



BERT NEURAL NETWORK AND ITS APPLICATION IN BUILDING SYNTACTIC ANALYZERS: CURRENT STATUS AND PROSPECTS

Aziz Otemisov Zarlikbaevich

Head of the Department of Applied Philology, Associate Professor,
Karakalpak State University (KSU)
e-mail: utemisov.aziz@mail.ru

Biybihajar Mambetullaeva Azat qızı

2nd-year Master's Student, Karakalpak State University
mambetullaevabiybihajar@gmail.com
<https://doi.org/10.5281/zenodo.20508822>

ARTICLE INFO

Received: 25th May 2026

Accepted: 30th May 2026

Online: 31st May 2026

KEYWORDS

BERT, syntactic analysis,
dependency parsing,
probing, transformers,
natural language
processing.

ABSTRACT

This article is dedicated to a comprehensive analysis of applying the BERT model to natural language syntactic parsing tasks. It examines the architectural features that determine BERT's effectiveness in extracting syntactic information, as well as various approaches to integrating this model into syntactic analyzers. Special attention is paid to probing methods, which allow for the investigation of syntactic structure encoding mechanisms within the model's hidden layers, and the practical implementation of BERT-based syntactic parsers. Quantitative assessments of the effectiveness of such solutions for various languages, including Russian and English, are presented, and the limitations and future development prospects of this research direction are analyzed.

BERT NEYRON TARMOG'I VA UNING SINTAKTIK ANALIZATORNI QURISHDA QO'LLANILISHI: HOZIRGI HOLATI VA ISTIQBOLLARI

Aziz O'temisov Zarliqbaevich

QDU, Amaliy filologiya kafedrasini mudiri, dotsent,
+998907352791

Biybihajar Mambetullaeva Azat qızı

Qoraqalpoq Davlat Universiteti 2-kurs magistranti
+998945942009 mambetullaevabiybihajar@gmail.com
<https://doi.org/10.5281/zenodo.20508822>

ARTICLE INFO

Received: 25th May 2026

Accepted: 30th May 2026

Online: 31st May 2026

KEYWORDS

BERT, sintaktik tahlil,
bog'liqlikni tahlil qilish,
o'rganish,
transformatorlar, tabiiy
tilni qayta ishlash.

ABSTRACT

Maqola tabiiy til sintaktik tahlili masalalarida BERT modelini qo'llashning har tomonlama tahliliga bag'ishlangan. Sintaktik ma'lumotlarni ajratib olish uchun BERTning samaradorligini belgilovchi arxitektura xususiyatlari, shuningdek, ushbu modelni sintaktik analizatorlarga integratsiyalashning turli yondashuvlari ko'rib chiqiladi. Modelning yashirin qatlamlarida sintaktik tuzilmalarni kodlash mexanizmlarini tadqiq qilish imkonini beruvchi probing usullariga va BERT asosidagi sintaktik



parserlarning amaliy tatbiqiga alohida e'tibor qaratilmoqda. Turli tillar, jumladan rus, ingliz va ingliz tillari uchun bunday yechimlar samaradorligining miqdoriy baholari keltirilgan, shuningdek, ushbu yo'nalishning cheklovlari va rivojlanish istiqbollari tahlil qilingan.

Sintaktik tahlil gapning grammatik tuzilishini tiklashdan iborat bo'lgan tabiiy tilni qayta ishlashning asosiy vazifalaridan biridir. Sintaktik tahlilning sifati amaliy masalalarni yechish samaradorligiga bevosita ta'sir ko'rsatadi: mashina tarjimasini, ma'lumotlarni ajratib olish, savol-javob tizimlari va boshqa ko'plab masalalar.

Uzoq vaqt davomida sintaktik tahlilning asosiy yondashuvi rekurrent neyron tarmoqlari va konvolyutsion arxitekturalardan foydalanishga asoslangan edi. Biroq, transformator arxitekturasining paydo bo'lishi va ayniqsa, BERT (Bidirectional Encoder Representations from Transformers) modelining paydo bo'lishi NLP sohasida inqilob yasadi. Ular sirasiga quyidagilar kiradi [6:361]:

- BERT (Bidirectional Encoder Representations from Transformers) — tushib qolgan so'zlarni ikki tomondan ham kontekstni hisobga olib, tiklash masalasiga o'rganiladi. Klassifikatsiyalash, matnni izohlash, ma'lumot olish masalalarida keng qo'llaniladi.

