
The Position of Spanish Among the World's Languages in Contemporary Computational Linguistics

Rakhmatov Sardorkhuja
sardorrakhmatov188@gmail.com

Master's student,
Uzbekistan State World Languages University

Annotation *This article examines the role of the Spanish language in modern computational linguistics and highlights its significance in the fields of Natural Language Processing (NLP), machine translation, corpus linguistics, automatic speech recognition, and artificial intelligence systems. The primary aim of the research is to identify the position of Spanish in the global digital environment and to evaluate its function within multilingual technological platforms. In the contemporary era of rapid technological development, Spanish has become one of the most influential languages used in digital communication and AI-based language processing systems. The study is based on research conducted by major technological institutions such as Google, OpenAI, Meta, Microsoft, and the Real Academia Española. Theoretical foundations of the article are supported by Wilhelm von Humboldt's linguistic philosophy regarding language as a reflection of national thought, as well as Noam Chomsky's theories of generative grammar. Particular attention is devoted to large linguistic corpora such as Corpus del Español and CORPES XXI, which serve as essential resources for machine learning and neural language models. The findings demonstrate that Spanish occupies a strategic position in computational linguistics due to its grammatical consistency, extensive digital corpus resources, and widespread global use. Furthermore, the article emphasizes that Spanish has become one of the key languages in machine translation technologies and multilingual large language models, contributing significantly to the advancement of global artificial intelligence systems and cross-cultural digital communication.*

Keywords *Computational linguistics, Spanish language, natural language processing, artificial intelligence, machine translation, corpus linguistics, language models*

Hozirgi kompyuter lingvistikasidagi ispan tilining dunyo tillari orasidagi o'рни

Raxmatov Sardorxo'ja
sardorrakhmatov188@gmail.com

Magistratura talabasi,
O'zbekiston davlat jahon tillari universiteti

Annotatsiya *Ushbu maqolada ispan tilining zamonaviy kompyuter lingvistikasidagi o'рни, uning tabiiy tilni qayta ishlash (Natural Language Processing – NLP), mashina tarjiması, korpus lingvistikasi, avtomatik nutq tanish tizimlari hamda sun'iy intellekt asosidagi til modellari rivojıdagi ahamiyati ilmiy jihatdan tahlil qilinadi. Tadqiqotning asosiy maqsadi ispan tilining global texnologik makondagi mavqegini aniqlash va uning ko'ptilli raqamli platformalarda tutgan o'rnini baholashdan iborat. Bugungi kunda ispan tili dunyodagi eng ko'p qo'llaniladigan tillardan biri sifatida xalqaro axborot almashinuvi va raqamli kommunikatsiya jarayonlarida muhim rol o'ynamoqda. Maqolada Google, OpenAI, Meta, Microsoft hamda Real Academia Española*

tomonidan olib borilgan tadqiqotlar asosida ispan tilining sun'iy intellekt tizimlariga integratsiyalashuvi yoritiladi. Shuningdek, tilshunos olim V. fon Gumboldtning "Til – xalq tafakkurining aksi" degan nazariyasi va N. Xomskiyning generativ grammatika haqidagi qarashlari tadqiqotning nazariy asosini mustahkamlaydi. Tadqiqot natijalari shuni ko'rsatadiki, ispan tili grammatik barqarorligi, keng korpus bazasi va ko'ptilli tizimlardagi yuqori samaradorligi sababli kompyuter lingvistikasining strategik tillaridan biriga aylangan. Shuningdek, maqolada ispan tilining mashina tarjimasi va neyron tarmoqlar asosidagi modellar uchun qulay lingvistik resurs ekanligi ham asoslab beriladi.

Kalit so'zlar *Kompyuter lingvistikasi, ispan tili, NLP, sun'iy intellekt, mashina tarjimasi, korpus lingvistikasi, til modellari*

