

# Označevalne smernice

*Ustvarjanje slovenske množice SI-NLI za sklepanje o pomenskem sosledju besedil*

Razpis CLARIN.SI 2022

Matej Klemen, Aleš Žagar, Jaka Čibej, Marko Robnik Šikonja

Različica: 1.0

Zadnja posodobitev: 2022-07-08

CC-BY-SA 4.0

## 1 Uvod

Sklepanje o pomenskem sosledju besedil (ang. *natural language inference*, NLI) je naloga, ustvarjena za ocenjevanje zmožnosti naprednih strojnih modelov, da razumejo naravni jezik.

Na podlagi podanega para besedil (predpostavka in hipoteza) je cilj ugotoviti:

- ali drugo besedilo pomensko **izhaja** iz prvega (izhajanje, ang. *entailment*);
- ali drugo besedilo prvemu pomensko **nasprotuje** (nasprotovanje, ang. *contradiction*);
- o pomenskem odnosu med besediloma **ni mogoče sklepati** (nevtralnost, ang. *neutral*).

Vsak par besedil je glede na to označen z natanko eno od teh treh oznak.

Oglejmo si spodnji primer: za isto predpostavko navedemo tri primere hipotez (po eno za vsak pomenski odnos).

### **PREDPOSTAVKA:**

Preiskave med sodnim postopkom so pokazale, da so se dogodki odvijali bistveno drugače.

—

### **HIPOTEZA 1:**

Preiskovalci so ugotovili, da poročanja o dogodku niso bila resnična.

Hipoteza **sledi** iz predpostavke (*entailment*), ker tako predpostavka kot hipoteza govorita o neresničnosti (prvotno opisanih) dogodkov.

—

### **HIPOTEZA 2:**

Obdukcija je pokazala precej drugačno sliko.

O relaciji med predpostavko in hipotezo **ni mogoče sklepati** (*neutral*), saj ni nujno, da je prav obdukcija vodila k spoznanju, da poročanja niso bila resnična.

—

### **HIPOTEZA 3:**

Zgodba se je izkazala za resnično tudi po opravljenih preiskavah.

Hipoteza pomensko **nasprotuje** (*contradiction*) predpostavki - predpostavka trdi, da so se dogodki odvijali bistveno drugače, hipoteza pa trdi, da so se dogodki ("zgodba") zgodili tako, kot so bili predstavljeni pred preiskavo.

## 2 Cilj in potek naloge

Glavni namen naloge je pripraviti učno množico iz parov slovenskih povedi za predstavljene pomenske odnose. Primeri morajo biti zahtevni, a naravni.

Pri nalogi dobimo dve podobni povedi (**predpostavko** in **hipotezo**), cilj pa je ugotoviti, katera oznaka (*entailment*, *contradiction* ali *neutral*) je ustrezna za hipotezo. Nato v skladu z navodili tvorimo dodatni dve povedi za manjkajoča odnosa (bodisi s *spreminjanjem hipoteze* bodisi s *tvorjenjem novih povedi*). Če je druga poved v prvotno podanem paru neustrezna, je treba tvoriti povedi za vse tri pomenske odnose.

**POZOR!** Same predpostavke (tj. prve povedi iz izvornega para) nikoli ne spreminjamo!

Pazimo, da se ne zapletemo v prekomerno razmišljanje o tem, ali gre pri primeru za izhajanje, nasprotovanje ali nevtralnost. Vodilo naj bo, da za vsak tak primer lahko porabimo največ 30 sekund. Če je primer problematičen, lahko to zapišemo v stolpec za komentarje in/ali pošljemo sporočilo na [clarin2022-sinli@googlegroups.com](mailto:clarin2022-sinli@googlegroups.com).

## 3 Navodila za tvorjenje povedi

V nadaljevanju natančneje opredelimo navodila za ustvarjanje primerov.

