JEZIKOSLOVLJE
23.2 (2022): 193-214

UDK 81'367.335.1
811.163.42'367.335.1
Pregledni članak
Prihvaćen za tisak 11. 10. 2022.
https://doi.org/10.29162/jez.2022.7

Daša Farkaš
Matea Filko
Filozofski fakultet Zagreb

Obilježavanje koordinacije u ovisnosnim bankama stabala

U ovome ćemo radu prikazati na koji se način obilježava koordinacija, i koordinacija surečenica i koordinacija skupina (engl. *phrases*), u ovisnosnim bankama stabala. Banke stabala temelje se na ovisnosnim pristupima sintaksi i preduvjet su za oblikovanje parsera, alata za automatsko sintaktičko označavanje rečenica. Posebno ćemo se pozabaviti označavanjem koordinacije unutar projekta *Universal Dependencies* (UD) (https://universaldependencies.org/). Projekt UD teži ujednačenome označavanju gramatičkih struktura u jezicima svijeta. Dosad je u sklopu projekta prikupljeno gotovo 200 banaka stabala za više od 100 jezika, među kojima je i jedna hrvatska ovisnosna banka stabala – Croatian UD. Prije nje za hrvatski je izrađena i Hrvatska ovisnosna banka stabala – HOBS (hobs.ffzg.hr), pri čijemu se obilježavanju slijedio pristup primijenjen pri izradi Praške ovisnosne banke stabala. Pristup primijenjen u izradi tih dviju banaka stabala razlikuje se u obilježavanju određenih sintaktičkih struktura. Prikazat ćemo temeljne razlike u obilježavanju koordinacije u dvjema hrvatskim ovisnosnim bankama stabala, a zatim ćemo se usredotočiti na problematične slučajeve koordinacije i rješenja za njihovo označavanje u dvjema bankama stabala te pokazati koje su prednosti i nedostaci ponuđenih rješenja.

Ključne riječi: ovisnosne banke stabala, koordinacija, Universal Dependencies, PDT, HOBS

JEZIKOSLOVLJE
23.2 (2022): 193-214

1. Uvod

Sintaktički označeni korpusi ili banke stabala (engl. *treebanks*) temelj su za razvoj računalnolingvističkih alata za obradu prirodnoga jezika, ponajprije sintaktičkih i semantičkih parsera. Međutim, osim široke upotrebe u razvoju alata za obradu prirodnoga jezika, banke stabala mogu se upotrebljavati i u drugim lingvističkim istraživanjima, napose psiholingvističkim i tipološkim, pri čemu su posebice korisne za istraživanja reda riječi u jezicima svijeta (de Marneffe i ostali 2021: 256). U ovisnosnim bankama stabala (engl. *dependency treebanks*) sintaktičko se označavanje korpusa provodi u skladu s pristupima ovisnosne gramatike ili gramatike zavisnosti ili dependencijalne gramatike (engl. *dependency grammar*)[1] L. Tesnièrea (1959). To znači da se u vrhu acikličkoga grafa – stabla – kojim se prikazuje struktura rečenice nalazi glagol, točnije glavni dio predikata, o kojemu onda ovise svi drugi rečenični dijelovi. Osim vršnoga čvora u kojemu se nalazi glavni dio predikata, i svaku preostalu riječ u rečenici predstavlja jedan čvor na stablu kojemu je dodijeljena jedna sintaktička funkcija. Upravo zbog takve strukture ovisnosnih banaka stabala koordinacija predstavlja izazov u označavanju. Naime, na jednu se ovisnosnu sintaktičku poziciju (na jedan čvor ovisnosnoga stabla) treba postaviti cijela koordinirana struktura koja se sastoji od najmanje triju elemenata: dvaju koordiniranih članova i veznika (npr. *snalažljiv i pametan*). Koordinacija dodatno predstavlja specifični sintaktički odnos kod kojega jedan član nije nadređen drugomu, kao što je slučaj u uobičajenome ovisnosnom odnosu. Stoga nam je u ovome radu cilj prikazati kako je obilježavanje koordinacije, i na razini rečenice i na razini skupina (engl. *phrases*), riješeno u hrvatskim ovisnosnim bankama stabala. Pritom ćemo se posebno usredotočiti na problematične slučajeve, poput ugniježđenih koordinacija ili koordinacija sa zajedničkim dopunama.

Kako bismo mogli pokazati razlike u obilježavanju koordinacije, u sljedećemu ćemo poglavlju najprije opisati dvije postojeće hrvatske ovisnosne banke stabala, a zatim ćemo prikazati na koji se način različiti tipovi koordinacije obilježavaju u dvama pristupima na kojima se te hrvatske ovisnosne banke stabala temelje. Na kraju ćemo se posvetiti razlikama u obilježavanju problematičnih slučajeva te pokazati koji je pristup ujednačeniji.

[1] Iako se u lingvističkoj terminologiji *dependencijalan* najčešće prevodi kao *zavisan*, u ovome ćemo se radu koristiti terminom ovisnosna gramatika s obzirom na to da se termin ustalio u računalnolingvističkim pristupima sintaktičkom označavanju korpusa nakon izrade Hrvatske ovisnosne banke stabala – prvoga sintaktički označenoga korpusa hrvatskoga jezika (Šojat 2008: 10).

JEZIKOSLOVLJE
23.2 (2022): 193-214

2. Hrvatske ovisnosne banke stabala

Kako je bez izrade ovisnosnih banaka stabala gotovo nemoguća računalna obrada prirodnoga jezika na sintaktičkoj razini, radi izrade alata za računalno sintaktičko obilježavanje hrvatskoga jezika dosad su izrađene dvije velike ovisnosne banke stabala[2] – Hrvatska ovisnosna banka stabala (HOBS, Agić i ostali 2014; Tadić 2007, http://hobs.ffzg.hr/hr/) te Croatian UD (Agić i Ljubešić 2015, https://github.com/UniversalDependencies/UD_Croatian-SET). Iako se obje temelje na principima ovisnosne gramatike, ipak postoje razlike u njihovoj strukturi i pravilima za označavanje pa ćemo ih ukratko opisati u sljedećim potpoglavljima.

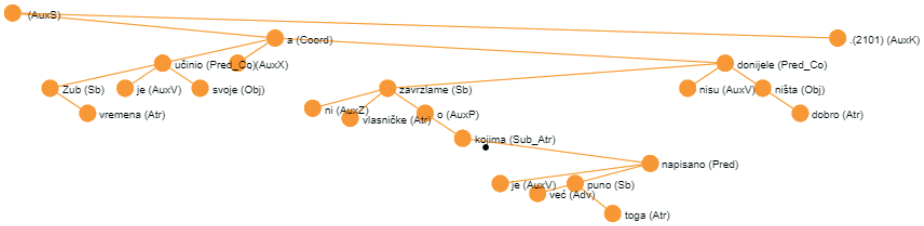
2.1. *Hrvatska ovisnosna banka stabala – HOBS*

Hrvatska ovisnosna banka stabala nastala je po uzoru na Prašku ovisnosnu banku stabala (Hajič i ostali 2000), koja je jedan od prvih i najznačajnijih sintaktički i semantički obilježenih korpusa temeljen na Tesniéreovoj ovisnosnoj gramatici. Poslužila je kao model za izgradnju banaka stabala za ostale morfološki bogate jezike, među kojima je i HOBS. HOBS se sastoji od 4 626 rečenica izlučenih iz potkorpusa novinskih tekstova CW2000 Hrvatskoga nacionalnog korpusa (HNK) (Tadić 2009). Rečenice su lematizirane i morfosintaktički označene u skladu s MulTextEast preporukama za hrvatski jezik.[3] Ukupno je 100 000 pojavnica u tim rečenicama označeno praškim formalizmom na sintaktičkoj (ili u praškoj terminologiji: analitičkoj) razini označavanja, pri čemu se prikazuju ovisnosni odnosi među elementima rečenice i označavaju sintaktičke funkcije tih elemenata.

Struktura rečenice vizualno se prikazuje stablom u obliku acikličkoga grafa. To drugim riječima znači da u stablu nema petlji i da između svaka dva čvora postoji točno jedan put. Svaka pojavnica, uključujući interpunkcijske znakove, predstavlja jedan čvor u stablu. Dakle, struktura pojedinačne rečenice prikazuje se jednim strukturnim stablom, u kojemu svaka pojavnica čini jedan čvor, a među čvorovima su obilježene poveznice koje odražavaju ovisnosne odnose. Primjer jedne označene rečenice u HOBS-u prikazan je na slici 1.

