

Špela Arhar Holdt,^{*} Magdalena Gapsa,[♦]
Polona Gantar,[°] Iztok Kosem[•]

Potencial ChatGPT pri razvoju Slovarja sopomenk sodobne slovenščine

IZVLEČEK

V raziskavi preverjamo, kako dobro se ChatGPT-4 odreže pri dveh slovaropisnih nalogah: (a) čiščenju seznama strojno pridobljenih sopomenskih kandidatov in umeščanju sopomenskega gradiva pod besedne pomene ter (b) izdelavi slovarskega gesla, vključno s pomensko členitvijo, definicijami in zgledi, na podlagi različnih vhodnih podatkov. Kot zlati standard upoštevamo slovaropisne odločitve, vključene v Digitalno slovarsko bazo za slovenščino. V prvem preizkusu analiziramo rezultate za 246 slovarskih iztočnic in ugotavljamo, da je ChatGPT podatke uredil povsem enako kot slovaropisci pri 41,9 odstotka iztočnic, pri 58,1 odstotka pa se je v odločitvi razlikoval. Pri presojanju relevantnosti sopomenskih kandidatov je bil ChatGPT popustljivejši od zlatega standarda. Razlike v razvrščanju sopomenk (umestitev pod drug pomen pri 14,6 odstotka iztočnic, manjkajoča umestitev pri 19,9 odstotka) deloma pripisujemo značilnostim vhodnih podatkov, kot sta kompleksnost naloge in kratkost pomenskih indikatorjev. V drugem preizkusu preverjamo zmožnost ChatGPT za samostojno izdelavo slovarskih gesel za 116 iztočnic. Analiza kakovosti generiranih pomenskih členitev in definicij kaže, da sistem deluje zmerno dobro: v 57 odstotkih primerov je zaznal vse pomene,

^{*} Dr., znan. sod., Univerza v Ljubljani, Filozofska fakulteta, Aškerčeva cesta 2, Ljubljana; Fakulteta za računalništvo in informatiko, Večna pot 113, 1000 Ljubljana, spela.arharholdt@ff.uni-lj.si; ORCID: 0000-0003-0565-0531

[♦] Inform. spec., Centralna tehniška knjižnica Univerze v Ljubljani, Trg republike 3, 1000 Ljubljana, magdalena.gapsa@ctk.uni-lj.si; ORCID: 0000-0003-2763-4495

[°] Dr., znan. sod., Univerza v Ljubljani, Filozofska fakulteta, Aškerčeva cesta 2, Ljubljana, apolonija.gantar@ff.uni-lj.si; ORCID: 0000-0001-5822-6414

[•] Dr., viš. znan. sod., Univerza v Ljubljani, Filozofska fakulteta, Aškerčeva cesta 2, Ljubljana; Institut »Jožef Stefan«, Jamova cesta 39, Ljubljana, iztok.kosem@ijs.si; ORCID: 0000-0002-4282-9031

skoraj 80 odstotkov generiranih gesel je doseglo povprečno oceno 3,5 ali več, 19 odstotkov pa najvišjo oceno obeh ocenjevalcev. Glavni izzivi so pretirano drobljenje pomenov, neprepoznane prenesene rabe in manjša predvidljivost rezultatov. Sklenemo lahko, da ima ChatGPT potencial za pohitritev ročnega slovaropisnega dela, če se njegovi rezultati ustrezno preverjajo in nadgrajujejo.

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

ABSTRACT

THE POTENTIAL OF CHATGPT IN THE DEVELOPMENT OF THE THESAURUS OF MODERN SLOVENE

In this study, we examine how well ChatGPT-4 performs in two lexicographic tasks: (a) cleaning the list of automatically retrieved synonym candidates and assigning synonymic material to lexical senses, and (b) generating dictionary entries, including sense division, definitions, and examples, based on different input data. As a gold standard, we consider the lexicographic decisions recorded in the Digital Dictionary Database for Slovene. In the first experiment, we analyse the results for 246 dictionary entries and find that ChatGPT processed the data identically to lexicographers in 41.9 % of cases, while in 58.1 % of cases, it made different decisions. When assessing the relevance of synonym candidates, ChatGPT was more permissive than the gold standard. Differences in synonym placement (assignment to a different sense in 14.6 % of entries, missing placement in 19.9 %) can be partly attributed to input data characteristics, such as task complexity and the brevity of semantic indicators. In the second experiment, we test ChatGPT's ability to autonomously generate dictionary entries for 116 headwords. The analysis of generated sense divisions and definitions reveals that the system performs moderately well: in 57 % of cases, it identified all senses, almost 80 % of generated entries received an average score of 3.5 or higher, and 19 % received the highest score from both evaluators. The main challenges include excessive splitting of senses, failure to recognise figurative meanings, and reduced predictability of results. We conclude that ChatGPT has potential for speeding up manual lexicographic work if its results are properly monitored and refined.

Keywords: digital lexicography, ChatGPT, synonyms, word senses, Slovenian language

Uvod

Generativna umetna inteligenca, ki temelji na velikih jezikovnih modelih, je prek klepetalnih vmesnikov, kakršen je ChatGPT, postala široko dostopna za številne z jezikom povezane naloge.¹ Med področji, ki zadnja leta preizkušajo moč in omejitve novih tehnologij, je tudi slovaropisje.

Kot pričajo Rundell,² Lew,³ Bartosz et al.,⁴ McKean in Fitzgerald⁵ ter Tiberius et al.,⁶ se dosedanje preizkusi rabe ChatGPT za slovaropisne namene osredotočajo na generiranje bolj ali manj celostnih slovarskih gesel za (pogosto dokaj priložnostno) izbran nabor iztočnic. De Schryver v svojem kritičnem pregledu prvih prispevkov na temo z umetno inteligenco podprtega slovaropisja poroča, da je trenutno največ pozornosti posvečene definicijam in primerom rabe.⁷ Skoraj vse študije oziroma preizkusi pa so bili izvedeni v angleščini in za angleščino, čeprav Jakubiček in Rundell naslavljata tudi problem večjezičnosti.⁸

Obstoječim raziskavam dodajamo dva preizkusa za slovenščino: (a) preizkus, kako dobro se ChatGPT-4 odreže pri čiščenju seznama strojno pridobljenih sopomenskih kandidatov in umeščanju sopomenskega gradiva pod besedne pomene, ter (b) preizkus izdelave slovarskega gesla (s pomensko členitvijo, definicijami in zgledi) na podlagi različnih vhodnih podatkov. Delo 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, kot opisujejo Krek, Laskowski in Robnik-Šikonja.⁹ Od objave leta 2018 se slovar postopoma ročno pregleduje in

1 »ChatGPT (veliki jezikovni model),« OpenAI, pridobljeno 31. 5. 2024, <https://chatgpt.com>.

2 Michael Rundell, »Automating the Creation of Dictionaries: Are We Nearly There?«, v: *Proceedings of the 16th International Conference of the Asian Association for Lexicography* (Yonsei University, 2023), 9–17, pridobljeno 20. 5. 2025, <https://www.asialex.org/pdf/Asialex-Proceedings-2023.pdf>.

3 Robert Lew, »ChatGPT as a COBUILD Lexicographer,« *Humanities and Social Sciences Communications* 10 (2023), pridobljeno 20. 5. 2025, <https://doi.org/10.1057/s41599-023-02119-6>.

4 Bartosz Ptasznik, Sascha Wolfer in Robert Lew, »A Learners' Dictionary versus ChatGPT in Receptive and Productive Lexical Tasks,« *International Journal of Lexicography* 37, št. 3 (2024): 322–36, pridobljeno 20. 5. 2025, <https://doi.org/10.1093/ijl/ecae011>.

5 Erin McKean in Will Fitzgerald, »The ROI of AI in Lexicography,« *Lexicography* 11, št. 1 (2024): 7–27, pridobljeno 20. 5. 2025, <https://utppublishing.com/doi/abs/10.1558/lexi.27569>.

6 Carole Tiberius et al., »LLMs and Evidence-based Lexicography,« v: Simon Krek, ur., *Large Language Models and Lexicography*, 2024, 44–48, pridobljeno 25. 1. 2025, https://www.cjvt.si/wp-content/uploads/2024/10/LLM-Lex_2024_Book-of-Abstracts.pdf.

7 Gilles-Maurice de Schryver, »Generative AI and Lexicography: The Current State of the Art Using ChatGPT,« *International Journal of Lexicography* 36, št. 4 (2023): 355–87, pridobljeno 20. 5. 2025, <https://doi.org/10.1093/ijl/ecad021>.

