
Od rječnika kao korpusa do hiponimije i meronimije¹

Naučni rad

DOI: 10.18485/judig.2025.1.ch28

Maja Matijević²,  0000-0003-3990-0826

Apstrakt

Rad se bavi hijerarhijskim leksičko-semantičkim odnosima, hiponimijom i meronimijom, te poluautomatskim prepoznavanjem tih odnosa u rječničkim definicijama. Hiponimija se pritom definira kao odnos između leksičke jedinice koja označuje vrstu i leksičke jedinice koja označuje predstavnika vrste (npr. hiperonim *životinja* > hiponim *pas*), a meronimija kao odnos između leksičke jedinice koja označuje cjelinu i leksičke jedinice koja označuje dio te cjeline (npr. holonim *tijelo* > meronim *ruka*). Definira se teorijski okvir te se prikazuje mogućnost automatiziranoga prepoznavanja u rječničkim definicijama. Metodološko polazište čine leksičko-sintaktički obrasci, tj. tipične strukture s određenim sintaktičkim pravilnostima i leksičkim jedinicama s pomoću kojih se mogu prepoznati leksičko-semantički odnosi. Analizom rječničkih definicija dobiveno je više obrazaca te su istaknuti padežni i prijedložni izrazi, važni za izricanje odnosa dijela i cjeline. U radu se ističe da su rječnici zbog ujednačene strukture bogat i pouzdan izvor za prepoznavanje hijerarhijskih odnosa, što otvara mogućnost za daljnja jezična istraživanja i analize.

Ključne riječi: leksičko-semantički odnos, hiponimija, meronimija, leksičko-sintaktički obrazac, rječnik

¹ Ovaj rad napisan je na istraživačkome projektu *Hrvatski mrežni rječnik (Mrežnik): 2. faza*, koji financira Europska unija – NextGenerationEU. Za iznesene stavove i mišljenja odgovorna je samo autorica te ti stavovi ne odražavaju službena stajališta Europske unije ili Europske komisije. Ni Europska unija ni Europska komisija ne mogu se smatrati odgovornima za njih.

² Institut za hrvatski jezik, mmatijevic@ihjj.hr

1. Uvod

Hiponimija i meronimija hijerarhijski su leksičko-semantički odnosi, tj. odnosi koji ustrojavaju leksičke hijerarhije. Hiponimija je pritom odnos između leksičke jedinice koja označuje vrstu općenito i predstavnika te vrste (npr. *životinja* > *pas*), a meronimija odnos između leksičke jedinice koja označuje cjelinu i leksičke jedinice koja označuje dio te cjeline (npr. *tijelo* > *ruka*).

Glavna je značajka obaju odnosa višerazinski ustroj, iz čega proizlazi to da leksičke jedinice u tim odnosima nisu zrcalne, simetrične (kao što su primjerice dva antonima jedan drugomu antonimi, npr. *crn* – *bijel*). Točnije, leksičke jedinice koje su na višoj razini u hiponimijskoj hijerarhiji nazivaju se hiperonimi (npr. *životinja*), a jedinice na nižoj razini hiponimi (*pas*, *mačka*). Dva hiponima istoga hiperonima nazivaju se kohiponimi (*pas* – *mačka*). Leksička jedinica koja u meronimijskome odnosu označuje cjelinu naziva se holonim (npr. *tijelo*), a jedinica koja označuje dio meronim (*ruka*, *noga*). Dva meronima istoga holonima nazivaju se komeronimi (*ruka* – *noga*).

Hijerarhijske odnose tematiziraju još grčki filozofi, zatim srednjovjekovni mislioci, a pitanja poput djeljivosti ostaju važna i u daljnjim epohama filozofske misli (usp. Materynska 2019: 256). Odnosi su važni i u antropološkim radovima, računalnoj znanosti (primjerice u sustavima upita i odgovora), no dugo nisu bili opisivani u jezikoslovlju. Hiponimija i meronimija počinju se sustavnije proučavati 80-ih i 90-ih godina, i to najviše u psiholingvističkim radovima koji se bave značajkama paradigmi, npr. meronimijskih (Winston i dr. 1987, Iris i dr. 1988, Gerstl i Pribbenow 1996 i dr., usp. Materynska 2019: 256).

Naglasak na proučavanje leksičko-semantičkih odnosa u jezikoslovlju se stavlja od samih početaka strukturalizma, 30-ih godina 20. stoljeća (Geeraerts 1999: 126, Geeraerts 2010: 49–51), no sustavno se opisuju krajem strukturalizma, 60-ih godina (npr. Lyons 1963, Lyons 1977). Leksičko-semantički odnosi računalno su najdetaljnije i najsustavnije povezani u semantičkim mrežama odnosno računalnim tezaurusima, među kojima se osobito ističe *WordNet* (Miller 1990, 1995, Fellbaum 1998). Skupovi sinonima koji predstavljaju jedno značenje (sinseti) povezani su u različite odnose: antonimija, troponimija (među glagolima), uzročnost, logička nužnost te, vrlo zastupljeni i oprimjereni odnosi, hiponimija i meronimija. Pritom se odnosi dijele na podvrste, pa postoji meronimija u kojoj je dio (1) komponenta cjeline, (2) član cjeline ili (3) materijal cjeline (tvar od koje je cjelina načinjena) (Miller 1990:

256). Semantičke mreže razvijaju se i za južnoslavenske jezike, npr. *CroWn* za hrvatski, *SrpWn* za srpski, *BulNet* za bugarski, *SloWNet* za slovenski.

