

UZPRESUPP: O'ZBEK TILIDA PRESUPPOZITSION BIRLIKLARNI ANIQLASH VA PRAGMATIK TEGLASH UCHUN ANNOTATSIYALANGAN KORPUS

Sh.Sh.O'dayeva¹ 

1. Buxoro davlat universiteti Mustaqil izlanuvchisi

Mas'ul muallif: O'dayeva
Shahnoza Shuxratovna, Buxoro
davlat universiteti Mustaqil
izlanuvchisi.

Received: 28 Sentabr 2026

Accepted: 03 Oktabr 2026

Published: 08 Oktabr 2026

Copyright: © 2026 by the
author(s). Licensee Buxoro xalqaro
universiteti ilmiy axborotnomasi.

Iqtibos uchun: O'dayeva, S. S.
UZPRESUPP: O'ZBEK TILIDA
PRESUPPOZITSION BIRLIKLARNI
ANIQLASH VA PRAGMATIK
TEGLASH UCHUN
ANNOTATSIYALANGAN KORPUS /
S. S. O'dayeva // Buxoro xalqaro
universiteti ilmiy axborotnomasi.
— 2026. — T. 10, № 1. — B.
296–300.

Annotatsiya. Ushbu tadqiqotda o'zbek tilidagi presuppozitsion birliklarni korpus tamoyillari asosida qayd etishga mo'ljallangan UzPresupp nomli ko'p qatlamli annotatsiyalangan resurs konsepsiyasi taklif qilinadi. Tadqiqotda asosiy e'tibor alohida triggerlar ro'yxatini tuzishga emas, balki tabiiy matn, trigger, undan kelib chiqadigan fon propozitsiyasi hamda ushbu propozitsiyaning kontekstdagi pragmatik holatini o'zaro bog'langan korpus obyektlari sifatida ifodalashga qaratiladi. Korpus uchun badiiy nasr, yangilik, ilmiy-ommabop matn va dialogik materiallardan muvozanatli tanlama shakllantirish, hujjat-gap-token identifikatorlari, trigger chegarasi va turi, presuppozitsiyaviy mazmun, uning kontekstda tasdiqlanishi yoki bekor bo'lishi, manba janri hamda annotator qarorini saqlash modeli ishlab chiqildi. Shuningdek, korpusni Universal Dependencies bilan bog'lash, annotatorlararo kelishuvni tekshirish, murakkab holatlardagi kelishmovchiliklarni ekspert tomonidan hal etish, versiyalash va resursni qidiruvga tayyor formatda tarqatish tamoyillari yoritildi. UzPresupp hozircha tayyor statistik korpus sifatida emas, balki o'zbek pragmatikasi uchun tekshiriladigan va bosqichma-bosqich kengaytiriladigan korpus loyihasi sifatida tavsiflanadi.

Kalit so'zlar: uzPresupp, presuppozitsiya, pragmatik annotatsiya, korpus lingvistikasi, trigger, fon propozitsiyasi, ko'p qatlamli annotatsiya, Universal Dependencies, o'zbek tili, NLP.

UZPRESUPP: АННОТИРОВАННЫЙ КОРПУС ДЛЯ ВЫЯВЛЕНИЯ И ПРАГМАТИЧЕСКОЙ РАЗМЕТКИ ПРЕСПОЗИЦИОННЫХ ЕДИНИЦ В УЗБЕКСКОМ ЯЗЫКЕ

Аннотация. В данном исследовании предлагается многоуровневая концепция аннотированного ресурса под названием UzPresupp, предназначенная для регистрации пресуппозиционных единиц в узбекском языке на основе корпусных принципов. Основное внимание в исследовании уделяется не составлению списка отдельных триггеров, а представлению естественного текста, триггера, вытекающего из него фонового утверждения и прагматического состояния этого утверждения в контексте как взаимосвязанных корпусных объектов. Для корпуса была разработана сбалансированная подборка литературной прозы, новостных, научно-популярных текстов и диалогических материалов, разработаны идентификаторы документ-речь-токен, границы и типы триггеров, пресуппозиционное содержание, его подтверждение или неподтверждение в контексте, жанр источника и модель хранения решений аннотаторов. Также были рассмотрены принципы связывания корпуса с универсальными зависимостями, проверки согласованности между аннотаторами, экспертного разрешения разногласий в сложных случаях, версионирования и распространения ресурса в формате, доступном для поиска. В настоящее время UzPresupp описывается не как готовый статистический корпус, а скорее как корпусный проект по узбекской прагматике, который будет тестироваться и постепенно расширяться.

