

## KAKO DOBER JE CHATGPT PRI UMEŠČANJU SOPOMENK POD BESEDNE POMENE

Magdalena GAPSA,<sup>1</sup> Špela ARHAR HOLDT,<sup>1,2</sup> Iztok KOSEM<sup>1,2</sup>

<sup>1</sup>Filozofska fakulteta, Univerza v Ljubljani

<sup>2</sup>Fakulteta za računalništvo in informatiko, Univerza v Ljubljani

V raziskavi preverjamo, kako dobro se ChatGPT-4 odreže pri čiščenju seznama strojno pridobljenih sopomenskih kandidatov in umeščanju sopomenskega gradiva pod besedne pomene. Kot zlati standard upoštevamo slovaropisne odločitve, ki so bile sprejete pri nadgrajevanju Slovarja sopomenk sodobne slovenščine v različico 2.0. V prispevku analiziramo rezultate za 246 slovarskih iztočnic. Za 41,9 % iztočnic je ChatGPT podatke uredil povsem enako kot slovaropisci, za 58,1 % pa se je v odločitvi razlikoval: 43,5 % iztočnic je vsebovalo razlike pri odstranjevanju neustreznih sopomenskih kandidatov, 28,9 % pa pri razvrščanju sopomenk pod pomene. Pri presojanju relevantnosti sopomenskih kandidatov je bil ChatGPT popustljivejši od zlatega standarda (priklic 0,33), medtem ko je bila natančnost višja (0,75), vendar razlike težje pojasnljive. Razlike v razvrščanju sopomenk (umestitev pod drug pomen pri 14,6 % iztočnicah, manjkajoča umestitev pri 19,9 %) deloma pripisujemo značilnostim vhodnih podatkov, kot sta kompleksnost naloge in kratkost pomenskih indikatorjev. Bodoče delo bo usmerjeno v preizkus implementacije strojnega postopka za pohitritev slovaropisnega dela.

**Ključne besede:** digitalno slovaropisje, ChatGPT, sopomenke, besedni pomen, slovenščina

### 1. UVOD

Generativna umetna inteligenca, ki temelji na velikih jezikovnih modelih, je prek klepetalnih vmesnikov, kakršen je ChatGPT (OpenAI, 2024), postala široko dostopna za številne z jezikom povezane naloge. Med področji, ki zadnji dve leti preizkušajo moč in omejitve novih tehnologij, je tudi slovaropisje.

Dosedanji preizkusi rabe ChataGPT za slovaropisne namene se osredotočajo na generiranje bolj ali manj celostnih slovarskih gesel za (pogosto dokaj priložnostno) izbran nabor iztočnic (prim. de Schryver in Joffe, 2023; Barrett, 2023 v de Schryver, 2023; Rundell, 2023; Jakubíček in Rundell, 2023; Lew,

2023). Kot v svojem kritičnem pregledu prvih prispevkov (člankov in posnetkov predavanj oz. referatov) na temo z umetno inteligenco podprtega slovaropisja poroča de Schryver (2023), je trenutno največ pozornosti posvečene definicijam in primerom rabe (de Schryver, 2023, str. 17), manj pa preprostejšim slovaropisnim nalogam. (Skoraj) vse študije oz. preizkusi so bili izvedeni v angleščini in za angleščino, čeprav Jakubiček in Rundell (2023) naslavljata tudi problem večjezičnosti (str. 522–523).

Obstoječim raziskavam dodajamo preizkus, kako dobro se ChatGPT-4 odreže pri čiščenju seznama strojno pridobljenih sopomenskih kandidatov in umeščanju sopomenskega gradiva pod besedne pomene. Preizkus se povezuje z nadgrajevanjem Slovarja sopomenk sodobne slovenščine, velike zbirke slovenskih sopomenk, ki je bila v prvem koraku pripravljena povsem strojno iz podatkov Velikega angleško-slovenskega slovarja Oxford®-DZS in referenčnega korpusa Gigafida (Krek in sod., 2017), od objave leta 2018 pa se ročno pregleduje in čisti v sodelovanju med strokovnjaki za slovaropisje ter zainteresirano uporabniško javnostjo (Arhar Holdt in sod., 2018).

Ob nadgradnji slovarja v različico 2.0 je bilo med slovaropisnimi nalogami tudi umeščanje strojno pripravljenih sopomenskih kandidatov pod besedne pomene. Cilj naloge je bil: (a) presoditi, ali je strojno pridobljena beseda oz. zveza sopomenska z iztočnico v vsaj enem od njenih slovarskih pomenov ali pa gre za neustrezno gradivo, ki ga je treba iz iztočnice odstraniti, ter (b) umestiti relevantno sopomensko gradivo pod vse ustrezne besedne pomene dane iztočnice (več v Arhar Holdt in sod., 2023; Gantar in sod., 2023).

V bodoče bi bilo v opisani slovaropisni postopek možno vključiti dodatno strojno predprocesiranje podatkov s pomočjo programa ChatGPT. Ta bi podatke uredil na način, primerljiv slovaropisnemu, čemur bi sledil končni ročni pregled. Uspešna integracija bi lahko pomembno pohitrila nadgrajevanje slovarja. Prvi korak pri morebitni optimizaciji postopka je ugotoviti, kakšne rezultate daje ChatGPT v primerjavi z jezikoslovci, čemur se posvečamo v tem prispevku. V nadaljevanju predstavimo metodologijo raziskave in primerjamo rezultate strojnega in ročnega dela.

