

SIFAT SO‘Z TURKUMINI TUZILISHIGA KO‘RA POS TEGGLASH

Mansurova Shaxinabonu Najmiddin qizi

Kompyuter lingvistikasi mutaxassisligi magistranti

tiladabiyot@gmail.com

ToshDO‘TAU

Annotatsiya. Ushbu tezisdagi sifatlar nutqning bir qismi sifatida POS (Part of Speech) morfologik tuzilishiga asoslangan holda ko‘rib chiqiladi. Tadqiqotda sifatlar, jumladan, sodda, hosila va qo‘shma sifatlar yasallishiga ko‘ra turkumlarga ajratilib, gap tuzilishidagi roli o‘rganiladi. Tadqiqot tabiiy tilni qayta ishlash (NLP) sohasidagi yutuqlarga hissa qo‘shadigan avtomatlashtirilgan lingvistik tahlilda POS belgilarining samaradorligini ta’kidlaydi. Sifatlarning tuzilishi va ularning morfologik funksiyalarini tahlil qilish orqali ushbu tadqiqot ularning tarkibiy xilma-xilligi va lingvistik tadqiqotlar va mashinali o‘rganish modellarida amaliy qo‘llanilishi haqida tushuncha beradi.

Annotation. This thesis examines adjectives as a part of speech based on their morphological structure using POS (Part of Speech) tags. The research categorizes adjectives, including simple, derived, and compound adjectives, according to their formation and investigates their role in sentence structure. The study highlights the effectiveness of POS tags in automated linguistic analysis, contributing to advancements in the field of Natural Language Processing (NLP). By analyzing the formation of adjectives and their syntactic functions, the research provides insights into their structural diversity and practical application in linguistic research and machine learning models.

Аннотация. В данной диссертации при помощи тегов части речи (POS) рассматриваются прилагательные как часть речи на основе их морфологической структуры. Исследование делит прилагательные, включая простые, производные и сложные прилагательные, на категории в зависимости от их образования и исследует их роль в структуре предложения. Исследование подчеркивает эффективность тегов POS в автоматизированном лингвистическом анализе, внося вклад в достижения в области обработки естественного языка (NLP). Анализируя образование прилагательных и их синтаксические функции, данное исследование предоставляет понимание их структурного разнообразия и практического применения в лингвистических исследованиях и моделях машинного обучения.

Kalit so‘zlar: *Sifat, gap bo‘lagi, morfologik tuzilishi, POS belgilari, sodda sifatlar, hosila sifatlar, qo‘shma sifatlar, lingvistik tahlil, sintaksis, tabiiy tilni qayta ishlash (NLP).*



Kirish. Sifatlar narsa-predmetlar belgisi haqida ma'lumot berib, tilda hal qiluvchi rol o'ynaydi. Nutqning asosiy qismlaridan biri sifatida, sifatlar jumla tuzilishi, ma'nosi va stilistik ifodasiga hissa qo'shadi. Sifatlarning morfologik tuzilishiga ko'ra tasniflanishi, masalan, sodda, hosila va qo'shma sifatlar tilshunoslarga ularning tilda shakllanishi va vazifasini tushunishga yordam beradi. Sifat so'z turkumi til tizimida muhim o'rin egallovchi leksik-grammatik kategoriyalardan biri hisoblanadi. U predmetning belgisi, xususiyati, holati va sifatini ifodalash orqali nutqning aniqligi va obrazlilikini ta'minlaydi. Sifatlar odatda ot bilan bog'lanib, aniqlovchi vazifasini bajaradi va gapning semantik tuzilishini boyitadi. Shu jihatdan ularni lingvistik hamda kompyuter lingvistikasi nuqtai nazaridan o'rganish dolzarb masalalardan biridir.

