

OILAVIY DISKURSNI NLP TEXNOLOGIYALARI ASOSIDA MODELASHTIRISH

Numonova Saidaxon Sotvoldiyevna

Andijon davlat chet tillari instituti tayanch doktoranti

saidanumonova76@gmail.com

<https://doi.org/10.5281/zenodo.21671147>

Annotatsiya: ushbu tezisdagi oilaviy diskursni nlp texnologiyalari asosida modelashtirish xususida tadqiq etilgan. Oila diskursi korpusini shakllantirish va uni keyingi lingvistik tahlilga tayyorlash jarayoni korpusni oldindan qayta ishlash bilan bog'liq usullarni qamrab oladi. Ular sarasiga: tokenizatsiya dialog matnlarini alohida so'z va tokenlarga ajratish, lemmatizatsiya so'zlarning lug'aviy shaklini aniqlash, morfologik va sintaktik tahlil esa so'z turkumlari hamda gaplarning grammatik tuzilishini belgilash kabi vazifalarni o'z ichiga oladi. Shuningdek, nomlangan obyektlarni aniqlash (Named Entity Recognition – NER) yordamida matnda uchraydigan shaxs nomlari, joy nomlari va boshqa nomlangan birliklar alohida qilib yoritib berilgan.

Kalit so'zlar: oilaviy diskurs, modelashtirish, lisoniy muloqot, dastlabki bosqich, NLP texnologiyalari, korpus lingvistikasi.

Tabiiy tilni qayta ishlash (Natural Language Processing – NLP) texnologiyasining shakllanishi 1950-yillarga borib taqaladi¹. 1950-yilda Alan Turing “Computing Machinery and Intelligence” nomli ilmiy maqolasini e'lon qilib, keyinchalik Tyuring testi (Turing Test) deb nomlangan sun'iy intellektni baholash mezonini taqdim etdi. O'sha davrda tabiiy tilni qayta ishlash sun'iy intellektdan alohida ilmiy yo'nalish sifatida shakllanganlik darajasi cheklangan bo'lsada, A.Tyuring tomonidan taklif etilgan test kompyuterning tabiiy tildagi axborotni avtomatik tarzda tushunishi va tabiiy tilda javob hosil qilishi bilan bog'liq vazifalarni amalga oshirgan. Ushbu g'oyalar keyinchalik tabiiy tilni qayta ishlash sohasining rivojlanishiga poydevor vazifasini o'tadi. Tabiiy tilni qayta ishlash bir necha bosqichlarni qamrab oladi. Dastlabki bosqichda NLP tizimlari lingvistik qoidalar va lug'atlarga asoslangan simvolik yondashuv asosida ishlab chiqilgan. 1954-yilda o'tkazilgan Georgetown tajribasi mashinaviy tarjima sohasidagi ilk muvaffaqiyatli tajribalardan biri bo'lib, NLP rivojlanishiga turtki bergan².

Tabiiy tilni qayta ishlash (Natural Language Processing – NLP) — bu kompyuter yordamida tabiiy tilda ifodalangan axborotni qayta ishlash jarayonidir. NLP kompyuter fanining alohida yo'nalishi bo'lib, sun'iy intellekt sohasi bilan chambarchas bog'liq. Shuningdek, u axborotni qidirish, bilimlarni ifodalash, kompyuter lingvistikasi hamda umumiy tilshunoslik kabi fanlar bilan uzviy aloqada rivojlanadi³. NLP tizimida bajariladigan asosiy qayta ishlash nutqni avtomatik tanib olish, matnni tasniflash, tabiiy tilni tushunish hamda tabiiy tilni generatsiya qilish kabi vazifalarni o'z ichiga oladi.

1990-yillardan boshlab qoidalarga asoslangan usullar o'rnini statistik modellar va korpuslarga tayangan mashinaviy o'rganish algoritmlari egallay boshladi⁴. Bu davr NLPda statistik

¹ **NLP – Overview. History** [Electronic resource]. – Stanford University. – URL: https://cs.stanford.edu/people/eroberts/courses/soco/projects/2004-05/nlp/overview_history.html (accessed: 02.07.2026).