Među leksičko-semantičkim odnosima općenito više se proučavaju sinonimija i antonimija, o čemu svjedoči i broj objavljenih monografija. Tako je primjerice u hrvatskome jezikoslovlju objavljena monografija o sinonimiji (Petrović 2005) te rječnik sinonima (Šarić i Wittschen 2008), cjelovita monografija napisana je i o antonimiji (Šarić 2007), a obuhvatni radovi o hiponimiji i meronimiji u hrvatskome jeziku, a ni u južnoslavenskim jezicima općenito, nisu objavljeni.

Ovaj rad definira dva važna leksičko-semantička odnosa i daje uvid u mogućnosti poluautomatskoga izlučivanja primjera iz rječničkoga korpusa koji zatim mogu poslužiti za dodatne analize. Rad se oslanja na postavke distribucijske semantike, tj. polazi od hipoteze da se značenje riječi može prepoznati prema kontekstu. Točnije, glavno su metodološko polazište leksičko-sintaktički obrasci (tako nazvani u Hearst 1992) kao tipične strukture prema kojima se mogu prepoznati parovi hiperonima i hiponima te holonima i meronima. U sljedećemu poglavlju detaljnije će se definirati leksičko-sintaktički obrasci te će se dati pregled proučavanja obrazaca. Treće poglavlje donosi osnovne informacije o korpusu koji se analizira, a u četvrtome donose se obrasci za hiponimiju i meronimiju.

2. Leksičko-sintaktički obrasci

Najutjecajniji predstavnici leksičke semantike u svojim kapitalnim radovima ističu da ostvareni obrasci daju naznake određenoga odnosa (Lyons 1977: 311) odnosno da postoje *dijagnostički okviri* ili *dijagnostičke formule* (Cruse 1986: 12). Cruse navodi da dijagnostički okviri služe za prepoznavanje značenja na osnovi distribucije i oblika te ih uspoređuje s lakmusovim papirom – čovjek ne može odrediti alkalnost tekućine, ali može procijeniti je li lakmusov papir plav ili crven i na temelju toga donijeti zaključak. U jeziku tako postoje dijagnostički okviri koji upućuju na što, pojednostavnjuju intuiciju i lakše dovode do zaključaka (Cruse 1986: 10–11). Upravo su hijerarhijski odnosi poslužili kao prvi primjer odnosa u kojima postoje dijagnostički okviri, odnosno Cruse (1986: 13, također i 2000: 153) najprije navodi okvir *X i ostali Y* (engl. *X and other Y*) za hiponimiju i *X je dio Y-a* (engl. *X is a part of Y*) za meronimiju.

Sljedećih godina dolazi do zamaha u proučavanju okvira, pa valja istaknuti da Justeson i Katz (1991) istražuju okoline u kojima se pojavljuju antonimi, a postoje i brojna druga istraživanja sintaktičkih okruženja u kojima se mogu pojaviti antonimi (npr. Jones i Murphy 2005, Jones 2006, Jones 2010, v. i Čulig Suknaić 2020: 57–58).

Ipak, u ovome kontekstu najvažniji rad objavljuje Hearst (1992), koja razvija metodu prepoznavanja hiponima s pomoću čestih *leksičko-sintaktičkih obrazaca*. Upućuje na to da se preko sintagmatskoga ostvaraja može doprijeti do paradigmatiskih značenjskih odnosa, a time i do razumijevanja rječničkoga ustroja. Točnije, apstrahira rečenice na leksičko-sintaktičke obrasce – primjerice iz rečenice *The bow lute, such as Bambara ndang, is plucked and has (...)*. (hrv. *Lutnja za gudalo, kao što je bambara ndang, trzalačka je i ima (...)*.) izlučuje obrazac NP_0 *such as* $\{NP_1, NP_2 \dots (and/or)\} NP_n$ (Hearst 1992: 539). Dakle, iz rečenice i obrasca može se zaključiti da je *bambara ndang* hiponim gudalu čak i ako se ne zna što je *bambara ndang*. Leksičko-sintaktički obrazac pretpostavlja da na određeno mjesto dolazi hiponim i na određeno mjesto hiperonim, tj. da određene leksičke jedinice otvaraju mjesto hiponimima, hiperonimima ili kojoj drugoj jedinici, ovisno o strukturi obrasca. Sintaktički gledano, leksička jedinica može doći u točnome padežu, glagol može imati različitu rekciju i sl., što je osobito važno u flektivnim jezicima kao što je hrvatski i drugi južnoslavenski jezici. S obzirom na navedeno i na ustaljenost naziva *leksičko-sintaktički obrazac* – njime se koriste i brojni drugi kasniji autori, npr. Ando i dr. (2004), Snow i dr. (2005), Girju i dr. (2006), Barbu Mititelu (2011), Yıldız i dr. (2016), Sahin (2017), Farrugia (2019) – u ovome radu prednost se daje tomu nazivu.

Leksičko-sintaktički obrasci za hiponimiju (kao i za meronimiju) mogu se izlučivati iz različitih vrsta tekstova. Hearst (1992) se u svojem pionirskom radu služi Grolierovom američkom enciklopedijom, a nadalje primjerice Rydin (2002) analizira novinske vijesti na švedskome, Ando i dr. (2004) analiziraju vijesti na japanskome. Hiponimno-hiperonimni parovi analiziraju se i u kineskome (Liu i dr. 2006), nizozemskome (Sang i Hofmann 2007), rumunjskome (Barbu Mititelu 2011), litavskome (Bielinskienė i dr. 2012), postoje brojni radovi za turski (npr. Yıldırım i Yıldız 2012a, 2012b, Sahin i dr. 2016) i španjolski (Gil-Berrozpe 2017 i Gil-Berrozpe i dr. 2017). Radovi postoje i za slavenske jezike – npr. Przepiórkowski i dr. (2007) za bugarski, češki i poljski, Baisa i Suchomel (2015) za češki, Vintar i Gorjanc (2003) za slovenski. Stanković i dr. (2021) iz velikoga broja definicija izlučuju pravilnosti (vrste definicija te leksičke i sintaktičke značajke) te (polu)automatskom analizom dolaze i do leksičko-semantičkih odnosa među kojima su hiponimija i meronimija. Vrijedan je doprinos u tome radu klasifikacija različitih rječničkih definicija (Stanković i dr. 2021: 698). Kad je riječ o meronimiji, obrasce spominju već Winston i dr. (1987), Iris i dr. (1988), Berland i Charniak (1999), Girju i dr. (2006). Obrasci za prepoznavanje meronimije također se istražuju u brojnim jezicima, također mnogo za turski jezik (v. npr. Yıldız i