- GPT (Generative Pre-trained Transformer) — kelgusi so'zni oldindan aytib berish orqali avtoregressiv tarzda o'rganiladi. Bir-biri bilan bog'liq matnlar va dialoglar hosil qilish qobiliyati bilan ajralib turadi.

- RoBERTa, XLNet, T5, ChatGPT va boshqalar yuqorida aytib o'tilgan

modellarning kengaytirilgan o'rganish va arxitekturasi yaxshilangan turlaridir.

Maqolaning maqsadi sintaktik analizatorlarni qurishda BERTni qo'llash bo'yicha zamonaviy bilimlarni tizimlashtirish, bunday foydalanishning nazariy asoslarini ko'rib chiqish, mavjud yondashuvlarni tahlil qilish va ularning samaradorligini baholashdan iborat.

BERT o'ziga-o'zi e'tibor berish (self-attention) mexanizmiga asoslangan ikki yo'nalishli til modelidir. Uning asosiy xususiyati — chap va o'ng kontekstni hisobga olgan holda so'zlarning kontekstlashtirilgan vektor ko'rinishlarini (embeddinglar) hosil qila olishidir. Bu imkoniyat BERTni sintaktik tahlil vazifalariga yanada yaqinlashtiradi.

Word2Vec yoki GloVe kabi so'zlarni ko'rsatishning an'anaviy usullari so'zning qo'llanilish muhitiga bog'liq bo'lmagan statik embeddinglarni hosil qiladi. BERT ushbu cheklovni o'z ichiga olib, dinamik kontekstli joylashmalarni yaratadi. So'zni o'rab turgan muhitga qarab uning vektor ko'rinishi o'zgaradi, bu esa so'zning semantik va sintaktik nuqtalarini his qilish imkonini beradi. Bu xususiyat sintaktik tahlil uchun prinsipial ahamiyatga ega, chunki so'zning sintaktik vazifasi gapdagi boshqa so'zlarga bevosita bog'liqdir. Sintaktik tahlil uchun BERTning eng muhim xususiyatlaridan biri uning ko'p qatlamli arxitekturasidir. Tadqiqotlar shuni ko'rsatadiki, modelning turli



qatlamlari turli darajadagi axborotni kodlaydi.

BERTda sintaktik tuzilmalar shakllanishining aniq ierarxiyasi kuzatiladi: quyi qatlamlarda mikrosintaktik tuzilmalar shakllanadi, yuqori qatlamlarda esa ular asta-sekin gapning yaxlit makrosintaktik tuzilmasiga integratsiyalashadi. Ushbu mexanizm BERTga sintaktik bog'liqliklar haqida chuqur tushuncha hosil qilish imkonini beradi, modelning niqoblangan so'zini to'g'ri bashorat qilish uchun uning grammatik rolini va gapning boshqa elementlari bilan aloqalarini bilish kerak. Ushbu qonuniyatni A.D. Mixnevaning diplom ishi ham tasdiqlaydi [3:199], u BERTning o'rta qatlamlari eng yuqori sezgirlikka ega ekanligini ko'rsatadi. Ushbu qatlamlarda chegaraning turiga bog'liq holda vektorlarning kosinuslik o'xshashligining aniq differentsiallashtirish kuzatiladi. O'rta qatlamlarni vektor ko'rinishida o'rgatilgan binar klassifikatorlar chegara turlarini ajratishda yuqori aniqlikka erishadi.

Sintaktik tahlilda BERTni qo'llashning eng ko'p uchraydigan usuli uni tokenlarning kontekstlashtirilgan ko'rinishlarini hosil qiluvchi encoder sifatida qo'llashdir. Ushbu tushunchalar sintaktik munosabatlarni oldindan bashorat qiluvchi tasniflagichga kiritiladi.

Shiyan va Markovlarning [7:225] ishida rus tilidagi matnlarni tahlil qilishning ko'p darajali usuli bayon etilgan bo'lib, unda RuBERT oilasining modellari so'zlarni embedinglarga aylantirishdan tortib, sintaktik va semantik tahlilgacha bo'lgan barcha bosqichlarda qo'llaniladi. Mualliflar bunday munosabatning yuqori

samaradorligi haqida sintaktik tahlilning aniqligi 96% deb ko'rsatadilar.

