты формада кездессе, түрік тілінде « -лар, -лер» деген екі формада ғана сөзге жалғанады. Жоғарыдағы мысалда «Моғоллар, түрклер» сөзіне жалғанған көптік жалғау екі тілде де кездеседі, алайда түрік тілі морфологиялық құрылымына жақын.

«ҚЫШҚЫСЫН» - «қыста» [5,156]

Қыс сөзі түрік тілінің «kış» лексикалық және морфологиялық нормасы негізінде жасалған.

«ЙҮЗ ҰРУ» - «бет алу, бет алып жүру» [5,154]

Мән берілу керек тұс «йүз» сөзі. Адамның жүзі деген мағынада қолданылады. Түрік тілінде дәл осы қолданыста. Фонетикалық өзгеріс алғашқы көрсеткен мысалдарымызға саяды.

Қазақ және түрік тіліндегі синтаксистік норма бірдей формада жасалады.

Мысалы, «Айша мені жетім қалған үйге жетектеп кіргізді». (Ш.Мұртаза «Ай мен Айша»)

Қазақ тіл білімінің негізін салушы А.Байтұрсынұлының «Тіл құрал» еңбегіне сүйенсек, бастауыш мүше сөйлемді бастап, баяндауыш мүше сөйлемді тиянақтап тұрады. Бұл қазақ тілі синтаксисінің негізі заңдылықтарының бірі. Дәл осы заңдылық түрік тілі синтаксисінің принципі.

Мысалы, «Halime Kadın salı sabahı erkenden kalktı».(«Sıcak ekmek» Gülten Dayıoğlu)

Демек, түрік тілінің синтаксисінде де бастауыш мүше сөйлемді бастап, баяндауыш мүше аяқтап тұрғандығын көруімізге болады.

Қазақ және түрік халықтарының ортақ қырлары жетерлік, бұл екі ұлттың түркі тектес болуымен түсіндіріледі. Мұндай ұқсастық белгілері олардың мәдениетінде, дінінде, дүниетанымында, салт-дәстүрлерінде ғана емес, тілдерінде де көрініс табатынын көрдік. Сонымен қатар, қазақ тілі түркі семьясына жататын тілдердің ішінде түрік тілімен салыстыра қарап, өз

кезегінде ішінара ұқсастықтарды(фонетикалық тұрғыда,лексика) тілдің элементтерін көрсету.Қорыта келгенде, тіліміздің тарихының қайнар көзі болған «Орхон-Енисей» ескерткіштерінде, Махмуд Қашқаридің “Диуани лұғат-ит-түрк” (Түркі тілдері сөздігі), Қадырғали Қосымұлы Жалайыри «Жами`ат-тауарих» атты жылнамалық еңбегінде қолданылған сөздер (лексика) қазіргі қазақ тілі мен түркі тілінен этногенезистік тұрғыда көп алшақтай қоймады.

Әр ұлттың, этностың тарих сахнасында өзіндік орнын ойып тұрып алуы оның тілі, ділі, мәдениетімен тығыз байланысты. Тарихта орын алған сан ғасырлық оқиғалардың нәтижесінде өз тілін сақтап қалған түркі тілдерінің, оның ішінде бүгінгі зерттеуге арқау болып отырған қазақ және түркі тілінің тарихы сонау тереңнен басталып, бүгінгі күнге дейін өз тіл тәуелсіздігін сақтап келеді.

Пайдаланылған әдебиеттер:

1. Сыздықова Р. "Қазақ әдеби тілінің тарихы" Алматы,"Ана тілі", 1997 ж.
2. Қашқари М. “Диуани лұғат-ит-түрк” (1- том. Аударған Егеубай А.Қ). Алматы,2017 ж.
3. "Қазақ әдеби тілінің тарихы" хрестоматия. Қалиев Ғ , Боранбаева. Ғ (Алматы, 2000 ж.)
4. Айдаров. Ғ «Күлтегін ескерткіші» (Алматы, 1995 ж.)
5. Сыздықова. Р «Қазақ тіліндегі ескіліктер мен жаңалықтар» (Алматы, 2009 ж.)
6. Қадырғали Жалайыри «Жами`ат-тауарих» (1997 ж.)
7. Мұратбек Б.Қ, Әмірғали Ж.О. «Қазақ-түрік тілдері ортақ лексикасының салыстырмалы сөздігі». Ақтөбе, 2022ж.

КӨНЕ ТҮРКІ ЕСКЕРТКІШТЕРІНІҢ ГРАММАТИКАСЫ

Кенебаева Ұлпан Шапағатқызы

Қ.Жұбанов атындағы Ақтөбе өңірлік
университетінің білім алушысы

Ғылыми жетекші: филология ғылымдарының кандидаты, доцент
(қауымдастырылған профессор) Мұратбек Бағыла Құрманбекқызы

Қазақ әдеби тілі сонау ерте түркі заманынан беріге дейін сүзгіден өткен сан ғасырлық тарихы бар тіл. Әдеби тіліміздің тарихын зерттегіміз келсе, тіліміздің тарихына аса мән беру керек. Қазақ тілі түркі тілдес халықтардың тобына жататындықтан, көне түркі ескерткіштерін зерттеуді негізге аламыз. Бүгінгі тіл топтары көші-қон нәтижесі немесе өмір сүру дағдысындағы өзгерістерден пайда болған мәдени диффузиялардың нәтижесі екенін көрдік. Жоғарыда түсіндіріп өткеніміздей, соңғы бес, он мыңжылдықтарда бұл өзгерістер өте қарқынды жүрді, бүгінгі күнгі тіл топтарының тарауына осынша уақыт кетті деп болжауымыздың бір себебі осы [1.45]. Қазақстанның көне жазба ескерткіштерін одан әрі анықтау мен зерттеудің түркі халықтарының көне мәдениетін және олардың әлемдік өркениетке қосқан үлесін тереңірек түсінуге мүмкіндік беретін жаңа ғылыми проблемаларды анықтаудың маңызы зор. Түркі руна жазбаларының графикасын жан-жақты зерттеу қазақ тілінің әдеби тілінің тарихымен кеңінен танысуға ықпал жасады. Сондықтан, Қазақстанның көне жазба ескерткіштерін одан әрі анықтау мен зерттеудің түркі халықтарының көне мәдениетін және олардың әлемдік өркениетке қосқан үлесін тереңірек түсінуге мүмкіндік беретін жаңа ғылыми проблемаларды әзірлеу үшін маңызы зор. Қазақ халқының ұлт болып қалыптасуына зор үлес қосқан халықтар мен ұлыстар, тайпалар мен тайпалар өз дәуірінде қолданыста болған тілдер қазіргі қазақ тілінің қалыптасуы мен дамуының қайнар көзі болып табылады. Бұл жұмыста көптеген деректік материалдарды пайдалана отырып, көне түркі руникалық жазбалары тілінің графикалық жүйесінің грамматикасын зерттеу міндеті қойылған.

Көне түркі ескерткіштері, түркі жазба және әдеби тілінің үлгілері көне түркілердің тілідес халықтарға ортақ болғанына бірден-бір дәлел. Ең көнелері VII ғасырдың аяғына жататын түркі руна ескерткіштерінің морфологиясы мен синтаксисінің даму дәрежесіне қарай, әдеби ескерткіштер түрінде өзіндік ерешеліктер мен ұқсастықтар бар. Көне түркі ескерткіштерінің Тоныкөк және Күлтегін тарихи шежіресінде сақталған түркі жазба және әдеби тілінің: оның да бұрынғылары болғаны сөзсіз. Стандартталған лексикалық, морфологиялық және синтаксистік формалар мен сөз тіркестері бар көптеген эпитафиялармен ұсынылған Енисей руникалық ескерткіштері сөзсіз тілдік дамудың ұзақ жолын көрсетеді белгілі бір стилі мен жазба және әдеби тілі, әсіресе үлкен Орхон ескерткіштерімен жарқын және толық бейнеленген, ішкі өмірі реттелетін

халықтың ірі тайпалық құрылымдарының ауызша - эпостық дәстүрі негізінде пайда болды және дамыды. Қазақ әдеби тілінің тарихында елеулі орны бар көне ескерткіштерге назар саламыз.

Негізгі бөлім

Түркітілдес халықтарға ортақ көне түркі ескерткіштер тілін «руна жазуы» деп атайды. «Руна» ескі скандинав тілінен шыққан, «жұмбақ» дегенді білдіреді. Бұл атауды қазіргі ғалымдар емес, руникалық жазбалардың шығу кұпиясын әлі де шешпеген кезде, белгілер, сиқырлы күштерді жатқызған ежелгі немістер қойған.

Түркі ханзадасы, кейінірек қағаны Йоллығ Тегін алғашқы атақты түркі ақыны, жазушысы және тарихшысы деп атауға болады. Ол түркі ханзадалары Күлтегін мен Білге қағанның, сондай-ақ Құтлұғ, Ілтеріс қағанның құрметіне жазылған бірқатар мемориалдық жазбалардың авторы болған деген болжам бар. Бұл жазулар ежелгі түріктердің мәдениетін, поэзиясын, прозасын, Шығыс Түрік қағанатының тарихи танымын және мемлекеттік идеологиясын көрсетеді. Ежелгі түркі жазуын сол кездегі диалектіліктен жоғары қойған (түркіттердің билеуші тайпасының тіліне негізделген, оны орхон-түркі тілі деп те атайды және белгілі бір дәрежеде әдеби деп те атауға болады) қолданған тіл. Руна жазулары Батыс Еуропаның әр жерінде кездеседі. Олардың көпшілігі Швецияда табылды - бірнеше мыңжылға тиесілі болды. Ең ерте руникалық мәтіндер Ютландия түбегі мен оған іргелес аралдар аумағында табылған - олар III ғасырға жатады. Сонымен қатар, бұл жазулар Англияда, Шотландияда, Норвегияда, Исландияда, мен Германия Федеративтік Республикасында, Францияда, Австрияда, Грецияда, Румынияда және Қара теңіздегі Березан аралында бар. Тіпті біздің елде, Волынь қаласында, Ковель қаласына жақын жерде руна жазуы бар найзаның ұшы табылды. Ежелгі түркі жазуын сол кездегі диалектіліктен жоғары қойған (түркіттердің билеуші тайпасының тіліне негізделген, оны орхон-түркі тілі деп те атайды және белгілі бір дәрежеде әдеби деп те атауға болады) қолданған тіл.

Түркі ханзадасы, кейінірек қағаны Йоллығ Тегін алғашқы атақты түркі ақыны, жазушысы және тарихшысы деп атауға болады. Ол түркі ханзадалары Күлтегін мен Білге қағанның, сондай-ақ Құтлұғ, Ілтеріс қағанның құрметіне жазылған бірқатар мемориалдық жазбалардың авторы болған деген болжам бар. Бұл жазулар ежелгі түріктердің мәдениетін, поэзиясын, прозасын, Шығыс Түрік қағанатының тарихи танымын және мемлекеттік идеологиясын көрсетеді (көне түрікше: Түрік немесе Оңтүстік Сібірден табылған көне түркі руна жазуының ескерткіштері Моңғолия, Қырғызстан мен Қазақстан – тіл, тарих, этногенез, география және рухани мәдениет бойынша аса құнды дереккөз көне түркі тайпаларының жазба дәстүрі, наным-сенімдері мен дүниетанымының бастауы.

Көне кезеңнің тілдік ерекшеліктерін белгілейтін мұралар көпшілік әдебиетте Орхон-Енисей-Талас жазба ескерткіштері деп аталып жүр. Ол

ескерткіштер көлемі жағынан әрқилы: кейбіреулері бірлі-жарым сөз тіркестері мен сөйлемдерден тұрса (Енисей, кейбірі Талас ескерткіштері), кейбірі үлкен көлемді текстер (Орхон ескерткіштері) болып келеді [2.89]. Орхон-Енисей ескерткіштерін шетелдік ғалымдармен қоса қазақ түркітанушы ғалымдары М.Томанов, Ғ.Айдаров, С.Аманжолов, Ә.Марғұлан, Ә.Қоңыратбаев, Ә.Құрышжанов т.б ғалымдар түркі ескерткіштерінің түркітілдес халықтардың тілінің тарихында алатын орны жайлы сөз қозғаған болатын. Сонымен қатар көне ескерткіштердің тіліне қатысты көлемді пікірталастардың орын алуы.

1. Диалектілік ерекшеліктермен әдеби тілдің арақатынасын нақтылауда қайшылықты көзқарастардың орын алуы.

2. Жазумен байланысты дамуы мен соған орай анықтамасының бірізді болмауының себебін нақтылау,

3. Қазақ әдеби тілін дәуірлеуге байланысты ғалымдар көзқарастардың арасындағы алшақтықтардың болуы[3.9].

Орхон-Енисей жазбалары сан ғасырлық мәдениеті бар түркі халықтарының тарихы тұрғысынан алғанда баға жетпес дүние. VII-VIII ғасырларда жазылған, Орхон мен Енисей өзендерінің бойынан табылған көне ескерткіштерді табылған мекеніне қарай ғалымдар топқа бөледі.

Ғалымдар көне түркі жазбаларын үш топқа бөледі[4.9].

1.Енисей ескерткіштері.

2.Талас ескерткіштері.

3.Орхон ескерткіштері.

1. Енисей ескерткіштері

Енисей ескерткіштері қатарына: Енисей өзені, Тува автономиялы республикасы, Хакас автономиялы облысы мен Краснояр өлкесі территориясынан табылған бірсыпыра үлкенді-кішілі ескерткіштер жатады. 1887-1888 жылдары Й.Аспелин бастаған фин ғалымдары Енисей өзені аңғарына барып, өз жеріндегі ескерткіштерді зерттеуден оралды. Бұл екі экспедицияның материалдары 1889 жылы «Енисей жазбалары» атты жеке еңбек болып жарияланды. Осыдан кейін 1891 жылы В.Радлов басқарған экспедициялар да Енисей өзендерін аралап, ескерткіштерді өз көздерімен көріп, оқуға қайтады. Экспедиция материалдары да жеке еңбек болып жарияланды.

2. Талас ескерткіштері

Талас жазу ескерткіштері қатарына Талас өзені бойынан табылған 12 ескерткіш жатады. Талас өзені бойындағы Айыртаң-Ой дейтін жерден (қазіргі Жамбыл облысы) 1896 жылы табылған еді. Олардың ең үлкені Уйбат өзенінің бойынан табылған. И. А. Батманов Талас ескерткіштерін арнайы зерттеп, өз алдына кітапша етіп шығарды. Бұл ескерткіштер зерттеушілердің топшылауы бойынша Талас жазба ескерткіштері VIII ғасырды қамтиды. Талас ескерткіштерінің басқа Орхон, Енисей ескерткіштеріне қарағанда, өзінің

кейбір графикалық ерекшеліктері бар. Көне кезеңнің тілдік ерекшеліктерін белгілейтін мұралар көпшілік әдебиетте Орхон-Енисей-Талас жазба ескерткіштері деп аталып жүр. Ескерткіштер көлемі жағынан әрқилы: кейбіреулері бірлі-жарым сөз тіркестері мен сөйлемдерден тұрса (Енисей, кейбірі Талас ескерткіштері), кейбірі үлкен көлемді мәтіндер (Орхон ескерткіштері) болып келеді. Бұдан, әрине, Орхон-Енисей-Талас жазба ескерткіштерінен бұрын дүниеге келген мұралар болмады деген пікір тұмайды. Бұл жерде көне түркі тілінің бізге жеткен жазба үлгілері сөз болып отыр.

3. Орхон ескерткіштері

Моңғолиядан табылған жазба ескерткіштер әдетте «Орхон ескерткіштері» деп аталады. Мұның себебі бар: Орхон өзені бойындағы аймақ Еуразия өркениетінің орталығы болған: ұлы ғұндардың (қытайша шаньюй) этносаяси бірлестігі (б.з.б. 209 – б.з. II ғ.), одан кейін сяньбилер (II-IV ғ.), Тоба, Жужан (IV-V ғғ.), Түрік империясы (VI-IX ғғ.), «Ұйғырлардың» (IX ғ.), «Қырғыздардың» (IX-X ғғ.) және Моңғол империясының (1189-XIV ғ.) этносаяси мемлекеттерінде табылған. Сөйтіп, бұл жер мыңдаған жылдар бойы көшпелілердің табиғи болмысының сарқылмас қайнар бұлағына айналды. Орхон маңынан көптеген тарихи ескерткіштер табылды. Орхон жазба ескерткіштері Алтай, Ханғай, Соен, Кентай тауларында, Гоби шөлінде және Орхон, Селенга, Туыл, Онгин, Керулен, Онон, Хануй, Хунуй, Тес, Қобдо өзендері мен Хупсгүл, Қырғыз, Қарасу көлдерінен табылған. Моңғолия жеріндегі көне түріктердің жазба ескерткіштері шығыстан батысқа, солтүстіктен оңтүстікке, барлық жерде кездеседі. Орхон жазба ескерткіштерінің ерекшелігі – олардың үлкен мемориалдық кешендерге жазылғандығы. Мұндай мемориалдық кешендерде стелалардан басқа тас мүсіндер, тасбақа тұрақтары, балбалдар, қорғандар, тағы басқа құрылыстар (саны 10 – 400-ден асады) кездеседі[5]. «Бітігші» шебері көне түркі жазуын арнайы жылтыратылған тастарға немесе зат, теңге және тастарға қолданған. Кейбір жартастарда боялған жазулар табылды. Моңғолия жерінен ағаш немесе қағаз бетіндегі жазулар (Турфан сияқты) әлі кездескен жоқ. Ордубалық, Байбалық, Тоғыбалық, т.б жүздеген қалалар мен елді мекендер белгілі болғанымен, оларда күрделі ғылыми-археологиялық жұмыстар жүргізілмеген.

Ескерткіштердің табылған уақыты, орналасқан жері болмаса, жалпы жазба таңбаларды пайдалану арасында айтарлықтай айырмашылық жоқ. Алайда жазылған уақытқа назар салсақ, Орхон, Енисей, Талас өзендерінің бойынан табылғанымен, жазу үлгісі, таңбалардың қолданылуы ұқсас, оқылуы бірдей. Мысалы бұл ескерткіштерде **тіл** және **таңдай** және **ерін дауыстыларының** үйлесімділігіне бағынады. Көрсетілген аффикстер дауысты дыбыстардан кейін қолданылған. Мысалы, септік жалғауы -



дауыссыз дыбыстардан кейін О формасы тандай және ерін дауыстыларының әрекетіне бағынатын жалғаулық дауысты дыбыс.

Түркі жазба әдебиетінің ескерткіштерінің ішінде көбірек зерттелгендері— «Күлтегін» және «Тоныкөк» ескерткіштері. Бұлар нағыз тарихи дастандар (шежіре жазбалар). Түрік қағанаты тұсында жазылған және көптеген тарихи оқиғаларды сипаттайтын шежіре. Бұл ескерткіштерде қазіргі түркі тілдерінде, әсіресе, қазақ тілінде басым сөздер, сөз тіркестері, жер-су атаулары, адам есімдері өте жиі кездеседі. Қазақ тіліндегі сөз тіркестері мен тіл құрылымының тарихы көне түркі дәуірінің тарихынан нәр алады, бастау алады. Орхон-Енисей жазбалары XVII-XVIII ғасырдың басында табылғанымен, бұл жазбалар әлі күнге дейін ғылыми ортаның назарында. Орхон-Енисей жазбаларының жетекші тілі, қазіргі түркі тілдерімен салыстыру, ғалымдардың жазу жанры туралы пікірталастары - ескерткіштердің ғылыми жаңалығын көрсету. Түріктердің ескерткішіндей әдемі Орхон-Енисей жазбалары тек халық үшін ғана емес, ғылыми зерттеулер, сонымен қатар жоғары оқу орындарында оқу үдерісінде белсенді түрде қолданылады, сонымен қатар олардың оқулықтар мен ғылыми оқу құралдарында қамтылған. Мысалы «Орхон-Енисей ескерткіштері», «Орхон-Енисей ескерткіштері», «Орхон жазбалары», «Көне түркі жазбалары», «көне түркі жазбалары», «көне түркі руникалық жазбалары» т.б. Орхон-Енисей жазбаларының жанрлық сипаты туралы пікірлердің палитрасының болуы. Ғалым С.Е. Малов Енисей жазбаларын «эпитафтар» немесе қазынаны «бищенская» поэзиясы деп атайды», В. Жирмунский Орхон ескерткіштерін Ф.Корш, А.Абид Орхон-Енисей ескерткіштерін «өлеңдегі прозой» деп атаған. Ғалым Т.Гаджиев жазуларға «көркем және деректі қаһармандық эпос» деген анықтама береді. Ғалым В.Осман оларды «тас шежіре» деп атайды. И.В. Стеблева, Орхон-Енисей жазбаларының мәтіндерін поэзияның үлгісі ретінде, сондай-ақ олардың батырлық эпостарды бейнелейтінін таңдауға көп көңіл бөледі. Орхон-Енисей жазбаларының мәтіндерін ашудың сенімділігін сақтай отырып, негізгі орыс мәтіндерінен басқа қытай тілінде қайталанатын жазу да болды. Орхон-Енисей жазбаларының тек түркілерге тиесілілігін анықтау үшін ғана емес, сонымен қатар көне түркі жазба кескіндемесінде Орхон жазбалары түркі әдебиетіне көшкен алашқы әдеби шығарма, тарихи факт.

Көне түркі жазба ескерткіштері тілінің грамматикалық жүйесі морфологиялық бірліктерден басқа тілдің барлық грамматикалық бірліктерін және грамматикалық категорияларда берілген грамматикалық формаларды қамтиды; 2) грамматикада – тілдің құрылымын зерттейтін тіл білімінің бөлімі – тілдік бірліктер өздерінің нақты (мағыналы) мағынасынан абстракцияланып, жалпыланады; 3) Грамматика термині кейде грамматикалық категорияларды (етістік грамматикасы, үстеу грамматикасы) белгілеу үшін қолданылады. Тілдің грамматикалық бірліктері және бұл бірліктердің басы. Осылайша, тілінде көптеген атаулар өздерінің формасын өзгертпей-ақ мазмұндық

мағыналарды бере алады. Сонымен бірге, Орхон Енесей тілінде зат есімдерден сын есім жасайтын бірнеше аффикстер қолданылады, яғни. заттық мағынадағы атаулардан. Зат есімнен, -ла аффикстері арқылы жасалған сын есімдер өнімсіз түрінің жасалуы. Мысалы: Сын есімдер сын есімдерден рақ аффиксі арқылы жасалып, атрибуттың мәнді білдіреді: татаглаграқ- тәттірек.

Тонькөк жырына назар аударсақ, ең бірінші байқайтынымыз сингармонизм заңының сақталуы. Қазақ тіліндегі үндестік заңының сонау түркі заманынан болғаны. Сингармонизм (грекше *syn* – бірге және *harmonia* – үндестік) – морфологиялық бірлік ретінде сөздің біркелкі дауысты (кейде дауыссыз) жасалуынан тұратын морфологиялық құбылыс. Сөздерде дауысты және дауыссыз дыбыстардың үндестігі болуы керек. Сингармонизмі бар тілдерде сөздің құрылымы дербес (басым) компонентпен (әдетте түбір) және тәуелді компоненттермен (аффикстер) ерекшеленеді. Яғни жуан дауысты түбірге жуан дауысты қосымшаның жалғануын айтамыз. Мысал ретінде: Бек+лер(беклер), аш+сыз(тамақсыз), тоқ+сық(тоқтық), қағанды+ың(қағандық), йағул+са(жаңылса),

Көне түркі ескерткіштерінде кездесетін септіктерге тоқталсақ, олардың саны алтау[6,56].

1. Yalın hali (атау септік)
2. Belirtme hali (табыс септік)
3. Yönelme hali (барыс септік)
4. Bulunma hali (жатыс септік)
5. Çıkış, ayrılma hali (шығыс септік)

6. Көмектес септік

Атау септік

Көне түркі жазба ескерткіштерінде атау септік заттық мағына түрінде берілген. Ешбір жалғаусыз, қосымшасыз келетін атау септігі, жер-су, жануа, зат және құбылыс атауларын білдірген. Көне ескерткіштердензаттық мағынадағы атау септікке мысал:

Алтун – алтын, барым – дүние, бай – бай, баш – бас, сүнүг – сүйек, Айғучысы – кеңесші, Барқ – үй, Бітіг – хат, жазу, Едгу – жақсылық, үгүз-өзен т.б.

Матаса байланысқан тіркестердің сыңарлары әртүрлі сөздер болып келумен қатар, құрамындағы сыңарлардың мағыналарына қарай бір-бірімен қатынастары арқылы да ерекшеленеді. Мәселен, олардың кейбірінде меншіктік, кейбірінде сапалық мағына басымдау екені байқалады. Бұл мағыналар көбіне атау септігі формаларының қолданылып-қолданылмауына байланысты өзгеріп отырса да, кейбір мысалларда атау септігі формаларының қолданылып қолданылмауына мағынаға аса әсер өте қоймайды.

Барыс септігі

Көне жазба ескерткіштер тілінде барыс жалғаулы сөз тіркестері ұшырасады. Ескерткіштерде кездесетін барыс септігі сөздері көбінесе

етістіктермен тікелей де, жанай да байланысып, қимылдың бағытын, бет алысын, мақсатын, мезгілін, білдіреді [3.3]. Көне түрік тілінен бастап түркі тілдерінің түрлі кезендерінде кең көлемде қолданылған ілік септігінің -ның, -нің формасы, түркі тілдерінің ішінде қазіргі саха (якут) тілінен өзге барлық тілдерде сақталған деуге болады. Ал түркі тілінде ілік септігінің формасының орнына монғол тіліндегідей -та, -те қолданылады. Олар барыс септігінің -ғару жалғауы арқылы байланысқан тіркестер. Бірақ ескерткіште бұл тіркестер сирек кездеседі және қолданыста көп қолданыла бермейді. Мысалы: Апа тарқанғару бере саб ыдмыс. Табғачқа йана ічікді. Қағанғару ол сабығ ыттым. Болчуқа таң үнтүрү тегдіміз. Неке тезер біз.

а) -қа, -ке: йылқықа жылқыға, іске – іске, тынлығқа - тіршілікке қанқа қанға, үйме – үйме, улуғқа - ұлыққа ие.

б). -ғару, -геру, -ғары, -гері: біргеру - бері қарай йергеру - жерге қарай т.б.

в) -ра, -ре: Өңре - алға, ташра - сыртқа, далаға.

г) -а, -е; жіктік есімдіктері мен тәуелдік жалғаулы сөздеінен кейін соң жалғанады: құтыңа – құтыңа, кучүне күшіне, башыңа басыңа, йеріңе – жеріңе т.б. Көне ескерткіштердегі барыс септікте келетін сөздердің мағыналары әр түрлі. Олар бет алысты, бағытты, мезгілді мекенді және көрсетеді. Сөйлем ішінде не толықтауыш, не пысықтауыш қызметін атқарады.

Табыс септік

Табыс жалғаулы сөз сабақты етістікпен тығыз байланысты болады. Ескерткіштер тілінде табыс жалғаулы сөз тіркестерінің бірнеше түрінің модельдері кездеседі: а) -ғ, -г, -ығ, -іг, -уғ, -уг тұлғалы жалғаулы сөз тіркестері. Мәселен: Күлтегініг аз ерін іртүрү ітіміз. Бү сүг елт. Қарығ сөкдім. Қырқ азығ уда басдымыз. Алтун йышығ аса келтіміз, Ертіс үгүзнүг кече келтіміз. ә) -н, -ын, -ін, -ун, -нн тұлғалы сөз тіркестері. Мысалы: Ақ адғырығ удлықын сыйу, урта. Чуғай қузын, Қара қумығ олурур ертіміз. б) -ны, -ні, -ы, -і тұлғалы сөз тіркестері. Мысал ретінде мына тіркестерді негізге аламыз:

Егер біз армиямызбен қарсы шықпасақ, олар бізді құртады – Аңару сүмшесер қач пен ерсер ол бізні өлүртечі. «Мен оғыз халқы өлтіреді», - дедім мен – Біні оғузығ+өлүртечі тірмен.

Жатыс септігі

Орхон-Енисей ескерткіштеріндегі жатыс септігінің берілу жолдары.