8 Miloš Jakubiček in Michael Rundell, »The End of Lexicography? Can ChatGPT Outperform Current Tools for Post-Editing Lexicography?«, v: *Electronic Lexicography in the 21st Century (eLex 2023): Proceedings of the eLex 2023 Conference*, ur. Marek Medved et al. (Lexical Computing CZ, 2023), 522–23, pridobljeno 20. 5. 2025, <https://elex.link/elex2023/wp-content/uploads/102.pdf>.

9 Simon Krek, Cyprian Laskowski in Marko Robnik-Šikonja, »From Translation Equivalents to Synonyms: Creation of a Slovene Thesaurus Using Word Co-occurrence Network Analysis,« v: Iztok Kosem et al., ur., *Electronic Lexicography in the 21st Century* (Leiden: Dutch Language Institute, Lexical Computing CZ s.r.o., Trojina, 2017), 93–109, pridobljeno 20. 5. 2025, <https://elex.link/elex2017/wp-content/uploads/2017/09/paper05.pdf>.

čisti v sodelovanju med strokovnjaki za slovaropisje ter zainteresirano uporabniško javnostjo. Različico 1.0 predstavljajo Arhar Holdt et al.,¹⁰ različico 2.0 pa Arhar Holdt et al.¹¹ in Gantar et al.¹²

Ideja pričujočega prispevka temelji na realnih potrebah nadaljnje slovarske gradnje. V prihodnje bi bilo v slovaropisne postopke mogoče 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 strojne podpore bi lahko pomembno pohitrila nadgrajevanje slovarja, s tem pa pripravo odprto dostopnega sopomenskega gradiva, ki je dragoceno tudi za razvoj številnih nadaljnjih jezikovih virov in tehnologij za sodobno slovenščino. Da bi lahko izbrali ustrezno metodologijo tovrstne strojne podpore, je v prvem koraku treba ugotoviti, kakšne rezultate daje ChatGPT v primerjavi s slovaropisci za različne avtentične slovaropisne naloge.

Prispevek je razširjena različica konferenčnega prispevka, v katerem je bil predstavljen prvi zgoraj navedeni preizkus.¹³ Za razširjeno različico smo dodali še drugi preizkus in članek ustrezno posodobili in nadgradili. V nadaljevanju zaporedno predstavimo metodologijo in rezultate obeh preizkusov, strnemo ugotovitve in napovemo nadaljnje delo na obravnavanem področju.

Prvi preizkus: selekcioniranje sopomenk in razvrščanje pod pomene

Metodologija

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 1.0¹⁴ (SSSS 1.0) in drugih prosto dostop-

10 Špela Arhar Holdt et al., »Thesaurus of Modern Slovene: By the Community for the Community,« v: Jaka Čižbej et al., ur., *Proceedings of the XVIII EURALEX International Congress, Lexicography in Global Contexts* (Ljubljana: Znanstvena založba Filozofske fakultete, 2018), 401–10, pridobljeno 20. 5. 2025, <https://doi.org/10.4312/9789610600961>.

11 Špela Arhar Holdt et al., »Thesaurus of Modern Slovene 2.0,« v: Marek Medved et al., ur., *Electronic Lexicography in the 21st Century (eLex 2023)* (Brno: Lexical Computing CZ, 2023), 366–81, pridobljeno 20. 5. 2025, <https://elex.link/elex2023/wp-content/uploads/82.pdf>.

12 Polona Gantar et al., »Sopomenke 2.0 in Kolokacije 2.0: Novi koraki za slovenske odzivne slovarje,« *Jezik in slovstvo* 68, št. 4 (2023): 157–75, pridobljeno 20. 5. 2025, <https://doi.org/10.4312/jis.68.4.157-175>.

13 Magdalena Gapsa, Špela Arhar Holdt in Iztok Kosem, »Kako dober je ChatGPT pri umeščanju sopomenk pod besedne pomene,« v: Špela Arhar Holdt in Tomaž Erjavec, ur., *Jezikovne tehnologije in digitalna humanistika: Zbornik konference* (Ljubljana: Inštitut za novejšo zgodovino, 2024), 144–62, pridobljeno 20. 5. 2025, <https://zenodo.org/records/13912515>.

14 Simon Krek et al., *Thesaurus of Modern Slovene 1.0* (Repozitorij raziskovalne strukture CLARIN.SI, 2018), pridobljeno 20. 5. 2025, <http://hdl.handle.net/11356/1166>.

pnih leksikalnih virih, kot opisuje Gapsa.¹⁵ 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 (oziroma odstranjeni) in razvrščeni pod identificirane pomene.

Za izbranih 266 iztočnic je bilo v prvem koraku iz baze SSSS 1.0 izluščenih skupno 1049 sopomenskih kandidatov (z morebitnimi področnimi slovarskimi oznakami). V drugem koraku so bile iz Digitalne slovarske baze za slovenščino¹⁶ izvožene pomenske členitve s pomenskimi indikatorji (tj. kratkimi opisi za ločevanje pomenov, kot pojasni Gantar¹⁷) 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 2.0¹⁸ (SSSS 2.0) pridobili slovaropisno pripravljene pomenske členjene iztočnice z razvrščenimi sopomenkami.

V prvem koraku analize je bilo med 266 iztočnicami prepoznanih 20 iztočnic, kjer se pomenska členitev iz DSBS ne ujema s SSSS 2.0 (npr. iztočnica *bonbon* ima v DSBS 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.

Struktura poziva za ChatGPT

Za izbrane iztočnice smo pripravili poziv za ChatGPT (Priloga 1), 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 zelenega 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.

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 povzemanje datoteko CSV in Excelovo datoteko z vsemi zbranimi podatki.

15 Magdalena Gapsa, »But why?? Evaluation of User-Suggested Synonyms in the Thesaurus of Modern Slovene,« *Lang Resources & Evaluation* (2025), pridobljeno 20. 5. 2025, <https://doi.org/10.1007/s10579-025-09821-8>.

16 Iztok Kosem, Simon Krek in Polona Gantar, »Semantic Data Should No Longer Exist in Isolation: The Digital Dictionary Database of Slovenian,« v: Zoe Gavrilidou et al., ur., *EURALEX XIX: Congress of the European Association for Lexicography* (Democritus University of Thrace, 2021), 81–83, pridobljeno 20. 5. 2025, <https://euralex.org/wp-content/uploads/2022/04/ABS2020.pdf>.

17 Polona Gantar, *Leksikografski opis slovenščine v digitalnem okolju* (Ljubljana: Znanstvena založba Filozofske fakultete, 2015), pridobljeno 20. 5. 2025, <https://doi.org/10.4312/9789612377922>.

18 Simon Krek et al., *Thesaurus of Modern Slovene 2.0* (Repozitorij raziskovalne strukture CLARIN.SI, 2023), pridobljeno 20. 5. 2025, <http://hdl.handle.net/11356/1916>.

Slovaropisna ekipa je določala sopomenskost na podlagi korpusne analize možnosti zamenjave sopomenskih besed v sobesedilu. V poziv nismo vključili celotnih smernic, ki jim je sledila slovaropisna ekipa, saj bi s tem v postopek vnesli preveč informacij in spremenljivk, kar bi privedlo 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 izvečnik najpomembnejših navodil. Navodila, ki jih nismo vključili v poziv, navajamo ob analizi rezultatov, kadar olajšajo interpretacijo razlik med ročnim in strojnim delom.

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. Vendar pa rezultati nakažejo, da je v določenih primerih rešitev, ki jo ponudi ChatGPT, drugačna od slovaropisne, vendar kljub temu sprejemljiva. Če bodo s ChatGPT pripravljene podatki vključeni v slovaropisne delotoke, bo v prihodnje treba presoditi, kako v praksi obravnavati take primere skladno z izbranim slovaropisnim konceptom.

Splošna uspešnost

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

Podatke s primeri iztočnic prikazuje Tabela 1, v kateri podajamo tudi povprečno število kandidatov in slovarskih pomenov v posamezni skupini. 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 ter uspešno uredi kompleksnejše iztočnice.

Tabela 1: Ujemanje med slovaropisci in ChatGPT s številom iztočnic, primeri in povprečnim številom sopomenskih kandidatov ter besednih pomenov na skupino

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

Vir: lastno delo

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

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 (26,2 odstotka) od 951 kandidatov. ChatGPT je odstranil le 110 kandidatov (11,6 odstotka). 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 pa sopomenski kandidat, o katerem je ChatGPT presojal.