Ключевые слова: uzPresupp, пресуппозиция, прагматическая аннотация, корпусная лингвистика, триггер, фоновое предложение, многослойная аннотация, универсальные зависимости, узбекский язык, НЛП.

UZPRESUPP: AN ANNOTATED CORPUS FOR IDENTIFYING AND PRAGMATICALLY ANNOTATING PRESUPPOSITIONAL UNITS IN THE UZBEK LANGUAGE

Abstract. This study proposes the concept of "UzPresupp," a multi-layered annotated resource designed to document presuppositional units in the Uzbek language based on corpus principles. The research focuses not merely on compiling a list of triggers, but on representing the natural text, the trigger, the resulting background proposition, and the proposition's pragmatic status within the context as interconnected corpus objects. A model was developed to create a balanced selection of materials –including literary prose, news, popular science texts, and dialogue –and to store data such as document-sentence-token identifiers, trigger boundaries and types, presuppositional content, its confirmation or cancellation within the context, source genres, and annotator decisions. Additionally, the

study outlines principles for integrating the corpus with Universal Dependencies, assessing inter-annotator agreement, resolving disagreements in complex cases via expert review, implementing version control, and distributing the resource in a search-ready format. At present, UzPresupp is characterized not as a finished statistical corpus, but as a corpus project for Uzbek pragmatics that allows for verification and incremental expansion.

Keywords: uzPresupp, presupposition, pragmatic annotation, corpus linguistics, trigger, multilayer annotation, Uzbek, NLP.

Kirish. Presuppozitsiya gapda bevosita tasdiqlanmagan, ammo nutqni talqin qilishda fon ma'lumoti sifatida qabul qilinadigan mazmun bilan bog'liq. Kompyuter lingvistikasi nuqtayi nazaridan masala faqat bunday mazmun mavjudligini aniqlash bilangina cheklanmaydi. Uni real matnda qayerdan boshlanib qayerda tugashi, qaysi trigger bilan bog'lanishi va kontekst ta'sirida qanday o'zgarishini ham izchil belgilash zarur. NOPE korpusida tabiiy kontekstda uchraydigan o'n turdagi presuppozitsiya triggeri o'rganilib, trigger bilan kontekst o'rtasidagi munosabat alohida muammo sifatida ko'rsatilgan [4].

Adabiyotlar tahlili. Presuppozitsiyani hisoblash modellarida tekshirish uchun yaratilgan resurslarning vazifasi bir xil emas. Jeretič va hammualliflar ishlab chiqqan IMPPRES diagnostik to'plami implikatura va presuppozitsiyani tabiiy til inferensiyasi doirasida sinashga mo'ljallangan 32 ming yarim avtomatik tuzilgan gap juftini o'z ichiga oladi [2]. Uning muhim jihati nazorat qilinadigan diagnostik qarama-qarshiliklarni yaratish imkoniyatidir. NOPE esa sun'iy minimal juftlardan ko'ra tabiiy kontekstga e'tibor qaratib, presuppozitsion xulosaning trigger turi va undan oldingi diskursga qay darajada bog'liqligini tekshiradi [4]. Shu bois UzPresupp uchun har ikki tajriba muhim bo'lib, ular trigger turi nazorat qilinishi, annotatsiya esa tabiiy kontekstdan ajratilmasligi kerakligi bilan izohlanadi. UzPresuppning vazifasi avvalgi "teglangan dataset yaratish" modelini takrorlashdan iborat emas. Dataset muayyan mashinaviy vazifa uchun tayyorlangan misollar to'plami bo'lishi mumkin, korpus esa manba, janr, hujjat tuzilishi, tokenlash, lingvistik qatlamlar, annotatsiya tarixi va qidiruv imkoniyatlarini ham saqlaydigan barqaror resurs bo'lishini talab etadi. Shu sababli maqolada asosiy e'tibor "qaysi tegni berish kerak?" degan savoldan "presuppozitsion hodisani korpusda qanday reprezentatsiya qilish, tekshirish, qidirish va qayta ishlatish mumkin?" degan savolga ko'chiriladi.