Место испанского языка среди языков мира в современной компьютерной лингвистике

Рахматов Сардорходжа
sardorrakhmatov188@gmail.com

*Студент магистратуры,
Узбекский государственный университет
мировых языков*

Аннотация *В данной статье рассматривается место испанского языка в современной компьютерной лингвистике, а также его значение в области обработки естественного языка (NLP), машинного перевода, корпусной лингвистики, автоматического распознавания речи и систем искусственного интеллекта. Основная цель исследования заключается в определении роли испанского языка в глобальном цифровом пространстве и анализе его функций в многоязычных технологических системах. В условиях стремительного развития искусственного интеллекта испанский язык становится одним из наиболее востребованных языков для автоматической обработки данных и международной коммуникации. В статье анализируются исследования компаний Google, OpenAI, Meta, Microsoft и Королевской академии испанского языка (RAE), посвящённые цифровой обработке испанского языка. Теоретической основой исследования служат идеи Вильгельма фон Гумбольдта о языке как отражении национального мышления, а также концепции генеративной грамматики Ноама Хомского. Особое внимание уделяется корпусам Corpus del Español, CORPES XXI и другим цифровым базам данных, активно применяемым в лингвистических исследованиях и обучении нейронных моделей. Результаты исследования показывают, что благодаря грамматической устойчивости, широкой распространённости и богатой цифровой базе испанский язык занимает стратегическое место в современной компьютерной лингвистике. Кроме того, подчёркивается, что испанский язык играет важную роль в развитии технологий машинного перевода и глобальных языковых моделей нового поколения.*

Ключевые слова *Компьютерная лингвистика, испанский язык, обработка естественного языка, искусственный интеллект, машинный перевод, корпусная лингвистика, языковые модели*

Kirish

XXI asrda axborot texnologiyalarining jadal rivojlanishi tilshunoslik fanining yangi yo'nalishlarini shakllantirdi. Ana shunday istiqbolli yo'nalishlardan biri kompyuter lingvistikasi hisoblanadi. Mazkur fan til va texnologiya o'rtasidagi o'zaro bog'liqlikni o'rganib, inson nutqini kompyuter orqali avtomatik qayta ishlash, tahlil qilish va tarjima qilish masalalarini tadqiq etadi. Bugungi kunda sun'iy intellekt, mashina tarjimasi, ovozli interfeyslar hamda yirik til modellari rivojida tabiiy tillarning o'rnini tobora ortib bormoqda. Shu nuqtai nazardan qaraganda, ispan tili kompyuter lingvistikasida strategik global tillardan biri sifatida muhim ahamiyat kasb etmoqda.

Tilshunos Ferdinand de Sossyur tilni "ijtimoiy belgilar tizimi" sifatida ta'riflagan bo'lsa, zamonaviy kompyuter lingvistikasida ushbu tizim sun'iy intellekt orqali avtomatik ravishda qayta ishlanmoqda. Noam Xomskiy esa generativ grammatika nazariyasini ilgari surib, tilning universal grammatik strukturalarga ega ekanligini asoslab berdi. Bugungi NLP (Natural Language Processing) tizimlari aynan ushbu nazariy tamoyillarga tayanadi.

Ispan tili bugungi kunda dunyoda eng ko'p qo'llaniladigan tillardan biri hisoblanadi. Ethnologue ma'lumotlariga ko'ra, 500 milliondan ortiq inson ispan tilida ona tili sifatida so'zlashadi. Ushbu omil ispan tilining sun'iy intellekt va kompyuter lingvistikasi tizimlarida strategik til sifatida shakllanishiga xizmat qiladi. Google, OpenAI, Meta va Microsoft kabi yirik texnologik kompaniyalar o'z til modellarida ispan tilini asosiy ishlov beriladigan tillar qatoriga kiritgan.

Shuningdek, O'zbekiston Respublikasida xorijiy tillarni, xususan lingvistika va zamonaviy texnologiyalar bilan bog'liq yo'nalishlarni rivojlantirish bo'yicha keng ko'lamli islohotlar amalga oshirilmoqda. O'zbekiston Respublikasi Prezidenti Shavkat Mirziyoyevning 2021-yil 19-

maydagi PQ-5117-son qarori xorijiy tillarni o'rganish tizimini yangi bosqichga olib chiqishda muhim huquqiy asos bo'ldi. Mazkur qarorda sun'iy intellekt, raqamli texnologiyalar va innovatsion metodlar asosida xorijiy tillarni o'qitish zarurligi alohida ta'kidlangan.

Prezident Shavkat Mirziyoyev: "Bugungi zamonda xorijiy tilni bilish – bu faqat bilim emas, balki global taraqqiyotga yo'l ochuvchi muhim vositadir", deb ta'kidlagan. Ushbu fikr kompyuter lingvistikasi va ko'ptilli sun'iy intellekt tizimlarining rivojida xorijiy tillarning strategik ahamiyatini yaqqol ifodalaydi.

