

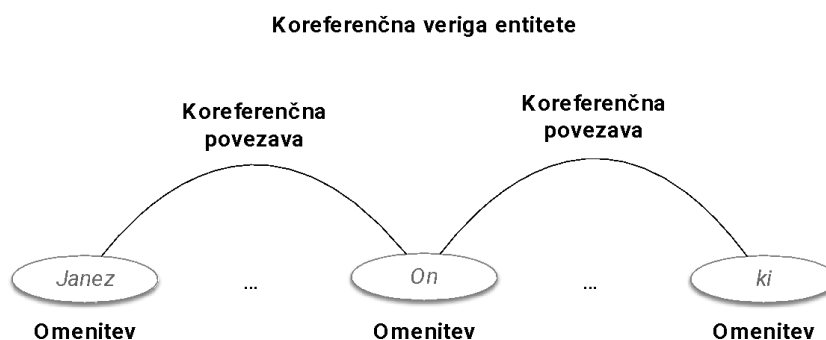
# Smernice za označevanje koreferenčnosti v slovenskem jeziku

Slavko Žitnik, Špela Arhar Holdt, Nejc Robida, Neli Blagus

| Kdaj         | Verzija | Kdo            | Opis  |
|--------------|---------|----------------|---|
| 29.12.2020   | 1.0     | Slavko Žitnik  | Vsebina za odkrivanje koreferenčnosti (prenos srbskih smernic). |
| 8.1.2021     | 1.1     | Slavko Žitnik  | Dopolnjena vsebina in dodane utemeljitve.                       |
| 28.1.2021    | 1.2     | Slavko Žitnik  | Malenkostni popravki  |
| 23. 9. 2022  | 1.3     | Špela A. Holdt | Premik poglavij za boljšo rdečo nit smernic                     |
| 24. 9. 2022  | 1.4     | Špela A. Holdt | Členitev vsebine in vsebinska nadgradnja                        |
| 15. 10. 2022 | 1.5     | Nejc Robida    | Slogovni popravki in razdvoumljanje                             |
| 22. 2. 2023  | 1.6     | Špela A. Holdt | Čistopis za projekt RSDO  |

Identifikacija koreferenčnosti je eden izmed težjih problemov pri ekstrakciji informacij. Za izdelavo smernic za označevanje neposredno sledimo srbskim smernicam za označevanje koreferenčnosti (RELDI, 2018), pripravljenim v okviru projekta RELDI, te se med drugim zgledujejo tudi po smernicah ACE (LDC, 2008). Za razliko od srbskih imajo slovenske smernice preurejeno in natančneje členjeno vsebino, nove primere označevanja, prav tako pa ne označujemo skladijskih lastnosti (poglavje Vrste omenitev), saj lahko slednje pridobimo iz sistema skladijskega označevanja slovenščine.

Shema označevanja koreferenčnosti, ki jo definirajo te smernice, je prikazana na sliki spodaj:



**Type**  
Copular-N, Copular-S, Appositive, Zero anaphora, Clause, Regular

**Referent** (\* only for Copular-S, otherwise NA)  
Subject/Object/NA

**Discontinuous**  
True/False

**Ellipsis**  
True/False

Slika 1: Prikaz sheme označevanja koreferenčnosti. Prikazane so tri omenitve, ki se sklicujejo na določeno entiteto. Povezave med zaporednimi omenitvami, ki se sklicujejo na isto entiteto, so koreferenčne povezave in skupaj z vsemi omenitvami tvorijo koreferenčno verigo. Vsaka omenitev ima določene še lastne oznake.

## 1. Uvod in splošna navodila

O koreferenčnosti govorimo, ko se več delov besedila (besed, besednih zvez ali stavkov) nanaša na istega referenta v zunajjezikovni stvarnosti. Z referentom označujemo različne vrste predmetnosti, npr. osebe, živali, rastline, predmete, kot tudi lokacije, dogodke, funkcije, koncepte.

Dele besedila, ki se nanašajo na istega referenta, imenujemo *omenitve* (angl. *mentions*). Omenitve se lahko pojavljajo preko različnih stavkov, povedi, odstavkov ali drugih delov besedila. Pri označevanju jih povežemo v t. i. koreferenčno verigo (angl. *coreference chain*). V teh navodilih koreferenčne verige nakazujemo z različnimi barvami, kot na primer v naslednjih povedih.

[1.a] Peter ima dva psa.

[1.b] On se velikokrat igra z njima.

Omenitvi *Peter* in *On* se nanašata na istega referenta, tj. isto konkretno osebo, kar pomeni, da sta koreferenčni. Enako velja za omenitvi *njima* in *dva psa*, ki se nanašata na isti par živali.

Naloga pri označevanju koreferenčnosti je identifikacija in povezovanje omenitev. Označuje se samo omenitve, ki so koreferenčne z drugo omenitvijo. Delov besedila, ki niso koreferenčni z nobenim drugim delom besedila, torej ne označujemo.

Označevanje koreferenčnosti se izvaja na nivoju dokumenta. To pomeni, da je potrebno označiti vse koreference, ki se nahajajo preko različnih stavkov ali odstavkov. Zato je pomembno, da se pred označevanjem prebere celoten dokument, vključno z naslovom ali drugimi pripadajočimi deli, v katerih se lahko pojavi koreferenčnost.

Delo na nivoju dokumenta obenem pomeni, da kakršnekoli semantične podobnosti med različnimi dokumenti niso pomembne. Če je referent A omenjen v dokumentu X in se nato pojavi še v dokumentu Y, torej ni potrebno označiti koreferenčnosti med njima. Posebej se označijo koreferenčne omenitve v dokumentu X in posebej v dokumentu Y.

Pri označevanju omenitev se načeloma označi celotno zvezo, ki se nanaša na referenta. V primeru [1.c] je pravilno označiti zvezo *albanski minister za notranje zadeve Bujar Nišani* kot eno omenitev (in ne le imena ministra). Po drugi strani zveza *direktor ruske zvezne službe za kontrolo narkotikov Viktor Ivanov* koreferira z lastnim imenom *Ivanovim* in ne lastnim imenom skupaj s predlogom – z *Ivanovim*. O tem, katere dele vključevati in katerih ne, natančneje pišemo v nadaljevanju smernic.

[1.c] Sporazum so podpisali albanski minister za notranje zadeve Bujar Nišani in direktor ruske zvezne službe za kontrolo narkotikov Viktor Ivanov. Nišani je z Ivanovim sklenil, da ...

[1.č] Marija je videla človeka, ki je ukradel direktorjevo vozilo. Niso ga še našli.

[1.d] Uradniki z ministrstva so sklicali tiskovno konferenco. Oni so rekli ...

[1.e] Jutranji sestanki so najproduktivnejši. Na takšnih sestankih pa je pogosto kdo odsoten.

## 2. Identifikacija relevantnih povezav

Koreferenčnost je zapleten jezikovni pojav in v teoriji obstaja več tipov koreferenčnih odvisnosti. Zaradi konsistentnosti bomo označevali le tip povezave, ki ga imenujemo identiteta (ang. *Identity*).

### 2.1. Identiteta

Identiteta predstavlja neusmerjeno in simetrično koreferenčno povezavo med omenitvami, ki se nanašajo na istega referenta. Takšne povezave so tranzitivne: če je A identično z B in B identično s C, potem sta tudi A in C identična. Vse omenitve določenega referenta v dokumentu predstavljajo koreferenčno verigo.

**[2.1.a]** Donald Trump je zmagal na preteklih predsedniških volitvah v Ameriki. On je na njih osvojil večino elektorskih glasov, Hillary Clinton pa je dobila večino glasov volivcev. Novi predsednik je prisegel v januarju 2017.

Omenitve *Donald Trump*, *novi predsednik* in *on* se nanašajo na istega referenta. Analogno se zveza *preteklih predsedniških volitvah v Ameriki* in zaimsek *njih* nanašata na drugega referenta.