**POZOR!** Upoštevajte, da so to le smernice, ne stroga pravila! Navodila smo preizkusili, a jim morda v določenih primerih ne bo mogoče slediti. V primeru dvomov pišite na skupino ([clarin2022-sinli@googlegroups.com](mailto:clarin2022-sinli@googlegroups.com)).

### 3.1 ✗ Neustrezni načini tvorjenja povedi

Za vsak pomenski odnos je mogoče ustvariti (preveč) preproste primere, iz katerih pa se model ne bo naučil kompleksnejših pravil razumevanja jezika. Takim primerom se izogibamo.

#### 3.1.1 Splošno o neustreznih primerih

##### 3.1.1.1 Prevelika prekrivnost

Če je več kot 70 % hipoteze enakih predpostavki (npr. vsebuje skoraj enake besede), je primer neustrezen. Vsaj tretjina povedi mora biti drugačna.

[P] Janez je v trgovini kupil jabolka.

[H] Janez je v trgovini nabavil jabolka. ✘

### 3.1.1.2 Nestandardna slovenščina

V hipotezo ne vnašamo zatipkov in nestandardne slovenščine (razen v primerih, ko je nestandardna slovenščina že v izvornem paru).

[P] Janez je v trgovini kupil jabolka.

[H] Janez je u trgovin kupu japka. ✘

### 3.1.1.3 Preprosta zamenjava števil, imenskih entitet ipd.

Ni dovolj, da v primerjavi s predpostavko v hipotezi le zamenjamo nekaj števil ali imen.

[P] Janez je kupil 5 jabolk.

[H] Janez je kupil 10 jabolk. ✘

[H] Metka je kupila 5 jabolk. ✘

## 3.1.2 ✘ Neustrezni primeri izhajanja

### 3.1.2.1 Občutno prekrivanje/podniz

Hipoteza vsebuje del predpostavke (z minimalnimi popravki) ali pa ponovno uporablja mnogo besed iz predpostavke.

[P] Poiskali smo nekakšno srednjo pot.

[H] Ubrali smo srednjo pot. ✘

### 3.1.2.2 Hipoteza je le skrajšana predpostavka

Hipoteza je le skrajšana različica predpostavke (to je tudi pogosta posledica tega, da ustvarjamo podnize predpostavk.) oz. je vedno občutno krajša od prepostavke.

[P] Janez je kupil 5 jabolk, da bi pripravil jabolčni zavitek.

[H] Janez je kupil 5 jabolk. ✘

## 3.1.3 ✘ Neustrezni primeri nevtralnosti

### 3.1.3.1 Občutna sprememba teme

Če hipoteza govori o popolnoma drugi temi kot predpostavka, je primer neustrezen. Ko tvorimo nevtralno hipotezo, moramo tvoriti poved, ki je še vedno pomensko povezana s predpostavko.

[P] Janez je kupil 5 jabolk.

[H] Zunaj pada dež. ✘

### 3.1.3.2 Preprosto dodajanje novih informacij

Ni ustrezno, da hipotezo napravimo nevtralno samo tako, da predpostavki dodamo neko preprosto novo informacijo (npr. naštevane novih predmetov).

[P] Janez je kupil 5 jabolk.

[H] Janez je kupil 5 jabolk in 7 hrušk. ✘

Poved je nevtralna takrat, ko o njej ne moremo sklepati iz predpostavke - največkrat zato, ker je njen pomen podrobnejši ali ožji. Najlažje to dosežemo z dodajanjem novih informacij v poved, a se bo v tem primeru model naučil kot nevtralne razpoznavati le tiste povedi, ki v primerjavi s predpostavko vsebujejo nekaj novega. Naš cilj je, da mu ponudimo tudi drugačne primere: pomen lahko npr. napravimo podrobnejši oz. bolj ozek tudi z uporabo drugačnih besed - v spodnjem primeru npr. v hipotezi trdimo, da je oseba čebulo sekljala; v predpostavki piše, da jo je rezala, kar ni nujno isto (velikost koščkov je drugačna; če čebulo sekljamo, smo jo v vsakem primeru narezali; če smo jo narezali, pa ni nujno, da smo jo nasekljali).