## 2. METODOLOGIJA

### 2.1 Opis podatkov

Preizkus temelji na delu podatkovnega vzorca za doktorsko raziskavo Sopomenskost v Slovarju sopomenk sodobne slovenščine in izbranih različicah wordneta, tj. seznamu 546 samostalnikov, ki se kot iztočnice pojavijo v podatkovni bazi Slovarja sopomenk sodobne slovenščine 1.0 (SSSS 1.0, Krek in sod., 2018) in drugih prosto dostopnih leksikalnih virih (prim. Gapsa, 2022). Ta nabor je bil omejen na 266 iztočnic, ki so bile ob posodobitvi SSSS 1.0 v verzijo 2.0 slovaropisno urejene, kar pomeni, da imajo v verziji 2.0 pripisano pomensko členitev, strojno pridobljeni sopomenski kandidati iz verzije 1.0 pa so bili ročno pregledani, potrjeni (oz. odstranjeni) in razvrščeni pod identificirane pomene.

Za izbranih 266 iztočnic je bilo v prvem koraku iz baze SSSS 1.0 izluščenih skupno 1.049 sopomenskih kandidatov (z morebitnimi področnimi slovarskimi oznakami). V drugem koraku so bile iz Digitalne slovarske baze (DSB, Kosem in sod., 2021) izvožene pomenske členitve s pomenskimi indikatorji (tj. kratkimi opisi za ločevanje pomenov) za izbrane iztočnice. Podatki so bili pretvorjeni v tabelo, kjer je posamezna vrstica vsebovala izvožene podatke po vzoru: iztočnica – pomenska členitev – sopomenski kandidati. Tabela je služila kot nabor vhodnih podatkov za preizkus s sistemom ChatGPT. Za preverbo uspešnosti naloge smo iz baze Slovarja sopomenk sodobne slovenščine 2.0 (SSSS 2.0, Krek in sod., 2023) pridobili pomensko členjene iztočnice z razvrščenimi sopomenkami.

V prvem koraku analize je bilo med 266 iztočnicami odkritih 20 iztočnic, kjer se pomenska členitev iz DSB ne ujema s SSSS 2.0 (npr. iztočnica *bonbon* ima v DSB en pomen, v SSSS 2.0 sta dva). Ti primeri so posledica dejstva, da se DSB dinamično razvija s podatki iz različnih virov, in so bili za ohranitev koherentnega zlatega standarda odstranjeni iz nadaljnje analize.

### 2.2 Struktura poziva za ChatGPT

Za izbrane iztočnice smo pripravili poziv za ChatGPT, pri čemer smo uporabili API model GPT-4. Poziv je bil pripravljen v angleščini in je bil med razvojem postopka večkrat testiran z uporabo brezplačne verzije sistema.

Med testiranjem se je izkazalo, da ChatGPT vrne boljše rezultate, če je v poziv vključen primer želenega rezultata. Posledično smo v poziv dodali primer vhodnih podatkov, tj. večpomensko iztočnico s sopomenskimi kandidati, in zelene izhodne podatke, tj. pravilno razporejene sopomenske kandidate po pomenih.

You are a lexicographer preparing a comprehensive language resource. You work in the Slovenian language. You should respond in the Slovenian language and only provide output that is relevant and valid for Slovenian.

You are given a word with its various meanings and its synonyms. Assign given synonyms to suitable meaning from one of the provided ones.

Not all meanings have synonyms. One synonym can suit multiple meanings, you can assign it to more than one meaning. You can discard synonyms that are not suitable for any of the meanings.

Each prompt represents a word and its synonyms and should be treated as unit. You will provide a response for each unit.

Unit "argument" looks like this:

word: argument

meanings:

1. utemeljen razlog
2. neodvisna spremenljivka

synonyms: dokazni razlog, neodvisna spremenljivka [matematika],<sup>1</sup> razlaga, utemeljitev, razmišljanje, smerni kot, udeleženska vloga [jezikoslovje]

Your response should look like this in a valid YAML format:

word: argument

meanings:

- id: 1

meaning: utemeljen razlog

synonyms:

- dokazni razlog
- utemeljitev
- razlog

---

<sup>1</sup> V oglatih oklepajih navajamo področne slovarske oznake.

- id: 2

meaning: neodvisna spremenljivka

synonyms:

- neodvisna spremenljivka [matematika]

discarded\_synonyms:

- razlaga

- razmišljanje

- smerni kot

- udeleženska vloga [jezikoslovje]

Only answer in YAML format. Write nothing else.

Izbrani primer za poziv je samostalnik *adaptacija*, ki prinaša več pomenov, slovarske oznake in primer odstranjenih (ang. *discarded*) sopomenskih kandidatov.

Here is unit “adaptacija”:

word: adaptacija

meanings:

1. prenova stavbe
2. priredba [v umetnosti]
3. prilagoditev [ponavadi v športu]

synonyms: predelava [glasba], priredba [glasba], preureditev, prilagoditev, prenova, aranžma [glasba]

word: adaptacija

meanings:

- id: 1

meaning: prenova stavbe

synonyms:

- preureditev

- prenova

- id: 2

meaning: priredba [v umetnosti]

synonyms:

- priredba [glasba]

- aranžma [glasba]

- id: 3

meaning: prilagoditev [ponavadi v športu]

synonyms:

- prilagoditev

discarded\_synonyms:

- predelava [glasba]

Odgovori so bili vrnjeni v formatu YAML, sledila je pretvorba v format JSON. Na podlagi teh podatkov smo za raziskovalne analize in evalvacijo ustvarili še povzemačno CSV datoteko in Excelovo datoteko z vsemi zbranimi podatki.

V poziv nismo vključili celotnih smernic, ki jim je sledila slovaropisna ekipa, saj bi s tem v postopek vnesli preveč informacij in spremenljivk, kar privede do neuporabnih in težje razločljivih rezultatov. Prav tako v poziv nismo vključili možnosti dodajanja ali spreminjanja besednih pomenov, ki jih je imela slovaropisna ekipa, saj smo želeli, da pomenska členitev ostane metodološko transparentna, rezultati pa dovolj enoznačni za analizo. Testiranja so pokazala optimalno delovanje poziva, ki je izvleček najpomembnejših navodil. Navodila, ki jih nismo vključili v poziv, navajamo ob analizi rezultatov, kadar olajšajo interpretacijo razlik med ročnim in strojnim delom.