² Hutchins J. *The History of Machine Translation in a Nutshell* [Electronic resource]. – 2005. – URL: <https://www.hutchinsweb.me.uk/MTN-14.pdf> (accessed: 02.07.2026).

³ **Eisenstein J.** *Introduction to Natural Language Processing*. – Cambridge, MA : The MIT Press, 2019. – 544 p.

⁴ Bengio Y., Ducharme R., Vincent P., Janvin C. A Neural Probabilistic Language Model // *The Journal of Machine Learning Research*. – 2003. – Vol. 3. – P. 1137–1155.

yondashuvning keng qo'llanilishi bilan tavsiflanadi. 2010-yillardan e'tiboran esa chuqur o'rganish va neyron tarmoqlar⁵ asosidagi usullar NLP rivojlanishining yangi bosqichini boshlab berdi. Hozirda zamonaviy NLP tizimlari katta hajmdagi matn korpuslari yordamida o'qitilib, mashinaviy tarjima, matn tahlili, hissiyotlarni aniqlash va dialog tizimlari kabi ko'plab vazifalarni yuqori aniqlikda bajarishga imkon yaratmoqda. NLPda uchta asosiy yondashuvni ajratiladi: a) simvolik (symbolic), b) statistik (statistical), v) konneksionistik (connectionist). Simvolik yondashuv grammatik qoidalar va lug'atlarga asoslanadi. Statistik yondashuv katta hajmdagi matn korpuslari yordamida ehtimollik modellarini yaratadi va zamonaviy NLP tizimlarining asosini tashkil etadi. Konneksionistik yondashuv esa sun'iy neyron tarmoqlar asosida ishlaydi⁶. Hozirgi kunda ushbu yondashuvlarning afzalliklarini birlashtiruvchi gibrid modellar keng qo'llanilmoqda.

Tabiiy tilni qayta ishlash tizimlari inson tilidagi matn va nutqni avtomatik tahlil qilish hamda qayta ishlashga qaratilgan turli vazifalarni bajaradi. Ularning asosiy turlari bilan quyida tanishib chiqamiz:

- Optik belgilarni tanish (Optical Character Recognition – OCR) – tasvir yoki hujjatlardagi matnni raqamli matnga almashtirish;
- Nutqni tanib olish (Speech Recognition) – og'zaki nutqni matn ko'rinishiga yo'naltirish;
- Tokenizatsiya (Tokenization) – matnni alohida so'zlar yoki tokenlarga ajratish⁷;
- Lemmatizatsiya (Lemmatization) – so'zlarni lug'aviy (asosiy) shakliga keltirish⁸;
- Morfologik va sintaktik tahlil – so'z turkumlari hamda gapning grammatik tuzilishini aniqlash⁹;
- Nomlangan obyektlarni aniqlash (Named Entity Recognition – NER) – matndan shaxs, joy, tashkilot kabi nomlarni ajratib olish¹⁰;
- Hissiyotlarni tahlil qilish (Sentiment Analysis) – matndagi ijobiy, salbiy yoki neytral emotsional mazmunni aniqlash¹¹;
- Semantik tahlil – so'z va gaplarning kontekstga mos ma'nosini tasniflash;
- Diskurs tahlili – matndagi gaplar va muloqot o'rtasidagi mantiqiy bog'lanishlarni o'rganish.

Yuqorida keltirilgan NLP vazifalari tabiiy tilni avtomatik qayta ishlashning turli yo'nalishlarini qamrab olsa ham, ularning barchasini mazkur tadqiqot doirasida qo'lashning imkoni yo'q. Tadqiqotning asosiy maqsadi oilaviy diskursdagi kommunikativ xususiyatlarni o'rganish bo'lganligi sababli hissiyotlarni tahlil qilish (Sentiment Analysis) va diskurs tahlili (Discourse Analysis) ustuvor ahamiyatga ega. Sentiment tahlili er-xotin muloqotlarida aks etadigan ijobiy, salbiy va neytral hissiyotlarni tavsiflashga imkon yaratsa, diskurs tahlili esa dialog tarkibidagi replikalarning o'zaro ta'sirlashuvligini, kommunikativ strategiyalar va pragmatik xususiyatlarni aniqlash uchun asos bo'ladi. Boshqa NLP vazifalari, jumladan, optik belgilarni tanish

⁵ Goodfellow I., Bengio Y., Courville A. *Deep Learning*. – Cambridge, MA : The MIT Press, 2016. – 775 p.