[2] Postoje još dvije hrvatske ovisnosne banke stabala o kojima ovdje nećemo govoriti. V. bilješku 8.

[3] MTE specifikacija prema kojoj je označen potkorpus CW2000 dostupna je na: http://nl.ijs.si/ME/Vault/V3/msd/html/msd.html#SECTION05800000000000000000.



Slika 1. Prikaz sintaktički označene koordinirane rečenice u HOBS-u

Iako radi jasnoće prikaza na stablu nisu prikazani svi elementi, svaki čvor u stablu ima trodijelnu strukturu koju čine: 1) izvorni oblik riječi, 2) morfološka oznaka i lema, 3) sintaktička oznaka. Sveukupno postoji skup od 28 analitičkih funkcija – sintaktičkih uloga u rečenici.[4] Pravila za označavanje na analitičkoj razini slijede osnovne principe ovisnosne sintakse. Glavni je cilj osigurati konzistentnost i eksplicitno prikazivanje odnosa među rečeničnim članovima. Iako se načelno slijede postavke tradicionalnih gramatika, pravila su često prilagođena i proširena za pojave koje nisu opisane u gramatikama. Naime, tradicionalne gramatike nisu računalnolingvistički usmjerene te niz čestih jezičnih pojava u njima nije eksplicitno, sustavno i dosljedno opisan. S obzirom na to da je jedna od temeljnih funkcija ovisnosnih banaka stabala treniranje alata za automatsko sintaktičko označavanje rečenice – parsera, ovisnosne banke stabala moraju sadržavati uobičajene rečenice prirodnoga jezika, kakve će kasnije i parser samostalno označavati. Naravno, takve su rečenice često sintaktički složene i sadržavaju sintaktičke strukture koje mogu biti neuobičajene, uobičajene, ali neopisane gramatikama, ili čak negramatične, no parser bi ih trebao neovisno o tome moći obraditi. Na temelju HOBS-a razvijen je i prvi parser za hrvatski jezik (Agić 2012), s početnom točnošću od 74,53 % u povezivanju pojavnica uz dodjelu sintaktičkih funkcija (LAS – *labeled attachment score*).

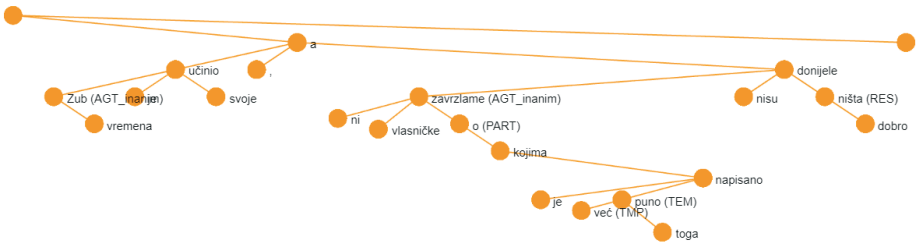
Potkorpus HOBS-a od 3 500 rečenica označen je i na semantičkoj razini. Glagolskim argumentima dodijeljene su semantičke uloge iz skupa od 17 uloga prema specifikaciji izrađenoj za hrvatski jezik.[5]

Osim rečenica iz novinskoga potkorpusa HNK-a, HOBS se sastoji od dodatnog manjeg korpusa rečenica iz tečajeva za hrvatski jezik dostupnih na mrežnome portalu HR4EU (Farkaš i dr. 2016, www.hr4eu.hr, www.hr4eu.eu). Oko 500 rečenica označeno je na sintaktičkoj i semantičkoj razini prema istome modelu kako bi se omogućila primjena računalnolingvističkih resursa i u glotodidaktici. Rečenice iz

[4] Popis sintaktičkih funkcija kojima je označen HOBS dostupan je na: http://hobs.ffzg.hr/static/docs/HOBS.pdf.

[5] Popis semantičkih uloga dostupan je na: http://hobs.ffzg.hr/static/docs/SRL_tagset.pdf.

obaju korpusa mogu se pretraživati SynSem vizualizatorom za prikaz sintaktičkih i semantičkih stabala. Semantičko stablo rečenice na slici 1 prikazano je na slici 2.



Slika 2. Prikaz semantički označene koordinirane rečenice u HOBS-u

2.2. *Croatian UD*

Croatian UD ovisnosna je banka stabala za hrvatski jezik izrađena u sklopu projekta *Universal Dependencies* (UD) (de Marneffe i dr. 2021). Projekt istovremeno pruža okvir za višejezično konzistentno morfosintaktičko označavanje korpusa, ali je i zajedničko otvoreno nastojanje da se sastave usporedivi korpusi za mnoge jezike te zbirka takvih korpusa. U samo šest godina u sklopu projekta prikupljene su 183 banke stabala za 104 jezika, u čijoj je izradi sudjelovalo više od 400 istraživača. Temeljni je cilj UD-a oblikovati jezične prikaze koji će biti korisni za morfosintaktička i semantička istraživanja te za razvoj alata za računalnu obradu prirodnoga jezika. Pri tom je naglasak na uočavanju sličnih struktura u različitim jezicima neovisno o redu riječi, morfološkim obilježjima te prisutnosti ili odsutnosti funkcionalnih riječi (de Marneffe i dr. 2021: 255–56). Upravo je zbog mogućnosti uočavanja sličnih jezičnih struktura UD pogodan za razvoj višejezičnih parsera, učenje stranih jezika ili tipološka istraživanja.

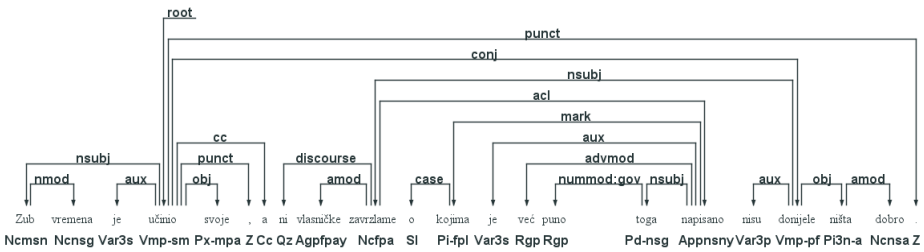
Iako se često promatra kao praktičan alat za označavanje, UD je koherentna morfološka i sintaktička teorija koja se jednim dijelom gradi na terminologiji i konceptima mnogih ranijih gramatičkih teorija, ali je nadograđena i oblikovana tako da je primjenjiva na računalni gramatički opis jezika svijeta (de Marneffe i dr. 2021: 256). Dizajn UD-a rezultat je kompromisa kako bi se zadovoljilo šest zahtjeva:

- mora biti jezično utemeljen i u skladu s lingvističkim analizama za pojedine jezike
- mora pružati dobru podlogu za tipološka istraživanja kako bi se mogle iščitati sličnosti (i razlike) među jezicima i jezičnim porodicama
- mora biti pogodan za brzo i konzistentno ručno označavanje

- mora biti lako razumljiv i upotrebljiv za nelingviste (npr. one koji uče jezik ili računalne stručnjake kojima treba osnovna obrada prirodnoga jezika; upravo je stoga dizajn UD-a naklonjen tradicionalnim gramatičkim opisima i terminologiji)
- mora biti pogodan za računalno parsanje s visokom točnošću
- mora dobro podupirati krajnje zadatke jezičnoga razumijevanja (izvlačenje odnosa, čitanje s razumijevanjem, strojno prevođenje…) (de Marneffe i dr. 2021: 302–3).[6]

Cjelovita gramatička teorija na kojoj počiva UD prikazana je u de Marneffe i dr. (2021), a mi ćemo se usredotočiti na temeljne postavke označavanja sintaktičkih odnosa, koje su primijenjene i za označavanje Croatian UD-a.

Sintaktički se odnosi nazivaju gramatičkim odnosima (engl. grammatical relations between words) i ostvaruju se binarno unutar skupina. Svaki element ovisan o glavi skupine i svaka funkcionalna riječ koja pripada glavi povezani su s glavom jednom od 37 sintaktičkih oznaka. Pritom se razlikuje glava imenske skupine i glava surečenice, dok ovisni element može biti imenska riječ (engl. nominal), surečenica (engl. *clause*) ili modifikator (engl. *modifier*). UD dopušta i jezično specifične podvrste postojećih odnosa ako je potrebno, no skup je dopuštenih odnosa zatvoren (de Marneffe i dr. 2021: 265). Tako postavljen univerzalni inventar kategorija zajedno s uputama za označivače osigurava konzistentnost pri označavanju sličnih konstrukcija u različitim jezicima.