Označevalec/-ka mora označiti koreferenčnost, tudi če se z mnenjem avtorja/-ice besedila ne strinja:

**[2.1.b]** Ne glede na pridobljene glasove, menim, da si ona ni zaslužila, da postane kanclerka.

Koreferenčne povezave pa ne označujemo, ko ni popolnoma jasna, se nanjo le namiguje ali o njej dvomi, je zanikana, skratka v primerih, kadar je sobesedilo ne potrjuje zares. Na primer:

**[2.1.c]** Možno je, da je organizirani kriminal najpomembnejše vprašanje.

**[2.1.č]** Skoraj so ga imenovali za direktorja korporacije.

**[2.1.d]** Če bi mu dali proste roke, bi bil on najuspešnejši minister.

**[2.1.e]** To trčenje ni bila nesreča.

### 2.2. Skorajšnja identiteta (angl. *near-identity*)

Pojavljajo se tudi primeri, med katerimi je določena pomenska bližina, vendar referent ni zares enak:

**[2.2.a]** V Beogradu so cene poskočile za 2,4 %, medtem ko so v Zagrebu ostale stabilne.

Čeprav je izraz *cene* enak, tu ne označujemo koreferenčnosti, ker se besedi v prvem in drugem delu povedi nanašata na po lokaciji različna referenta (in sicer na cene v Beogradu in cene v Zagrebu). Podobno se lahko zgodi, ko se različni primeri istega izraza časovno razlikujejo:

**[2.2.b]** Zneski davka na nepremičnine so se letos v primerjavi z lanskimi dohodki povečali za 20 %.

Na drugi strani pa označujemo omenitve, pri katerih je referent predstavljen z dveh različnih vidikov:

**[2.2.c]** »Po njegovi smrti se je vse spremenilo. Pogrešam igralca, še bolj pa očeta.«

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

## 2.3. Ostali tipi koreferenčnosti

V literaturi je opisanih več tipov koreferenčnosti, katerih obravnava je odvisna od posameznega označevalnega projekta. V nadaljevanju prikažemo, kako označujemo nekaj glavnih tipov.

### 2.3.1. Sopomenke

V okviru sopomenk se velikokrat pojavlja tip identitete, ki ga je potrebno označiti, kot na primer:

**[2.3.1.a]** V antikvariatu je kupil velik predalnik. Starodavno omaro smo komaj preselili v stanovanje.

**[2.3.1.b]** Tat je pred stavbo oropal starko. Na srečo je babica uspela prepoznati napadalca.

**[2.3.1.c]** Za Pariz smo napovedali nove turistične aranžmaje – obiščite Mesto luči po posebni ceni!

### 2.3.2. Nadpomenke/podpomenke

Pri odnosih nadpomenka – podpomenka označujemo koreferenčnost, ko nadpomenka ali podpomenka jasno označujeta istega referenta. Na primer:

**[2.3.2.a]** V našem živalskem vrtu so levi najbolj popularni. Te živali so fascinirale tudi mene.

Možni so tudi primeri, ko nadpomenka označuje skupino, podpomenka pa njeno podskupino. V takšnih primerih koreferenčnosti ne označujemo:

**[2.3.2.b]** Čeprav redko jem sadje, so jabolka izjema.

**[2.3.2.c]** Matematika je lahko težavna, zlasti trigonometrija.

**[2.3.2.č]** Sladkarijam se je potrebno izogibati, vendar sem vseeno pojedel ploščico čokolade.

### 2.3.3. Metonimija

Metonimije tipa *vino - steklenica* ali *avtor - delo* (spodaj) kljub pomenskemu prenosu označujemo kot koreferenčne omenitve.

**[2.3.3.a]** Iz hladilnika sem vzel steklenici vina, a ju je spil.

**[2.3.3.b]** Začel sem brati Šantića. On je napisal veliko število pesmi.

Metonimija se pogosto uporablja tudi v političnih novicah, kjer se imena glavnega mesta, države ali sedeža državne uprave uporabljajo za vlado te države, na primer:

**[2.3.3.c]** Washington je pozdravil sestanek: »ZDA ostajajo zavezane k stabilnosti Balkana. To srečanje lahko le prispeva k sodelovanju in ima popolno podporo Bele hiše.«

Včasih se lahko metonimije na prvi pogled zdijo koreferenčne, vendar niso. Čeprav je *Politika* omenjena v obeh povedih [2.3.3.č], je iz sobesedila jasno, da se prva omenitev nanaša na specifično številko/izdajo časopisa *Politika*, medtem ko se druga nanaša na uredništvo časopisa, zato koreferenčnosti tu ni. Podoben primer je [2.3.3.d], v katerem se *Kitajska* najprej nanaša na zemljepisno enoto, nato pa na vlado države.

**[2.3.3.č]** Zjutraj sem kupil Politiko in Novosti. Ljiljana Smajlovič je odstopila z mesta odgovorne urednice Politike.

**[2.3.3.d]** Teroristi so prevzeli nadzor nad letalom, ki je letelo na Kitajsko. Kitajska je poslala svoje lovce za prestrežanje.

#### 2.3.4. Meronimija

Meronimov načeloma ne označujemo.

**[2.3.4.a]** Med najpomembnejšimi dogovori, doseženimi na dvodnevem vrhu, sta odločitev o začetku programa usposabljanja iraških varnostnih sil in **obljuba, da se bo povečala velikost vojaških sil v Afganistanu.**

V primeru 2.3.4.a ne označimo segmentov *sta odločitev o začetku programa usposabljanja iraških varnostnih sil in obljuba, da se bo povečala velikost vojaških sil v Afganistanu* in *najpomembnejšimi dogovori, doseženimi na dvodnevem vrhu*, saj lahko razberemo, da *odločitev* in *obljuba* predstavljata elemente širšega koncepta – dogovorov.

#### 2.4. Dodatne opombe

V tem razdelku opisujemo postopek označevanja koreferenčnosti v specifičnih situacijah.

##### 2.4.1. Uporaba splošnega znanja

Včasih podrobnosti situacije v besedilu niso eksplicitno navedene, lahko pa jih dopolnimo na osnovi splošnega znanja. Na primer, naslednji primer nikjer ne navaja, da je Donald Trump predsednik ZDA, vendar se to lahko dovolj zanesljivo sklepa:

**[2.4.1.a]** **Donald Trump** se je danes srečal z japonskim premierjem Shinzom Abeom. **Predsednik ZDA** je rekel, da ...

##### 2.4.2. Časovni in prostorski izrazi

Tudi časovni in prostorski izrazi lahko nastopajo kot koreference, na primer:

**[2.4.2.a]** John je **tri leta** preživel v zaporu. **Tekom tega časa** je ....

**[2.4.2.b]** **Zagreb** je priznal neodvisnost **Kosova leta 2008**. **Istega leta** je **vzpostavil** diplomatske odnose s Prištino.

**[2.4.2.c]** Lansko leto smo **prepotovali Indijo**. **Država** nas je navdušila.

##### 2.4.3. Splošne omenitve

Da se zmanjša zahtevnost označevanja koreferenčnosti, besed, ki predstavljajo splošne pojme, kot so *svet, vesolje, umetnost, glasba* in podobno, ne označujemo kot omenitve.

##### 2.4.4. Pojmi

Pri samostalniških besednih zvezah, ki predstavljajo pojme, kot so *svoboda medijev* ali *proizvodnja orožja*, običajno ni potrebno označevati koreferenčnosti.