[P] Pred kosilom si jo največkrat videl za kuhinjskim pultom, kako **reže** čebulo.

[H] Med pripravo kosila je v kuhinji pogosto **sekljala** čebulo. (nevtralnost)

### 3.1.3.3 Hipoteza je podaljšana predpostavka

Ni ustrezno, da je hipoteza le podaljšana različica predpostavke (to je pogosta posledica preprostega dodajanja informacij). Pazimo tudi, da hipoteza ni vedno konsistentno daljša od predpostavke, sicer se model za ugotavljanje pomenskega odnosa preveč zanaša samo na dolžino povedi.

[P] Janez je kupil 5 jabolk.

[H] Janez je kupil 5 jabolk, da bi pripravil jabolčni zavitek. ✘

## 3.1.4 ✘ Neustrezni primeri nasprotovanja

### 3.1.4.1 Preprosto zanikanje predpostavke

Hipoteza je le zanikana oblika predpostavke, tvorjena s preprostimi menjavami posameznih besed ("sem" -> "nisem", "je" -> "ni").

[P] Oddal sem domačo nalogo.

[H] Nisem oddal domače naloge. ✘

### 3.1.4.2 Uporaba zelo omejenega nabora besed

Paziti moramo, da hipoteze ne zanikamo vedno na povsem enak način oz. da se ne zatekamo ves čas k enim in istim besedam - spodnji primeri so neustrezni, ker se vedno znova zatekamo k besedi "spati", da zanikamo to, kar je osebek počel v predpostavki. Model bi se iz takšnih primerov naučil le to, da mora biti pozoren na besedo "spati".

[P] Janez je zjutraj igral nogomet.

[H] Janez je zjutraj **spal**. ❌

[P] Janez je pojedel zajtrk.

[H] Janez je **spal**. ❌

## 3.2 Ustrezni načini tvorjenja povedi

V tem razdelku podajamo še nekaj vzorcev ustvarjanja ustreznih primerov. Kombinacija teh vzorcev naj vam služi kot izhodišče pri ustvarjanju primerov.

**POZOR!** Seznam **ni izčrpen**: obstajajo tudi drugi ustrezni načini ustvarjanja primerov. Ne osredotočajte se le na eno kategorijo tvorjenja povedi, saj je cilj naloge, da pridobimo čimveč raznolikih primerov.

### 3.2.1 Zdravorazumsko sklepanje

Hipoteza vsebuje vsebuje dodatno znanje, ki je ljudem samoumevno (v spodnjem primeru npr. vemo, da če človek pritisne na zavoro, se hitrost zmanjša, čeprav to ni izrecno navedeno v predpostavki).


[P] Janez je med vožnjo pritisnil na zavoro.

[H] Hitrost se je zmanjšala.  (izhajanje)

### 3.2.2 Netrivialno sklepanje o številih

Ustrezno je, če v hipotezi uporabimo sklepanje o številih, ki ni trivialno in ki zahteva nekaj dodatnega znanja. V spodnjem primeru npr. iz predpostavke lahko razberemo, da ima oseba rojstni dan tri mesece po juniju, torej najverjetneje septembra. Za to, da pridemo do tega sklepa, nam ni treba znati le šteti, temveč potrebujemo tudi dodatno znanje o mesecih v letu, zato primer ni trivialen.

[P] Z načrtovanjem je začel že junija, tri mesece pred njenim rojstnim dnevom.

[H] Rojstni dan ima jeseni, natančneje septembra.  (izhajanje)

### 3.2.3 Vključevanje splošnega znanja o svetu

Ustrezno je, če v hipotezo vključujemo (splošno znana) dejstva o svetu. V spodnjem primeru npr. iz splošne razgledanosti vemo, da "vijoličasti" predstavljajo Maribor in "zeleni" Olimpijo.