Slika 3. Prikaz koordinirane rečenice obilježene u skladu s Croatian UD-om[7]

Croatian UD (Agić i Ljubešić 2015) sastoji se od ukupno 6 914 rečenica prikupljenih iz SETimes-HR korpusa,[8] koji se, kao i CW2000 na temelju kojega je izrađen

6 V. i https://universaldependencies.org/introduction.html.

7 Rečenica je označena u alatu DGA (Dependency Grammar Annotator; http://medialab.di.unipi.it/Project/QA/Parser/DgAnnotator/) i nije dio Croatian UD-a.

8 Croatian UD nadogradnja je banke stabala SETimes.HR (Agić, Ljubešić 2014). Rečenice iz te banke stabala prebačene su u UD format obilježavanja i dopunjene novim rečenicama. SETimes.HR i Babel Dependency Treebank of Public Messages in Croatian (Merkler i dr. 2013) dvije su dodatne ovisnosne banke stabala za hrvatski jezik označene pojednostavljenim skupom od 15 analitičkih oznaka, međutim

JEZIKOSLOVLJE
23.2 (2022): 193-214

HOBS, sastoji od novinskih tekstova. Rečenice su lematizirane i morfosintaktički označene te prilagođene UD formatu oznaka za vrste riječi i morfosintaktičke kategorije. Sintaktički odnosi ručno su označeni u skladu s uputama za označavanje u sklopu projekta UD. Na temelju banke stabala Croatian UD napravljen je model za lematizaciju, morfosintaktičko označavanje i parsanje hrvatskoga jezika koji je dostupan kao dio alata za lančanu obradu prirodnoga jezika (engl. *pipeline*) UD pipe (https://lindat.mff.cuni.cz/services/udpipe/). UD pipe dakle omogućava automatsku analizu hrvatskih tekstova i prikaz rečeničnih struktura u obliku ovisnosnih stabala u skladu s UD-om. Rečenica prikazana na slici 1 u HOBS-u označena u UD formatu prikazana je na slici 3.

Kao što možemo primijetiti ako usporedimo sliku 2 i sliku 3, koordinacija se u dvjema hrvatskim ovisnosnim bankama označava različito. U sljedećemu ćemo poglavlju stoga objasniti razlike između tih dvaju pristupa koordinaciji.

3. Opis koordinacije u različitim pristupima izradi ovisnosnih banaka stabala

Koordinacija predstavlja izazov za označavanje u ovisnosnim bankama stabala iz dvaju razloga. Prvo, sintaktičke se funkcije dodjeljuju jednoj riječi na jednome čvoru i to u binarnome odnosu, a kod koordinacije se radi o cijeloj koordiniranoj strukturi od najčešće barem triju članova (dva koordinirana člana i veznik), što nadilazi mogućnosti formalizma utemeljenoga na odnosima između dvaju članova. Drugo, kod koordinacije se ne radi o uobičajenome ovisnosnom odnosu kod kojega je jedan član nadređen drugomu, a u ovisnosnim je pristupima to temeljno načelo obilježavanja odnosa. Osim toga, u ovisnosnim bankama stabala moraju se označiti i interpunkcijski znakovi, pa se i zarezi u koordinativnoj funkciji moraju dosljedno i jednoznačno označiti u cijelome korpusu. Upravo su zbog tih izazova tvorci različitih ovisnosnih banaka stabala pronašli različita rješenja za ujednačeno označavanje koordinacije.

3.1. *Označavanje koordinacije u Praškoj ovisnosnoj banci stabala*

U uputama za označavanje Praške ovisnosne banke stabala – PDT-a (Hajič i dr. 1999), na temelju koje je izrađen HOBS, detaljno se opisuju različiti tipovi koordi-

njihovi se autori ne bave detaljnije pristupom koordinaciji i problematičnim slučajevima. Koordinacija je riješena tako da se vezniku ili zarezu koji ima vezničku ulogu dodjeljuje oznaka Co, a koordinirani članovi dobivaju oznaku sintaktičke funkcije koju imaju u rečenici i vežu se na veznik kao glavu skupine. Time nastavljaju tradiciju PDT-a da funkcionalni elementi budu glave skupina koje uvode te svode sve tipove koordinacije na jedan oblik uveden veznikom (Merkler i dr. 2013: 494–495). Pliće oznake omogućavaju im veću točnost parsanja, no time se gubi uvid u specifičnosti pojedinih sintaktičkih struktura. Kako je SETimes.HR nadograđen na UD kao noviji i perspektivniji format označavanja, tom se ovisnosnom bankom stabala ovdje nećemo baviti.

nacija i načini njihova označavanja.[9] Koordinacija se može pojaviti 1) između dvaju ili više članova rečenice, pri čemu u koordinaciji mogu biti članovi neovisno o funkciji (subjekti, objekti, atributi)[10] (npr. *Kupio je **kruške i jabuke**.*) ili 2) između dviju ili više surečenica, pri čemu se u koordinaciji mogu naći i nezavisne (npr. *Ispričao se i otišao.*) i zavisne surečenice (npr. *Rekao je da će doći i da će donijeti kolače.*). Međutim, pri označavanju se ta dva tipa koordinacije ne razlikuju, odnosno i koordinaciji na razini rečeničnih članova i koordinaciji na razini rečenica dodjeljuje se ista sintaktička funkcija, baš kao što se ne razlikuju ni različite vrste koordinacija (sastavna, rastavna i sl.).

Osnovno je pravilo pri označavanju koordinacije u PDT-u da je korijen koordiniranih članova, odnosno glava koordinirane skupine veznik, i on je nadređeni član koji nosi sintaktičku funkciju *Coord*. Ako se radi o koordinaciji nezavisnih surečenica, veznik će ujedno biti i prvi čvor ispod vršnoga čvora stabla (v. sliku 1 – vršni član u PDT-u nije predikat nego prazni čvor *AuxS* koji označava rečenicu). U ostalim slučajevima veznik ovisi o čvoru o kojemu ovisi cijela koordinirana skupina, primjerice veznik koji povezuje koordinirane objekte ovisit će o predikatu, veznik koji povezuje atribute ovisit će o imenici na koju se atributi odnose, a veznik koji povezuje zavisne surečenice ovisit će o predikatu glavne surečenice. Ostali članovi koordinacije dobivaju analitičku funkciju pripadajuće sintaktičke kategorije s dodatkom *_Co*. Tako će koordinirani subjekti imati oznaku *Sb_Co*, a koordinirani objekti *Obj_Co*. Važno je da su svi članovi koordinirane skupine obilježeni istom analitičkom funkcijom (v. i bilješku 10).

U slučaju koordiniranih lista (npr. *Proučavali smo sintaktičke funkcije: subjekte, objekte, predikate itd.*) nositelj je koordinirane skupine završni član (*itd.*), koji u tome slučaju dobiva oznaku *Coord*. Zarezi koji povezuju koordinirane članove označavaju se istom funkcijom kao i ostali zarezi (*AuxX*), a vežu se na veznik kao glavu koordinirane skupine.

S obzirom na to da označavanje u ovisnosnim bankama stabala ne prelazi rečenične granice, dodatno je opisan slučaj jednostruke koordinacije, pri čemu se prvi član koordinacije nalazi u prethodnoj rečenici (npr. *Ali ja to nisam čuo.*). U tome slučaju označavanje slijedi sva pravila označavanja koordinacije s najmanje dvama članovima, osim što je označen samo jedan član (čuo = Pred_Co).

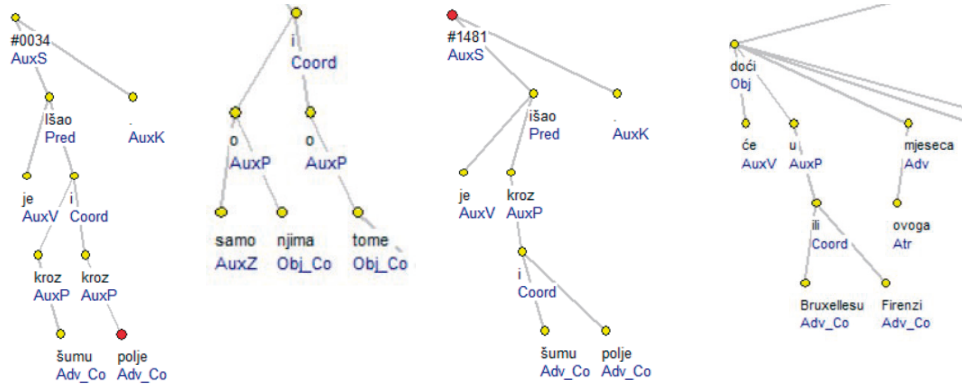
[9] Upute za obilježavanje koordinacije na analitičkoj razini dostupne su na: https://ufal.mff.cuni.cz/pdt2.0/doc/manuals/en/a-layer/html/ch03s04.html#cozachyc i prema njima je izrađeno ovo potpoglavlje, s time da su se za sve vrste koordinacije navodili vlastiti hrvatski primjeri.