⁶ Liddy E.D. *Natural language processing*. – 2001.

⁷ **Natural Language Processing (NLP): A Complete Guide** [Electronic resource]. – DeepLearning.AI, 11 Jan. 2023. – URL: <https://www.deeplearning.ai/resources/natural-language-processing/> (accessed: 02.07.2026).

⁸ **What Is Natural Language Processing? Intro to NLP in Machine Learning** [Electronic resource]. – GyanSetu!, 06 Dec. 2020. – URL: <https://www.gyansetu.in/what-is-natural-language-processing-intro-to-nlp-in-machine-learning/> (accessed: 02.07.2026).

⁹ Kishorjit N., Vidya R. R. K., Nirmal Y., Sivaji B. Manipuri Morpheme Identification // *Proceedings of the 3rd Workshop on South and Southeast Asian Natural Language Processing (SANLP), COLING 2012*. – Mumbai, 2012. – P. 95–108.

¹⁰ Kariampuzha W., Alyea G., Qu S., Sanjak J., Mathé E., Sid E., Chatelaine H., Yadaw A., Xu Y., Zhu Q. Precision Information Extraction for Rare Disease Epidemiology at Scale // *Journal of Translational Medicine*. – 2023. – Vol. 21, No. 1. – Art. 157.

¹¹ **Natural Language Processing (NLP): A Complete Guide** [Electronic resource]. – DeepLearning.AI. – 2023. – January 11. – URL: <https://www.deeplearning.ai/resources/natural-language-processing/> (accessed: 02.07.2026).

(OCR), nutqni tanib olish (Speech Recognition) hamda semantik tahlil tadqiqotimizning asosiy obyekti hisoblanmaganligi uchun amaliy bosqichda qo'llashimiz ehtimoldan holi. Chunki tadqiqot audio yoki tasvirli ma'lumotlarga emas, balki oldindan yig'ilgan yozma dialoglar korpusiga tayanadi. Tadqiqotimizda korpusni tayyorlash va modellash jarayonida asosan tokenizatsiya, lemmatizatsiya, morfologik va sintaktik tahlil, nomlangan obyektlarni aniqlash, hissiyotlarni tahlil qilish hamda diskurs tahlili kabi NLP usullaridan foydalanishimiz maqsadga muvofiq deb topdik. Yuqorida nomlari ta'kidlab o'tilgan usullar oila diskursi korpusini tizimli tashkil etish va keyingi lingvistik hamda pragmatik tahlillarni amalga oshirish uchun zarur metodologik asos bo'lib xizmat qiladi deb o'ylaymiz.

Tabiiy tilni tushunishning asosiy tahlil darajalarini K.R. Chowdhury yettita o'zaro bog'liq turlarga ajratadi: fonetik (yoki fonologik), morfologik, leksik, sintaktik, semantik, diskursiv va pragmatik tahlil¹². Mazkur darajalar birgalikda kompyuter tizimlariga matn yoki nutqning mazmunini chuqurroq tushunishga ko'maklashadi. Morfologik tahlil so'zlarning ichki tuzilishini, ya'ni morfemalar, prefikslar va suffikslarni hamda ularning grammatik xususiyatlarini aniqlashga yordam beradi. So'zning asosiy shakl (lemma), so'z turkumi va grammatik kategoriyalari aniqlanib, chuqur tahlil qilinadi. Sintaktik tahlil esa gapning grammatik tuzilishini va undagi so'zlar o'rtasidagi sintaktik munosabatlarni tadqiq etishga qaratilgan bo'lib, ega, kesim, to'ldiruvchi, aniqlovchi va hol kabi gap bo'laklari hamda ularning o'zaro aloqadorligi tadqiq etiladi. Diskursiv tahlil matn yoki dialogning umumiy tuzilishi, gaplar o'rtasidagi mazmuniy bog'lanish va izchillikni tekshiradi. Muloqotning kommunikativ maqsadi, matnning yaxlitligi va gaplar o'rtasidagi semantik aloqalar yaqqol ochib beriladi. Pragmatik tahlil esa muloqot vaziyati, kontekst, so'zlovchining kommunikativ maqsadi va tashqi dunyo kabi matndan tashqari omillarni hisobga olgan holda matn mazmunini tavsiflaydi. Shuning uchun pragmatik tahlil tabiiy tilni to'liq tushunishning muhim bosqichlaridan biri hisoblanadi¹³.