### **2.3 Postopek analize gradiva**

Pridobljeni podatki so bili organizirani v preglednice. Strojno pripravljene rezultate smo primerjali s slovaropisnimi rešitvami in najprej ugotovili, katere iztočnice so obravnavane povsem enako in katere vsebujejo razlike. Razlike smo nato natančneje analizirali v dveh korakih: (a) katere vrste odstopanja se pojavljajo pri odstranjevanju neustreznih sopomenskih kandidatov in kako pogosto in (b) katere vrste odstopanja se pojavljajo pri umeščanju neodstranjenega gradiva pod besedne pomene in kako pogosto.

V raziskavi rešitve slovaropisne ekipe obravnavamo kot zlati standard, kar pomeni, da odstopa načeloma razumemo kot neželene. Rezultati sicer nakažejo, da je v določenih primerih rešitev, ki jo ponudi ChatGPT, drugačna od slovaropisne, vendar kljub temu sprejemljiva. Če bodo s ChatomGPT pripravljene podatki vključeni v slovaropisne delotoke, bo v prihodnje treba presoditi, kako v praksi obravnavati take primere skladno z izbranim slovaropisnim konceptom.

## 2.4 Omejitve

Strojni postopek, ki ga preizkušamo, je odvisen od izbranega poziva, vhodnih podatkov in različice uporabljenega sistema. Pri delu s sistemom ChatGPT oblikovanje ubeseditve pomembno vpliva na rezultate, pri čemer vpliva ni mogoče povsem predvideti ali pojasniti. Vpliva lahko tudi, da smo oblikovanje poziva testirali s pomočjo brezplačne, ne pa tudi plačljive različice sistema. V nadaljnjem delu je mogoče poziv nadalje razvijati ali postopek razdeliti na dva dela (ločeno čiščenje in umeščanje).

Pomenski indikatorji so krajši in jedrnatejši od polnih pomenskih definicij, oblikovani tako, da izražajo glavne razlike med pomeni, ki so za sistem morda premalo povedne. Predvidevamo, da bi lahko raba daljših, sistematično strukturiranih slovarskih definicij rezultate izboljšala.

## 3. REZULTATI IN DISKUSIJA

### 3.1 Splošna uspešnost

Pri analiziranih 246 iztočnicah je ChatGPT v 103 primerih (41,9 %) podatke uredil povsem enako kot slovaropisci, v 143 primerih (58,1 %) pa se je v odločitvi tako ali drugače razlikoval.

Podatke s primeri iztočnic prikazuje Tabela 1, v kateri podajamo tudi povprečno število kandidatov ter slovarskih pomenov v posamezni od skupin. V skupini ustrezno urejenih sopomenskih podatkov sta obe povprečji nižji, kar je skladno s pričakovanji, saj se s številom sopomenk za razvrstitev in številom besednih pomenov viša možnost za razlike v odločitvah. Povezava ni povsem enoznačna, saj se ChatGPT (lahko) razlikuje tudi pri iztočnicah z malo pomeni in sopomenkami in uspešno uredi kompleksnejše iztočnice.

Tabela 1: Ujemanje med slovaropisnimi odločitvami in odločitvami ChataGPT s številom iztočnic, primeri in povprečnim številom sopomenskih kandidatov ter besednih pomenov na skupino.

<i>Vrsta rezultata</i>	<i>Primeri</i>	<i>Št. iztočnic</i>	<i>Povpr. št. kandidatov</i>	<i>Povpr. št. pomenov</i>
Strojni rezultat enak ročnemu	adolescenca, aerodinamika, agonija, alkohol, ambicija,	103	2,2	1,7

	anatomija			
Strojni rezultat drugačen od ročnega	adaptacija, anonimnost, aplikacija, arbiter, arhitektura, arhiv	143	5,1	2,4
Skupaj analiziranih		246	3,9 (vseh kandidatov: 951)	2,1 (vseh pomenov: 516)

Natančnejša analiza je pokazala, da se med 143 iztočnicami pojavlja 107 takih, ki kažejo razlike na ravni odstranjevanja neustreznih sopomenskih kandidatov (43,5 % analiziranih iztočnic), 71 takih, ki kažejo razlike na ravni razvrščanja pod pomene (28,9 %), od tega pa je 35 primerov (14,2 %), kjer se pojavljajo tako razlike prvega kot drugega tipa.

### 3.2 Razlike v odstranjevanju neustreznih sopomenskih kandidatov

Prva naloga za ChatGPT je bila odstraniti sopomenske kandidate, ki ne sodijo pod nobenega od pomenov izbrane iztočnice. V zlatem standardu je bilo na ta način odstranjenih 249 od 951 (26,2 %) kandidatov. ChatGPT je odstranil le 110 kandidatov (11,6 %). Rezultati so prikazani v Tabeli 2, kjer so navedeni primeri, ki jih je ChatGPT glede na zlati standard ustrezno obdržal (*true negatives*, *TN*), ustrezno odstranil (*true positives*, *TP*), neustrezno obdržal (*false negatives*, *FN*) ali neustrezno odstranil (*false positives*, *FP*). V tabeli je najprej navedena iztočnica, nato sopomenski kandidat, o katerem je ChatGPT presojal.

Tabela 2: Primeri in število pravih in napačnih odločitev pri presojanju ChataGPT, ali je sopomenski kandidat ustrezen za dano iztočnico ter pomen ali ne.