[10] Naravno, važno je da oba člana imaju istu funkciju: u koordinaciji mogu biti dva ili više subjekta, dva ili više objekta, dva ili više atributa…, ali ne mogu biti članovi koji imaju različite sintaktičke funkcije, npr. subjekt i objekt ili objekt i atribut.

Autori PDT-a opisuju i nekoliko posebnih slučajeva koordinacije: koordinaciju prijedložno-padežnih izraza, koordinaciju surečenica te slučajeve zajedničke dopune koordiniranih članova.

3.1.1. Koordinacija prijedložno-padežnih izraza

Pri koordinaciji prijedložno-padežnih izraza označavanje se razlikuje ovisno o tome pojavljuje li se prijedlog uz oba koordinirana člana (npr. *Išao je kroz šumu i kroz polje.*) ili samo uz jedan koordinirani član (npr. *Išao je kroz šumu i polje.*), no u obama slučajevima prijedlozi kao funkcionalne riječi (*AuxP*) ne dobivaju oznaku koordiniranih članova, nego je dobivaju punoznačne riječi uz oznaku sintaktičke funkcije koju imaju u rečenici (*Sb_Co*, *Adv_Co*, *Obj_Co*…). Kada se prijedlog pojavljuje uz oba koordinirana člana, glava koordinacije je veznik, prijedlozi se vežu izravno na veznik kao glavu koordinacije, a imenski dio prijedložno-padežnoga izraza ovisi o njima. Međutim, kada se prijedlog pojavljuje samo uz jedan (prvi) član koordinacije, onda je taj prijedlog glava cijele skupine (iako i dalje ima oznaku *AuxP*), o njemu ovisi veznik, a o vezniku ovise koordinirani članovi koji uz svoju sintaktičku funkciju dobivaju oznaku *_Co* (v. sliku 4)).

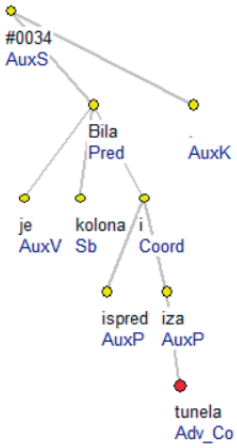


Slika 4. Označavanje koordinacije sa zajedničkim prijedlogom u HOBS-u[11]

Zanimljiv je i slučaj koordinacije prijedložno-padežnih izraza u kojima se pojavljuju dva prijedloga, ali samo drugi koordinirani član, dok je prvi član u elipsi (npr. *Bila je gužva ispred i iza tunela.*). U takvim je slučajevima veznik glava skupine na koju se vežu oba prijedloga, dok se koordinirani član veže na drugi prijedlog i dobiva oznaku svoje sintaktičke funkcije uz dodatnu oznaku *_Co* (v. sliku 5).[12]

[11] Na slikama su prikazani naši primjeri iz teksta i nakon njih stvarni primjeri iz HOBS-a.

[12] Iako nismo našli zabilježen takav primjer u HOBS-u, koordinacija prijedloga zabilježena je u hrvatskome jeziku od početka 20. stoljeća naovamo (v. Petrović 2021) te je danas potvrđena u hrvatskim korpusima (Petrović 2021, Katunar 2014).



Slika 5. Koordinacija prijedložno-padežnih izraza s elipsom prvoga koordiniranog člana

3.1.2. Koordinacija surečenica

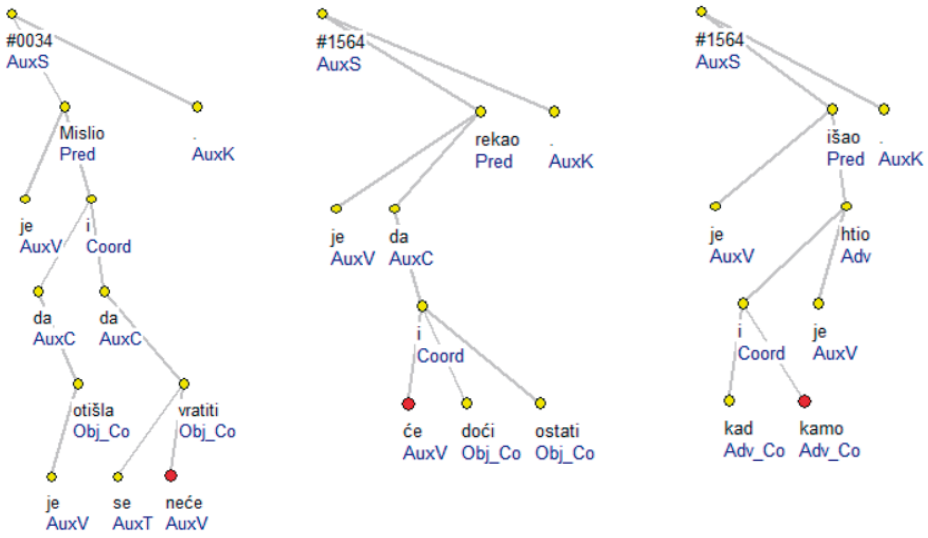
Koordinacija surečenica[13] (npr. *Mislio je da je otišla i da se neće vratiti.*) označava se prema istome načelu kao i koordinacija prijedložno-padežnih izraza. Glava koordinacije i ovdje je veznik te izravno ovisi o predikatu glavne surečenice. O nezavisnosloženome vezniku ovise oba zavisnosložena veznika kao glave zavisnih surečenica, no oni ne dobivaju oznaku *_Co*, nego zadržavaju svoju uobičajenu oznaku *AuxC*. O veznicima zavisnih surečenica izravno ovise predikati zavisnih surečenica, koji dobivaju oznaku ovisno o vrsti zavisne surečenice uz dodatak *_Co*. Tako će predikati objektnih surečenica (kao u našemu primjeru) dobiti oznake *Obj_Co*, predikati subjektnih surečenica *Pred_Co*, a predikati priložnih surečenica *Adv_Co* (v. sliku 6).

U nekim slučajevima veznik zavisnih surečenica pojavljuje se samo jednom, ispred prve surečenice (npr. *Rekao je da će doći i ostati.*). U tome slučaju veznik zavisne surečenice izravno ovisi o predikatu glavne surečenice, a nezavisnosloženi veznik ovisi o zavisnosloženome vezniku kao glava koordinirane skupine. Predikati zavisnosloženih surečenica i ovdje dobivaju oznaku ovisno o vrsti zavisne surečenice uz dodatak *_Co* koji ukazuje na to da se radi o koordiniranim članovima (u našemu primjeru *Obj_Co*, v. sliku 6).

[13] U uputama za označavanje Praške ovisnosne banke stabala taj se tip odnosa naziva koordinacijom umetnutih surečenica. Međutim, umetnute surečenice u praškoj terminologiji ne poklapaju se u potpunosti s umetnutim surečenicama u hrvatskoj gramatikološkoj tradiciji, gdje se umetnutom rečenicom smatra ona surečenica koja prekida glavnu surečenicu i od nje je s obiju strana odvojena zarezom: *Marko, koji je mislio da je otišla i da se neće vratiti, potpuno je pogriješio.* V. npr. https://gramatika.hr/pravilo/zavisnoslozene-recenice/90/. U praškoj tradiciji to obuhvaća koordinaciju zavisnih surečenica.