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

Primeri		Vsota
Ustrezno obdržanih (TN)	adaptacija – preureditev, adolescence – 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 – vzdrževalec telefonskega omrežja	167
Neustrezno odstranjenih (FP)	adaptacija – predelava, anonimnost – nepoznanost, aplikacija – prekritje, atentat – umor, bife – prehranjevalnica, cenzura – predelava [tiskarstvo]	28
Skupaj		951

Vir: lastno delo

Tabela 3: Natančnost (kolikšen delež odstranjenih primerov je dejansko neustreznih sopomenskih kandidatov), priklic (kolikšen delež vseh neustreznih kandidatov je bil identificiran) in F1 (harmonična sredina obeh vrednosti)

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

Vir: lastno delo

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.¹⁹ Kot smo zapisali v razdelku Struktura poziva za ChatGPT, 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 ustrezne 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 pogosto pojavljajo v rabi (**dražba* – *razpis del*). Razlike v navodilih pojasnijo del razlik v rezultatih. Pri morebitni uporabi ChatGPT za pohitritev ročnega dela bi bila ta odstopanja predvidljiva, hitro opazna in enostavno rešljiva.

19 Gantar et al., »Sopomenke 2.0: Kolokacije 2.0: Novi koraki za slovenske odzivne slovarje,« 161.

V naboru neustrezno prepoznanih so tudi mejni primeri, ki so bili zahtevni že za slovaropisno odločitev. Pri teh bi raba ChatGPT za pohitritev ročnega dela lahko doprinesla k lažjim, morda še širše vključujočim odločitvam. Na drugi strani so problematične neprepoznane sopomenske besede, kot denimo *atentat – umor, debelost – obilnost, kaos – razdejanje*. Pri takšnih primerih bi bila pri morebitni rabi postopka potrebna pozornost.

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 oziroma vse ustrezne pomene glede na zlati standard. Umestitev pod neustrezen pomen smo prepoznali pri 36 iztočnicah (14,6 odstotka analiziranih iztočnic), manjkajočo umestitev pri 49 iztočnicah (19,9 odstotka), od tega je 14 (5,7 odstotka) takih, kjer se pojavljata obe vrsti problema, tj. umestitev pod neustrezen pomen in 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 ali je ni umestil. Kot smo opozorili v razdelku Postopek analize gradiva, ustreznost oziroma neustreznost razumemo v razmerju do zlatega standarda, vendar se med rezultati pojavljajo tudi mejni primeri, kjer je lahko poleg slovaropisne odločitve sprejemljiva tudi odločitev ChatGPT.

Tabela 4: Primeri v iztočnicah, kjer je ChatGPT umestil sopomenko pod napačen 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.

Vrsta rezultata	Primeri	Št. napak	Št. iztočnic	Povpr. št. kandidatov	Povpr. št. pomenov
Umeščeno pod neustrezen pomen	bazar – sejem [ekonomija]: pod 'orientalska tržnica' namesto 'prireditve'; hazarder – igralec na srečo: pod 'kdor rad veliko tvega' namesto 'kdor rad stavi'	55	36	6,7	2,8
Neumeščeno pod pomen	bolnik – pacient: ustrezno pod 'kdor je bolan', manjka pri 'kdor je neprijeten ali krut [izraža negativen odnos]; gneča – množica: ustrezno pri 'o ljudeh', manjka pri 'o stvareh'	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 lažje razumljivi, vendar se to ne kaže enoznačno: kot primer uspešno urejenega kompleksnega gesla lahko navedemo iztočnico *jagoda* s štirimi kratkimi in medsebojno podobnimi indikatorji (1. rastlina, 2. plod, 3. okrogel plod, 4. droben predmet), h katerim je sistem ustrezno razvrstil sopomenske kandidate *biser*, *korald*a, *kroglica*. Podobno je bilo tudi pri iztočnici *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* in *telefonski klic*.

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 ChatGPT. Razlike so prikazane z ležečim tiskom.

Tabela 5: Prikaz obravnave najtežjih primerov z opredelitvami razlik med odločitvami

Iztočnica	Slovaropisno	ChatGPT
adaptacija	1. prenova stavbe: prilagoditev, prenova, preureditev	1. prenova stavbe: preureditev, prenova
	2. priredba [v umetnosti]: <i>predelava</i> [glasba], priredba [glasba], <i>preureditev</i> [glasba], aranžma [glasba]	2. priredba [v umetnosti]: priredba [glasba], aranžma [glasba]
	3. prilagoditev [ponavadi v športu]: PRAZNO	3. prilagoditev [ponavadi v športu]: <i>prilagoditev</i>
	Odstranjeni kandidati: /	Odstranjeni kandidati: <i>predelava</i> [glasba]
čarovnica	1. pravljичno bitje: čarodejka, coprnica	1. pravljīčno bitje: čarodejka, coprnica, <i>vračarica</i> , <i>vešča</i>
	2. ženska, obtožena čarovništva: <i>vračarica</i>	2. ženska, obtožena čarovništva: <i>vešča</i> , <i>hudičevka</i>
	3. zlobna, starejša ženska [izraža negativen odnos]: <i>vešča</i> , babura, <i>coprnica</i> , <i>hudičevka</i> , stara baba, zlobnica	3. zlobna, starejša ženska [izraža negativen odnos]: babura, stara baba, zlobnica
	4. ženska z nadnaravnimi sposobnostmi [lahko izraža negativen odnos]: <i>čarodejka</i> , <i>coprnica</i>	4. ženska z nadnaravnimi sposobnostmi [lahko izraža negativen odnos]: <i>iluzionistka</i>
	Odstranjeni kandidati: <i>iluzionistka</i>	Odstranjeni kandidati: /

Iztočnica	Slovaropisno	ChatGPT
gurman	1. ljubitelj hrane in pijače: sladokusec, uživač, gastronom, poznavalec hrane, epikurejec	1. ljubitelj hrane in pijače: sladokusec, uživač, <i>bonvivan</i> , poznavalec hrane, gastronom, epikurejec
	2. poznavalec: <i>sladokusec</i>	2. poznavalec: <i>bonvivan</i> , <i>poznavalec hrane, gastronom</i>
	Odstranjeni kandidati: <i>bonvivan</i>	Odstranjeni kandidati: /
kandidat	1. za službo ali funkcijo: prosilec, aspirant, prijavljenec, <i>interesent [ekonomija]</i> , prijavljeni	1. za službo ali funkcijo: prosilec, aspirant, prijavljenec, <i>potencialna kandidatka</i> , <i>možna kandidatka</i> , <i>predlaganec</i> , prijavljeni, <i>vložnik</i>
	2. na izpitu: <i>prijavljenec</i> , <i>prijavljeni</i> , <i>udeleženec</i>	2. na izpitu: <i>izpitna kandidatka</i> , <i>izprašanec</i>
	3. za nagrado: <i>predlaganec</i> , nominiranec	3. za nagrado: nominiranec, <i>volilna agitatorka</i> , <i>volilni agitator</i>
	Odstranjeni kandidati: <i>potencialna kandidatka</i> , <i>možna kandidatka</i> , <i>izpitna kandidatka</i> , <i>vložnik</i> , potencialni kupec [ekonomija], možni kupec [ekonomija], potencialna stranka [ekonomija], možna stranka [ekonomija], <i>izprašanec</i> , <i>volilna agitatorka</i> , <i>volilni agitator</i> , tekmovalac [šport], anketiranec, intervjuvanec	Odstranjeni kandidati: <i>interesent [ekonomija]</i> , potencialni kupec [ekonomija], možni kupec [ekonomija], potencialna stranka [ekonomija], možna stranka [ekonomija], udeleženec, tekmovalac [šport], anketiranec, intervjuvanec
	1. močna naklonjenost: naklonjenost, nagnjenje, posebna naklonjenost, posebno nagnjenje, eros	1. močna naklonjenost: naklonjenost, <i>strast</i> , nagnjenje, posebna naklonjenost, <i>posebna nagnjenost</i> , posebno nagnjenje, <i>oboževanje</i> , <i>občudovanje</i> , <i>čustva</i>
ljubezen	2. o intimnem odnosu: romanca, ljubezensko razmerje	2. o intimnem odnosu: ljubezensko razmerje, romanca
	3. ljubljena oseba: draga, ljuba, dragi, ljubi, fant, punca	3. ljubljena oseba: draga, ljubi, dragi, ljuba, fant, punca
	4. o dejavnosti: <i>naklonjenost</i> , <i>strast</i> , <i>nagnjenje</i> , <i>posebna naklonjenost</i> , <i>posebno nagnjenje</i>	4. o dejavnosti: PRAZNO
	5. seks: <i>strast</i>	5. seks: <i>eros</i>
	Odstranjeni kandidati: nežnost, toplina, <i>posebna nagnjenost</i> , <i>oboževanje</i> , <i>občudovanje</i> , <i>čustva</i>	Odstranjeni kandidati: nežnost, toplina