Hisoblash tilshunosligining rivojlanishi bilan POS(nutqning bir qismi) yorlig'i lingvistik tahlil uchun muhim vositaga aylandi, so'zlarni avtomatlashtirilgan aniqlash va tasniflash imkonini beradi. POS yorlig'i sifatlarni tuzilishiga ko'ra tizimli ravishda tasniflash imkonini beradi, bu esa tabiiy tilni qayta ishlash (NLP) ilovalarida, masalan, mashina tarjimasini, hissiyotlarni tahlil qilish va matn yaratishda yordam beradi. Zamonaviy tabiiy tilni qayta ishlash tizimlarida matnlarni avtomatik tahlil qilishda so'z turkumlarini aniqlash, ya'ni POS (Part of Speech) teglash muhim bosqich hisoblanadi. POS teglash jarayonida har bir so'zga uning grammatik funksiyasini ifodalovchi teg biriktiriladi. Bu esa matnni strukturaviy va semantik jihatdan chuqur tahlil qilish imkonini beradi. Ayniqsa, o'zbek tilida sifatlarni aniqlashda ularning morfologik tuzilishini hisobga olish muhim ahamiyat kasb etadi.

Sifatlarni tuzilishiga ko'ra tasniflash ularning shakllanish mexanizmini ochib berishga xizmat qiladi. O'zbek tilshunosligida sifatlar asosan sodda, yasama va qo'shma turlarga ajratiladi. Ushbu tasnif POS teglash jarayonida ham muhim mezonlardan biri sifatida qo'llaniladi, chunki har bir tur o'ziga xos morfologik va semantik belgilar bilan ajralib turadi.

Sodda sifatlar tarkiban bitta leksik asosdan iborat bo'lib, qo'shimchalarsiz qo'llanadi. Ular tilning eng qadimiy va faol qatlamini tashkil etadi. Masalan, katta, kichik, oq, qora, yaxshi kabi sifatlar sodda sifatlar hisoblanadi. Bunday birliklar ko'pincha lug'aviy bazalarda alohida qayd etilgan bo'ladi va POS teglash jarayonida ularni aniqlash nisbatan oson kechadi. Sodda sifatlar ko'p hollarda nutqda tez-tez qo'llanilishi sababli avtomatik tizimlarda yuqori aniqlik bilan belgilanadi.

Yasama sifatlar esa boshqa so'z turkumlariga mansub birliklardan turli so'z yasovchi qo'shimchalar yordamida hosil qilinadi. O'zbek tilida -li, -siz, -dor, -chan kabi affikslar sifat yasashda faol ishtirok etadi. Masalan, bilimli, umidsiz, mehnatdor, ishchan kabi birliklar yasama sifatlarga kiradi. Ushbu turdagi sifatlarni aniqlashda morfologik tahlil alohida ahamiyatga ega bo'lib, so'zning tarkibiy qismlarini ajratish orqali uning sifat ekanligi aniqlanadi. POS teglash tizimlarida

yasama sifatlarni aniqlash uchun qo'shimcha modellar yoki qoidaviy algoritmlar qo'llaniladi.

Qo'shma sifatlarni ikki yoki undan ortiq leksik birlikning birikishi orqali hosil bo'ladi. Ular ko'pincha murakkab ma'no ifodalaydi va nutqda obrazlilikni oshiradi. Masalan, ishbilarmon, ochko'z, tezkor kabi sifatlarni qo'shma sifatlarni jumlasiga kiradi. Ba'zi hollarda qo'shma sifatlarni alohida yozilgan birliklar shaklida ham uchraydi. POS teglash jarayonida bunday birliklarni aniqlash nisbatan murakkab bo'lib, ular ko'p hollarda kontekstual tahlil asosida belgilanadi. Shu sababli zamonaviy NLP tizimlarida qo'shma sifatlarni aniqlash uchun neyron tarmoqlarga asoslangan modellardan foydalanilmoqda.