Neyron tarmoqlariga asoslangan zamonaviy sintaksis analizatorlari ko'pincha 2017-yilda Dozat va Manning tomonidan taklif etilgan biaffin tasniflash arxitekturasidan foydalanadi. Ushbu arxitektura gapdagi har bir juft so'z o'rtasidagi sintaktik aloqaning mavjudligini va turini samarali bashorat qilish imkonini beradi.

SuPar kutubxonasi [8] bir nechta sintaktik parserlarning amalga oshirilishini BERTni encoder sifatida foydalanish imkoniyati bilan ta'minlaydi. Ingliz tili uchun biaffine-dep-bert-en modeli yuqori natijalarni ko'rsatmoqda: UAS — 96.69%, LAS — 95.15%. CTB7 korpusidagi xitoy tili uchun BERT qo'llanilishi UASni 88,77% dan 91,81% ga, LASni 85,63% dan 88,94% ga ko'tarish imkonini beradi.

Modelni bir vaqtning o'zida bir nechta lingvistik masalalarda birgalikda o'rganish muhim metodologik tamoyil hisoblanadi. Shiyan va Markov qo'shma yo'qotishlar (joint losses) nuqtayi nazaridan foydalanadi, bunda model har bir vazifaning ulushini hisobga olib, bir vaqtning o'zida bir nechta maqsadli funksiyalarni optimallashtiradi. Bu modelni tahlilning barcha darajalarida bir vaqtning o'zida o'rganish imkonini beradi, bu esa natijalarning aniqligi va izchilligini oshiradi.

Yana bir usul sintaktik tuzilmalardagi qo'shni gaplarni qo'llash bilan bog'liq [9]. Ishda abzaslardagi grammatik xatolarni sintaktik belgilarning differensial birikishiga asoslanib tuzatish usuli taklif etilgan.

Metod uch bosqichdan iborat:



1. BERT yordamida gaplarni vektorlashtirish;

2. Dependency parser yordamida sintaktik strukturani ajratish;

3. Differensial tahlil: grammatik xatolarni ko'rsatishi mumkin bo'lgan muhim belgilarni aniqlash uchun kosinusli o'xshashlik yordamida qo'shni gaplar orasidagi sintaktik xatolarni ko'rsatish.

Eksperiment natijalari ushbu usul amaldagi grammatikani tuzatish tizimlaridan yuqori ekanligini ko'rsatadi: CoLA ma'lumotlar to'plamida aniqlik 0.88 (BERT-GC dan 3% yuqori), FCE da 0.89. Ingliz tili uchun sintaksis parserlarida BERTdan foydalanish eng yuqori natijalarni ko'rsatadi. Ko'rib chiqilgan ishda ta'kidlanganidek, BERT va uning modifikatsiyalari sintaktik tahlil masalalarida yangi sifat standartlarini belgilaydi.

SuPar kutubxonasi PTB (Penn Treebank) korpusida ingliz tili bo'yicha quyidagi ko'rsatkichlarni keltiradi (1-rasm):

Name	Metric	Performance	Speed (Sents/s)
biaffine-dep-en	UAS/LAS	96.03	94.37
biaffine-dep-bert-en	UAS/LAS	96.69	95.15
crfnp-dep-en	UAS/LAS	96.01	94.42
crf-dep-en	UAS/LAS	96.12	94.50
crf2o-dep-en	UAS/LAS	96.14	94.55

1-rasm. PTB korpusida ingliz tili bo'yicha ko'rsatkichlar.

Jadvaldan ko'rinib turibdiki, BERTdan foydalanish aniqlikni 0,5-1 foizlik punktga oshirish imkonini beradi, lekin qayta ishlash tezligining sezilarli darajada pasayishi bilan birga olib boriladi (biaffin parser uchun sekundiga 1826 tadan 646 tagacha so'z). Bu natija

va tezlikni bir-biriga mos kelishini ko'rsatadi.