J.Hirschberg va Ch.D. Manning ta'kidlashicha, tabiiy tilni qayta ishlash (NLP) inson tilini tushunish, tahlil qilish va qayta ishlashga xizmat qiluvchi bir qator texnologiyalarni qamrab oladi. Asosiy yo'nalishlardan biri mashina tarjimai bo'lib, u matnlarni bir tildan ikkinchi tilga avtomatik o'giradi. Nutqni avtomatik tanish inson nutqini matnga aylantirsa, matnni nutqqa aylantirish yozma matnni sun'iy nutqqa o'zgartiradi. Dialog tizimlari esa chatbotlar va virtual yordamchilar orqali inson va kompyuter o'rtasidagi muloqotni ta'minlaydi¹⁴. Mashina orqali o'qish matndan bilimlarni ajratib olishga xizmat qilsa, axborotni ajratib olish esa matndagi muhim obyektlar va ular o'rtasidagi munosabatlarni farqlaydi. Sentiment tahlili matnning ijobiy, salbiy yoki neytral munosabatini ochib bersa, hissiyotlarni aniqlash matnda ifodalangan hissiyotlarni belgilaydi. Ijtimoiy tarmoq ma'lumotlarini tahlil qilish esa ijtimoiy tarmoqlardagi katta hajmdagi matnlarni tahlil qilish, jamoatchilik fikri va hissiyotlarini aniqlash imkonini beradi¹⁵. Ushbu texnologiyalarni zamonaviy NLPning eng muhim yo'nalishlari desak mubolag'a bo'lmaydi. E.Liddy tabiiy tilni qayta ishlash bir necha lingvistik tahlil bosqichlarini qamrab ololmasligini ta'kidlaydi. Ular fonologik, morfologik, leksik, sintaktik, semantik, diskurs va pragmatik tahlilni o'z ichiga oladi¹⁶. Har bir bosqichni matn mazmunini tushunishda muhim vazifasi bor. Masalan, morfologik tahlil so'z

¹² Chowdhury K.R. Natural language processing //Fundamentals of artificial intelligence. – 2020. – C. 603-649.

¹³ Chowdhury K. R. Natural language processing //Fundamentals of artificial intelligence. – 2020. – C. 603-649.

¹⁴ Hirschberg J., Manning C. D. Advances in natural language processing //Science. – 2015. – T. 349. – №. 6245. – C. 261-266.

¹⁵ Hirschberg J., Manning C.D. Advances in natural language processing //Science. – 2015. – T. 349. – №. 6245. – C. 261-266.

¹⁶ Liddy E.D. Natural language processing. – 2001.

tuzilishini, sintaktik tahlil gapning grammatik tuzilishini, semantik tahlil esa soʻz va gaplarning maʼnosini yoritadi. Diskurs va pragmatik tahlil esa matnning umumiy mazmuni hamda kontekstini hisobga olib, chuqurroq tahlil qilish imkonini beradi¹⁷. Yuqori sifatli NLP tizimlari ushbu bosqichlarning bir nechtasini yoki barchasini birgalikda foydalanadi.