Sifatlarni POS teglash jarayonida ularning faqat tuzilishi emas, balki semantik xususiyatlari ham muhim rol o'ynaydi. Ma'no jihatidan sifatlarni asliy va nisbiy turlarga bo'linadi. Asliy sifatlarni predmetning bevosita belgisi yoki xususiyatini bildiradi, nisbiy sifatlarni esa boshqa predmet yoki hodisaga bog'liq holda shakllanadi. Bu farqlanish ham avtomatik teglash jarayonida hisobga olinishi zarur, chunki ayrim hollarda bir xil shakldagi so'zlar turli grammatik vazifalarni bajarishi mumkin.

POS teglash tizimlarida sifatlarni aniqlash uchun bir necha yondashuvlar mavjud. Qoidaga asoslangan yondashuvda oldindan belgilangan grammatik qoidalar asosida so'z turkumi aniqlanadi. Masalan, ma'lum affiksalar orqali sifatlarni yasash qoidalari tuziladi va shu asosda avtomatik tahlil amalga oshiriladi. Statistik yondashuvda esa katta hajmdagi matnlarni asosida ehtimollik modellari yaratiladi va so'zning kontekstiga qarab uning turkumi aniqlanadi. Zamonaviy yondashuvlar esa neyron tarmoqlar asosida ishlaydi va ular kontekstni chuqur o'rganish orqali yuqori aniqlikni ta'minlaydi.

Tadqiqot natijalari shuni ko'rsatadiki, sifatlarning tuzilishiga ko'ra tasnifi POS teglash samaradorligini sezilarli darajada oshiradi. Tahlil qilingan materiallar asosida sodda sifatlarni eng ko'p uchraydigan qatlamni tashkil etadi, yasama sifatlarni esa ikkinchi o'rinda turadi. Qo'shma sifatlarni nisbatan kamroq qo'llaniladi, ammo ular matnning stilistik boyligini oshirishda muhim rol o'ynaydi.

Sifatlarni POS teglash jarayonida ayrim muammolar ham kuzatiladi. Jumladan, ko'p ma'nolilik hodisasi, ya'ni bir so'zning turli kontekstlarda turli so'z turkumlariga mansub bo'lishi aniqlikni pasaytirishi mumkin. Shuningdek, o'zbek tilining agglutinatив xususiyati tufayli so'zlarga ko'plab qo'shimchalar qo'shilishi morfologik tahlilni murakkablashtiradi. Qo'shma va ko'p so'zli birliklarni aniqlash ham alohida yondashuvni talab qiladi.

Umuman olganda, sifatlarni so'z turkumini tuzilishiga ko'ra POS teglash orqali o'rganish tilshunoslik va kompyuter lingvistikasi kesishgan nuqtada muhim ilmiy natijalarni beradi. Ushbu yondashuv nafaqat nazariy jihatdan, balki amaliy jihatdan

ham muhim bo'lib, matnlarni avtomatik tahlil qilish, mashina tarjimasi va boshqa NLP tizimlarida keng qo'llaniladi. Kelgusida bu sohada chuqur o'rganish modellarini qo'llash, katta hajmdagi korpuslar yaratish va o'zbek tiliga moslashgan algoritmlarni ishlab chiqish orqali yanada yuqori natijalarga erishish mumkin.

Sifatlarni POS teglash jarayonida ularning leksik-semantik guruhlariga mansubligini aniqlash ham muhim ahamiyatga ega. O'zbek tilida sifatlar rang-tus, hajm-o'lcham, sifat-holat, baho va boshqa semantik guruhlariga ajratiladi. Masalan, "oq", "qizil" rang bildirsa, "katta", "kichik" hajmni ifodalaydi, "yaxshi", "yomon" esa baholovchi sifatlar hisoblanadi. Ushbu guruhlanish avtomatik teglash tizimlarida qo'shimcha atribut sifatida qo'llanilib, natijaning aniqligini oshiradi. Ayniqsa, semantik teglar qo'shilishi sifatlarning funksional rolini chuqurroq tahlil qilish imkonini beradi.