### 2.4.5. Koreferenčnost identičnih omenitev

Če se ena beseda pojavlja na več mestih v besedilu, še ne pomeni nujno, da gre za koreferenčnost, na primer:

**[2.4.5.a]** Nekateri delajo ves dan, nekateri pa uživajo.

**[2.4.5.b]** Obstajajo matere in matere.

### 2.4.6. Razlike med anaforami in koreferencami

Anafora je sklicevanje na nek že omenjen del besedila. Čeprav je anaforičnost zelo podobna koreferenčnosti, ne pomeni nujno koreferenčnosti. V tem projektu nas zanimajo le koreferenčne omenitve, na katere se sklicujemo, zato naslednjih primerov ne označujemo:

**[2.4.6.a]** V napadu na Madrid je poškodbe utrpelo 200 ljudi več kot v Londonu.

**[2.4.6.b]** Slika iz Louvra mi je bolj všeč od te iz Rima.

**[2.4.6.c]** Vsak človek, ki je prepričan vase, je pripravljen zagovarjati svoja stališča.

**[2.4.6.č]** Vsak s časom razume, kakšne so mere njegovih sposobnosti.

**[2.4.6.d]** Večina ljudi ima svoje otroke najrajši na svetu.

**[2.4.6.e]** Mnogi sanjajo, da bi se videli na vrhu.

**[2.4.6.f]** Nihče ni rekel, da je lačen.

**[2.4.6.g]** Nobena oseba ne bi smela doživeti, da se ji reče kaj takšnega.

**[2.4.6.h]** Nekateri ljudje znajo poskrbeti za svoje stvari.

**[2.4.6.i]** Pesniki radi recitirajo svoje pesmi.

**[2.4.6.j]** Vsak profesor rad opravlja svoj posel.

V primeru [2.4.6.j] denimo obstaja anaforičnost med svojilnim zaimkom *svoj* in besedno zvezo *vsak profesor*. Koreferenčnost pa ne obstaja, ker se zaimek ne nanaša na skupino vseh profesorjev, ampak posamično na katero koli entiteto, ki spada v skupino. Za koreferenčnost bi moralo biti besedilo na primer tako, kot je v [2.4.6.č] - ali v zadnjih dveh primerih, kjer koreferenčnost označujemo, saj se zaimek nanaša na celotno skupino:

**[2.4.6.k]** Profesor Marko rad opravlja svoj posel, prav tako profesor Nenad svojega.

**[2.4.6.l]** Nekateri psi so bili tako ljubki, da sem jih božala.

**[2.4.6.m]** Protestirajoči so podali svoje zahteve.

## 3. Vrste besed, ki nastopajo v koreferencah

### 3.1. Lastnoimenski samostalniki in izpeljani pridevniki

Lastna imena vedno predstavljajo omenitev referenta, na katerega se nanašajo. Takšne omenitve je najlažje zaznati, poleg njih pa je potrebno označiti tudi ustrezajoče izlastnoimenske pridevnike:

**[3.1.a]** Eden od tradicionalnih beograjskih dogodkov je tudi maraton. Ta se letošnje leto organizira že 27. leto. Beograd bo ...

**[3.1.b]** Filip je sam v stanovanju. Filipovi starši so odšli v Grčijo.

[3.1.c] **Hrvaška** planira povečanje blagovne izmenjave s **Slovenijo**. **Slovenski** predsednik je povabil **hrvaške** gospodarstvenike na posvet.<sup>1</sup>

[3.1.č] **Pink** je najavil odprtje dveh novih kanalov. **Pinkova** programska shema za osnovne kanale ostaja nespremenjena.

[3.1.d] **Novak Đoković** je eden najboljših tenisačev na svetu. Čeprav so mi teniške tekme zanimive, so **Novakove** nekaj posebnega.

[3.1.e] **Hram Svete Save** se nahaja na vračarskem platoju. Čeprav ima Beograd polno cerkva in hramov, je **svetosavski** daleč največji.

Ponekod se pojavljajo primeri, ko se izlastnoimenski pridevniki (tipično s samostalniškim jedrom), nanašajo na drugega referenta:

[3.1.f] **Sibirski tiger** je na robu izumrtja. Krivolovci so zmanjšali število **tigrov** na manj kot 300.

[3.1.g] **Jagodinski muzej voščениh lutk** je dobil **novost** – **figuro Vladimirja Putina**. Število figur na **tej turistični atrakciji** se je povečalo na 30.

### 3.2. Ostali samostalniki in pridevniki

Tudi občnoimenski samostalniki in zveze lahko nastopajo kot omenitve:

[3.2.a] **Obtožen** je pranja denarja in **prometa z mamili**. Za **trgovino z drogo** bi moral dobiti kazen do 10 let zapora.

[3.2.b] **Prodaja prenosnih in tabličnih računalnikov** se je v letošnjem letu povečala za 20 %. Za razliko od **prometa prenosnih računalnikov** tržišče stacionarnih računalnikov stagnira.

Pridevniki, ki ne izhajajo iz lastnih imen, ne predstavljajo omenitev sami po sebi oz. je njihova možnost koreferiranja odvisna od konteksta. Ko se v besedilu govori o specifičnem referentu, lahko takšna beseda predstavlja omenitev:

[3.2.c] **Madera** že dolgo slovi kot **najbolj znana beograjska restavracija**. **Njena** vrata dnevno prestopajo mnoge znane osebnosti. **Restavracijski** vrt je še posebej privlačen poleti.

[3.2.č] **Sovražne** izjave na **Balkanu** – **regionalne** razmere ponovno v centru pozornosti.

Zaradi zmanjšanja kompleksnosti označevanja smo se odločili, da lahko takšni pridevniki tvorijo koreferenčno verigo le v primeru, ko je v verigi vsaj en samostalnik. Z drugimi besedami: koreferenčnih verig ne bi smeli oblikovati le iz opisnih pridevnikov.

#### 3.2.1. Omenjanje področja, države ali ljudstev

Imena držav/področij imajo včasih enako obliko kot vrstni pridevniki (običajno na *-ski/-ški*). Takšne omenitve povezujemo v isto koreferenčno verigo. Na primer:

[3.2.1.a] Na **Srbskem** in tudi v tujih muzejih ter galerijah so ta mesec promovirali **srbsko** srednjeveško umetnost.

---

<sup>1</sup> Če v besedilu obstaja lastnoimenski samostalnik, verižimo vse primere izpeljanega pridevnika.

### 3.2.2. Kratice

Kratice in okrajšave je potrebno upoštevati tako kot vse ostale omenitve in označiti njihovo koreferenčnost.

[3.2.2.a] Nova komisija **Mednarodnega denarnega sklada** bi morala obiskati Srbijo v naslednjem tednu. Iz sedeža **MDS** so sporočili ...

### 3.3. Zaimki

Zaimki predstavljajo najpogostejše besede, ki sodelujejo v koreferenčnostih.

#### 3.3.1. Osebni in svojilni zaimki

Osebne in svojilne zaimke je potrebno obravnavati kot omenitve, če so koreferenčne kakoli drugi omenitvi v besedilu. Primeri:

[3.3.1.a] Lansko leto sem zadnjič videl **Tanjo**. **Ona** sedaj študira v Ameriki.

[3.3.1.b] S **Petrom** grem v kino. **Njemu** bi moral biti ta film všeč.

[3.3.1.c] **Vaše** finančno stanje je stvar **vaših** odločitev, **moja** vloga je samo, da **vam** svetujem.

Enako označujemo tudi navezne oblike osebnih zaimkov.

[3.3.1.d] **Katja** je imela včeraj rojstni dan. **Zanjo** sem kupil darilo.

#### 3.3.2. Splošna uporaba ti/vi

Včasih se zaimek druge osebe ednine/množine uporablja v splošnem, tako da se dejansko ne nanaša na osebo/osebe, ki se jih nagovarja. V takšnih primerih splošne uporabe zaimkov ali generične ničelne anafore ne označujemo, na primer:

[3.3.2.a] Frist se spominja svojega dela: »Pogosto ste imeli bolnika v kritičnem stanju in ste želeli začeti operacijo presaditve, vendar ste vedeli, da ne morete, dokler ni zagotovljeno, da bo organ prišel v operacijsko dvorano.«

[3.3.2.b] »V Beogradu imate mnogo linij javnega prevoza, vendar je gneča takšna, da pogosto čakate, potujete.«

#### 3.3.3. Povratno svojilni zaimki

Povratno svojilne zaimke je treba obravnavati enako kot osebne zaimke. Enako velja za svojilni zaimek *svoj*:

[3.3.3.a] **Ona** spoštuje le **sebe**.

[3.3.3.b] **Zbrani študenti** so se oglasili s **svojim** predlogom.

Prav tako je potrebno posebno pozornost nameniti označevanju zaimkov *sebe* in *svoj* zaradi razlik med anaforo in koreferenčnostjo – primeri in dodatni opisi so na voljo v razdelku Razlike med anaforami in koreferencami.



Nasprotno pa zaimek se ne sme biti označen kot omenitev, tudi če gre za del glagola in kadar se zdi, da izraža koreferenčnost:

[3.3.3.c] Marija se je pobožala po roki.

[3.3.3.č] Mi se imamo radi.

[3.3.3.d] Pokesal se je.

#### 3.3.4. Kazalni zaimki

Kazalni zaimki so tipičen primer besed, ki predstavljajo omenitve in jih je potrebno označiti. Kazalni zaimki so lahko koreferenčni tudi s celotnimi stavki.

