



UDK:10.15(86)

Dilnoza XAMDAMOVA,
Qarshi davlat universiteti Amaliy ingliz tili kafedrasida tayanch doktoranti
E-mail: dilnozka0322@gmail.com

Qarshi DU dotsenti, PhD N.Mahmudova taqrizi asosida

SYNTACTIC ANNOTATION OF SIMPLE SENTENCES IN UZBEK AND ENGLISH: METHODOLOGICAL PRINCIPLES AND PRACTICAL CHALLENGES

Annotation

This paper offers a methodological overview of syntactic annotation of simple sentences in the Universal Dependencies (UD) framework, comparing English and Uzbek. It summarizes key UD conventions that support cross-linguistic comparability, including lexical head selection, attachment of functional elements, treatment of case markers, and coordination. The analysis highlights recurrent annotation difficulties in borderline constructions and argues that consistent annotation requires both universal UD rules and clearly documented language-specific decisions, especially for Uzbek postverbal complexes and morphologically driven relations, and for English auxiliary chains and secondary predication.

Key words: Universal Dependencies, syntactic annotation, dependency parsing, English, Uzbek, simple sentences, treebanks.

СИНТАКСИЧЕСКАЯ АННОТАЦИЯ ПРОСТЫХ ПРЕДЛОЖЕНИЙ В УЗБЕКСКОМ И АНГЛИЙСКОМ ЯЗЫКАХ: МЕТОДОЛОГИЧЕСКИЕ ПРИНЦИПЫ И ПРАКТИЧЕСКИЕ ТРУДНОСТИ

Аннотация

Статья представляет методологический обзор синтаксической аннотации простых предложений в рамках Universal Dependencies (UD) на материале английского и узбекского языков. Обобщаются ключевые конвенции UD, обеспечивающие межъязыковую сопоставимость, включая выбор лексической вершины, прикрепление функциональных элементов, трактовку маркеров case и разметку координации. Рассматриваются типичные трудности аннотирования в пограничных конструкциях и показывается, что стабильная разметка требует сочетания универсальных правил UD с заранее задокументированными языково-специфическими решениями. Для узбекского это особенно важно в зоне поствербальных комплексов и морфологически обусловленного определения отношений, для английского при анализе цепочек auxiliaries и конструкций вторичной предикации.

Ключевые слова: Universal Dependencies, синтаксическая аннотация, зависимости, английский язык, узбекский язык, простые предложения, дериванки.

O‘ZBEK VA INGLIZ TILLARIDA SODDA GAPLARNING SINTAKTIK ANNOTATSIYASI: METODOLOGIK TAMOYILLAR VA AMALIY QIYINCHILIKLAR

Annotatsiya

Ushbu maqola Universal Dependencies (UD) doirasida sodda gaplarning sintaktik annotatsiyasiga bag‘ishlangan metodologik sharh bo‘lib, ingliz va o‘zbek tillari misolida asosiy UD konvensiyalarini umumlashtiradi. Unda leksik cho‘qqini tanlash, funksional birliklarni biriktirish, case markerlari hamda koordinatsiyani belgilash qoidalari yoritiladi. Shuningdek, chegaraviy konstruksiyalarda tez-tez uchraydigan razmetka xatolari tahlil qilinadi va izchil annotatsiya uchun universal qoidalar bilan birga tilga xos yechimlarni ham oldindan hujjatlashtirish zarurligi ko‘rsatiladi. O‘zbek tilida bu postverbal komplekslar va morfologiyaga tayangan munosabatlarni belgilashga, ingliz tilida esa yordamchi fe‘llar zanjiri hamda ikkilamchi predikatsiyaga taalluqlidir.

Kalit so‘zlar: Universal Dependencies, sintaktik annotatsiya, bog‘lanishlar sintaksisi, ingliz tili, o‘zbek tili, sodda gap, daraxtbank.

Kirish. Sintaktik razmetka korpus lingvistikasi va tabiiy tilni qayta ishlash (NLP) uchun asosiy vosita hisoblanadi, chunki u gapning grammatik tuzilishini formallashtiradi va shu tariqa sintaktik tahlilchilarning (parserlarning) o‘qitilishi hamda ularni taqqoslab baholashni qo‘llab-quvvatlaydi. Yagona unifikatsiyalashgan annotatsiya sxemalari paydo bo‘lishidan avval ko‘p tilli tadqiqotlar tillar o‘rtasidagi razmetka tizimlarining o‘zaro mos kelmasligi sababli murakkablashgan: sxemalar o‘rtasida ko‘chirish, moslashtirish hatto parser yuqori sifatli bo‘lgan taqdirda ham formal favovutlar sonining ortishiga olib kelgan.