Rus tili uchun RuBERT modellari yuqori samaradorlikni ko'rsatmoqda. Ishda matnlarni kompleks tahlil qilishdagi aniqlikning quyidagi ko'rsatkichlari haqida xabar berilgan:

Vazifasi	Aniqlik
lemmatizatsiya	98%
So'z turlarini aniqlash	97%
Morfologik belgilari	97%
Sintaksis tahlili	96%
Semantik tahlil	94%

Ushbu natijalar BERTni qo'llash morfologik masalalar uchun natijalarning amaliy jihatdan aniqligini va sintaktik tahlil uchun juda yuqori ko'rsatkichlarga erishish imkonini berishini tasdiqlaydi.

BERTni qo'llashda yuzaga keladigan muhim savol uning kam resursli tillar uchun samaradorligi bilan bog'liq. Qadimgi ingliz tili materiallari asosida o'tkazilgan tadqiqotlar transformator modellarining hamma joyda ustunligi haqidagi bashorat har doim ham o'zini oqlay olmasligini ko'rsatadi.

Unda uchta yondashuv taqqoslanadi [3:199]:

1. Ixtiyoriy ravishda initsializatsiyalanadigan spaCy bazaviy naychasi.

2. Eski ingliz tili korpusida tok2vec o'rgatilgan komponentli payplayn.

3. MobilBERT asosida "noldan" o'rgatilgan model.

Oldindan o'qitilgan tok2vec modeli UAS 83,24% va LAS 74,23% ga yetdi, transformator modeli esa ancha yomon natijalarni ko'rsatdi. Mualliflar buni model parametrlari soni (25,3 mln) va mavjud ma'lumotlar (20 mingta o'qish so'zi) o'rtasidagi nomuvofiqlik, qadimgi



ingliz tilining orfografik va morfologik xususiyatlari bilan bog'laydilar.

Ushbu tadqiqot muhim metodologik tamoyilni ko'rsatadi: arxitekturani tanlash nafaqat potensial sifat bilan, balki mavjud o'quv ma'lumotlari hajmi bilan ham belgilanishi kerak. Kam resursli tillar uchun monolingval oldindan o'rgatiladigan murakkab bo'lmagan modellar yirik transformatorlarga qaraganda samaraliroq bo'lishi mumkin.

Lingvistik zondlash (probing) neyron tarmoq modellarining ichki ko'rinishlarini tadqiq etish usuli. Probingning vazifasi turli qatlamlardagi modelning yashirin holatlarida qanday lingvistik axborot mavjudligini aniqlashdan iborat.

Sintaktik tuzilmalar shakllanishining to'liqroq tahlili bir qancha tadqiqot ishlarida tavsiya etiladi [4], unda derivativ probing tushunchasi kiritiladi. Ushbu usul mikrosintaktik tuzilmalarning quyi qatlamlarda shakllanishi va yuqori qatlamlardagi makrosintaktik tuzilmalarga asta-sekin kirib borishini kuzatish imkonini beradi. Tadqiqotlar shuni ko'rsatadiki, BERTdagi sintaktik tuzilmalar quyidan yuqoriga (bottom-up) qarab quriladi: quyi qatlamlarda bog'liqliklar shakllanadi, yuqori qatlamlarda esa global gap tuzilishiga birlashadi. Ushbu yangilik BERTning sintaktik tahlil masalalarida qanday qo'llanilishi mumkinligini tushunish uchun katta ahamiyatga ega.

Sintaksis analizatorlarida BERT qo'llanilishidagi asosiy kamchilik hisoblash murakkabligining yuqori bo'lishi hisoblanadi. BERTni biffin parserda qo'llash ishlov berish tezligini ikki baravar kamaytiradi. Bu katta hajmdagi matnni real vaqt rejimida qayta

ishlashni talab qiladigan amaliy ilovalar uchun muhim omil bo'lishi mumkin. Qadimgi ingliz tilini o'rganishda ko'rsatilganidek, BERTning samaradorligi asosan o'rganish ma'lumotlarining hajmiga bog'liq. Cheklangan korpusli tillar uchun soddaroq modellar sifat va hisoblash ko'rsatkichlarining yaxshi nisbatini ko'rsatishi mumkin.