O'zbek tili uchun bunday pragmatik qatlamni mavjud morfologik-sintaktik resurslardan alohida shakllantirib, ayni paytda ular bilan moslashtirish maqsadga muvofiq. Universal Dependencies tizimidagi Uzbek-UT 500 gapdan iborat bo'lib, yangilik va badiiy janrdagi materiallarni qamrab oladi; unda lemma, so'z turkumi, morfologik belgilar va sintaktik munosabatlar annotatsiyalangan [1]. Uzbek-UzUDT esa 684 gap va 7 582 tokenni o'z ichiga olgan, badiiy hamda o'quv/akademik materiallarga tayangan qo'lda annotatsiyalangan treebank sifatida tavsiflanadi [7]. Uzbek-TueCL 148 gapdan iborat parallel turkiy resursdir [6]. Ushbu manbalar presuppozitsiyani tayyor ko'rinishda taqdim qilmaydi, biroq token, lemma, morfologiya va sintaksis darajasida UzPresupp bilan moslashtirish mumkin bo'lgan tayanch qatlamni taqdim etadi.

Ko'p qatlamli korpuslar bilan ishlash tajribasi ham UzPresupp uchun ahamiyatli. Zeldes va Simonson GUM materiali asosida janr hamda gap turining annotatsiya sifatiga ta'sirini tekshirgan [9]. Ushbu yondashuv UzPresuppda faqat umumiy kelishuv ko'rsatkichini hisoblash bilan cheklanmasdan, kelishuvni trigger turi va janr kesimida ham alohida tahlil qilish zarurligini ko'rsatadi. INCEPTION esa span, relation va semantik annotatsiyalarni yagona muhitda boshqarish hamda annotatorlar ishini tashkil etish imkonini beruvchi umumiy platforma sifatida tavsiflangan [3].

Metodologiya: korpus tarkibi va ma'lumot modeli. UzPresupp uchun qo'yiladigan birinchi metodologik talab janr muvozanatidir. Presuppozitsiyaning namoyon bo'lishi faqat grammatik shaklga emas, balki diskurs vazifasiga ham bog'liq bo'lishi mumkin. Shu bois korpusning dastlabki relizini bitta muallif yoki bitta janr materiali bilan cheklash ilmiy jihatdan maqsadga muvofiq emas. Taklif qilinayotgan modelda badiiy nasr, yangilik, ilmiy-ommabop matn va dialog/intervyu materiallari alohida subkorpuslar sifatida saqlanadi. Har bir hujjatning mualliflik huquqi va litsenziya holati metadata qatlamida qayd etiladi; ochiq tarqatish mumkin bo'lmagan matnlarda esa faqat annotatsiya va identifikatorlarni tarqatish masalasi alohida hal qilinishi lozim.

Korpusning bazaviy birligi alohida gap emas, balki hujjat ichidagi kontekst bilan bog'langan annotatsiya hodisasidir. Har bir hodisa document_id, sentence_id va token_id orqali matnga bog'lanadi. Trigger bir yoki bir nechta tokenni qamrab olishi mumkin, presuppozitsion mazmun esa triggerning oddiy nusxasi bo'lmay, annotator tomonidan qisqa propozitsiya shaklida qayta tiklanadi. Masalan, "Anvar yana qaytdi" jumlasida "yana" trigger sifatida, "Anvar ilgari ham qaytgan" esa presuppozitsion propozitsiya sifatida qayd etiladi. Bunda mazkur propozitsiyaning matnda eksplisit tarzda ifodalanmaganini alohida ko'rsatish muhim.

1-jadval.

UzPresupp yozuvining asosiy maydonlari

Maydon	Vazifasi	Namunaviy qiymat
document_id	Manba hujjat identifikatori	UZP_FIC_00017
sentence_id	Gapning barqaror identifikatori	s0042
trigger_span	Triggering token oralig'i	yana
trigger_class	Lingvistik trigger sinfi	ITER
presuppositsiyaviy	Qayta tiklangan fon mazmuni	Anvar ilgari ham qaytgan

Maydon	Vazifasi	Namunaviy qiymat
context_status	Kontekstda amal qilish holati	SUPPORTED
projection	Operator ostida saqlanishi	PROJECTS
genre	Subkorpus turi	fiction
annotator	Qaror manbai	A2
decision	Yakuniy holat	ADJUDICATED

Trigger sinflari korpusda qidiruv indeksi vazifasini bajaradi. Dastlabki inventarda ekzistensial/definit birliklar (EXIST), faktiv predikatlar (FACT), holat o'zgarishi fe'llari (COS), iterativ vositalar (ITER), temporal konstruksiyalar (TEMP), presuppozitsiyali so'roq tuzilmalari (QUES) va qiyosiy konstruksiyalar (COMP) ajratiladi. Ushbu inventar universal yoki yopiq ro'yxat sifatida qaralmaydi. Pilot annotatsiya davomida o'zbek tiliga xos muntazam hodisalar aniqlansa, yangi teg uning ta'rif, musbat va salbiy misollari hamda boshqa toifalardan chegaralash mezonini ishlab chiqilgandan keyingina kiritiladi.