[3.3.4.a] Beograd se nahaja na izlitju Save v Donavo. V tem mestu ...

[3.3.4.b] Spet sem jo videla na televiziji. Ta oseba je zelo ...

[3.3.4.c] Spomenik je bil postavljen v čast padlim borcem. Oni so v boju za domovino dali svoja življenja.

[3.3.4.č] Peter je planiral oditi. To je včeraj končno storil.

[3.3.4.d] Novega blaga nismo prejeli že cel mesec. Ta situacija ne koristi nikomur, to je sigurno.

[3.3.4.e] Novega blaga nismo prejeli že cel mesec. Ta situacija ne koristi nikomur, to je sigurno.

[3.3.4.f] Psi se igrajo na vrtu. Tam so tudi vse njihove igračke.

[3.3.4.g] Pred nekaj leti smo šli na Japonsko. Tedaj se je zgodil potres, ki je prizadel Fukušimo.

[3.3.4.h] Prišel je točno, ker je vozil hitro. Vendar tako je treba, kadar se dogajajo takšne stvari.

[3.3.4.i] Na štedilniku imate dva kosa mesa. Želite, da vam pripravim več? – Ni potrebno, tudi toliko mi je dovolj.

[3.3.4.j] Ves denar je zapravil za nov TV. Zaradi tega si ni mogel privoščiti novih zvočnikov.

#### 3.3.5. Oziralni zaimki

V nekaterih jezikih (npr. srbsščina) se oziralni zaimki lahko pišejo brez vejice in imajo v tem primeru skupaj z identifikacijskim delom omenitve dopolnilno vlogo ter se označujejo skupaj kot ena omenitev. Na podlagi tega lahko ločijo tudi omejevalno in neomejevalno vlogo. V slovenščini oziralni zaimek (če se avtor pri pisanju ne zmoti) načeloma tvori podredni stavek, zato zaimek v tem primeru predstavlja svojo omenitev, na primer:

[3.3.5.a] Film, ki sem ga gledal, je bil dober. Tako mi je bil všeč, da sem ga priporočil ostalim.

[3.3.5.a.1] Film, ki sem ga gledal, je bil dober. Tako mi je bil všeč, da sem ga priporočil ostalim.

[3.3.5.a.2] Film, ki sem ga gledal, je bil dober. Tako mi je bil všeč, da sem ga priporočil ostalim.

Primeri dopolnilne vloge, kjer stavek referenta le dodatno opiše:

[3.3.5.b] Hilary, ki se je povzpел na Mount Everest leta 1953, je bil največji raziskovalec gora vseh časov.

[3.3.5.c] John je videl Jane, s katero se je takoj objel.

[3.3.5.č] Hans živi v Berlinu, kjer ima polno prijateljev.

[3.3.5.d] Obtoženi so se pojavili pred sodiščem v Beogradu, čigar sodbe so bile v preteklosti kontroverzne.

[3.3.5.e] Gradnja mostu je trajala tri leta, kolikor je bilo kasneje potrebno tudi za dostopne poti.

Primeri dopolnilne vloge, kjer stavek opiše referenta, a se v njem pojavi še ena dodatna omenitev na referenta:

[3.3.5.f] Legendarni film Orsona Wellesa »Državljeni Kane,« ki je bil nekoč razglašen za najboljši film vseh časov, bo na ogled jutri zvečer v Jugoslovanski kinoteki.

[3.3.5.g] Peter Johnson, ki je predsednik družbe, je bil imenovan za menedžerja leta.

Dodatni primeri:

[3.3.5.h] Ona obožuje Pariz, kjer je bila že velikokrat.

[3.3.5.i] S poti se vrnem naslednji četrtek, kadar tudi ti.

[3.3.5.j] Izstrelitev rakete je bila brez problemov, kakor je bilo tudi planirano.

[3.3.5.k] Dobil je povečanje plače za 15 %, kolikor je tudi pričakoval.

### 3.4. Glagoli

Ker glagoli označujejo dejanja, sami po sebi niso predmet označevanja koreferenčnosti. Ko se glagol sklicuje na dejanje, izraženo s pomočja glagola ali glagolnika, je potrebno to omenitev označiti, na primer:

[3.4.a] Vsak ima pravico, da protestira proti trenutni vladi, vendar morajo biti ti protesti organizirani na odgovoren način.

[3.4.b] Velikokrat se mi dogaja, da veliko jem, čeprav vem, da to prenajedanje ni zdravo.

[3.4.b.1] Velikokrat se mi dogaja, da veliko jem, čeprav vem, da to prenajedanje ni zdravo.

[3.4.b.2] Velikokrat se mi dogaja, da veliko jem, čeprav vem, da to prenajedanje ni zdravo.

V primeru, da se glagol in glagolnik nanaša na splošne koncepte, potem koreferenčnosti ne označujemo, saj ne obstaja specifičen primerik referenta, na katerega bi se nanašale takšne omenitve:

[3.4.c] Jaz rad berem, ker me branje umirja.

[3.4.č] Čeprav mu pisanje ne gre od rok, je Andrej uspel napisati pismo.

### 3.5. Številске vrednosti

Tudi številске vrednosti lahko nastopajo v koreferenčnostih, na primer:

[3.5.a] Ponudba za 150.000 \$ je zavrnjena. Upravni odbor se tej ceni le nasmiha.

### 3.6. Posebni primeri

V tem razdelku opisujemo postopek označevanja koreferenčnosti v specifičnih situacijah.

#### 3.6.1. Citati

V primerih, ko se v besedilu pojavi citirana izjava, je potrebno omenitve označiti tako kot v drugih delih besedila. Pri tem je potrebno pozorno označiti koreferenčnosti, saj se lahko med citatom in ostalim delom besedila spremenijo oblikoslovne značilnosti omenitev:

**[3.6.1.a]** »Ni nobenega razloga za **moj** odhod,« je **povedal**.

**[3.6.1.b]** »Nekateri so rekli, da je nespoštljivo do Partenona, da nima dorskih stebrov, vendar **jaz** nisem želel imitirati Partenona,« je odgovoril **Čumi**.

Pogosto se prva oseba množine pretvori v tretjo osebo množine izven citata, navadno kadar predstavnik neke skupine (države, podjetja, ...) poda izjavo:

**[3.6.1.c]** »**Mi** želimo, da se postopek nadaljuje pod okriljem OZN, tako da se najde vzajemno sprejemljiva rešitev,« je rekel tiskovni predstavnik **grške vlade**.

**[3.6.1.č]** Tiskovni predstavnik **Air Srbije je** izjavil: »**Mi** mislimo, da ...«

#### 3.6.2. Določne in nedoločne omenitve

Za razliko od angleščine, kjer se uporabljata določni in nedoločni člen, je v slovenščini prepoznavanje določnosti možno le v redkih slovničnih strukturah. V prvem primeru spodaj ne vemo, ali gre za isti zeleni avto ali ne. Da bi lahko povezali koreference, je treba poiskati informacijo o tem v širšem sobesedilu. V drugem primeru pa že določna oblika pridevnika nakazuje, da gre za istega referenta.

**[3.6.2.a]** Pri nas imamo **zelen** in rdeč avto. Zjutraj sem videl **zelen** avto.

**[3.6.2.b]** Pri nas imamo **zelen** in rdeč avto. Zjutraj sem videl **zeleni** avto.

#### 3.6.3. Izpust ali elipsa

Elipsa je izpustitev posameznih delov besedila zaradi ponavljanja. V nekaterih primerih se lahko izpuščen del nanaša nanaša na referenta (t.j. koreferenčnost). V primerih, ko to ne velja, se koreferenčnosti ne označuje:

**[3.6.3.a]** Markovo darilo mi je bilo bolj všeč od Marijinega.

**[3.6.3.b]** Tvoja disertacija je napisana bolje kot moja.

V navedenih primerih Markovo darilo ne pomeni iste entitete kot Marijino darilo. Analogno pri disertaciji. Zaradi tega v tem primeru ni koreferenčnosti.

V nasprotju s tema primeroma pa je potrebno koreferenčno razmerje označiti v primeru, ko se omenitve nanašajo na istega referenta. Eliptični omenitvi je potrebno nastaviti atribut Ellipsis na True. Pri vseh drugih omenitvah pa mora biti privzeta vrednost tega atributa nastavljena na False.