Ispan tilining global til sifatidagi o'rnini

Ispan tili bugungi globallashuv davrida xalqaro kommunikatsiyaning eng muhim vositalaridan biriga aylangan. Tilshunos olim Vilgelm fon Gumboldt "Til – xalq ruhining ko'zgasidir" deb ta'kidlagan. Haqiqatan ham, ispan tili o'zida nafaqat kommunikativ imkoniyatlarni, balki boy tarixiy va madaniy merosni ham mujassamlashtiradi. Ispan tilining xalqaro nufuzi uning geografik tarqalishi bilan chambarchas bog'liq. Bugungi kunda ispan tili Ispaniya, Meksika, Argentina, Kolumbiya, Chili, Peru va boshqa ko'plab davlatlarning rasmiy tili hisoblanadi. BMT, Yevropa Ittifoqi va boshqa xalqaro tashkilotlarda ham ispan tili rasmiy maqomga ega.

E. Sapir "Har bir til dunyoni o'ziga xos tarzda idrok etish vositasidir" degan fikrni ilgari surgan. Shu jihatdan ispan tili nafaqat lingvistik birlik, balki global axborot makonining muhim tarkibiy qismi sifatida namoyon bo'ladi. Internet va raqamli kommunikatsiya rivoji natijasida ispan tilidagi elektron resurslar, matn korpuslari va multimodal ma'lumotlar hajmi keskin oshdi.

Shuningdek, ispan tilining standartlashgan yozuv tizimi va grammatik barqarorligi ham uning texnologik tizimlarda samarali qo'llanilishiga imkon yaratadi. Real Academia Española (RAE) tomonidan ishlab chiqilgan normativ qoidalar tilni avtomatik

qayta ishlash jarayonida muhim metodologik asos bo'lib xizmat qiladi.

Ispan tili va tabiiy tilni qayta ishlash (NLP)

Tabiiy tilni qayta ishlash (Natural Language Processing – NLP) kompyuter lingvistikasining eng muhim yo'nalishlaridan biridir. NLP inson tilini kompyuter orqali tahlil qilish, tushunish va generatsiya qilish bilan shug'ullanadi. Bugungi kunda ispan tili NLP tizimlarida eng faol ishlatilayotgan tillardan biri hisoblanadi. Google Translate, DeepL, ChatGPT, Gemini va boshqa til modellari ispan tilini yuqori aniqlik bilan qayta ishlash imkoniyatiga ega. Bu esa tilning grammatik tizimi va morfologik barqarorligi bilan izohlanadi.

Noam Xomskiyning generativ grammatika nazariyasi zamonaviy NLP tizimlari uchun nazariy asos yaratdi. Uning fikricha, inson tilidagi sintaktik strukturalar universal xususiyatga ega. Shu sababli kompyuter algoritmlari grammatik modellar yordamida til birliklarini avtomatik tahlil qila oladi.

Kompyuter lingvistikasi sohasida ispan tili uchun turli yirik til korpuslari yaratilgan. Ularning eng mashhurlari – Corpus del Español, CREA (Corpus de Referencia del Español Actual) va CORPES XXI korpuslaridir. Ushbu bazalar ispan tili grammatikasi, leksikasi va sintaksisini avtomatik tahlil qilishda asosiy manba sifatida qo'llaniladi. Tilshunos Jon Sinkler korpusni "real tilning elektron modeli" deb ta'riflagan bo'lib, aynan shu jihat korpus lingvistikasining zamonaviy sun'iy intellekt tizimlari uchun naqadar muhim ekanligini ko'rsatadi. NLP sohasidagi ko'ptilli tadqiqotlarda ham ispan tili muhim o'rin tutadi. Europarl Corpus, OPUS Project va Tatoeba kabi ochiq bazalarda ispancha materiallar eng katta segmentlardan biri hisoblanadi. Nutqni avtomatik tanish (ASR – Automatic Speech Recognition) tizimlarida esa Ispaniya, Meksika, Argentina va Kolumbiya lahjalari bo'yicha alohida fonetik modellar ishlab chiqilgan. Mashina tarjimasida ham ispan tili yetakchi o'rin tutadi. Google Translate, DeepL va Bing Translator tizimlarida ispan tili eng yuqori aniqlikka ega tillardan biri

sifatida qayd etiladi. Bu til morfologik jihatdan muntazam tuzilishga ega bo'lgani sababli neyron tarjima modellarida samarali ishlaydi. Roman Yakobson tarjimini "ma'no transformatsiyasi" sifatida talqin qilgan bo'lsa, bugungi neyron tarjima tizimlari aynan shu tamoyil asosida faoliyat yuritmoqda. Shu bilan birga, sun'iy intellekt asosidagi yangi til modellarida, jumladan GPT-5, LLaMA 3 va Gemini 1.5 tizimlarida ispan tili uchun maxsus tokenlash va morfologik tahlil modullari ishlab chiqilgan. Ispan tili NLP tadqiqotlarida kross-lingvistik, ya'ni ko'ptilli o'rganishlarda ham muhim o'rin egallaydi. Inglizcha-ispancha juftliklar mashina tarjimasida eng ko'p o'rganilgan ma'lumot to'plamlaridan biri hisoblanadi.

