

KOREFERENSIYA VA ANAFORA TALQINI

Elova Dilrabo Qudratullayevna

PHD, ToshDO'TAU katta o'qituvchisi

elova@navoiy-uni.uz

Abdisalomova Shahlo Abdimurod qizi

ToshDO'TAU tayanch doktoranti

abdisalomovashahlo@gmail.com

Annotatsiya. Ushbu maqola NLPning muhim komponenti bo'lgan koreferensiya tadqiqiga bag'ishlangan. Unda koreferensiyaning mohiyati, o'ziga xos xususiyatlari tavsiflanadi, koreferensiya va anaforaning o'xshash va farqli jihatlari yoritiladi, koreferensiya tadqiqiga doir adabiyotlar tahlil qilinadi.

Kalit so'zlar: *NLP, kompyuter lingvistikasi, koreferensiya, anafora, referent, antisident, havola bo'lak.*

Abstract. This article is devoted to the study of coreference, an important component of NLP. The essence and specific features of coreference are described, the similar and different aspects of coreference and anaphora are accentuated, and the literature on the study of coreference is analyzed.

Keywords: *NLP, computational linguistics, coreference, anaphora, referent, antecedent, reference*

Абстракт. Данная статья посвящена изучению кореференции, важного компонента НЛП. Описаны сущность и особенности кореференции, выделены сходные и различные стороны кореференции и анафоры, проанализирована литература по изучению кореференции.

Ключевые слова: НЛП, компьютерная лингвистика, кореференция, анафора, референт, антисидент, референциальный фрагмент.

Tabiiy tilni qayta ishlash (NLP) – tabiiy til va nutqni hisoblash texnikasi yordamida tahlil va sintez qilish sohasidir. Tilning noaniqlik xususiyati tabiiy tilni qayta ishlash jarayonida qiyinchiliklarni yuzaga keltiradi. NLPning bir qator vazifalari ushbu muammolarni bartaraf etishga qaratilgan.

Koreferensiya tabiiy tilni qayta ishlash jarayonining tarkibiy qismi, matnni semantik tahlil qilishda muhim bo'lgan ishlov berish bosqichidir. Koreferensiya jahon tilshunosligi va kompyuter lingvistikasida keng tadqiq etilgan. Tilshunoslik nuqtayi nazaridan, B.Russell, J.Searle va K.Donnellan koreferensiyaga “bir xil referent haqida turli yo'llar bilan gapirish”, G.Frege “murojaat qilish usuli”, N.D.Arutyunova, V.Quine, Sh.A.Kripke, Y.V.Paduchevalar “referentga ishoraning aniqlik turlari” deya ta'rif bergan. J.Layonz, R.Bogrand asarlaridagi matn semantikasida koreferensiyaning ikki turi ko'rib chiqilgan: anaforik (anafora referentdan keyin keladi) va kataforik (anafora referentdan oldin keladi). Halliday va Hasanning “Cohesion in English” asarida koreferensiyaga matn uzviyligi –

kogeziyaning mustaqil vositasi sifatida qaraladi. Kognitiv lingvistikada esa “suhbatdoshlarning diqqat-e’tiborini rivojlantirish” deb tushuniladi [1].

Kompyuter lingvistikasida **koreferensiya** obyekt haqidagi matn ma’lumotlarini birlashtirishning asosiy texnologiyasi [2], matndagi bir xil obyektga ishora qiluvchi barcha havola bo’laklarni aniqlash jarayonidir [3]. U anafora tushunchasi bilan chambarchas bog’langan. Elango P. [2], Sukthanker R. [4], Singh S. [5] tadqiqotlarida koreferensiya va anafora terminlari ko’pincha bir-birining o’rnida ishlatilishi, koreferensiyaning qat’iy ta’rifiga amal qilinmaganligi ta’kidlanadi hamda koreferensiya va anaforani hal etish vazifasining chegarasini belgilashga urg’u beriladi. Dmitrey Kiselyovning [6] anafora va koreferensiya munosabatlariga bag’ishlangan maqolasida ushbu hodisalarning tafovutlari ochib berilgan. Unga ko’ra, anafora til hodisasi, matn qismlari o’rtasidagi sinsemantik, assimetriya, qaytmaslik munosabatlari bilan tavsiflanadi, aynan korrelyatsiya jarayoni; koreferensiya tildan tashqaridagi hodisa, avtosemantik, simmetriya munosabatlarini qamrab oladi. Yuqoridagi fikrlarga tayanib aytish mumkinki, koreferensiya *matn tarkibidagi barcha havola bo’laklarni ekvivalent sinflarga guruhlash vazifasidir*. Bunda ma’lum bir sinf ichidagi havola bo’laklar bitta nutq obyektiga tegishli bo’ladi. Quyidagi namuna asosida koreferensiya hodisasini tahlil qilamiz:

Hoshimni₁ taniysizmi? Ro’zivoy akaning sartaresh o’g’lini₂ tanimaysizmi? Yo’q taniysiz. Hozir esingizga tushiraman. Uning₃ to’g’risida To’xtaboyev degan yozuvchi kitob ham yozgan.

Yuqorida indekslangan so’zlardan *Hoshim* – nutq obyekt (referent), *Ro’zivoy akaning sartaresh o’g’li, u* birliklari referentga turli ko’rinishda murojaat qilayotgan havola bo’laklar (koreferentlar)dir. Bunda barcha koreferentlar bitta ekvivalentlik guruhini tashkil etmoqda.

Anafora – etimologiyasi *ana* (yun. “orqa”) va *pheri* (yun. “olib ketmoq”) bo’lib, “takrorlash” degan ma’noni bildiradi. Kompyuter tilshunosligida anafora nutqda ilgari aytib o’tilgan narsalarga havola yoki Mitkov ta’riflaganidek [7], “orqaga ishora” sifatida izohlanadi. Anafora – mohiyati o’zidan oldingi boshqa so’zga bog’liq bo’lgan ifoda va antisidentni topish vazifasi. Bunda **anafora** – antisidentga ishora qiluvchi so’z, **antisident** – anafora nazarda tutilgan shaxsdir. Koreferensiya esa matndagi qaysi (yo’naltiruvchi) ot iboralari bitta nutq obyektiga ishora qilishini aniqlashga qaratilgan va shu asosida asosiy referent zanjirlarini yaratadi. Demak, anafora natijalari antisident-anafora juftliklari bo’lsa, koreferensiya natijalari bir xil nutq obyektiga tegishli bo’lgan har xil turdagi (yo’naltiruvchi olmoshlar va ularning oldingi so’zlari, qavm-qarindoshlik otlari, murojaat birliklari, kasb-hunarni atovchi so’zlar va b.) birliklar to’plamidir. Koreferensiya anaforaga nisbatan umumiy, kengroq tushunchani anglatadi, ya’ni koreferensiyaning hal etish jarayoni anaforani hal etishni ham o’z ichiga oladi.

Aravakash zerikdi, shekilli, astagina ashula boshladi. **Uning** (1) ashulasi ham xirgoyidan nariga o'tmaganidan keyin **Salti** ko'tarildi: – O'lmasxon, bormisiz? Durustroq aytsangiz-chi! **Qumri** **uning** (2) so'zini quvvatladi.

Ushbu parchada anafora jarayonini aniqlaymiz: bu yerda anafora natijasi **aravakash-uning** (1), **Salti-uning** (2) juftliklari bo'lib, **aravakash** va **Salti** so'zlari **antisidentlar**, **uning**(1) va **uning** (2) birliklari bu so'zlarga ishora qilgani bois **anaforalardir**. Muammo shundaki, matndagi **O'lmasxon** birligi ham **aravakashga** tegishli havola bo'lak, biroq anafora jarayoni uni aniqlashni e'tibordan chetda qoldiradi. Koreferensiya uchun esa matndagi referentga tegishli bo'lgan barcha havola bo'laklarni aniqlash muhimdir. Бодрова tadqiqotida [8] koreferensiyaning aniqlashning ikki bosqichi keltiriladi:

1. obyektga ishora qiluvchi barcha nomzodlarni belgilash;
2. klasterlash.