dr. 2016). Obrasci, dakako, ovise o jeziku i mogu biti vrlo raznoliki, o čemu svjedoče primjeri iz prikazani na Slici 1.

1	$NP_X PP_Y$ — PP_Y starts with <i>of</i>	173	door of his car the executive of the new government
2	$NP_X PP_Y$ — PP_Y starts with <i>in</i>	61	people in the world
3	$NP_X PP_Y$ — NP_X ends with <i>branch</i> — PP_Y begins with <i>of</i>	14	They organized the executive branch of government
4	$NP_X PP_Y$ — PP_Y starts with <i>from</i>	9	oxygen from air people from all over the world
5	$NP_X PP_Y$ — PP_Y starts with <i>throughout</i>	6	people throughout the world
6	$NP_X PP_Y$ — PP_Y starts with <i>at</i>	5	window at the rear of the building
7	$NP_X PP_Y$ — PP_Y starts with <i>on</i>	16	five fingers on one hand
8	$NP_X PP_Y$ — PP_Y starts with <i>to</i>	3	the door to the house
9	$NP_X PP_Y$ — PP_Y starts with <i>around</i>	2	people around the world
10	$NP_X PP_Y$ — PP_Y starts with <i>all over</i>	1	people all over the world

Slika 1. Neki leksičko-sintaktički obrasci za meronimiju u engleskome i primjeri za svaki obrazac (Girju i dr. 2006: 122)

Za izlučivanje leksičko-sintaktičkih obrazaca dobru podlogu čine rječnici, koji s jasnom strukturom definicija nude mogućnost za (polu)automatsko izlučivanje leksičkih jedinica koje su u hiponimijskome ili meronimijskome odnosu. Za to su osobito korisni mrežni rječnici s dostupnim definicijama, a jedan je od njih za hrvatski jezik *Hrvatski mrežni rječnik – Mrežnik*.