Б. А. Серебренников -да тұлғасының бойында болған лока тив мәндерінің морфологиялық тәсіл тұрғысынан жіктелуін да негізінде жаңа жалғаудың қалыптасуымен түсіндіреді -дан жалғауының қалыптасуымен бірге сын мәні соған ауысты да, локатив мәні -да тұлғасында орнығып қалды [5.243]. Мысалы:

-де, -та, -те түрінде қолданылады: ата-да//ада-да - әке; таш-та//таш-да-тастау; сен де- сенде; бұл арада - тап сол кезде; бун-да – бұл жерде;

Орхон-Енисей ескерткіштерінде жатыс септігінің берілу жолдары: Жатыс септік. Қазіргі түркі тілдеріндегі жатыс жалғауының варианттарын салыстыра қарағанда, дауысты дыбыстар ыңға йынан болсын (кейбірінде бұлардың еріндік варианттары бар), дауыссыз дыбыстар жағынан болсын, негізінен -да тұлғасының айналасына топтасады. Әрине, басқы дауыссыз кейбір тілдерде тек-д (азербайжан, түркмен) болып келетіні де бар. Қалай болғанда да жатыс жалғауы сыртқы тұлғалық жағынан бізге белгілі ежелгі дәуірден біздің тілімізге дейін көп өзгеріссіз жеткенін айту керек. Тіл дамуының барысында бұл септік жалғауына қатысты өзгерістер қолданылу мағыналары мен етістіктерді меңгеру жүйесінен ғана байқалады.

Іс-әрекеттің, заттың орын-мекенін білдіреді: Түрк сір будун Перінте бод қалмады (Түрк сір халқының жерінде тірі жан қалмады). Уда ташла қалмысы құбранып, йеті бүз болты (Ағаш, таста қалғандары жиналып жеті жүз болды). Іс-қимылдың, әрекеттің уақыт, мезгілін білдіреді: Кул тегін ол сүнүшде отуз дашайур ерті (Ол соғыста Күлтегін отыз жасар еді) т.б.

Шығыс септігі

Шығыс септік. Шығыс септік тұлғасы қазіргі түркі тілдерін де әр түрлі варианттарда кездеседі: бірсыпыра тілдерде тек д дыбысынан ғана басталса, кейбірінде тек қана қатаң г дыбысынан басталады. Ал, қазақ тілі мен тағы да кейбір тілдерде т, д, н дыбыстарынан басталған, дауыстылардың ыңғайына қарай алты түрлі вариантта ұшырасады. Түркі тілдерінің өте ескі кезеңінде шығыс мәнін беретін арнаулы қосымша болды ма дейтін мәселеде бірсыпыра зерттеушілер бұл қызметті сөйлем құрамында жатыс септік атқарған деп қарайды, Мұндай көзқарасқа дәлел болатын Күлтегін мен Тонықұқ ескерткіштерінің тілі. Бұлардың екеуінде де шығыс септіктің арнаулы тұлғасы өте сирек ұшырайды. Ал, оның қызметін, мәнін, осыдан бұрын айтылғандай, көбінесе жатыс жалғау атқарған. Осыған сүйене отырып Радлов та, кейін сол жазу ескерткіштерін жинап, грамматикалық комментарий берген.) С.Малов та ол дәуірде шығыс септіктің арнаулы қосымшасы бар деп айтпайды) Дегенмен, мына сөйлемдер құрамындағы бір-екі мысалға тоқталайық: Өңден қағанғару су йорыдым. (Оң жақтан хандыққа қарай әскер аттанды). Парақлығ қантан келіп йаңа елтді? Сүнгліг қантан келіпен сүре елтді? (Жарақтылар қайдан келіп толқындатты? Сүнгілілер қайдан келіп ерітті?) Осындағы өңден (оңынан), қантан сөздері шығыс тұлғалы екені, мекендік (іс-әрекеттің шығар көзі) мәнде қолданылғаны айқын. Ұйғыр жазбаларында да осы жалғаудың қолданылғаны бар: Іл-Теміртін бір күрі күнчіт алтым . (Іл-Темірден бір өлшем күншіт алдым). м). Сенден қамар сундылач, менде тынар қарғылач (Ш). (Сенен қа- шар сандуғаш, менде тынар) тыным алар қарлығаш. Бурунғы кечкен Саин хан чағыдин (Бұрынғы өлген Сайын хан шағынан).

Шығыс септік қосымшаларын ескерткіштер бойынша салыстырып қарағанда мына жайт байқалады: көне түркі жазбаларында қысаң дауысты варианттар да (-тын, тік), ашық дауысты варианттар да (-тан, -тен-ден) кездеседі) Қашқари сөздігінде солай. Сірә, бұл Қашқари сөздігінде сол кездегі түркі тайпалары тілдерінің барлығының материалы берілуіне байланысты болса керек. Орхон-Енисей жазбаларында жатыс септік тұлғасы шығыс мәнінде де қолданылған. Шығыс септік қызметіндегі жатыс септік тұлғасы іс-қимылдың шығар шешімін, мекенін білдіреді: Қыр қызда йактымыз. Түргес қағанта көруг келті т.б.

Біздің байқауымызша, шығыс септік ол дәуірде қалыптаса бастаса керек. Өйткені, шығыс септіктің орнына көп жерде жатыс септіктің жұмсалатыны байқалады. Мысалы: Қырқ азда йантымыз. Түргеш қағанта көруг келті.

Көмектес септік

Көмектес септік. Көне түркі ескерткіштерінде кездескенімен, қазіргі түркі тілдерінің барлығында бірдей кездеспейтін септік жалғауы осы көмектес септік. Зерттеуші ғалымдар көмектес септіктің мәніне жақын септік жалғауының якут, чуваш тілінде барлығын айтады. Дегенмен, арнаулы грамматикалық тұлғасы бар көмектес септік қазақ тілінің ғана ерекшелігі. Басқа түркі тілдерінде бұл септік жалғаудың қызметін біле (білен) шылауы атқарады. Көне түркі тілінде қазақ тіліндегі -мен септік жалғауы мен басқа түркі тілдеріндегі біле шылауы

алмастырады. Көмектес септіктің жалғауы көне ескерткіштердің арасында - ын,- ін,- н. Ол сөзге қимыл-әрекеттің белгілі бір құралын, амалын білдіреді. Мысалы, елігін күче колынмен күште; адақын келті аяғымен келді, кулғақын ешіделім құлақпен естиік. Мысалы: Тонью - Кук бірле келті— Тонью-Кукпен бірге келді. Бұл айтылғандардан байқап отырғанымыз, қазіргі көмектес септіктің мен, бен, пен жалғаулары қазіргі түркі жазба қолданысында кездеспейді. Алайда, бұл тұлға ескерткіштер тілінде бірде шылау сөзімен жарыса қолданылған. Сын, (-ін, -н) тұлғасы мынадай негізгі мағыналар туғызған: іс-әрекеттің болып өткен жолын, орнын білдіреді: Ол йолын йорысар, унчтідім (Ол жолмен жүрсе, мүмкін дедім). Іс- қимылдың болған, болатын мекенін білдіреді: Алтун йысыр Подсузын астым, Ертіс үгүзүг кечіесізін кечдіміз (Алтын жерін жолсызбен астым, Ертіс өзенін кешпес ізбен кештім). Іс- әрекеттің болу құралын объектісін білдіреді: інгек көлүкін Тоғ лада Оғуз келті (Жүк тиелген көлікпен оғыздар Тоғлада келді. Тағыг укруқын егмес, теңізні қайрықын бүкмес. (Тауды арқанмен имес, теңізді қайықпен бекітпес). Біргелік, ортақтастықты білдіреді: Кул тегін аз ерін іртүрү ытымыз (Кул тегінді аз кісімен ілгері жібердік).

Көмектес септігінің қосымшасы қатар да сөз тіркестері жасалған. Олар тексте айтарлықтай көп емес: Інгек көлүкін Тоғлада оғуз келті – Сиыр көлігімен Тоғлада оғыз келді. Сүңүгін ачдымым – Сүңгімен аштық т.б.

Сөз таптары

Орхон-Енисей ескерткіштерінде кездескен сөз таптары: зат есім, сын есім, есімдік, етістік, үстеу, жалғаулық, құпия сөздер, сөз тудырғыш және сөз түрлендірушілер. Сөз құрамы мынадай түрлерге бөлінеді:

1. Сөз құрамы мынадай түрлерге бөлінеді:

а) түбір сөздер: будун(халық), біліг(білік), емгек(еңбек), бедіз (өрнек), баға(бақа), йарық(жарық), алку (бөрі), утлы(сыйлық), йонт(жылқы)

ә) туынды сөздер: істе(істе), етікчі – қарыздар, білігсіз – біліксіз, йолчы жолшы, бөлүг – бөлік, күчлүг – күшті, батсық – батыс, қорқыныш - қорқыныш

в) күрделі сөздер: іс-күч(күш-жігер), иіні-йіген(іні-жиен), біс түмен(елу мың),

Иал Тұрған - Иалтурған (кісі аты) Іл Темір - Елтемір (кісі аты), бу күн – бүгін т.б. сөздерді айта аламыз.

Орхон-Енисей ескерткіштерінде кездескен сөз таптары: зат есім, сын есім, есімдік, етістік, үстеу, жалғаулық, құпия сөздер, сөз тудырғыш және сөз түрлендірушілер. Зат есімдердің грамматикалық категориялары бар: көптік,, қатыстылық, предикативтілік. Олардың көптеген лексикалық және семантикалық категориялары бар. Олардың ең көп саны мыналар: 1) жалқы есім және жалпы есімдер; 2) адамзат және ғаламзат есім; 3) деректі және дерексіз зат есімдер[6].

Зат есім

Заттың, заттың, құбылыстың атаулары зат есім. Семантикасы мен мағынасына қарай зат есімдер екіге бөлінеді. Олар: нақты және абстракті (деректі және дерексіз).

Зат есімдердің нақты (деректі) зат есімдерге қатысты: оғул– ұл, уд – сиыр, лашын – лашын, йонт – жылқы, куш- құс сөздерін мысалға аламыз.

Адамның ойлау қабілеті арқылы ғана біле алатын абстрактілі(дерексіз) зат есімдер. Олар көзге көрінбейді тек заттық ұғымды білдіреді. Мысалға:

Алқыш (алғыс), аңыз (аңыз), атағ, (атақ), иыл (жыл), обут (ұят), ыдук (қасиетті),

Іді (ие,қожа), бұнтуг (қамқоршы), біліг (білік) т.б.

Зат есімдер өзінің морфологиялық жағынан септеледі, тәуелденеді, көптеледі, кейде жіктеледі. Соның ішінде, көптік жалғауына тоқсалсақ, Тоныкөк жырынан көптік мағынаның 3 түрлі тәсілмен берілетінін байқауға болады. Бірінші, морфологиялық тәсілмен: зат есімге көптік жалғауы жалғануы арқылы жасалады. Мысалы: Беглер - бектер, қадашлар - жолдастар, ерлер – ерлер, күдегүлер – күйеулер т.б.

Екіншісі, лексикалық әдіс. Ескерткіштер тілінде көптік лексикалық категория екі түрде жүзеге асады: алдымен көптік мағына лексикадағы жеке сөздердің мағына жағынан байланысуы болады және көптік мағына сөз тіркестері арқылы да беріледі. Орхон-Енисей ескерткіштері тілінде

төмендегідей сөздер көптік жалғауы жалғанбай-ақ көптік ұғымды білдіреді. Мысалы: белгілі бір елдің, тайпаның, халықтың аттарын білдіретін кейбір жинақты ұғымы бар сөздер: түрк будун (түркі халқы), оғуз (оғыз), қыбчак (қыпшақ), көз (көз), қулғақ (құлақ), адағ (аяқ), алтун (алтын), сач (шаш), күмүш (күміс), інгек (сиыр) т.б. кездеседі.

Үшінші, синтаксистік тәсіл:сөздердің қосарлануы, сөз алдына көптік немесе сандық ұғым қосылуы арқылы жасалды. Көне ескерткіштерден мысал келтірсек, йеті йүз кісі (жеті жүз кісі), он түмен су (он мың әскер), коп соғды (көп соғды), үкүл ер (бүкіл ер), қыз-қудуз (қыз-қырғын) т.б. сөздер ешбір көптік жалғау жалғанбай-ақ көптік мәнді білдіреді.

-шы түп түбір арқылы өрнектелген затқа қатысты тұлғаларды білдіретін зат есімдер, олардың терістеушілік қасиеті төл түбір мағынасына байланысты: аңшы аңшы, қайықшы аңшы, теңізші ,балықшы, сабшы «хабаршы», таңбалы сақтаушы, елші билеушісі, иерчи «жолбасшы, танричи дін қызметкері», эочи «әйелі» (үй шаруасымен айналысатын), номшы діни тәлімгер», гоатигчи «саудагер, егінші», талдықшы «олуга», блучи «өлтіруші», өлумчи «жанкешті», (өлімге кесілген адам), Йонакчи жалашы»[7,88-93]. т.б.

-дық-дук мен мағынасын білдіретін зат есімдер жасайды. Мұнда бастапқы негіз арқылы белгіленетін заттар және сын есімдерден дерексіз мағыналы зат есімдер бар: ағылық қазына, ізмишляк бағы, чәчәктік гүл бағы, кинау орны, 'түрме', қазлұқ қаз көп тұратын жер (көлдің аты).

-ку, -гу дерексіз мағыналы зат есімдер жасайды:

сын есімдерден: инчку тыныштық, есен саулық (есенгу). Сызықша есімді білдіретін зат есімдер қатысы бар тұлғалардың: қадаш туысы.» -қақ-ақ кішірейткішпен зат есім жасайды. Кішірейткіш мағынасы: еркек адам» (тарихи түрде бұл кішірейткіш мағынаға ие болған).Сөз -сук зат есімнен белгілі бір мағыналы зат есім жасайды: бегарсук ішегі (бағыр «бауыр»).

Әдетте, түркі тілдеріндегі сын есім тудыратын қосымшалар аффикстері болып табылады, олардың қолданылуы осы аффикстердің көмегімен жасалған формаларда туынды және флексияның өзара тығыз байланыста болатынын көрсететін бірқатар белгілерге ие.

-дың аффиксі бастапқы негіз арқылы білдірілгеннің иелігін білдіретін атрибутивтік формаларды жасайды. Түбір белгілі бір затты білдіретін зат есім болса, онда -лығ аффиксі сапалық сын есімдер жасайды: күш-қуатты, атлығ даңқты, татағлағ-тәтті формалар «-лығ» аффиксі арқылы жасалған өзендер аффиксімен тіркесіп, артықтығын білдіре алады.

Сын есім

Көне жазба ескерткіштердегі сын есім-қосымшалардың айтарлықтай тобы бар. Сын есім тудыратын қосымшалардың көпшілігі қазіргі басқа түркі тілдерінде де, ескі жазбаларда да кездеседі. Сын есім жасаушы қосымшаларды былай қойғанда, қосымшалардың басым көпшілігі бұрынғы семантикасын,

дыбыстық формасын өзгеріссіз сақтап қалған. Сын есім-білімдік қосымшалардың мағыналық дамуы мен тұлғалық өзгерістерін анық көру үшін оларды тілдің ескі мұраларымен салыстырған жөн. Сын есім жасаушы қосымшалар:

Зат есім түбірлерінен сын есім жасайтын affикстер: -лығ, ліг (-лық, -лік). Көне түркі тілінде зат есімді де, сын есімді де жасайтын жұрнақ. Дегенмен, -лық пен -лық қосымшаларының арасында айырмашылық бар, -лық тек зат есім жұрнағы, ал -лығ жұрнағы сын есім жасауға қатысады. Мысалы: туғлығ (тулы), атлығ (атты), йарақлығ (жарақты), чөлліг (шөлдік), йумғылығ (жинақты) сөздерін айтуға болады.

Келесі, -қы, -кі, -ғы, -гі affикстері арқылы жасалған сын есім. Орхон-Енисей жазбаларынан морфологиялық тұлғаны ескере отырып, сын есімдердің мезгілдік және көлемдік-кеңістіктік мағыналы сөздерге айналатынын білеміз. Яғни, бұл affикстер мезгіл, мекен мәнді сын есім жасайтын қатыстық сын есім жасайды.

Ескі түркілік -қы affиксі мезгіл мәнді мынадай сын есімдер құрамынан ұшырасады: жазғы, күзгі, қысқы, бүгінгі, ертеңгі, былтырғы, кешегі, күндізгі, таңғы, алғашқы, ішкі, тысқы, сыртқы, астыңғы, үстіңгі, төргі, ауызғы (үй) т. б. Ескі түркілік -қы affиксі мезгіл мәнде мынадай сын есімдерден мына күйде ұшырасады: жазғы, күзгі, қысқы, бүгінгі, ертеңгі, былтырғы, кешегі, күндізгі, таңғы, алғашқы, ішкі, тысқы, сыртқы, астыңғы, үстіңгі, төргі, ауызғы (үй) т. б. Мысалы: ашнуқы (алғашқы), ең ілкі (ең ілкі), табғачғы (табғашқы), башғу (басқы) т.б.

-сыз, -сіз жұрнағы арқылы жасалған сын есім. Орхон-Енисей жазбаларынан бастап ескерткіштер тілінде зат есім түбірден сын есім жасайтын өнімді қосымша. Білігсіз қаған (білімсіз хан), түзсіз (дұрыстықсыз), ашсыз (ассыз), түтінсіз (түтінсіз), башсыз (бассыз), айпсыз (айыпсыз), аңсыз (ойсыз, сезімсіз), атасыз (атасыз), бұрынсыз (мұрынсыз), көрүксіз (көріксіз). -сыз affиксі өзінің шығу тегі жағынан -сығ affиксімен және монғолдық -сар -сер affиксімен ұялас. р-з сәйкестігі жалпы алтай тілдері арасында ғана емес, түркі тілдерінің өз арасында да бар. Көне түркі (Орхон-Енисей) ескерткіштерінде кездесетін сыра (қаған сыра - қансыз қалу) қосымшасы осы ұядан тараған. Н. А. Баскаковтың зерттеулеріне қарағанда, -сыз affиксі құранды, екі түрлі affикстің біріккен түрі -сы және з элементтеріне бөліне -діз, -з болымсыздықты білдіретін арнаулы қосымша да, -сы бір кезде актив субъектіні білдіретін қосымша» деген пікірді ескерген жөн[6.187].

-дақы, -дікі, -тақы, -тікі жұрнағы арқылы жасалған сын есімдер: көңүлдекі – көңілдегі, булундақы төрт бұрыштағы, алықдақы – қаладағы, йердекі – жердегі, өңрекі – алдағы, чірекі — ішкі т.б.

Сын есімдер сын есімдерден -рак affиксі арқылы жасалып, атрибуттың артықтығын білдіреді: бая рек жоғары, татаглаграк тәтті. Қарқынды формалар пликацияны ішінара азайту арқылы қалыптасады: қалпақ толығымен қара».

бастапқы негізбен өрнектелетін нәрсеге ие болу тілінде аффикс -лығ жиі қатыстық байланыс көрсеткіші қызметін атқарады және ол тіркесетін зат есімнің синтаксистік қызметін ғана көрсетеді: Оқунмәклиг әрт өкініш оты.

-она аффиксі төл түбір арқылы білдірілгеннің жоқтығын білдіретін формалар жасайды: Извукоуз тынлығлар-Күнәсіз тіршілік иелері. Бу өгсүз этузда- мынау ессіз денеде. Билигсіз таро қылынч-Ақылға қонымды, жалған істер.

Сын есімдер сын есімдерден -рак аффиксі арқылы жасалып, атрибуттың артықтығын білдіреді: татаглаграк- тәттірек.

Шырай категориясы

Көне түркі тілінде де шырай категориясы бар екенін және бұл категория сын есімнің ерекше белгісі екенін тіл ғалымдарының ешқайсысы жоққа шығармайды. Дегенмен, сапалық немесе қатыстық сын есім қандай дәрежеде көрінетінін білу барысында жіктеу мәселесінде ортақ негіз болар пікір жоқ. Көне ескерткіштерде сын есімнің шырай мәнінің берілу жолы және сын есімдерге -рак, -лау аффикстері жалғану арқылы немесе сын есімдердің алдынан тіркесе айтылатын арнаулы сөздер арқылы берілетіні мәлім. Тарихи тұрғыдан бұл тәсілдердің қайсысы да жаңа құбылыс емес, ертеден орнығып қалған жүйе. Бұған көне түркі ескерткіштерінің материалдарын мысалға келтіруге болар еді. Көне ұйғыр жазба ескерткіштері тілінде төмендегідей үш түрлі шырайлар кездеседі. Жай шырай: узун – ұзын, қара қара, пашыл жасыл

Салыстырмалы шырай: бегрек- бегірек, артуқрак - артығырак

Күшейтпелі шырай:, ең кішігі- ең кішісі, ең ұлуғы - ен үлкені, едгу – игі, көк көк, кунтег – күндей т.б.

Сын есімдер әдетте сөйлем ішінде анықтауыш болады. Сын есімдер көне ескерткіштерде де түрленіп келіп сөйлем ішінде түрлі синтаксистік қызметтер атқарған. Кейде сын есімдер етістіктен бұрын болған мүшеден бұрын келсе, пысықтауыш та болады. Сын есімдер сөйлем соңында көмекші етістіктермен тіркесіп, баяндауыш та болады. Заттанғанып келіп, сын есімдер сөйлем ішінде бастауыш та, толықтауыш та бола алады.

Сан есім

Жалпы сан есім — белгілі бір заттың санын, мөлшерін, ретін, шамасын білдіретін сөз табы. Ескерткіштерде заттың санын, ретін білдіретін сан есімдер, де кездеседі. Сан есімнің өзіне тән семантикалық және грамматикалық ерекшеліктері болады. Олар жеке тұрғанда абстрактілік ұғымды білдіреді. Бірақ олар тілде, негізінде, затты не болмаса оқиғаны сан жағынан анықтап береді. Сол кезде ғана олардың мағынасы анықтала түседі. Көне ұйғыр жазба нұсқаларында кездесетін сын есімдер мағынасына қарай екі түрлі.

1. Сапалық сын есімдер.

2. Қатыстық сын есімдер.

Сапалық сын есімдер. Ескерткіштерде ешбір аффиксиз түбір күйінде заттың түр-түсін, сыр-сипатын көрсететін сапалық сын есімдер де кездеседі. Олар төмендегідей топқа бөлінеді:

1. Заттың түр-түсін білдіретін сын есімдер:

2. Заттың көлемі мен аумағын білдіретін сын есімдер: боз - боз, сарыг-сары, йашыл- жасыл, қызыл - қызыл, көк - көк, артуқ - артық, йумчақ- жұмсақ т.б.

3. Заттың сапа, сыр-сипатын білдіретін сын есімдер: ач – аш, тіріг – тірі, арығ – таза, ысық - ыстық, ыссы, йаблақ - жаман, төмен.

Қатыстық сын есімдер. Текстерде қатыстық сын есімдер аф- акстер жалғану арқылы басқа сөз таптарынан жасалады. Сын імдер жасайтын аффикстердің бастылары төмендегідей: -лы, -лі, теңрілі – тәңрілі [7.37].

VI-VIII ғасырда өмір сүрген Орхон-Енисей түркілері мәдениетті тайпалардың бірі болған. Ескерткіштер тіліндегі сан есімдер сол дәуірдің өзінде-ақ әбден қалыптасқан сөз таптарының бірі болған. Ескерткіштер тілінде қазіргі түркі тілдерінде кездесетін сан есімдердің бәрі де кездеседі. Текстердегі сан есімдер құрамына қарай екіге бөлінеді.

а) дара сан есімдер, б) күрделі сан есімдер.

Дара сан есімдер. Бір ғана сөздерден жасалады, олардың лексикалық мағыналары әр түрлі. Мысалы: бір (бір), екі, ікі (екі), үч (үш), он (он) т. б.

Күрделі сан есімдер: екі үч быч (екі-үш мың), алты бың (алты мың), йеті иүз (жеті жүз), төрт түмен (қырық мың) т.б.

Көне ескерткіштерден мысал: Курийа он оққа кірті - Батыстағы (түркілердің) он одағына қосылды. Он оқ қағаны йағымыз ерті - он оқ қағаны жауымыз еді. Елігче ер тутдымыз - Елудей ерді тұткын еттік. Күлтегін ол сунушде отуз йашайур ерті - Күлтегін ол соғыста отыз жасаған еді. Иетміс ер болты - Жетпіс ер болды. Улуғ Кулі Чор секіз он йасап, йоқ болды.- Ұлы Күлі Чор сексен жасап, кайтыс болды. Түрк қыбчақ еліг олурмыс -Түркі қыпшақ, елу жыл мекендеді.

Енді көне Ұйғыр жазба ескерткіштерінде сан есімнің мағыналық түрлеріне қарай төмендегідей топтарға бөлінеді:

1) есептік: иеті, төрт (төрт), беш – (бес), алты, алды (алты), йеті, йіті (жеті),

Секіз (сегіз), тоқиз(тоғыз), отуз отыз т.б.

2) реттік: екінті – екінші, учуня - үшінші бісінч – бесінші, төртінч төртінші т.б.

3) жинақтық: учегу-үшеуі, екігу – екеуі, бірегу – біреуі т.б.

4) болжалды: йүзче-жүздей.

5) Бөлшекті: бірер - бір ғана, жеке; үчер-үштен; ікірер – екіден; онар – ондап т.б.

Зат есімнің негізінде сан сөйлемде пысықтауыш қызметін атқарады. Бірақ зат есімдер сөйлемнің басқа мүшелері де бола алады.

Етістік

Етістік – адамның, жануардың, белгілі бір заттың, құбылыстың іс-әрекеттің атауын білдіретін сөз табы. Көне ескерткіштер тілін қазіргі тіл материалдарымен салыстыратын болсақ, ескерткіштер тіліне сөзжасамның қазіргі әдістерінің барлығы дерлік кіреді. Өзгеріс көбінесе аффикс мағынасының басқа күрделі аффикске негіз болуына немесе аффикстің әр түрлі мағынадағы варианттарының, кейде, керісінше, шығу тегі жағынан әр түрлі қосымшалардың синонимге айналуына ғана әкеледі. Кейде зат есімнің түбіріне де, етістік түбіріне де бір ғана аффикс қосылып, туынды етістік жасайтыны байқалады. Сонымен, зат есім мен етістіктің синкретизміне негізделген қосымшалардың омонимиясы белгілі бір тарихи кезеңнің айғағы қызметін атқарады. Қазіргі қазақ тілінде –сы//си аффиксі есімдік түбірге де, етістік түбірге де жалғанады. Бірақ зат есімге қосымша мен етістікке қосымша бір деңгейде емес. Зат есімге -сы жұрнағы жалғанып туынды етістік жасайды және ондай етістік салыстырмалы мағынада жұмсалады. -сы аффиксімен қалыптасу негізгі жағынан да, мағынасы жағынан да орайлы болатын аффикс «-сыра» Орхон-Енисей жазбалары тілінде -қағансыра т. б. туынды етістіктер ұшырасады. Орхон-Енисей дәуірінде өнімділігін жоғалтқан -ын, -ын, -н (-сін, -сын), -ғары, -гері формалары қазақ тілінде сөз түбірінде жиі кездесетін *рудимент* ретінде ғана кездеседі. тілімен қатар тіл де: *бұрын, жаз, қыс, ертең, баяу, білінбейтін (келетін), жасырын, жақын, жақын, қиын, еркін, биік, озық, өйткені, кішкентай, жақын, жыртылған (болуы), алдында, астында, кейін, ішінде, сыртында*, т.б. Бұл сөздер кем дегенде мың жыл бұрын қалыптасып, жалпы тілімізге енген, өйткені V-VIII ғасырлардағы жазба тілінде де бұл үстемелер бір-екі сөзде ғана кездесіп, бүгінде бастапқы мағынасын жоғалтқан. мағынасы мен қызметі алшақтайды[9]. Қазіргі біздің тілімізде де «қансыра», «жалғызсыра» тәрізді есім туынды етістіктер жасайды. «сыра» аффиксін туынды сын есім жасайтын сыз аффиксімен жақындастыруға да негіз бар. Соңғы -а кейін жамалған, етістік жасайтын аффикс. Екінші элементке келсек, сында да, тарихи тұрғыдан керек болады. қазіргі түркі тілдері арада сәйкестігі бар екенін еске алу. Түркі тілдерінде ежелден етістік жасайтын аффикстердің бірі -ла// -ле. Бұл аффикстің қазақ тілінің заңдылығына лайық. Көне түркі ескерткіштерінен етістіктің негізгі және туынды түріне көп мысал келтіруге болады[9].

1. Негізгі етістіктер 2. Туынды етістіктер

1. Негізгі етістіктер. Текстерде негізгі етістіктер қатарына қимыл мен қозғалысты, жай-күйді және амалды білдіретін төмендегідей етістіктер жатады: *буз – бұз, біл- біл, бат жат, туш – түс, олтур -отыр, тоқы – ұр* т.б.

2. Туынды етістіктер. Туынды етістіктердің көпшілігі түбір етістіктен де, басқа сөз таптарынан да аффикстер жалғану арқылы жасалады. Текстердегі туынды етістіктер есім және етістік негізді болып екіге бөлінеді.

Ескерткіштердегі туынды етістіктер жасайтын аффикстердің негізгілері төмендегідей: 1. -а, -е, -у, :ата- ата, куче-күште; 2. -ад, -ед, бошу – боша, қутад - құтты болу, муқад – мұнай; 3. -ла, -ле: ігле- ауыр, куле – мақта; 4. сыра, сіре: күчсіре – күшсіре, татығысыра – тәттісіре т.б.