Probing tadqiqotlarining yutuqlariga qaramay, BERT sintaktik tuzilmalarini kodlashning ichki mexanizmlari yetarli darajada o'rganilmagan. Ammo uning qatlamlari va sintaktik tuzilmalari o'rtasidagi korrelyatsiya, kuz mexanizmlari tez-tez tadqiqot obyektiga aylanmoqda.

Eng istiqbolli yo'nalishlardan biri sintaktik qurilishni hisobga oladigan oldindan o'qitishni rivojlantirish va BERTning afzalliklarini saqlab qoluvchi yanada samarali arxitekturalarni ishlab chiqishdir. Ushbu yondashuvni sintaktik tahlil masalalariga qo'llash sifatga qo'shimcha o'sish berishi mumkin. Bunday istiqbolli arxitekturalar qatoriga DistilBERT, MobileBERT, shuningdek, bilimlarni distillash va kvantlashning turli usullari kiradi.

BERT oilasining ko'p tilli modellarining rivojlanishi bir vaqtning o'zida o'nlab tillar bilan ishlaydigan sintaksis analizatorlarini yaratish imkoniyatlarini ochadi. Bu, ayniqsa, katta annotatsiyalangan korpuslarga ega bo'lmagan tillarga taalluqlidir. Shuningdek, sintaktik bilimlarni BERTni o'rganish jarayoniga aniq integratsiyalash istiqbolli yo'nalish hisoblanadi. Bunga oldindan o'qitish jarayonida sintaktik bayon qilingan ma'lumotlardan foydalanish, yordamchi



sintaktik vazifalarni qo'shish yoki sintaktik tuzilmalarni hisobga olish uchun arxitekturani o'zgartirish kiradi.

Xulosa qilib aytganda, BERT modeli sintaktik tahlil sohasida inqilob yaratib, chap va o'ng kontekstni hisobga olgan holda so'zlarning kontekstlashtirilgan ko'rinishini olish imkonini berdi. Ko'p qatlamli BERT arxitekturasi tabiiy ravishda turli darajadagi sintaktik axborotni kodlaydi. Mikrosintaktik tuzilmalar quyi qatlamlarda shakllanadi va yuqoridagi makrosintaktik tuzilmalarga asta-sekin integratsiyalashadi. BERT asosidagi sintaktik analizatorlarning amaliy amalga oshirilishi, masalan, SuPar

kutubxonasiidagi biaffin parserlar yuqori aniqlik darajasini ko'rsatmoqda. Lekin, BERTni qo'llashdagi to'siqlardan biri ma'lumotlar kam bo'lgandagi xato-kamchiliklarning yo'l qo'yilishidir.

Rivojlanishning istiqbolli yo'nalishlari qatoriga sintaktik ongni oldindan o'qitish, samarali arxitekturalarni ishlab chiqish va sintaktik bilimlarni o'qitish jarayoniga integratsiyalash kiradi. BERT sintaktik tuzilmalarni qanday kodlashini tushunish probing usullari orqali chuqurlashib bormoqda, bu esa yangi avlod sintaktik analizatorlarini yanada takomillashtirishga zamin yaratadi.

References:

1. Abdurahmonova N.Z, ÓtemisovA.Z. Kompyuter lingvistikasina kirisiv. — Nókis, 2024.
2. Abjalova M, Tahrir va tahlil dasturlarining lingvistik modullari. — Toshkent, 2020.
3. Javier Martín Arista, Ana Elvira Ojanguren López, Sara Domínguez Barragán. // Big Data and Cognitive Computing. — 2025. — № 9.
4. Taiga Someya., Ryo Yoshida., Hitomi Yanaka., Yohei Oseki. Derivational Probing: Unveiling the Layer-wise Derivation of Syntactic Structures in Neural Language Models. // arXiv:2506. — 2025.
5. Михнева А. Синтаксический пробинг границ клауз на внутренних слоях BERT. — Москва, 2025.
6. Умаров А. Обработка естественного языка с использованием нейросетей. // Innovations in Science and Technologies. — 2025. — №2.
7. Шиян В.И., Марков В.Н. Комплексный анализ русскоязычных текстов на основе нейросетевых моделей трансформерного типа. // Инженерный вестник Дона, 2025, — №5.
8. <https://pypi.org/project/supar/1.0.1a1>
9. <https://pmc.ncbi.nlm.nih.gov/articles/PMC12266432/>