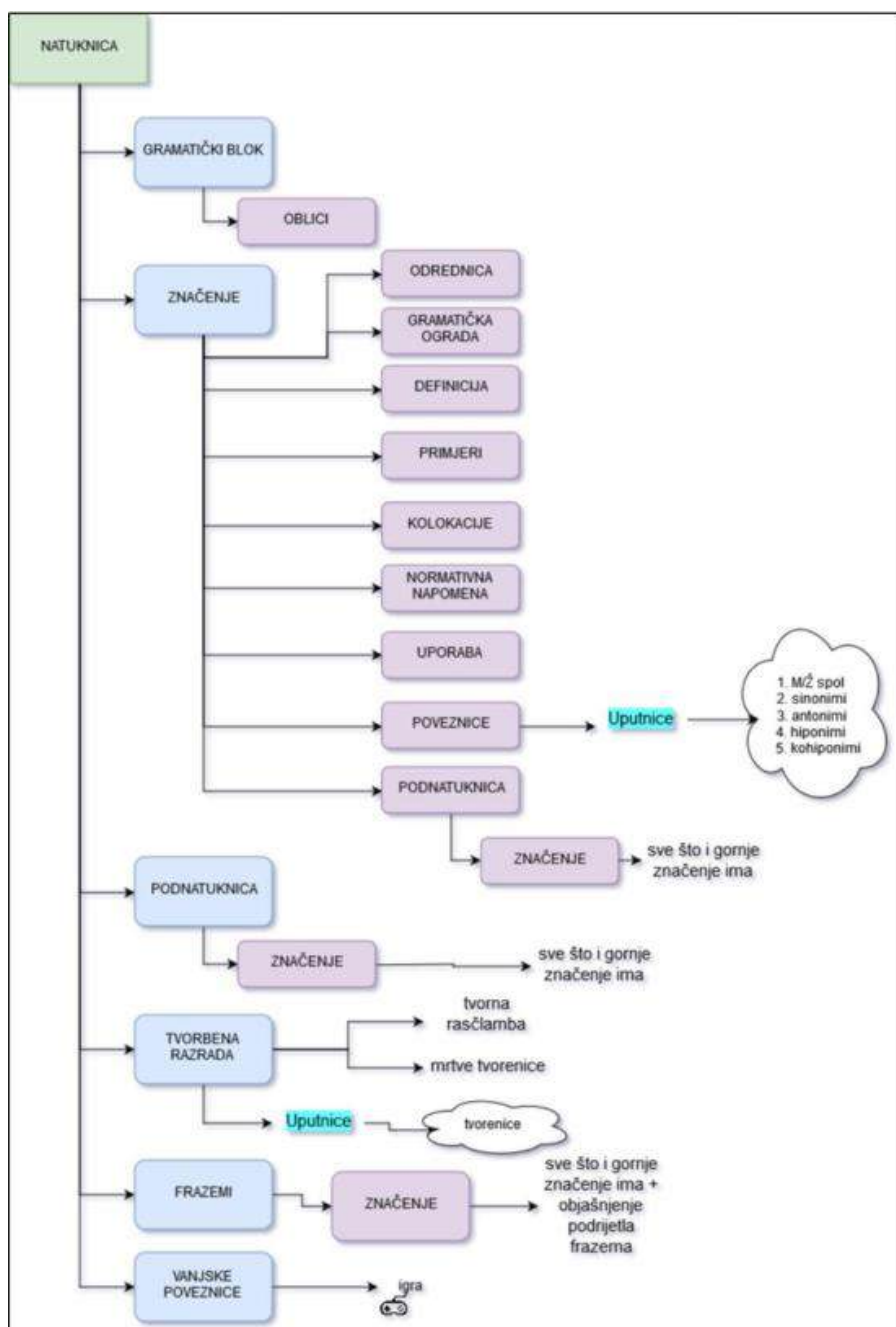
3. Korpus za istraživanje

Rječnička definicija kao jedan od glavnih dijelova rječničkoga članka sadržava podatke o leksičko-semantičkim odnosima, a Zgusta (1971: 253–254) navodi da su u njoj važni upravo hiperonim i potom značenjska razlika, ono što su u logičkoj definiciji *genus proximum* i *differentia specifica*. Osim u rječnicima važnost definicije ističe se u jeziku općenito – definicija (definiranje) osnovna je jezična aktivnost jer njome jezik opisuje sam sebe (Barnbrook 2002). Još se 80-ih godina prošloga stoljeća autori služe rječnicima kako bi došli do leksičko-semantičkih odnosa (npr. Amsler 1980, Markowitz i dr. 1986, Ahlswede i Evens 1988, Iris i dr. 1988), a u novije doba navode da je rječnik dobar korpus za takva istraživanja jer se (1) sastoji od definicija koje pružaju mnogo informacija,

(2) rječničke definicije pišu na isti način, tj. imaju istu strukturu te zbog toga jasne i česte sintaktičke obrasce (Šerbetći i dr. 2011: 13). Stanković i dr. (2021: 698), osim toga, ističu da modeli rječničkih definicija i njihove lokalne gramatike omogućuju povezivanje drugih leksičkih resursa (u njihovu slučaju za srpski jezik) te izdvajanje definicija iz strukovnih korpusa.

Radi lakšega pristupa podacima kao glavni korpus za ovo istraživanje poslužio je prvi mrežni rječnik hrvatskoga jezika: *Hrvatski mrežni rječnik – Mrežnik*.³ U vrijeme pisanja ovoga rada (u srpnju 2025. god.) *Mrežnik* sadržava više od 10 000 natuknica, i to svih vrsta riječi. Natuknica se u osnovnome modulu sastoji od gramatičkoga bloka, značenjskoga bloka, podnatuknica, tvorbenoga bloka, frazema, etimološke obrade i poveznica (više o *Mrežniku* v. u Hudeček i dr. 2024). Cijela struktura natuknice prikazana je na Slici 2.

³ *Hrvatski mrežni rječnik – Mrežnik* objavljuje se na stranici rjecnik.hr/mreznik i trenutačno su dostupne natuknice od A do K. Cijeli rječnik (koji se izrađuje u *TLexu*) dostupan je autorici ovoga rada kao suradnici na projektu.



Slika 2. Maksimalna struktura rječničkoga članka u *Mrežniku* (Hudeček i Mihaljević 2021: 22)

4. Dobiveni obrasci

U samoj strukturi rječničkoga članka naglasak je stavljen na rječničke definicije, koje su u prvoj etapi ovoga istraživanja izvezene u poseban dokument kako bi se mogle analizirati te su uočeni jasni obrasci za hiponimiju i meronimiju. Definicije su u najvećoj mjeri sortirane u Pythonu pronalaženjem definicija koje imaju jednaku treću ili četvrtu riječ, ovisno o strukturi definicije. Točnije, u prvome koraku sortirane su definicije prema trećoj riječi tako da se sve definicije koje imaju zajedničku treću riječ pojavljuju jedna ispod druge (tako su dobivene skupine kohiponima, tj. definicija kohiponima). Rezultati su potom pregledani ručno kako bi se provjerilo jesu li obuhvaćene i višerječne leksičke jedinice te su dobivene pravilnosti, tj. leksičko-sintaktički obrasci. Iako pojedini autori ističu da obrasci imaju slab odziv i nisku preciznost (Navigli i Velardi 2010), leksičko-sintaktičke obrasce ipak vrijedi istaknuti jer mogu poslužiti kao osnova za stvaranje razvijenijih modela – posebno obrasce za hrvatski jezik jer dosad nisu opisani.

4.1. Hiponimija

Najveći broj definicija imao je hiponimijsku strukturu, i to tako da se hiponim (podcrtan) izravno definira hiperonomom (podebljan), primjerice:

Agresivnost je **osobina** onoga koji silom djeluje protiv čije volje ili prava.

Bahatost je **osobina** onoga koji je bahat.

Besmrtnost je **osobina** onoga koji je besmrtan, koji živi vječno.

Akcelerator je **uređaj** za povećanje kinetičke energije električno nabijenih čestica.

Alkoholometar je **uređaj** s pomoću kojega se mjeri koncentracija alkohola u organizmu.

Ampermetar je **uređaj** za mjerenje jakosti električne struje.