Eliptična omenitev nastane, ko je iz omenitve izpuščen glavni del besedne zveze, na primer:

**[3.6.3.c]** Tako hrastov kot **bukov les** se dolgo uporabljata za izdelavo pohištva, vendar je **bukov** v zadnjem času bolj popularen.

Možno je celo, da so v neki verigi vse omenitve eliptične, kot na primer spodaj:

**[3.6.3.č]** Pohvalili so **mi rdečo** in **modro majico**. **Modra mi** je prevelika, vendar **mi rdeča** popolnoma odgovarja.

Problem pri označevanju eliptičnih omenitev je ta, da zaradi izpustitev besed lahko pride do tega, da se ista beseda nanaša na dva različna referenta. V teh primerih to isto besedo ločeno označimo v okviru več koreferenčnih verig. V spodnjem primeru *njeno* označimo kot omenitev torte in ločeno kot omenitev osebe (tj. nje). Analogno velja za *mojo*.

**[3.6.3.d]** Če ona poskusi *mojo torto*, želim jaz *njeno*. Njena torta zagotovo ne vsebuje skrivnih sestavin, zaradi katerih je *moja* tako okusna.

**[3.6.3.d.1]** Če ona poskusi *mojo torto*, želim jaz **njeno**. **Njena torta** zagotovo ne vsebuje skrivnih sestavin, zaradi katerih je *moja* tako okusna.

**[3.6.3.d.2]** Če **ona** poskusi *mojo torto*, želim jaz **njeno**. **Njena** torta zagotovo ne vsebuje skrivnih sestavin, zaradi katerih je *moja* tako okusna.

**[3.6.3.e]** **Ona** je boljša kuharica, zato je *moja sarma* slabša od **njene**.

## 4. Gnezdenje koreferenčnosti

V nekaterih primerih je možno na določenih delih besedila označiti več vzporednih ali vgnezenih koreferenčnosti, na primer:

**[4.a]** Podjetje ACME ima mnogo zaposlenih, ki programirajo v Pythonu. Peter je njihov guru za Python, čeprav je začel uporabljati ta programski jezik šele, ko je prišel v ACME.

**[4.a.1]** **Podjetje ACME** ima mnogo zaposlenih, ki programirajo v Pythonu. Peter je **njihov** guru za Python, čeprav je začel uporabljati ta programski jezik šele, ko je prišel v **ACME**.

**[4.a.2]** Podjetje ACME ima mnogo zaposlenih, ki programirajo v Pythonu. **Peter** je **njihov guru za Python**, čeprav je **začel** uporabljati ta programski jezik šele, ko je **prišel** v ACME.

**[4.a.3]** Podjetje ACME ima mnogo zaposlenih, ki programirajo v **Pythonu**. Peter je njihov guru za **Python**, čeprav je začel uporabljati **ta programski jezik** šele, ko je prišel v ACME.

V zgornjem primeru lahko opazimo, da deli omenitev predstavljajo omenitve več različnih referentov.

Možni so tudi primeri, ko so omenitve v vseh primerih prekrivne z drugimi širšimi omenitvami. V takšnih primerih je potrebno označiti vse nivoje koreferenčnosti, na primer:

[4.b] Predsednik Srbije je položil venec na Spomenik neznanemu junaku na Avali. Predsednik države je izpostavil ...

[4.b.1] Predsednik Srbije je položil venec na Spomenik neznanemu junaku na Avali. Predsednik države je izpostavil ...

[4.b.2] Predsednik Srbije je položil venec na Spomenik neznanemu junaku na Avali. Predsednik države je izpostavil ...

Podobno je možno, ko se v besedilu pojavlja kopularna omenitev ali pa je omenitev izpuščena:

[4.c] Jovan se je javil na telefon. Povedali so mu, da je bil imenovan za vodjo oddelka.

[4.c.1] Jovan se je javil na telefon. Povedali so mu, da je bil imenovan za vodjo oddelka.

[4.c.2] Jovan se je javil na telefon. Povedali so mu, da je bil imenovan za vodjo oddelka.

#### 4.1. Lastna imena

Za razliko od preostalega besedila je potrebno na lastna imena gledati kot na nedeljive celote, med deli imena in kakršnokoli drugo omenitvijo pa ne sme biti nobenih povezav. V naslednjem primeru ne sme biti nobene koreferenčnosti, ker celotna zveza Narodna banka Srbije predstavlja ime institucije:

[4.1.a] Predpise glede finančnega poslovanja Srbije določa Narodna banka Srbije.

#### 4.2. Sestavljene entitete

Sestavljene omenitve so iz več manjših samostojnih omenitev. Navadno nastanejo tudi s pomočjo veznika in. V takšnih primerih je potrebno celotno takšno omenitev obravnavati ločeno od ostalih. Na primer:

[4.2.a] Janez in Katarina sta se prejšnjo nedeljo poročila. Javila sta se nam in oba nam sporočata, da njima je na medenem mesecu na Karibih prekrasno.

Če se v besedilu ločeno omenja posamezna omenitev iz skupine, je potrebno označiti tudi njo:

[4.2.b] Janez in Katarina sta se prejšnjo nedeljo poročila. Javila sta se nam in oba nam sporočata, da njima je na medenem mesecu na Karibih prekrasno. Janez obožuje srfanje, a Katarina ves dan uživa na plaži.

[4.2.b.1] Janez in Katarina sta se prejšnjo nedeljo poročila. Javila sta se nam in oba nam sporočata, da njima je na medenem mesecu na Karibih prekrasno. Janez obožuje srfanje, a Katarina ves dan uživa na plaži.

[4.2.b.2] Janez in Katarina sta se prejšnjo nedeljo poročila. Javila sta se nam in oba nam sporočata, da njima je na medenem mesecu na Karibih prekrasno. Janez obožuje srfanje, a Katarina ves dan uživa na plaži.

Kot že omenjeno: če se pojavi le ena omenitev sestavljene omenitve, je ne označujemo posebej:

[4.2.c] Janez in Katarina sta se prejšnjo nedeljo poročila. Janez je nosil mornarsko modro obleko, Katarina pa belo poročno obleko.

## 5. Prekinjene omenitve

Po izvornih srbskih smernicah je treba posebej označevati, ali je omenitev prekinjena ali ne. V slovenščini tega ne označujemo, vseeno pa ohranjamo to poglavje, saj vsebuje koristne zglede.

Če je omenitev v besedilu razbita v več delov, jo je potrebno zaradi tehničnih razlogov označiti kot eno neprekinjeno celoto, ki vključuje tudi besede, ki logično gledano ne sodijo v celoto. Na primer:

**[5.a]** Jovan se je trudil celo življenje. Sedaj je končno tudi uspel.

Da se takšne omenitve ločijo od ostalih, je potrebno pri omenitvi *je končno tudi uspel* nastaviti atribut `Discontinuous` na `True`. V vseh ostalih primerih je privzeta vrednost tega atributa nastavljena na `False`.

Eden redkih primerov, ko lahko vzpostavimo koreferenčno razmerje med omenitvijo in njenim delom je na primer:

**[5.b]** Najboljši mi je on prijatelj bil.

**[5.b.1]** Najboljši mi je on prijatelj bil.

**[5.b.2]** Najboljši mi je on prijatelj bil.

Ta rešitev izhaja iz koreferenčne enakovrednosti in besedne zveze (t.j. najboljši prijatelj v tem primeru), ko je vrstni red besed spremenjen. Primer z drugačnim vrstnim redom besed ne bi vseboval prekinjenih omenitev:

**[5.c]** On mi je najboljši prijatelj.

Pri tem anafore s prekinjeno predhodno omenitvijo ne smejo biti označene:

**[5.č]** Peter je srečal Marijo pred fakulteto. Oba sta od tam šla na predavanja.

Ne glede na to, da oba pomeni Petra in Marijo, nista predhodno omenjena kot skupna omenitev, zato se zaradi tega ne označi kot omenitev in posledično v tem primeru nimamo dodatne koreferenčnosti.