Shu asosida keltirilgan nutqiy parchadagi koreferensiyaning aniqlaymiz:

1. Obyektga ishora qiluvchi barcha nomzodlarni belgilash.

Ushbu bosqichda nomzodlar: olmoshlar, otlar, otli birikmalar aniqlab olinadi: **aravakash, uning** (1), **Salti, O'lmasxon, Qumri, uning** (2)

2. klasterlash. Bu bosqichda nomzodlar ekvivalentlik sinflariga ajratiladi:

aravakash-uning (1)-**O'lmasxon**

Salti – uning (2)

Qumri

Ko'rish mumkinki, nomzodlar 2 ta ekvivalentlik guruhini tashkil etdi. **Qumri** referenti esa klasterlash uchun mos nomzodlarga ega bo'lmagan sababli **singlton** hisoblanadi. **Singlton** – bu koreferenti mavjud bo'lmagan referentdir [9].

Ingliz, nemis, fransuz, ispan, arab, italyan, rus, turk va boshqa ko'plab tillarda koreferensiyaning aniqlash masalasi tadqiq etilgan. Bugungi kunda koreferensiyaning aniqlash tizimlari qo'lda yaratishga asoslangan usullardan neyron tarmoq tizimlariga [8, 9] o'tdi. Mitkov [7] anaforani hal qilish vazifasi uchun ushbu sohadagi eng dolzarb usullar va ta'riflarni bayon qilgan. Vincent Ng [2] koreferensiyaning hal qilish bo'yicha 20 yil davomida amalga oshirilgan tadqiqotlarni tahlil qilgan, yutuq va kamchiliklar muhokama qilingan. Hassan Haji Mohammadi [10] maqolasida mavjud korpuslar va baholash ko'rsatkichlari ko'rib chiqildi, asosiy algoritmlarning umumiy ko'rinishi taqdim etildi, fors tilidagi koreferensiya va olmoshlarni aniqlash tizimlari tekshirildi. Lata [11] koreferensiya tizimlari uchun barcha turdagi koreferentlikni aniqlash usullarini batafsil ko'rib chiqqan. Zhekovaning [12] ilmiy ishida ko'p tilli koreferensiyaning aniqlash tizimi taklif etildi. Grobol og'zaki fransuz tilining koreferensiyaning hal etish tizimini yaratdi. Tolpegin rus tilidagi matnlarda uchinchi shaxs olmoshlari asosida koreferensiyaning aniqlash bo'yicha nomzodlik ishini himoya qilgan [13]. Tüfekci [14] ilmiy izlanishlarida turk tilida uchinchi shaxs olmoshlarini hal qilish uchun Hobbsning sodda yondashuvi va Xomskiyning tobelik nazariyasiga tayanilgan tizimni ishlab chiqqan.

Koreferensiya mashina tarjimasida, savollarga javob berish, matnning umumlashtirish, hissiyotlarni tahlil qilish, matnning tasniflash, nutqni tanish, nomlangan obyektni aniqlash, chatbot kabi NLP yo'nalishlariga tatbiq etiladi. Koreferensiyani hal etish uchun quyidagi bosqichlar bajarilgan bo'lishi lozim:



O'zbek tilshunosligida koreferensiya va anafora masalasiga oid tadqiqotlar ozchilikni tashkil etadi. Matn tilshunosligi va pragmatik lingvistik sohalarida koreferensiya, anafora xususida nazariy ma'lumotlar uchraydi. "Tilshunoslik terminlarining izohli lug'ati"da anaforik so'z haqida shunday deyilgan: "Oldingi so'z bildirgan shaxs yoki narsaga, oldin aytilganga ishora qiluvchi so'z" [15]. Bu yerda A.Hojiyev asosan olmoshlarni anaforik so'z sifatida talqin qilgan. Sh.Safarov monografiyasining [16] "Deyksis va nutqiy muloqot matni yoxud matnda mazmunning ifoda topishi" qismida anafora mavzusiga to'xtalgan. Olim anaforaga matndagi bog'lanishning bir turi, o'rin almashtirish vazifasi, kontekst deyksisi sifatida yondashadi, koreferentlikka nisbatan hamreferentlik atamasini qo'llaydi. Olimning fikricha, "o'rin almashtirish vazifasining bajarilishi nutqiy aktga yoki matnda koreferentlik (hamreferentlik) hodisasi mavjud bo'lishini talab qiladi", "anafora nutqiy tuzilma bo'laklarining mazmunan bog'lanishini ta'minlaydi, u koreferentlik haqida ma'lumot beruvchi lisoniy hodisadir. Tilshunoslar anafora hodisasiga tavsif berayotib, an'anaga binoan anaforik olmoshlarning matnning oldingi qatorlarida eslatilgan referentini izlashgan". Keltirilgan fikrlar Sh.Safarovning anafora masalasiga lingvistik yondashuvi jahon tilshunoslarining qarashlari bilan hamohang ekanligini ko'rsatmoqda. Koreferensiya masalasi **o'zbek kompyuter tilshunosligida, NLPda tadqiqot obyekti sifatida o'rganilmagan**. Demak, koreferensiyani hal etish masalasi bugungi o'zbek kompyuter tilshunosligining, NLPning dolzarb vazifasidir.