UzPresuppda trigger tegi va pragmatik teg bir xil vazifani bajarmaydi. Trigger tegi presuppozitsiyani faollashtirishi mumkin bo'lgan lisoniy shaklni tasniflash, pragmatik qatlam ayni kontekstda fon mazmunining qanday maqomda ekanini ko'rsatadi. Ushbu farq korpusning markaziy tamoyillaridan biridir. NOPE natijalarida ayrim triggerlarga berilgan inson baholari kontekstga qarab o'zgarishi ko'rsatilgan. Shu sababli trigger aniqlanishi bilan oq presuppozitsiyani avtomatik ravishda "mavjud" deb belgilash to'g'ri emas [4].

2-jadval.

UzPresupp uchun pragmatik holat teglari

Pragmatik teg	Mazmuni	Annotator savoli
SUPPORTED	Fon mazmuni kontekst bilan mos	Oldingi diskurs uni qo'llab-quvvatlaydimi?
ACCOMMODATED	Oldindan aytilmagan, ammo qabul qilinadi	Tinglovchi uni fon sifatida qo'sha oladimi?
CANCELLED	Kontekst fon mazmunini bekor qiladi	Keyingi/oldingi matn unga zidmi?
UNCERTAIN	Talqin yetarlicha barqaror emas	Ikki asosli o'qish saqlanib qoladimi?
NOT_PRESUPP	Nomzod trigger bu holatda presuppozitsiyaviy emas	Shakliy moslik noto'g'ri signalmi?

Proyeksiya holati alohida maydonda qayd etiladi. Presuppozitsiya inkor, modal operator yoki boshqa embedding muhitlarida saqlanishi mumkin, biroq proyeksiyani mutlaq va mexanik test sifatida qo'llab bo'lmaydi. Masalan, "Anvar yana kelmadi" jumlasida "yana" odatda oldingi kelish haqidagi fon mazmunini saqlab qoladi. Faktiv yoki boshqa yumshoq triggerlarda esa kontekst bu fon mazmunini zaiflashtirishi mumkin. Shu bois UzPresuppda NEG_TEST singari diagnostik belgi annotatorga yordamchi vosita bo'lib xizmat qiladi, ammo yakuniy pragmatik qarorning o'rnini bosmaydi.

O'zbek tilining morfologik tuzilishi span chegarasini belgilashda alohida e'tiborni talab etadi. UDning o'zbek tiliga oid tavsifida kelishik, egalik, zamon, mayl, aspekt va boshqa morfologik belgilar tizimli ravishda qayd etilishi ko'rsatilgan [5]. UzPresuppda trigger imkon qadar token darajasida saqlanadi. Biroq presuppozitsion effekt affiks, yordamchi fe'l yoki analitik konstruksiya bilan bog'liq bo'lgan hollarda trigger_span bir nechta tokenni qamrab olishi mumkin. Bunday yechim keyinchalik lemma, morfologik xususiyat va dependency munosabatlarini birlashtirgan holda qidiruv olib borish imkonini beradi.

Masalan, korpus foydalanuvchisi faqat ITER tegi bilan belgilangan barcha holatlarni emas, balki "ITER + VERB + negation" yoki "FACT + fiction + CANCELLED" kabi bir nechta qatlam kesishgan so'rovlarni ham olish imkoniga ega bo'lishi kerak. Shu nuqtayi nazardan, UzPresuppning qiymati teglarning ko'pligida emas, balki turli qatlamlarning o'zaro bog'langanligida namoyon bo'ladi. UzPresuppni yaratish jarayoni ikki bosqichli annotatsiya va kelishmovchiliklarni ekspert tomonidan majburiy ko'rib chiqish asosida tashkil etiladi. Dastlab annotatorlar bir xil yo'riqnomaga tayangan holda matnni mustaqil ravishda belgilaydi. Shundan so'ng trigger chegarasi, trigger sinfi va pragmatik holat yuzasidan kelib chiqqan kelishmovchiliklar ajratib olinadi. Uchinchi bosqichda ekspert bahsli misolni kontekst bilan birga ko'rib chiqib, yakuniy qarorni qayd etadi. Korpusda faqat yakuniy tegni emas, dastlabki annotator qarorlarini ham audit izi sifatida saqlash foydali. Bu keyingi yo'riqnomada tahrirlarda qaysi kategoriyalar ko'proq muammo tug'dirayotganini aniqlash imkonini beradi.