Najviše je hiponima definirano hiperonimima žena, osobina, svojstvo, uređaj, naprava, mjesto, pripadnik, stanje, predmet, muškarac, znak, boja, prostorija, prostor, način, sustav, riječ, rečenica, razdoblje, zgrada, građevina itd. i hiponimija u takvim definicijama mogla bi se nazvati izravnom hiponimijom. U korpusu osim definicija takva tipa zamjetne su definicije u kojima se između hiponima i hiperonima

pojavljuje i dodatna riječ koja upućuje na to da je riječ o hiponimiji, svojevrsna spona između dviju jedinica koja govori u njihovu značenju i koja je stoga nazvana značenjska spona (usp. Matijević 2024). Kako navode još rani autori poput Crusea (npr. u Cruse 1986: 137), najprozirnija je značenjska spona za hiponimiju vrsta (engl. sort, kind). U Mrežniku se tom riječju u definicijama ustrojavaju brojne hiponimijske hijerarhije, primjerice:

Morseova abeceda vrsta je **telegrafske abecede** koja poruku prikazuje s pomoću točaka i crtica.

Akrostih je vrsta **pjesme** u kojoj početna slova stihova čitana odozgo nadolje daju (...).

Puni pansion vrsta je **ugostiteljske usluge** koja nudi noćenje i tri dnevna obroka.

Osim spone *vrsta* pojavljuju se i spone *rod*, *tip*, *red*, *razred*, *carstvo*, *skupina* – važne u biološkoj sistematici – koje upućuju na hiponimijski odnos. Osim tih spona i na druge se načine upućuje na hiponimiju. Na primjer, izraz *jedan od* također može upućivati na hiponimiju:

Množitelj je jedan od **faktora u računskoj operaciji množenja**, broj kojim se množi.

Svinja je jedan od **znakova kineskoga horoskopa**.

Iako se u radovima o hiponimiji najveći naglasak stavlja na imenice, u analizi je uočena i pravilnost u definicijama glagola:

Cijukati znači **glasati se** glasom koji podsjeća na ciju.

Hukati znači **glasati se** glasom koji podsjeća na hu-hu.

Tuliti znači **glasati se** dugačkim i jednoličnim zvukom.

Hodati znači **kretati se** pješice.

Šetati se znači **kretati se** pješice i polako bez određena cilja radi odmora i razonode.

Trčati znači **kretati se** skačući s noge na nogu.

Klizati se znači **kretati se** po ledu ili po skliskoj površini.

Pritom se kod glagola može uočiti da hiponim i hiperonim imaju jednak glagolski vid, da može postojati i više razina u glagolskoj hijerarhiji (*kretati se* > *hodati* > *šetati se*), ali da i analize glagolskih hiperonima mogu

biti korisne u podjeli glagola na semantičke skupine (navedeni glagoli glasanja i glagoli kretanja).

Za ostale vrste riječi i hiponimiju valja istaknuti da su primjerice brojevi definirani metajezlično, kao *brojevi*, kako su definirani i usklici, čestice i neki veznici te kratice:

Jedan je **broj** koji se obilježuje s 1, za jedan manji od broja dva.

Aha je **usklik** kojim se izriče iznenadno shvaćanje čega.

Npr. je **kratica** od *na primjer*, uvodi primjer.

Definiranje jedne vrste riječi drugom tematiziraju i Lyons (1977: 289–299) i Cruse (1986: 149) te odnos između leksičkih jedinica koje pripadaju različitim vrstama riječi nazivaju kvaziodnosima, tj. navedeni odnos (između npr. *usklik* > *aha*) kvazihiponimijom. U definicijama se mogu primijetiti i druge vrste kvazihiponimije koje autori navode: (1) neslaganje hiponima i hiperonima u broju (npr. *Voće su različiti jestivi, najčešće slatki, plodovi voćaka. – plodovi > voće*) i (2) nepodudaranje u strukturi leksičkih jedinica, kad je hiponim jednorječna leksička jedinica, a hiperonim višerječna ili obratno (npr. *Bacačica diska atletičarka je koja se natječe (...). – atletičarka > bacačica diska*).

U definicijama za hiponimiju važno je primijetiti i definicije u kojima se pojavljuje *istoimen*, među ostalima:

Pauk je pripadnik istoimenoga reda.

Smreka je pripadnica istoimene skupine.

Leksikografi tako upućuju na to da se značenje odnosi na drugo značenje u rječničkom članku i drže se načela da se natuknica ne ponavlja u definiciji, iz čega se zapravo zaključuje da je riječ o hiponimiji unutar jedne leksičke jedinice. Cruse (2000: 110) ističe da jedna riječ može imati općenitije i specifičnije, tj. uže značenje i taj odnos naziva autohiponimijom. U rječničkom korpusu hrvatskoga jezika uočeno je da se ona pojavljuje uglavnom među leksičkim jedinicama koje označuju pripadnike biljnoga i životinjskoga svijeta (npr. *mačke > mačka, smreke > smreka* i dr.).

4.2. Meronimija

U prethodnome poglavlju istaknuto je da se hiponimija često ostvaruje izravno, npr. (*Akcelerator je uređaj (...) – uređaj > akcelerator*), a za meronimiju je najprije važno istaknuti da se ne ostvaruje izravno, nego nužno uz značenjsku sponu. Najčešća je leksička jedinica koja upućuje na meronimiju *dio* (kako je istaknuto u brojnim radovima, npr. još u Winston i dr. 1987: 417–418, Iris i dr. 1988: 269). Najveći broj meronimijskih definicija u *Mrežniku* i ima upravo tu jedinicu (podcrtan meronim, podebljan holonim):

Bat je dio **zvona** koji udaranjem po njegovoj stijenci proizvodi zvuk.
Klip je dio **kukuruza** s malim žutim plodovima u obliku bobica.