JEZIKOSLOVLJE
23.2 (2022): 193-214

Isto tako, opisan je i slučaj u kojemu imamo dva zavisnosložena veznika i samo jedan predikat (npr. *Išao je kad i kamo je htio.*). S obzirom na to da se u hrvatskome jeziku udvajaju veznici koji su po postanku druga vrsta riječi, oni se prema pravilima PDT-a ne označavaju kao veznici, nego im se dodjeljuje sintaktička funkcija koju bi imali u jednostavnoj surečenici pa samim time i ovise o predikatu zavisne surečenice, a međusobno su u odnosu koordinacije (v. sliku 6).[14]



Slika 6. Različite mogućnosti koordinacije surečenica

3.1.3. Zajednička dopuna koordiniranih članova

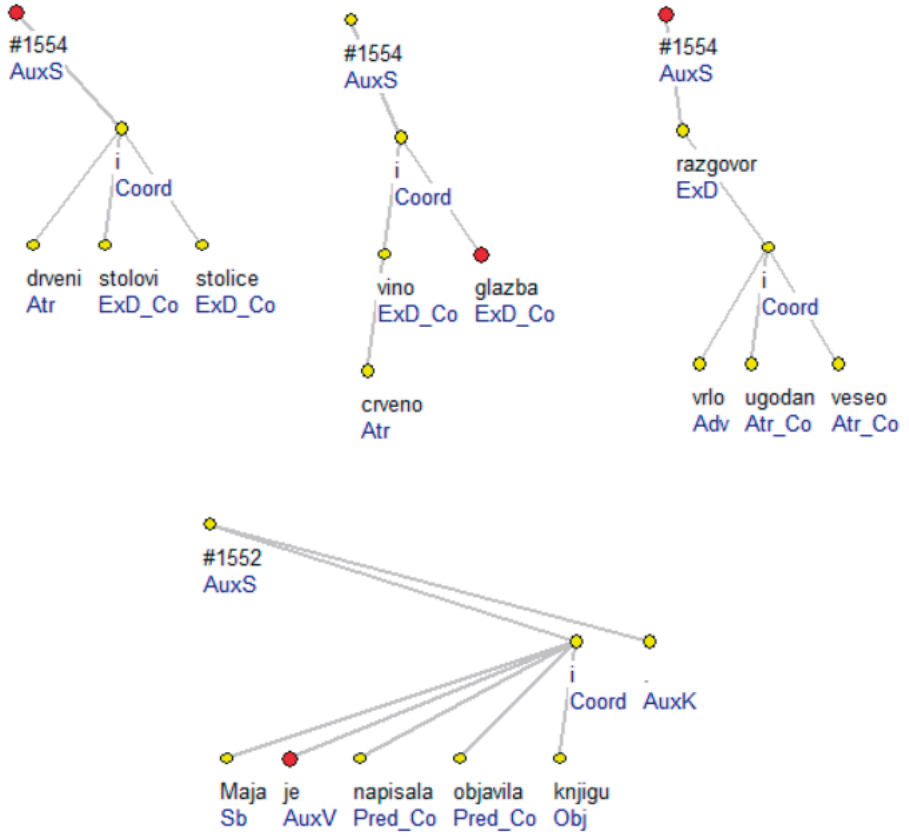
Zajednička dopuna koordiniranih članova također predstavlja problematičan slučaj pri obilježavanju u ovisnosnim bankama stabala zato što bi u strukturi trebala biti prikazana kao ovisna o koordiniranome članu, ali time se gubi podatak da dopuna istovremeno modificira oba člana. Kako bi se taj podatak sačuvao, u PDT-u se u takvim slučajevima dopuna obilježava kao izravno ovisna o nezavisnome vezniku te dobiva oznaku sintaktičke funkcije koju ispunjava, ali bez dodatne oznake *_Co* kako bi bilo jasno da se ne radi o koordiniranome članu, nego o modifikatoru.

[14] U PDT-u se opisuje slučaj koordinacije „pravih" veznika, ali svako takvo udvajanje rezultira pleonazmom: šel, ***protože*** *a* ***poněvadž*** *chtěl* 'išao je **zato** i **jer** je htio'. Takve primjere nismo zabilježili u hrvatskim ovisnosnim bankama stabala, ali obilježavaju se tako da nezavisnosloženi veznik izravno ovisi o predikatu glavne surečenice, o njemu ovise zavisnosloženi veznici (oznaka *AuxC*), a predikat zavisne surečenice ovisi o drugome vezniku i dobiva oznaku ovisno o vrsti zavisnosložene surečenice s dodatnom oznakom *_Co* (u gornjemu slučaju *chtěl* = *Adv_Co*). Problematično je to da će se, kao što vidimo, koordinacija veznika od slučaja do slučaja obilježavati različito.

U slučaju zajedničkoga pridjevskoga atributa (npr. ***drveni*** *stolovi i stolice*) veznik je glava skupine, a o njemu izravno ovise i atribut (označen kao *Atr*) i koordinirani članovi (dakle *drveni* neće ovisiti o *stolovi*, nego o vezniku *i*). Time se strukturno obilježava da atribut modificira cijelu koordiniranu skupinu, za razliku od primjera u kojima atribut modificira samo jedan koordinirani član (npr. *crveno vino i glazba*), gdje će atribut *crveno* izravno ovisiti o koordiniranome članu *vino* (v. sliku 7).

Na isti će način biti označen i priložni modifikator uz pridjevske atribute (npr. *vrlo veseo i ugodan razgovor*). S obzirom na to da koordinirana skupina opisuje imenicu *razgovor*, veznik će izravno ovisiti o imenici, a o njemu će ovisiti oba koordinirana člana (*Atr_Co*) i priložni modifikator (*Adv*) (v. sliku 7).

U slučaju zajedničkih argumenata (npr. *Maja je napisala i objavila knjigu.*), argumenti će, kao i koordinirani predikati (*Pred_Co*), biti povezani na nezavisnosloženi veznik i dobiti oznaku svoje sintaktičke funkcije (*Maja = Sb*, *knjigu = Obj*) (v. sliku 7).



Slika 7. Obilježavanje zajedničke dopune koordiniranih članova

JEZIKOSLOVLJE
23.2 (2022): 193-214

3.2. *Označavanje koordinacije u sklopu projekta* Universal Dependencies

Osnovno pitanje kojim su se vodili istraživači okupljeni oko projekta *Universal Dependencies* jest može li se koordinacija, kao specifičan tip strukture koji ne uključuje ovisnosni, nego ravnopravni odnos, svesti na pojednostavljenu strukturu ovisnosnoga stabla, a da se pritom ne izgube važne informacije (de Marneffe i ostali 2021: 276–77).[15] Naime, format ovisnosnoga stabla ne dopušta simetrične odnose (kakav je koordinacija) pa je bilo potrebno svesti koordinaciju na specifični asimetrični odnos unutar kojega će biti jasno obilježeni svi elementi koordinacije: koordinirani članovi, veznici i zarezi. Za označavanje koordinacije tako su osmislili posebnu oznaku: *conj*, kojom se svi koordinirani članovi u koordinaciji vežu na prvi koordinirani član, dok prvi član u koordinaciji dobiva oznaku sintaktičke funkcije koju ima u rečenici. Prvi je član koordinacije tako glava cijele koordinirane skupine i o njemu ovise svi drugi koordinirani članovi. Nezavisnosloženi veznici označeni su oznakom *cc* i vežu se na član koji im se nalazi prvi s desne strane. Takvim se označavanjem, kao i u PDT-u, omogućava označavanje rečenica koje počinju nezavisnosloženim veznicima. Isto pravilo kao i za veznike vrijedi i za zareze (oznaka: *punct*) – svaki se zarez veže na koordinirani član koji mu je prvi zdesna. Uniformno označavanje zareza posebno je važno zbog slučajeva asindetske koordinacije (v. i sliku 12.).

Na isti se temeljni način obilježava koordinacija pojedinačnih članova (*jabuke i kruške*) i koordinacija surečenica (*Uzeo je godišnji i dobro se odmorio.*). No, autori UD-a dodatno opisuju i dva posebna tipa koordinacije: koordinaciju sa zajedničkim dopunama te ugniježđenu koordinaciju, koje ćemo opisati u sljedećim potpoglavljima.

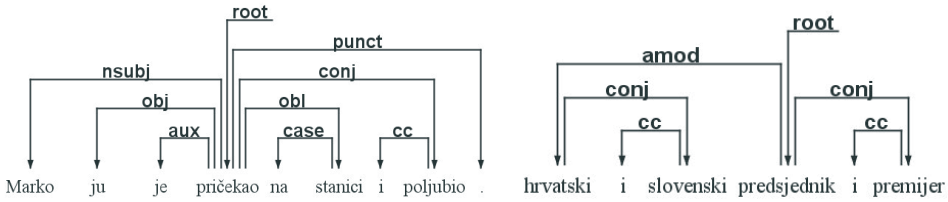
3.2.1. Koordinacija sa zajedničkim dopunama

Označavanje koordinacije u UD-u ne omogućava razlikovanje između dopune prvomu koordiniranom članu i zajedničke dopune cijele koordinacije. Primjerice, u koordiniranoj rečenici s elipsom *Marko ju je pričekao na stanici i poljubio*. imamo čak dvije zajedničke dopune: subjektnu dopunu *Marko* i objektnu dopunu *ju*. U osnovnome prikazu obje su dopune vezane na prvi koordinirani član, u ovome slučaju na predikat prve nezavisnosložene surečenice *pričekao* (v. sliku 8). Međutim, u obogaćenome prikazu moguće je dodatnim vezama spojiti *Marko* kao subjekt i *ju* kao objekt predikata *poljubio*.