Annotatorlararo kelishuvni bitta umumiy ko'rsatkich bilan ifodalash yetarli emas. Spanlarni aniqlashda chegaralarning mosligi, kategorik teglar uchun esa moslik koeffitsiyenti alohida hisoblanishi kerak. GUM bo'yicha olib borilgan tadqiqot janr va gap turi annotatsiya sifatiga ta'sir qilishi mumkinligini ko'rsatadi [9]. Shu bois UzPresuppda kelishuv trigger turi va janr kesimida ham alohida hisobot qilinadi. Pilot bosqichda kelishuv darajasi past bo'lgan kategoriya avtomatik ravishda "yomon teg" sifatida chiqarib tashlanmaydi. Avvalo bunga yo'riqnomadagi noaniqlik, misollarning yetishmasligi yoki haqiqiy pragmatik gradatsiya sabab bo'layotgani tekshiriladi.

3-jadval.

UzPresupp sifat nazoratining asosiy bosqichlari

Tekshiruv	Nima tekshiriladi	Natija
Format validatsiyasi	ID, span, teg qiymati, bo'sh maydon	Texnik xatolar ro'yxati
Ikki tomonlama annotatsiya	Mustaqil qarorlar	Kelishuv jadvali
Ekspert ko'rigi	Bahsli misollar	Gold qaror
Janr auditi	Subkorpuslar muvozanati	Tarqatish profili
Versiya auditi	Yo'riqnoma va teg o'zgarishi	CHANGELOG + version

Versiyalash korpusning ilmiy takrorlanuvchanligini ta'minlash uchun zarur. Har bir relizda korpus versiyasi, annotatsiya qo'llanmasining versiyasi, hujjatlar va tokenlar soni, annotatsiyalangan presuppozitsion hodisalar miqdori hamda teglar inventari alohida ko'rsatiladi. Mazkur maqolada bu raqamlar keltirilmaydi, chunki real korpus annotatsiyasi yakunlanmasdan turib ularni taxminiy ko'rsatish ilmiy jihatdan asosli emas. Shu tariqa UzPresupp doirasida rejalashtirilgan arxitektura bilan amalda yig'ilgan korpus statistikasi qat'iy ravishda bir-biridan ajratiladi.

Muhokama va natijalar. Korpus tayyor bo'lgach, undan foydalanishning dastlabki vazifalaridan biri presuppozitsiyaviy birliklarni tegishli misollar bilan topish bo'ladi. Oddiy konkordans qidiruvi "yana" yoki "hamon" so'zlarini aniqlashi mumkin, biroq ularning har bir qo'llanishi presuppozitsion xususiyatga ega ekanini kafolatlamaydi. UzPresuppda esa foydalanuvchi trigger shakli va sinfi, pragmatik holat, janr hamda sintaktik xususiyatlarni bitta so'rov doirasida birlashtira oladi. Masalan, ITER sinfiga mansub, badiiy nasrda uchragan va SUPPORTED sifatida ekspert tomonidan tasdiqlangan misollarni alohida ajratib olish mumkin.

Ikkinchi imkoniyat presuppozitsiya bilan morfosintaksis o'rtasidagi munosabatni tekshirishdir. Universal Dependencies turli tillardagi so'z turkumlari, morfologik xususiyatlar va sintaktik bog'lanishlarni umumiy tamoyillar asosida belgilashga intiladi [8]. UzPresupp UD identifikatorlari bilan moslashtirilsa, tadqiqotchi faktiv predikatlarning qaysi komplement turlari bilan kelishini, iterativ markerlarning qaysi fe'l shakllari bilan birikishini yoki presuppozitsion so'roqlarda qaysi sintaktik tuzilmalar ko'proq uchrashini korpus materiali asosida tekshira oladi. Bunda pragmatik teg UD qatlamining o'rnini bosmaydi, balki mavjud grammatik qatlamga qo'shimcha sifatida xizmat qiladi.