Osim jedinice *dio* pojavljuju se jedinice element (*Žila je element lista i stabljike koji služi za (...)*), sastavnica (*Mikroprocesor je sastavnica silikonskoga čipa koja izvršava (...)*), odjel (*Prodaja je odjel u kojoj tvrtki ili djelatnosti koji (...)*) i druge, koje također upućuju na to da je riječ o odnosu dijela i cjeline. Te se jedinice mogu pojaviti i nakon izraza jedan od, koji je spomenut kod hiponimije, pa se pronalazi primjerice:

Struk je jedan od dijelova **sidra**.

Prst je jedan od krajnjih pokretljivih dijelova **šake** i **stopala**.

U definicijama za meronimiju važni su i glagoli, primjerice *sastojati se* (*Jezgra je dio atoma koji se sastoji od protona i neutrona. – atom > jezgra > proton, neutron*), *sadržavati*, *dijeliti se*, ali su za flektivne jezike važni i primjerice *padeži*, tj. sintagmatska razina. Točnije, odnos dijela i cjeline može se izricati pridjevom i imenicom tako da je pridjevom označena cjelina (*rečenični dio – rečenica > dio*) ili dio (sastojak) (*čokoladni kolač – kolač > čokolada*), genitivom *s od* (*ogrtač od balonske svile – ogrtač > balonska svila*), besprijedložnim genitivom (*organ ženskoga tijela – žensko tijelo > organ*), instrumentalom *sa s* (*kolač s nadjevom – kolač > nadjev*), usp. Matijević 2024: 161.

Za meronimiju među glagolima pronađeno je znatno manje primjera nego za hiponimiju, no dva glagola koji stoje kao meronimi (*genus* u definiciji) mogu upućivati na to da radnja koja se definira (natuknica) ima podradnje:

Čitati znači raspoznavati i povezivati slova ili znakove te razumjeti ono što je napisano.

Pušiti znači uvlačiti i ispuštati duhanski dim.

Trepnuti znači nakratko zatvoriti i otvoriti očne kapke.

Među definicijama uočena je kvazimeronimija, najčešće u tome što meronim može biti višerječna leksička jedinica, a holonim jednorječna ili obratno (npr. *Vilica je dio pribora za jelo kojim se nabada hrana. – pribor za jelo > vilica*). Jedinicom *istoimen*, spomenutom kod hiponimije, također se može izraziti da jedna leksička jedinica ima značenje koje označuje cjelinu i jedno koje označuje dio, primjerice:

Badem je jestiv plod gorkaste jezgre istoimene biljke.

Celer je zeleno lišće istoimene biljke koje se upotrebljava kao začim.

Cikla je tamnocrveni jestivi korijen istoimene biljke.

5. Zaključak

Hiponimija i meronimija kao leksičko-semantički odnosi koji ustrojavaju leksičke hijerarhije vrlo su važni za razumijevanje organizacije rječničkoga sustava i mehanizama ljudske kategorizacije. Rad prikazuje kako leksičko-sintaktički obrasci mogu biti dobar putokaz za prikupljanje primjera hiperonima i hiponima te holonima i meronima, koji zatim mogu poslužiti za različita istraživanja, primjerice razradu taksonomije odnosa, uočavanje granica i dodira tih odnosa te za izgradnju računalnih resursa i alata.

Rječnik kao korpus koji se sastoji od strukturiranih i usklađenih definicija može biti dobro polazište za prepoznavanje obrazaca, osobito ako je riječ o računalno čitljivu i dostupnu rječniku koji se može lako pretraživati. Iz *Hrvatskoga mrežnog rječnika – Mrežnika* izvezene su definicije te su uočeni najčešći leksičko-sintaktički obrasci u definicijama. Za hiponimiju su to *HIPONIM je HIPERONIM* (izravna hiponimija) te *HIPONIM je vrsta/rod/tip/red/razred/carstvo/skupina HIPERONIMA*, *HIPONIM je jedan od HIPERONIMA* (gdje postoje značenjske spona između hiponima i hiperonima). Za meronimiju najčešći je obrazac *MERONIM je dio HOLONIMA*, a pojavljuju se i *MERONIM je element/sastavnica/odjel HOLONIMA* ili *MERONIM je jedan od dijelova HOLONIMA*.

Pretraživanje drugih vrsta tekstova, primjerice općih i stručnih korpusa, može dati i nove obrasce – za hiponimiju primjerice *vrsta/tip HIPERONIMA: HIPONIM, (HIPONIM, HIPONIM...), HIPERONIM (HIPONIM,*

HIPONIM...), neki od HIPERONIM je(su) / kao što su: HIPONIM, HIPONIM, HIPONIM (...), HIPERONIM poput HIPONIM i dr. – no detaljnijim proširivanjem analize ovaj bi rad premašio zadani opseg.

Zaključno, istraživanje i dobiveni rezultati (pa tako i primjeri kvaziodnosa i autooodnosa) pokazuju da su rječnici, zahvaljujući strukturi i sustavnosti, plodno tlo za različita jezična istraživanja.

Захвалница: Ovaj rad napisan je na istraživačkome projektu *Hrvatski mrežni rječnik (Mrežnik): 2. faza*, koji financira Europska unija – NextGenerationEU. Za iznesene stavove i mišljenja odgovorna je samo autorica te ti stavovi ne odražavaju službena stajališta Europske unije ili Europske komisije. Ni Europska unija ni Europska komisija ne mogu se smatrati odgovornima za njih.