Universal Dependencies (UD) loyihasi aynan shu nomutanosiblikni bartaraf etish maqsadida ishlab chiqilgan: UD tillararo izchillikka va natijalarni qiyosiy baholashga yo‘naltirilgan yagona munosabatlar hamda tamoyillar tizimini belgilaydi. Shu bilan birga, metodologik jihatdan muhim qoida shundan iboratki, muayyan tilda mavjud bo‘lmagan birliklarni

boshqa tillarga “sun‘iy ravishda kiritish” kerak emas; chunki qiyoslanuvchanlik soxta tugunlar yaratish orqali emas, balki grammatik ma‘nolarning ifodalanish vositalari turlicha bo‘lgan holatlarda ham bir xil qo‘llanadigan qoidalarni tanlash orqali ta‘minlanadi: bu vositalar yordamchi so‘zlar, morfologik ko‘rsatkichlar yoki nol kopula bo‘lishi mumkin.

Ushbu maqolaning maqsadi – Universal Dependencies (UD) tamoyillarini ingliz va o‘zbek tillaridagi sodda gaplar misolida tizimlashtirish hamda sintaktik razmetkada uchraydigan tipik qiyinchiliklarni aniqlashdan iborat. Mazkur tillarning tanlanishi ularning tipologik kontrasti bilan izohlanadi: so‘z tartibi nisbatan qat‘iy bo‘lgan analitik ingliz tili so‘z tarkibi va shakllanishi jihatidan agglutinatив, gap bo‘laklari joylashuvi nisbatan erkin hamda morfologiyasi rivojlangan o‘zbek tiliga qarama-qarshi qo‘yiladi; shu bois ularni qiyoslash UD doirasida universal va tilga xos

yechimlarni sinovdan o'tkazish uchun ko'rsatkichli material bo'lib xizmat qiladi.

Tadqiqot materiali va metodlari. Tadqiqotning metodologik asosi Universal Dependencies (UD) tamoyillariga tayanadi. UD doirasida gapning sintaktik tuzilishi bog'lanishlar daraxti (dependency tree) ko'rinishida tasvirlanadi. So'zlar "bosh so'z → tobe so'z" tipidagi yo'naltirilgan munosabatlar orqali bog'lanadi, klauzaning cho'qqisi esa root sifatida belgilanadi.

Ushbu ish nazariy-metodologik xarakterga ega bo'lib, uch turdagi manbalarga tayangan holda olib boriladi. Birinchidan, UD bo'yicha me'yoriy va umumlashtiruvchi tadqiqotlar asos qilib olinadi; bu yo'nalishda asosiy konvensiyalar va ularning motivatsiyasi J. Nivre hamda hammualliflarining ishlarida, shuningdek UD'ning umumiy tavsifini bergan M.C. de Marneffe, C. D. Manning, J. Nivre va D. Zeman tomonidan tayyorlangan ishlarda bayon etilgan. Ikkinchidan, o'zbek tilining birinchi UD daraxtbanki (treebank) – Uzbek-UT bo'yicha nashr qilingan materiallar, ya'ni korpusning hajmi, janr tarkibi, razmetka strategiyasi hamda muammoli konstruksiyalar yuzasidan qabul qilingan yechimlar A. Akhundjanova va L. Talamo ishlariga tayangan holda ko'rib chiqiladi. Uchinchi manba guruhi sifatida esa o'zbek tilida sintaktik munosabatlarni belgilashda morfologik axborotning ahamiyatini va dependency parsing masalalarini yorituvchi tadqiqotlar, xususan N. Abdurakhmonovanning morfologik ko'rsatkichlarning sintaktik tahlildagi roliga bag'ishlangan ishlari jalb etiladi.