Uchinchi yo'nalish hisoblash modellarini baholash uchun maxsus to'plam hosil qilish bilan bog'liq. IMPPRES nazorat qilinadigan diagnostik misollar orqali NLI modellarining pragmatik inferensiya imkoniyatlarini tekshirgan [2], NOPE esa tabiiy kontekstda yuzaga keladigan murakkabliklarni ko'rsatgan [4]. Kelgusida UzPresupp asosida train/dev/test bo'linmalarini shakllantirish mumkin, biroq ularni hujjat birligi bo'yicha ajratish zarur. Aks holda bitta hujjatga mansub juda o'xshash kontekstlarning bir vaqtning o'zida train va test qismlariga tushishi model natijasini sun'iy ravishda oshirishi mumkin.

4-jadval.

UzPresuppdan foydalanishning asosiy korpus va NLP ssenariylari

So'rov turi	Namunaviy shart	Tadqiqot vazifasi
Trigger qidiruvi	trigger_class=ITER	Iterativlarning inventari
Pragmatik qidiruv	status=CANCELLED	Bekor bo'lish kontekstlari
Janr qiyosi	FACT × fiction/news	Janrga bog'liq farqlar
Morfosintaktik qidiruv	FACT × ccomp	Trigger-komplement munosabati
Model baholash	gold status + held-out docs	Avtomatik tasnifni sinash

Korpusning ochiq va qayta foydalanishga yaroqli bo'lishi tanlanadigan formatga ham bog'liq. Amaliy yechim sifatida asosiy matn va UD qatlamini CoNLL-U formatida, pragmatik span hamda relationlarni esa JSONL yoki standoff formatida saqlash mumkin. INCEpTION singari platformalar semantik annotatsiya va murakkab qatlamlarni boshqarishga mo'ljallangan [3]. Shu sababli pilot annotatsiyada bunday muhitdan foydalanish texnik jihatdan asosli variant hisoblanadi. Biroq yakuniy format muayyan platformaga qaram bo'lmasligi va eksport qilish imkonini berishi kerak.

Xulosa. UzPresupp konsepsiyasi o'zbek tilidagi presuppozitsiyani faqat alohida triggerlar ro'yxati ko'rinishida emas, balki tabiiy kontekst bilan bog'langan ko'p qatlamli korpus hodisasi sifatida ifodalashni nazarda tutadi. Taklif qilinayotgan modelda matnning o'zi, morfosintaktik qatlam, trigger span, trigger sinfi, qayta tiklangan presuppozitsion propozitsiya, pragmatik holat, proyeksiya va annotatsiya qarori o'zaro bog'langan holda saqlanadi. Bunday tuzilma lingvistik tavsifni hisoblash jarayonida qayta foydalanish mumkin bo'lgan ma'lumot modeli bilan birlashtiradi.

Foydalanilgan adabiyotlar ro'yxati

- [1] Akhundjanova A., Talamo L. Universal Dependencies Treebank for Uzbek // Proceedings of RESOURCEFUL-2025. – Tallinn, 2025. – P. 1–6.
- [2] Jeretič P., Warstadt A., Bhooshan S., Williams A. Are Natural Language Inference Models IMPPRESsive? Learning IMpliciture and PRESupposition // Proceedings of ACL 2020. – 2020. – P. 8690–8705. DOI: 10.18653/v1/2020.acl-main.768.
- [3] Klie J.-C., Bugert M., Boullosa B., Eckart de Castilho R., Gurevych I. The INCEpTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation // COLING 2018 System Demonstrations. – P. 5–9.
- [4] Parrish A., Schuster S., Warstadt A., Agha O., Lee S.-H., Zhao Z., Bowman S. R., Linzen T. NOPE: A Corpus of Naturally-

Occurring Presuppositions in English // Proceedings of CoNLL 2021. – 2021. – P. 349–366. DOI: 10.18653/v1/2021.conll-1.28.

[5] Universal Dependencies. UD for Uzbek: tokenization, morphology and syntax guidelines. – UD v2, 2026.

[6] Universal Dependencies. UD_Uzbek-TueCL: parallel Universal Dependencies corpus for Turkic languages. UD v2.18 resource description. – 2026.

[7] Universal Dependencies. UD_Uzbek-UzUDT: manually annotated Uzbek treebank. UD v2.18 resource description. – 2026.

[8] Universal Dependencies. Universal Dependencies framework and treebank collection. – Version 2.18, 2026.

[9] Zeldes A., Simonson D. Different Flavors of GUM: Evaluating Genre and Sentence Type Effects on Multilayer Corpus Annotation Quality // Proceedings of LAW-X 2016. – P. 68–78. DOI: 10.18653/v1/W16-1709.