Sun'iy intellekt va avtomatik nutq tanish tizimlari rivojida ham ispan tili faol qo'llanilmoqda. Google Assistant, Amazon Alexa va Apple Siri tizimlarida ispancha ovoqli interfeyslar eng barqaror va keng qo'llanilgan variantlardan biridir. Bundan tashqari, Ispaniya, Meksika, Argentina va Kolumbiya variantlari uchun alohida fonetik modellar ishlab chiqilgani tilning dialektal xilma-xilligini texnologik jihatdan ifodalash imkonini beradi.

Kompyuter lingvistikasida ispan tilining kuchli jihatlaridan biri uning standartlashgan yozuv tizimi va akademik normativ asosidir. Real Academia Española (RAE) hamda Asociación de Academias de la Lengua Española (ASALE) tomonidan ishlab chiqilgan til me'yorlari sun'iy intellekt tizimlarida tilni to'g'ri qayta ishlash uchun barqaror metodologik asos yaratadi. Shu bilan birga, ispan tilshunoslari Computational Linguistics Association of Spain (SEPLN) doirasida ko'plab xalqaro NLP loyihalarida faol ishtirok etib kelmoqda.

Ispan tili bugungi kunda sun'iy intellekt tizimlarining ko'ptilli rivojlanishida muhim omilga aylangan. U ingliz tili bilan bir qatorda global til modellari uchun "asosiy o'quv tili" maqomiga ega. Meta AI (2023) va OpenAI (2024) hisobotlarida ispan tili global sun'iy

intellekt ekotizimida eng ko'p ishlov berilgan beshta til qatoriga kirgani alohida ta'kidlanadi.

Bugungi kunda sentiment analysis, chatbot, ovozli yordamchilar va avtomatik tarjima tizimlarida ispan tili keng ishlatilmoqda. Google Assistant, Siri va Alexa tizimlarining ispancha interfeyslari eng rivojlangan variantlardan biri hisoblanadi.

Mashina tarjimasi va ispan tili

Mashina tarjimasi kompyuter lingvistikasining amaliy jihatdan eng muhim yo'nalishlaridan biridir. Ushbu tizimlar sun'iy intellekt yordamida matnlarni bir tildan ikkinchi tilga avtomatik tarjima qiladi. Ispan tili mashina tarjimasi tizimlarida eng yuqori aniqlikka ega tillardan biri sifatida e'tirof etiladi. Google Translate va DeepL tizimlarida inglizcha-ispancha tarjimalar eng mukammal juftliklardan biri hisoblanadi. Zamonaviy neyron tarjima tizimlari aynan ushbu tamoyil asosida ishlaydi. Neyron tarmoqlar matn ma'nosini kontekst asosida aniqlab, tarjima jarayonini tabiiylashtiradi.

Ispan tilining mashina tarjimasidagi ustunliklari quyidagilar bilan izohlanadi:

- Morfologik tizimning aniq va muntazamligi.
- Sintaktik strukturalarning standartligi.
- Katta parallel korpuslarning mavjudligi.
- Inglizcha-ispancha ma'lumot bazalarining kengligi.

Europarl Corpus, OPUS Project va Common Crawl kabi ochiq ma'lumotlar bazalarida ispan tili eng faol tillardan biri hisoblanadi. Ushbu korpuslar neyron tarjima modellarini o'qitishda asosiy manba sifatida ishlatiladi. Bugungi kunda GPT, LLaMA, Gemini kabi yirik til modellarida ispan tili uchun maxsus tokenlash va morfologik tahlil algoritmlari ishlab chiqilgan. Bu esa ispan tilining sun'iy intellekt tizimlarida strategik o'ringa ega ekanligini ko'rsatadi.

Korpus lingvistikasi va ispan tili

Korpus lingvistikasi kompyuter lingvistikasining muhim tarkibiy qismi bo'lib, elektron matn bazalari orqali til birliklarini avtomatik tahlil qilishni o'rganadi.

Ispan tili uchun yaratilgan yirik korpuslar bugungi ilmiy tadqiqotlarda keng qo'llanmoqda. Ulardan eng mashhurlari:

- Corpus del Español;
- CORPES XXI;
- CREA;
- Tatoeba;
- Europarl Corpus.