Sodda gaplarning sintaktik razmetkasining asosiy qoidalari. Nivre va boshqalar ta'kidlaganidek, UD doirasida sintaktik daraxtning tayanch karkasi, avvalo, leksik jihatdan mazmunli so'zlar o'rtasidagi bog'lanishlar asosida qurilishi lozim, yordamchi birliklar esa grammatik yoki semantik jihatdan aniqlashtirayotgan mazmunli so'zga tobe sifatida birlashtiriladi [4]. Nivre va boshqalarga ko'ra, bunday yondashuv tillararo qiyoslanuvchanlikni oshiradi, chunki ayrim tillarda grammatik ma'nolar yordamchi so'zlar orqali ifodalansa, boshqalarida ular morfologik ko'rsatkichlar bilan yoki umuman nol kopula shaklda realizatsiyalanishi mumkin. UD'ning umumlashtiruvchi tavsifida de Marneffe va hammualliflar ham leksik markazni, ya'ni leksik bosh so'zni tanlash tipologik jihatdan keskin farqlanuvchi tillarni qiyoslashni osonlashtirishini alohida qayd etadi, chunki bu bog'lanishlar darajasida tuzilmalarni yanada parallel ko'rinishga keltiradi [3].

Shundan kopula haqidagi qoida kelib chiqadi: Nivre va boshqalar bayon qilgan UD sxemasiga ko'ra, bog'lama fe'l klauzaning cho'qqisi hisoblanmaydi – klauzaning cho'qqisi sifatida leksik predikat tanlanadi, kopula esa unga cop munosabati orqali tobe qilib birlashtiriladi. Shu bois ingliz tilidagi Ivan is the best dancer gapida UD tahlili quyidagicha beriladi: root(dancer); nsubj(dancer, Ivan); cop(dancer, is); det(dancer, the); amod (dancer, best). de Marneffe va boshqalar alohida ta'kidlashicha, bunday konvensiya bog'lama fe'l umuman ifodalalmagan yoki mavjud bo'lmashligi mumkin bo'lgan tillar va konstruksiyalar bilan qiyosiy moslikni saqlash imkonini beradi[3]. Xuddi shu sababli o'zbek tilidagi U talaba ("U – talaba") misolida nol kopula alohida birlik sifatida kiritilmaydi: root(talaba); nsubj(talaba, U) natijada har ikki tildagi realizatsiyada ham predikativ markaz leksik birlik sifatida saqlanadi.

UD shu mantiqni yordamchi fe'llarga ham tatbiq etadi. Nivre va boshqalar ko'rsatganidek, zamon, modallik va aspekt kabi grammatik ma'nolarni yordamchi so'zlar orqali kodlaydigan birliklar mazmuniy fe'lga aux munosabati bilan tobe qilib birlashtiriladi va klauzaning cho'qqisi sifatida tanlanmaydi. Shu bois ingliz tilidagi Ivan will participate in the show gapida UD tahlili quyidagicha beriladi: root(participate); nsubj(participate, Ivan); aux(participate,

will); obl(participate, show); case(show, in); det(show, the). O'zbek tilida qiyoslanadigan natija shundan kelib chiqadiki, ko'plab grammatik ma'nolar morfologik ko'rsatkichlar orqali ifodalangan va cho'qqi sifatida mazmuniy predikat saqlanib qoladi, ya'ni Ivan shouda qatnashadi gapida root(qatnashadi); nsubj(qatnashadi, Ivan); obl(qatnashadi, shouda)[4].

Keyingi qoida predlog va postpozitsiyalar (posleloklar) haqida bo'lib, u ham Nivre va boshqalar tomonidan aniq belgilangan. Kelishik-rol funksiyasini bildiruvchi markerlar ot (nominal) cho'qqiga case munosabati bilan tobe qilib tahlil qilinishi kerak, shunda asosiy bog'lanish mazmunli so'zlar o'rtasida qayd etiladi. Shunga ko'ra, ingliz tilidagi the office of the Chair birikmasida asosiy bog'lanish nmod(office, Chair) tarzida beriladi, of esa case(Chair, of) sifatida ajratiladi. To'liq razmetka quyidagicha ko'rinadi: root(office); det(office, the); nmod(office, Chair); case(Chair, of); det(Chair, the). O'zbek tilidagi U men uchun ishlaydi ("U men uchun ishlaydi") misolida ham ayni konvensiya qo'llanadi va natija root(ishlaydi); nsubj(ishlaydi, U); obl(ishlaydi, men); case(men, uchun) tarzida ifodalangan [4].