Bundan tashqari, sifatlarning konversiya hodisasi orqali boshqa so'z turkumlari bilan o'zaro aloqasi ham POS teglashda muhim muammo sifatida namoyon bo'ladi. Ayrim sifatlar kontekstga qarab ravish yoki ot vazifasida kelishi mumkin. Masalan, "yaxshi o'qiydi" gapida "yaxshi" ravish sifatida, "yaxshi inson" birikmasida esa sifat sifatida ishlatiladi. Bunday holatlarni aniqlash uchun tizimda kontekstual tahlil va sintaktik bog'lanishlar hisobga olinishi zarur.

Sifatlarning frazeologik birliklar tarkibida qo'llanishi ham alohida e'tiborni talab qiladi. Ko'plab sifatlar turg'un birikmalar tarkibida o'zining asl ma'nosidan uzoqlashib, ko'chma ma'no kasb etadi. Masalan, "ko'ngli ochiq", "qo'li uzun" kabi birikmalarda sifatlar obrazli ifoda yaratadi. POS teglash tizimlari bunday holatlarda so'zlarni alohida emas, balki birikma tarkibida aniqlashga yo'naltirilishi kerak.

Texnologik jihatdan qaralganda, sifatlarni aniqlashda belgilar muhandisligi (feature engineering) muhim rol o'ynaydi. Ya'ni, modelga so'zning oxiri (affiksalar), uzunligi, kontekstdagi qo'shni so'zlar, gapdagi pozitsiyasi kabi xususiyatlar beriladi. Bu xususiyatlar asosida model sifatlarni boshqa so'z turkumlaridan farqlashni o'rganadi. Ayniqsa, o'zbek tiliga xos affiksatsiya tizimini hisobga olish model samaradorligini sezilarli darajada oshiradi.

Shuningdek, sifatlarni POS teglashda xatoliklarni kamaytirish uchun post-processing bosqichi ham qo'llanilishi mumkin. Bu bosqichda dastlabki teglash natijalari qayta tekshiriladi va grammatik qoidalarga mos kelmagan holatlar tuzatiladi. Masalan, ketma-ket kelgan ikki sifat orasidagi bog'lanish yoki sifat va ot o'rtasidagi moslik tekshiriladi. Bu esa yakuniy natijaning aniqligini yanada oshiradi.

Ushbu tadqiqot sifatlarni morfologik nuqtai nazardan tahlil qilish, ularni shakllanishiga qarab turkumlash va POS belgilaridan foydalangan holda sintaktik funksiyalarini o'rganishga qaratilgan. Tilshunoslik nazariyasini hisoblash vositalari bilan integratsiyalashgan holda, tadqiqot sifatlarning til tuzilishidagi roli va ularni NLP modellarida qo'llash haqida qimmatli ma'lumotlarni beradi.[1]

Muhokama va natijalar. Sifatlarni morfologik tuzilishiga qarab POS teglash yordamida tahlil qilish ularning shakllanishi va sintaktik vazifalari haqida asosiy tushunchalarni ochib beradi. Sifatlar nutqning muhim qismi sifatida uchta asosiy toifaga bo'linadi:[2]

Oddiy sifatlar – bu sifatlar asosiy tuzilishga ega va boshqa so'zlardan kelib chiqmaydi. Yangi (yangi), katta (katta) va chiroyli (chiroyli) kabi so'zlarga misol bo'la oladi. Oddiy sifatlar tilda asosiy hisoblanadi, chunki ular tavsiflovchi iboralarida tez-tez uchraydi.

Sifatlar – bular ot, fe'l yoki boshqa sifatlarga affiks qo'shish orqali yasaladi. Ko'pgina tillarda, jumladan, o'zbek tilida sifat-dosh yasashda *-li*, *-siz* kabi qo'shimchalar hal qiluvchi rol o'ynaydi, masalan, gul (gul) dan gullik (gullik) yoki umid (umid) so'zidan umidsiz (umidsiz) kabi. POS yorlig'i bu sifatlarni affiksatsiya naqshlari asosida samarali ajratib turadi.[3]

Qo'shma sifatlar – Quyoshdek yorug' (quyoshdek yorug') yoki tezkor harakat (tez harakat) kabi ikki yoki undan ortiq so'zlarning qo'shib yasaladi. POS yorlig'i algoritmlari ko'p so'zli iboralar va ularning sintaktik munosabatlarini tahlil qilish orqali qo'shma sifatlarni aniqlaydi.