Prekinjene omenitve so možne tudi v primerih tipa `Clause`. Ko stavek ali del stavka vsebuje nekatere informacije, ki ne pripomorejo k identifikaciji omenitve (npr. kot je *Beriša*), jih moramo vseeno označiti, da lahko označimo celotno omenitev (t.j. odstranitev zaposlenih iz plačnega seznama):

**[5.d]** ... pogosto zaposlene, kot je Beriša, izbrišejo iz plačnega seznama, kar vodi do številnih protestov pred Agencijo.

## 6. Vrste omenitev

Po izvornih srbskih smernicah je treba za vsako omenitev določiti, kateri vrsti pripada. Ker so v slovenščini te informacije vključene v označevanje odvisnostne skladnje, jih ne označujemo, vseeno pa ohranjamo to poglavje, saj vsebuje koristne zglede. Vrste omenitev, ki jih uporabljajo, so naslednje:

- Copular – za imenske in pomožne predikate
- Apositive – za apozicije
- Zero anaphora – za ničelno anaforo
- Clause – za stavke in povedi
- Regular – za vse ostale omenitve

### 6.1. Copular – imenske entitete in pomožni predikati

Koreferenčnost med osebkom in samostalniškim delom imenske omenitve ali dopolnitvijo z glagolom je potrebno označiti, ko samostalnik/komplementarni predikat služi za identifikacijo referenta, na katero se osebek nanaša, na primer:

[6.1.a] Marko je moj oče.

[6.1.b] On je postal direktor podjetja.

V zgornjih stavkih bi bili omembi moj oče in direktor podjetja tipa Copular. Ker ima slovenščina prosti besedni red, je bilo za lažje označevanje sprejeto linearno pravilo. To pomeni, da je omenitev tipa Copular označena, ker se pojavi kasneje v stavku – za glagolom, omemba pred glagolom pa se obravnava kot osebek. Zaradi tega bi bila v naslednjem primeru omenitev tipa Copular pokrovitelj foruma, Makedonski predsednik pa tipa Regular. V stavku [6.1.č] pa ravno obratno.

[6.1.c] Makedonski predsednik je pokrovitelj foruma.

[6.1.č] Pokrovitelj foruma je makedonski predsednik.

Za razlikovanje omenitev samostalnikov in dopolnilnih predikatov, samostalniške označimo kot Copular-N, dopolnilne (npr. povedkove vezi) pa kot tip Copular-S. Povedkove vezi je potrebno označiti s tipom Copular, ko predstavljajo identifikacijo referenta, na primer:

[6.1.d] Teododosić je postal najboljši igralec tekmovanja.

Da bi razlikovali omenitve, ki predstavljajo osebek ali predmet, je potrebno pri označevanju omenitev Copular-S dodatno definirati še atribut Referent na vrednost Subject ali Object. Ta atribut ni potreben sicer, saj vedno nastopa kot osebek.

Težko je definirati vse povedkove vezi, ki jih je potrebno upoštevati (nekateri glagoli so polkopalitivni le v sekundarnem pomenu), vendar je pomembno poudariti, da so pomembni le tisti glagoli, ki omogočajo vzpostavitev koreference. Ostalih primerov ne označujemo, kot na primer:

[6.1.e] Đorđe se počuti kot najsrečnejši človek na svetu.

### 6.1.1. Časovno omejene koreferenčnosti

Zaradi enostavnosti označevanja je potrebno označevati tipe omenitev, ko je koreferenčnost časovno omejena, na primer:

**[6.1.1.a]** Donald Trump je izvoljen za predsednika ZDA.

**[6.1.1.b]** Ratzinger je bil papež do leta 2013.

Med drugim je potrebno razločevati, kdaj se katera od omenitev nanaša na specifičnega referenta. V primeru, da govorimo o več referentih z istimi funkcijami, vendar časovno omejenimi, jih moramo ustrezno povezati, kot na primer:

**[6.1.1.c]** Donald Trump je predsednik ZDA. Tudi Barack Obama je bil predsednik ZDA. Predsednik ZDA ima veliko odgovornosti.

V zgornjem primeru se zadnji izraz predsednik ZDA ne nanaša na določeno osebo, vendar na dolžnost predsednika v splošnem ali na skupino ljudi, ki to dolžnost opravlja. Četudi bi v besedilu nastopala le ena oseba, izraza ne bi označili za koreferenčnega, na primer:

**[6.1.1.č]** Donald Trump je predsednik ZDA. Predsednik ZDA ima veliko odgovornosti.

Prav tako koreferenčnosti ne označujemo, če razmerje še ni resnično oz. bo postalo resnično v prihodnosti. S tem se tudi izognemo združevanju časovno povezanih omenitev v eno verigo/skupino. V naslednjem primeru bi torej bilo napačno označiti koreferenčnost med omenitvami Donald Trump in predsednik ZDA.

**[6.1.1.d]** Zatem, ko Obama zaključi svoj mandat, bo Donald Trump postal predsednik ZDA.

Takšni primeri ilustrirajo, da je označevanje koreferenčnosti veliko bolj zapleteno, kot je le preprosto ujemanje enakih delov besedila v dokumentu.

### 6.1.2. Določljive ali kvalitativne omenitve

V nasprotju s predhodnimi primeri lahko omenitev le opisuje referenta (t.j. kvalitativna) in ne služi njeni identifikaciji (angl. določljiva). V takšnih primerih omenitev ne označujemo za koreferenčne, kot na primer:

**[6.1.2.a]** Moja hiša je rdeča.

**[6.1.2.b]** Ta sladoled je bil okusen.

**[6.1.2.c]** Petar Petrović je bil izbran za policaja.

**[6.1.2.č]** Naredili so ga močnega.

Kot pomoč pri razlikovanju med določljivimi in kvalitativnimi omenitvami uporabimo tudi sledeče pravilo. Če je omenitev v nadaljevanju uporabljena tako, da je jasno nakazana koreferenčnost na referenta, potem omenitev označimo (t.j. jo štejemo kot določljivo). V nasprotnem primeru jo štejemo kot kvalitativno in je ne označimo, na primer:



**[6.1.2.d]** Tamara je moja najboljša prijateljica ... Ker je bilo družinsko kosilo odpovedano, sem se jaz odločila, da grem s svojo najboljšo prijateljico v kino.

Besedna zveza (*moja*) najboljša prijateljica v nadaljevanju besedila nedvoumno nakazuje na referenta Tamara. Razlog za to je superlativ najboljša, saj se s tem enolično določa le en referent. Poleg tega referentapa je potrebno v tem stavku seveda označiti tudi druge referente, t.j. avtorja izjave:

**[6.1.2.d.1]** Tamara je jaz najboljša prijateljica ... Ker je bilo družinsko kosilo odpovedano, sem se jaz odločila, da grem s svojo najboljšo prijateljico v kino.

Nasprotno pa takšna vrsta identifikacije s kvalitativnimi omenitvami ni mogoča:

**[6.1.2.e]** Mirko je dober plavalec. ... Otroka je iz morja rešil dober plavalec, ki se je zadrževal v bližini.

Glede na podani kontekst zgoraj ne moremo biti prepričani, da je zadnja omenitev dober plavalec mišljena za Mirka, ker je mogoče, da se nanaša na kakega drugega plavalca. Iz tega primera opazimo predvsem, da je potrebno koreferenčnost označiti le v primeru, ko smo lahko prepričani, da se omenitev zagotovo nanaša na izbranega referenta.

V primeru, če za omenitev ni jasno, da je določljiva ali kvalitativna, je potrebno vedno upoštevati širši kontekst v besedilu. Poglejmo naslednji primer:

**[6.1.2.f]** Na ulici Knez Mihajlova lahko najdete različne vrste restavracij. Via del Gusto je picerija. Platonova akademija je knjigarna. ...

**[6.1.2.g]** Via del Gusto je picerija. Med njenimi pogostimi gosti so tudi znane osebnosti. Picerijo so obiskali že Novak Đoković, Vuk Bojović, Vladimir Veličković, ...

Potrebno je opozoriti, da je možno, da ima omenitev kvalitativno funkcijo in zato ne predstavlja kopularne koreferenčnosti, vendar se kljub temu na referenta sklicuje kot regularna omenitev, na primer:

**[6.1.2.h]** Ana, Peter in Frane so odlični učenci, takšne imajo vsi profesorji radi.

## 6.2. Apositive – apozicije

Omenitev se označi za apozicijsko, če je za drugo omenitvijo in običajno od nje ločena z ločili (npr. vejica, vezaj, oklepaj), kjer sta apozicijska in osnovna omenitev samostalniška zveza, ki se nanaša na istega referenta. Apozicijske in osnovne omenitve so v koreferenčnem odnosu, kot na primer.