Iztočnica	Slovaropisno	ChatGPT
	1. v magiji: čarovnija, prekletstvo, zakletev, zaklinjanje, <i>zarotitev</i> , magične besede, coprnija, <i>zarekanje</i> , čarovniški izrek, čarobni napev	1. v magiji: čarovnija, prekletstvo, zakletev, zarotitev, magične besede, coprnija, zaklinjanje, <i>abrakadabra</i> , <i>vudu</i> , <i>čaranje</i> , <i>čarobna formula</i> , čarovniški izrek, čarobni napev, <i>čarobni simbol</i>
urok	2. o neuspehu [pogosto v športnem kontekstu]: <i>prekletstvo</i>	2. o neuspehu [pogosto v športnem kontekstu]: <i>nesreča</i> , <i>smola</i> , <i>zarekanje</i>
	Odstranjeni kandidati: <i>nesreča</i> , <i>smola</i> , <i>abrakadabra</i> , <i>vudu</i> , <i>čaranje</i> , <i>čarobna formula</i> , <i>čarobni simbol</i> , molitveni obrazec [religija]	Odstranjeni kandidati: molitveni obrazec [religija]

Vir: lastno delo

Primeri pregledno prikazujejo različne 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 'pravljico bitje'). Prav tako so lahko mestoma zavajajoče ubeseditve v pomenskih indikatorjih, ki jih slovaropisna ekipa lahko interpretira na podlagi preostalih podatkov v DSBS, kot so na primer kolokacije, v nalogi za ChatGPT pa so bili predstavljeni brez dodatnega konteksta (denimo indikator poznavalec pri iztočnici *gurman*, ki je v opoziciji do 1. pomena /'ljubitelj hrane in pijače'/ in se v prenesenem pomenu ne navezuje več na hrano/pijačo, česar ChatGPT ne razbere). Nekaj je primerov, pri katerih slovaropisci upoštevajo smernice, ki ChatGPT niso bile podane v razdelku Struktura poziva za ChatGPT, na primer pri (ne) vključevanju moško-ženskih parov (*kandidat* – *izpitna kandidatka*). Najti pa je tudi razlike, kjer so odločitve ChatGPT težko razložljive, denimo *kandidat* v pomenu 'za nagrado' – *volilna agitatorica*, *volilni agitator*.

Drugi preizkus: izdelava novih pomenskih členitev

Ker številne iztočnice v Slovarju sopomenk sodobne slovenščine še nimajo izdelane pomenske členitve in pomenskih opisov, smo se odločili ChatGPT preizkusiti še pri nalogi pomenskega členjenja, ki je vključevala tudi oblikovanje definicij²⁰ in ne zgolj pomenskih indikatorjev.

²⁰ V tem prispevku uporabljamo termin slovarska *definicija* (v pozivu *definition*) tudi za definicije cobuildskega tipa, čeprav bi zanje po obliki in vsebini ustrežal tudi termin *razlaga*. S tem sledimo predhodnim študijam, ki so bile zasnovane za podobne namene (gl. Razdelek 1). Sprememba poziva z navodilom za *explanation* bi v konkretni raziskavi uvedla novo spremenljivko in drugačne rezultate, zanimivo pa bi jo bilo preizkusiti pri nadaljnjem delu.

Metodologija

Za preizkus smo iz DSBS izbrali 116 ročno izdelanih iztočnic (63 samostalnikov, 32 pridevnikov in 21 glagolov), pri čemer je bil glavni pogoj, da so vključene tudi v Kolokacijski slovar sodobne slovenščine,²¹ saj je to pomenilo, da so vsebovale zadostno količino kontekstualnih podatkov (kolokacij in zgledov). Čeprav smo zaradi poudarka na pomenski členitvi in razlikovanju med pomeni v nabor vključili predvsem večpomenske iztočnice (55 z dvema pomenoma, 34 s tremi, 12 s štirimi, po dve s petimi in šestimi ter eno s sedmimi pomeni), smo dodali tudi deset enopomenskih.

- Za vsako iztočnico smo iz baze izvozili ročno pregledane kolokacije in avtomatsko izluščene zglede (po en zgled na kolokacijo):
- Za vsak pomen smo izvozili do 20 kolokacij in zgledov.
- Pri izbiri kolokacij smo upoštevali podatek o jakosti logDice.
- Pri izbiri zgleда za kolokacijo smo izbrali tistega z najvišjo oceno kakovosti dobrega zgleда v orodju GDEX, ki ga opisujejo Kosem, Husak in McCarthy.²²
- Glede na izsledke predhodnih kolokacijskih raziskav smo dali poudarek na izvozu kolokacij za pomensko bolj obvestilne skladišne strukture:
 - Za samostalnike smo za strukture glagol + samostalnik v tožilniku, pridevnik + samostalnik in samostalnik + samostalnik v rodilniku izvozili po pet kolokacij in zgledov, preostanek smo zapolnili s kolokacijami in zgledi iz preostalih struktur.
 - Za pridevnike smo za strukturo pridevnik + samostalnik izvozili po deset kolokacij in zgledov, preostale iz drugih struktur.
 - Za glagole smo za strukturo glagol + samostalnik izvozili po sedem kolokacij in zgledov, za strukturo prislov + glagol po pet kolokacij in zgledov, preostale iz drugih struktur.

V primerih, ko prioritete strukture niso vsebovale dovolj kolokacij, smo jih nadomestili s kolokacijami iz drugih struktur.

Druga informacija, ki smo jo pripravili, so bile slovarske definicije, ki smo jih pridobili iz dveh virov: semantičnega slovenskega leksikona Open Slovene Wordnet 1.0²³ in Angleško-slovenskega slovarja Bridge.²⁴ V obeh primerih smo pridobljene definicije še dodatno prilagodili oziroma izboljšali:

- V slovenskem Wordnetu so slovenske definicije zgolj avtomatski prevod angleških definicij in so v številnih primerih kratke in slabo obvestilne, na primer

21 Iztok Kosem et al., *Kolokacijski slovar sodobne slovenščine* (Ljubljana: Znanstvena založba Filozofske fakultete, 2018–), pridobljeno 20. 5. 2025, <https://viri.cjvt.si/kolokacije/slv/#>.

22 Iztok Kosem, Miloš Husak in Diana McCarthy, »GDEX for Slovene«, v: Iztok Kosem in Karmen Kosem, ur., *Electronic Lexicography in the 21st Century: New Applications for New Users* (Ljubljana: Trojina, Institute for Applied Slovene Studies, 2011), 150–59, pridobljeno 20. 5. 2025, http://www.trojina.si/elex2011/elex2011_proceedings.pdf.

23 Jaka Čibej et al., *Open Slovene WordNet OSWN 1.0* (Slovenian language resource repository CLARIN.SI, 2023), pridobljeno 20. 5. 2025, <http://hdl.handle.net/11356/1888>.

24 Angleško-slovenski slovar Bridge (Ljubljana: Državna založba Slovenije, 2000).

humanost → *kakovost človeškosti*; *forma* → *določen način, na katerega se nekaj izrazi*. Pri obdelavi s ChatGPT-4 smo iskali daljše, celostavčne definicije. Primer pretvorbe za *prevajati*:

- Izvorna definicija: *restate (words) from one language into another language*.
- Avtomatski slovenski prevod: *ponovno izraziti (besede) iz enega jezika v drugem jeziku*.
- Izboljšana definicija: *Prevajati pomeni izražati ali podajati pomen besedil ali izrazov iz enega jezika v drugega, tako da ohranjamo njihov pomen*.
- V Angleško-slovenskem slovarju Bridge so definicije na voljo v celostavčni obliki, vendar pa vsebujejo angleške iztočnice. V tem primeru smo do definicij za naš preizkus prišli po sledečem postopku:
 - Najprej smo angleške iztočnice avtomatsko zamenjali s slovenskimi prevodi, npr. *Kar je huge, je izjemno veliko po obsegu, količini ali stopnji*. → *Kar je velikanski, je izjemno po obsegu, količini ali stopnji*. in *Kadar nekaj browns ali is browned, postane temnejše barve*. → *Kadar nekaj porjaveti ali porjaveti, postane temnejše barve*.
 - Nato smo v analizi najprej izločili neproblematične definicije, pri preostalih pa smo prepoznali pet vzorcev težav, od takih, ki so zahtevale samo popravek sklona, do takih, kjer smo morali odpraviti podvajanje iztočnice v definiciji ali celo daljši del ubeseditve. Na podlagi tega smo za odpravo napak prilagodili sistemske pozive za ChatGPT-4, tako da smo dobili izboljšane definicije, na primer *Kar je velikansko, je izjemno veliko po obsegu, količini ali stopnji*. in *Kadar nekaj porjavi, postane temnejše barve*.