Na isti se način mogu označiti zajednički atributi. U primjeru dvostruke koordinacije *hrvatski i slovenski predsjednik i premijer* oba pridjeva odnose se na obje

[15] Pri prikazu koordinacije u UD-u, osim navedenim člankom, vodili smo se objašnjenjima u uputama za označivače: https://universaldependencies.org/u/dep/all.html#al-u-dep/conj.

imenice (da se odnose samo na jednu, svaki bi pridjev stajao uz svoju imenicu). Pri označavanju najprije ćemo u jednu koordiniranu skupinu povezati pridjeve *hrvatski i slovenski*, tako da *slovenski* povežemo s *hrvatski* i dodijelimo mu oznaku *conj* te isto učinimo s veznikom *i*, samo što njemu dodijelimo oznaku *cc*. Zatim u drugu koordiniranu skupinu povežemo *predsjednik i premijer*, opet tako da *premijer* povežemo s *predsjednik* kao glavom skupine oznakom *conj*, a veznik *i* oznakom *cc*. U osnovnome prikazu zatim glavu koordinirane pridjevske skupine *hrvatski* vežemo kao pridjevski atribut (*amod*) na glavu koordinirane imenske skupine *predsjednik*. (v. sliku 8). U obogaćenome prikazu dodatno vežemo *hrvatski* na *premijer*, a *slovenski* i na *predsjednik* i na *premijer*, sve s oznakom pridjevskoga atributa. Time se eksplicitno označavaju veze koje se inače mogu i algoritamski pronaći na temelju osnovnoga prikaza, za razliku od primjera s dijeljenim subjektom i objektom među surečenicama, gdje informacije iz obogaćenoga prikaza ne bi bile dostupne samo na temelju osnovnoga prikaza.



Slika 8. Obilježavanje koordinacije sa zajedničkom dopunom u UD-u

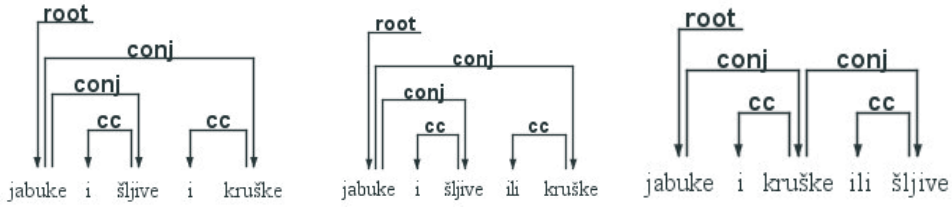
3.2.2. Ugniježđena koordinacija

Osnovni način obilježavanja koordinacije ne može prikazati razlike među različitim tipovima ugniježđene koordinacije (engl. *nested coordination*). Ugniježđena ili umetnuta koordinacija podrazumijeva slučajeve koordinacije unutar koordinacije. Ugniježđena koordinacija razlikuje se od višestruke koordinacije po tome što kod višestruke koordinacije imamo više od dvaju koordiniranih članova na istoj razini (npr. *Marko, Matija i Ivan* pomoći će mi oko zadatka), dok kod ugniježđene koordinacije članovi nisu na istoj razini (npr. *Marko i Matija ili Ivan pomoći će mi oko zadatka.*[16]). Ako uzmemo tri moguća slučaja koordinacije sa samo trima članovima:

- A, B, C: *jabuke i šljive i kruške*
- (A, B), C: *jabuke i šljive ili kruške*
- A, (B, C): *jabuke i kruške ili šljive*

[16] Dakle, u ovome slučaju imamo dvije koordinacije. Prvi je koordinirani član prve koordinacije *Marko*, a drugi je koordinirani član koordinacija *Matija ili Ivan*. Zato kažemo da se radi o koordinaciji unutar koordinacije.

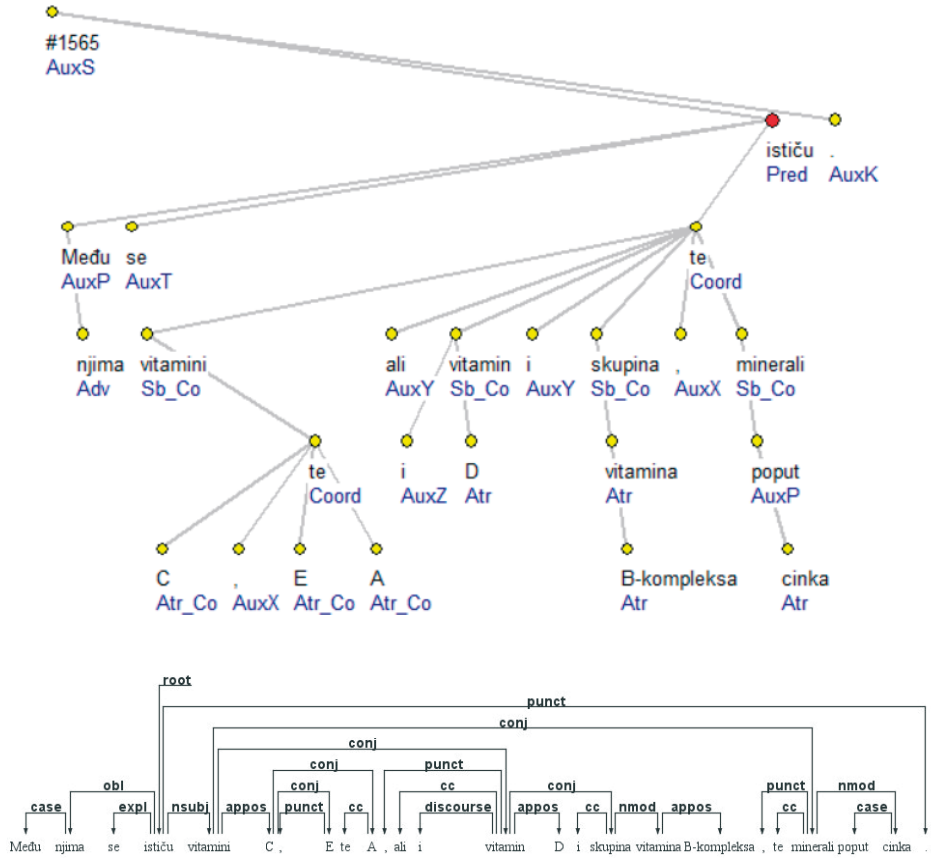
i pokušamo ih prikazati ovisnosnim stablom, vidjet ćemo da će prve dvije mogućnosti rezultirati potpuno istom strukturom,[17] dok će se samo treća od njih i strukturno razlikovati (v. sliku 9).



Slika 9. Obilježavanje ugniježđene koordinacije u UD-u

Ipak, pravila za obilježavanje ugniježđene koordinacije u UD-u omogućavaju obilježavanje finih značenjskih nijansi u slučajevima višestruke i ugniježđene koordinacije. Kao što možemo vidjeti na slici 10, hijerarhijski se *vitamin D i skupina vitamina B-kompleksa* izdvajaju kao izdvojena koordinirana skupina koja je kao cjelina drugi koordinirani član višestruke koordinacije na višoj razini, uz članove *vitamini* i *minerali*. Slučajevi ugniježđene koordinacije nisu obuhvaćeni priručnikom za označavanje PDT-a, a isto tako nijedan opisani tip koordinacije koji bi se mogao primijeniti na primjere ugniježđene koordinacije (koordinacija više nezavisnosloženih surečenica, koordinacija listi) ne može opisati razliku u hijerarhiji nekoliko koordinacija u istoj rečenici pa je označavanje ugniježđene koordinacije prednost UD pristupa koordinaciji.

[17] Smatramo kako bi u UD-u razlika između tih dviju struktura mogla biti prikazana u obogaćenome prikazu tako da u drugome slučaju *kruške* bude vezano oznakom *conj* i na *jabuke* i na šljive. Više o ugniježđenoj koordinaciji i mogućim rješenjima, od kojih nijedno nije potpuno zadovoljavajuće, v. u (Przepiórkowski i Patejuk 2019).