Етістіктің рай категориясы

Етістіктің рай категориясы көне ескерткіштерде де көрініс табады. Рай деп әдетте етістіктің модальділік мән беретін тұлғаларын атайды. Сондай рай деп аталатын етістік категориялары әрі грамматикалық класс, әрі семантикалық топ құрады. Арнаулы грамматикалық көрсеткіштер райлық топ құрайды. [8.101]. Жалпы рай категориясы сөйлеушің қимылдың болмысқа қатысы туралы түсінігін, көзқарасын білдіреді. Негізі рай категорияларының көне түркі ескерткіштерінде төрт түрін кездестіре аламыз.

Бұйрық рай

Бұйрық райдың негізгі берілу тәсілі етістіктерге негізделетіні бізге мәлім. Бұл тәсіл көне түркі ескерткіштер тіліне де тән. Кейде осы тұлға көпше мәнде де қолданылғаны байқалады: Тоғыз оғуз бегілері будуны бу сабымын едгүті есід, қатырды тыңла. Бұл сөйлем бір адамға емес, көпшілікке арналса да етістік түбіріне арнайы қосымша жалғанбаған. Әдетте етістіктің жалаң түбірі деп түркі тілдерінде бұйрық райдың II жақ анайы (немесе қатаң бұйрық) тұлғасы айтылады. Қазақ тіліндегі бұйрық райдың II жақ тұлғасы дегеніміз, сонымен бірге, етістіктің түбірі деп те ұғынылады.

Бірінші, бұйрық райда етістіктің негізіне -ғыл аффиксі қосылып, бұйрық райдың мағынасын күшейте алады. *Екінші*, бұйрық райда (-сун -зун) әрқашан еріндік дауысты дыбыс болады. Шартты рай -сар аффиксі арқылы жасалған. Руна және көне ұйғыр ескерткіштерінің грамматикалық құрылысындағы айырмашылықтар өте аз. Ежелгі ұйғыр ескерткіштерінде –*тек*, –*тік* жалғау аффиксі – енді дауысты түбірлерден кейін ғана емес, дауыссыз дыбыстардан кейін де формаға ие. –*ны* септік жалғауы бар, бірақ аффиксі септік жалғауымен қатар төл септіктің өзі –*дан/-дын* аффиксі де кеңінен қолданылады. Айта кетсек, бастапқы формасы руникалық ескерткіштерде (Тоныкөк ескерткіші) де кездеседі. Бұл руникалық ескерткіштерге қарағанда кеңірек қолданылады. Ол тек қана кеңейтілген анықтама қызметін ғана емес, *рай* аффикстерін қабылдай отырып, сөйлемнің басқа мүшелерінің қызметін де атқара алады. Сонымен, руникалық және көне ұйғыр ескерткіштерінің белгілі бір айырмашылығы бар, бірақ олар мардымсыз. Екі топтағы ескерткіштердің де қазіргі түркі тілдерінен ерекшеленетін ортақ белгілері көп. Руна ескерткіштерінің тілі мен дәстүрлі ескі ұйғыр деп аталатын көне ұйғыр ескерткіштерінің тілі арасындағы тығыз байланыс динамикалы, бірақ

келіспейтін тұстары да бар. Түрколог ғалым В.М.Насилов былай деп жазады: «Этногенетикалық қауымдастық пен ұзақ мерзімді тарихи жақындық ежелгі ұйғырлардың тілімен байланысты». Дәл түркі руникалық жазу негізінде көне ұйғыр тілі дамыды. Себебі, Күлтегін жырында айтылғандай түркі халқы ұйғыр халқымен тығыз қарым-қатынаста болған. Руна мәтіндерінің тілі *стандартты тіл* қызметін атқарды. Алайда ұйғыр тайпасының ықпалымен бірте-бірте өзгерді. Мысалға тілдік диалектілердегі өзгешелік пен ұқсастық. Жалпы руникалық ескерткіштер тілінің грамматикалық құрылымының негізгі белгілерін Енисей-Орхон және ұйғыр жазбаларының ескерткіштерінен алады. Бұл екі жазудың тілдері арасында негізгі айырмашылықтардың бар екендігіне жеткілікті дәлел келтірмеген, дегенмен, біз олардың әрқайсысының ерекше ерекшеліктері бар екенін жоққа шығара алмаймыз [10].

Ашық рай

Ашық рай іс-әрекетті қай шақта тұрғанын көрсетеді. Ашық рай шақ формалары арқылы берілетіні мәлім. Шақ мәнін беретін тұлғалардың үлкен бір тобы есімше мен көсемшелер. Есімше мен көсемшелер қай шақты білдіруіне қарай өз іштерінде жеке-жеке топтарға жіктеледі, сол топтар белгілі шақ жайлы ұғым берудің негізгі тәсілдері болып табылады. Арнайы тоқталуды қажет ететін шақ тұлғасы -ды (-ді, -ты, -ті). Әдетте қазіргі түркі тілдерінің грамматикаларында жедел өткен шақ деп аталатын бұл тұлға Орхон-Енисей жазбалары уақытынан бері өткен шақ мәнін берудің басты тәсілі болып келеді. Орхон-Енисей жазбалары тіліндеды формантының түбірлерге жалғануы түбірдің соңғы дыбысының сапалық қасиетіне бірде үйлессе, бірде үйлеспейді. Яғни түбірдің соңғы дыбысыз (үнді) болуына қарамай, ол бірде ұяң дауыссыздан, бірде қатаң дауыссыздан басталып отырады. Мәтінде істелетін істің субъектісі көрсетіліп, істің істелгенін, не алдағы кезде істелетіндігін білдіреді. Көне ескерткіштерде кездесетін ашық райдың үш шағы бар.

а) Осы шақ. Ашық райдың осы шағы істің дәл қазір, осы сәтте істеліп жатқандығын көрсетеді. Табғачығ өлүртегі табғашты өлтіреді. Өтүнүчүмін есідү берті - Өтінішімді ести берді. Алы олур - Ала отыр. Кімге іліг қазғанурмен, не қағанқа ісіг - күчіг бірүрмен. Кімге елді пайдаланамын, кімге күш жігерім- ді жұмсаймын. Анта ічрекі будун көп көрүр Ондағы ішкі халықтың көбі бағынды.

б) Келер шақ. Ашық райдың келер шақ түрі істің әлі болмағандығын, тек болашақта, алдағы уақытта болатындығын білдіреді. Мысалы: келур – келер, көрүр – көрер, ерсер-егер жүрсе, барсар – барса, тодсар-тойса т.б. Бұйрық райдың болымсыз түрі -ма, -ме жұрнақтары жалғану арқылы жасалады: целме - желме, басытма - бастырма, олурма - отырма т. б.

с) Өткен шақ. Іс-әрекеттің айтылған сәттен бұрын болғандығын білдіретін ашық рай. Мысалы: Өңрекі будун келті- Алғашқы халық келді. Өрленде сүнүшдім, сүсін санчдым - Өрпенде соғыстым, әскерін шаныштым.

Ертис үгүзүг кечігсізін кечдіміз, түн қатдымыз. Болчуқа таң үнтүрү тегдіміз - Ертис өзенін өткелсіз кештік, түн қаттық, Болшұққа таң ата жеттік.

Қалау рай

Көне ескерткіштерде қалау рай істеушінің іске деген белгілі ықылас-ынтасын, ниетін білдіріп, төмендегідей жұрнақтар арқылы жасалды: қалау райдың жекеше түрі: -айын, -егч, -йын, -йін жұрнақтары жалғану арқылы жасалады. Мысалы: айтайын- айтайын, қысайын-қысайын, тегейін- тиейін, йоғлатайын –жоқтайын. Қалау райдың көпше түрі: -алым, -елім, -лым, -лім жұрнақтарының жалғануы арқылы жасалады. Мысалы: йоққысалым - құртайық, қабысалым- бірігейік, сулелім – соғысайық, баралым- барайық, йаналым –қайтайық т.б.

Бен саңа не айайын - Мен саған не айтайын. Білге қағанға адырылмалым - Білгіш қағаннан айрылмайық. Мәтінде кейбір қалау рай -ғай жұрнағы жалғану арқылы да жасалатыны байқалады: болғай – болғай.

Берілген мысалдардан етістіктің рай категориясының жан-жақты, әр тарапты, кең көлемді екенін байқауға болады. Сонымен қатар, етістіктің рай категориялары етістіктің барлық формасымен тығыз байланысты.

Шартты рай

Шартты рай – белгілі бір іс-әрекеттің орындалу шартын білдіретін етістік түрі. Көне түркі V-VIII ғасыр ескерткіштерінде (Орхон-Енисей жазбаларында) шартты рай тұлғасы -са, -се -сар, -сер түрінде ұшырасады: ол йергеру барсар, түрк будун, өлтөчісен. Бұл тұлғаның қазіргі тілдегі тұлғадан негізгі айырмашылығы тәуелдік жалғауларды қабылдамаған. Қай жақта айтылғаны постпозициялық қолданыстағы жіктеу есімдіктері арқылы берілген. Көне түркілік –сар, -сер аффиксі кейінгі дәуірлерге қатысты жазбалар тілінде кездеседі. Мәтіндік қолданысына қарай бірде шартты, бірде істің істелуге тиіс уақыт, мезгілін білдіріп отырған. Мысалы: ысар – жіберсе, ергер – егер жүрсе, тодсар – тойса, сүлемесер – соғыспаса. Күлтегін йоқ ерсер, коп өлтөчті ертігіз – Күлтегін жоқ болса, көп өлетін едіңіз[11].

Қорыта айтқанда, түркі тіл біліміндегі грамматика саласының ұғымдық ерекшеліктері мен метатеориялық негіздері тілдік иерархияның іштей саралануы, тіл құрылымының салалық-деңгейлік жүйеленуі тұрғысынан талдаып, тіл мен тіл иесі арасындағы көпқатпарлы, көпвекторлы байланыс глоттохронологиялық өріс аясындағы тұлғалық-мазмұндық ұқсастығы арқылы көне ескерткіштерімізді зерттеу қазақ әдеби тілінің тарихына үлкен әсері бар екені сөзсіз.

Қорытынды:

Қазақ әдеби тілі тарихының тамырының тереңдігіне түркітанушы ғалымдардың зерттеулері де жоғарыда келтірілген мысалдардың қазақ тілімен



ўқастығы да жеткілікті дәлел бола алады. Әдеби тіліміз көне түркі кезеңінде қалыптасып, талай дәуір, талай кезең тармағынан өтіп, бүгінгі күні дербес тіл ретінде танылып отыр. Көне түркі ескерткіштерін зерттеу арқылы тек қазақ әдеби тілінің ғана емес, түркітілдес халықтардың тілінің тарихы, грамматикасы және олардың қазақ тілімен ұқсастығын көреміз. Енисей және Орхон руникаларына зерттеулер жүргізу арқылы тіліміздің әдеби тіліміздің тарихына, қалыптасуына, тілдік негіз алуының бастауы ретінде қараймыз. Қазақстанның көне жазба ескерткіштерін одан әрі анықтау мен зерттеудің түркі халықтарының көне мәдениетін және олардың әлемдік өркениетке қосқан үлесін тереңірек түсінуге мүмкіндік беретін жаңа ғылыми проблемаларды әзірлеу үшін маңызы зор. Кірме сөздердің, терминдердің, варваризмдердің қазақи баламасы белең алған. Сөздік қордың өсуімен бірге тілдің грамматикалық құрылымы да жетілдірілді.

XV ғасырға қарай қазақ тілі өзіндік ерекше белгілері бар толық дербес тілге айналып үлгерді – мысалы, нақты диалектілік бөліну болмады, тек аз ғана лексикалық-фонетикалық ерекшеліктер ғана болды. Қазіргі қазақ әдеби тілін бұрынғы кезең бастауларымен салыстырғанда, үлкен өзгеріс пен жетістікті байқауға болады. Ұлттық тіліміздің қалыптасуының өзі қазақ әдеби тілінің кең тыныс алуына себепкер болды. Сан ғасырлық тарихы бар әдеби тіл, қазіргі кезеңге дейін сүзгіден өтіп, тілдік жүйесі мен грамматикалық құрылымы бірқалыпқа түсті.

Тарихқа көз жүгіртсек, қазақ жерінің өзінен басқа, оның ішінде Қазақстан территориясы мен көршілес елдер – Ресей, Өзбекстан, Түркіменстан, Қырғызстан, сондай-ақ Қытай, Моңғолия, Ауғанстан, Пәкістан, Түркия елдерінде қазақ тілі кеңінен таралған. Бүгінде қазақтар Еуропаның көптеген елдерінде, Таяу Шығыста, АҚШ пен Канадада, Оңтүстік-Шығыс Азия елдерінде тұрақты немесе уақытша тұрады. Демек, қазақ тілі табысты дамып қана қоймай, сөйлейтіндер саны мен таралу аймағы жағынан да өсіп келеді деп нық сеніммен айта аламыз.

Пайдаланылған әдебиетер:

1. Янсон Торе Тіл тарихы: Кіріспе Алматы: «Ұлттық аударма бюросы» қоғамдық қоры, 2019. – 244 б.
2. Момынба Б. Қ. Қазақ әдеби тілінің тарихы: оқулық. – Алматы: Қазақ университетті, 2011. – 283б.
3. Исаев С. Қазақ тіл білімінің мәселелері: оқулық. – Алматы: Абзал-Ай, 2014. – 640б.
4. Айдаров Ғ. Көне түркі жазба ескерткіштерінің тілі: оқулық. – Алматы: Мектеп, 1986. – 183б.
5. Серебренников Б.А. Роль человеческого фактора в языке: Язык и мышление: книга. – Москва: Наука, 1988г. – 248с.



6. Томанов М. Қазақ тілінің тарихи грамматикасы: оқулық. –Алматы: Мектеп, 1981. – 207б.
7. Щербак А. М. *Очерки по сравнительной морфологии тюркских языков:* (Имя) – Ленинград: отделение: Наука, 1977г. – 321б.
8. Хасанова С. Қазақ жазуы мен оқуының бастаулары (көне түркі және араб, латын жазбалары). — Алматы: 2007. — 264 б.
9. Гаджиева Н.З. Основы синтактической структуры пюркских языков. – Москва: Наука, 1973г. – 408с.
10. Баскаков Н.А. Историко-типологическая характеристика структуры тюркских языков. Словосочетание и предложение. – Москва: Наука, 1975г. – 287б.
11. Кононов Н. *Грамматика современного узбекского литературного языка* / Ответственный редактор И. А. Орбели. М. Л.: Издательство АН СССР, 1960. 447 б.

ЦИФРЛЫҚ БІЛІМ БЕРУ РЕСУРСТАРЫН ОҚЫТУДА ҚОЛДАНУ

Абирова Жұлдыз Дархановна

П.Дутов атындағы орта мектеп шағын орталығымен
Жетісу облысы

Аннотация. В данной статье рассматривается значение использования цифровых образовательных ресурсов в образовательном процессе, описание методических особенностей, анализ содержания и структуры ЦОР, разработанных в рамках проекта системы электронного обучения. Кроме того, представлены требования к ЦОР, принципы, содержание пакетов и методы его применения.

Abstract. This article examines the importance of using digital educational resources in the educational process, describes the methodological features, analyzes the content and structure of the e-learning centers developed within the framework of the e-learning system project. In addition, the requirements for the data center, principles, package contents and methods of its application are presented.

Ключевые слова: цифровой образовательный ресурс, система электронного обучения, электронный учебник, информационные технологии.

Білім және ғылым мемлекетіміздің әлеуметтік-экономикалық күшеюуіне өркендеуінің негізгі қозғаушы күші болып табылады. Білім саласын дамыту инновацияны ендіруге септігін тигізіп, жастарды қазіргі еңбек нарығында бәсекеге қабілетті болуға қажетті керекті білімдер мен дағдылармен қамтамасыз етеді.

Цифрлық білім беру ресурстары (бұдан әрі — ЦБР) білім беру мазмұнын анықтайтын электрондық оқыту жүйесі компоненттерінің бірі болып табылады. Білім берудің жоғары сапасын қамтамасыз ету үшін, оқу үдерісінде өскелең ұрпақтың ЦБР белсенді қолдану, бүгінгі таңда берілген бағдарлама аясындағы педагогикалық қоғамдастықтың алдында өзекті мәселелердің бірі болып табылады.

2011 жылы Ұлттық ақпараттандыру орталығы жалпы орта білім беру мекемелеріндегі электрондық оқыту жүйесі үшін цифрлық білім ресурстарын дайындау стандартын әзірлеген болатын. Ол жалпы орта білім беру мекемелеріндегі электрондық оқыту жүйесі үшін цифрлық білім ресурстарын дайындауға қойылатын педагогикалық, психологиялық, техникалық және дизайн-эргономикалық талаптардың жиынтығын ұсынады және электрондық басылымдарды дайындаушыларға арналған және электрондық оқыту жүйесі (e-learning) үшін әзірленетін барлық цифрлық білім ресурстары үшін қолданылады [www.ripkso.kz].

Осы стандарт бойынша цифрлық білім ресурстары — бұл білімдік процесс субъектілерінің интерактивтік қашықтықтан өзара әрекеттестік

ретінде электрондық оқитудың ақпарат қатынастық білімдік ортасының жинағын құруды қамтамасыз ететін электрондық тасуыштардағы дидактикалық материалдар деп анықталған.

Стандарттағы жалпы талаптар:

- ✓ Қазақстандық білім беру мазмұнын анықтайтын мемлекеттік құжаттарға сәйкестігі;
- ✓ Оқыту процесінің заңдылықтары мен қағидаларын есепке алу;
- ✓ Пән саласы білімдеріндегі педагогика ғылымының жетістіктерін есепке алу;

✓ ЦБР-дың мазмұны ұлттық дәстүрлер ескеріліп дайындалынады.

Цифрлық білім ресурстарға қойылатын жалпы педагогикалық талаптар:

Ғылымилық қағида бойынша:

- ✓ оқу материалының мазмұнын баяндаудың ғылыми дұрыстығы;
- ✓ пайдаланылатын терминологиялардың қазіргі заманғы түсіндірмелерге сәйкестігі;
- ✓ оқу материалын баяндау түсінікті, нақты, толық және қайшылықсыз болуы тиіс.

Түсініктілік қағидасы

- ✓ оқу материалын оқушылардың нақты жас ерекшеліктеріне бейімді түрде ұсыну.
- ✓ оқушыларда парасаттық, адамгершілік, физикалық артық жүктемелердің болмауы.

Көрнекілік қағидасы бойынша:

- ✓ Оқу материалын қабылдауға және қайта өңдеуге сезім мүшелерін мақсатқа лайықты қатыстыру;
- ✓ Оқу материалын барынша көрінерлік ету.

Жүйелілік қағидасы бойынша: ЦБР-дың оқу процесінің белгілі бір құрылымдық құрамдас бөлігіне сәйкестігі:

- ✓ құлшыныстық-мақсаттық;
- ✓ мағналы мазмұндық;
- ✓ операциялық іс әрекеттік немесе бағалаушылық нәтижелік.

Саналылық қағидасы мен интерактивтік бойынша:

- ✓ пайдаланушылар үшін оқу міндеттерінің айқын қойылымы.
- ✓ оқушылар мүддесіне тірек және ілім түрткілерін қалыптастыру. — оқушылардың белсенді ақыл-ой іс-әрекетін ұйымдастыру.
- ✓ шынайы уақыт режімінде өзара әрекеттестікті және кері байланыстарды ұйымдастыру.

Теорияның практикамен байланыс қағидасы бойынша:

- ✓ Оқу материалының практикалық бағдарланғыштығы.
- ✓ ЦБР-дың қоғам мен экономиканың қазіргі заманғы даму процесіне бағдарлануы.

Аталган педагогикалық талаптардан басқа Интерфейстің бір түрлілігі, яғни жазбалар, анықтамалар жүйесі, ЦБР әр түрлі құрамдас бөліктерін айқын ажыратуды қамтамасыз ететін бір типті таңбашаларды (икондарды) және басқа арнайы символдарды пайдалану, пайдаланылатын түстердің тұрақтылығы, фонға қатысы бойынша кескін қанықтылығының ұтымдылығы; орнықты көру бірлестігіне түстердің сәйкестігі, бір экранда толқын ұзындықтары әр түрлі төрттен артық түстерді пайдаланудың мүмкін еместігі және т. б. дизайн-эргономикалық талаптар да қарастырылады. Сондай-ақ, стандартта ЦБР-ға қойылатын техникалық талаптар, декларация құрылымы мен ресурстың сипаттамасы және де ресурстардың рұқсат етілген форматтары да берілген [Роберт, 2010].

Оқу үдерісінде ЦБР тиімді пайдалану үшін, алдымен ЦБР топтамасымен танысып, ондағы қажетті ресурсты таңдай білген жөн. Біздің елімізде ең ауқымды ЦБР топтамасы Ақпараттандырудың Ұлттық орталығының www.psi.kz сайтында орналасқан. ЦБР топтамасы қарастыратын негізгі мәселелер — мұғалімге сабақты даярлау барысында көмектесу; сабақ барысындағы көмек; оқушыларға үй тапсырмасын орындауда көмектесу. Мұғалімге сабаққа даярлық барысында көмектесу:

- ✓ жекеленген сандық объектілерден сабақты құрастыру және моделдеу;
- ✓ пән туралы білімді тереңдетуге арналған қосымша және анықтамалық кең көлемді ақпараттың болуы;
- ✓ ЦБР топтамасында ақпаратты тиімді іздестіру;
- ✓ бақылау және өзіндік жұмыстарды даярлау;
- ✓ шығармашылық тапсырмаларды даярлау;
- ✓ цифрлық объектілермен байланысты сабақ жоспарларын даярлау;
- ✓ Интернет арқылы басқа мұғалімдер қызметтерінің нәтижелерімен алмасу.

Сабақ барысындағы көмек:

- ✓ мультмедиялық проектор арқылы даярланған цифрлық объектілерді демонстрациялау;
- ✓ лабораториялық жұмыстарда виртуальды лабораториялар мен интерактивті моделдерді қолдану;
- ✓ оқушыларды компьютерлік тестілеу және білімдерін бағалауға көмектесу;
- ✓ сабақта оқушылардың ЦБР-мен жеке зерттеушілік және шығармашылық жұмыстары;

Оқушыларға үй тапсырмасын орындауда көмектесу:

- ✓ материалды жаңаша көрсету арқылы оқушылардың пәнге деген қызығушылығын арттыру;
- ✓ оқушылардың өздеріне ыңғайлы кез келген уақытта өзін-өзі автоматты бақылауы;

- ✓ Баяндамалар, презентациялар, рефератлар және тағы сол сияқтыларды даярлауға үшін қажетті объектілердің үлкен қоры;
- ✓ Энциклопедиялық сипаттағы қосымша ақпараттарды тез арада алу мүмкіндігі;
- ✓ Пәндік виртуалды ортада оқушылардың шығармашылық әлеуетін дамыту;
- ✓ Оқушының пәнді оқуын өзіне ыңғайлы темпте және материалды игеруді өзінің жеке қабылдау ерекшелігіне байланысты таңдалған деңгейде ұйымдастыруына көмектесу;
- ✓ Оқушыларды заманауи ақпараттық технологияларға үйрету, АТ меңгерудегі талпыныстарын және олармен үнемі жұмыс істеуді қалыптастыру [www.nci.kz].

ЦБР топтамасының мазмұны: Ақпаратты алу блогы: ғылыми-танымал мақалалар, қайнар көздерінің мәтіндері, оқулықтардың үзінділері, иллюстрациялар, анимациялар, басқа да мультимедия компоненттері, мультимедиялық сабақ-презентациялары. Виртуальды галерея: бейнеүзінділер, анимациялар, нақты және синтездік бейнелер. Дыбыстық объектілер. Виртуальды лаборатория: интербелсенді модельдер, интербелсенді анимациялар, виртуальды лаборатория. Анықтамалық материалдар: графиктер мен диаграммалар, ғалымдардың өмірбаяндары, түйіндемелері бар Интернет-сілтемелер. Терминдер, анықтамалар және заңдардың сөздігі. Аттестация: сұрақтар мен есептердің жиынтығы, зерттеушілік іс-әрекетке арналған тапсырмалар. ЦБР жиынтығында шартты түрде мынадай блоктарды атап көрсетуге болады: итербелсенді компоненттер; демонстрациялық графика; мәтіндер; мұғалімге арналған материалдар [Савелова, 2008].

Білім беру үдерісінде ЦБР-ды қолданудың негізгі әдістері:

Біріншіден, жаңа материалды түсіндіргенде немесе жаңа сабақты бекіткенде қолдану әдісі. Бұл жағдайда анимациялық, бейне үзінділер, дыбыстық файлдар, графикалық кескіндерді көрсету сияқты ЦБР-ды пайдаланған дұрыс.

Екіншіден, оқушының өзіндік оқу іс-әрекетін ұйымдастырғанда. Бұл жағдайда оқу комплексінің барлық материалдары пайдалы болуы мүмкін.

Үшіншіден, ЦБР-ды әртүрлі бақылауларды (кіріс, ағымдық, кесінді, қорытынды) ұйымдастырғанда. Мұнда бақылау-диагностикасы оқушылардың компьютерлік тестіленуі.

Төртіншіден, мультимедиялық құралдарды пайдалану арқылы ЦБР-ды нақты бір пән не бағыт бойынша топтастыруға мүмкіндік беретін әдіс.

Бесіншіден, ЦБР-ды лабораториялық жұмыстарды орындау кезінде. Мұндай ЦБР — дың артықшылығы қолдануы өте үлкен және қымбат тұратын құрылғылардан тұратын оқу кластары мен лабораторияларды алмастыруға мүмкіндік береді.

Алтыншыдан, жаратылыстану ғылыми циклінің пәндері мен информатика және АКТ үшін интерактивті ЦБР-ды жаттықтырушы ретінде қолдану әдісін айтуға болады.

Жетіншіден, оқушыларға мұғалімнің қатысуынсыз өздеріне ығайлы уақытта, өз темпінде теориялық материалдармен танысуға, лабораториялық жұмыстар мен тестік тапсырмаларды орындауға мүмкіндік беретін қашықтықтан оқытуды ұйымдастыру әдісін айтуға болады. [6]

Қорыта келе, сандық білім ресурсы мұғалімді алмастыра алмайды, бірақ мұғалімге қосымша материалдарды ұсынады, яғни сабақ мазмұнын ақпараттық коммуникациялық технологиялардың жаңа мультимедиялық мүмкіндіктерімен толықтыруға, оқушылардың назарын аса маңызды оқу тақырыптарына аударуға, қажет болған жағдайда оқушылардың назарын зерделенген көріністердің ерекшеліктеріне шоғырландыруға, оны көрнекі түрде көрсетуге, сабақтың мазмұнын қоғамда болып жатқан өзгерістермен, өмірлік тәжірибелермен, оқушылардың пәнге деген қызығушылықтарымен және т. б. құбылыстармен байланыстыруын қамтамасыз етуге мүмкіндік береді деп айта аламыз.

Әдебиеттер:

1. Білім беру жүйесінің басшы және ғылыми-педагогикалық кадрлары біліктілігін арттыратын республикалық институт (www.ripkso.kz)
2. Роберт И. В. Теория и методика информатизации образования (психолого-педагогический и технологический аспекты). 3-е изд., доп. — М.: ИИО РАО, 2010. — 356 с.
3. АҚ «Білім беруді ақпараттандырудың ұлттық орталығы» (www.nci.kz)
4. Савелова Е. В. Цифровые образовательные ресурсы в школе: методика использования. Обществознание. Сборник учебно-методических материалов для педагогических вузов. — М.: Университетская книга, 2008. — 224 с.

ВОЗМОЖНОСТИ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ ДЛЯ АКТИВИЗАЦИИ ПОЗНАВАТЕЛЬНОЙ ДЕЯТЕЛЬНОСТИ УЧАЩИХСЯ

Жисува Эльмира Сарбалаевна

учитель русского языка и литературы
средняя школа имени Имангали Билтабанова

Аннотация. Бул мақалада орыс тілі мен әдебиетін оқытуда оқушылардың танымдық белсенділігін, қызығушылықтарын арттыру үшін қолданылатын ақпараттық технологиялар жайында айтылады. Нақты жұмыс түрлері көрсетіледі.

Abstract. This article discusses information technologies used in teaching the Russian language and literature to increase the cognitive activity and interest of students. Specific types of work are indicated.

Тірек сөздер: білім беру, оқыту технологиясы, цифрлық мектеп, электронды ресурстар.

Наступила эра смешанного обучения. Учителей призывают использовать цифровые ресурсы. Цифровая школа - наше будущее. И что-то мне подсказывает, что многие коллеги ради цифры как таковой будут попросту терять время урока. Очень хочется помочь таким учителям, потому что есть эффективные компьютерные упражнения для уроков русского и литературы. Не обязательно жертвовать качеством работы ради отчёта. Напротив, качество можно повысить.