**[6.2.a]** Britansko kraljico, Elizabeto II, sem spoznal na praznovanju njenega diamantnega jubileja.

V tem primeru bi bila omenitev Elizabeto II označena kot apozicija, medtem ko bi bila Britansko kraljico regularna omenitev. Za identifikacijo apozicijskih omenitev so ločila zelo pomembna. Na primer:

**[6.2.b]** Britansko kraljico Elizabeto II sem spoznal na praznovanju njenega diamantnega jubileja.

V primeru brez ločil torej ne bi bilo nobenega razloga, da bi omenitev Britansko kraljico Elizabeto II razbili na dve manjši, ampak bi bila celotna samostalniška zveza označena kot ena regularna omenitev.

### 6.2.1. Posebnosti apozicijskih omenitev

Pri označevanju apozicijskih omenitev je potrebno navesti njihovo specifičnost – torej ali je omenitev glavna (angl. head) ali sekundarna (angl. secondary) glede na regularno omenitev. Glavna omenitev je tista, ki referenta opredeli na najbolj jasen, natančen način. V okviru tega projekta uporabljamo tri stopnje specifičnosti omenitev, ki so odvisne od vrst besed, ki sestavljajo omenitev:

- Imenske entitete – samostalniki, ki vedno predstavljajo glavno omenitev, saj nedvoumno identificirajo referenta. Enako velja za samostalniške besedne zveze, ki vključujejo imenske entitete.
- Druge vrste samostalnikov – predstavljajo glavne omenitve, če ni definiranih imen, ki se nanašajo na opazovani referent. Enako velja za samostalniške besedne zveze, sestavljene iz njih.
- Zaimki – so najmanj informativni, ker se po svoji funkciji lahko nanašajo na kateri koli referent.

V zgornjem primeru bi bila zato omenitev Elizabete II označena kot glavna omenitev. Regularne omenitve Britansko kraljico ni potrebno označevati kot sekundarno, ker je to jasno določeno iz omenitve Elizabete II. Specifičnost regularnih omenitev se zato ne označuje posebej.

V naslednjem primeru je glavna omenitev samostalniška besedna zveza znanstveniki in raziskovalci, ker je v hierarhiji specifičnosti nad zaimki:

**[6.2.1.a]** Mi, znanstveniki in raziskovalci, menimo, da se morajo izdatki za znanost povečati.

Če imata tako regularna kot apozicijska omenitev enako stopnjo specifičnosti, je potrebno za glavno omenitev šteti levo, t.j. regularno omenitev. Zaradi tega bi bila v naslednjih dveh primerih apozicijska omenitev označena kor sekundarna:

**[6.2.1.b]** Naš edini tekmovalec v kategoriji mladincev, beograjski prvak v judu, je odpotoval na tekmovanje v Rusijo.

**[6.2.1.c]** Beograjski prvak v judu, naš edini tekmovalec v kategoriji mladincev, je odpotoval na tekmovanje v Rusijo.

### 6.2.2. Določljive ali kvalitativne omenitve

Tako kot v primeru kopularnih omenitev, je potrebno označevati tudi apozicijske omenitve, če nastopajo v identifikacijski funkciji. V primerih, ko apozicija nastopa v kvalitativni funkciji, se ne označuje kot omenitev in niti ne sme ustvarjati koreferenčne povezave z osnovno omenitvijo:

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

**[6.2.2.a]** Med dobitniki letošnje nagrade so Veljko Luković, policaj, in Vesna Milić, medicinska sestra.

V primeru, ko ni jasno, če apozicija identificira referenta ali ga le opisuje, je potrebno upoštevati kontekst v besedilu.

### 6.2.3. Ostalo

Preтворbe enot in datumi, ki se nanašajo na nekega referenta, so lahko določljive. Na primer:

**[6.2.3.a]** Iz banke je bilo včeraj ukradenih milijon funtov (približno 1,5 milijona dolarjev).

**[6.2.3.b]** Tekma je potekala v torek, 20. marca.

Primeri struktur, ki so podobne apozicijam, vendar jih ne označujemo, so na primer starosti oseb, kot v naslednjih primerih:

**[6.2.3.c]** Ivan Marković, 26, in Petar Milinković, 28, so bili aretirani v jutranji policijski akciji.

**[6.2.3.č]** Ivan Marković star 26 let, in Petar Milinković, star 28 let, so bili aretirani v jutranji policijski akciji.

Možno je tudi, da takšne strukture razdvajajo apozicijske omenitve, ki jih je potrebno označiti:

**[6.2.3.d]** Melisa Margarita Calderon, 30, najvplivnejša vodja mamilarskega kartela v Mehiki, je bila aretirana v soboto.

## 6.3. Zero anaphora – ničelna anafora

Ničelno anaforo je potrebno označiti, če referent ni jasno definiran s svojo omenitvijo, vendar je iz konteksta razvidno, da več predikatov kaže na isto entiteto:

**[6.3.c]** Odšel je v trgovino. Ko je prišel nazaj, je ugotovil, da so mu ukradli kolo.

Omenitve ničelne anafore vključujejo tudi kataforo, t.j. primere, ko je referent omenjen šele po nastopu glagola, na primer:

**[6.3.č]** Zatem, ko je odšel iz trgovine, je Marko ugotovil, da so mu ukradli kolo.<sup>2</sup>

Če omenitev vsebuje kopulo, jo je vedno potrebno označiti kot del omembe, četudi to pomeni, da so zaradi vrstnega reda besed v stavku vključene nekatere besede, ki ne spadajo v omenitev (za podrobnosti glej razdelek Prekinjene omenitve), na primer:

---

<sup>2</sup> Vedno označujemo samo eno omenitev istega referenta znotraj enega stavka. Prednost pri označevanju imajo samostalniki, nato so zaimki, nato je glagol.

[6.3.d] **Marko** je šel v trgovino. Ko je potem **prišel** nazaj, je **ugotovil**, da so mu ukradli kolo.

V primeru pomožnih glagolov, ko je le ta sestavljen iz vezi in povedkovega določila, je prav tako potrebno označiti koreferenčnost. V vlogi vezi nastopa nepolnopomenska oblika glagola, v vlogi povedkovega določila pa polnopomenska beseda (samostalniška, pridevniška beseda ali nedoločnik).

[6.3.e] **Nikola** je opravil zadnji izpit. **Bil** je utrujen in srečen.

[6.3.f] **Nikola** je opravil zadnji izpit. **Počutil** se je srečnega.

Tudi naklonske glagole je potrebno obravnavati kot vse druge glagole:

[6.3.g] **Ana** se je zbudila ob zori. Še vedno je **bila** zaspana, a je **morala** iti na delo.

[6.3.h] Nadrejeni **mu** je že velikokrat pogledal skozi prste. Ni **mu mogel** odreči usluge. (drugi »mu« nastopa v obeh omenitvah)

[6.3.i] **Moram** iti na sestanek, saj vedno čakajo **name**.

[6.3.j] Napaka **je** moja in **želim**, da to javno **povem**.

Označiti je potrebno tudi »da-« konstrukcije glagolov:

[6.3.k] »Ali bo **predsednik** rešil ta problem? **Mislim**, da ne **bo**.« je rekel Tatar.

[6.3.l] »Prepričan **sem**, da si tega ne bi **upali** storiti,« je **Filip** dodal **kolegom**.

[6.3.m] **Verjamem**, da sem **zaslužil** to nagrado.

Ničelna anafora se lahko pojavi tudi sledeče:

[6.3.n] **Mi** menimo, da je trditev, da **poskušamo** zaobiti postopke, smešna.

[6.3.o] **Mi** menimo, da je trditev, da **poskušamo**, da bi **zaobšli** postopke, smešna.

## 6.4. Clause – stavki in povedi

Stavčne omenitve, so omenitve, ki vključujejo celotne stavke (neodvisne ali odvisne). V sledečih primerih so to primeri stavkov *da Izrael sabotira misijo, Predsednik je razpustil parlament, Pozabil sem prinesiti denarnico, da bo vodil državo po poti približevanja Evropi*:

[6.4.a] Aktivisti so izjavili, **da Izrael sabotira misijo**, vendar so oblasti **to trditev** zavrnile.