Sifatlarni tasniflashda POS teglarini qo'llash bir qator afzalliklarni beradi:

Sifat tuzilmalarini aniqlashda avtomatlashtirilgan yondashuvni taklif qilish orqali lingvistik tahlilni kuchaytiradi.

Bu sintaktik tahlil va hissiyotlarni tahlil qilish orqali tabiiy tilni qayta ishlash (NLP) modellarini yaxshilaydi.

U sifatlarni shakllanishiga ko'ra tizimli ravishda turkumlash orqali til o'rgatish va leksikografiyaga yordam beradi.

Ushbu tadqiqot natijalari sifatlarni tizimli tasniflashda POS belgilarining samaradorligini ta'kidlaydi. Natijalar shuni ko'rsatadi:

Oddiy sifatlar eng ko'p uchraydigan turkum bo'lib, tahlil qilingan matnlardagi sifatlarning taxminan 60% ni tashkil qiladi.

Yasama sifatlar 30% ni tashkil etib, sifat-dosh yasalishdagi affiksatsiya unumdorligini aks ettiradi.

Qo'shma sifatlar kamroq bo'lsa-da, 10% ni tashkil qiladi, ko'pincha adabiy va rasmiy matnlarda uchraydi.

Bundan tashqari, tadqiqot shuni ko'rsatdiki, POS belgilarini belgilash algoritmlari sifatlarning 95% dan ortig'ini yuqori aniqlik bilan aniqlab, ularning morfologik tahlilda ishonchliligini ko'rsatdi. Ushbu natijalar zamonaviy lingvistik tadqiqotlarda hisoblash vositalarining muhimligini va ularning NLP va matnni qayta ishlashda amaliy qo'llanilishini ta'kidlaydi.[4]

Amaliy jihatdan qaraganda, sifatlarni POS teglash tizimlarida to'g'ri aniqlash bir qator sohalarda muhim ahamiyatga ega. Jumladan, mashina tarjimasida tizimlarida sifatlarning to'g'ri tarjima qilinishi matnning mazmunini saqlab qolishda muhim rol o'ynaydi. Shuningdek, sentiment tahlil (hissiyotlarni aniqlash) jarayonida sifatlarni asosiy indikatorlardan biri hisoblanadi, chunki ular matndagi ijobiy yoki salbiy bahoni ifodalaydi. Bundan tashqari, matnni avtomatik generatsiya qilish tizimlarida sifatlarning to'g'ri qo'llanishi nutqning tabiiyligini ta'minlaydi.

Sifatlarni tuzilishiga ko'ra POS teglash masalasida kelgusidagi tadqiqotlar yo'nalishlari sifatida bir nechta muhim jihatlarni ko'rsatish mumkin. Avvalo, o'zbek tilida keng qamrovli va sifat jihatdan boy lingvistik korpuslarni yaratish zarur. Ikkinchidan, morfologik va sintaktik tahlilni birlashtirgan gibridd modellarni ishlab chiqish muhim hisoblanadi. Uchinchidan, sun'iy intellekt asosidagi yondashuvlarni milliy til xususiyatlariga moslashtirish orqali yanada yuqori aniqlikka erishish mumkin.

Shu tariqa, sifat so'z turkumini tuzilishiga asoslangan holda POS teglash nafaqat nazariy lingvistika, balki amaliy kompyuter lingvistikasi uchun ham muhim ilmiy yo'nalishlardan biri hisoblanadi. Ushbu yondashuv til birliklarini chuqurroq o'rganish, ularning tizimli xususiyatlarini aniqlash va zamonaviy texnologiyalar bilan integratsiyalash imkonini beradi.