	Primeri	Vsota
Ustrezno obdržanih (TN)	adaptacija – preureditev, adolescenca – odraščanje, aerodinamika – aerodinamičnost, agonija – trpljenje, ambicija – želja po uspehu, anatomija – telesna zgradba	674
Ustrezno odstranjenih (TP)	arbiter – posrednik, argument – razlaga, avto – vagon, birokrat – velika živina, čajnik – kavnik, cedilo – posodica za kuhinjske odpadke	82
Neustrezno obdržanih (FN)	arbiter – gospodar, arhiv – arhivi, avtoriteta – premoč, dedek – babica, dražba – razpis del, električar –	167

	vzdrževalec telefonskega omrežja	
Neustrezno odstranjenih (FP)	adaptacija – predelava, anonimnost – nepoznanost, aplikacija – prekritje, atentat – umor, bife – prehranjevalnica, cenzura – predelava [tiskarstvo]	28
Skupaj		951

Tabela 3 prikazuje natančnost (kolikšen delež odstranjenih primerov so dejansko neustrezni sopomenski kandidati), priklic (kolikšen delež vseh neustreznih kandidatov je bil identificiran) in F1 (harmonično sredino obeh vrednosti).

Tabela 3: Natančnost (*precision*), priklic (*recall*) in F1 za obravnavo neustreznih sopomenskih kandidatov.

Natančnost	Priklic	F1
0,7455	0,3293	0,4568

Iz rezultatov je razvidno, da je ChatGPT pri presojanju relevantnosti sopomenskih kandidatov opazno popustljivejši od zlatega standarda, čeprav so uredniška načela SSSS že izhodiščno naravnana k širšemu razumevanju sopomenskosti in odločitvi za karseda široko vključevanje kandidatov (Gantar in sod., 2023, str. 161). Kot smo zapisali v Razdelku 2.2, poziv za strojno obdelavo ni vseboval celotnih slovaropisnih smernic, po katerih velja, da se moške in ženske slovnične oblike ne obravnavajo kot neposredne sopomenke, ampak se uvrščajo pod spolsko ustrezajoče iztočnice (npr. *dedek* – *stari oče*, *babica* – *stara mama*, ne pa *\*dedek* – *babica*), da se množinske oblike ne upoštevajo kot sopomenke, razen če so za to v rabi utemeljeni razlogi (*\*arhiv* – *arhivi*) in da se opisne, definicijam podobne zveze obdržijo le, če se kot take pojavljajo v rabi (*\*dražba* – *razpis del*). Razlike v navodilih pojasnijo del razlik. Pri morebitni uporabi ChataGPT za pohitritev ročnega dela bi bila ta odstopanja predvidljiva, hitro opazna in enostavno rešljiva.

V naboru neustrezno obdržanih so tudi mejni primeri, ki so bili zahtevni že za slovaropisno odločitev. Pri teh bi raba ChataGPT za pohitritev ročnega dela lahko doprinesla k lažjim, morda še širše vključujočim odločitvam. Na drugi strani so problematične odstranitve, ki se v rezultatih sicer pojavljajo relativno redko, kot denimo *atentat* – *umor*, *debelost* – *obilnost*, *kaos* – *razdejanje*. Pri tovrstnih primerih bi bila pri morebitni rabi postopka potrebna pozornost.

### 3.3 Napake v razvrščanju sopomenk

Pri analizi razvrščanja sopomenk pod pomene smo ločili dve vrsti razlik: (a) ChatGPT je sopomenko umestil pod neustrezen besedni pomen in (b) ChatGPT sopomenke ni umestil pod ustrezen pomen oz. vse ustrezne pomene glede na zlati standard. Umestitev pod neustrezen pomen se pojavi pri 36 iztočnicah (14,6 % analiziranih iztočnic), manjkajoča umestitev pri 49 iztočnicah (19,9 %), od tega je 14 (5,7 %) takih, kjer se pojavljata obe vrsti problema, tj. umestitev pod neustrezen pomen ter manjkajoča umestitev. V Tabeli 4 so prikazani primeri, število razlik in iztočnic ter povprečno število kandidatov in slovarskih pomenov v posamezni od skupin. Pri primerih je najprej navedena iztočnica, sledi sopomenka, o kateri je ChatGPT presojal, in pomen, pod katerega jo je oz. je ni umestil. Kot smo opozorili v Razdelku 2.3, ustreznost oz. neustreznost razumemo v razmerju do zlatega standarda, vendar se med rezultati pojavljajo tudi mejni primeri, kjer je lahko poleg slovaropisne odločitve sprejemljiva tudi odločitev ChataGPT.

Tabela 4: Primeri, število napak v iztočnicah, kjer je ChatGPT umestil sopomenko pod napačni pomen ali je ni umestil pod vse pomene. V stolpcih 3-6 je navedeno število napak, število iztočnic, povprečno število sopomenskih kandidatov in besednih pomenov za obe skupini.

<i>Vrsta rezultata</i>	<i>Primeri</i>	<i>Št. napak</i>	<i>Št. iztočnic</i>	<i>Povpr. št. kandi-datov</i>	<i>Povpr. št. pomenov</i>
Umeščeno pod neustrezen pomen	<ul style="list-style-type: none"> <li>bazar – sejem [ekonomija]: pod 'orientalska tržnica' namesto 'prireditve'</li> <li>depresija – deprimiranost: pod 'bolezen' namesto 'potrtost'</li> <li>hazarder – igralec na srečo: pod 'kdor rad veliko tvega' namesto 'kdor rad stavi'</li> <li>nakup – kupčija: pod 'nakupljeno blago' namesto 'dejanje'</li> <li>pes – klinec: pod 'žival' namesto 'izprijena'</li> <li>ničvredna oseba [izraža]</li> </ul>	55	36	6,7	2,8

	<ul style="list-style-type: none"> <li>negativen odnos]'</li> <li>recept – formula: pod 'za izdajo zdravila' namesto 'navodilo; pravilo'</li> </ul>				
Neumeš- čeno pod pomen	<ul style="list-style-type: none"> <li>arhitektura – stavbarstvo: ustrezno pod 'veda in dejavnost', manjka pri 'o gradnji stavb'</li> <li>bajka – zgodbica: ustrezno pod 'pripovedka [književnost]', manjka pri 'izmišljotina'</li> <li>bolnik – pacient: ustrezno pod 'kdor je bolan', manjka pri 'kdor je neprijeten ali krut [izraža negativen odnos]</li> <li>dialog – dvogovor: ustrezno pri 'pogovor med osebama', manjka pri 'med nasprotnima stranema'</li> <li>gneča – množica: ustrezno pri 'o ljudeh', manjka pri 'o stvareh'</li> <li>gneča – truma: ustrezno pri 'o ljudeh', manjka pri 'o stvareh'</li> </ul>	78	49	5,3	2,9
Skupaj		133	71	5,3	2,8