Электронные упражнения не дадут должного эффекта, если преподаватель не понимает, что и зачем он использует. Использование цифровых ресурсов не должно быть самоцелью. Давайте разберёмся, зачем использовать цифровые технологии при изучении гуманитарных предметов.

Цели использования электронных упражнений на уроках русского и литературы

- Создание мотивации (ученики получают интересные разнообразные задания в том виде, к которому они привыкли).
- Повышение качества образования для всех (возможность тиражировать и использовать в любой школе материалы лучших специалистов).
- Доступность во время дистанционного обучения
- Экономия времени (использование упражнений, которые можно организовать без компьютера, но с большими временными затратами).
- Эффективность (использование эффективных упражнений, которые невозможно организовать без компьютера).

- Формирование универсальных навыков.

Возможности, которые дают цифровые технологии:

- автоматическая проверка (экономия времени преподавателя, возможность задавать на дом тренировочные диктанты) ЯКЛАСС, Лернинг апс,
- мультимедийность (возможность вставить в учебник звук, видео, что помогает лучше запомнить материал);
- интерактивность (включение в активную работу одновременно всех учеников группы, что даёт экономию времени);
- совместная работа группы над проектом (возможность совместно создавать документы, презентации, ментальные карты и т.п. исследовательской группой учащихся);
- разнообразие (возможность быстро генерировать большое и даже бесконечное количество однотипных упражнений);
- экономия времени учителя, повышение качества образования для ученика.

Под использованием цифровых технологий преподаватели пока понимают лишь показ презентаций и обучающих видеороликов, раздачу ссылок на текстовые материалы да компьютерное тестирование. Что ещё можно сделать:

- предлагать упражнения на сортировку слов, терминов, портретов писателей и т.д. с автоматической проверкой,
- предлагать упражнения типа «вставь букву» с автоматической проверкой,
- предлагать зрительные диктанты с автоматическим таймером и проверкой,
- предлагать задания на поиск соответствий (определений и терминов, предложений и схем, слов с одним признаком, названий произведений и писателей),
- предлагать ученикам самостоятельно создавать подобные упражнения,
- создавать общий текст со множеством гиперссылок при анализе текста (например, в гуглдоках создать файл или презентацию с гипертекстом и ссылками на источники),
- создавать общие презентации,
- создавать ментальные карты,
- создавать профили литературных героев в соцсетях,
- создавать хронологические ленты (история литературы),
- создавать лендинги - презентации литературных произведений,
- создавать слайды с инфографикой по произведениям и авторам,
- создавать облака тегов (слова на одно правило, слова произведения, логические задачи),
- создавать читательские дневники в блогах,

- создавать видеоролики по стихам и коротким рассказам,
- анализировать тексты с помощью компьютерных программ, подсчитывающих частотные слова и выражения, выделяющих общие элементы в разных текстах и т.п. (очень помогает при анализе мотивной структуры),

Этот перечень можно продолжать и продолжать. Технологии дают возможности, дело за вашей фантазией.

Остановимся на некоторых из них.

Сколько времени вам потребовалось бы на организацию зрительного диктанта для всего класса без компьютера? Если давать один вариант на всех, надо заранее написать предложение на доске, закрыть его и последить, чтоб на перемене никто не подсмотрел. На уроке открыть доску, засечь время, снова закрыть, собрать листочки для проверки или надеяться на качественную самопроверку. Максимум можно дать 2 таких предложения, потому что укромных мест на доске ровно два (если доска вообще закрывается). Чтобы раздать каждому свой вариант, надо потратить минут пять, а потом следить, чтоб все перевернули листочки и не списывали. В общем, одна возня, поэтому мало кто использует зрительные диктанты, хотя все понимают, что такая работа очень полезна. Автоматизированный зрительный диктант вместе с проверкой занимает не более 2 минут на одну фразу, причём фраз предлагается огромный набор(презентация)

А вот пример использования гуглкарт: путеводитель по городу N (по комедии Н. В. Гоголя «Ревизор» (8 кл.). Тут и визуализация, и метапредметность, и работа в группах, и использование технологий.

Какие сайты можно использовать для создания интерактивных заданий?

Казалось бы, как далеки русский язык и информатика! Однако появляется всё больше сервисов, которые позволяют сделать урок русского языка и литературы не только нескучным, современным и технологичным, но и более эффективным.

Самое простое, что нам подарили информационные технологии, - создание презентаций к уроку. Если вы знаете, где брать хорошие иллюстрации, если вы без труда обрабатываете эти иллюстрации (например в Фотошопе), если можете с нуля создать шаблон оформления, и главное - если у вас есть время на эту работу, то вам не нужны какие-то особые программы для создания презентаций к уроку. Хватит и стандартных средств типа PowerPoint. Существуют интернет-сервисы, которые значительно упрощают процесс создания презентаций и экономят время учителя. Популярен сайт Canva

Canva - это графический сервис, который позволяет создавать любые виды презентаций прямо в браузере без установки какого-либо программного обеспечения на компьютер. Ресурс ориентирован на непрофессиональных пользователей, что отражается не только в простоте использования сервиса,

но и в детальной системе обучения, которая расскажет о всех возможностях работы с проектами.

Canva предлагает своим пользователям более 300 000 бесплатных изображений для фона презентаций, более 50 тысяч уже готовых шаблонов и более 35 разных видов шрифтов от рукописных до стандартных.

Сервис предусматривает совместную работу нескольких людей над одним проектом, использование собственных медийных материалов при оформлении слайдов. А это значит, что одну презентацию одновременно могут делать несколько учеников.

Дополнительный функционал: создание плана урока, грамоты, оформление рефератов и выпускных альбомов. Кроме того, если вы ведёте учительский блог, в этом сервисе можно быстро сделать стильные иллюстрации к вашим статьям. Все необходимые размеры уже заданы, выбираете соцсеть и фон, вводите текст - и всё готово!

Инструмент для анализа текста - Национальный корпус русского языка.

Сайт представляет собой громадную библиотеку текстов и механизмы для анализа этих текстов. Например, можно быстро найти фразу, которую использовали разные авторы, что очень помогает для сравнительного анализа текстов. Моментально собирается материал для мотивного анализа по ключевому слову.

Интересна статистика употребления слов. Можно быстро получить график, который покажет, как изменялась «мода» на то или иное слово в течение столетий. Например, этот график показывает, как устаревал глагол «извольте». График составил за доли секунды.

Игры со словами

Для активизации словарного запаса можно использовать различные браузерные игры со словами. «Балда» всем известна и доступна в соцсетях, но есть более сложные и интересные игры. Некоторые из них предлагают создатели сайта <http://www.iqfun.ru/>. Там так много всего, что даже не возьмёмся описывать. Просто сами зайдите и поиграйте, и вам захочется предложить такие задания ученикам.

Запомните новое слово - майндмэппинг. Это удобная и эффективная техника визуализации. Ментальные карты можно применять для фиксации идей, анализа и систематизации информации. Что-то типа создания очень разветвлённого плана, только не в линейной записи. Предлагайте ученикам создавать ментальные карты для закрепления материала. Например, можно устроить конкурс на лучшую карту по видам придаточных предложений или по языковым средствам выразительности.

Легко устроить мозговой штурм на уроке литературы и, используя совместно одну доску, быстро добавить много примеров из текста на заданную тему. Просто шикарно для анализа мотивной структуры произведения. Тут вам и совместная групповая работа. Карты размещаются на бесконечных

досках и могут расширяться в разные стороны. Поэтому не сложно теорию подкрепить многочисленными примерами. Примеры могут быть в прикрепленных файлах, открытых для совместного редактирования. Так вы используете ещё одну технологию. Обычно сервисы создания ментальных карт позволяют добавлять прямо на доску прикрепленные файлы для скачивания, фото и видео. К анализу эпизода можно присоединить отрывок из фильма-экранизации, а ещё лучше - из нескольких фильмов, чтоб возникло поле для сравнения и дискуссий.

Стрелки могут показывать сеть очень сложных взаимосвязей блоков. Это существенное отличие от линейного плана.

Чем карта отличается от обычного способа работы с информацией?

Во-первых, информация визуализируется, во-вторых, систематизируется, в-третьих, проще выделить основную идею, которая размещается в центре, проблемный вопрос, в-четвертых, это мыслительный процесс, демонстрирующий индивидуальный способ восприятия. Достоинства сервиса я вижу в том, что он позволяет работать как индивидуально, так и группой, применим в урочной и внеурочной деятельности для учащихся 5-11 классов на разных этапах урока, при подготовке к СОР, СОЧ, итоговой аттестации. Ментальные карты можно создавать по определенным темам и параграфам учебника, по биографиям поэтов и писателей, по правилам русского языка. Создавая ментальную карту, ученик работает с информацией, перерабатывает ее, самостоятельно изучает учебный и справочный материал. Для создания карты нужно вдумчиво прочитать произведение.

Личные орфографические словарики.

Создайте вместе с детьми личные орфографические словарики. Для начала добавьте те слова, которые запоминаете в этой четверти, потом можно приучить учеников пополнять словарики, работая над ошибками. Вы сразу получите несколько бесплатных автоматизированных упражнений со словами, попавшими в словарики, в разделе Тренинг. Если слово есть в нашей базе, оно попадёт в озвученные задания, если нет - только в зрительные диктанты. Программа анализирует частоту ошибок и предлагает проблемные слова чаще.

Справа от вас на этой странице круляш «Выучить 15 слов».

Например, на этой странице подборка из трёх упражнений. Сначала слабый ученик выполняет очень простую задачу: растаскивает карточки по двум полям в соответствии с выделенными буквами. Если карточка попадёт не на место, прозвучит звуковой сигнал. Ребёнок точно справится с таким заданием, получит позитивный опыт и запомнит по ходу дела несколько слов. Затем эти же слова предлагаются в другом порядке и уже с пропущенными буквами. Можно предложить классу проговаривать слова хором. В момент

нажатия на карточку появляются пропущенные буквы и ученики могут проверить себя.

Страничка литературного героя в соцсетях

На уроках литературы для полного понимания задумки автора текста, образов главных героев мы часто должны заниматься кропотливой рутинной работой. Но всегда ли это нравится нашим ученикам? Мне кажется, что нет. И причин для такого ответа множество. Поэтому каждый раз, подходя к изучению серьезного произведения, я продумываю, какое же интересное задание можно использовать, чтобы неинтересная работа заиграла другими красками.

Я считаю применение информационных технологий необходимым на уроках русского языка, литературы и мотивирую это тем, что они способствуют совершенствованию практических умений и навыков, позволяют эффективно организовать самостоятельную работу и индивидуализировать процесс обучения, повышают интерес к урокам русского языка и литературы, активизируют познавательную деятельность учащихся, осовременивают урок.

К современным методам образования можно добавить геймификацию. Термин геймификация (игрофикация) появился не так давно и часто вызывает недопонимание. Геймификация — это процесс использования игрового мышления и динамики игр для вовлечения аудитории и решения задач, превращение чего-либо в игру. Если сказать просто, геймификация — это инструмент, позволяющий усилить мотивацию и вовлеченность.

Значит, создавая геймифицированную систему, мы можем превратить процесс выполнения задачи в удовольствие.

Геймификация учебного процесса напоминает компьютерную игру, когда за каждое правильно выполненное задание участнику начисляется определенное количество баллов, которыми игрок распоряжается по своему усмотрению.

Геймификация отличается от дидактических игр тем, что специально разработанные игровые элементы и техники пронизывают весь процесс обучения.

Положительные черты:

- делает обучение более интересным,
- использует привычные игровые действия,
- повышает мотивацию,
- побуждает ребенка к самостоятельному обучению,
- применение знаний на практике,
- развивают организованность и командную работу.

Отрицательные черты:

- обучение может стать азартным, потерять обучающую суть,

- развитие конкуренции и соревновательного отношения между обучающимися,

- подготовка разработок требует от преподавателя много времени.

Что нужно учитывать при создании учебной игры:

- придумать сюжет,
- определить цель,
- разработать игровую механику,
- распределить роли,
- использовать привычные вещи в необычном ключе.

Наряду с деловыми и ролевыми играми, геймификация учебного процесса способствует развитию коммуникабельности, целеустремленности, познавательной и интеллектуальной активности обучающихся и так далее.

Среди образовательных технологий достойное место заняла ИКТ - технология, так как является главнейшим средством доступа к различным источникам информации и мотивируют к самостоятельному поиску, анализу и использования полученной информации. Практическое применение информационных технологий. При создании уроков – игр были использованы следующие программы и сервисы:

1. <http://LearningApps.org/>
2. <http://youtube.ru/>
3. <http://www.wordle.net/>
4. Prezi
5. Playcast
6. Test Master
7. Hot Potato
5. Виртуальная доска IDroo
6. Презентации по шаблону Комаровского
7. Презентация в PowerPoint

Разумеется, учебная игра отличается от традиционных детских игр. Потому и введён в педагогику термин «геймификация» как введение в жизнь (читай, урок) игровых механик с целью повышения познавательного интереса и активности обучающихся.

Уроки литературы содержат большие возможности для использования приемов геймификации. Здесь актуальны инсценировки, рассказы от имени героя, прогнозирование финала, интервью с персонажем, писателем, викторины и литературные диктанты по произведениям.

В рамках предметной недели использую игру «Литературный музей», где учащиеся сами подбирают и демонстрируют классу предметы, «принадлежащие» персонажам изученных литературных произведений.

В старших классах провожу урок -конференцию по обзорным темам, в ходе которого класс делится на группы: критики, оппоненты, журналисты.

Выступающие готовят доклады по заранее выбранным темам и выступают перед аудиторией, вступая в дискуссию, отстаивая свою позицию. Такого рода уроки способствуют формированию навыков публичной речи, учат ребят держаться на публике.

Это далеко не полный перечень возможностей геймификации на уроках русского языка и литературы. Главное, что они помогают разнообразить урок, заинтересовать детей, повысить их познавательную активность.

Достоинство геймификации еще и в том, что данную технологию одинаково результативно можно использовать как в начальной и средней школе, так и в процессе профессионального образования, причем на всех этапах учебного занятия: в начале, в ходе изложения нового материала, при закреплении знаний.

Литература:

1. Попова Н.В. Мотивация учебной деятельности учащихся. Дифференцированный подход к обучению. – Режим доступа: <http://festival.1september.ru/articles/508356/>. – Загл. с экрана.
2. <https://www.hr-director.ru/geymifikatsiya>
3. Образовательные веб-квесты. Быховский Ярослав Семенович. [Электронный ресурс]: Режим доступа: <http://ito.edu.ru/1999/III/1/30015.html>.
4. Игра захватывает школу // Кусочки утопии: нас ждет геймификация образования. [Электронный ресурс]: Режим доступа: <http://luckyea77.livejournal.com/831284.html>.

İLKOKUL ÖĞRENCİLERİ İÇİN ELEKTRONİK SÖZLÜK OLUŞTURMANIN ÖNEMİ

Barnoyeva Zarina Saifiddinovna

stajyer-araştırmacı,

barnoyevazarina072@gmail.com

Buhara Devlet Üniversitesi

Özet: Bu makalede ilkokul öğrencileri için eşanlamlılar sözlüğü oluşturmanın önemi ve elektronik sözlük oluşturmanın öneminden bahsedilmektedir. Öğrencilerin ilkokul ders kitaplarından okuma-yazma ve anadil bilimi alıştırmalarında verilen eşanlamlılarla ilgili görevleri tamamlamaları zordur. Bu amaçla öğrencinin bir eşanlamlılar sözlüğüne ve bunun yaratılışının dilsel temeline ihtiyacı vardır.

Abstract: This article talks about the importance and relevance of creating a vocabulary of synonyms for elementary school students. It is difficult for students to complete tasks related to synonyms given in reading literacy and mother tongue science exercises from elementary school textbooks. For this purpose, the student needs a dictionary of synonyms and the linguistic basis of its creation.

Аннотация: В данной статье говорится о важности и актуальности создания электронного словаря синонимов для учащихся младших классов. Учащимся сложно выполнять задания, связанные с синонимами, данными в упражнениях по грамотности чтения и знанию родного языка из учебников начальной школы. Для этого студенту необходим словарь синонимов и лингвистическая основа его создания.

Anahtar kelimeler: eşanlamlılık, eğitimsel kelime bilgisi, terminoloji, sözlükbilimi, varyant, eşanlamlı seriler, referans şemaları, ifade şemaları.

Giriş. Dünya dilbiliminde, geçtiğimiz yüzyılın ortalarına gelindiğinde, dili öğretme amaçlı tanımlayan dilsel-didaktik araçların araştırılmasına önem verilmiştir. Antik felsefeden günümüze kadar düzenli olarak üzerinde durulan ve düşüncenin etkili, açık ve çekici bir şekilde ifade edilmesinin önemli bir aracı olarak kabul edilen anlambilim olgusunun incelenmesinde yeni bir dönemin başlangıcı, Eşanlamlı sözlüklerin modern dil paradigmaları ışığında incelenmesi, antroposentrik sözlükbilimin yenilikçi sözlükbilimsel geleneklerine dayalı gelişmeler ve modern öğrenenlerin gereksinimleri, dilbilimsel, psikodilbilimsel yönlerin açıklığa kavuşturulması gerektiğini göstermek açısından önemlidir.

Özbek dilinin sözlükbilimi ve anlambilimi alanında yapılan son çalışmalarda eşanlamlılığın tamamen dilbilimsel bir olgu olduğu, dilde ayrı bir sözcüksel-anlam sistemi oluşturduğu ve sözcüksel-anlamsal bir paradigma olduğu tespit edilmiştir. Özellikle eşanlamlılığa dil-konuşma farklılaşması açısından yaklaşan H.Ne'matov

ve R. Rasulov'un yazdığı "Özbek dili sistemi sözlükbiliminin temelleri" kitabında eşanlamlılık, anlam bakımından farklı olan bir kelime dizisidir. Biçim ve aynı içeriğe sahip - birkaç dil biriminin belirli bir benzerliğe dayalı olarak bu birimler arasındaki birbirine bağlanması, belirli farklılıkları, bunların doğasını ve görünümleriyle ilgili durumları kapsayan sözcüksel bir paradigma olarak kabul edildi.

Ana bölüm. Eş anlamlılar, konuşmanın çekiciliğini, benzersizliğini, zenginliğini ifade eden başlıca sanatsal imaj araçlarından biridir ve dilbilimin ilk dönemlerinde bunların özünü ortaya çıkarmaya, anlamlarını kaydetmeye ve listeler yapmaya yoğun bir ilgi vardı. Dilbilim tarihinde oluşturulan ilk kaynakların sözlükler olduğu, sözlükbilim tarihindeki en eski sözlüklerin çoğunun ise eşanlamlıların açıklanmasına adanmış şerhler ve sözlükler olduğu vurgulanabilir. Kaynaklar, eşanlamlıların incelenmesi ve ilgili sözlüklerin derlenmesiyle ilgili uzun bir tarihten bahsediyor; örneğin, MS başlarında Çinliler, şairlerin kullanması için dünyanın ilk eşanlamlılar sözlüğünü derledi.

Bağımsızlık yıllarında ülkemizin sosyo-politik ve kültürel hayatında yaşanan iniş çıkışlar, Özbek sözlükbiliminin, özellikle de eğitim sözlükbiliminin gelişmesinde önemli bir dönemi başlatmıştır. Eş anlamlılar, paronimler, zıt anlamlılar, düzey kelimeler ve ilgili kelimelerin sözlüğü gibi çeşitli kelime sistemlerinin yanı sıra bazı eğitici eş anlamlı sözlükler oluşturulmuştur. Özellikle M. Olimov'un Özbekçe dil bilimi, yazılı eserlerin dilinin ilk eşanlamlı sözlüğünü, filoloji bilginleri ve öğretmenlerine, Özbek filolojisi ve gazetecilik fakültesi öğrencilerine yönelik sözlükleri oluşturmuştur.

Özbekçe dil bilgisi derslerinden özellikle ilköğretim öğrencilerinin karşısına en çok çıkan konulardan birisi de eş anlamlı kelimelerdir. Eş anlamlı kelimeler yazılışları ve okunuşları ayrı ancak anlamları aynı olan sözcüklerdir. Söz konusu kelimeler Özbekçede birbirinin yerine kullanılabilir. Günlük konuşma dilinde herkesin sıklıkla karşılaştığı eş anlamlı kelimelerin dilimizde önemi büyüktür. Bu tür kelimelerin listesi oldukça uzundur. Cümle içerisinde birbirlerinin yerine kullanıldıklarında anlam değiştirmezler. Türkçe eş anlamlı kelimeler listesi için “Eş Anlamlı Kelimeler Sözlüğü” içeriğimize göz atabilirsiniz. Eş anlamlı kelimeler yazılışları ve okunuşları birbirinden farklı olan ancak aynı anlamı taşıyan sözcüklerdir. 1. 2. 3. 4. sınıf öğrencilerinin sıklıkla karşılaştığı konulardan birisidir. Aynı anlamı taşıdıkları için bu kelimeler birbirlerinin yerine kullanılabilirler. Cümle anlamında değişiklik olmaz. Konuşma dilinde de kullanılan eş anlamlı kelimelerin listesi oldukça uzundur.

Eş anlamlı, yazılışları farklı olduğu halde anlamları aynı veya çok yakın olan sözcükler. Bu tür kelimeler genelde birbirlerinin yerini tutabilir. Özellikle öğrencilerin Özbekçe derslerinde gördüğü başlıklar arasında yer alır.

Özbekçe dil bilgisi derslerinde, özellikle 1. Sınıf, 2. Sınıf, 3. Sınıf ve 4. Sınıf öğrencilerinin karşısına en çok çıkan konulardan birisi eş anlamlı kelimelerdir. Eş anlamlı kelimeler listesi, öğrencilerin öğrenmesi için ödev olarak verilebildiği gibi,

derslerde etkinlik yoluyla da o'g'renmeleri hedeflenmektedir. Bilindiği gibi, eş anlamlı kelimeler, yazılışları ve okunuşları ayrı ancak anlamları aynı olan sözcükleri ifade eder. Bu kelimeler, Özbekçede birbirlerinin yerine kullanabilir; çünkü taşıdıkları anlam aynıdır. Eş anlamlı kelimeler listesi, binlerce kelimenin eş anlamlısını içermektedir.

Ders kitaplarında verilen eşanlamlılarla ilgili uygulamalı alıştırmalar aşağıdaki unsurlardan oluşur:

- a) Okunan metinlerden eş anlamlı sözcükler bulmak, anlamlarını, özellikle anlam yönlerini, eş anlamlı sözcükler arasındaki farkları tespit etmek;
- b) bu kelimelerin yerine geçebilecek eşanlamlıların seçimi ve kullanımlarındaki farkın tanımı;
- d) eşanlamlılar üzerine özel alıştırmalar (eş anlamlıların sıralanması, vb.);
- e) sözlü ve yazılı konuşmada eşanlamlıların aktif kullanımı
- f) Konuşma hatalarının (uygunsuz kullanılan kelimeler) düzeltilmesi, kelimenin eşanlamlı, metne daha uygun başka bir kelimeyle değiştirilmesi.

İlköğretim ana dili ders kitaplarında eşanlamlı kelimeler ayrı bir konu olarak ayrılmamaktadır. Kelime gruplarına ilişkin alıştırmalar ve görevler kapsamında eş anlamlı kelimelerle ilgili alıştırmalar verilmektedir.

Sözlük maddesi ayrı bir sistem olup, karşılıklı iç ilişkilere (çekirdek, merkez, bir çizginin çevresi) dayalı anlamsal bir paradigma, genel görünümü, her bir üyenin sözlükbilimsel açıklamasından oluşan bir mikrosistem olduğu belirlenmiştir. her eşanlamliya verilen tanımı doğrulayan paradigma ve açıklayıcı materyal;

Eşanlamlıların eğitici açıklayıcı sözlüğünün yapısal düzeni, derlemi, kelime dağarcığı - iç ilişkilere dayalı anlamsal paradigmalardan oluşturulması ve bu sistemde yer alabilecek eşanlamlıların ve açıklayıcı referansların seçimi gibi kriterler;

Özbek dilinde yeni nesil eşanlamlı eğitim sözlüklerinin derlenmesi, mega, makro, mikro yapıların geliştirilmesi.

Ayrıca geçen yüzyılın sonuna kadar kelimeler arasındaki anlamsal ilişkiler eşanlamlı ve zıt anlamlılıkla sınırlıydı, sistem analizi yöntemine dayanarak yapılan daha sonraki gözlemler, dilde anlamsal ilişkilerin çeşitli olduğunu, dereceli, alt anlamlı, eş anlamlı olduğunu ortaya çıkardı. , işlevsellik, hiyerarşiklik gibi türlerin tanımlandığı açıklanmakta, bu sözlük söz varlığının oluşumunda eşanlamlıları bu ilişkilerden ayırması açısından önceki sözlüklerden farklılık göstermektedir: bir tür şey-öge, işaret türünü temsil eden sözlüklerarası bir ilişkidir. -özellik, eylem durumu. Derecelendirme, aynı genel terime sahip ancak belirli bir terimin farklı düzeylerine sahip sözlüklerdir.

Sonuç ve öneriler. Sonuç olarak öğrencilerin alıştırmalarda verilen kelimelerin tamamına anlamlı bir kelime bulamayabilecekleri söylenebilir. Bu nedenle eşanlamlılar sözlüğüne ihtiyaç duyarlar. Ancak ilköğretim öğrencilerine yönelik eşanlamlılar sözlüğü şu anda mevcut değildir. Bu nedenle şimdilik Nizomiddin Mahmudov'un (2022) editörlüğünü yaptığı "Özbekçe Eş Anlamlıların Büyük Açıklamalı Sözlüğü" ve Azim Hajiye'v'in (1974) derlediği "Özbek Dili Eş



Anlamlıları Açıklamalı Sözlüğü"nden" kullanılmalıdır. Bu nedenle bilimsel çalışmalarımızın sonunda ilkokul öğrencilerine yönelik eşanlamlılar sözlüğü oluşturma hedefini belirledik.

KİTABIN:

1. Mirkhanova G.R. Özbek dilinin eşanlamlıları sözlüğünü oluşturma dilsel temeli: Filoloji. Bilim. Doktor (Doktora) diss. - 2020'ye karşı., - 143 s.
2. Adizova N.B. Anadil öğretiminin metodolojisi. - Buhara: Durdona. - 2021., 113 s.
3. Aydarova U.B. Okuryazarlık okumak. Çalışma Kitabı Bölüm I. - Taşkent: Novda. -34 s
4. Bakhridinova B.M. Özbekistan'da Eğitim: dilsel temeller, tarih ve perspektifler: Filol. Bilim. Doktor (DSc) incelemesi. - Semerkant 2020., -234 s.
5. Daniyarov B. Özbekçe sözcük eşanlamlılarının dil ve konuşma ilişkisi, sözlük bilimi çalışması. - Taşkent: Sano standardı., 2019.-S.-159.

ОҚЫТУДА ҚОЛДАНЫЛАТЫН ОНЛАЙН ВЕБ-РЕСУРСТАР

Кальжанова Айгуль Ермухановна

№2 Қандыағаш қалалық ЖББМ
қазақ тілі мен әдебиеті пәндерінің мұғалімі

Аннотация. В статье рассказывается об использовании онлайн веб-ресурсов в условиях дистанционного обучения. Одной из новых форм организации учебного процесса является обучение с помощью дистанционных технологий обучения. Дистанционное обучение-это форма обучения, которая проходит между учителем и учеником на некотором расстоянии в сопровождении ресурсов интернета, то есть обучение на определенном расстоянии с помощью сетей интернета.

Abstract. The article describes the use of online web resources in the context of distance learning. One of the new forms of organizing the educational process is learning using distance learning technologies. Distance learning is a form of learning that takes place between a teacher and a student at a certain distance, accompanied by Internet resources, that is, learning at a certain distance using Internet networks.

Ключевые слова: онлайн обучение, дистанционное обучение, веб-ресурсы, технологии обучения.

Оқытуда қолданылатын әдіс-тәсілдер, әдістемелер, технологиялар білім беру талаптарына байланысты жаңартылып отырады. Сапалы терең білімді меңгеруге талпынған өскелең ұрпақтың жеткілікті дәрежеде сандық сауаттылығы бар, себебі олар өмір жағдайлары барысында жаңа технологияның барлық мүмкіндіктерін пайдаланады. Сол мүмкіндіктерді нәтижелі қолдану мұғалімнің біліктілігіне, біліміне, тәжірибесіне байланысты [Ахметова, Караев, Мухамбетжанова, 2013].