[6.4.b] **Predsednik je razpustil parlament**, **čemu** so se uprle vse opozicijske stranke.

[6.4.c] **Pozabil sem prinesiti denarnico**, **kar** je povzročilo kup težav.

[6.4.č] **Novi predsednik je obljubil**, **da bo vodil državo po poti približevanja Evropi**. Med drugim **mu** ne bo lahko, da **to** stori.

Zadnji primer vsebuje tudi ničelno anaforo in regularne omenitve, ki so označene s sivo.

Omenitve v obliki stavkov so lahko koreferenčne ne samo z zaimki, vendar tudi z neodvisnimi samostalniki ali besednimi zvezami:

Operacijo Razvoj slovenščine v digitalnem okolju sofinancirata Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020.

[6.4.d] Hekerji so uspeli ukrasti podatke o več kot 485.000 kreditnih karticah. Krajo so odkrili šele po ...

[6.4.e] V preteklih letih je evropska ekonomija hitro rasla, pri čemer je rasti doprinesel ...

[6.4.f] Na ulici Mihajla Pupina sta se zalela avtomobil in avtobus. Do težke nesreče je prišlo, ko je ...

[6.4.g] Na fakulteti sem srečala prijateljico. Zato sem zamudila.

[6.4.g.1] Na fakulteti sem srečala prijateljico. Zato sem zamudila.

[6.4.g.2] Na fakulteti sem srečala prijateljico. Zato sem zamudila.

Omenitve stavkov omejimo na omenitve samih glagolov, na primer:

[6.4.h] Jaz sem pred vsemi temi dogodki želel, da pojem, vendar sem s tem sedaj prenehal.

[6.4.i] Od veselja je želel skočiti in sedaj je to storil.

[6.4.i.1] Od veselja je želel skočiti in sedaj je to storil.

[6.4.i.2] Od veselja je želel skočiti in sedaj je to storil.

Upoštevati je potrebno, da stavki, ki imajo funkcijo dodatnega opisa omenitve in predstavljajo omenitev skupaj z omenitvijo, ne spadajo v tip Clause. V slovenskem jeziku do takšnih primerov ne prihaja, saj so podredni stavki načeloma ločeni z ločilom.

Prav tako v primeru, da stavek vsebuje nekatere podatke, ki ne prispevajo k identifikaciji referenta, ga je potrebno v celoti označiti kot omenitev. Na primer:

[6.4.j] Še vedno vsako jutro pijem kavo. To je za mene ritual.

[6.4.j.1] Še vedno vsako jutro pijem kavo. To je za mene ritual.

[6.4.j.1] Še vedno vsako jutro pijem kavo. To je za mene ritual.

Zaradi konstitentnosti označevanja smo se odločili, da posamezna koreferenčna veriga ne more biti le iz omenitev tipa Clause, vendar mora vsebovati vsaj še kak drug tip omenitev, da lahko obstaja.

## 6.5. Regular – ostali tipi omenitev

Vse ostale omenitve, ki ne spadajo v katerega od zgornjih tipov, se označuje kot regularne (t.j. običajne) omenitve. Ta tip je tudi najbolj pogost (kar je moč opaziti tudi v dosedanjih primerih).

V besedilih se ponekod pojavijo omenitve, ki so podobne kopularnim omenitvam. Na primer:

[6.5.a] Patrick Johnson bo na poziciji predsednika korporacije.

[6.5.b] Patrick Johnson bo opravljal dolžnosti kot predsednik korporacije.

Vse takšne omenitve je potrebno označiti kot običajne, ne glede na pomensko podobnost s kopularnimi omenitvami (npr. Patrick Johnson je bil imenovan za predsednika korporacije).

## 7. Literatura

### 7.1. Navodila za označevanje koreferenčnosti na drugih projektih

1. RELDI: Uputstvo za anotiranje koreferenci, Verzija 1.1, Januar 2018.
2. Richer Event Description (RED) Annotation Guidelines (v.1.7), Martha Palmer, Will Styler, Kevin Crooks, Tim O’Gorman, <https://github.com/timjogorman/RicherEventDescription/blob/master/guidelines.md>
3. Coreference Corpus Annotation Guidelines for ACL Anthology papers, Ulrich Schäfer, Christian Spurk, Jörg Steffen, <http://take.dfki.de/C12-2103-dataset/AnnotationGuidelines.html>
4. Co-reference Guidelines for English OntoNotes (version 6.0), BBN Technologies, <https://catalog.ldc.upenn.edu/docs/LDC2007T21/coreference/english-coref.pdf>
5. 2006 Guidelines for Annotation of Within-document NP Coreference, L.Hasler, K. Naumann, C. Orasan, [http://rgcl.wlv.ac.uk/projects/NP4E/NP\\_guidelines\\_2006.pdf](http://rgcl.wlv.ac.uk/projects/NP4E/NP_guidelines_2006.pdf)
6. MUC-7 Coreference Task Definition (Version 3.0), L. Hirschman, N. Chinchor, [http://www-nlpir.nist.gov/related\\_projects/muc/proceedings/co\\_task.html](http://www-nlpir.nist.gov/related_projects/muc/proceedings/co_task.html)
7. EuroParl Named Entity Annotation Guidelines (v1.0), Andreea Bodnari, Pierre Zweigenbaum, <http://web.mit.edu/andreeab/www/corpus/annotationGuidelines.pdf>
8. Automatic Content Extraction 2004 Evaluation – Annotation Guidelines for Entity Link Tracking (LNK) (version 3), <https://catalog.ldc.upenn.edu/docs/LDC2005T09/guidelines/EnglishLNKV3-0.PDF>
9. Clinical Coreference Annotation Guidelines (with excerpts from ODIE guidelines and modified for SHARPN/THYME), Arrick Lanfranchi, Kevin Crooks, and Mariah Hamang, 2013, [http://clear.colorado.edu/compsem/documents/coreference\\_guidelines.pdf](http://clear.colorado.edu/compsem/documents/coreference_guidelines.pdf)

### 7.2. Raziskovalni članki

1. M. Ogrodniczuk, M. Zawistawska, K. Głowińska, and A. Savary, “Coreference Annotation Schema for an Inflectional Language,” in Proceedings of the 14th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2013), 2013, pp. 394–407.
2. M. Ogrodniczuk, K. Głowińska, M. Kopeć, A. Savary, and M. Zawistawska, “Interesting Linguistic Features in Coreference Annotation of an Inflectional Language,” in Proceedings of the 12th China National Conference on Computational Linguistics (CCL 2013) and the First International Symposium on Natural Language Processing Based on Naturally Annotated Big Data (NLP-NABD 2013), 2013, pp. 97–108.
3. S. Pradhan, A. Moschitti, N. Xue, O. Uryupina, and Y. Zhang, “CoNLL-2012 Shared Task: Modeling Multilingual Unrestricted Coreference in OntoNotes,” in Proceedings of the Joint Conference on EMNLP and CoNLL: Shared Task, 2012, pp. 1–40.
4. M. Recasens, M. A. Martí, and C. Orasan, “Annotating Near-Identity from Coreference Disagreements,” Proceedings of LREC 2012, pp. 165–172, 2012.
5. M. Recasens, “Coreference: Theory, Annotation, Resolution and Evaluation,” PhD dissertation, 2010.

6. U. Schäfer, C. Spurk, and J. Steffen, "A Fully Coreference-annotated Corpus of Scholarly Papers from the ACL Anthology," in Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012), 2012, pp. 1059–1070.
7. L. Hasler, C. Orasan, and K. Naumann, "NPs for Events: Experiments in Coreference Annotation," in Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006), 2006, pp. 1167–1172.
8. K. N. Björkenstam, "SUC-CORE: A Balanced Corpus Annotated with Noun Phrase Coreference," Northern European Journal of Language Technology, vol. 3, pp. 19–39, 2013.
9. M. Garcia and P. Gamallo, "Multilingual corpora with coreferential annotation of person entities," in Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014), 2014, pp. 3229–3233.
10. A. Ghaddar and P. Langlais, "WikiCoref: An English Coreference-annotated Corpus of Wikipedia Articles," in Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016), 2016, pp. 136–142.