[P] Vijoličasti so z zelenimi remizirali.

[H] Odvila se je nogometna tekma med Mariborom in Olimpijo. ✓ (izhajanje)

### 3.2.4 Časovno/prostorsko sklepanje

Ustrezno je, če v hipotezo vključujemo svoje sklepanje o prostoru in času. V spodnjem primeru iz predpostavke izvemo, da je reka široka pet metrov. Ker vemo, da ljudje ne morejo skočiti tako daleč, lahko sklepamo, da hipoteza, da je nekdo skočil z enega brega na drugi breg, predpostavki nasprotuje.

[P] Med bregoma teče pet metrov široka reka.

[H] Skočil sem z enega na drugi breg. ✓ (nasprotovanje)

### 3.2.5 Frazemi, idiomi in metafore

Uporaba metaforičnih izrazov in frazemov je ustrezna (npr. v spodnjem primeru "prestaviti" - "načrti so padli v vodo").

[P] Sestanek je bil prestavljen zaradi nepredvidenih okoliščin.

[H] Načrti za sestanek so padli v vodo. ✓ (izhajanje)

## 3.3 ! Pogojno ustrezni načini tvorjenja povedi

V tem razdelku navajamo načine tvorjenja povedi, ki so sicer dovoljeni in ustrezni, a moramo paziti, da povedi ne tvorimo le z enim takšnim načinom (npr. zamenjamo le eno besedo s sopomenko). Vodilo je, da mora biti v najslabšem primeru vsaj tretjina povedi drugačna, v najboljšem pa morajo biti povedi karseda raznolike.

### 3.3.1 Uporaba sopomenk, nadpomenk, protipomenk

[P] Ko sem še živel v Frankfurtu, sem se pogosto vozil z vlakom.

[H] Ko sem živel v Frankfurtu, sem se pogosto prevažal z vlakom. ✗

Zamenjali smo "voziti se" s "prevažati se" in odstranili "še" - sprememba ni zadostna, povedi sta preveč prekrivni.

[H] V Nemčiji sem večkrat uporabljal javni potniški promet. ✓ (izhajanje)

Vključili smo splošno znanje, da je Frankfurt v Nemčiji; "vlak" smo zamenjali z nadpomenko "javni potniški promet"; "pogosto" smo zamenjali z "večkrat"; hipoteza ni občutno krajša od predpostavke in se od nje dovolj razlikuje.

### 3.3.2 Zamenjava vrstnega reda besed, tvornik/trpnik

Pazimo, da zamenjava vrstnega reda ni edina sprememba. Velja tako za zamenjavo vrstnega reda besed kot za npr. zamenjavo vršilca dejanja in prizadetega (osebek - predmet; tvornik - trpnik).

**[P]** Vse vilice in nože, ki so jih pobrali iz lijaka, so zložili v predale v omari.

**[H]** V predale v omari so zložili vse vilice in nože, ki so jih pobrali iz lijaka. ❌

Sprememba ni zadostna - spremenili smo le vrstni red.

**[H]** Lijak je bil poln pribora, zato so ga pospravili v omaro. ✅ (izhajanje)

Spremenili smo vrstni red in obenem uporabili sopomenski izraz (zložili - pospravili) ter nadpomenko (pribor).

**[P]** Janez je Metki podaril rože.

**[H]** Metki je Janez podaril rože. ❌

Nezadostna sprememba - zamenjan je le vrstni red.

**[H]** Metka je od Janeza v dar prejela šopek cvetic. ✅ (izhajanje)

Uporabili smo izraz, ki ponuja drugačen pogled na dejanje z vidika vršilca dejanja (podariti - prejeti v dar), uporabili smo sopomenski izraz ("šopek cvetic") in zamenjali vrstni red.