Оқу үдерісін ұйымдастырудың жаңа формасының бірі – қашықтықтан оқыту технологиялары арқылы оқыту. Қашықтықтан оқыту дегеніміз — мұғалім мен оқушы арасында қандай да қашықтықта интернет ресурстарының сүйемелдеуімен өтілетін оқытудың формасы, яғни интернет желілерінің көмегімен белгілі бір арақашықтықта оқыту [<https://bilimdinews.kz>].

Қашықтықтан сабақ өзінің құрылымындағы (дидактикалық бөлім) күндізгі сабақпен бірдей болатындығын есте сақтау қажет, атап айтқанда:

- Ұйымдастыру сәті және міндеттер мен мақсаттардың құрылымы - оқушыларға материалды игеру уақыты, сабақ кестесі және т.б. көрсетілген нақты тапсырмалармен анықталады.
- Үй тапсырмасын тексеру - оқушылардың жетістіктері мен сәтсіздіктеріне жалпы шолу, қиын сәттерді түсіндіру және т.б.

- **Білімнің өзектілігі** - жаңа материалды - бейнероликтерді, оқу тапсырмасын оқуға деген ынтасын арттыру үшін түрлі әдістерді қолдану. Егер сіз бұл кезеңді сыныппен бір уақытта жұмыс түрінде өткізгіңіз келсе - бұл маңызды емес, сіз сұрақтардың жауабын талап етпей жай ғана жариялай аласыз, өйткені олар өздері баланы жаңа білім беру тапсырмасына дайындайды.

- **Жаңа білімді игеру** - бұл жалпыға қолжетімді ресурстарға орналастырылған білім беру порталдарының немесе мұғалімнің әзірлемелерінің ресурстарын пайдалану. Алайда, материалды оқымас бұрын оқушылар олардан не талап етілетіндігін нақты білуі керек, оған не назар аудару керек, қандай білім және практикалық дағдылар тексерілетінін көрсету қажет; тұтастай алғанда, оқушыларға жаңа материалды өз бетінше зерттеу алдында міндеттер нақты қойылуы керек.

- **Сабақты түсіну мен қорытындылаудың алғашқы сынағы** - бұл жаңа материалды зерделеу, білім мазмұнын тапсырмалармен, интерактивті модельдермен және өзін-өзі тексеруге арналған сұрақтармен тікелей байланыстыру кезінде осы кезеңді жүзеге асырудың тамаша мүмкіндігі.

- **Сабақты меңгеру, жіберген қателерді талдау және түзету** - тесттер, кіріккен интерактивті модельдер және жауаптарды жинау формаларын қолдану арқылы жүзеге асырылады. Тапсырмаларды уақтылы тексеруді және оқушыларға олардың жетістіктері мен қателіктері туралы көпшілікке немесе жеке ақпарат беруді ұмытпаңыз.

- **Үй тапсырмасы және оны орындауға арналған нұсқаулық.**

- **Рефлексия және кері байланыс.** [<https://edunews.kz>]

Осы жоғарыда аталған құрылымдағы сабақты түсіну, меңгеру кезеңдерінде оқушылардың білімін тексеріп, бағалауды онлайн режимде тиімді ұйымдастыруға көмектесетін келесі веб-сервистерді қолдануға болады: Quizizz, Socrative, Google forms, Learningapps, Polleverywhere.

Quizizz оқушыларды бағалауға арналған интернет-құрал болып табылады. Ол Kahoot веб-сервисіне ұқсайды, бірақ оларды бір-бірінен ерекшелейтін айырмашылықтары да бар. Бұл сервисе мұғалім өзі тест құрып, оны редакторлеу мүмкіндігіне ие болады. Ал оқушы болса, компьютер, ноутбук немесе смартфонды пайдаланып, ПИН-код пен өз атын енгізеді, яғни тест кез келген Интернетке қосылған құрылғы арқылы ойнатылады. Quizizz сервисінің Kahoot-тан айырмашылығының бірі – мұғалім сыныпты дәлірек басқара алады, әрбір оқушының жеке жұмысын бақылай алады. Барлық оқушылар бірдей тапсырмаларды алады, бірақ әр оқушы өз құрылғысында кездейсоқ таңдалған сұрақтарға жауап беріп, өз темпінде тестпен жұмыс жасай алады. Kahoot сервисіндегі сияқты әртүрлі алаңдататын факторлары жоқ, сондай-ақ оқушылар дисплейінде түгел сұрақпен бірге сурет толық көрсетіледі, әрі масштабталады.

Тестти аяқтаган соң нәтижесін тек көріп қана қоймайсыз, Excel кестесінде барлық мәліметтерді алуға мүмкіндігі бар.

Socrative бағдарламасы арқылы оқушы қашықтықтан тесттермен, түрлі бақылаулармен жұмыс істей алады. Ғаламторды тиімді пайдаланып, қысқа уақытта өз білімін толықтыру мүмкіндігіне ие. Socrative қосымшасы қолданыста өте қарапайым. Тестілеу немесе статистикалық сауалнама Интернетке қосылған кез келген құрылғы арқылы жүзеге асады. Тапсырма оқушыларға үш түрлі нұсқада беріледі: тест түрінде, дұрыс/дұрыс емес нұсқасы және қысқаша жазбаша жауап түрінде.

Оқытушы әрекеттерінің алгоритмі:

1. Жұмысты бастау үшін мұғалімге тіркелу қажет, ол үшін t.socrative.com сайтына кіріп логин мен парольді енгізесіз.

2. Бірнеше сауалнама сұрақтарын құрыңыз. Manage Quizzes батырмасын басыңыз.

3. Жұмысты бастамас бұрын оқушыларға өз виртуалды бөлмеңіздің нөмірін айтыңыз. Оқушылардың саны 50-ден аспауы қажет. Оқушы ресурсқа кіріп оқытушы айтқан нөмірді, аты-жөнін, фамилиясын енгізеді.

4. Тест қорытындысын Google кестесі түрінде берілген интерактивті есеп арқылы немесе Excel файлын электронды поштасынан көре алады. [<https://ortalyq.kz>]

LearningApps - бұл әр түрлі пәндер бойынша интерактивті оқу-әдістемелік қосымшалар жасауға мүмкіндік беретін білім беру мекемелеріндегі білім беру процесін қолдау үшін арналған сервис деп есептеймін.

Сервистің ерекшелігі, пайдаланушылар жасаған барлық тапсырмалар галереяда жиналған және жалпыға қол жетімді болып табылады. LearningApps құралдары әртүрлі интерактивті тапсырмаларды жасауға мүмкіндік береді: викторина, сұрыптау, топтастыру, жіктеу, мәтінді енгізу, кроссворд, уақытша таспа және т.б.

Іздестіруге оңай болу үшін барлық интерактивті тапсырмалар санат бойынша топтастырылған (пәндер бойынша). Категория бойынша таңдаудан басқа, қиындығына қарай да топтастырылған тапсырмаларды таңдап алуға мүмкіндік бар. Дайын тапсырмаларды серверге тіркелмегендер де пайдалануға мүмкіндігі бар. Жеке материалдарды жасау және сақтау қарапайым тіркеуден өткен соң ғана мүмкін болады. Сонымен қатар, өз материалдарын жасай отырып, пайдаланушы оларды жеке қолдану және жалпы пайдалану үшін сақтай алады. [<http://zkoipk.kz>]

Google form сияқты веб-сервистерді, аралық және ағымдық бақылау жүргізуге сәтті пайдалануға болады. Оқушылар аккаунттарына кірместен, желі әлсіздігіне қарамай бірден ашылып, пайдалануда және тест жұмыстарын жасауда қарапайымды, қолжетімді, тегін ұсынылатын сервис. Оқушыларды бағалаудың тестін жасап, жауаптарын уақытты үнемдей отырып қадағалауға мүмкіндік береді.



Polleverywhere сервисін оқушылардан кері байланыс алу мақсатында сабақтың рефлексия бөлімінде қолдануға болады. Бұл сервис өз алдына сауалнама жасауға арналған.

Әрбір оқушының білім деңгейін талапқа сай көтеру үшін озық әдістемелік технологиялар қажет. Сол жаңа озық әдістемелік технологиялармен қаруланған, заман талабына сай оқытудың жаңа әдістерін, яғни ақпараттық – коммуникациялық технологияларды толық меңгерген педагог қана оқушылардың сапалы білім алуына мүмкіндік жасайды.

Осы қосымшалардың бірін таңдау сабақ барысында мақсаттарға жету мен білімді игерту көзделеді және бұл қосымшаларды сабақтың білімді тиянақтау, топтық және жұптық, жаңа материалды игерту, үй жұмысын тексеру кезеңдерінде көптеп қолдануға болады деп есептеймін.

Пайдаланылған әдебиеттер мен сілтемелер:

1. Білім беру ұйымдарына электрондық оқыту жүйесін енгізу жағдайында педагогтардың біліктілігін арттыруды ұйымдастыру әдістемесі / Ахметова Г.К., Караев Ж.А., Мухамбетжанова С.Т. // Алматы:АҚ «ҰБАО «Өрлеу», 2013.
2. <https://bilimdinews.kz/?p=115010>
3. <http://edunews.kz/shcollnews/1803-ashytytan-oytu-zaman-talaby.html>
4. <https://ortalyq.kz/onlajn-o-ytudy-tiimdi-t-silderi/>
5. <http://zkoipk.kz/kz/smartconf2018/2-section/4194-conf.html>

ҚАЗАҚ ТІЛІ САБАҚТАРЫНДА АҚПАРАТТЫҚ- КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАРДЫ ҚОЛДАНУДЫҢ ТИІМДІЛІГІ

Молдабаева Зухра Мирамбековна

Қ.Жұбанов атындағы Ақтөбе өңірлік университеті

Жолдасова Назира Женисовна

Қ.Жұбанов атындағы Ақтөбе өңірлік университеті

Аннотация. В данной статье одним из наиболее эффективных путей развития сферы образования является информатизация образования, поясняются вопросы применения информационных технологий в системе образования. Применение информационных технологий в преподавании казахского языка - это необходимость, которая способствует совершенствованию практических навыков и умений; позволяет эффективнее организовать самостоятельную работу и индивидуальный процесс обучения; повышает интерес к занятиям; активизирует познавательную деятельность учеников, развивает интеллектуальные творческие способности.

Abstract. In this article, one of the most effective ways to develop the field of education is the informatization of education, explains the issues of using information technology in the education system. The use of information technologies in teaching the Kazakh language is a necessity that contributes to the improvement of practical skills and abilities; allows you to more effectively organize independent work and an individual learning process; increases interest in classes; activates the cognitive activity of students, develops intellectual creativity.

Ключевые слова: инновационная технология, обучение, образование, мультимедийная презентация.

Оқу процесінде ақпараттық-коммуникациялық технологияларды (АКТ) пайдалану қазіргі заманғы мектептегі білім берудің өзекті мәселесі болып саналады. Бүгінгі таңда кез-келген мектеп пәні бойынша әр мұғалім АКТ-ны қолдана отырып сабақ дайындап, өткізе алуы керек, өйткені қазір мұғалімге сабақты жарқын әрі қызықты етуге мүмкіндік берілді.

Жалпы білім беретін орта мектептердегі заманауи сабақ туралы педагогикалық практика тәжірибесіне сүйене отырып, біз төмендегідей жетекші идеяны анықтадық: сабақта оқушының рөлін пассивті тыңдаушыдан оны оқу процесіне белсенді қатысушы ету.

Осыған байланысты сабақтың қарқындылығын, оның нақтылығын арттыру мәселесі туындайды. Бұл мәселені шешудің бір жолы заманауи ақпараттық технологиялар болуы мүмкін.

Оқу процесіне АКТ енгізу екі негізгі бағытқа ие:

«Біріншісі, компьютер оқу процесіне оқыту жүйесінің дәстүрлі әдістері шеңберінде «қолдау» құралы ретінде енгізіледі. Бұл жағдайда компьютер оқу процесін күшейту құралы ретінде әрекет етеді.

Екіншісі, бұл іс жүзінде оқу процесін кең мағынада технологияландыру, оқытудың компьютерлік-ақпараттық модельдерін әзірлеу және енгізу Оқытудың заманауи техникалық құралдарын қолдану қажетті нәтижеге қол жеткізуге мүмкіндік береді» [Амандыкова, 2007; Садуова, 2006]

Табысты оқытудың маңызды компоненттерінің бірі - оқушының мотивациясы. Сабақтарда заманауи ақпараттық технологияларды қолдану кез-келген жастағы оқушы үшін оқуды жарқын, есте қаларлық, қызықты етеді, тақырыпқа эмоционалды оң көзқарас қалыптастырады.

Ақпараттық технологияларды қолдану мақсаттары:

1. Білім алушының жеке басын дамыту, ақпараттық қоғам жағдайында дербес өнімді қызметке дайындау:

- компьютермен қарым-қатынас ерекшеліктеріне байланысты сындарлы, алгоритмдік ойлауды дамыту;

- репродуктивті белсенділіктің үлесін азайту арқылы шығармашылық ойлауды дамыту;

2. Қазіргі қоғамды ақпараттандыруға негізделген әлеуметтік тапсырысты іске асыру: ақпараттық технологиялар арқылы білім алушыларды тәуелсіз танымдық қызметке дайындау.

3. Оқу-тәрбие процесін ынталандыру:

- ақпараттық технологиялардың мүмкіндіктерін іске асыру есебінен оқыту процесінің сапасы мен тиімділігін арттыру;

- танымдық белсенділікті белсендіру ынталандыруларын анықтау және пайдалану.

Қазақ тілінің мектеп курсын оқуда компьютерді пайдалану негізделген бірнеше негізгі бағыттарды бөліп көрсетуге болады:

- лингвистикалық құбылыстар туралы көрнекі түсінік;

- емле мен тыныс белгілерін зерттеу;

- тестілік бақылау жүйесі;

- ҰБТ және сынақтарға дайындық.

Компьютерді кеңінен қолдану оқуды көрнекі, түсінікті және есте қаларлық етеді.

Тестілеу жүйесін қолдана отырып, мұғалім оқушының білімін тексере алады, сонымен қатар баланың өзі материалды игеру дәрежесін басқара алады.

АКТ қолдану формалары.

1. Дайын электронды өнімдерді пайдалану мұғалім мен оқушының белсенділігін арттыруға, пәнді оқыту сапасын арттыруға мүмкіндік береді; көрнекілік принципін өмірге әкеледі.

2. Мультимедиялық презентацияларды пайдалану. Презентация - кестелер, диаграммалар, суреттер, иллюстрациялар, аудио және

бейнематериалдар ўсинылуы мўмкин слайд тўриндеги материалды беру формасы [Садуова, 2006].

Презентация оқу материалын жан-жақты қўрылымдық ақпаратпен толтырылған жарқын тїрек кескіндер жўйесї ретїнде ўсинуға мўмкіндік бередї.

Бўгїнгї таңда мен Қазақ тїлі сабақтарына арналған жеке тақырыптар бойынша бїрқатар презентациялар әзїрледїк. Сонымен қатар, сабақтарды дайындау және өткїзу кезїнде қолдануға болатын оқушылар жасаған презентациялардың кейбїрі жинақталған. Бїз оларды әрї қарай пайдалану ўшін презентациялар банкїн қўрамыз.

Өз тәжірибемде қолданамын:

1. Жалпылау сабақтары.

Презентацияның көмегімен сїз жалпылама сабақтарды дайындай аласыз. Сабақтың бұл тўрїнің мїндетї талдау процесїнде жасалған барлық бақылауларды тақырыпты бїртўтас қабылдау жўйесїне жинау, бїрақ терең тўсіну деңгейїнде, қазїрдің өзїнде қозғалған мәселелерден асып, бўкїл тақырыпты эмоционалды тўрде қамту. Диаграммалар, кестелер, материалдың тезистїк орналасуы уақытты ўнемдеуге және ең бастысы, зерттелген материалды тўтастай ўсинуға мўмкіндік бередї. Сонымен қатар, қорытындылар мен схемалар оқушылардың талқылауынан немесе сауалнамасынан кейїн бїртїндеп пайда болуы мўмкин. Мўғалїм презентацияның арқасында класс жўмысын ўнемі бақылай алады.

Осылайша, оқушыларда мемлекеттїк бїлім беру стандарттарымен ўсинылатын негїзгї қўзыреттїліктер қалыптасады:

- қызығушылық тудыратын тақырып бойынша ақпаратты жалпылау, талдау, жўйелеу мўмкіндігї;
- топта жўмыс їстей бїлу;
- әр тўрлі дерекөздерїнен ақпарат таба бїлу;
- коммуникативтїк қўзыреттїлік;
- алынған бїлім мен дағдылардың пайдалылығын тўсіну.

Презентациялармен жўмыс жасауда оқытуға жеке көзқарас жўзеге асырылады, әлеуметтену, жеке тұлғаны өзін-өзі растау процесї белсенді жүредї, тарихи, ғылыми және табиғи ойлау дамиды.

2. Интерактивтї тақтаны пайдалану.

Интерактивтї тақта - қазақ тїлін оқытудың қўнды қўралы. Интерактивтї тақталарды қолдану көптеген қосымша мўмкіндіктерді ашады. Қазақ тїлін ўйрену кезїнде интерактивтї тақтаны пайдалану - бұл пәнге деген қызығушылықты арттырудың, орфографиялық қырағылық пен жалпы сауаттылықты арттырудың тағы бїр қадамы.

Мўғалїмге арналған артықшылықтар :

- оқытушыларға класс орталығынан жаңа материалды тўсїндіруге мўмкіндік бередї;

-мўғалімдерге кез-келген қосымшалар мен веб-ресурстардың үстіне сурет салуға және жазбалар жасауға мүмкіндік беру арқылы импровизация мен икемділікті ынталандырады;

- көп уақыт пен күш жұмсамай және үйренген материалды тексеруді жеңілдетпей, сабақ кезінде жасалған кез келген жазбаларды қоса, тақтадан суреттерді сақтауға және басып шығаруға мүмкіндік береді;

- оқытушыларды оқытудың жаңа тәсілдерін іздеуге шабыттандырады, кәсіби өсуді ынталандырады.

Оқушыларға арналған артықшылықтар:

- сабақтарды қызықты етеді және мотивацияны дамытады;

- ұжымдық жұмысқа қатысуға, жеке дағдыларды дамытуға көбірек мүмкіндіктер береді;

- оқушылар материалды неғұрлым айқын, тиімді және динамикалық беру нәтижесінде неғұрлым күрделі материалды түсіне бастайды;

- оқытудың әртүрлі стильдерін қолдануға мүмкіндік береді, оқытушылар белгілі бір қажеттіліктерге бейімделе отырып, барлық ресурстарға жүгіне алады;

- оқушылар шығармашылықпен жұмыс істей бастайды және өзіне сенімді болады;

- бұл жабдықпен жұмыс істеу үшін пернетақтаның қажеті жоқ, осылайша оқушылардың оқу процесіне қатысуы артады.

Интерактивті тақталар компьютерлік мәдениет енгізген материалды ұсынудың таза презентациялық түрінен аулақ болуға мүмкіндік береді, ескертуден бас тарту арқылы сабақ уақытын үнемдейді. Интерактивті тақталар оқытуға бөлінген уақытты барынша тиімді пайдалануға және жалпы білім берудің тиімділігін арттыруға көмектеседі.

Бұл технологиялар бойынша жұмыс жалпы білім беру циклінің құрылымын сақтап қана қоймай, білім беру мазмұнының міндетті минимумының талаптарына толық сәйкес келеді, сонымен қатар:

- пәнге танымдық қызығушылықты арттыруға ықпал етеді;

- оқушылардың пән бойынша үлгерімінің өсуіне ықпал етеді;

- оқушыларға өздерін жаңа рөлде көрсетуге мүмкіндік береді;

- дербес өнімді қызмет дағдыларын қалыптастырады;

- әр оқушы үшін сәттілік жағдайын жасауға ықпал етеді.

Қазақ тілі сабақтарында АКТ кез-келген басқа сабақтар сияқты, мүмкін ғана емес, сонымен қатар қажет екеніне сенімдіміз. Біз ақпараттық қоғамда өмір сүріп жатырмыз, сондықтан да ақпаратқа ие адам әлемді иеленеді. АКТ-ны қолдану қазіргі сабақтың мүмкіндіктерін едәуір кеңейтеді, бұған біз АКТ-ны сабақта қолдану арқылы тағы да көз жеткіздік.

Қазақ тілі сабақтарында оқытушылар қолданатын ақпараттық компьютерлік технологиялар:

Интернет ресурстар. Интернет-ресурстарды пайдалану мүмкіндіктері өте зор. Ғаламдық Интернет желісі студенттер мен оқытушыларға Жер шарының кез келген жерінде қажетті кез келген ақпаратты алуға жағдай жасайды.

PowerPoint көмегімен жасалған компьютерлік дәріс-бұл экранда немесе мониторда көрсетілетін ақпараттық объектілердің тақырыптық және логикалық байланысты тізбегі. Дәріс барысында әртүрлі ақпараттық Нысандар қолданылады: суреттер (слайдтар), дыбыстық және бейне фрагменттер. Слайдтармен, суреттермен және басқа демонстрациялық материалдармен жұмыс тиімділігі оларды диаграммалар мен кестелерді көрсетумен толықтырған кезде әлдеқайда жоғары болады. Осындай сабақтардан кейін зерттелген материал студенттердің жадында жарқын сурет ретінде қалады және мұғалімге студенттің танымдық белсенділігін ынталандыруға көмектеседі. Көбінесе аралас типтегі сабақтар өткізіледі, онда үй тапсырмасын сұрау және жаңа материалды түсіндіру бар.

Электрондық оқулықтар. Электрондық оқулықтардың артықшылықтары: біріншіден, олардың ұтқырлығы, екіншіден, компьютерлік желілердің дамуымен байланыстың болуы, үшіншіден, қазіргі ғылыми білімнің даму деңгейіне сәйкестігі. Екінші жағынан, электронды оқулықтар жасау ақпараттық материалды үнемі жаңартып отыру сияқты мәселелерді шешуге де ықпал етеді. Олар сондай-ақ көптеген жаттығулар мен мысалдарды қамтуы мүмкін, әртүрлі ақпарат түрлерін динамикада егжей-тегжейлі суреттейді. Сонымен қатар, электронды оқулықтардың көмегімен білімді бақылау – компьютерлік тестілеу жүзеге асырылады.

Осылайша, білім беруде заманауи технологияларды қолдану оқушылардың жеке басын қалыптастыруға қолайлы жағдай жасайды және қазіргі қоғамның сұраныстарына жауап береді. Интернет желісінің ақпараттық ресурстарын пайдалана отырып, оларды оқу процесіне біріктіре отырып, қазақ тілі мен әдебиеті сабақтарында бірқатар дидактикалық міндеттерді тиімдірек шешуге болады:

әртүрлі күрделілік дәрежесіндегі желі материалдарын тікелей пайдалана отырып, оқу дағдыларын қалыптастыру;

мұғалім немесе оқушы ұсынған желі материалдарын проблемалық талқылау негізінде монологтық және диалогтық сөйлеу дағдыларын жетілдіру;

серіктестердің бірлескен іс-әрекетінің эсселерін, басқа эпистолярлық өнімдерін дайындауға қатыса отырып, серіктестерге жеке немесе жазбаша жауап беру арқылы жазбаша сөйлеу дағдыларын жетілдіру.

Мынадай қорытынды жасауға болады: қазақ тілі сабақтарында компьютерлерді қолдану оқуға деген қызығушылықты арттыру құралдарының бірі болып табылады. Бұл пәнді оқуда жеке қызығушылықты тудырады және кез-келген кезеңде оқудың сәттілігінің маңызды факторы болып



табылады. Бірақ компьютер ешқашан мұғалімді алмастыра алмайды, ол тек оқытудың сапасы мен бақылаудың тиімділігін арттыратын көмекші.

Әдебиеттер:

1. Амандыкова Г. Шет тілін оқыту әдістемесі. Алматы, 2007. 265 б.
2. Садуова Ж. Инновациялық: технологияларды білім беруде қолдану - Алматы: Білім. № 5, 2006. 22-24 б.

ҚАШЫҚТАН ОҚУ ЖАҒДАЙЫНДА ҚОЛДАНАТЫН БАҒДАРЛАМАЛАР

Акимбекова Айгүл Мырзахановна

№42 орта мектебі МКМ

қазақ тілі мен әдебиеті пәні мұғалімі

Асқар Тоқпанов ауылы Іле ауданы

Алматы облысы

Резюме. В статье рассматривается возможность использования в образовательном процессе дистанционных технологий обучения. Представлены средства обеспечения дистанционного обучения, достоинства его применения в процессе обучения.

Summary. The article considers the possibility of using distance learning technologies in the educational process. Means of providing distance learning, advantages of its application in the learning process are presented.

XXI ғасырдың басталуымен адамзат өркениеті үшін жаңа даму дәуірі басталды – ақпараттық қоғам дәуірі, ол жаңа ақпараттық-телекоммуникациялық революцияның дамуымен, ақпараттық технологиялардың тез таралуымен, әлеуметтік процестердің жаһандануымен, халықаралық конвергенциямен және көп салалы кооперациямен сипатталады. Қазіргі уақытта заманның даму талаптарына сәйкес ақпараттық-коммуникациялық технологиялар арқылы білімді бағалау және пайдалану жүйесін өзгертіп отыруды талап етеді. Осыған байланысты оқытуда қолданылатын әдіс-тәсілдер, әдістемелер, технологиялар білім беру талаптарына байланысты жаңартылып отырады. Сапалы терең білімді меңгеруге талпынған өскелең ұрпақтың жеткілікті дәрежеде сандық сауаттылығы бар, себебі олар өмір жағдайлары барысында жаңа технологияның барлық мүмкіндіктерін пайдаланады. Сол мүмкіндіктерді нәтижелі қолдану мұғалімнің біліктілігіне, біліміне, тәжірибесіне байланысты.

Сабақтарын алдын ала видео жазбаға түсіріп аудиторияға сапалы дүниелерді ұсыну мұғалімнің жаңа технологияларды тиімді пайдалана алатындығының белгісі. Видеодәріс пен видеопрактикалық сабақтарды түсіруге төмендегі скринкастинг (экранды түсіру) бағдарламаларды қолданса болады:

Movie Screen Recorder бағдарламасы - экраннан бейне жазуға арналған қосымшадан және Movavi бейне редакторынан тұратын скринкастингке арналған шартты тегін бағдарлама. Movavi экран жазғыш бағдарламасы компьютер экранын толығымен немесе бөлек терезелермен жазуға мүмкіндік 106 береді, сонымен қатар Веб-камера мен микрофонға жазуға мүмкіндік береді. Алынған жазбаны қосымша сәттерді кесу арқылы өңдеуге болады.

Сондай-ақ, Screen Recorder көмегімен скриншоттарды жасауға және жазбаны бейне файл ретінде

Пайдалану ыңғайлылығы. Барлық қажетті функциялар негізгі панельде, ал қалғандарын мәзірден табуға болады;

Интерфейстің қарапайымдылығы. Егер сіз экранды бірінші рет жазып жатсаңыз да, барлық түймелер интуитивті болады;

Танымал форматтарды қолдау. MP4, AVI, MOV, MKV, MP3, GIF;

Функционалдылық. Барлық қарапайымдылығына қарамастан, бағдарламада көптеген пайдалы нұсқалар бар.

Screencastify браузер қосымшасы арқылы сабақты түсіру. Screencast бағдарламалық жасақтамасы жұмыс үстелін немесе толық Windows бағдарламалық жасақтамасын жазуға мүмкіндік береді. Google Chrome-ға Screencastify қосымшсын басты бетке қосу керек. Бұл веб-сайттарды, жұмыс үстелін жазуға немесе веб-камерадан бейнелер түсіруге болатын кеңейтім. Бұл веб-сайт беттерін жазуға арналған керемет қосымша және жұмыс үстелінде скринкаст жасау үшін кейбір бағдарламаларға жақсы балама. Screencastify-тің екі нұсқасы бар. Lite нұсқасы тегін қол жетімді, онда 10 минуттық бейнелерді жаза алады. Google Chrome-ға Screencastify Lite қосу үшін осы бетті ашыңыз. Премиум нұсқасы жылына 24 доллардан сатылады, ол сізге шексіз жазу уақытын береді және бейнені өңдеуге/кесуге қосымша мүмкіндік береді.

Тағы бір видео жазуға арналған бағдарламаның бірі - OBS Studio бағдарламасы. OBS Studio-бүгінгі таңда ағындық саладағы ең жақсы бағдарлама, әдетте, фрагменттерді немесе бүкіл бейнелерді онлайн режимінде жазу (және тарату) үшін қолданылады. OBS Studio-бұл экранды компьютерге түсіруге және жазуға арналған бағдарлама.

OBS studio орнату өте қарапайым екенін атап өткен жөн. Жұмысқа кіріспес бұрын бағдарламаны ресми сайттан жүктеу керек, ол тегін: <https://obsproject.com/ru/download>.