Pri združitvi definicij iz dveh virov smo opazili, da definicije niso deloma prekrivne zgolj med viroma, temveč tudi znotraj posameznega vira, kot kaže primer za glagol *degradirati*:

SLOVENSKI WORDNET

- ***Degradirati*** pomeni opraviti dejanje, s katerim zmanjšamo stopnjo, rang ali vrednost nečesa, zaradi česar je to nekaj manj cenjeno ali spoštovano.
- ***Degradirati*** pomeni povzročiti zmanjšanje nivoja zemlje, na primer zaradi erozije.
- ***Degradirati*** pomeni uradno ali neformalno znižati nekoga ali nekaj v oceni, vrednosti ali ugledu, zaradi dejanskega dejanja, situacije ali presoje.
- ***Degradirati*** pomeni uradno prenesti nekoga na nižjo pozicijo ali mu uradno zmanjšati čin.
- ***Degradirati*** pomeni zmanjšati nečiji ali nečesa stopnjo ali rang, ali povzročiti, da se nekdo znajde v neprijetni ali nedostojni situaciji.

SLOVAR BRIDGE

- Če človek, ki ima oblast, **degradira** nekoga, mu podeli nižji položaj, pogosto kot znamenje kazni.
- **Degradirati** pomeni dati osebi ali stvari manj pomemben položaj ali veljavo.

Hkrati je analiza pokazala, da tudi izboljšane definicije lahko vsebujejo slovnične ali druge napake, zato smo v poziv za ChatGPT dodali navodilo, naj podane definicije po potrebi združuje in izboljša.

Za deset iztočnic definicij nismo imeli na voljo, kar se je izkazalo za koristno, saj smo tako preverili tudi delovanje ChatGPT samo s podanimi kolokacijami in zgledi.

Struktura poziva za ChatGPT

Pri pripravi poziva smo najprej opravili obsežno testiranje, pri čemer smo prišli do podobnih ugotovitev kot pri razvrščanju sopomenk pod pomene, da bolje deluje poziv v angleščini in s primerom vhodnih in želenih podatkov. Pri tem poizkusu smo že uporabili novejši model GPT-4o.

Poziv je bil razdeljen na sistemsko navodilo, ki je bilo vedno enako, je bilo pa nekoliko drugačno za vsako besedno vrsto zaradi vključenih vzorcev definicij. Primer sistema navodila za pridevniške iztočnice predstavlja Priloga 2.

V glavni del poziva smo potem vključili definicije, kolokacije in zglede:

Here are definitions, and collocations and their examples for the Slovenian word <iztočnica>. Collocations are numbered. Definitions come from various sources, and need to be improved, merged, and even omitted if they are referring to the same sense. Using all this data, create senses with definitions, distributing collocations and examples under senses. Provide only numbers of collocations, do not repeat the entire text of collocations and examples.

DEFINITIONS:

COLLOCATIONS AND EXAMPLES:

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 povzemanje datoteko CSV in Excelovo datoteko z vsemi zbranimi podatki.

Analiza

Pri analizi smo preverili tri vidike podatkov, pridobljenih s ChatGPT: pokritost pomenov v Digitalni slovarski bazi za slovenščino, splošno ustreznost generiranih gesel in splošno ustreznost generiranih definicij.

Analizo pokritosti pomenov je opravil en slovaropisec, pri čemer je uporabil lestvico od 0 do 5 (0 – ni bil zaznan noben pomen, neuporabni podatki; 1 – zaznani redki pomeni; 2 – zaznana približno polovica pomenov; 3 – zaznana več kot polovica pomenov; 4 – zaznani skoraj vsi pomeni ali pa vsi pomeni, a nekateri le delno; 5 – zaznani vsi pomeni). Morebitne pomanjkljivosti, kot je pretirano drobljenje pomenov, prekrivnost definicij in podobno, niso bile upoštevane, zanimalo nas je samo, ali so bili vsi ročno zaznani pomeni tudi avtomatsko identificirani.

Splošno ustreznost generiranih gesel in ustreznost definicij sta ocenjevala dva slovaropisca. Pri splošni ustreznosti gesel so bila gesla ocenjena z uporabniškega vidika, torej smiselnosti, razumljivosti in dodelanosti. Pri ocenjevanju nismo upoštevali primerjav s pomensko členitvijo v DSBS, saj ChatGPT ni dobil podatkov o načelih in pravilih, ki jim pri izdelavi pomenov sledijo slovaropisci. Uporabljena je bila ocenjevalna lestvica od 0 do 5:

- 5 – V celoti je geslo zelo informativno, definicije dobre, pomenska členitev ustrezna; možne so manjše pomanjkljivosti, npr. napačno razporejen zgled, slovnična napaka v definiciji ipd.
- 4 – V celoti je geslo dobro izdelano, je pa pomanjkljivo v enem ali dveh elementih, npr. neustrezne definicije pri določenih pomenih, več zgledov napačno razvrščenih, preveč pomenov.
- 3 – Geslo je dokaj informativno, pomenska členitev deloma neustrezna, a posreduje relevantne informacije; pomeni se delno prekrivajo, določene definicije so lahko problematične za razumevanje.
- 2 – Posamezni pomeni so ustrezni in smiselni, pomenska členitev je pretežno neustrezna, definicije so prekrivne, možna je neustrezna razdelitev zgledov ali ubeseditve definicij ipd.
- 1 – Pomenska členitev je nejasna oz. nelogična, med pomeni je težko ali nemogoče razlikovati, razporeditev zgledov je neustrezna, ubeseditve definicij so pretežno neustrezne.
- 0 – Geslo je povsem neustrezno, npr. ne pojasnjuje besede v iztočnici, prevladuje tuj jezik, definicije so povsem neustrezne.

Tudi pri ocenjevanju ustreznosti definicij smo uporabili lestvico od 0 do 5, pri čemer nismo upoštevali morebitnih nezaznanih pomenov, napačno umeščenih zgledov ali soodvisnosti z indikatorji. Uporabljena lestvica:

- 5 – Definicije so dokaj dobro ubesedene in pomeni jasno razločeni (manjše napake toleriramo).
- 4 – Ubeseditve definicij so lahko problematične ali pa so pomensko prekrivne.
- 3 – Nekatero definicije so slabo ubesedene, prihaja tudi do pomenske prekrivnosti.
- 2 – Večina ali vse definicije so slabo ubesedene, nekateri ali vsi pomeni so prekrivni in slabo razlikovalni.
- 1 – Večina ali vse definicije so slabo ubesedene, niso razlagalne, pa tudi med njimi je slaba razlikovalnost.
- 0 – Definicije pojasnjujejo napačne pomene ali iztočnice.

Rezultati

Analiza pokritosti generiranih pomenov v Digitalni slovarski bazi za slovenščino (Tabela 6) je pokazala zmerno dobre rezultate, pri čemer je bila pri več kot 93 odstotkih iztočnic zaznana polovica ali več pomenov, od tega so bili pri 57 odstotkih iztočnic zaznani vsi pomeni. V primeru dveh iztočnic (*bakren*, *padalski*) z oceno 0 je šlo za jasno napako modela, ki je ponudil podatke za povsem napačno iztočnico.

Tabela 6: Pokritost pomenov v Digitalni slovarski bazi za slovenščino

Ocena	Število iztočnic	Odstotek
5 – zaznani vsi pomeni	66	57
4	21	18
3	21	18
2	5	4
1	1	1
0 – zaznan ni noben pomen	2	2

Vir: lastno delo

Rezultati ocenjevanja splošne ustreznosti gesel, generiranih s ChatGPT, ki so prikazani v Tabeli 7, kažejo precejšnjo uporabnost modela, saj je skoraj 80 odstotkov gesel dobilo povprečno oceno 3,5 ali več in 19 odstotkov najvišjo oceno obeh ocenjevalcev. Izračun strinjanja ocenjevalcev sicer kaže na relativno nizko ujemanje (Krippendorff’s Alpha²⁵ (Ordinal Scale): 0.497, Bootstrap Confidence Interval (95% CI): [0.328, 0.631]), vendar pa je pri šeststopenjski lestvici in številnih elementih ocenjevanja to pričakovano.

Tabela 7: Splošna ustreznost gesel, generiranih s ChatGPT

Povprečna ocena	Število iztočnic	Odstotek
5	22	19
4,5	24	20,7
4	24	20,7
3,5	22	19
3	6	5,2
2,5	7	6
2	8	6,9
1	1	0,8
0	2	1,7

Vir: lastno delo

25 Gabriele Marzi, Marco Balzano in Davide Marchiori, »K-Alpha Calculator – Krippendorff’s Alpha Calculator: A User-Friendly Tool for Computing Krippendorff’s Alpha Inter-Rater Reliability Coefficient,« *MethodsX* 12 (2024), 102545, pridobljeno 20. 5. 2025, <https://doi.org/10.1016/j.mex.2023.102545>.

Primeri gesla s povprečno oceno 5:

bičati

- o fizičnem udarjanju z bičem; Kadar kdo biča nekoga, ga večkrat udari z bičem, šibo ali palico, da mu povzroči bolečino ali poškodbo.
- o kritiziranju ali obsodbi; Kadar kdo biča napake, grehe ali oblast, ostro kritizira in obsoja ti dve ali osebe.
- o udarjanju vremena ali naravnih elementov; Ko naravni elementi, kot sta veter in dež, bičajo nekaj, to pomeni, da silovito udarjajo nanj.