Podatki v tabeli 4 kažejo, da se razlike pri razvrščanju pojavljajo pri iztočnicah, ki so v povprečju kompleksnejše glede števila sopomenk za razvrstitev ter števila besednih pomenov. Sklepati je mogoče, da na razlike vpliva tudi abstraktnost pomenskih indikatorjev, ki so človeku morda laže razumljivi (gl. Razdelek 3.5), vendar se to ne kaže enoznačno: kot primer uspešno urejenega kompleksnega gesla lahko podamo iztočnico *jagoda* s štirimi kratkimi in medsebojno podobnimi indikatorji (1. *rastlina*, 2. *plod*, 3. *okrogel plod*, 4. *droben predmet*), h katerim je sistem ustrezno razmestil sopomenske kandidate *biser*, *koralda*, *kroglica* ali iztočnico *telefon* s šestimi indikatorji (1. *naprava*, 2. *telekomunikacijska povezava*, 3. *napeljava*, 4. *o storitvi*, 5. *telefonska številka [neformalno]*, 6. *družabna igra*), kamor sta bila ustrezno razvrščena kandidata *telefonska slušalka*, *telefonski klic*.

### 3.4 Najtežji primeri

Med vsemi analiziranimi iztočnicami je šest takšnih, pri katerih so se pojavile težave tako na ravni odstranjevanja neustreznih kandidatov kot umeščanja sopomenskih kandidatov pod vse ustrezne pomene. Te obravnavamo kot težke primere in jih v nadaljevanju ponazarjamo v celoti, tako v slovaropisni različici kot po presoji ChataGPT. Razlike so obarvane z zeleno.

Tabela 5: Prikaz obravnave najtežjih primerov z oznakami razlik med odločitvami.

<i>Iztočnica</i>	<i>Slovaropisno</i>	<i>ChatGPT</i>
adaptacija	<p>1. prenova stavbe: <i>prilagoditev</i>, <i>prenova</i>, <i>preureditev</i></p> <p>2. priredba [v umetnosti]: <i>predelava [glasba]</i>, <i>priredba [glasba]</i>, <i>preureditev [glasba]</i>, <i>aranžma [glasba]</i></p> <p>3. prilagoditev [ponavadi v športu]: PRAZNO</p> <p>Odstranjeni kandidati: /</p>	<p>1. prenova stavbe: <i>preureditev</i>, <i>prenova</i></p> <p>2. priredba [v umetnosti]: <i>priredba [glasba]</i>, <i>aranžma [glasba]</i></p> <p>3. prilagoditev [ponavadi v športu]: <i>prilagoditev</i></p> <p>Odstranjeni kandidati: <i>predelava [glasba]</i></p>
čarovnica	<p>1. pravljіčno bitje: <i>čarodejka</i>, <i>coprnica</i></p> <p>2. ženska, obtožena čarovništva: <i>vračarica</i></p> <p>3. zlobna, starejša ženska [izraža negativen odnos]: <i>veščca</i>, <i>babura</i>, <i>coprnica</i>, <i>hudičevka</i>, <i>stara baba</i>, <i>zlobnica</i></p> <p>4. ženska z nadnaravnimi sposobnostmi [lahko izraža negativen odnos]: <i>čarodejka</i>, <i>coprnica</i></p> <p>Odstranjeni kandidati: <i>iluzionistka</i></p>	<p>1. pravljіčno bitje: <i>čarodejka</i>, <i>coprnica</i>, <i>vračarica</i>, <i>veščca</i></p> <p>2. ženska, obtožena čarovništva: <i>veščca</i>, <i>hudičevka</i></p> <p>3. zlobna, starejša ženska [izraža negativen odnos]: <i>babura</i>, <i>stara baba</i>, <i>zlobnica</i></p> <p>4. ženska z nadnaravnimi sposobnostmi [lahko izraža negativen odnos]: <i>iluzionistka</i></p> <p>Odstranjeni kandidati: /</p>
gurman	<p>1. ljubitelj hrane in pijače: <i>sladokusec</i>, <i>uživac</i>, <i>gastronom</i>, <i>poznavalec hrane</i>, <i>epikurejec</i></p> <p>2. poznavalec: <i>sladokusec</i></p>	<p>1. ljubitelj hrane in pijače: <i>sladokusec</i>, <i>uživac</i>, <i>bonvivan</i>, <i>poznavalec hrane</i>, <i>gastronom</i>, <i>epikurejec</i></p> <p>2. poznavalec: <i>bonvivan</i>, <i>poznavalec hrane</i>, <i>gastronom</i></p>