Ең маңызды артықшылығы-Hitbox, Good Game және, әрине, YouTube сияқты ең танымал бейне хостинг платформаларымен үйлесімділік береді. Оның құрамында керемет құралдар жиынтығы бар. Мысалы, ыңғайлы, сіз көріністер арасында тез қалпына келтіре аласыз, сонымен қатар веб-камерадан трансляция жасай аласыз, экранға донаттар мен басқа көмекші элементтерді қосуға болады. Аппараттық жеделдетумен жұмыс істеу мүмкін. Жазбаларды ең өзекті форматтарда сақтауға болады: *.flv және *. mp4.

Қарапайым тегін бағдарламаның бірі, скринкастты тез және қиындықсыз жазуға мүмкіндік береді (экраннан бейне) — oCam Screen Recorder. Бұл бағдарлама қарапайымдылығымен және түсініктілігімен жақсы, ол тез жұмыс істейді.

Жазу аяқталғаннан кейін сіз бірден интернетте орналастыруға болатын дайын бейне файлды (mp4) аласыз. Осылайша сіз нұсқауларды, шолуларды, сабақтарды жазып, экрандағы ақпаратты көрсете аласыз және осының бәрін

дауыстық түсініктемелермен қоса аласыз. Жалғыз ескеретін жайт бұл бағдарламада жазылған бейнені, мысалы, Camtasia Studio-да өндеуге мүмкіндік жоқ. Сонымен, оcam Screen Recorder Бағдарламасын басқару оңай, қысқа интерфейсі бар, ал функционалды және салмағы аз. Сіз ресми сайттан жүктей аласыз — <http://ohsoft.net/eng/> Бұл үйде коммерциялық емес пайдалану үшін ақысыз.

Қай бағдарламаны алмайық өзіндік мүмкіндіктерімен қолданушыларына білімді ерекше ұсына алу тәжірибесін көрсетеді. Бұдан бөлек онлайн жағдайында қандай бағдарламаларды қолданса өз аудиторияңыздың тақырыпқа деген қызығушылықтары артады? Кері байланыс жасау үшін wooclar, ahaslides бағдарламасын қолдансаңыз болады. Әрине бұл бағдарламалар ағылшын тілінде болса да оның мүмкіндігімен аудиторияңызбен тікелей кері байланыс жасау мүмкіндігін туғызады. Мысалы, wooclar сайтында 20-дан көп аудиториямен байланыс орнатсаңыз, ahaslides-та ақылы бағдарлама болғаны үшін тегін қолданылым саны аз, 6 адам ғана өз ойын білдіре алады. Одан көп аудиторияны қамтуға сіз ақша төлейсіз.

Есеп шығаруға, топтық жұмысты бірлесе орындауға арналған электронды тақта қызметін сізгі – Google Jamboard бағдарласы ұсынады. Бұл миға шабуыл жасауға және команданың жұмысын жақсартуға көмектеседі. Интернеттен суреттер мен мазмұнды жүктеп, оларды тікелей тақтаға апарады. Бұл құрал жазбаларды жасауға және бірлесіп жұмыс істеуге, тақталарды араластыруға және әріптестеріңізбен нақты уақыт режимінде өзгерістер енгізуге мүмкіндік береді. Тақырып соңын қортынды жасауға арналған тест құрастыруға арналған сайттарда баршылық. Kahootты формативты бағалауда қолдану, Google-форм суммативты бағалау кезінде қолдануға болады. Kahoot-білім беру жобаларына арналған қосымша. Оның көмегімен сіз тест, сауалнама, оқу ойынын жасай аласыз немесе білім марафонын ұйымдастыра аласыз. Бағдарлама жұмыс үстелінде де, смартфондарда да жұмыс істейді. Google-форм - Google-дің сауалнама нысандары мен тесттер құруға арналған Онлайн – сервисі пайдаланушыға интуициялық түсінікті интерфейс пен жауаптарды автоматты түрде талдай отырып,

Ал Flubaroo қосымшасын орнатқан жағдайда-сауалнама нәтижелерін автоматты түрде тексере отырып, интерактивті сауалнама нысандарын әзірлеу үшін кең мүмкіндіктер береді. Менің ойымша, Google формаларының басты артықшылығы-бұл қызметті тегін пайдалану мүмкіндігі (сізге тек Google есептік жазбасы қажет) және оны өз қалауыңыз бойынша конфигурациялау. Бірнеше минут ішінде сіз сауалнама жасай аласыз, қонақтар тізімін жасай аласыз, жаңалықтар тарату үшін электрондық пошта мекенжайларын жинай аласыз және тіпті викторина жасай аласыз. Қорыта келе, қашықтықтан білім беру технологияларын оқу процесіне енгізу тәжірибесі тыңдаушылар мен студенттер тарапынан да, оқытушылар тарапынан да осы технологияға жоғары сұранысты көрсетеді.

Қазіргі адам өмірінің динамикалық сипаты, халықаралық байланыстардың кеңеюі, сапалы жаңа байланыс құралдары және барлық қызмет салаларының техникалық қанықтылығы белгілі бір географиялық орынға тұрақты сілтеме жасамай, мобильді оқыту қажеттілігін тудырады. Қоғамды ырықтандыру мен ізгілендіруге байланысты оқуға мұқтаж, бірақ аудиторияға баруға мүмкіндігі жоқ азаматтардың жаңа санаттары пайда болады. Мұндай жағдайда қашықтықтан оқытуды дамыту бағдарламасы бұрынғыдан да өзекті. Екінші жағынан, қашықтықтан оқыту ашық білім беру тұжырымдамасын жүзеге асырудың ең қолайлы нысаны болды, ол білім алушы үшін оқу кезінде тандаудың айтарлықтай еркіндігін қамтамасыз етеді.

Әдебиеттер:

1. Сиговцев Г.С., Чарута М.А. Қашықтықтан білім беру. Оқытушыларға арналған әдістемелік материал. Петрозаводск: Изд-во ПетрГУ. 2009. 294 с.
2. Матвиюк В.М. Қоғам мен адам өмірінде ақпараттық-коммуникациялық технологиялардың қолданылуын талдау //В сборнике: Социально-гуманитарные и психологические науки: теоретико-методологические и прикладные аспекты. Материалы Всероссийской научно-практической конференции. 2015. С. 84-86.
3. Муромцева А.В. Қашықтықтан оқыту жүйесіндегі мультимедиялық құралдар // Вестник МГОУ. Серия «Лингвистика». №1, 2011. с.195 — 198

O'zbek tili Semantik analizatori.

Sultonov Shohnazar

sultonovshonazar@gmail.com

ToshDO‘TAU Kompyuter lingvistikasi
mutaxassisligi magistranti

Annotatsiya. Ushbu maqolada o'zbek tiliga moslashtirilgan semantik analizatorni ishlab chiqish va amalda qo'llash masalalari ko'rib chiqilgan. Ilg'or tillarni qayta ishlash texnikasi va lingvistik nazariyalardan foydalangan holda, tadqiqot o'zbek tilining o'ziga xos morfologik va sintaktik xususiyatlarini ko'rib chiqishga qaratilgan. Tadqiqot mavjud muammolar haqida tushuncha beradi, qo'llanilgan usullarni belgilaydi va prototip sinovlaridan olingan natijalarni baholaydi.

Abstract. This article examines the development and practical application of a semantic analyzer adapted to the Uzbek language. Using advanced language processing techniques and linguistic theories, the study aims to examine the specific morphological and syntactic features of the Uzbek language. The study provides an understanding of the existing problems, identifies the methods used, and evaluates the results obtained from prototype testing.

Kalit so'zlar: O'zbek tili, semantik tahlil, tabiiy tilni qayta ishlash, lingvistik morfologiya, hisoblash tilshunosligi.

Kirish

Asosan O'zbekistonda so'zlashadigan turkiy til bo'lgan o'zbek tili o'zining aglutinativ tuzilishi va boy morfologiyasi tufayli semantik tahlil uchun o'ziga xos muammolarni keltirib chiqarmoqda. Sodda sintaktik va morfologik tuzilishga ega bo'lgan tillardan farqli o'laroq, o'zbek tili hisoblash tilshunosligida ma'noni samarali qayta ishlash uchun maxsus yondashuvlarni talab qiladi. Ushbu hujjat o'zbek tilidagi matnlarni qayta ishlash va talqin qila oladigan semantik analizatorni yaratish va joriy etishni o'rganadi, bu esa mashina tarjimasini, hissiyotlarni tahlil qilish va axborot qidirish vazifalarini takomillashtirishga qaratilgan.

Adabiyot Tahlili

Tabiiy tillarni qayta ishlashdagi semantik tahlil ingliz, Mandarin va ispan kabi yirik tillar uchun keng o'rganilgan. Ammo o'zbek tili uchun lingvistik manbalar va hisoblash vositalarining rivojlanishi nisbatan cheklangan. Mavjud tadqiqotlar sintaktik tahlilchilar va morfologik analizatorlarga qaratilgan, ammo semantik ishlov berishga qaratilgan manbalarda bo'shliq mavjud. Tegishli ishlar [Muallif nomi] kabi tadqiqotchilar tomonidan morfologik analizatorlarni ishlab chiqish va tillarni qayta ishlashning boshqa turkiy tillarga tatbiq etilishini o'z ichiga oladi. Semantik rollar va ramka semantikasining hisoblash modellariga integratsiyalashuvi Fillmor va Levin kabi tadqiqotchilar tomonidan o'rganilib, bu nazariyalarni o'zbek tiliga tatbiq etish uchun asos yaratadi.

Tadqiqot usullari

Semantik analizator qoidalarga asoslangan va mashinani o'rganish yondashuvlarining kombinatsiyasi yordamida ishlab chiqilgan:

- Korpusni tayyorlash: izohli O'zbekcha matnlar korpusi tuzilib, unda turli janrlar, jumladan, yangiliklar, adabiyotlar va suhbat matnlari jamlangan.
- Morfologik tahlil: o'zbek tilining murakkab so'z yasalishi, ildiz so'zlar, qo'shimchalar va grammatik belgilarni aniqlash uchun morfologik analizator birlashtirildi.
- Semantik tahlil: semantik rollarni aniqlash imkonini beruvchi so'zlar o'rtasidagi munosabatlarni xaritalash uchun qaramlik daraxtlari qurildi.
- Mashinani o'rganish modellari: nazorat ostidagi o'quv modellari jummalarni tasniflash va semantik nuanslarni aniqlash uchun etiketli ma'lumotlardan foydalangan holda o'qitildi.
- Baholash ko'rsatkichlari: semantik analizatorning ishlashini baholash uchun aniqlik, aniqlik, eslash va F1 ballari ishlatilgan.

Natijalar

O'zbek tili uchun semantik analizatorlar matnni so'zlarni, iboralarini va ularning o'zaro munosabatlarini kontekstda tahlil qilish orqali uning ma'nosini tushunish va izohlash uchun qayta ishlaydigan vositalardir. Bu tarjima, hissiyotlarni tahlil qilish, ma'lumot olish va savollarga javob berish tizimlari kabi turli xil tillarni qayta ishlash dasturlarida juda muhimdir. Quyida o'zbek tili uchun semantik analizatorlar haqida asosiy fikrlar keltirilgan:

Morfologik Tahlil

- Fleksion til: o'zbek tili aglutinativ til bo'lib, grammatik munosabatlar va ma'nolarni etkazish uchun ko'plab affikslardan foydalanadi. Semantik analizator avval so'z tuzilmalarini buzish va tushunish uchun morfologik tahlilni amalga oshirishi kerak.
- Ildiz va affikslar: aniq semantik tahlil uchun ildiz so'z va uning affikslarini tanib olish juda muhimdir.
- Leksik Semantika
- So'z Ma'nosi: analizator so'zlarning ma'nosini, shu jumladan kontekstga asoslangan ko'p ma'noli so'zlarni tushunish uchun boy leksikaga ega bo'lishi kerak.
- Sinonimlar va antonimlar: tizim tushunishni yaxshilash va yanada nozik tahlil qilish uchun sinonim va antonimlarni taniy olishi kerak.
- Sintaksis va tahlil qilish
- Grammatik qoidalar: predmetlar, predikatlar va predmetlar o'rtasidagi munosabatlarni aniqlash uchun jumlar tuzilishini tahlil qilish.
- Bog'liqlikni tahlil qilish: so'zlarning bir-biri bilan qanday bog'liqligini tushunish, izchil jumla ma'nosini hosil qilish.
- Kontekstli Tushunish

- Kontekstli so'zlarni joylashtirish: Bert kabi o'zbek tiliga moslashtirilgan, so'zlarning kontekstga xos ma'nolarini bajara oladigan modellardan foydalanish.
- Ko-mos yozuvlar qarori: turli ifodalar matndagi bir xil mavjudotga qachon murojaat qilishini aniqlash.
- Semantik rollar va munosabatlar
- Rol yorlig'i: harakatlar ma'nosini tushunish uchun gapdagi agent, bemor va vosita kabi semantik rollarni aniqlash.
- Tematik tahlil: matnning asosiy g'oyalarini tushunish uchun so'zlar va iboralarni mavzular yoki tushunchalar bo'yicha guruhlash.
- O'zbek tiliga xos muammolar
- Dialektlar va mintaqaviy tafovutlar: dialektal farqlarni ko'rib chiqish qiyin bo'lishi mumkin, chunki o'zbek tilida turli xil nutq shakllari mavjud.
- Cheklangan tillarni qayta ishlash resurslari: ingliz tili kabi asosiy tillar uchun NLP yaxshi rivojlangan bo'lsa-da, o'zbek tili uchun resurslar nisbatan cheklangan bo'lib, keng qamrovli semantik analizatorlarni yaratishni qiyinlashtiradi.

Ilovalar

- Mashina tarjimasi: chuqurroq semantik tuzilmalarni egallash orqali tarjimalarni takomillashtirish.
- Hissiyotlarni tahlil qilish: matndagi his-tuyg'ular va fikrlarni tushunish.
- Chatbotlar va Virtual yordamchilar: foydalanuvchi ma'lumotlarini o'zbek tilida tushunish orqali aniq javoblarni taqdim etish.

Muhokama

Natijalar shuni ko'rsatadiki, qoidalarga asoslangan tahlil va mashinani o'rganishni birlashtirgan gibrid yondashuv o'zbek tilidagi matnlarning semantik mazmunini qayta ishlash uchun samarali hisoblanadi. Fillmorning ramka semantikasi kabi lingvistik nazariyalarning integratsiyasi kontekstual ma'noni tushunish uchun foydali bo'ldi. Shu bilan birga, tadqiqot yanada kengroq o'quv ma'lumotlari va takomillashtirilgan algoritmlarni talab qiladigan so'zlashuv tili va noyob iboralar bilan ishlash kabi ba'zi cheklovlarni ta'kidladi. O'zbek tiliga xos leksik ma'lumotlar bazalarini yanada kengroq qamrab olish zarurligi yaqqol namoyon bo'ldi.

Xulosa

Tadqiqot o'zbek tili uchun dastlabki semantik tahlil vositasini ishlab chiqish orqali sohaga o'z hissasini qo'shmoqda. Ushbu vosita NLP ilovalaridagi keyingi ishlanmalar, jumladan, mashina tarjimasi, hissiyotlarni tahlil qilish va o'zbek kontenti uchun optimallashtirilgan qidiruv tizimlari uchun asosiy model bo'lib xizmat qilishi mumkin. Kelajakdagi ish o'quv korpusini kengaytirishga, transformatorlar kabi neyron tarmoq arxitekturalarini o'z ichiga olishga va murakkab til tuzilmalarini yanada samarali boshqarish uchun algoritmlarni takomillashtirishga



qaratilishi kerak. Tilshunoslar va til mutaxassislari bilan birgalikdagi sa'y-harakatlar vositaning semantik tushunish qobiliyatini oshirish uchun juda muhimdir.

- O'zbek tiliga xos bo'lgan yanada kengroq, izohli korpuslarni ishlab chiqish.
- -Ishlashni yaxshilash uchun transformatorlar kabi kontekstdan xabardor modellarni birlashtirish.
- Mintaqaviy dialektlarga murojaat qilish va ularning semantik tahlilga ta'siri.

Adabiyotlar.

1. A.A.Abduazizov. Tilshunoslik nazariyasiga kirish. T.,2010.
2. Golovin BN.Tilshunoslikka kirish. "Oliy maktab" nashriyoti. M., 1977 yil.
3. I.Yo'ldoshev, O'.Sharipova. Tilshunoslik asoslari. T.,2017.
4. N.Mahkamov, I.Ermatov. Tilshunoslik terminlarining izohli lug'ati. "Fan", T., 2013.
5. N.Uluqov. Tilshunoslik nazariyasi. T., 2016.

FE'L SO'Z TURKUMIDAGI LEKISIK OMONIMLARNI ANIQLASH YONDASHUVLARI

Malikov Elbek

malikovelbekk@gmail.com

ToshDO'TAU Kompyuter lingvistikasi
mutaxassisligi magistranti

Annotatsiya. Ushbu maqola omonimlikni aniqlash metodlarini ikki asosiy yondashuvga – statistik va qoidalarga asoslangan usullarga ajratadi. Statistik usullar matnlardagi omonimlarni morfologik tahlil orqali aniqlashda qo'llaniladi, unda grammatik xususiyatlarning birgalikda yuzaga kelish statistikasi ishlatiladi. Qoidalarga asoslangan tizimlar esa sintaktik tahlil bosqichida omonimlikni aniqlaydi va ko'proq mutaxassislar ishtirokini talab qiladi. Gibrid yondashuvlar, statistik metodlar va kam sonli qoidalar kombinatsiyasidan tashkil topgan. Tadqiqotlar rus va ingliz tillaridagi matnlarda turli metodlarning samaradorligini ko'rsatadi. Maqola, shuningdek, o'zbek tilida omonimiyani semantik farqlashga qaratilgan yangi yondashuvlarni muhokama qiladi.

Abstract. This article discusses methods for resolving homonymy, categorizing them into statistical and rule-based approaches. Statistical methods are used for morphological analysis to identify homonymy in texts, relying on the co-occurrence statistics of grammatical features. Rule-based systems address homonymy at the syntactic analysis stage and require significant input from specialists. Hybrid approaches combine statistical methods with a limited number of rules. Research demonstrates the effectiveness of various methods in texts in Russian and English. The article also explores new approaches for semantic differentiation of homonyms in the Uzbek language.

Аннотация. Данная статья рассматривает методы устранения омонимии, разделяя их на два основных подхода – статистические и основанные на правилах. Статистические методы используются для морфологического анализа, чтобы идентифицировать омонимию в текстах, опираясь на статистику совместного появления грамматических характеристик. Системы, основанные на правилах, устраняют омонимию на этапе синтаксического анализа и требуют значительного участия специалистов. Гибридные подходы состоят из комбинации статистических методов и небольшого количества правил. Исследования показывают эффективность различных методов на текстах на русском и английском языках. Статья также обсуждает новые подходы к семантическому различению омонимов в узбекском языке.

Kalit soʻzlar: omonimiya, tilshunoslik, statistik tahlil, morfologik tahlil, semantik farqlash, algoritmlar, tilni qayta ishlash, grammatik tahlil.

Kirish. Omonimlar tahlili tilshunoslikning dolzarb masalalaridan biridir, chunki ular tilning koʻplab aspektlarini, xususan, semantik va sintaktik oʻziga xosliklarini oʻrganishda katta ahamiyatga ega. Bitta soʻzning bir nechta maʼnoga ega boʻlishi nafaqat matnlarni tahlil qilish, balki avtomatik tarjima, tilni qayta ishlash (NLP) va kompyuter lingvistikasida muhim muammo hisoblanadi. Omonimiyalarni toʻgʻri farqlash, matnlarning aniq va toʻgʻri talqin qilinishini taʼminlash uchun zamonaviy metodlar va algoritmlar zarur. Statistik usullar, qoidalarga asoslangan tizimlar va gibril yondashuvlar bu masalani samarali hal qilishda asosiy vositalar hisoblanadi.

Oʻzbek tilida omonimlarni tahlil qilishda yuzaga keladigan muammolar va ularni bartaraf etish uchun qoʻllaniladigan metodlar haqida tadqiqotlar davom etmoqda. Bu masala nafaqat tilshunoslar, balki kompyuter olimlari va NLP mutaxassislari uchun ham jiddiy eʼtibor talab qiladi. Shuningdek, ingliz va rus tillaridagi tajribalar asosida oʻzbek tilida ham samarali yondashuvlar ishlab chiqilishi kerak. Omonimlarni aniq farqlash va tahlil qilishda ilmiy yondashuvlar, metodologiyalar va ilgʻor texnologiyalar birlashtirilishi lozim.

Asosiy qism. Omonimlarni farqlash metodlari qoidalarga tayanadigan statistik metodlar va gibril yondashuvlarga boʻlinadi.

Statistik usullar asosida qurilgan tizimlar morfologik tahlil bosqichida matnlarda omonimiyani bartaraf etadi, bunda qoʻl bilan belgilangan holatlardan grammatik xususiyatlarning birgalikda yuzaga kelish statistikasi qoʻllaniladi.

Qoidaga asoslangan tizimlar omonimlarni sintaktik tahlil bosqichida bartaraf etadi. Bunday tizimlarning rivojlanishi juda uzoq vaqt talab qiladigan jarayon boʻlib, soha mutaxassislari jalb qilish talab etiladi.

Gibril yondashuvlar statistik usullarga asoslangan va juda kam sonli qoidalar bilan taʼminlangan boʻladi.

Tajriba uchun rus tili milliy korpusi (RTMK) ishlatilgan. Podkorpusning bir qismida modelni oʻqitish olib borilgan, boshqa qismi esa omonimlarni farqlash usullarini sinab koʻrish uchun ishlatilgan. Ish jarayonida ikkita tajriba oʻtkazildi. Tayanch vektor metodi uchun ham, YMM uchun ham omonimning lokal (mahalliy) konteksti ishlatilgan. Ushbu ishda lokal kontekst ostida ikkitadan ortiq boʻlmagan qoʻshni chap va oʻng soʻz teglari qabul qilinadi.

Birinchi tajriba omonimiyani qisman bartaraf etish uchun oʻtkazildi. Eslatib oʻtamiz, omonimiyaning bu turi ingliz tili matnlariga xos. Ushbu tajriba rus tilidagi

matnlarda omonimiyani bartaraf etish mumkin bo'lgan usullarni aniqlash uchun o'tkazildi.

Ikkinchi tajriba grammatik shakllar majmuini aniqlash uchun o'tkazildi. Bunda so'z turkumlariga kelishik, shaxs-son, zamon, nisbat grammatik shakllari qo'shilgan hamda tajriba amalga oshirilgan. Leksik omonimlarni farqlash uchun ma'no tushunchasi kiritilgan. Ma'no deganda, predmetni aniq tasvirlaydigan satr konstantasi tushuniladi.

Omonimlikni aniqlash usullari belgilangan korpusning bir qismida sinovdan o'tkazildi. Omonimlarga lug'at bo'yicha tegishli bo'lgan teglar va lemma tanlandi. So'ngra har bir usul bilan bir ehtimoliy tegni aniqlab, korpusdagi teg bilan solishtirib ko'rildi.

Birinchi tajriba uchun so'z qismi to'g'ri aniqlangan bo'lsa, omonimiya bartaraf etilgan deb hisoblandi. Ikkinchi tajriba uchun yuqorida bayon etilgan grammatik xususiyatlar va bartaraf etilgan leksik omonimiyani ko'rsatkichi sifatida urg'u to'g'ri kelsa, omonimiyani bartaraf etish natijasi to'g'ri hisoblangan bo'ladi. Tajriba davomida olingan natijalar 1 va 2-jadvallarda berilgan. 1-jadvalda qisman omonimiyani, 2-jadvalda esa to'liq omonimiyani bartaraf etish natijalari aks ettirilgan¹⁶:

Metod	Qisman to'g'ri mansubligini aniqlash, %		
	barcha so'zlar	leksik omonimiyani hisobga olgan holda barcha so'zlar	notanish so'zlar
YaMM	92,78 93,56	90,56	87,32
Support vektor metodi		90,98	89,13

2.1-jadval. Rus tili matnlarida qisman omonimiyani bartaraf etish metodi natijalari

Metod	Teglarning to'g'riligini aniqlash, %		
	barcha so'zlar	leksik omonimiyani hisobga olgan holda barcha so'zlar	notanish so'zlar
YaMM	87,73	84,39	64,25
Support vektor metodi	86,12	82,89	62,41

2.2-jadval. Rus tili matnlarida omonimiyani aniqlash metodlari natijalari

¹⁶ Порохнин А. Анализ статистических методов снятия омонимии в текстах на русском языке // <http://www.mathnet.ru/vagtu286>

Shunday qilib, har ikkala metod ham rus tilidagi matnlarda omonimiyani qisman bartaraf etish ishini uddalay oladi. Biroq rus tilidagi matnlarda qisman omonimiyani bartaraf etish natijalari ingliz tilidagi matnlarga qaraganda ancha past. Rus tilidagi matnlarda omonimiyani aniqlash uchun Yashirin Markov modeli Support vektor metodiga qaraganda aniqroq natija beradi.

Omonimlarni semantik farqlashning ko'plab zamonaviy usullari bilan tajriba o'tkazishda nafaqat teglarni, balki o'ziga xos iboralarni ham qo'llash va omonimlarni turiga qarab bosqichma-bosqich bartaraf etishdan foydalanish ko'zda tutiladi.

M.Abjalova dissertatsiyasida omonimlarni tahlil qilish texnologiyasiga to'xtalib o'tadi. U Brill usuli va Yashirin Markov modelining vazifasi hamda afzalliklarini yoritadi. O'zbek milliy korpusining yaratilmaganligi, rasmiy va ilmiy uslubdagi matnlar bazasi to'la shakllantirilmaganligi e'tiborga olinib, o'zbek tilidagi omonim shakllarni tahlil qilishda N-grammga bog'liq ravishda omonim-tahlil usuli qo'llanishini ta'kidlaydi. Omonim-tahlil usuli muayyan omonim so'zga boshqa so'zning ma'lum grammatik ko'rsatkich orqali bog'lanishi bilan tushuntiriladi va bu jarayon tahlil imkoniyati hamda tahlilning aniqligini oshirishi ko'rsatiladi¹⁷.

Ingliz tilidagi matnlarda omonimiyani bartaraf etishni qiyoslash maqsadida ikki metod tanlab olingan: Support vektor metodi(SVM) va Yashirin Markov modeli (Hidden Markov Model HMM). Bu metodlar lug'atda bo'lmagan so'zlar uchun omonimiyani aniqlash imkoniyatini beradi.

Mashhur Yashirin Markov modeli (HMM) omonimlarni ehtimoliy yondashuv asosida aniqlashga qaratilgan. Algoritmning asosiy g'oyasi shundan iboratki, jumlagi kiritilgan har bir so'z uchun quyidagi funktsiyani maksimallashtiradigan grammatik sinfni (tegni) tanlash kerak:

$$P(\text{word}|\text{tag}) * P(\text{tag}|\text{previous } n \text{ tags})$$

$P(\text{tag}|\text{previous } n \text{ tags})$ – shartli ehtimollik, ushbu teg oldingi n teg aniqlangandan keyin hosil bo'ladi.

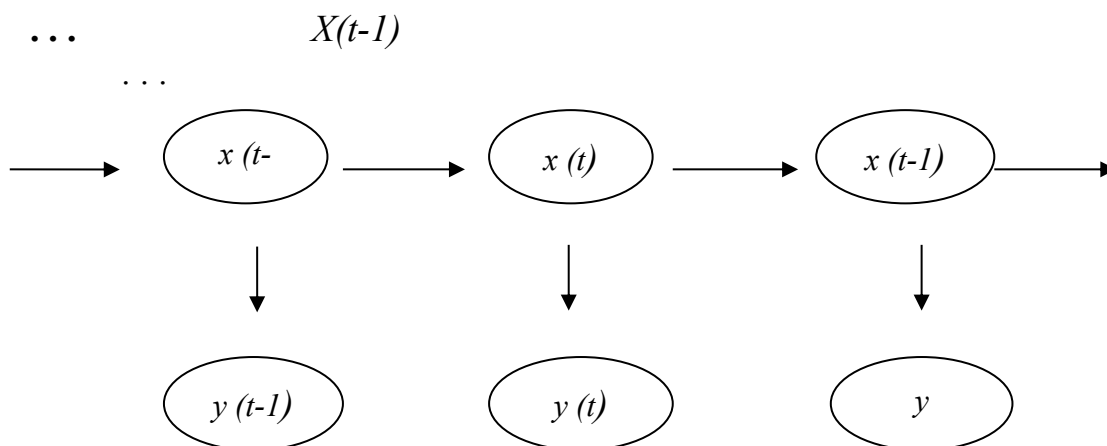
$P(\text{word}|\text{tag})$ – shartli ehtimollik, berilgan joyda hosil bo'lgan *word* so'zi, berilgan grammatik sinf tegiga ega.