Slika 10. Obilježavanje višestruke koordinacije s ugniježđenom koordinacijom u HOBS-u i UD-u

Možemo uočiti kako se i u jednome i u drugome pristupu koordinacija izdvaja kao poseban slučaj i izazov upravo zato što se ravnopravnim položajem koordiniranih članova razlikuje od ovisnosne strukture svih ostalih tipova sintaktičkih odnosa. Temeljna razlika između dvaju pristupa proizlazi iz različita tretiranja funkcionalnih riječi. Naime, u PDT-u se prijedlozi i veznici promatraju kao glave struktura koje uvode (prijedložno-padežnih skupina i surečenica), dok u UD-u funkcionalne riječi ovise o punoznačnim riječima: prijedlozi o imenskim riječima (i dobivaju oznaku *case*), a veznici o predikatima surečenica koje uvode. Ta je struktura u obama slučajevima ostala očuvana i pri označavanju koordinacije (v. sliku 11). PDT time postiže da su koordinirani članovi na istoj razini u stablu, ali im problem stvaraju slučajevi sa zajedničkim prijedlogom te će se razlikovati struktura koordiniranih skupina *Išao je* ***kroz šumu i kroz polje*** i *Išao je* ***kroz šumu i polje*** (v. poglavlje 3.1.1.). U UD-u će osnovna struktura takvih skupina ostati očuvana jer je to bila i namjera pri odluci o

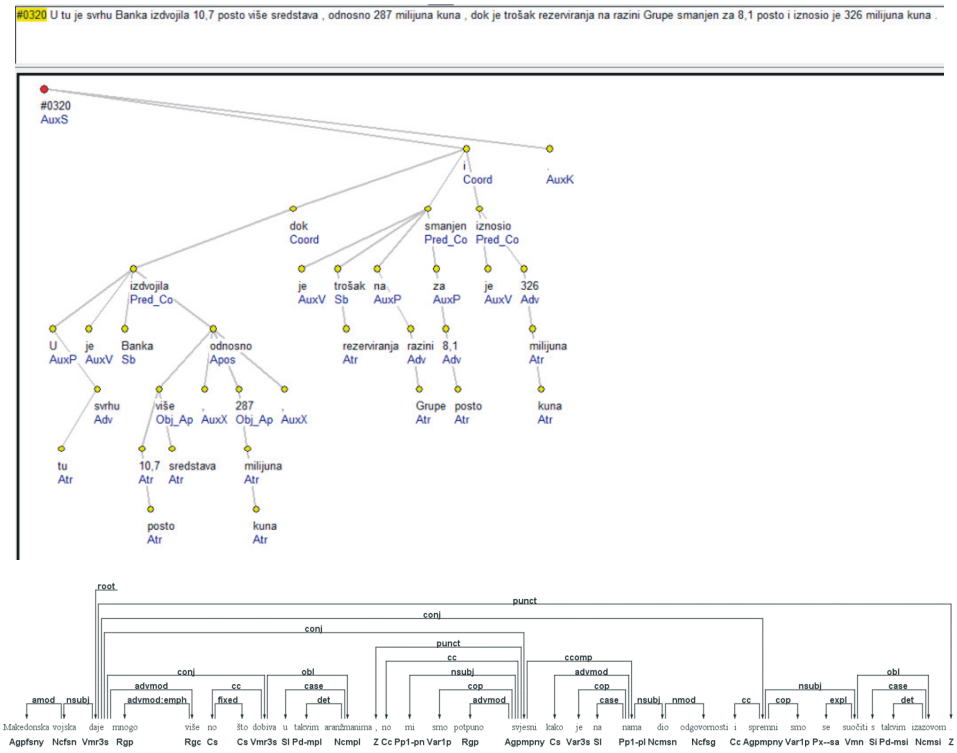
JEZIKOSLOVLJE
23.2 (2022): 193-214

položaju funkcionalnih riječi. Naime, kako je jedan od ciljeva UD-a uvid u tipološke sličnosti i razlike među jezicima, htjela se postići ujednačenost analize u morfološki bogatim jezicima tako da se, primjerice, imenske riječi koje dolaze bez prijedloga nalaze na istome mjestu u strukturnome stablu kao i imenske riječi uvedene prijedlogom (ili poslijelogom, vrsta adpozicije ne utječe na strukturu stabla, što je još jedna prednost takve odluke).[18] Isto tako, činjenica da svi koordinirani članovi ovise o prvome članu, a ne o vezniku, olakšava označavanje višestrukih nezavisnosloženih rečenica u UD-u, za razliku od PDT-a, gdje se mora odlučiti koji će koordinirani član doći niže u strukturi, što je u nekim slučajevima potpuno proizvoljno i ne odražava stvarne odnose među članovima (v. sliku 11).

Nadalje, u PDT-u će oba koordinirana člana dobiti oznaku sintaktičke funkcije koju imaju u rečenici s dodatnom specifikacijom *_Co*, dok će u UD-u prvi član dobiti oznaku sintaktičke funkcije koju ima u rečenici, a ostali će članovi s njim biti povezani oznakom *conj*. Oba pristupa omogućavaju izdvajanje koordiniranih članova, ali mala je prednost pristupa u UD-u to što omogućava manji skup oznaka, a to je opet u skladu s jednom od težnji UD-a da omogući brzo i konzistentno ručno označavanje.

Isto tako, UD omogućava i obogaćeni prikaz, u kojemu se u slučaju zajedničkih dopuna može jasno odrediti na koje se sve članove u koordinaciji dopune odnose, a potencijalno može riješiti i problem ugniježđene koordinacije (v. bilješku 17), što u PDT-u nije moguće.

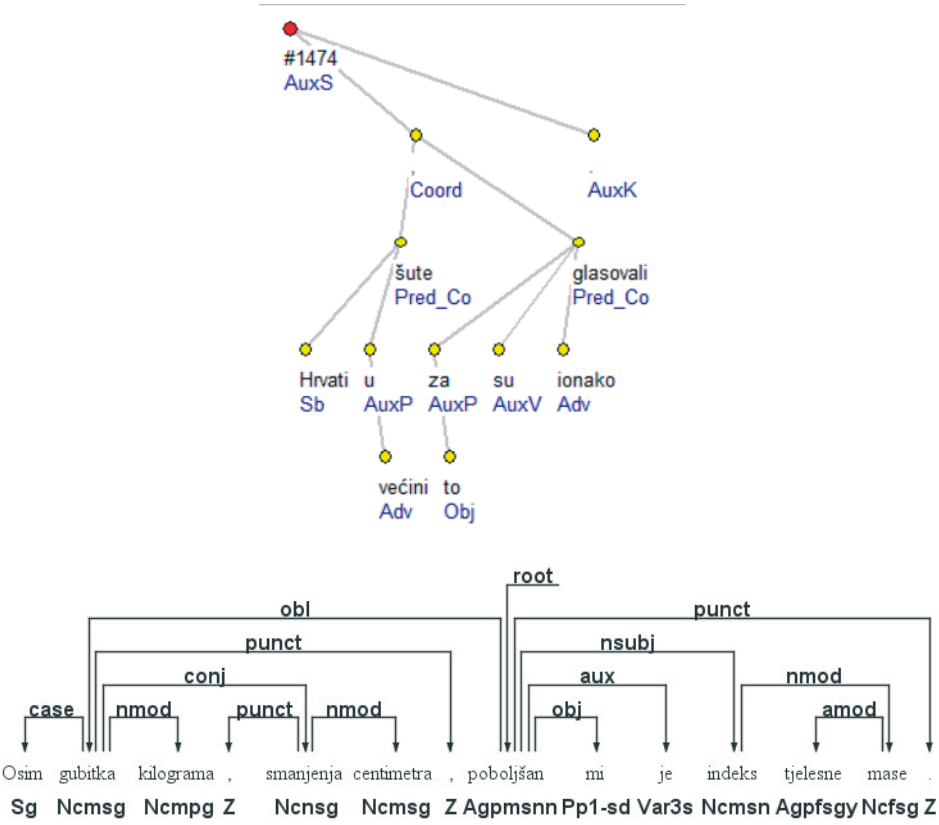
[18] Više o uniformnome označavanju specifičnih sintaktičkih struktura u slavenskim jezicima v. u (Zeman 2015).



Slika 11. Razlika u obilježavanju višestrukih nezavisnosloženih rečenica

Osim toga, u PDT-u će pri obilježavanju veznik biti vršni čvor stabla (točnije, najviši čvor odmah iza čvora AuxS kojim se uvodi rečenica (v. sliku 11), što nije u skladu s temeljnim teorijskim postavkama ovisnosne gramatike, prema kojima je vršni čvor glagol. Pristup koordinaciji u UD-u u skladu je s tim pravilom te glagol ili imenski dio imenskoga predikata ostaje vršni čvor rečenice.