Koordinatsiyani belgilashda UD texnik kelishuvdan foydalanadi. Nivre va boshqalar izohlashicha, bog'lanishlar daraxti formatida koordinatsiyaning simmetriyasini bevosita kodlash mumkin emas. Shu sababli bosh birlik sifatida birinchi kon'yunkt tanlanadi, qolganlari conj orqali birlashtiriladi, bog'lovchi esa cc bilan belgilanadi. de Marneffe va boshqalar bu yechim formatga xos konvensiya ekanini va uni izchil qo'llash zarurligini alohida ta'kidlaydi. Masalan, gold, silver and bronze birikmasi root(gold); conj(gold, silver); conj(gold, bronze); cc(bronze, and); punct(gold,); punct(silver,) tarzida tahlil qilinadi. O'zbek tilidagi oltin, kumush va bronza misoli ham parallel tarzda beriladi: root(oltin); conj(oltin, kumush); conj(oltin, bronza); cc(bronza, va); punct(oltin,); punct(kumush,) [4].

Nihoyat, o'zbek tili materiali uchun postverbal komplekslarga doir yechimlarni alohida qayd etish muhimdir. Bu yerda "UDning umumiy normasi"ga tayanishning o'zi yetarli emas, chunki mazkur soha turli talqinlarga ochiq. Akhundjanova va Talamo Uzbek-UT tavsifida ko'rsatganidek, postverblar uchun takror qo'llash mumkin bo'lgan tahlil compound munosabati orqali beriladi. Bunda postverbal element bosh birlik sifatida olinadi, undan oldingi komponent esa tobe unsur sifatida birlashtiriladi. Shu bilan birga, aux munosabati modal va kopula fe'llar uchun alohida rezerv qilinadi. Mazkur yondashuvga ko'ra to'xtab qol- ("to'xtab qolmoq", ya'ni 'batamom to'xtash/ yakun topish') birikmasi root(qol); compound (qol, to'xtab) tarzida razmetka qilinadi [2].

Sodda gaplarning razmetkasidagi xos qiyinchiliklar UD amaliyotini umumlashtirar ekan, de Marneffe va boshqalar shuni qayd etadiki, yagona sxema qo'llanilganda ham asosiy xatolar odatda annotator ichida raqobatlashadigan ikki xil intuitiv yondashuv to'qnashadigan nuqtalarda yuzaga keladi. Bularning birinchisi formal-grammatik yo'nalish bo'lib, unda zamon, shaxs va bog'lamalilik kabi ko'rsatkichlarni qaysi element "olib yurishi" asosiy mezon sifatida olinadi. Ikkinchisi esa leksik-semantik yo'nalish bo'lib, unda predikatsiya va mazmunning tashuvchisi qaysi birlik ekani ustuvor hisoblanadi. UD qiyosiy moslikni yuqori darajada ta'minlash uchun ikkinchi yondashuvni ongli ravishda afzal ko'radi, biroq bu chegaraviy konstruksiyalarda qat'iy izchillikni talab etadi [3].

Eng ko'p farqlanish kuzatiladigan birinchi soha bog'lama va yordamchi elementlar bilan bog'liq. Nivre va boshqalar ta'kidlashicha, "leksik predikatni bosh birlik sifatida tanlash" strategiyasi tillararo parallelizmni maksimal darajada oshiradi. Bu ayniqsa nol bog'lama mavjud bo'lgan holatlarda yoki grammatika morfologik yo'l bilan ifodalangan vaziyatlarda muhimdir. Amaliyotda esa yangi annotatorlar

ko'pincha finitiv kopula yoki yordamchi fe'lni cho'qqi sifatida belgilab qo'yadi, chunki grammatik ko'rsatkichlar aynan shu birliklarda jamlangandek ko'rinadi. Natijada tahlil bog'lamasiz konstruksiyalar bilan ham, grammatik ma'nolar alohida so'z bilan emas, boshqa vositalar orqali ifodalanadigan tillar bilan ham qiyoslanmaydigan bo'lib qoladi [4].