Yashirin Markov modeli – noma'lum parametrlarga ega bo'lgan jarayonning ishlashini simulyatsiya qiluvchi statistik model. Noma'lum parametrlarni o'zgaruvchan kuzatuvlar asosida hal qilish vazifa qilib qo'yiladi¹⁸.

¹⁷ Абжалова М.А. Ўзбек тилидаги матнларни таҳрир ва таҳлил қилувчи дастурнинг лингвистик модуллари (Расмий ва илмий услубдаги матнлар таҳрири дастури учун). Филол. фан. бўйича фалсафа доктори (PhD) диссер. – Фарғона, 2019. – 164 б.

¹⁸ Ветров Д. Скрытые марковские модели //

http://www.machinelearning.ru/wikiimages/8/83/GM12_3.pdf.



3.1-rasm. YMMning umumiy strukturasi.

Bunda $x(t)$ yashirin o'zgaruvchining qiymati faqat $x(t-1)$ yashirin o'zgaruvchining qiymatiga bog'liq ekanligi ravshan bo'ladi. Bu bog'liklik (tobelik) "Markov xususiyati" deb ataladi. Shu bilan birga, kuzatilgan o'zgaruvchining qiymati $y(t)$ faqat yashirin o'zgaruvchining $x(t)$ qiymatiga bog'liq bo'ladi. Lug'atda bo'lmagan so'zlarni tahlil qilish uchun uch harfli postfikslar ishlatiladi. Natijada, morfologik xususiyatlarni aniqlash muammosini hal qilish yashirin qiymatlarning eng munosib qiymatlari ketma-ketligini hisoblash uchun kamayadi:

$$u'(t) = \operatorname{argmax} P(y|x)$$

YMM algoritmi juda yuqori hisoblash murakkabligiga ega (klassik Viterbi dinamik dasturlash usuli bilan amalga oshiriladi, shunga o'xshash Levenshteyn masofani hisoblash uchun ishlatiladi) va amalda, odatda, uni kamaytirish uchun turli soddalashtiruvchi taxminlar ishlatiladi (masalan, uch so'zli ketma-ketliklardan ortiq bo'lmisligi kerak). Ingliz tili uchun YaMM algoritmining aniqligi 96%.

Morfologik noaniqlikni bartaraf etishning yana bir yondashuvi deterministik va ehtimoliy taggerlarning eng yaxshi xususiyatlarini birlashtiradigan yondashuvdir. Bunday turdagi eng mashhur yondashuvlardan biri – TBL (Transformation-Based Learning) tagger yoki Brillning morfologik analizatori (Brill tagger)¹⁹.

Bu usul mashinasozlik g'oyalariga asoslangan bo'lib, unda grammatik sinflarni so'zlarga tayinlash qoidalari oldindan belgilangan korpusdan avtomatik ravishda chiqariladi va keyin deterministik algoritmlar qayta ishlangan matnga qo'llaniladi.

Ushbu algoritmnini o'rgatishda, avvalo, korpuslarning belgilangan barcha so'zlariga eng ehtimoliy teg qiymatlari beriladi. Keyin barcha mumkin bo'lgan teglar (ishlab chiqarish qoidalari) o'rganiladi; (allaqachon belgilangan korpus nuqtai

¹⁹ Cristopher D. Manning, Hinrich Schutze. Foundation of Statistical Natural Language Processing, 1999.

nazaridan) tahlil sifatini maksimal yangilaydiganlari qoldiriladi. O'qitish natijasida yangi holatlarni qayta ishlashda yana qo'llash mumkin bo'lgan qoidalarning tartiblangan ro'yxati (ajralish tartibi) yaratiladi²⁰.

Ko'rinadiki, ingliz tilida matnlarda omonimiyani bartaraf etish masalasida ehtimoliy metodlardan yuqori darajada foydalanilgan. Ingliz tilida gapda so'zlar ketma-ketligi aniq bo'lib, mazkur usullarni qo'llash soddalashtirilgan. Tadqiqotlardan kuzatiladiki, ingliz tili matnlarida omonimiyani bartaraf etish metodlaridan Yashirin Markov modeli va tayanch vektor metodi yetakchilik qiladi.

Sh. Gulyamova dissertatsiyasida omonimlarni turli so'z turkumlari doirasidagi guruhlariga ajratgan. Ularni ikki so'z turkumi doirasidagi omonim so'zlarni farqlashda “Qoidalarga asoslangan teglash usuli”dan foydalanishni taklif qilgan²¹. Ushbu teglash usuli aniq qoidalarga tayanadi. Ya'ni ikkita so'z turkumi doirasidagi omonim so'zlarning ma'nolarini farqlash ma'lum qoidalar orqali amalga oshiriladi.

Ingliz olimi Markov bir turkum doirasidagi omonimlikni aniqlashda so'z turkumlarini teglashtirish va ehtimollar nazariyasidan foydalanishni taklif qilgan. Bizning tadqiqotimizda o'zbek tilidagi fe'l so'z turkumi doirasidagi omonimlarni farqlashda mana shu metodlardan o'rinli va maqsadli foydalanildi.

O'zbek tilining xalqimiz ijtimoiy hayotida va xalqaro miqyosdagi obro'-e'tiborini tubdan oshirish, unib-o'sib kelayotgan yoshlarimizni vatanparvarlik, milliy an'ana va qadriyatlarga sadoqat ruhida tarbiyalash, mamlakatimizda davlat tilini to'laqonli joriy etishni ta'minlash, O'zbekistondagi millat va elatlarning tillarini saqlash va rivojlantirish, davlat tili sifatida o'zbek tilini o'rganish uchun shart-sharoitlar yaratish, o'zbek tili va til siyosatini rivojlantirishning strategik maqsadlari, ustuvor yo'nalish va vazifalarini hamda istiqboldagi bosqichlarini belgilash maqsadida²² bir qator chora-tadbirlar belgilanishi ona tilimiz mavqeini yanada ko'tarishga, uning amaliy qo'llanishini kengaytirishga qaratilgan.

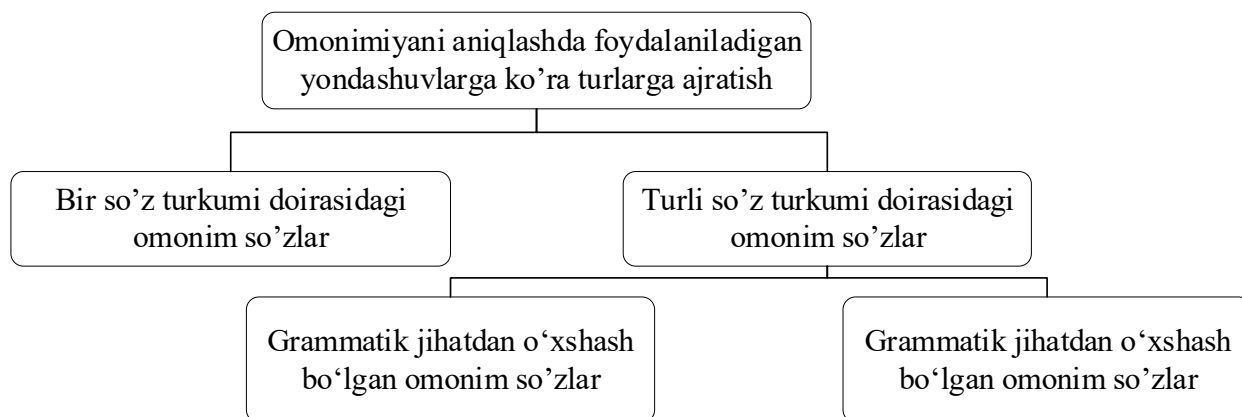
Gulyamova tomonidan taklif etilgan lingvistik ta'minot bo'yicha tadqiqotchi X.Axmedova o'zining “O'zbek tilidagi gaplarni semantic analiz qilishning modeli, algoritmlari va axborot tizimini ishlab chiqish” deb nomlangan PhD dissertatsiyasi avtoreferatida omonim so'zlarni semantik farqlashga alohida

²⁰ Зеленков Ю.Г., Сегалович И.В., Титов В.А. Вероятностная модель снятия морфологической омонимии на основе нормализующих подстановок и позиций соседних слов // http://www.dialog-21.ru/media/2444/zelenkov_segalovich.pdf

²¹ Gulyamova Sh. O'zbek tili semantic analizatori ishlab chiqishning lingvistik ta'minoti. Filol.fan.doktori (DcS) disser. –Farg'ona, 2022.

²² O'zbekiston Respublikasi Prezidenti Shavkat Mirziyoyevning “Mamlakatimizda o'zbek tilini yanada rivojlantirish va til siyosatini takomillashtirish chora-tadbirlari to'g'risida”gi PF-60-84-son farmoni // <https://lex.uz/docs/-5058351>

to'xtalgan²³. Olima omonim so'zlarni semantik farqlashda ularni so'z turkumlari doiraga ajratgan holda turlicha yondashuvlardan foydalangan. Ya'ni bir so'z turkumi, turli so'z turkumlari doirasidagi omonim so'zlarni farqlash uchun turlicha algoritmlarni taklif qilgan.



3.2-rasm: Omonim so'zlarni ularni farqlovchi yondashuvlarga ko'ra guruhleri

X.Axmedova bir so'z turkumi doirasidagi omonim so'zlarni farqlashda, ularni birikuvchilari yordamida farqlash mumkinligi ta'kidlab, YMM modelini taklif qilgan. Turli so'z tukumlari doirasidagi omonim so'zlarni ham o'z navbatida ikki guruhga ajratgan: grammatik jihatdan o'xshash va o'xshash bo'lmagan so'z turkumlari doirasidagi omonimlarni alohida ajratgan. X.Axmedova grammatik jihatdan o'xshasha bo'lmagan omonim so'zlarni tilning grammatik qoidalari asosida farqlovchi matematik modellar ishlab chiqqan. Grammatik jihatdan o'xshash bo'lgan so'z turkumlari doirasidagi omonim so'zlarni farqlovchi qat'iy qoida mavjud emas deya baholaydi va bu turdagi omonimlarni farqlashda statistik usullardan foydalanishni taklif qilgan.

Xulosa. Omonimlarni aniqlash va ularni to'g'ri tahlil qilish tilshunoslik va NLP sohasida katta ahamiyatga ega. Ingliz va rus tillarida bu masala uchun samarali metodlar va algoritmlar ishlab chiqilgan bo'lsa-da, o'zbek tilida bu soha hali ham rivojlanish jarayonida. O'zbek tilida omonimlarni aniqlash va farqlash uchun ilg'or statistik metodlar, morfologik tahlil usullari, va gibrid yondashuvlarni qo'llash kerak. Ushbu metodlar tilshunoslik va kompyuter fanlarining integratsiyasini ta'minlaydi va tilning yangi imkoniyatlarini ochadi.

²³ Axmedova X. O'zbek tilidagi gaplarning semantik analiz qilishning modeli, algoritmlari va axborot tizimini ishlab chiqish// texnika fanlari falsafa doktori (PhD) disser. – Toshkent, 2023. 120-b.

Omonimlarni samarali bartaraf etish uchun ko'p jihatdan texnologik yondashuvlarning va ilmiy izlanishlarning davom ettirilishi muhim. Shuningdek, o'zbek tilining o'ziga xos xususiyatlarini hisobga olish orqali, yangi va innovatsion metodlar ishlab chiqilishi lozim. Bu jarayonda tilshunoslar, kompyuter olimlari va dasturchilarning hamkorligi tilshunoslik va NLP sohasida katta yutuqlarga olib kelishi mumkin. Omonimiyalarni to'g'ri tahlil qilish, tilni avtomatik qayta ishlashda yanada aniq va samarali natijalarga erishish imkonini beradi.

Foydalanilgan adabiyotlar

1. Hamroyeva Sh. O'zbek tili morfologik analizatorining lingvistik. – GlobeEdit, 2020;
2. Гулямова Ш. Ўзбек тили семантик анализаторининг лингвистик асослари: Филол.фан.док-ри...дисс... – Фарғона, 2022. – 215 б.;
3. Hamroyeva Sh. O'zbek tili mualliflik korpusini tuzishning lingvistik asoslari: Filol.fan. bo'yicha falsafa doktori (PhD)...diss. – Qarshi, 2018.–250 b;
4. Abjalova M.A. O'zbek tilidagi matnlarni tahrir va tahlil qiluvchi dasturning lingvistik modullari (Rasmiy va ilmiy uslubdagi matnlar tahriri dasturi uchun): Filol.fan.bo'yicha falsafa doktori (PhD)...diss. – Farg'ona, 2019. – 164b;
5. Axmedova D. Atov birliklarini o'zbek tili korpuslari uchun leksik- semzoantik teglashning lingvistik asos va modellari.Filologiya fanlari bo'yicha fal.dok-ri.diss.- Buxoro, 2020. – 163 b.;
6. Bakayev I.I. «O'zbek tili so'z shakllarini morfologik tahlil qilish modellari va algoritmlari»: texnika fan. fan. bo'yicha falsafa doktori (PhD)...diss.avtoref. – Toshkent, 2021. –12-22 b.;
7. X.Axmedova. O'zbek tilidagi gaplarni semantik analiz qilishning modeli, algoritmlari va axborot tizimini ishlab chiqish// texnika fanlari falsafa doktori (PhD)... diss.avtoref.– Toshkent, 2023.
8. Нурмонов А. Ўзбек тилшунослиги тарихи. – Тошкент: Ўзбекистон, 2002. – 177 б. – Б. 83.

TABIIY TILGA ISHLOV BERISH TIZIMLARI

Sultonov Shohnazar

sultonovshonazar@gmail.com

ToshDO‘TAU Kompyuter lingvistikasi
mutaxassisligi magistranti

Annotatsiya: Mazkur maqola tabiiy tilni ishlov berish tizimlariga bag'ishlangan bo'lib, bu tizimlarning ilmiy va amaliy ahamiyatini yoritadi. Tabiiy tilni ishlov berish (TTIB) so'zi kompyuterlar va boshqa sun'iy intellekt tizimlari yordamida inson tilini tushunish, qayta ishlash va tarjima qilish jarayonlarini o'z ichiga oladi. Maqolada TTIB tizimlarining asosiy konseptlari, texnologiyalari va usullari tahlil qilingan. Shuningdek, TTIB tizimlarining turli sohalarda, jumladan, tilni avtomatik tarjima qilish, matnni tahlil qilish, xabarlarni tushunish va sentiment tahlili kabi amaliy qo'llanilishini ko'rib chiqishadi. Tizimlarning samaradorligini oshirish uchun ishlatiladigan zamonaviy usullar, xususan, mashina o'rganish va chuqur o'rganish texnologiyalari haqida ham so'z yuritiladi. Maqolada tilni ishlov berish tizimlarining rivojlanishi va ularning kelajakdagi imkoniyatlari ham o'rganiladi. TTIB sohasidagi eng yangi yutuqlar, ularning ijtimoiy, madaniy va texnologik sohalarda qo'llanilishi hamda yuzaga kelayotgan muammolar ham maqolada batafsil ko'rib chiqiladi.

Abstract: This article is devoted to natural language processing systems and highlights the scientific and practical significance of these systems. The term natural language processing (NLP) includes the processes of understanding, processing and translating human language using computers and other artificial intelligence systems. The article analyzes the main concepts, technologies and methods of NLP systems. It also considers the practical application of NLP systems in various fields, including automatic language translation, text analysis, message understanding and sentiment analysis. Modern methods used to improve the efficiency of systems, in particular, machine learning and deep learning technologies, are also discussed. The article also studies the development of language processing systems and their future potential. The latest achievements in the field of NLP, their applications in social, cultural and technological spheres, as well as emerging problems are also considered in detail in the article.

Kalit so'zlar: tizimli analiz, tilni qayta ishlashning amaliy dasturlari, ko'p tildagi tizimlar, tabiiy tilni yaratish, tilni qayta ishlash

Keywords: systems analysis, language processing applications, multilingual systems, natural language generation, language processing

Kirish. Tabiiy tilga ishlov berish (TTIB) tizimlarining muvaffaqiyati, katta qismda, matnlarning ma'nosini to'g'ri tushunish qobiliyatiga bog'liq. Semantik analiz – TTIB ning muhim tarkibiy qismi bo'lib, matnning semantik tuzilishini, so'zlar va iboralarning ma'nosini, ular orasidagi munosabatlarni aniqlashga xizmat qiladi. Ushbu bobda ko'rib chiqilganidek, semantik analizning turli usullari – leksik-semantik resurslardan foydalanish, kontekstual tahlil, mashinaviy o'rganish algoritmlari – matnni tushunishda muhim rol o'ynaydi. Ammo, O'zbek tili kabi kam resursli tillar uchun semantik analizni amalga oshirishda hali ham muammolar mavjud. Lexik-semantik resurslarning yetishmasligi, murakkab grammatik tuzilmalar va ko'p ma'noli so'zlarning ko'pligi semantik analizning aniqligi va samaradorligiga ta'sir qiladi. Kelajakdagi tadqiqotlar O'zbek tili uchun yuqori aniqlikdagi semantik analizatorlarni yaratishga, mavjud resurslarni boyitishga va yangi, zamonaviy usullarni qo'llashga qaratilishi lozim.

Asosiy qism. Tabiiy tilga ishlov berish (Natural Language Processing - NLP) tizimlari, kompyuterlarning inson tilini tushunishi, tahlil qilishi va ishlab chiqishi uchun ishlatiladi. Bu sohada quyidagi asosiy yo'nalishlar mavjud:

1. *Tilni tushunish (Natural Language Understanding - NLU):* Bu inson tilidagi ma'lumotlarni analiz qilish va ularni kompyuter tushunadigan formatga aylantirish jarayonidir. Masalan, savolni aniqlash yoki kontekstni tushunish.

2. *Tilni ishlab chiqarish (Natural Language Generation - NLG):* Bu jarayon kompyuterga inson tilida ma'lumotlar yaratish imkonini beradi. Masalan, avtomatik matn yozish yoki xulosalar tayyorlash.

3. *Matnni tahlil qilish (Text Analysis):* Bu jarayonda matn ichidagi ma'lumotlarni, tuzilmani va kontekstni tahlil qilish uchun turli metodlar va algoritmlar qo'llaniladi.

4. *Muloqot tizimlari (Conversational Agents):* Chatbotlar va virtual yordamchilar bu toifaga kiradi. Ular foydalanuvchilar bilan interaktiv muloqot qilish imkoniyatini beradi.

5. *Mashina tarjimasi (Machine Translation):* Bu texnologiya bir tildagi matnni boshqa tilga avtomatik ravishda tarjima qilish uchun ishlatiladi, masalan, Google Translate.

6. *Ovozli tanish va generatsiya:* Ovozli buyruqlarni tanish va ovozli javoblarni ishlab chiqish.

Quyida tabiiy tilga ishlov berish sohasining ushbu yo'nalishlariga alohida to'xtalib o'tamiz.

Tilni tushunish (Natural Language Understanding - NLU) — bu tabiiy tilga ishlov berishning bir bo'limi bo'lib, kompyuterlarning inson tilidagi matn yoki nutqni tushunishi va tahlil qilishini ta'minlaydi. NLU jarayonlari quyidagi asosiy komponentlardan iborat:

1. **Maqsadni aniqlash:** Foydalanuvchining so'rovi yoki savolining maqsadini aniqlash. Masalan, foydalanuvchi "Bugun havo qanday?" deb so'raganda, bu savol havo ma'lumotlariga qiziqish bildiradi.
2. **Obyektlarni tanib olish:** Matndagi muhim elementlarni, masalan, odamlar, joylar yoki vaqtlarni aniqlash. Bu jarayon "Entity Recognition" deb ataladi.
3. **Sentiment tahlili:** Matnning ruhiyatini yoki hissiyotini aniqlash. Misol uchun, ijobiy, salbiy yoki neytral munosabatlarni tushunish.
4. **Sintaktik tahlil:** Matnning grammatik tuzilishini tahlil qilish. Bu jarayonda jumladagi so'zlar va ularning o'zaro aloqalari tahlil qilinadi.
5. **Semantik tahlil:** Matnning ma'nosini chuqurroq tushunish. Bu bosqichda so'zlar orasidagi bog'lanishlar va kontekst o'rganiladi.
6. **Kontekstni tushunish:** Ushbu jarayon, matnning oldingi va keyingi qismlarini hisobga olgan holda, foydalanuvchining maqsadini aniqlashni o'z ichiga oladi.

Tabiiy tilni ishlab chiqarish (Natural Language Generation - NLG) — bu tabiiy tilga ishlov berishning bir qismi bo'lib, kompyuterlarning ma'lumotlarni inson tilida ifodalashini ta'minlaydi. NLG tizimlari avtomatik ravishda matn yaratish uchun turli jarayonlarni o'z ichiga oladi. Asosiy komponentlar quyidagilardan iborat:

1. Ma'lumotni tahlil qilish: Avval, NLG tizimi uchun kerakli ma'lumotlar yig'iladi va tahlil qilinadi. Bu bosqichda qaysi ma'lumotlar kerakligini aniqlash muhimdir.

2. Rejalashtirish: Bu bosqichda tizim ma'lumotlardan qanday xulosa chiqarish kerakligini va qanday jummalarni yaratishi kerakligini rejalashtiradi. U qanday tuzilishga ega bo'lishi, qanday ma'lumotlarni ta'kidlashi kerakligi haqida qaror qabul qiladi.

3. Boshqaruv (Scripting): Rejaga muvofiq, tizim turli matn elementlarini va jummalarni birlashtiradi. Bu bosqichda grammatika va stil jihatlari inobatga olinadi.

4. Tilni ishlab chiqarish: Ushbu jarayon, tanlangan ma'lumotlarni, rejalashtirilgan tuzilmani va boshqaruvni birlashtirib, inson tilida to'g'ri va tushunarli matn yaratishni o'z ichiga oladi.

5. Post-ishi: Yaratilgan matnni tahrir qilish va yaxshilash. Bu bosqichda ko'proq ravonlik, mantiqiylik va grammatik to'g'rilik nazorat qilinadi.

NLG tizimlari ko'plab sohalarda qo'llaniladi, masalan:

- Avtomatik xulosa tayyorlash: Ma'lumotlar to'plamidan xulosalar chiqarish, masalan, hisobotlar va statistikalar.
- Chatbotlar va virtual yordamchilar: Foydalanuvchilarga javob berish uchun matn yaratish.
- Ovozli yordamchilar: Ovozli javoblar ishlab chiqarish.
- Ma'lumotlar vizualizatsiyasi: Murakkab ma'lumotlarni oddiy va tushunarli tarzda taqdim etish.

Matnni tahlil qilish (Text Analysis) — bu matndagi ma'lumotlarni, tuzilishni va ma'noni tahlil qilish jarayoni. Ushbu jarayon tabiiy tilga ishlov berish (NLP) ning

muhim qismi bo'lib, ko'plab sohalarda qo'llaniladi, jumladan, axborot qidirish, sentiment tahlili, va muloqot tizimlarida. Matnni tahlil qilish jarayoni bir necha asosiy bosqichlardan iborat:

1. *Preprocessing (dastlabki ishlov berish)*: Matnni tahlil qilishdan oldin, uni tayyorlash zarur. Bu bosqichda quyidagi jarayonlar amalga oshiriladi:

- Tozalash: Keraksiz belgilar va so'zlarni olib tashlash (masalan, nuqtalar, raqamlar, turli belgilar, internet linklari, html teglar).

- Tokenizatsiya: Matnni so'z yoki iboralar kabi kichik birlashtiruvchi qismlarga bo'lish.

- Stemming yoki lemmatization: So'zlarni ularning ildiz shakliga qaytarish yoki grammatika jihatidan to'g'ri shaklga keltirish.

2. *Tahlil qilish*: Bu bosqichda turli metodlar va algoritmlar yordamida matnning ma'lum jihatlari tahlil qilinadi. Masalan:

- *Nomlangan obyektlarni aniqlash (Named Entity Recognition)*: Matndagi nomlangan obyektlarni ya'ni odamlar, joylar, ko'chalar, binolar yoki tashkilotlar kabi obyektlarni ajratib olish.

- *Sentiment tahlili*: Matnning ijobiy, salbiy yoki neytral ruhiyatini aniqlash.

- *Sintaktik tahlil*: Matnning grammatik tuzilishini o'rganish.

3. *Ma'lumotlarni vizualizatsiya qilish*: Tahlil qilingan ma'lumotlar vizual shaklda taqdim etilishi mumkin, masalan, grafiklar yoki diagrammalar yordamida.

4. *Natijalarni chiqarish*: Tahlil natijalari asosida xulosalar chiqarish va ulardan foydalanish. Bu bosqichda foydalanuvchilarga yoki tizimlarga ma'lumotlar bilan bog'liq qarorlar qabul qilishda yordam beriladi.

Matnni tahlil qilish jarayonlari ko'plab sohalarda qo'llaniladi, masalan:

- *Marketologiya*: Mijozlar fikrlarini va sharhlarini tahlil qilish.

- *Ijtimoiy media monitoringi*: Foydalanuvchilarning postlarini va fikrlarini tahlil qilish.

- *O'qituvchilar va ilmiy tadqiqotchilar*: Matnlarni tahlil qilish orqali ilmiy tadqiqotlar va natijalarni chiqarish.

Matnni tahlil qilish jarayonlari uchun ko'plab vositalar va kutubxonalar mavjud, masalan, Python dasturlash tilida NLTK, spaCy, va TextBlob.

Muloqot tizimlari yoki suhbat agentlari — sun'iy intellekt asosida ishlaydigan dasturiy ta'minot bo'lib, ular odamlar bilan tabiiy tilda muloqot qilishga mo'ljallangan. Bu tizimlar ko'pincha chatbotlar, virtual yordamchilar yoki muloqot botlari sifatida tanilgan va turli vazifalarni bajarishga yordam beradi. Quyida muloqot tizimlarining turlarini ko'rishimiz mumkin:

- *Qoidaga asoslangan agentlar*: Ushbu agentlar oldindan belgilangan qoidalar va sxemalar asosida ishlaydi. Ular aniq skriptlar va muloqotlar ketma-ketligini bajaradi, lekin kutilmagan yoki yangi savollarga javob bera olmaydi.

- *Sun'iy intellekt asosidagi agentlar*: Ushbu agentlar tabiiy tilga ishlov berish (NLP) texnologiyalari yordamida ishlaydi. Ular foydalanuvchi maqsadini aniqlash,

mashinali o'qitish va chuqur o'qitish algoritmlaridan foydalanib, murakkab va noaniq savollarga javob bera oladi.

Muqolqot tizimlarini turli sohalarida qo'llash mumkin:

Xizmat ko'rsatish: Kompaniyalar muloqot agentlarini mijozlarga xizmat ko'rsatish uchun ishlatib, tez va qulay javoblar taqdim etishadi.

Savdo va marketing: Chatbotlar mahsulotlar haqida ma'lumot berish, mijozlarni jalb qilish va savdo jarayonini avtomatlashtirish uchun qo'llaniladi.

Ta'lim: Virtual yordamchilar talabalarga mavzuni tushuntirish, mashqlar berish va savollariga javob berishda yordam beradi.

Sog'liqni saqlash: Tibbiyotda bemorlarga maslahat berish, dori-darmonlar bo'yicha eslatmalar berish va diagnostik ma'lumotlar yig'ish uchun ishlatiladi.

Muloqot tizimlari yordamida insonlar bilan doimiy va samarali muloqot qilish imkoniyatlari kengayib, ularni turli sohalarida qo'llash imkoniyati yanada ortmoqda.

Mashina tarjimasi — bu kompyuter dasturlari yordamida matn yoki nutqni bir tildan boshqa tilga avtomatik ravishda tarjima qilish jarayonidir. Bu texnologiya sun'iy intellekt va tabiiy tilga ishlov berish (NLP) sohalaridagi rivojlanishlarga asoslanadi. Mashina tarimasini uch xil yondashuv asosida amalga oshirish mumkin:

- *Qoidalarga asoslangan mashina tarjimasi (Rules based on Machine Translation RBMT):* Ushbu yondashuv oldindan belgilangan grammatik qoidalar va lug'at asosida ishlaydi. U manba tilining grammatik tuzilmalarini maqsad tiliga o'zgartiradi. Bu usul aniq va qattiq qoidalarga asoslangan, ammo yangi yoki murakkab iboralar bilan ishlashda qiyinchiliklarga duch kelish mumkin.

- *Statistik mashina tarjimasi (Statistical Machine Translation SMT):* Bu yondashuv katta hajmdagi ikki tilli korpuslardan statistik ma'lumotlarni yig'ib, tarjima jarayonini amalga oshiradi. SMT biror jumlaning qaysi ehtimol bilan qanday tarjima qilinishini aniqlash uchun statistik modellarni qo'llaydi.