Primer gesla s povprečno oceno 3:

vešč

- Oseba ali način izvedbe, ki obvlada znanje ali spretnosti; Kdor je vešč česa, ima znanje in sposobnosti, ki mu omogočajo, da nekaj dela dobro.
- Oseba, ki je seznanjena s čim ali ima potrebna sredstva. Če si vešč nečesa, si s tem seznanjen ali imaš potrebna sredstva, znanje ali pooblastila za uspešno izvedbo določenega dejanja.

Primer gesla s povprečno oceno 2:

zatrjevati

- dati vtis ali trditi brez dejanske osnove; Ko kdo zatrjuje nekaj, daje vtis, da ima določene lastnosti ali trdi nekaj, kar je pogosto na neprepričljiv ali neresničen način.
- formalno predstaviti argumente ali trditve v pravnem kontekstu; Ko kdo v pravnem kontekstu zatrjuje nekaj, formalno predstavi svoje argumente ali trditve pred sodiščem ali v uradnih dokumentih.
- jasno trditi ali potrjevati nekaj; Ko kdo zatrjuje, potrjuje ali jasno izraža neko trditev ali prepričanje, pogosto z namenom, da prepriča druge, da je to resnica.
- potrditi brez dvoma; Ko nekdo zatrjuje nekaj, to pomeni, da to izjavi z gotovostjo in brez dvoma, pogosto kot dejstvo.
- zagotoviti ali izjaviti nasprotje nečesa drugega; Ko kdo zatrjuje nasprotno, trdi nekaj, kar je v nasprotju z obstoječimi informacijami ali prepričanji.

Uporabnost modela pri generiranju definicij potrjujejo tudi rezultati analize splošne ustreznosti generiranih definicij, ki je podana v Tabeli 8. Kot lahko vidimo, je skoraj 90 odstotkov gesel za ustreznost definicij dobilo povprečno oceno 3 ali več. Strinjanje ocenjevalcev je bilo tokrat še nekoliko nižje kot pri ocenjevanju splošne ustreznosti gesel (Krippendorff's Alpha (Ordinal Scale): 0.395, Bootstrap Confidence Interval (95 % CI): [0.226, 0.547]).

Tabela 8: Splošna ustreznost definicij, generiranih s ChatGPT

Povprečna ocena	Število iztočnic	Odstotek
5	12	10,4
4,5	28	24,1
4	26	22,4
3,5	19	16,4
3	18	15,5
2,5	5	4,3
2	3	2,6
1,5	3	2,6
0	2	1,7

Vir: lastno delo

Problemi generiranih podatkov

Pri analizi smo opazili nekatere ponavljajoče probleme v generiranih podatkih. Med najbolj opaznimi je težnja ChatGPT k pretiranemu drobljenju pomenov. Vendar pa statistični podatki kažejo nekoliko drugačno sliko, saj je samo pri 25 odstotkih iztočnic (29) ChatGPT zgeneriral več pomenov, kot jih najdemo v DSBS (pri tem je zanimiv podatek, da je med njimi kar osem od desetih iztočnic v DSB enopomenskih). Podrobnejša analiza je namreč pokazala, da so bili pri številnih iztočnicah, tudi ko je ChatGPT generiral enako število ali manj pomenov, kot jih je v DSB, generirani pomeni prekrivni ali skoraj identični (pri čemer so lahko nekateri drugi pomeni manjkali), kot denimo pri iztočnici *civilist*, kjer je ChatGPT zgeneriral sledeča pomena:

- človek, ki ni vojak; Civilist je človek, ki ni član oboroženih sil.
- oseba zunaj vojske; Civilist je oseba, ki ne pripada vojaškim silam in ne sodeluje v vojaških dejavnostih.

V drugih primerih je bila pomenska členitev prepodrobna, na primer definicije so pojasnjevale posamezno kolokacijo oziroma njihovo rabo. Dober primer tega je pridevnik *kamelji*, ki ima v DSB štiri pomene, ChatGPT pa jih je zgeneriral kar deset (Tabela 9).

Tabela 9: Pomenska členitev za *kamelji*, kot jo je generiral ChatGPT

Št. pomena	Generirani pomenski indikator	Generirana slovarska definicija
1	o barvi ali odtenku;	Barva ali odtenek, ki je kamelji, je podobna barvi kamelje dlake.
2	o dlaki ali volni;	Kar je kamelje, je narejeno iz dlake kamele ali jih kako drugače vključuje.
3	o festivalih ali dirkah;	Dogodek, ki je kamelji, vključuje kamelje dirke ali festivale, povezane s kamelami.
4	o glavi ali kosteh;	Predmeti, ki so kamelji, so narejeni iz delov telesa kamele, kot so glave ali kosti.
5	o hrani in pijači;	Kadar je hrana ali pijača kamelja, je narejena iz kameljih sestavin ali se nanaša na kamele.
6	o hrbtih;	Kar je kamelje, se nanaša na hrbte kamel in njihov specifičen videz.
7	o mladih ali drugih lastnikih;	Kar je kamelje, se nanaša na mlade kamele ali ljudi, ki skrbijo za kamele.
8	o prevozu in potovanjih;	Prevoz, ki je kamelji, vključuje uporabo kamel kot prevozno sredstvo ali se nanaša na relacije povezane z njimi.
9	o trgovini ali sejmu;	Kadar je nekaj kamelje, je povezano z dejavnostmi, povezanimi s prodajo kamel.
10	o usnju ali koži;	Kar je kamelje, je narejeno iz usnja ali kože kamele.

Vir: lastno delo

Druga pogosta težava so bili neprepoznani pomeni, med katerimi gre pogosto za preneseno rabo. Na primer pri *deževati* je ChatGPT zaznal pomen ‘vremenski pojav’ in pomen ‘padanje predmetov’, ne pa tudi pomena ‘nenadna pojavitev velike količine česa’ (na primer *Pritožbe in grožnje zdaj dežujejo z vseh strani.*).

Kot pogosta težava se je izkazala tudi umestitev zgledov pod pomene, kar je bilo v številnih primerih posledica (hkratne) neustreznosti pomenske členitve oziroma prekrivnosti generiranih pomenov. Tu smo prepoznali tako umestitev zgledov pod napačne pomene kot tudi podvajanje pri umeščanju, tj. umestitev istega zgleda pod več kot en pomen.

Čeprav je bila ubeseditve definicij glede na navodila²⁶ celostno gledano precej ustrezna, pa smo vseeno zaznali kar nekaj primerov problematične ubeseditve. Po eni

26 Gantar, *Leksikografski opis slovenščine v digitalnem okolju*. Ustreznost definicij smo ocenjevali skladno s slovarskimi navodili, ki smo jih oblikovali pri izdelavi LBS, in sicer smo definicije opredelili ločeno za posamezno besedno vrsto, pri čemer smo za glagolske pomene preferirali navedbo stavčne definicije, ki naj vključuje vse ključne skladenjsko-pomenske elemente posameznega pomena, tj. udeležence in okoliščine kot tudi konotativne in pragmatične pomenske elemente.

strani take definicije niso sledile vzorcem iz sistemskih navodil, bile so tudi predolge ali zelo kratke. Nekatere so vsebovale slovnične ali skladenjske napake. Naleteli smo tudi na nekaj primerov neustreznih oziroma napačnih definicij, na primer za pomen samostalnika *kajak*: *Ko govorimo o izposoji ali najemu **kajakov**, mislimo na možnost, da plovilo najdemo za določeno obdobje proti plačilu.*

Med redkejšimi težavami smo zaznali uporabo angleških indikatorjev in generiranja povsem napačne definicije (za neko drugo iztočnico).

Problemi vhodnih podatkov

Analiza rezultatov je pokazala tudi nekatere pomanjkljivosti vhodnih podatkov, ki so lahko privedli do slabših rezultatov pri pomenskem členjenju (oblikovanju definicij, razporeditvi kolokacij in zgledov ipd.). Predvsem gre tu za pomensko ustreznost in kakovost avtomatsko pridobljenih zgledov. Na primer, pri večpomenskih kolokacijah, ki se potrjujejo z zgledi, se lahko zgodi, da jih večina potrjuje samo določen pomen. Za drugi pomen, ki mu pripada enaka kolokacija, pa zgledov ni ali pa jih je malo. Primer tega je samostalnik *agonija* in večpomenska raba številnih kolokacij (*huda agonija, mučna agonija, podaljšati agonijo* ipd.), pri čemer smo za kar 33 zgledov kolokacij pri pomenu 'umiranja' ugotovili, da spadajo v pomen 'težavnega obdobja' (na primer *Obljube se niso izpolnile, letališče pa je zapadlo v še hujšo agonijo.*). To je pomenilo, da je imel model za pomen 'umiranja' na voljo zelo malo zgledov. Povezana težava je slaba kakovost nekaterih zgledov, predvsem smo zaznali težavo pomanjkljivega konteksta ali referenta, na katerega se definicija nanaša. Nekaj primerov:

- 1. Sodeč po fotografijah, ki jih prejemamo v uredništvu, so že lepo **košati**.
- 2. Nekatere so čisto **benigne**, nekatere pa ogrožajo celo človeška življenja.
- 3. Označujeta jo izviren izraz in bogata **barvitost**.
- 4. So tudi raj za prave **gurmane**.