Ikkinchi muammoli soha predlog-kelishik va postpozitsiyali guruhlar bilan bog'liq. Nivre va boshqalar ta'kidlaganidek, predlog va postpozitsiyalar ot birligiga case sifatida tobe qilib birlashtirilishi kerak, biroq amaliyotda ko'p uchraydigan xato shundan iboratki, bu marker butun guruhning bosh birligi sifatida belgilanib qo'yiladi. Ingliz va o'zbek tillarini qiyoslashda bu masala prinsipial ahamiyatga ega, chunki case tahlili qo'llangandagina ingliz tilidagi for me va o'zbek tilidagi men uchun mazmunli so'zlar o'rtasida bir xil "tayanich" bog'lanishni beradi [4].

Uchinchi qiyinchilik koordinatsiya bilan bog'liq. de Marneffe va boshqalar qayd etishicha, "birinchi kon'yunkt bosh birlik bo'ladi" degan konvensiya semantik emas, balki texnik xususiyatga ega. Shuning uchun uni "yaxshilash"ga urinishlar, masalan semantik jihatdan eng muhim elementni bosh qilib tanlash, razmetkaning beqarorlashuviga va daraxtda kaskad xatolarga olib keladi. Demak, bu yerda izchillik ayniqsa muhim, koordinatsiya aniqlangan zahoti annotator UD konvensiyasiga qat'iy amal qilishi lozim [4].

O'zbek tili uchun esa eng muhim amaliy murakkablik sintaktik tahlilning morfologiyaga kuchli bog'liqligidir. N. Abdurakhmonova dependency parsing materiallari asosida shuni ko'rsatadiki, so'z tartibi nisbatan erkin bo'lgan sharoitda sintaktik munosabatlarni belgilash uchun asosiy signallar aynan so'z shaklining morfologik ko'rsatkichlaridan kelib chiqadi. Bular jumlasiga kelishik ko'rsatkichlari va predikativ affiksalar ham kiradi. Shundan metodik xulosa kelib chiqadiki, bahsli holatlarda annotator, avvalo, morfologik markerlarga tayanishi, komponentning gapdagi chiziqli o'rnini esa ikkilamchi mezon sifatida ko'rishi kerak [1].

O'zbek tili uchun alohida xavf zonasi postverbal komplekslar hisoblanadi. Akhundjanova va Talamo maxsus qayd etishicha, grammatik belgilar ko'pincha postverbal elementning o'zida markalanadi. Shu sababli uni ikkilamchi aux sifatida talqin qilish predikatsiya tuzilmasini noto'g'ri aks ettirishi mumkin. Bundan tashqari, turkiy tillar bo'yicha UD resurslarida bu masalada turlicha yechimlar uchraydi, natijada korpuslararo qiyoslanuvchanlik pasayadi. Shuning uchun bu yerda amaliy xulosa faqat "qaysi tahlilni tanlash" bilan cheklanmaydi. Tanlangan yechimni annotator qo'llanmasida aniq hujjatlashtirish va uni izchil qo'llash ham zarur [2].

Ingliz tili materialida esa sodda gaplarni ikkilamchi predikatsiya konstruksiyalari murakkablashtiradi. de Marneffe va boshqalar umumlashtirgan tavsifda bunday holatlar qo'shimcha predikativ komponent formal ravishda ergash gap shaklida ifodalanmagan bo'lsa-da, UD tahlilida klauzaga oid xcomp munosabatini talab qilishi mumkin bo'lgan tuzilmalar

sifatida izohlanadi. Bu holatni metodik jihatdan oldindan belgilab qo'yish muhim, chunki an'anaviy maktab grammatikasidagi ma'noda "sodda gap" deb hisoblangan tuzilma UD taqdimotida klauzaviy bog'lanishlar mavjud bo'lmasligini avtomatik tarzda kafolatlamaydi [3].

Umumiy va tilga xos murakkabliklar. Ingliz va o'zbek tillarini qiyoslash shuni ko'rsatadiki, sodda gaplarning razmetkasida uchraydigan qiyinchiliklarning sezilarli qismi annotatsiyaga xos umumiy muammolar bilan bog'liq. Bular qatoriga cho'qqini tanlash, yordamchi birliklarni to'g'ri birlashtirish hamda qabul qilingan konvensiyalarga qat'iy amal qilish kiradi. Aynan shu jihat UD'ning tillararo sxema sifatidagi asosiy mazmunini belgilaydi.