- *Neural mashina tarjimasi (Neural Machine Translation NMT):* Bu yondashuv chuqur o'qitish texnologiyalaridan foydalangan holda ishlaydi. NMT tizimlari neyron tarmoqlar yordamida tilni tushunish va tarjima qilishni o'rgatadi, bu esa ularning kontekstni tushunish va yanada tabiiyroq tarjimalar qilish imkoniyatini beradi.

Mashina tarjimasi tizimini ishlab chiqishning afzalliklari quyidagilardan iborat:

- *Tezlik va samaradorlik:* Mashina tarjimasi qo'l bilan tarjima qilishga qaraganda ancha tezroq bo'ladi, bu katta hajmdagi matnlarni qisqa vaqt ichida tarjima qilish imkonini beradi.

- *Qulaylik:* Bu texnologiya turli tabiiy tillardagi foydalanuvchilar o'rtasida aloqa o'rnatishni osonlashtiradi va global muloqot uchun qulay sharoit yaratadi.

- *Doimiy o'rganish va yaxshilanish:* Nevron tarmoq modellari asosida ishlovchi tizimlar o'z tajribalaridan o'rganib, vaqt o'tishi bilan o'z tarjima sifatini yaxshilay oladi.

Mashina tarjimasini amalga oshirish jarayonida ham qator qiyinchiliklar va cheklovlar mavjud:

- *Til va madaniyat farqlari:* Ba'zan matnni to'g'ri tarjima qilish uchun tilning nozikliklari va madaniy kontekstni tushunish zarur bo'ladi, bu esa mashina tarjimasini uchun qiyinchilik tug'dirishi mumkin.
- *Qiyin iboralar va murakkab grammatik tuzilmalar:* Mashina tarjimasini tizimlari murakkab sintaksis va ko'p ma'noli so'zlar bilan ishlashda xatolarga yo'l qo'yishi mumkin.
- *Korpus va ma'lumot yetishmovchiligi:* Tarjima tizimining sifatli ishlashi uchun katta hajmdagi yuqori sifatli o'quv ma'lumotlari talab etiladi, lekin har doim ham bunday ma'lumotlar mavjud bo'lmasligi mumkin.

Mashina tarjimasining turli sohalarda qo'llash mumkin:

Internet tarjima xizmatlari: Google Translate, DeepL, Microsoft Translator kabi mashhur onlayn tarjima xizmatlari dunyo bo'ylab millionlab foydalanuvchilar uchun tezkor va qulay tarjima imkoniyatlarini taqdim etadi.

Ovozli tarjima qurilmalari: Portativ qurilmalar yoki smartfon dasturlari yordamida biror nutqni real vaqt rejimida boshqa tilda tinglash va tushunish imkoniyati mavjud.

Professional tarjima yordamchilari: Ba'zi dasturlar professional tarjimonlarga qo'shimcha vosita sifatida ishlatiladi, ularning ish jarayonini tezlashtiradi va ularga tahliliy ma'lumotlar taqdim etadi.

Mashina tarjimasini kelajakda yanada rivojlanib, yanada aniqroq, tabiiyroq va samaraliroq tarjimalar qilish imkoniyatiga ega bo'lishi kutilmoqda. Jahonda ko'plab tabiiy tillarda bunday tizimlar ishlab chiqilgan va bugungi kunda foydalanilmoqda. NLP tizimlari odatda mashinali o'qitish (machine learning) va sun'iy intellekt (AI) metodlari yordamida rivojlanadi.

Xulosa: Ta'qdim etilgan ma'lumotda matnni tahlil qilish va muloqot tizimlari, ularning asosiy tamoyillari, qo'llanilishi va sohalari o'rganildi. Matnni tahlil qilish, tabiiy tilga ishlov berish (NLP)ning asosiy qismi bo'lib, axborot qidirish, sentiment tahlili va muloqot tizimlarida muhim o'rin tutadi. U bir necha bosqichlardan iborat: dastlabki ishlov berish, tahlil qilish, ma'lumotlarni vizuallashtirish va natijalarni chiqarish.

Matnni tahlil qilish jarayonida turli texnikalar qo'llaniladi, jumladan, matnni tozalash, tokenizatsiya, stemming yoki lemmatizatsiya, nomlangan obyektlarni aniqlash, sentiment tahlili va sintaktik tahlil. Bu texnikalar ma'lumotlarni yig'ish, tahlil qilish va xulosalar chiqarishda yordam beradi.

Muloqot tizimlari esa, sun'iy intellektga asoslangan dasturiy ta'minot bo'lib, insonlar bilan tabiiy tilda muloqot qilish imkonini beradi. Ular qoidalarga asoslangan va sun'iy intellektga asoslangan agentlarga bo'linadi. Sun'iy intellektga asoslangan agentlar NLP texnologiyalari yordamida murakkab va noaniq savollarga



javob berishga qodir. Muloqot tizimlari xizmat ko'rsatish, savdo, marketing, ta'lim va sog'liqni saqlash kabi turli sohalarda keng qo'llanilmoqda.

Foydalanilgan adabiyotlar

1. Abdurahmonova N. Automatic morphological analyze for English-Uzbek system // Известия Кыргызский государственный технический университет им. И.Раззакова. Теоретический и прикладной научно-технический журнал. № 2 (38) – Бишкек, 2016. – С. 12-18.;
2. Mengliyev B., Hamroyeva Sh. Turkiy tillardagi morfologik analizatorlar xususiyada// O'zbek tilini dunyo miqyosida keng targ'ib qilish bo'yicha hamkorlik istiqbollari // Xalqaro ilmiy-amaliy anjuman materiallari. – Toshkent, 2020, oktabr. – B. 360-364;
3. Hamroyeva Sh. O'zbek tili morfologik analizatorining lingvistik. – GlobeEdit, 2020;
4. Elov B. Hamroyeva SH. MORFOLOGIK ANALIZATORNI YARATISH USULLARI//Til va madaniyat 2022. 4-son.

AVTOMATIK LINGVISTIK VA EKSTRALINGVISTIK TEGlash VOSITALARI UMUMIY TAVSIFI

UDK:81:33

Yuldashev Aziz Uyg'un o'g'li
yuldashevaziz@navoiy-uni.uz
ToshDO 'TAU o'qituvchisi

Annotatsiya. Ushbu maqolada avtomatik lingvistik va ekstralingvistik teglash vositalari umumiy tavsifi keltiriladi. Mystem, pymorphy2 kabi avtomatik morfologik teglash vositalari; Stanford CoreNLP, spaCy avtomatik sintaktik teglash vositalari; WordNet, ruWordNet kabi avtomatik semantik teglash vositalari tavsiflanadi. Korpus tuzish, interfeys va qidiruv tizimidan foydalanish uchun ham tayyor dasturiy vositalar bo'lib, Sketch Engine, AntConc kabi vositalarning umumiy xususiyatlari, imkoniyatlari, afzalliklari tahlil qilinadi.

Kalit so'zlar: avtomatik lingvistik teglash, avtomatik ekstralingvistik teglash, Mystem, pymorphy2, Stanford CoreNLP, spaCy, WordNet, ruWordNet, Sketch Engine, AntConc.

Автоматические лингвистические и экстралингвистические средства тегирования: общий обзор

Абстракт. В данной статье дается общее описание инструментов автоматической лингвистической и экстралингвистической маркировки. Инструменты автоматического морфологического тегирования, такие как Mystem, pymorphy2; Stanford CoreNLP, инструменты автоматической синтаксической разметки spaCy; Описаны инструменты автоматической семантической разметки, такие как WordNet и ruWordNet. Представлены готовые программные инструменты для создания корпусов, интерфейс и использование поисковых систем, а также проанализированы общие характеристики, возможности и преимущества таких инструментов, как Sketch Engine и AntConc.

Ключевые слова: автоматическая лингвистическая разметка, автоматическая экстралингвистическая разметка, Mystem, pymorphy2, Stanford CoreNLP, spaCy, WordNet, ruWordNet, Sketch Engine, AntConc.

Automatic Linguistic and Extralinguistic Tagging Tools: A General Overview

Abstract. This article provides a general description of automatic linguistic and extralinguistic tagging tools. Automatic morphological tagging tools such as Mystem, pymorphy2; Stanford CoreNLP, automatic syntactic markup tools spaCy;

Automatic semantic markup tools such as WordNet and ruWordNet are described. Ready-made software tools for creating corpora, interface and use of search engines are presented, and general characteristics, capabilities and advantages of such tools as Sketch Engine and AntConc are analyzed.

Keywords: automatic linguistic markup, automatic extralinguistic markup, Mystem, pymorphy2, Stanford CoreNLP, spaCy, WordNet, ruWordNet, Sketch Engine, AntConc.

KIRISH

Til korpuslarini teglashda Mystem, pymorphy2 kabi avtomatik morfologik teglash vositalari; Stanford CoreNLP, spaCy avtomatik sintaktik teglash vositalari; WordNet, ruWordNet kabi avtomatik semantik teglash vositalari ahamiyatli va zaruriy vositalar sanaladi.

Shuningdek, korpus tuzish, interfeys va qidiruv tizimidan foydalanish uchun ham tayyor dasturiy vositalar bo'lib, Sketch Engine, AntConc kabi vositalarning umumiy xususiyatlari, imkoniyatlari, afzalliklari tahlil qilinadi.

ASOSIY QISM

Lingvistik teglash tabiiy tilni qayta ishlashda muhim bosqich bo'lib, bu jarayonda ko'plab yordamchi vositalar mavjud. Quyida shunday dasturiy vositalarning tavsifi keltiriladi.

1. Avtomatik morfologik teglash vositalari:

Mystem: Yandeks tomonidan ishlab chiqilgan, Ingliz va rus tilining morfologik tahlilini taqdim etadi. Tez va aniq ishlaydi.

pymorphy2: morfologik tahlil uchun Python kutubxonasi. Rus, ukrain va boshqa tillarini qo'llab-quvvatlaydi. Moslashuvchan va foydalanish oson.

2. Avtomatik sintaktik teglash vositalari:

Stanford CoreNLP: Stanford universitetining tabiiy tilni qayta ishlash uchun vositalar to'plami. Sintaktik tahlil (parsing) dasturi bo'lib, nomlangan obyektlarni aniqlash (Named Entity Recognition) va boshqa vositalarni o'z ichiga oladi, ko'p tillarni qo'llab-quvvatlaydi.

spaCy: tabiiy tilni qayta ishlash uchun ilg'or Python kutubxonasi. Tez va samarali ishlaydi, sintaktik tahlil (parsing), nomlangan obyektlarni aniqlash (Named Entity Recognition) va boshqa funksiyalarni taqdim etadi.

3. Avtomatik semantik teglash vositalari:

WordNet: Ingliz tilining leksik ma'lumotlar bazasi. So'zlarni ularning ma'nolari va munosabatlari bo'yicha guruhlaydi. Semantik tahlil va matnni qayta ishlashda ishlatiladi.

ruWordNet: WordNet ma'lumotlar bazasining ruscha versiyasi.

Mazku ma'lumotlar bir qancha tillarni avtomatik morfologik, sintaktik va semantik tahlil qilsa-da, o'zbek tiliga samarali qo'llab bo'lmaydi, chunki o'zbek tili morfologik va sintaktik xususiyatlariga ko'ra ingliz va rus tillaridan tubdan farq qiladi. Shu sababli o'zbek tili birliklarini sifatli avtomatik tahlil qilish uchun dasturlarga ehtiyoj mavjud.

Korpus tuzish, interfeys va qidiruv tizimidan foydalanish uchun ham tayyor dasturiy vositalar mavjud:

Rus tili milliy korpusi (NKRYA): Tilshunoslik belgilari bilan rus tili matnlarining ulkan korpusi. Matnni qidirish va tahlil qilish uchun kuchli vositalarni taqdim etadi. Tilshunoslik belgilarining turli darajalarini ko'rish imkonini beradi.

Sketch Engine: Matn korpuslari bilan ishlash uchun onlayn vosita. Qidirish, tahlil qilish va ma'lumotlarni vizuallashtirish funksiyalarini taqdim etadi.

AntConc. AntConc – matn korpuslarini tahlil qilish uchun bepul vosita. U lingvistik tadqiqotlarda, ayniqsa, korpus lingvistikasida so'zlarning chastotasi, kollokatsiyalari, konkordanslar va boshqa til hodisalarini tahlil qilishda keng qo'llaniladi. AntConc-ning asosiy xususiyatlari va qo'llanish sohalari quyidagicha ko'rsatish mumkin:

Asosiy funksiyalari:

Konkordans: korpusdagi muayyan so'z yoki iboralarni qidirish, ularni kontekstda ko'rsatish imkonini beradi.

Chastotali tahlil: korpusdagi so'zlarning chastotasini ko'rsatadigan ro'yxatni taqdim etadi.

Kollokatsiyalar: byerilgan so'z bilan tez-tez uchraydigan so'zlarni ko'rsatadi.

Klasterlar: korpusdagi takrorlanuvchi so'zlar ketma-ketligini aniqlaydi.

N-grammalar: n-so'zlar ketma-ketligini tahlil qiladi.

Asosiy so'zlar: turli korpuslardagi so'zlarning chastotasini solishtirish va asosiy so'zlarni aniqlash imkonini beradi.

AntConc dasturining qo'llanish sohalari sifatida quyidagilarni ko'rsatish mumkin, shuningdek, mazkur ro'yxat uning ahamiyatini ham ko'rsatadi:

Lingvistik tadqiqotlar: lyeksika, grammatika va uslubni tahlil qilish, kollokatsiyalar va frazeologik birliklarni o'rganish, diskursni tadqiq qilishda foydalaniladi.

Xorijiy tillarni o'rganish: horijiy til ta'limida o'quv qo'llanmalari va topshiriqlarni ishlab chiqishda til materialini tahlil qilish, o'rganilayotgan tildagi so'z va kollokatsiyalarning chastotasini kuzatish, tarjimada yordam berishga qaratilgan.

Matnlarni tahlil qilish: matnlardagi asosiy mavzular va tushunchalarni aniqlash, muallif uslubini tahlil qilish, katta hajmdagi matn ma'lumotlarini tahlil qilishda qo'l keladi.

AntConc dasturining afzalliklari sifatida bepul va foydalanish uchun qulayligi, sodda va intuitiv tushunarli interfeysga egaligi, matn ma'lumotlarini tahlil qilish uchun kuchli vosita ekanligi, Windows, Mac va Linuxda ishlay olishini

ta'kidlash zarur. AntConc matn ma'lumotlari bilan ishlaydiganlar, ayniqsa, lingvistika va chet tillarni o'rganish sohasidagilar uchun juda foydali vositadir.

Avtomatik lingvistik teglash vositani tanlashda til, teglash turi va ishlash talablarini hisobga olish lozim. Vositalardan maksimal darajada samarali foydalanish uchun hujjatlarni va foydalanish imkoniyatlarini, yo'riqnomani o'rganish talab etiladi. Yaxshi natijalarga erishish uchun avtomatik va qo'lda belgilash usullarining kombinatsiyasidan foydalanish mumkin.

Turli janrdagi matnlarni avtomatik ekstralingvistik (nolisoniy omillarga ko'ra) tegash murakkab, lekin amalga oshirish mumkin bo'lgan vazifadir. Bunda quyidagi asosiy jihatlarni hisobga olish kerak:

Avtomatlashtirish imkoniyatlari:

1. Metama'lumotlar: ko'p matnlar allaqachon avtomatik ravishda olib tashlanishi mumkin bo'lgan metama'lumotlarni o'z ichiga oladi. Masalan:

Veb-sahifalar: nashr etilgan sana, muallif(lar), sarlavha. Yangiliklar maqolalari: manbalar, kategoriyalar, vaqt belgilari.

Ilmiy maqolalar: muallif(lar), jurnallar nom(lar)i, nashr etilgan sanalar.

Veb-skreypping va matnni qayta ishlash vositalari ushbu ma'lumotni avtomatik ravishda olib tashlash uchun ishlatilishi mumkin.

2. Janrni aniqlash. Mashina o'rganish modellari matnning lingvistik xususiyatlari va metama'lumotlari asosida matn janrini avtomatik ravishda aniqlash uchun o'rgatilishi mumkin. Matn janrini yetarli darajada aniqlik bilan topa oladigan matn tasnifi algoritmlari mavjud.

3. Bo'yoqdorlik va emotsionallikni tahlil qilish. Zamonaviy NLP vositalari matnning bo'yoqdorgini, shuningdek, uning emotsional xususiyatini avtomatik ravishda aniqlashga imkon beradi. Bu muallifning matnga yoki hikoya mavzusiga bo'lgan munosabati haqida xulosa chiqarishga imkon beradi.

4. Nomlangan obyektlarni aniqlash. Nomlangan obyektlarni aniqlash vositalari odamlar, joylar, tashkilotlar va ekstralingvistik teglash uchun ahamiyatli bo'lgan boshqa obyektlar haqidagi ma'lumotlarni avtomatik ravishda olib tashlashi mumkin.

5. Ijtimoiy tarmoqlarni tahlil qilish. Ijtimoiy tarmoqlar ma'lumotlari foydalanuvchilar, ularning qiziqishlari va ijtimoiy munosabatlari haqidagi ma'lumotlarni olish uchun avtomatik ravishda tahlil qilinishi mumkin.

Matnning ekstralingvistik xususiyatlarini avtomatik tahlil qilish va aniqlashning ayrim qiyinchilik va cheklovlari mavjud bo'lib, ular quyidagilar:

Janrlar xilma-xilligi. Turli janrdagi matnlar turli xususiyatlarga va ekstralingvistik teglash talablariga ega. Barcha janrlarni qayta ishlay oladigan universal vositani yaratish murakkab vazifadir.

Metama'lumotlarning yetishmasligi. Hamma matnlar ham to'liq va aniq metama'lumotlarni o'z ichiga olmaydi. Ba'zi hollarda metama'lumotlar mavjud bo'lmasligi yoki noto'g'ri bo'lishi mumkin.

Kontekst va talqin. Ekstralingvistik tellashning ba'zi jihatlari kontekstni chuqur tushunishni va talqin qilishni talab qiladi. Avtomatik tizimlar bunday vazifalarda qiyinchiliklarga duch kelishi mumkin.

Subyektivlik. Ekstralingvistik teglashning bir qismi subyektiv bo'lishi mumkin. Masalan, hissiyo bo'yoqdorlik va sentiment tahlil bunga misol bo'ladi.

Kuzatishlardan ma'lum bo'ldiki, ekstralingvistik teglash avtomatlashtirish usullari va yondashuvlari quyidagicha tasniflanadi:

1. Gibrid tizimlardan foydalanish, ya'ni avtomatik va qo'lda teglash usullaridan aralash holda foydalanish eng samarali usul bo'lishi mumkin. Avtomatik tizimlar asosiy ma'lumotlarni olish uchun ishlatilishi mumkin, keyin esa ekspertlar natijalarni tekshirishi va tuzatishi mumkin.

2. Mashinali o'rganish (Machine learning) yondashuvidan foydalanish, ya'ni ekstralingvistik ma'lumotlarni avtomatik ravishda olish va tasniflash uchun mashinali o'rganish modellaridan foydalanish bugungi kunda samalari natijalar berayotgani kuzatiladi.

3. Ma'lumotlar bazalari va ontologiyalarni qo'llash, ya'ni ekstralingvistik ma'lumotlarni saqlash va tashkil qilishda ma'lumotlar bazalari va ontologiyalardan foydalanish mumkin.

XULOSA

Xulosa sifatida aytish mumkinki, turli janrdagi matnlarni to'liq avtomatik ekstralingvistik teglash murakkab, lekin rivojlanayotgan sohadir. Zamonaviy vositalar va usullar bu jarayonning ba'zi jihatlarini avtomatlashtirishga imkon beradi, lekin yuqori aniqlik va to'liqlikka erishish uchun ko'pincha avtomatik va qo'lda teglash usullari kombinatsiyasidan foydalanish talab qilinadi.

Foydalanilgan manbalar:

1. <https://yandex.ru/dev/mystem/>
2. <https://pymorphy2.readthedocs.io/>
3. <https://github.com/stanfordnlp/CoreNLP>
4. <https://spacy.io/>
5. <https://wordnet.princeton.edu/>
6. <https://ruwordnet.ru/>
7. <https://ruscorpora.ru/>
8. <https://www.sketchengine.eu/>
9. <https://www.laurenceanthony.net/software/antconc/>



MUNDARIJA

SO‘Z BOSHI

Sirojiddinov Sh.S. O‘zbek tilini raqamlashtirish – davr talabi	4
Jamoldinova O.R. Kompyuter lingvistikasi: ta’lim yo‘nalishi va ilmiy izlanish sohalari haqida	5
Xalikov B.A. Kompyuter lingvistikasini rivojlantirish: ehtiyoj va zarurat	6
Normamatov S.E. O‘zbek tilining elektron resurslarini yaratish – zamon talabi	7

YALPI YIG‘ILISH MA’RUZALARI

Eʃref Adali Türkçenin Mantığı	8
Ashirboyev S., Turdialiyev A. O‘zbek shevalari leksikasini korpusga joylashtirishning dasturiy ta’minoti	24
Abdurahmonova N.Z., Vasliddinova K. O‘zbek tili parallel korpusining lingvistik masalalari	30
Toirova G.I. Oronimik qurilmalarning milliy korpusda berilish masalasi	39
Elov B.B., Xudayberganov N.U. O‘zbek tili korpusi matnlarini pos teglash usullari	47

TABIY TILNI QAYTA ISHLASH (NLP)

Sharipov M.S., Samandarova D.SH. O‘zbek tili matnlarini morfologik tahlil qilish uchun lingvistik manba yaratish va lemmalash	55
Ochilov M.M, Jurayev D.B, Narzullayev O.O. Matndagi tinish belgilarini tiklash muammolari va yechimlari	60
Sobirov O.O. O‘zbek tilidagi so‘zlarning lemmasini topish	67
Xudayberganov N.U. O‘zbek tili korpusiga morfologik ishlov berish	72
Jumanazarov M.D., Madatov X.A. Bert til modelidan foydalangan holda o‘zbek tili lingvistik muammolarini yechish	79
Baxodirov S.R. Matn tozalash tizimi ayrim kategoriyalarining tavsifi	83
Berdiyev J.B. Matnlarni tasniflash uchun Bert modeli va Tensorflow kutubxonasining o‘rni	90
Jumaqulova N.Q. NLPda koreferensiya masalasi	97

KORPUS LINGVISTIKASI

Qarshiyev A.B., Karimov S.A., Tursunov M.S. O‘zbek tilidagi so‘zlarni morfologik tahlil qiluvchi dastur	101
Elov B.B., Abdullayeva O.X. O‘zbek tili korpusini sintaktik teglash masalasi ..	108
Gulyamova Sh.Q., Nurboyeva M.V. Frazelogik polisemiya va uning korpus tilshunosligida modellashtirish masalalari	115
Elov B.B., Xusainova Z.Y. Til korpuslarini lingvistik teglash bosqichlari	122
Xolmuxamedov B.F. Qur’oni karim ma’nolari tarjimasi va tafsirining til korpusi – dasturiy ta’minoti interfeysi va qidiruv tizimidan foydalanish	129



Akramova Sh.I. Mualliflik korpusi keng qamrovli qidiruv tizimi	137
Yuldashev A.U. Til korpusi matnlari chastotasi, konkordans va KWIC	142
Yodgorov U.S. Korpus matnlarini teglash muammolari	148
Nuritdinov A.S. Jadid davri adabiy muhitiga doir asarlardan korpusda foydalanish	155
Masharipova B.R., Qutlimuratova B.X. O'zbek tili bolalar nutqi audio korpusini yaratish	162
O'dayeva Sh.Sh. O'zbek tili korpusida teglash metodlari va usullari	167
Alimbekova M.X. Abdurauf Fitrat asarlari korpusi matnlarini to'plash va ularga dastlabki ishlov berish muammolari	172
Qurbondurdiyeva L.M., Rahimbayeva M.D., Kuriyozov E.R. O'zbek tilidagi omonim so'zlar korpusini yaratish	178
Maxammadiyeva D.R. Mualliflik korpuslarini yaratish bosqichlari (Usmon Azim dramalarining til korpusi dasturining yaratilishi misolida)	182
Amirqulov M.A. Corpus processing tools in building multilingual corpora	187
Xudayarova S.Sh. Lingvistik annotatsiya turlari va uning til korpuslaridagi ahamiyati	192
Xolmonova I.A. Parallel korpus tuzishda stemming va lemmatizatsiyaning ahamiyati	198
Achilova N.X. O'lchov birliklarining milliy korpusda berilishi	208
Bekmuradova I.Z. Panxronik korpus tavsifi, imkoniyat va afzalliklari (rus tili panxronik korpusi misolida)	212
Muradova N.B. Zahiriddin Muhammad Bobur mualliflik korpusining tuzilishi va tarkibi	218
Muftillayeva R.T. Dialektal korpusning arxitekturasini va ma'lumotlar bazasi ...	226
Nizomova F.B., Qodirov K.Q. Til korpusida sodda gaplarni avtomatik aniqlash	234
MATNGA QAYTA ISHLOV BERISH TIZIMLARI	
Sharipov M.S., Amanbayeva R.X., Jumanazarova O.M. O'zbek tili so'zlarining stemini topish uchun lingvistik manba yaratish	245
Khayatova Z.M. Algorithms behind uzbek corpus paraphrasing: a closer look at linguistic support and methodologies	250
Matyakubova N. Understanding the distinction: word and phrase alignment ...	256
Abdullayeva O.X., Nizomova F.B., Qodirov K.Q. O'zbek tilida gap bo'laklari tartibini avtomatik aniqlash (sodda gaplar misolida)	262
Primova M.H. Matn ma'lumotlaridan kerakli ma'lumotlarni olish	269
Allanazarova S.Y. Matnlarning aspektga asoslangan hissiy tahlili	274
Sobirova N.G'. Tahrirlovchi dastur (spellchecker) texnologiyasi	279
Qodirova M. Mashina tarjimasini uchun ingliz va o'zbek tillarida grammatik ma'no ifodalashning o'ziga xosliklari	287
Mahmudjonova G.U. O'zbek matnlari uchun count vectorizer imkoniyatlari va kamchiliklari	293

Qosimova S.F. n-grammga asoslangan stemming algoritmini ishlab chiqish, tatbiq qilish va baholash	298
Bekmuradova I.Z. Tilshunoslikda “sleng” tushunchasi va unga yondosh hodisalar	305
Mamasoyilova S.M. O‘zbek tilida hislarni ifodalovchi birliklar va ularning ahamiyati	310
SEMANTIK TIZIMLAR VA LINGVISTIK ONTOLOGIYALAR	
Sharipov M.S., Mamatova M.D. O‘zbek tilida fe’l so‘z turkumi uchun VERBNET yaratish	315
Abdurahmanova M.T. Sun’iy intellekt tizimida omonimlarni tushunish va tarjima qilish muammolari	319
Hamroyeva Sh.M., Hamroqulova M.N. Matn birliklarini semantik teglashning ahamiyati	324
Xabibullayeva D.Q. Issues of implementing machine interpreting into the sphere of medicine	329
O‘rinova Z.J. Intertekstual tezauruslarning mental lug‘at sifatidagi ahamiyati ...	335
Po‘latova G.G’. O‘zbek tili tezaurusi uchun sifat leksemalarning semantik munosabatlari	340
Axmedova A. Zamonaviy tilshunoslik terminologiyasi muammolari	345
KOMPYUTER LEKSIKOGRAFIYASI VA LINGVODIDAKTIKASI	
Rahimbayeva M.D., Rajapboyeva Y.I. Turk tili va Xorazm shevalari o‘xshash leksemalarining qiyosiy-tarixiy va etimologik lug‘atini elektron shaklda yaratish	349
Мұратбек Б.Қ. Гуманитарлық білім берудегі ақпараттық технологиялар ...	354
Бисенгали І.Б. Қазақ және түрік тіліндегі ортақ лексика сипаты	358
Кенебаева Ұ. Ш. Көне түркі ескерткіштерінің грамматикасы	365
Абирова Ж. Д. Цифрлық білім беру ресурстарын оқытуда қолдану	385
Жисува Э.С. Возможности информационных технологий для активизации познавательной деятельности учащихся	390
Barnoyeva Z.S. Ilkokul oğrencileri için elektronik sözlük oluşturma önemi ..	398
Кальжанова А.Е. Оқытуда қолданылатын онлайн веб-ресурстар	402
Молдабаева З.М., Жолдасова Н.Ж. Қазақ тілі сабақтарында ақпараттық-коммуникациялық технологияларды қолданудың тиімділігі	406
Акимбекова А.М. Қашықтан оқу жағдайында қолданатын бағдарламалар	412
Sultonov Sh. O‘zbek tili Semantik analizatori.....	416
Malikov E. Fe’l so‘z turkumidagi lexisik omonimlarni aniqlash yondashuvlari..	420
Sultonov Sh. Tabiiy tilga ishlov berish tizimlari.....	428
Yuldashev A.U. Avtomatik lingvistik va ekstralingvistik teglash vositalari umumiy tavsifi	435