Literatura

- [1] Ahlswede, T., & Evens, M. (1988). Parsing vs. text processing in the analysis of dictionary definitions. U: *Proceedings of the 26th Annual Meeting of the Association for Computational Linguistics* (pp. 217–224). Association for Computational Linguistics. <https://doi.org/10.3115/982023.982050>
- [2] Amsler, R. (1980). *The structure of the Merriam-Webster pocket dictionary*. University of Texas.
- [3] Ando, M., Sekine, S., & Ishizaki, S. (2004). Automatic extraction of hyponyms from Japanese newspapers using lexico-syntactic patterns. U: M. T. Lino, M. F. Xavier, F. Ferreira, R. Costa, & R. Silva (ur.), *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)* (pp. 387–390). European Language Resources Association.
- [4] Baisa, V., & Suchomel, V. (2015). Corpus-based extraction of hypernyms in terminological thesaurus for land surveying domain. U: *Ninth Workshop on Recent Advances in Slavonic Natural Language Processing* (pp. 69–74). Tribun EU.
- [5] Barbu Mititelu, V. (2011). Hyponymy patterns in Romanian. *Memoirs of the Scientific Sections of the Romanian Academy*, 34, 1–13.
- [6] Barnbrook, G. (2002). *Defining language: A local grammar of definition sentences*. *Studies in Corpus Linguistics*. John Benjamins Publishing Company.
- [7] Berland, M., & Charniak, E. (1999). Finding parts in very large corpora. U: *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL 1999)* (pp. 57–64).
- [8] Bielinskienė, A., Boizou, L., Kovalevskaitė, J., & Utka, A. (2012). Towards the automatic extraction of term-defining contexts in Lithuanian. U: A. Tavast, K.

- Muischnek, & M. Koit (ur.), *Human language technologies: The Baltic perspective* (pp. 18–26). IOS Press.
- [9] Cruse, D. A. (1986). *Lexical semantics*. Cambridge University Press.
- [10] Cruse, D. A. (2000). *Meaning in language: An introduction to semantics and pragmatics*. Oxford University Press.
- [11] Čulig Suknaić, J. (2020). *Antonimija kao pojmovna kategorija značenjske suprotnosti u engleskome i hrvatskome jeziku*. University of Zagreb, Faculty of Humanities and Social Sciences. <https://doi.org/10.17234/diss.2020.201018>
- [12] Farrugia, J. P. (2019). *Exploring hypernym discovery*. University of Malta, Department of Artificial Intelligence.
- [13] Fellbaum, C. (1998). *WordNet – An electronic lexical database*. MIT Press.
- [14] Geeraerts, D. (1999). Hundred years of lexical semantics. U: M. Vilela & F. Silva (ur.), *Atas do 1º Encontro internacional de linguística cognitiva* (pp. 123–154). Faculdade de Letras.
- [15] Geeraerts, D. (2010). *Theories of lexical semantics*. Oxford University Press.
- [16] Gerstl, P., & Pribbenow, S. (1996). A conceptual theory of part-whole relations and its applications. *Data & Knowledge Engineering*, 20(3), 305–322.
- [17] Gil-Berrozpe, J. C. (2017). *Corpus-based identification of hyponymy subtypes and knowledge patterns in the environmental domain*. University of Granada, Faculty of Translation and Interpreting.
- [18] Gil-Berrozpe, J. C., León-Araúz, P., & Faber, P. (2017). Specifying hyponymy subtypes and knowledge patterns: A corpus-based study. U: I. Kosem, C. Tiberius, M. Jakubiček, J. Kallas, S. Krek, & V. Baisa (ur.), *Electronic lexicography in the 21st century: Proceedings of eLex 2017 conference* (pp. 63–92). Lexical Computing CZ s.r.o.
- [19] Girju, R., Badulescu, A., & Moldovan, D. (2006). Automatic discovery of part-whole relations. *Computational Linguistics*, 32(1), 83–135.
- [20] Hearst, M. (1992). Acquisition of hyponyms from large text corpora. U: *Proceedings of the 14th International Conference on Computational Linguistics (COLING-92)* (pp. 539–545). Nantes.
- [21] Hudeček, L., & Mihaljević, M. (2021). *Hrvatski mrežni rječnik – MREŽNIK: Monografija (radna inačica)*. Institut za hrvatski jezik i jezikoslovlje.
- [22] Hudeček, L., Mihaljević, M., & Jozić, Ž. (ur.). (2024). *Anatomija rječnika: Hrvatski mrežni rječnik – Mrežnik*. Institut za hrvatski jezik.
- [23] Iris, M. A., Litowitz, B. E., & Evens, M. (1988). Problems of the part-whole relation. U M. W. Evens (ur.), *Relational models of the lexicon: Representing knowledge in semantic networks* (pp. 261–288). Cambridge University Press.

