

## KOREFERENT BIRLIKLAR TASNIFI: O‘ZBEK TILI MISOLIDA IYERARXIK ELAK YONDASHUVI

Shahlo Abdusalomova,  
ToshDO‘TAU tayanch doktoranti  
abdusalomovashahlo@gmail.com  
ORCID: 0009-0003-9150-2603

DOI: <https://doi.org/10.5281/zenodo.17488774>

*Annotatsiya.* Koreferensiyani avtomatik aniqlash mashinalarning nutqni semantik jihatdan to‘g‘ri talqin qilishida muhim ahamiyatga ega. Bu murakkab jarayon bo‘lib, yuqori aniqlikka erishish maqsadida turli yondashuv va usullar tadrijiy taraqqiyot yo‘lini bosib o‘tgan. Ushbu maqolada o‘zbek tilidagi matnlarda koreferensiyani avtomatik aniqlash uchun qoidaga asoslangan iyerarxik elak yondashuvi tahlil qilinadi. Yondashuv bosqichlari o‘zbek tili xususiyatlarini hisobga olgan holda qayta ishlangan. Har bir elak moduli uchun real misollar va zarur bo‘lgan cheklovlar tavsiflanadi. Mazkur yondashuv kam resursli tillarda, ayniqsa, o‘zbek tili matnlarida koreferensiyani avtomatik aniqlashning tasniflash bosqichi uchun izchil va sodda usulni taklif etadi.

*Kalit so‘zlar:* koreferensiya, iyerarxik elak, koreferent, referent, qoida, matn.

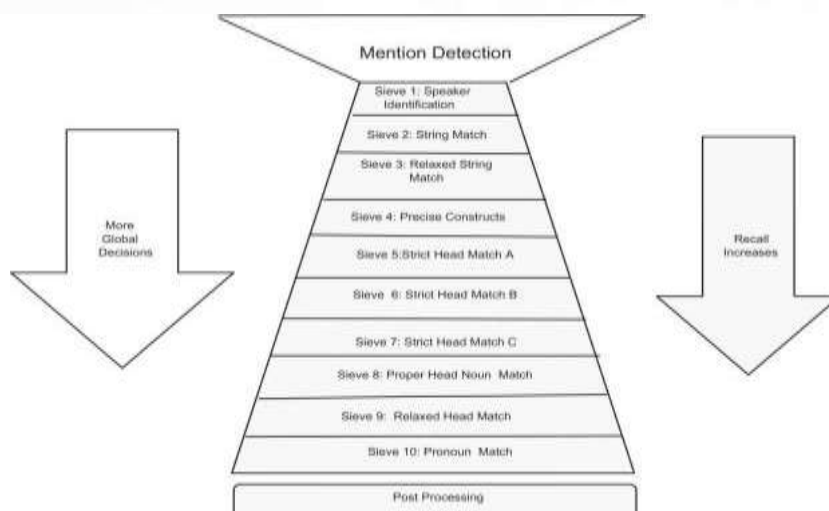
Koreferensiyani avtomatik aniqlash tabiiy tilni qayta ishlash jarayonidagi eng muhim vazifalardan biridir. Bu vazifa matnda aynan bitta shaxs va narsa-hodisalarga murojaat qilgan til birliklarini aniqlashga qaratilgan. Ushbu jarayonning vazifa qamrovi quyidagicha:

- Olmoshlarni aniqlash va bog‘lash (*u, ular, o‘zi*);
- Nom va unvonlarni bog‘lash (*Alisher Navoiy – shoir*);
- Sinonim va qisqartmalarni bog‘lash (*O‘FA – O‘zbekiston futbol assotsiatsiyasi*);
- Diskursdagi izchillikni ta‘minlash (*kim nima qilmoqda, qaysi obyekt qayerda kabi*).

Dastlab koreferensiya masalasi qoidalarga asoslangan lingvistik yondashuvlar orqali hal etilgan. Ushbu qoidalar ko‘rib chiqilayotgan matnning sintaktik va semantik xususiyatlariga asoslanadi. Shunday tizimlardan biri **Stenford CoreNLP deterministik koreferensiya tizimidir** [1]. Ushbu tizim yuqori aniqlikdan eng pastki aniqlikkacha tartiblangan deterministik qoidalarni bosqichma-bosqich qo‘llaydigan iyerarxik elakka asoslangan koreferensiya arxitekturasini taklif qiladi. Algoritm ikki bosqichdan iborat:

a) koreferent birliklarni aniqlash: koreferent birliklar bir nechta cheklovlarni qo‘llash orqali ajratib olinadi, saralanadi;