Na kraju, u slučajevima asindetske koordinacije zarez će postati vršni čvor te uvoditi rečenicu, a osim toga, njegova će funkcija u tom slučaju biti *Coord*, a ne *AuxX*, dok će u UD-u zarez imati isto mjesto u strukturi i istu funkciju neovisno o tome radi li se o asindetskoj ili sindetskoj koordinaciji (v. sliku 12).



Slika 12. Razlika u obilježavanju asindetske koordinacije

4. Zaključak

U ovome smo radu pokazali na koji se način pristupa koordinaciji kao specifičnoj sintaktičkoj strukturi u ovisnosnim bankama stabala. Naime, koordinacija predstavlja izazov u obilježavanju u sklopu ovisnosnih pristupa zbog svoje neovisnosne prirode, odnosno zbog simetričnosti odnosa između koordiniranih članova, za razliku od asimetričnih odnosa između članova ovisnosnih odnosa. Kako bi doskočili tomu problemu, autori Praške ovisnosne banke stabala (*Prague Dependency Treebank*, PDT) i projekta *Universal Dependencies* osmislili su različita pravila za obilježavanje koordinacije te se ta razlika očituje i u dvjema hrvatskim ovisnosnim bankama stabala: Hrvatskoj ovisnosnoj banci stabala (HOBS-u) i Croatian UD-u.

Na temelju detaljnoga prikaza i jednoga i drugoga pristupa označavanju koordinacije pokazali smo kako je – za razliku od pristupa u PDT-u, koji je pokušao sačuvati njezinu simetričnu prirodu – pristup primijenjen u UD-u, premda svodi

koordinaciju na ovisnosni odnos, pogodniji za ujednačen opis različitih vrsta koordinacije, što je posebno važno za označavanje koordinacije u ovisnosnim bankama stabala za brojne tipološki i genetski različite jezike. Isto tako, posredno smo pokazali kako pri usporedbi podataka u različitim ovisnosnim bankama stabala za iste jezike istraživači moraju biti svjesni razlika u označavanju kako bi mogli pravilno upotrebljavati podatke.

Na kraju, iako je koordinacija detaljno opisana u obama pristupima, neke osobitosti i dalje u ovome trenutku ostaje nemoguće prikazati u strukturnome stablu, posebno kada govorimo o ugniježđenoj koordinaciji (iako smo ponudili rješenje za taj problem u obogaćenome prikazu UD-a), čime se pokazuje slojevitost i kompleksnost koordinacije kao specifičnoga sintaktičkog fenomena.

Literatura

Agić, Željko. 2012. *Pristupi ovisnosnom parsanju tekstova*. Neobjavljena doktorska disertacija. Zagreb: Filozofski fakultet Sveučilišta u Zagrebu.

Agić, Željko; Berović, Daša; Merkler, Danijela; Tadić Marko. 2014. Croatian dependency treebank 2.0: New annotation guidelines for improved parsing. U Calzolari, Nicoletta i dr. (ur.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 2313–2319. Reykjavik: ELRA.

Agić, Željko; Ljubešić, Nikola. 2014. The SETimes.HR Linguistically Annotated Corpus of Croatian. U Calzolari, Nicoletta i dr. (ur.) *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 1724–1727. Reykjavik: ELRA.

Agić, Željko; Ljubešić, Nikola. 2015. Universal Dependencies for Croatian (that work for Serbian, too). U Piskorski, Jakub i dr. (ur.), *Proceedings of the 5th Workshop on Balto-Slavic Natural Language Processing*, 1–8. Hissar; INCOMA.

Farkaš, Daša; Filko, Matea, Tadić, Marko. 2016. HR4EU – Using Language Resources in Computer Aided Language Learning. U Institute for Bulgarian Language, Bulgarian Academy of Sciences (ur.), *Proceedings of the Second International Conference Computational Linguistics in Bulgaria,* 38–44. Sofija: Institute for Bulgarian Language, Bulgarian Academy of Sciences.

Hajič, Jan; Panevová, Jarmila; Buráňová, Eva; Urešová, Zdeňka; Bémová, Alla. 1999. *Prague Dependency Treebank 2.0. Annotations at analytical level. Instructions for annotators.* Prag: UK MFF ÚFAL. Dostupno na: https://ufal.mff.cuni.cz/pdt2.0/doc/manuals/en/a-layer/html/. Datum posjete stranici: 23. 3. 2022.

Hajič, Jan i dr. 2000. The Prague Dependency Treebank: A Three-Level Annotation Scenario. U Abeillé, Anne (ur.), *Treebanks: Building and Using Parsed Corpora*, 103–127. Amsterdam: Kluwer.

Hudeček, Lana; Mihaljević, Milica. 2017. *Hrvatska školska gramatika*. Zagreb: Institut za hrvatski jezik i jezikoslovlje. Dostupno na: gramatika.hr. Datum posjete stranici: 23. 3. 2022.

Katunar, Daniela. 2014. Prepositional antonymy in Croatian: a corpus approach. *Suvremena lingvistika* 40(78). 151–169.

de Marneffe, Marie-Catherine; Manning, Christopher D.; Nivre, Joakim; Zeman, Daniel. 2021. Universal Dependencies. *Computational Linguistics* 47(2). 255–308. DOI: 10.1162/coli_a_00402

Merkler, Danijela; Agić, Željko; Agić, Ana (2013). Babel Treebank of Public Messages in Croatian. *Procedia – Social and Behavioral Sciences* 95. 490–497.

Nivre, Joakim i dr. 2016. *Guidelines for Universal Dependencies v2.* Dostupno na: https://universaldependencies.org/guidelines.html. Datum posjete stranici: 23. 3. 2022.

Petrović, Ante. 2021. Sa ili bez istog padeža: katalipsa prijedložne dopune u koordiniranim konstrukcijama (u tisku).

Przepiórowski, Adam, Patejuk, Agnieszka. 2019. Nested coordination in Universal Dependencies. U: *Proceedings of the Third Workshop on Universal Dependencies (UDW, SyntaxFest 2019)*, 58–69. Pariz: Association for Computational Linguistics. DOI: 10.18653/v1/W19-8007

Šojat, Krešimir. 2008. *Sintaktički i semantički opis glagolskih valencija u hrvatskom*. Neobjavljena doktorska disertacija. Zagreb: Filozofski fakultet Sveučilišta u Zagrebu.

Tadić, Marko. 2007. Building the Croatian dependency treebank: The initial stages. *Suvremena lingvistika* 63.85–92.

Tadić, Marko. 2009. New version of the Croatian National Corpus. U Hlaváčková, Dana; Horák, Aleš; Osolsobě, Klara; Rychlý, Pavel (ur.), *After Half a Century of Slavonic Natural Language Processing*, 199–205. Brno: Masaryk University.

Tesnière, Lucien. 1959. *Élements de syntaxe structurale*. Pariz: Librarie C. Klincksieck.

Zeman, Daniel. 2015. Slavic Languages in Universal Dependencies. U Gajdošová, Katarina; Žáková, Adriána (ur.), Natural Language Processing, Corpus Linguistics, E-learning, 151–163. Lüdenscheid: RAM-Verlag.

JEZIKOSLOVLJE
23.2 (2022): 193-214

ANNOTATING COORDINATION IN DEPENDENCY TREEBANKS

In this paper, we present how coordination (both coordination of clauses and phrases) is annotated in dependency treebanks. Dependency treebanks are built in accordance with the dependency approaches to syntax. Special emphasis will be given to coordination annotation within the Universal Dependencies project (UD) (https://universaldependencies.org/). The UD project aims for consistent annotation of grammatical structures across world languages and has collected almost 200 treebanks in more than 100 languages so far, including the one for Croatian – the Croatian UD. Before the Croatian UD treebank was built, the first Croatian Dependency Treebank was built based on the modified Prague Dependency Treebank specification for annotation at the analytical level. The approach used in these two treebanks differs when it comes to the annotation of particular syntactic structures. We show the main differences in annotating coordination in the two Croatian dependency treebanks and focus on problematic cases of syntagmatic and clausal coordination.

Keywords: dependency treebanks, coordination, Universal Dependencies, PDT, HOBS

Adrese autorica:

Daša Farkaš
Filozofski fakultet Zagreb
HR – 10 000 Zagreb, Ivana Lučića 3
dfarkas@ffzg.hr

Matea Filko
Filozofski fakultet Zagreb
HR – 10 000 Zagreb, Ivana Lučića 3
matea.filko@ffzg.hr