	Odstranjeni kandidati: <i>bonvivan</i>	Odstranjeni kandidati: /
kandidat	<p>1. za službo ali funkcijo: <i>prosilec, aspirant, prijavljenec, interesent [ekonomija], prijavljeni</i></p> <p>2. na izpitu: <i>prijavljenec, prijavljeni, udeleženec</i></p> <p>3. za nagrado: <i>predlaganec, nominiranec</i></p> <p>Odstranjeni kandidati: <i>potencialna kandidatka, možna kandidatka, izpitna kandidatka, vložnik, potencialni kupec [ekonomija], možni kupec [ekonomija], potencialna stranka [ekonomija], možna stranka [ekonomija], izprašanec, volilna agitatorka, volilni agitator, tekmovalec [šport], anketiranec, intervjuvanec</i></p>	<p>1. za službo ali funkcijo: <i>prosilec, aspirant, prijavljenec, potencialna kandidatka, možna kandidatka, predlaganec, prijavljeni, vložnik</i></p> <p>2. na izpitu: <i>izpitna kandidatka, izprašanec</i></p> <p>3. za nagrado: <i>nominiranec, volilna agitatorka, volilni agitator</i></p> <p>Odstranjeni kandidati: <i>interesent [ekonomija], potencialni kupec [ekonomija], možni kupec [ekonomija], potencialna stranka [ekonomija], možna stranka [ekonomija], udeleženec, tekmovalec [šport], anketiranec, intervjuvanec</i></p>
ljubezen	<p>1. močna naklonjenost: <i>naklonjenost, nagnjenje, posebna naklonjenost, posebno nagnjenje, eros</i></p> <p>2. o intimnem odnosu: <i>romanca, ljubezensko razmerje</i></p> <p>3. ljubljena oseba: <i>draga, ljuba, dragi, ljubi, fant, punca</i></p> <p>4. o dejavnosti: <i>naklonjenost, strast, nagnjenje, posebna naklonjenost, posebno nagnjenje</i></p> <p>5. seks: <i>strast</i></p> <p>Odstranjeni kandidati: <i>nežnost, toplina, posebna nagnjenje, oboževanje, občudovanje, čustva</i></p>	<p>1. močna naklonjenost: <i>naklonjenost, strast, nagnjenje, posebna naklonjenost, posebna nagnjenje, posebno nagnjenje, oboževanje, občudovanje, čustva</i></p> <p>2. o intimnem odnosu: <i>ljubezensko razmerje, romanca</i></p> <p>3. ljubljena oseba: <i>draga, ljubi, dragi, ljuba, fant, punca</i></p> <p>4. o dejavnosti: PRAZNO</p> <p>5. seks: <i>eros</i></p> <p>Odstranjeni kandidati: <i>nežnost, toplina</i></p>
urok	<p>1. v magiji: <i>čarovnija, prekletstvo, zakletev, zaklinjanje, zarotitev, magične besede, coprnija, zarekanje, čarovniški</i></p>	<p>1. v magiji: <i>čarovnija, prekletstvo, zakletev, zarotitev, magične besede, coprnija, zaklinjanje, abrakadabra, vudu, čaranje,</i></p>

<p><i>izrek, čarobni napev</i></p> <p>2. o neuspehu [pogosto v športnem kontekstu]: <i>prekletstvo</i></p> <p>Odstranjeni kandidati: <i>nesreča, smola, abrakadabra, vudu, čaranje, čarobna formula, čarobni simbol, molitveni obrazec [religija]</i></p>	<p><i>čarobna formula, čarovniški izrek, čarobni napev, čarobni simbol</i></p> <p>2. o neuspehu [pogosto v športnem kontekstu]: <i>nesreča, smola, zarekanje</i></p> <p>Odstranjeni kandidati: <i>molitveni obrazec [religija]</i></p>
---	--

Primeri pregledno prikazujejo raznovrstne težave. Kot je razvidno, so določena razhajanja lahko posledica korpusnega gradiva, na osnovi katerega se oblikujejo pomenske členitve in potrjuje sopomenska raba (npr. *bonvivan*, ki se v rabi najbrž pojavlja preredko, da bi ga obdržali, ali *vešča* v pomenu pravljičnega bitja). Prav tako so lahko mestoma zavajajoče ubeseditve v pomenskih indikatorjih, ki jih slovaropisna ekipa interpretira na podlagi preostalih podatkov v DSB, pri nalogi za ChatGPT pa so bili predstavljeni brez dodatnega konteksta (npr. indikator 'poznavalec' pri iztočnici *gurman*, ki je v opoziciji do 1. pomena in se v prenesenem pomenu ne navezuje več na hrano/pijačo, česar ChatGPT ne razbere). Nekaj je primerov, pri katerih slovaropisci upoštevajo smernice, ki ChatuGPT niso bile podane (gl. 3.2), npr. pri (ne) vključevanju moško-ženskih parov (*kandidat – izpitna kandidatka*). Najti pa je tudi razlike, kjer so odločitve ChataGPT težko razložljive, npr. *kandidat 'za nagrado' – volilna agitatorka, volilni agitator*.

#### 4. SKLEP IN NADALJNJE DELO

V raziskavi smo preverili, kako dober je ChatGPT za umeščanje sopomenskega gradiva pod besedne pomene. Analizirali smo rezultate razvrščanja 951 sopomenskih kandidatov za 246 slovarskih iztočnic. Za 41,9 % slednjih je strojni postopek vrnil rezultate, povsem skladne s slovaropisnimi. Pri ostalih iztočnicah, ki so v povprečju kompleksnejše (prinašajo več sopomenskih kandidatov za razvrstitev in več slovarskih pomenov), se pojavljajo odstopanja različnih vrst.

Ob odstranjevanju neustreznih sopomenskih kandidatov se sistem razlikuje v 43,5 % analiziranih iztočnicah. Večina odstopanj je posledica popustljivosti

sistema do sopomenskih kandidatov, ki jih je slovaropisna ekipa odstranila. Ker koncept SSSS načelno teži k vključevanju gradiva, slovarski vmesnik pa omogoča reakcijo uporabniške skupnosti na neustrezne kandidate, so ti odstopi manj problematični. V 28,9 % analiziranih iztočnic se pojavijo napačne razporeditve sopomenk pod pomene ali neumestitve sopomenk pod vse ustrezajoče pomene. Ti odstopi so pogostejši pri kompleksnejših geslih, predvidevamo pa, da so vsaj delno (lahko) posledica ubeseditiv indikatorjev, kot tudi specifik korpusnega gradiva, ki v slovaropisnih delotokih pogojuje pomensko členjenje in preverbo sopomenskosti. Natančnejši pregled primerov, v katerih se pojavljajo raznovrstne razlike, pokaže, da se ChatGPT tudi pri najtežjih primerih ne razlikuje radikalno, razlike pa so lahko za slovaropisno delo tudi uporabne, saj omogočajo dodatne razmisleke, zlasti pri mejnih primerih. Skleniti je mogoče, da postopek deluje dokaj dobro in ima uporabno vrednost za pohitritev ročnega slovaropisnega dela.