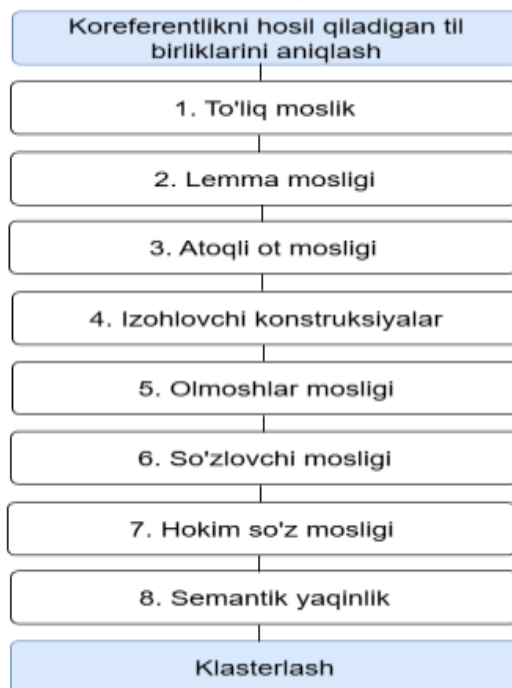
b) iyerarxik elakdan o‘tish: har bir elakda qo‘lda yozilgan ba‘zi qoidalar mavjud. Elaklarning har biri o‘zidan oldingi elakda qayta ishlangan ma‘lumotlarni klasterlash tamoyiliga tayanadi. Ya‘ni har bir elakda bir qator koreferentlik zanjirlari hosil bo‘ladi. So‘ngra zanjirlar o‘zaro birlashtirilishi uchun keyingi elakka o‘tkaziladi. Deterministik tizimda qo‘llaniladigan turli xil elaklarning tarkibi 1-rasmda ko‘rsatilgan:



### 1-rasm. Iyerarxik elak arxitekturasi

Ushbu tadqiqotda o‘zbek tilida koreferent birliklarni tasniflash uchun iyerarxik elak yondashuvining [Raghunathan, 2010: 495-498] moslashtirilgan varianti taqdim etiladi. Bu tizimning shakllanishiga A.Haghini va D.Klein tomonidan ishlab chiqilgan imkoniyati yuqori bo‘lgan qoidaga asoslangan deterministik koreferensiya tizimi (H va K modeli) turtki bo‘lgan [Haghini, 2009:1153-1157]. H.Lee boshchiligidagi bir guruh olimlar tushunarli va intuitiv misollar yordamida elaklarning ishlash prinsipini batafsil tasvirlab bergan [Lee, 2013: 887]. So‘nggi yillarda koreferensiyani aniqlash tizimlarini ishlab chiqishda elak arxitekturasi keng foydalanilmoqda. Ch.Chen va V.Ng [Chen, 2012: 58-59]; E.Fernandes, C.D.Santos va R.Milidiu [Fernandes, 2012: 42]; H.Shou va H.Zhao [Shou, 2012: 119-120]; B.Yuan va hammualliflarning [Yuan, 2012: 78] tadqiqotlari buning yaqqol dalilidir. Shuningdek, fors tilida [Mohammadi, 2023: 413], indoneziya tilida [Artari, 2021: 80-82] koreferensiya muammosini bartaraf etishda ham ushbu algoritim dasturulamal bo‘lib xizmat qilgan.

***O‘zbek tili matnlaridagi koreferensiya muammosi uchun iyerarxik elak yondashuvi.*** Dastlab o‘zbek tili matn fragmentlaridan koreferentlikni hosil qila oladigan birliklar: *ot, olmosh va otli birikmalar* avtomatik tarzda ajratib olinadi. So‘ngra har bir koreferent birlik uchun juftlik shakllantirilib, ularning koreferent bo‘lish ehtimoli iyerarxik elak qoidalari asosida tekshiriladi. O‘zbek tili matnlaridagi koreferent birliklar tasnifi uchun iyerarxik elak va lingvistik qoidalar quyida bayon qilinadi (2-rasm):



## 2-rasm. O‘zbek tili xususiyatlari asosida ishlab chiqilgan iyerarxik elak algoritmi

**1. To‘liq moslik.** Agar ikkita nomzod birlikdan biri ikkinchisining o‘z-o‘ziga havolasi (aynan takrorlangan varianti) bo‘lsa, ular koreferent. Masalan, “*Inson bilimni asosan kitob orqali oladi, uning tarbiyasi ham kitob yordamida shakllanadi*” parchasida [kitob, kitob] koreferent sifatida bir guruhga birlashtiriladi.

**2. Lemma mosligi.** Agar referentlar juftligida har ikki birlikning lemmasi bir xil bo‘lsa, ular koreferent: *Ko‘zguning paydo bo‘lish tarixi juda qadimga borib taqaladi. Qadimgi Rusda asl po‘lat ko‘zgu vazifasini bajargan jumlasida [Ko‘zguning, ko‘zgu] koreferent deb belgilanadi.*

**3. Atoqli ot mosligi.** Agar ism-familiyalar, joy va tashkilot nomlarining qisqa va to‘liq shakli mos bo‘lsa, ular koreferent: [Shuhratov Aziz, A.Shuhratov].