Foydalanilgan adabiyotlar:

1. <https://tproger.ru/articles/natural-language-processing/>
2. Ng V. Machine Learning for Entity Coreference Resolution: A Retrospective Look at Two Decades of Research. / Vol.31 No.1 (2017): Thirty-First AAAI Conference on Artificial Intelligence, pp.4877-4884. – San Francisco, California USA / <https://doi.org/10.1609/aaai.v31i1.11149>.
3. Elango P. Coreference Resolution: A Survey. – University of Wisconsin, Madison, WI, pp.1-8, 2005.

4. Sukthanker R., Poria S., Cambria E., Thirunavukarasu R. Anaphora and coreference resolution: A review. / *Inf Fusion* 59, pp.139–162, 2020. / <https://doi.org/10.1016/j.inffus.2020.01.010>
5. Singh S., Bhattacharjee K., Darbari H., Verma S. Analyzing Coreference Tools for NLP Application. / *IJCSE Vol. 7, Issue 5*, pp.608-615, 2019.
6. Киселёв Д.А. К проблеме соотношения анафоры и (ко) референции. / *So‘z san’ati xalqaro jurnali*, № 4, pp.34-38, 2019.
7. Mitkov R. Anaphora resolution. – Routledge, 2014.
8. Бодрова Анастасия Александровна. Разрешение кореференции методом кластеризации / маг. дисс. – Санкт-Петербург, 2016. – 41 с.
9. Recasens M.P. Coreference: Theory, Annotation, Resolution and Evaluation. / PhD thesis. – Universitat de Barcelona, 2010. – 259 p.
10. Zhou M., Duan N., Liu S., Shum H.Y. Progress in Neural NLP: Modeling, Learning and Reasoning. In *Engineering*, 2020. / <https://doi.org/10.1016/j.eng.2019.12.014>
11. Aralikatte R. Towards Better Neural Coreference Resolution. / PhD Thesis. – University of Copenhagen, 2021. – 165 p.
12. Mohammadi Hassan Haji, Talebpour Alireza, Aznavah Ahmad Mahmoudi, Yazdani Samaneh. Review of coreference resolution in English and Persian. / *Computation and Language; Artificial Intelligence*, pp.1-44, 2022. / <https://doi.org/1048550/arXiv.2211.04428>
13. Lata K., Singh P. & Dutta K. Mention detection in coreference resolution: survey. / *Applied Intelligence* 52, pp.9816-9860 (2022). / <https://doi.org/10.1007/s10489-021-02878-2>
14. Zhekova D. Towards Multilingual Coreference Resolution. / PhD dissertation. – Languages and Literary Studies University of Bremen, 2013. – 333 p.
15. Толпегин П.В. Автоматическое разрешение кореференции местоимений третьего лица русскоязычных текстов. / дисс.кандидат технических наук. – Москва, 2008. – 241 с.
16. Tüfekci P. Türkçe için sözdizim tabanlı bir adıl çözümleme sistemi. / Doktora tezi. – Trakya Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Ana Bilim Dalı, 2009. – 137 s.
17. Ҳоҷиев А. Тилшунослик терминларининг изохли луғати. – Тошкент: Ўзбекистон миллий энциклопедияси, 2002. – 168 б.
18. Сафаров Ш. Прагмалингвистика. / Монография. – Тошкент, 2008. – 366 б.