- [24] Jones, S. (2006). A lexico-syntactic analysis of antonym co-occurrence in spoken English. *Text & Talk*, 26(2), 191–216.
- [25] Jones, S. (2010). Using web data to explore lexico-semantic relations. U: P. Starjohann (ur.), *Lexical-semantic relations: Theoretical and practical perspectives*. John Benjamins.
- [26] Jones, S., & Murphy, L. M. (2005). Using corpora to investigate antonym acquisition. *International Journal of Corpus Linguistics*, 10(3), 401–422.
- [27] Justeson, J. S., & Katz, S. M. (1991). Co-occurrences of antonymous adjectives and their contexts. *Computational Linguistics*, 17(1), 1–19.
- [28] Liu, L., Cao, C., & Wang, H. (2006). Extracting hyponymic relations from Chinese free corpus. U: *Proceedings of the 5th WSEAS International Conference on Applied Computer Science (ACOS'06)* (pp. 962–968). WSEAS.
- [29] Lyons, J. (1963). *Structural semantics: An analysis of part of the vocabulary of Plato*. Basil Blackwell.
- [30] Lyons, J. (1977). *Semantics* (Vols. 1–2). Cambridge University Press.
- [31] Markowitz, J., Ahlswede, T., & Evens, M. (1986). Semantically significant patterns in dictionary definitions. U: *Proceedings of the 24th Annual Conference of the ACL* (pp. 112–119).
- [32] Materynska, O. (2019). Semantic relation of meronymy in languages of different structure (case study of semantics of body part names). *Advanced Education*, 12, 256–264.
- [33] Matijević, M. (2024). *Hijerarhijski leksičko-semantički odnosi u hrvatskome jeziku*. Filozofski fakultet Sveučilišta u Zagrebu.
<https://doi.org/10.17234/diss.2025.314038>
- [34] Miller, G. A. (1990). Nouns in WordNet: A lexical inheritance system. *International Journal of Lexicography*, 3(4), 246–264.
- [35] Navigli, R., & Velardi, P. (2010). Learning word-class lattices for definition and hypernym extraction. U: *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics* (pp. 1318–1327).
- [36] Petrović, B. (2005). *Sinonimija i sinonimičnost u hrvatskome jeziku*. Hrvatska sveučilišna naklada.
- [37] Przepiórkowski, A., Degórski, Ł., Spousta, M., Simov, K., Osenova, P., Lemnitzer, L., Kuboň, V., & Wójtowicz, B. (2007). Towards the automatic extraction of definitions in Slavic. U: *Proceedings of the Workshop on Balto-Slavonic Natural Language Processing 2007: Special theme – Information extraction and enabling technologies* (pp. 43–50). Association for Computational Linguistics.

- [38] Rydin, S. (2002). Building a hyponymy lexicon with hierarchical structure. U: *Proceedings of the ACL-02 Workshop on Unsupervised Lexical Acquisition* (pp. 26–33).
- [39] Sahin, G. (2017). Extraction of hyponymy, meronymy, and antonymy relation pairs: A brief survey. *International Journal on Natural Language Computing*, 6(2), 1–18.
- [40] Sang, E. T. K., & Hofmann, K. (2007). Automatic extraction of Dutch hypernym-hyponym pairs. *LOT Occasional Series*, 7, 163–174.
- [41] Şerbetçi, A., Orhan, Z., & Pehlivan, İ. (2011). Extraction of semantic word relations in Turkish from dictionary definitions. U: *Proceedings of the ACL 2011 Workshop on Relational Models of Semantics* (pp. 11–18). Association for Computational Linguistics.
- [42] Snow, R., Jurafsky, D., & Ng, A. (2005). Learning syntactic patterns for automatic hypernym discovery. U: *Advances in neural information processing systems* (pp. 1297–1304).
- [43] Stanković, R., Krstev, C., Stijović, R., Gočanin, M., & Škorić, M. (2021). Towards automatic definition extraction for Serbian. U: L. Mitits & S. Kiosses (ur.), *Proceedings of the XIX EURALEX Congress: Lexicography for Inclusion* (Vol. 2, pp. 695–703). Democritus University of Thrace.
- [44] Šarić, Lj. (2007). *Antonimija u hrvatskome jeziku: Semantički, tvorbeni i sintaktički opis*. Hrvatska sveučilišna naklada.
- [45] Šarić, Lj., & Wittschen, W. (2008). *Rječnik sinonima hrvatskoga jezika*. Jesenski i Turk.
- [46] Vintar, Š., & Gorjanc, V. (2003). Identifying markers of semantic relations in Slovene. *Strani jezici*, 32(1–2), 37–44.
- [47] Winston, M. E., Chaffin, R., & Herrmann, D. (1987). A taxonomy of part-whole relations. *Cognitive Science*, 11(4), 417–444.
- [48] Yıldırım, S., & Yıldız, T. (2012a). Corpus-driven hyponym acquisition for Turkish language. U: A. Gebukh (ur.), *Computational linguistics and intelligent text processing (CICLing 2012)* (pp. 29–41). Springer.
- [49] Yıldırım, S., & Yıldız, T. (2012b). Automatic extraction of Turkish hypernym-hyponym pairs from large corpus. U: *Proceedings of the International Conference on Computational Linguistics (COLING 2012)* (pp. 493–500).
- [50] Yıldız, T., Diri, B., & Yıldırım, S. (2016). Acquisition of Turkish meronym based on classification of patterns. *Pattern Analysis and Applications*, 19(2), 495–507.
- [51] Zgusta, L. (1971). *Manual of lexicography*. De Gruyter Mouton.

From a Dictionary as Corpus to Hyponymy and Meronymy

Maja Matijević

Summary

This study investigates two hierarchical lexical-semantic relations, hyponymy and meronymy, and their semi-automatic identification in dictionary definitions. Hyponymy is defined as the relation between a lexical unit denoting a category and a lexical unit denoting a member of that category (e.g., hyperonym *animal* > hyponym *dog*), while meronymy is defined as the relation between a lexical unit denoting a whole and a lexical unit denoting a part of that whole (e.g., holonym *body* > meronym *hand*). The study establishes a theoretical framework and demonstrates the potential of automated identification in dictionary definitions. The methodological approach is based on lexico-syntactic patterns, i.e., typical structures characterized by specific syntactic regularities and lexical units that enable the recognition of lexical-semantic relations. The analysis of dictionary definitions yielded several patterns, with particular emphasis on case forms and prepositional expressions relevant for expressing part-whole relations. The results highlight dictionaries as a rich and reliable source for the recognition of hierarchical relations, opening new perspectives for further linguistic research and analysis.

Keywords: lexico-semantic relations, hyponymy, meronymy, lexico-syntactic patterns, dictionary