**4. Izohlovchi konstruksiyalar.** Agar nomzod juftliklardan biri ikkinchisining izohlovchisi bo‘lsa, ular koreferent hisoblanadi. Izohlovchi qism otli birikma bo‘lishi shart: “*Shavkat Mirziyoyev, O‘zbekiston Respublikasi Prezidenti XXRga tashrif buyurdi*” jumlasida [Shavkat Mirziyoyev, O‘zbekiston Respublikasi Prezidenti] juftligi koreferentdir.

**5. Olmoshlar mosligi.** Agar nomzod juftliklardan biri olmosh bo‘lsa va uning grammatik xususiyatlari boshqa yaqin masofada joylashgan referent bilan mos kelsa, ular koreferent: *Ali uyga keldi. U ovqatlandi* jumlasida [Ali, u] koreferent birliklar.

**6. So‘zlovchi mosligi.** Agar so‘zlovchi ma‘lum bo‘lsa va subyekt aniqlanmagan bo‘lsa yoki subyekt tushirib qoldirilgan bo‘lsa, ular so‘zlovchi bilan bog‘lanadi: *Imron dedi: “Men vazifani bajara olmadim”* – [Imron, men].

7. **Hokim so‘z mosligi.** Agar ikkala referent ham o‘tli birikma bo‘lsa va ularning hokim qismi lemmasi bir xil bo‘lsa, koreferent deb qaraladi: *Bir kuni uyimizga tilanchi keldi. O‘sha tilanchi bor-budimizni tunab ketmoqchi bo‘lgani esimda* parchasida [tilanchi, o‘sha tilanchi] koreferensiya zanjirini hosil qiladi.

8. **Semantik yaqinlik (vektor asosida).** Agar ikki referent semantik jihatdan yaqin bo‘lsa, ularning koreferent bo‘lish ehtimoli yuqori: *O‘zbekiston delegatsiyasi keldi. Mamlakat vakillari muzokara olib bordilar* – [O‘zbekiston delegatsiyasi, mamlakat vakillari].

Iyerarxik elak yondashuvi o‘zbek tilida koreferensiya muammosi uchun tushunarli, izchil va moslashuvchan asos yaratadi. Ushbu yondashuv hali ham zamonaviy tillarni qayta ishlash uchun muhim vositadir. Iyerarxik elak arxitekturasi afzalligi shundaki, deterministik qoidalar aniqlikdan umumlashmaga qarab tartiblangan va modulli bo‘lib, unga o‘rnatilgan algoritmgaga yangi qoida va o‘zgarishlar kiritish qiyinchilik tug‘dirmaydi. Aniq tartiblangan elaklar qoidalarga asoslangan gibril tizimlarni ishlab chiqishning muhim usuli bo‘la oladi.

#### Foydalanilgan adabiyotlar:

1. <https://corenlp.run/>
2. Raghunathan K., Lee H., Rangarajan S., Chambers N., Surdeanu M., Jurafsky D., Manning Ch. A Multi-Pass Sieve for Coreference Resolution // Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, Cambridge. – 2010. – pp. 492–501.
3. Haghighi A., Klein D. Simple Coreference Resolution with Rich Syntactic and Semantic Features // Proceedings of EMNLP 2009, Singapore. – 2009. – pp. 1152–1161.
4. Lee H., Chang A., Peirsman Y., Chambers N., Surdeanu M., Jurafsky D. Deterministic Coreference Resolution Based on Entity-Centric, Precision-Ranked Rules // Computational Linguistics. – 2013. – Vol. 39, № 4. – pp. 885–916.
5. Chen Ch., Ng V. Combining the Best of Two Worlds: A Hybrid Approach to Multilingual Coreference Resolution // Proceedings of the CoNLL-2012 Shared Task, Jeju Island, Korea. – 2012. – pp. 56–63.
6. Fernandes E., Santos C.D., Milidiu R. Latent Structure Perceptron with Feature Induction for Unrestricted Coreference Resolution // Proceedings of the CoNLL-2012 Shared Task, Jeju Island, Korea. – 2012. – pp. 41–48.
7. Shou H., Zhao H. System Paper for CoNLL-2012 Shared Task: Hybrid Rule-Based Algorithm for Coreference Resolution // Joint Conference on EMNLP and CoNLL-Shared Task, Jeju Island, Korea. – 2012. – pp. 118–121.
8. Yuan B., Chen Q., Xiang Y., Wang X., Ge L., Liu Z., Liao M., Si X. A mixed deterministic model for coreference resolution // Joint Conference on EMNLP and CoNLL-Shared Task, Jeju Island, Korea. – 2012. – pp. 76–82.
9. Mohammadi H.H., Talebpour A., Aznavah A.M., Yazdani S. Mehr: A Persian Coreference Resolution Corpus // Journal of Artificial Intelligence and Data Mining (JAIDM). – 2023. – Vol. 11, № 3. – pp. 407–416.
10. Artari V.K.P., Mahendra R., Jiwanggi M.A., Anggraito A., Budi I. A Multi-Pass Sieve Coreference Resolution for Indonesian // Proceedings of Recent Advances in Natural Language Processing. – 2021. – pp. 79–85.