Večji izziv je nepredvidljivost postopka. Za razliko od preteklih metodologij strojnega pridobivanja jezikovnih podatkov je generiranje s programom ChatGPT manj gradivno in postopkovno transparentno, zato je nemogoče sklepati o vzrokih za posamezne odločitve. Ta značilnost pomembno omejuje evalvacijske raziskave, kot je naša, ne more pa biti razlog, da generativnih tehnologij v slovaropisju ne bi uporabljali in/ali ocenjevali. V nadaljevanju bi bilo smiselno ciljno preveriti hipotezo, da so pri preprostejših geslih rezultati zanesljivejši, preizkusiti pripravo gradiva za nove (preprostejše) iztočnice in nato testirati, ali strojna predpriprava slovaropisne odločitve pohitri ali ne.

Zanimivo bi bilo nadalje raziskati primere, ki jih ChatGPT za razliko od slovaropisne ekipe ni odstranil, in raziskavam, ki preverjajo razumevanje koncepta sopomenskosti med različnimi uporabniškimi skupinami slovarja (Gapsa, 2023; Gapsa in Arhar Holdt, 2023), dodati še »razumevanje« pri rabi ChataGPT. Raziskavo je mogoče ponoviti na zmogljivejših različicah ChataGPT ali drugih podobnih sistemih, z nadgrajenimi pozivi in na novem gradivu (npr. za razvrščanje uporabniško dodanih sopomenk ali protipomenk). V načrtu je raziskava, ki bo preverila, koliko na rezultate vpliva prisotnost oz. odsotnost slovarskih oznak. Preizkusiti je mogoče tudi druge naloge v podporo slovaropisnemu delu, tako za urejanje gradiva posameznega slovarja kot povezovanje leksikalnih podatkov iz različnih virov. Z ustreznimi

metodološkimi premisleki je mogoče preveriti in vključiti tudi ustvarjalne generativne naloge, kot je denimo predlaganje novih sopomenk in protipomenk za podane iztočnice.

## ZAHVALA

Zahvaljujemo se vsem, ki so omogočili izvedbo raziskave. Pri pripravi raziskovalnih podatkov je sodeloval Timotej Petrič. Projekt Nadgradnja temeljnih slovarskih virov in podatkovnih baz CJVT UL je v letih 2021–2022 financiralo Ministrstvo za kulturo Republike Slovenije. Raziskovalni program Jezikovni viri in tehnologije za slovenski jezik (št. P6-0411) in raziskovalni projekt Empirična podlaga za digitalno podprt razvoj pisne jezikovne zmožnosti (št. J7-3159) sofinancira Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije iz državnega proračuna.

## LITERATURA

- Arhar Holdt, Š., Čibej, J., Dobrovoljc, K., Gantar, P., Gorjanc, V., Klemenc, B., Kosem, I., Krek, S., Laskowski, C. in Robnik-Šikonja, M. (2018). Thesaurus of Modern Slovene: By the Community for the Community. V J. Čibej, V. Gorjanc, I. Kosem in S. Krek (ur.), *Proceedings of the XVIII EURALEX International Congress, Lexicography in Global Contexts, 17-21 July 2018, Ljubljana* (str. 401–410). Znanstvena založba Filozofske fakultete. <https://doi.org/10.4312/9789610600961>
- Arhar Holdt, Š., Gantar, P., Kosem, I., Pori, E., Robnik Šikonja, M. in Krek, S. (2023). Thesaurus of Modern Slovene 2.0. V M. Medved, M. Měchura, C. Tiberius, I. Kosem, J. Kallas, M. Jakubiček in S. Krek (ur.), *Electronic lexicography in the 21st century (eLex 2023), Proceedings of the eLex 2023 conference, 27–29 June 2023* (str. 366–381). Lexical Computing CZ. Pridobljeno 28. maja 2024, <https://elex.link/elex2023/wp-content/uploads/82.pdf>
- de Schryver, G.-M. (2023). Generative AI and Lexicography: The Current State of the Art Using ChatGPT. *International Journal of Lexicography*, 36(4), 355–387. <https://doi.org/10.1093/ijl/ecad021>
- de Schryver, G.-M. in Joffe, D. (2023). The End of Lexicography, Welcome to the Machine: On How ChatGPT Can Already Take over All of the Dictionary Maker's Tasks. *20th CODH Seminar, ROIS-DS Center for Open Data in the Humanities, Tokyo, Japan, 27 February 2023* [Referat]. Pridobljeno 28. maja 2024, <http://codh.rois.ac.jp/seminar/lexicography-chatgpt-20230227/>
- Gantar, P., Arhar Holdt, Š., Kosem, I. in Krek, S. (2023). Sopomenke 2.0 in Kolokacije