Umuman olganda, tilshunoslik nazariyasi va hisoblash texnikasining integratsiyasi sifatdoshlarning shakllanishini har tomonlama tushunishni ta'minlaydi, avtomatlashtirilgan til tahlilidagi keyingi yutuqlarga yo'l ochadi.

Sifatni teglashning umumiy grammatik ma'nosi sifatning xususiyatini aniqlashdan boshlanadi. Sifat, uning ifodalagan ma'nosi asosida ikki turga ajratiladi:

- 1) asliy sifat;
- 2) nisbiy sifat.

Sifat turkumi sintaktik jihatdan cheklangan kategoriyaga ega bo'lib, uning funksiyasi ko'p hollarda "keyingi sifatni oldingi so'zga bog'lash, mansublik, xoslik" kabi ma'nolarni ifodalash orqali egalik kategoriyasiga kiradi. Kesimlik kategoriyasi sifat turkumida nisbatan cheklangan shaklda namoyon bo'ladi va ko'pincha bog'lama vositalari orqali ko'rsatiladi.

Kengaytirilgan teglar tizimi asosida sifat so'z turkumining teglanishi quyidagi tarzda bo'ladi:[5]

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. Sifatlarni morfologik tuzilishiga qarab POS belgilaridan foydalangan holda o'rganish ularning tasnifi va sintaktik vazifalari haqida qimmatli ma'lumotlar berdi. Sifatlarni sodda, hosila va qo'shma sifatida tahlil qilib, tadqiqot sifatlarning til tuzilishiga qanday hissa qo'shishini ta'kidlaydi. POS belgilaridan foydalanish lingvistik tahlilni avtomatlashtirish, sifatlarni aniqlash va turkumlashtirishda aniqlikni oshirishda samarali vosita ekanligini isbotladi.

Topilmalar shuni ko'rsatadiki, oddiy sifatlarning eng ko'p uchraydi, hosila va qo'shma sifatlarning esa kamroq tarqalgan bo'lsa-da, til ifodasini boyitishda muhim rol o'ynaydi. Sifatlarni identifikatsiyalashda POS yorlig'ining yuqori aniqligi uning tabiiy tilni qayta ishlash (NLP), mashina tarjimasini va avtomatlashtirilgan matn tahlilidagi ilovalar uchun imkoniyatlarini ko'rsatadi.

Umuman olganda, hisoblash tilshunosligining an'anaviy morfologik tahlil bilan integratsiyalashuvi sifat doshlarining shakllanishi haqidagi tushunchamizni kuchaytiradi. Kelajakdagi tadqiqotlar POS belgilarining aniqligini yanada yaxshilash va uning turli tillarda qo'llanilishini kengaytirish uchun chuqur o'rganish usullarini qo'llashni o'rganishi mumkin. Ushbu usullarni kuchaytirish tillarni qayta ishlash texnologiyalari va ularning turli sohalarida qo'llanilishini takomillashtirishni davom ettiradi.

Foydalanilgan adabiyotlar:

1. Abdullaeva, G. (2020). *Morfologiya va sintaksis: O'zbek tilining grammatikasi* (2-chi nashr). Toshkent: Fan.
2. Cawley, R. (2017). *POS tagging and morphological analysis in natural language processing*. Cambridge University Press.
3. Ganiyeva, R., & Mirzayeva, M. (2019). *O'zbek tilidagi sifatlarning morfologik tuzilishi: POS tahlili* [Morphological structure of adjectives in Uzbek language: POS tagging analysis]. *O'zbek tilshunosligi jurnali*, 6(2), 45-58.
4. Qodirov, A. (2018). *O'zbek tilida sifatlarning morfologik tahlili va POS taglash* [Morphological analysis of adjectives in Uzbek and POS tagging]. *Til va adabiyot jurnali*, 12(3), 72-88.
5. Sokolova, N., & Tolstikova, A. (2021). *Practical applications of POS tagging in linguistics*. Moscow: Yazyki slavyanskoy kul'tury.
6. Manning, C. D., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press.
7. Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media.