**[P]** Prejšnji teden, morda v sredo, ji je poštar dostavil pismo v močno okrašeni kuverti.

**[H]** Pismo v močno okrašeni kuverti ji je bilo dostavljeno prejšnji teden, morda v sredo. ❌

Primer je neustrezen, ker gre le za spremembo trpnik - tvornik.

**[H]** Pismo v izumetničeni kuverti ji je bilo dostavljeno prejšnji teden okrog srede. !

Pogojno sprejemljiv primer (trpnik - tvornik; "morda v sredo" - "okrog srede", "okrašen" - "izumetničen"), a bi lahko bil še boljši.

**[H]** Pismo v kičastem in izumetničenem ovoju ji je bilo izročeno enkrat v sredini tedna. ✅

"ovoj" - "kuverta", "dostavljeno" - "izročeno", "okrog srede" - "enkrat v sredini tedna", trpnik - tvornik.

### 3.3.3 Uporaba morfoloških značilnosti jezika

Smiselno lahko npr. spreminjamo spol, število ipd., a nujno dodamo še druge spremembe. Pazimo tudi, da števila in spola ne spremenimo tako, da poved govori o nekom povsem drugem, ki v predpostavko ni bil vključen (npr. "Igrala je na kitaro." -> "Igral je na kitaro.").

**[P]** S kolesom sva se odpravila do starega mlina nedaleč od kraja, kjer se stekata obe reki.

**[H]** S kolesom sem se odpravil do starega mlina od kraja, kjer tečeta reki. ❌

Omenjamo le eno osebo namesto obeh (kar je ustrezno - če sva se odpravila oba, to nujno pomeni, da sem se odpravil tudi jaz) in nekoliko spremenimo zadnji del ("tečeta reki"), a je poved še vedno preveč prekrivna.

**[H]** Odkolesaril sem vse do mlina, ki stoji zelo blizu sotočja obeh rek. ✅ (izhajanje)

Zadostna sprememba - sopomenke ("odpraviti se s kolesom" - "odkolesariti"; "kraj, kjer se stekata obe reki" - "sotočje"), smiselna sprememba slovnične osebe (midva - jaz).

### 3.3.4 Uporaba kratic/okrajšav oz. njihovih razširjenih oblik

Ustrezno je, da uporabljamo ustaljene okrajšave oz. da jih razširjamo. Tudi kratice lahko zapisujemo s polno obliko. Ne smemo pa okrajšav uvajati ad hoc (npr. krajšamo besedo "temperatur" v "temp.").

**[P]** V ZDA so zaradi nizkih temperatur razglasili izredne razmere.

**[H]** Združene države Amerike so ob nizkih temperaturah razglasile izredne razmere. ❌

Razširjena kratica in spremenjen predlog ("ob") nista zadostna sprememba.

**[H]** Ker so Združene države Amerike v primežu hudega mraza, so tam uvedli izredne razmere. ✅ (izhajanje)

Razvezana kratica, sopomenski izrazi ("nizke temperature" - "hud mraz", "razglasiti" - "uvesti")

**[P]** Več o pripravi bezgovega sirupa si lahko preberete na str. 5.

**[H]** Več o pripravi sirupa si lahko preberete na strani 5. ❌

Nezadostna sprememba - samo razširjena kratica in okrajšan izraz ("bezgov sirup" - "sirup").



**[H]** Postopek izdelave sirupa iz bezgovih cvetov je podrobneje opisan na strani 5. ✓  
(izhajanje)

Sopomenski izraz ("bezgov sirup" - "sirup iz bezgovih cvetov"), razvezana okrajšava ("str." - "strani"), prestrukturirana poved ("več o pripravi si lahko preberete" - "postopek je opisan").

### 3.3.5 Anafore in koreferenčnost

Uporaba zaimkov in drugih anaforičnih sredstev je dopustna, ne sme pa biti edina sprememba.

**[P]** Po še enem dolgoveznem in nepotrebem govoru se je odločila, da podpredsednika na sestankih sploh ne bo več poslušala.