- 2.0: Novi koraki za slovenske odzivne slovarje. *Jezik in Slovastvo*, 68(4), 157–175. <https://doi.org/10.4312/jis.68.4.157-175>
- Gapsa, M. (2022). Ocenjevanje uporabniško dodanih sopomenk v Slovarju sopomenk sodobne slovenščine – pilotna študija. V D. Fišer in T. Erjavec (ur.), *Zbornik konference Jezikovne tehnologije in digitalna humanistika, 15.–16. september 2022, Ljubljana, Slovenija* (str. 308–316). Inštitut za novejšo zgodovino. Pridobljeno 28. maja 2024, [https://nl.ijs.si/jtdh22/pdf/JTDH2022\\_Gapsa\\_Ocenjevanje-uporabnisko-dodanih-sopomenk-v-Slovarju-sopomenk-sodobne-slovenscine.pdf](https://nl.ijs.si/jtdh22/pdf/JTDH2022_Gapsa_Ocenjevanje-uporabnisko-dodanih-sopomenk-v-Slovarju-sopomenk-sodobne-slovenscine.pdf)
- Gapsa, M. (2023). “But why??” Evaluation of user-suggested synonyms in the Thesaurus of Modern Slovene. [Preprint]. <https://doi.org/10.21203/rs.3.rs-2775161/v1>
- Gapsa, M. in Arhar Holdt, Š. (2023). How Lexicographers Evaluate User Contributions in The Thesaurus of Modern Slovene in Comparison to Dictionary Users. V M. Medved, M. Měchura, C. Tiberius, I. Kosem, J. Kallas, M. Jakubíček in S. Krek (ur.), *Electronic lexicography in the 21st century (eLex 2023): Invisible Lexicography. Proceedings of the eLex 2023 conference. Brno, 27–29 June 2023* (str. 178–200). Lexical Computing CZ. Pridobljeno 28. maja 2024, <https://elex.link/elex2023/wp-content/uploads/47.pdf>
- Jakubíček, M. in Rundell, M. (2023). The end of lexicography? Can ChatGPT outperform current tools for post-editing lexicography?. V M. Medved, M. Měchura, C. Tiberius, I. Kosem, J. Kallas, M. Jakubíček in S. Krek (ur.), *Electronic lexicography in the 21st century (eLex 2023), Proceedings of the eLex 2023 conference, 27–29 June 2023* (str. 518–533). Lexical Computing CZ. Pridobljeno 28. maja 2024, <https://elex.link/elex2023/wp-content/uploads/102.pdf>
- Kosem, I., Krek, S. in Gantar, P. (2021) Semantic data should no longer exist in isolation: the Digital Dictionary Database of Slovenian. V Z. Gavrilidou, L. Mitits in S. Kiosses (ur.), *EURALEX XIX: Congress of the European Association for Lexicography, Lexicography for inclusion, 7-9 September 2021, Virtual, Book of Abstracts* (str. 81–83). Democritus University of Thrace. Pridobljeno 29. maja 2024, <https://euralex.org/wp-content/uploads/2022/04/ABS2020.pdf>
- Krek, S., Laskowski, C. in Robnik-Šikonja, M. (2017). From translation equivalents to synonyms: creation of a Slovene thesaurus using word co-occurrence network analysis. V I. Kosem, C. Tiberius, M. Jakubíček, J. Kallas, S. Krek in V. Baisa (ur.), *Electronic lexicography in the 21st century. Proceedings of eLex 2017 conference: Lexicography from Scratch. Leiden, the Netherlands, 19–21 September 2017* (str. 93–109). Dutch Language Institut, Lexical Computing CZ

- s.r.o., Trojina. Pridobljeno 28. maja 2024, <https://elex.link/elex2017/wp-content/uploads/2017/09/paper05.pdf>
- Krek, S., Laskowski, C., Robnik-Šikonja, M., Kosem, I., Arhar Holdt, Š., Gantar, P., Čibej, J., Gorjanc, V., Klemenc, B. in Dobrovoljc, K. (2018). *Thesaurus of Modern Slovene 1.0*. Repozitorij raziskovalne strukture CLARIN.SI. <http://hdl.handle.net/11356/1166>
- Krek, S., Laskowski, C., Robnik-Šikonja, M., Kosem, I., Arhar Holdt, Š., Gantar, P., Čibej, J., Gorjanc, V., Klemenc, B., Dobrovoljc, K., Pori, E., Roblek, R. in Zgaga, K. (2023). *Thesaurus of Modern Slovene 2.0*. Repozitorij raziskovalne strukture CLARIN.SI. <http://hdl.handle.net/11356/1916>
- Lew, R. (2023). ChatGPT as a COBUILD Lexicographer. *Humanities and Social Sciences Communications* 10, članek 704 (2023). <https://doi.org/10.1057/s41599-023-02119-6>
- OpenAI. (2024). *ChatGPT* (31. 5. 2024) [veliki jezikovni model]. <https://chatgpt.com>
- Rundell, M. (2023). Automating the Creation of Dictionaries: Are We Nearly There?. V *Proceedings of the 16th International Conference of the Asian Association for Lexicography: "Lexicography, Artificial Intelligence, and Dictionary Users"*, 22-24 June 2023, Seoul, South Korea (str. 9–17). Yonsei University. Pridobljeno 28. maja 2024, <https://www.asialex.org/pdf/Asialex-Proceedings-2023.pdf>

## HOW GOOD IS CHATGPT AT PLACING SYNONYMS UNDER WORD SENSES

In this study, we test how well ChatGPT-4 cleans the list of automatically retrieved synonym candidates and distributes the synonyms under appropriate lexical senses. As a gold standard, we consider the lexicographic decisions made when updating the Thesaurus of Modern Slovene to version 2.0. In this paper, we compare the results for 246 dictionary entries. For 41.9% of entries, ChatGPT processed the data in the same way as lexicographers, while for 58.1%, it made a different decision: 43.5% of entries contained differences in the removal of noisy data, and 28.9% in the mapping of synonyms to lexical senses. When assessing the relevance of synonym candidates, ChatGPT is more permissive than the gold standard (recall 0.33), while precision is higher (0.75), but the differences are more difficult to explain. Differences in synonym placement (placement under a different sense in 14.6% of entries, missing placement in 19.9%) are partly attributed to features of the input data, such as task complexity and brevity of semantic indicators. Future work will focus on the validation of the method for speeding up lexicographic work.

**Keywords:** digital lexicography, ChatGPT, synonyms, word senses, Slovene language

To delo je ponujeno pod licenco Creative Commons: Priznanje avtorstva-Deljenje pod enakimi pogoji 4.0 Mednarodna.

This work is licensed under the Creative Commons Attribution-ShareAlike 4.0 International.

<https://creativecommons.org/licenses/by-sa/4.0/>