Har ikkala til uchun umumiy bo'lgan muammo predikatsiyaning leksik va funksional komponentlar o'rtasida "bo'linib" ketishidir. UD bu masalani tizimli ravishda hal qiladi, ya'ni leksik jihatdan mazmunli birlik bosh so'z sifatida tanlanadi va funksional markerlar unga tobe qilib birlashtiriladi. Xuddi shuningdek, predlog va postpozitsiyalarni nominal cho'qqi ostida case sifatida bir xil talqin qilish ingliz tilidagi predlogli guruhlar va o'zbek tilidagi postpozitsiyali guruhlarining qiyoslanishini ta'minlaydi.

Biroq ayrim qiyinchiliklar tilga xos bo'lib qoladi. O'zbek tili uchun, birinchidan, so'z tartibi o'zgaruvchan bo'lgan sharoitda sintaktik munosabatlarni belgilashda morfologiyaning yuqori darajadagi roli muhim hisoblanadi. Ikkinchidan, postverbal komplekslar masalasi va tanlangan konvensiyani, masalan, Uzbek-UT dagi kabi, aniq hujjatlashtirib qo'yish zarurati mavjud. Ingliz tilida esa shu kabi ta'sirga ega murakkablik ko'proq yordamchi fe'llar zanjiri va ikkilamchi predikatsiya konstruksiyalarida namoyon bo'ladi. Bunday holatlarda aux va cop munosabatlarini hamda xcomp tipidagi klauzaviy bog'lanishlarni ehtiyotkorlik bilan farqlash talab etiladi.

Xulosa. Ingliz va o'zbek tillaridagi sodda gaplarning sintaktik razmetkasi sifati hamda tillararo qiyoslanuvchanligi, avvalo, klauza cho'qqisini tanlash va yordamchi birliklarni to'g'ri birlashtirish bo'yicha izchil qarorlarga bog'liq. UD qiyoslanuvchanlikning nazariy asosini mazmunli so'zlar o'rtasidagi bog'lanishlarga ustuvorlik berish va cop, aux, case hamda koordinatsiya uchun yagona qoidalarni belgilash orqali ta'minlaydi.

Shu bilan birga, razmetkaning amaliy barqarorligi faqat universal tamoyillar sistematik tarzda muammoli bo'ladigan zonalar uchun oldindan hujjatlashtirilgan, tilga xos yechimlar bilan to'ldirilgan taqdirda yuzaga keladi. O'zbek tili uchun bu bog'lanishlarni belgilashda morfologik markerlarga tayangan holda ishlashni va postverbal komplekslar bo'yicha konvensiyani Uzbek-UT tajribasi asosida aniq belgilab qo'yishni anglatadi. Ingliz tili uchun esa "funktional" fe'llarni (aux, cop) ikkilamchi predikatsiya holatlaridan aniq farqlash, shuningdek xcomp talab qilinadigan konstruksiyalarni izchil ajratish muhim hisoblanadi.

ADABIYOTLAR

1. Abdurakhmonova, N. (2019). Dependency Parsing Based on Uzbek Corpus. Language Technologies for All (LT4All) konferensiyasi materiallari (164–168-betlar). European Language Resources Association (ELRA).
2. Akhundjanova, A., & Talamo, L. (2025). Universal Dependencies Treebank for Uzbek. Under-Resourced Languages and Domains uchun resurslar va reprezentatsiyalar bo'yicha uchinchi seminar (RESOURCEFUL-2025) materiallari (1–6-betlar). University of Tartu Library, Estonia.
3. de Marneffe, M.C., Manning, C. D., Nivre, J., & Zeman, D. (2021). Universal Dependencies. Computational Linguistics, 47(2), 255–308. https://doi.org/10.1162/coli_a_00402
4. Nivre, J., de Marneffe, M.C., Ginter, F., Goldberg, Y., Hajič, J., Manning, C. D., McDonald, R., Petrov, S., Pyysalo, S., Silveira, N., Tsarfaty, R., & Zeman, D. (2016). Universal Dependencies v1: A Multilingual Treebank Collection. Language Resources and Evaluation bo'yicha o'ninchi xalqaro konferensiya (LREC 2016) materiallari (1659–1666-betlar). European Language Resources Association (ELRA).
5. Universal Dependencies. UD Uzbek-UT: Treebank description (UD loyihasi; Uzbek-UT daraxtbanki tavsifi sahifasi).