**[H]** Odločila se je, da ga ne bo več poslušala. ✗

Neustrezen primer - Hipoteza je občutno krajša od predpostavke; gre za skrajšan del izvorne predpostavke; uporabljen je le zaimek namesto "podpredsednika".

**[H]** Spet je nakladal povsem brez potrebe in sklenila je, da med sestanki sploh ne bo več pozorna na njegove besede. ✓ (izhajanje)

Ustrezen primer - močno prestrukturiran stavek, sopomenski izrazi ("odločila" - "sklenila", "poslušati" - "biti pozoren na (njegove) besede"), anaforična sredstva ("podpredsednika" - "ga").

### 3.4 Dodatni primeri za boljšo predstavbo

V nadaljevanju navajamo še nekaj primerov s pojasnili, zakaj so (ne)ustrezni.

**[P]** Oči so se mi privadile na somrak.

**[H]** Mi je kar mrak padel na oči. ✗

Neustrezen primer, ki ga sploh ne moremo povezati s predpostavko - ima metaforičen pomen, ki pomeni nekaj povsem drugega. Obravnavamo ga kot občutno spremembo teme.

**[H]** Začel sem bolje zaznavati obrise v temi. ✓ (izhajanje)

Ustrezen primer za izhajanje - poved ni prekrivna s predpostavko, ni občutno krajša ali daljša, vsebuje splošno znanje (če se nam oči privadijo na mrak, bolje zaznavamo obrise v temi).

**[H]** V votlini je bilo temno. ! (nevtralnost)

Pogojno sprejemljiv primer za nevtralnost - v predpostavki ni (in)direktne omembe votline

ali česa podobnega, torej smo v hipotezo dodali neko novo informacijo. To ni najboljši način za tvorjenje nevtralnih povedi (glej 3.1.3.2).

**[H]** Običajno v temi nisem imel težav z vidom.  (nevtralnost)

Iz predpostavke ne moremo sklepati, ali oseba res običajno nima težav z vidom v mraku.

**[H]** V črnini nisem mogel videti ničesar.  (nasprotovanje)

Ustrezen primer za nasprotovanje - imamo sopomenske izraze ("somrak" - "črnina"), poved ni podniz predpostavke in je zadostno prestrukturirana, njen pomen pa nasprotuje predpostavki (če so se nam oči privadile na mrak, najverjetneje nekaj vidimo).

**[P]** O tem nam je danes povedal: "V nekem trenutku me je naš policist povabil na kratek pogovor."

**[H]** Na današnji dan nas je obvestil o policistovem vabilu na razgovor.  (izhajanje)

Ustrezen primer - sopomenski izrazi ("danes" - "na današnji dan", "pogovor" - "razgovor", "povedal" - "obvestil"), prestrukturirana poved (sprememba premege govora; glagolska struktura v samostalniško - "povabil" - "vabilo").

**[H]** Možje v modrem ga v tej zadevi niso nikoli kontaktirali.  (nasprotovanje)

Ustrezen primer za nasprotovanje - sorodni izrazi ("policist" - "možje v modrem", "kontaktirati" - "povabiti") in pomensko nasprotovanje predpostavki (če je oseba povedala, da ga je policist povabil na razgovor, ne more veljati, da ga možje v modrem nikoli niso kontaktirali).

**[H]** S policistom se je pogovoril o dogodku.  (nevtralnost)

Ustrezen primer za nevtralnost - ne vemo, ali se je oseba s policistom že pogovorila ali ne.

## 4 Pogosta vprašanja

**Predpostavka vsebuje nestandardno slovenščino (npr. sleng) ali tipkarsko napako (npr. "torej" -> "torj"). Ali jo lahko popravim?**

Ne, predpostavke nikoli ne popravljamo. Če predpostavka vsebuje nestandardno slovenščino, je najbolje, da tudi ustvarjena hipoteza vsebuje nestandardno slovenščino - ni pa nujno.