Tilshunos Jon Sinkler korpusni "real tilning elektron modeli" deb atagan. Haqiqatan ham, korpuslar orqali tilning grammatikasi, leksikasi va semantik xususiyatlarini statistik jihatdan tahlil qilish mumkin. Mark Davies tomonidan yaratilgan Corpus del Español bazasida milliardlab so'z birliklari mavjud bo'lib, u ispan tilining tarixiy va zamonaviy shakllarini o'rganishda asosiy manbalardan biri hisoblanadi. Korpus lingvistikasi nafaqat ilmiy tadqiqotlar, balki sun'iy intellekt tizimlarini o'qitishda ham muhim rol o'ynaydi. Til modellari aynan katta hajmdagi korpus ma'lumotlari asosida shakllanadi.

Sun'iy intellekt va ispan tilining strategik ahamiyati

Sun'iy intellekt texnologiyalarining rivojlanishi tillarning raqamli maqomini keskin o'zgartirdi. Bugungi kunda yirik til modellari ko'ptilli tizim asosida ishlaydi va ispan tili ushbu tizimlarda eng faol tillardan biri hisoblanadi. Meta AI va OpenAI hisobotlariga ko'ra, ispan tili global sun'iy intellekt tizimlarida eng ko'p ishlov berilgan tillar qatoriga kiradi. Bu esa uning global kommunikatsiya va texnologik rivojlanishdagi ahamiyatini yanada oshiradi.

Marshall Maklyuen "Dunyo global qishloqqa aylanmoqda" degan edi. Aynan raqamli kommunikatsiya va sun'iy intellekt texnologiyalari orqali ispan tili millionlab foydalanuvchilarni yagona axborot makonida birlashtirmoqda. Ispan tilining sun'iy intellekt tizimlaridagi muvaffaqiyati quyidagi omillar bilan belgilanadi:

- foydalanuvchilar sonining ko'pligi;
- global internet resurslarining mavjudligi;
- ko'ptilli platformalarda faol qo'llanilishi;
- akademik normativ tizimning barqarorligi.

Bugungi kunda ispan tilida ishlovchi chatbotlar, virtual yordamchilar, avtomatik subtitr tizimlari va generativ AI platformalari keng rivojlanmoqda.

Xulosa

Xulosa qilib aytganda, ispan tili zamonaviy kompyuter lingvistikasida strategik global tillardan biri sifatida shakllanmoqda. Uning tabiiy tilni qayta ishlash, mashina tarjimasini, korpus lingvistikasi va sun'iy intellekt tizimlaridagi faol qo'llanilishi tilning texnologik ahamiyatini yanada oshirmoqda. Ispan tilining grammatik muntazamligi, keng korpus bazalari va global foydalanuvchilar sonining

ko'pligi uni kompyuter algoritmlari uchun qulay tilga aylantirgan. Shu bilan birga, tilning madaniy va kommunikativ ahamiyati ham global raqamli kommunikatsiyada muhim rol o'ynamoqda. Tilshunos olimlar V. fon Gumboldt, Ferdinand de Sossyur, Noam Xomskiy, Roman Yakobson va E. Sapirning nazariyalari bugungi kompyuter lingvistikasi rivojida muhim metodologik asos bo'lib xizmat qilmoqda. Shunday qilib, ispan tili nafaqat xalqaro kommunikatsiya vositasi, balki sun'iy intellekt va global texnologik taraqqiyotning muhim lingvistik komponenti sifatida namoyon bo'lmoqda.

References:

1. O'zbekiston Respublikasi Prezidenti. (2021, May 19). *O'zbekiston Respublikasida xorijiy tillarni o'rganishni ommalashtirish faoliyatini sifat jihatidan yangi bosqichga olib chiqish chora-tadbirlari to'g'risida* (PQ-5117-son qaror).
2. Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.
3. Davies, M. (2021). *Corpus del Español: 2 billion words, 21 countries*. Brigham Young University.
4. Ethnologue. (2024). *Languages of the world: Spanish profile*. SIL International.
5. Humboldt, W. von. (1999). *On language and human diversity*. Cambridge: Cambridge University Press.
6. Jakobson, R. (1959). *On linguistic aspects of translation*. Harvard University Press.
7. Meta AI Research. (2023). *Massively multilingual models for speech and text*. Menlo Park.
8. OpenAI. (2024). *Language models and multilingual understanding: Technical report*. San Francisco.
9. Real Academia Española. (2020). *CORPES XXI: Corpus del Español del Siglo XXI*. Madrid: RAE.