Ena od pomanjkljivosti, ki je vplivala na število zaznanih pomenov pri sicer le nekaj iztočnicah, je bila zastopanost pomenov v vhodnih podatkih. V nekaterih primerih namreč pomen v DSBS še ni imel pripisanih kolokacij (mogoče so bili v bazi samo zgledi), zato ga tudi ni bilo mogoče vključiti v vhodne podatke za ChatGPT.

Glede definicij, ki smo jih kot vhodne podatke vzeli iz slovenskega Wordneta in slovarja Bridge, lahko rečemo, da so bile ne glede na morebitno prekrivnost in deloma slabšo kakovost pogosto v pomoč, saj jih je v številnih primerih ChatGPT uporabil dobesedno. Pri iztočnicah, za katere nismo imeli na voljo definicij, nismo opazili izstopajočih značilnosti, saj so bila generirana gesla različnih ocen splošne ustreznosti, je pa mogoče pomenljiv podatek, da sta bila med njimi obe problematični iztočnici s popolnoma napačnimi podatki (*bakren* in *padalski*).

Sklep in nadaljnje delo

V raziskavi smo preverili, kako uspešen je ChatGPT pri umeščanju sopomenskega gradiva pod besedne pomene in pri generiranju slovarskih gesel. Analizirali smo rezultate razvrščanja 951 sopomenskih kandidatov za 246 slovarskih iztočnic ter kakovost generiranih pomenskih členitev in definicij za 116 iztočnic.

Pri prvem poizkusu je strojni postopek v 41,9 odstotka primerov vrnil rezultate, povsem skladne s slovaropisnimi. Pri drugih 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 odstotka analiziranih iztočnic. Večina odstopanj je posledica popustljivosti sistema do sopomenskih kandidatov, ki jih je slovaropisna ekipa odstranila. Ker koncept SSSS načelno teži k širokemu vključevanju gradiva, slovarski vmesnik pa omogoča odziv uporabniške skupnosti na neustrezne kandidate, so ti odstopi manj problematični. V 28,9 odstotka 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 kratkosti oziroma specifične vloge indikatorjev znotraj DSBS, pa tudi specifik korpusnega gradiva, ki v slovaropisnih delotokih pogojuje pomensko členjenje in preverbo sopomenskosti. Natančnejši pregled primerov, v katerih se pojavljajo različna odstopanja, pokaže, da se ChatGPT tudi pri najtežjih primerih ne razlikuje radikalno od slovaropisne presoje, 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.

V drugem preizkusu smo testirali zmožnost ChatGPT za samostojno izdelavo slovarskih gesel. Analiza kakovosti generiranih gesel kaže, da je sistem zaznal vse pomene v 57 odstotkih primerov, skoraj 80 odstotkov generiranih gesel je doseglo povprečno oceno 3,5 ali več, 19 odstotkov pa najvišjo oceno obeh ocenjevalcev. Pri generiranju pomenov se je kot težava izkazala pretirana granularnost, zlasti kot posledica ponovljenih ali pretirano podrobnih pomenov. Med problematične vidike spadajo tudi neprepoznane prenesene rabe ter težave pri razvrščanju zgledov pod ustrezne pomene. Pri ocenjevanju definicij smo ugotovili, da so bile nekatere neustrezno oblikovane ali premalo informativne, druge so vsebovale slovnične napake, v redkih primerih pa so bile generirane definicije povsem napačne. Kljub temu so v večini primerov definicije sledile pričakovanim smernicam in so bile ocenjene kot uporabne. Podatki kažejo, da so bili rezultati zanesljivejši, kadar so bili vhodni podatki bogatejši, še posebej v primerih, kjer so bile na voljo kakovostne kolokacije in zgledi.

Eden izmed ključnih izzivov obeh preizkusov je nepredvidljivost postopka. ChatGPT kot generativni model ne deluje po strogo določenih pravilih strojnega procesiranja podatkov, kar pomeni, da rezultati niso nujno ponovljivi ali povsem razločljivi. Ta značilnost pomembno omejuje domet evalvacijskih raziskav, kot je

naša, ne more pa biti razlog, da generativnih tehnologij v slovaropisju ne bi uporabljali in/ali ocenjevali.

Prvi korak za nadaljnje delo je s preizkušeno metodologijo pripraviti nove rezultate in testirati, ali delo s strojno predpripravo slovaropisne odločitve dejansko pohitri ali ne. Ker je strojni postopek, ki ga preizkušamo, odvisen od izbranega poziva, vhodnih podatkov in različice uporabljenega sistema, je raziskavo mogoče ponoviti na zmogljivejših različicah ChatGPT ali drugih podobnih sistemih, z nadgrajenimi pozivi in na novem gradivu (na primer za razvrščanje uporabniško dodanih sopomenk ali protipomenk). Jezikoslovno preglednejše in jasnejše rezultate bi lahko dobili, če bi se omejili na homogeno gradivo, denimo celoten razred vrstnih pridevnikov tipa *kame-lji* (tudi *slonji*, *krokodilji* itd.) ali glagolov s primerljivimi vezljivostnimi značilnostmi. Preizkusiti pa 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. V širšem smislu bi bilo zanimivo raziskavam, ki preverjajo razumevanje koncepta sopomenskosti med različnimi uporabniškimi skupinami slovarja,²⁷ dodati še »razumevanje« pri rabi ChatGPT oziroma umetne inteligence.

Zahvala

Raziskovalna programa št. P6-0411 (*Jezikovni viri in tehnologije za slovenski jezik*) in št. P6-0215 (*Slovenski jezik – bazične, kontrastivne in aplikativne raziskave*) ter raziskovalni projekt *Veliki jezikovni modeli za digitalno humanistiko* (GC-0002) sofinancira Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije iz državnega proračuna.

Viri in literatura

Literatura

Angleško-slovenski slovar Bridge. 2000. Ljubljana: Državna založba Slovenje.

Arhar Holdt, Špela, Jaka Čibej, Kaja Dobrovoljc, Polona Gantar, Vojko Gorjanc, Bojan Klemenc, Iztok Kosem, Simon Krek, Cyprian Laskowski in Marko Robnik-Šikonja. »Thesaurus of Modern Slovene: By the Community for the Community.« V: *Proceedings of the XVIII EURALEX International Congress, Lexicography in Global Contexts, 17–21 July 2018, Ljubljana*, uredili Jaka Čibej, Vojko Gorjanc, Iztok Kosem in Simon Krek, 401–10. Ljubljana: Znanstvena založba Filozofske fakultete, 2018. Pridobljeno 20. 5. 2025. <https://doi.org/10.4312/9789610600961>.