Tadqiqotimizning yuqoridagi fasllarida biz korpus haqida maʼlumotlar toʻpladik va ularni tasnifladik. Korpus bilan NLP bir biri bilan chambarchas bogʻliq ekanligi taʼkidlansa ham, ularning har biri lingvistikada aynan qanday maʼno kasb etishini xulosalashimizni muhim deb topdik. Bunda korpus – bu lingvistik tadqiqotlar uchun maxsus yigʻilgan va tizimlashtirilgan matnlar majmuasi boʻlib, u tadqiqotning asosiy maʼlumotlar bazasi vazifasini oʻtasa, NLP esa kompyuter yordamida mazkur matnlarni avtomatik qayta ishlash, tahlil qilish va modellashtirishni amalga oshiruvchi texnologiyalar toʻplami sifatida baholandi. Yana ham sodda tilda solishtiradigan boʻlsak, Korpus – bu kitob boʻlsa, NLP esa ustoz vazifasini bajaradi. Ustoz kitobdagi maʼlumotlarga tayangan holda, matnlarni tahlil qiladi. Korpus va NLP texnologiyasini mana shu kitob-ustoz tushunchalarining zamonaviy koʻrinishi deb talqin qilsak maqsadga muvofiq boʻladi. Mazkur tadqiqotda er-xotin muloqotlaridan iborat korpus shakllantirilishi va ushbu korpus NLP texnologiyalari yordamida qayta ishlanib, tokenizatsiya, lemmatizatsiya, morfologik belgilash hamda hissiyotlar va diskurs tahlilini amalga oshirish vazifasini belgilangan edi. Shu sababli korpus va NLP bir-birini toʻldiruvchi, ammo vazifasi jihatidan farq qiluvchi tushunchalar ekanligini inkor etmaymiz. Biz oila diskursida asosi oʻrin tutuvchi er-xotin muloqotlarini kompyuter lingvistikasi asosida tahlil qilar ekanmiz, buni korpus va NLP texnologiyalarisiz amalga oshirish imkonsiz ekanligini kuzatamiz. Zamonaviy tabiiy tilni qayta ishlash (NLP) tizimlari katta hajmdagi matnlar toʻplami – korpuslar asosida ishlaydi. Korpuslar real til namunalarini oʻz ichiga olgani sababli, ular NLP modellarini yaratish va takomillashtirish uchun asosiy maʼlumot manbai hisoblanadi¹⁸. Korpuslardan statistik modellarni yaratish, Machine Learning modellarini oʻqitish, matndagi hissiyotlarni aniqlash, grammatik va sintaktik tuzilmalarni tahlil qilish hamda turli tillar uchun yangi til modellarini ishlab chiqishda foydalaniladi. Shu sababli korpuslar zamonaviy NLP texnologiyalarining muhim nazariy va amaliy asosini tashkil etadi.

NLP – matnlarni tahlil qilishimiz uchun asosiy vosita boʻlib xizmat qilar ekan, tadqiqotlarimiz natijasida NLPning vazifaviy jihatdan quyidagi asosiy tasniflarini ishlab chiqdik.

Adabiyotlar, References, Литературы:

- 1.Звягинцева В. В. Обращение в семейном дискурсе: диссертация на соискание ученой степени кандидата филологических наук. – Екатеринбург, 2006. – 189 с.
2. Ismatullayeva I.I. Vatan lingvistik-madaniy konseptining qiyosiy-chogʻishtirma modeli va universal, milliy-madaniy konseptual belgilari // Ilm sarchashmalari, -Urganch, 2024. -№ 3. -B.122-125.
3. Исмоилов И.А. Туркий тилларда қавм-қариндошлик терминлари. – Тошкент: Фан, 1966. – Б.27.
4. Хайтбаева В. ҚОРАҚАЛПОҒИСТОНДАГИ ҚАРИНДОШЛИК АТАМАЛАРИ (Қорақалпоғистондаги ўзбек-туркман шевалари мисолида)// <https://academicsbook.com/index.php/konferensiya/article/view/3017>

¹⁷Liddy E.D. Natural language processing. – 2001.

¹⁸ Hirschberg J., Manning C. D. Advances in natural language processing //Science. – 2015. – Т. 349. – №. 6245. – С. 261-266.

5. Севортян Э.В. Этимологический словарь тюркских языков. Москва, Наука. 1974.
<http://dx.doi.org/10.26739/2181-9297-2019-3-2>