27 Gapsa, »But why?? Evaluation of User-Suggested Synonyms in the Thesaurus of Modern Slovene.«

- Arhar Holdt, Špela, Polona Gantar, Iztok Kosem, Eva Pori, Marko Robnik Šikonja in Simon Krek. »Thesaurus of Modern Slovene 2.0.« V: *Electronic Lexicography in the 21st Century (eLex 2023), Proceedings of the eLex 2023 Conference, 27–29 June 2023*, uredili Marek Medved, Michal Měchura, Carole Tiberius, Iztok Kosem, Jelena Kallas, Miloš Jakubiček in Simon Krek, 366–81. Brno: Lexical Computing CZ, 2023. Pridobljeno 20. 5. 2025. <https://elex.link/elex2023/wp-content/uploads/82.pdf>.
- de Schryver, Gilles-Maurice. »Generative AI and Lexicography: The Current State of the Art Using ChatGPT.« *International Journal of Lexicography* 36, št. 4 (2023): 355–87. Pridobljeno 20. 5. 2025. <https://doi.org/10.1093/ijl/ecad021>.
- Gantar, Polona. *Leksikografski opis slovenščine v digitalnem okolju*. 1. izd., elektronska izd. Ljubljana: Znanstvena založba Filozofske fakultete, 2015. Zbirka Sporazumevanje. Pridobljeno 20. 5. 2025. <https://doi.org/10.4312/9789612377922>.
- Gantar, Polona, Špela Arhar Holdt, Iztok Kosem in Simon Krek. »Sopomenke 2.0 in Kolokacije 2.0: Novi koraki za slovenske odzivne slovarje.« *Jezik in slovstvo* 68, št. 4 (2023): 157–75. Pridobljeno 20. 5. 2025. <https://doi.org/10.4312/jis.68.4.157-175>.
- Gapsa, Magdalena, Špela Arhar Holdt in Iztok Kosem. »Kako dober je ChatGPT pri umeščanju sopomenk pod besedne pomene.« V: *Jezikovne tehnologije in digitalna humanistika: Zbornik konference, 19.–20. september 2024, Ljubljana, Slovenija*, uredila Špela Arhar Holdt in Tomaž Erjavec, 144–62. Ljubljana: Inštitut za novejšo zgodovino, 2024. Pridobljeno 20. 5. 2025. <https://zenodo.org/records/13912515>.
- Gapsa, Magdalena. »But why?? Evaluation of User-Suggested Synonyms in the Thesaurus of Modern Slovene.« *Lang Resources & Evaluation* (2025). Pridobljeno 20. 5. 2025. <https://doi.org/10.1007/s10579-025-09821-8>.
- Jakubiček, Miloš in Michael Rundell. »The End of Lexicography? Can ChatGPT Outperform Current Tools for Post-Editing Lexicography?« V: *Electronic Lexicography in the 21st Century (eLex 2023): Proceedings of the eLex 2023 Conference*, uredili Marek Medved, Michal Měchura, Carole Tiberius, Iztok Kosem, Jelena Kallas, Miloš Jakubiček in Simon Krek, 518–33. Lexical Computing CZ, 2023. Pridobljeno 20. 5. 2025. <https://elex.link/elex2023/wp-content/uploads/102.pdf>.
- Kosem, Iztok, Simon Krek in Polona Gantar. »Semantic Data Should No Longer Exist in Isolation: The Digital Dictionary Database of Slovenian.« V: *EURALEX XIX: Congress of the European Association for Lexicography, Lexicography for Inclusion, 7–9 September 2021, Virtual, Book of Abstracts*, uredili Zoe Gavrilidou, Lydia Mitits in Spyros Kiosses, 81–83. Democritus University of Thrace, 2021. Pridobljeno 20. 5. 2025. <https://euralex.org/wp-content/uploads/2022/04/ABS2020.pdf>.
- Kosem, Iztok, Husak, Miloš in McCarthy, Diana. »GDEX for Slovene.« V: *Electronic Lexicography in the 21st Century: New Applications for New Users: Proceedings of eLex 2011, 10–12 November 2011, Bled, Slovenia*, uredila Iztok Kosem in Karmen Kosem, 150–159. Ljubljana: Trojina, Institute for Applied Slovene Studies, 2011. Pridobljeno 20. 5. 2025. http://www.trojina.si/elex2011/elex2011_proceedings.pdf.
- Krek, Simon, Cyprian Laskowski in Marko Robnik-Šikonja. »From Translation Equivalents to Synonyms: Creation of a Slovene Thesaurus Using Word Co-occurrence Network Analysis.« V: *Electronic Lexicography in the 21st Century. Proceedings of eLex 2017 Conference: Lexicography from Scratch*, uredili Iztok Kosem, Carole Tiberius, Miloš Jakubiček, Jelena Kallas, Simon Krek in Vit Baisa, 93–109. Leiden: Dutch Language Institute, Lexical Computing CZ s.r.o., Trojina, 2017. Pridobljeno 20. 5. 2025. <https://elex.link/elex2017/wp-content/uploads/2017/09/paper05.pdf>.
- McKean, Erin in Will Fitzgerald. »The ROI of AI in Lexicography.« *Lexicography* 11, št. 1 (2024): 7–27. Pridobljeno 20. 5. 2025. <https://utppublishing.com/doi/abs/10.1558/lexi.27569>.
- Lew, Robert. »ChatGPT as a COBUILD Lexicographer.« *Humanities and Social Sciences Communications* 10 (2023), Article 704. Pridobljeno 20. 5. 2025. <https://doi.org/10.1057/s41599-023-02119-6>.

- Marzi, Gabriele, Marco Balzano in Davide Marchiori. »K-Alpha Calculator—Krippendorff's Alpha Calculator: A User-Friendly Tool for Computing Krippendorff's Alpha Inter-Rater Reliability Coefficient.« *MethodsX* 12 (2024), 102545. Pridobljeno 20. 5. 2025. <https://doi.org/10.1016/j.mex.2023.102545>.
- Ptasznik, Bartosz, Sascha Wolfer in Robert Lew. »A Learners' Dictionary versus ChatGPT in Receptive and Productive Lexical Tasks.« *International Journal of Lexicography* 37, št. 3 (2024): 322–36. Pridobljeno 20. 5. 2025. <https://doi.org/10.1093/ijl/eca011>.
- Rundell, Michael. »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, 9–17. Yonsei University, 2023. Pridobljeno 20. 5. 2025. <https://www.asialex.org/pdf/Asialex-Proceedings-2023.pdf>.
- Tiberius, Carole, Kris Heylen, Jesse de Does, Bram Vanroy, Vincent Vandeghinste in Job van Doeselaar. »LLMs and Evidence-based Lexicography.« V: *Large Language Models and Lexicography, Book of Abstracts, 8th October 2024, Cavtat, Croatia*, uredil Simon Krek, 44–48. 2024. Pridobljeno 25. 1. 2025. https://www.cjvt.si/wp-content/uploads/2024/10/LLM-Lex_2024_Book-of-Abstracts.pdf.

Spletni viri

- Čibej, Jaka, Luka Terčon, Simon Krek, Andraž Repar, Erik Novak, Polona Gantar, Iztok Kosem, Špela Arhar Holdt, Kaja Dobrovoljc, Amadea Berginc, Irena Hvala, Damijan Klement, Manja Kolenc, Ana Močnik, Tina Munda, David Pavlas, Anamari Pečan, Aleksandra Poljak, Davorin Sečnik, Jure Šešet, Jan Štumberger, Tina Toličič in Laura Trpin. *Open Slovene WordNet OSWN 1.0*. Slovenian language resource repository CLARIN.SI, 2023. Pridobljeno 20. 5. 2025. <http://hdl.handle.net/11356/1888>.
- Kosem, Iztok, Špela Arhar Holdt, Simon Krek, Polona Gantar, Eva Pori, Urška Kamenšek, Primož Ponikvar, Rebeka Roblek, Jure Šešet, Petra Zaranšek, Karolina Zgaga, Jaka Čibej, Bojan Klemenc, Cyprian Laskowski, Kaja Dobrovoljc, Vojko Gorjanc in Nikola Ljubešić. *Kolokacijski slovar sodobne slovenščine*. Ljubljana: Znanstvena založba Filozofske fakultete, 2018–. Pridobljeno 20. 5. 2025. <https://viri.cjvt.si/kolokacije/slv/#>.
- Krek, Simon, Cyprian Laskowski, Marko Robnik-Šikonja, Iztok Kosem, Špela Arhar Holdt, Polona Gantar, Jaka Čibej, Vojko Gorjanc, Bojan Klemenc in Kaja Dobrovoljc. *Thesaurus of Modern Slovene 1.0*. Repozitorij raziskovalne strukture CLARIN.SI, 2018. Pridobljeno 20. 5. 2025. <http://hdl.handle.net/11356/1166>.
- Krek, Simon, Cyprian Laskowski, Marko Robnik-Šikonja, Iztok Kosem, Špela Arhar Holdt, Polona Gantar, Jaka Čibej, Vojko Gorjanc, Bojan Klemenc, Kaja Dobrovoljc, Eva Pori, Rok Roblek in Klemen Zgaga. *Thesaurus of Modern Slovene 2.0*. Repozitorij raziskovalne strukture CLARIN.SI, 2023. Pridobljeno 20. 5. 2025. <http://hdl.handle.net/11356/1916>.
- OpenAI. »ChatGPT (veliki jezikovni model).« Pridobljeno 31. 5. 2024. <https://chatgpt.com>.

**Špela Arhar Holdt, Magdalena Gapsa,
Polona Gantar, Iztok Kosem**

THE POTENTIAL OF CHATGPT IN THE DEVELOPMENT OF THE THESAURUS OF MODERN SLOVENE

SUMMARY

This study examines the potential of ChatGPT-4 to support lexicographic work by evaluating its performance in two tasks: filtering and assigning synonym candidates to their corresponding lexical senses, and generating complete dictionary entries, including sense distinctions, definitions, and usage examples. The evaluation is based on a comparison with expert lexicographic decisions recorded in the Digital Dictionary Database for Slovene. The goal is to determine how closely ChatGPT's outputs align with established lexicographic practices and to explore whether the model can reliably contribute to streamlining dictionary compilation. By assessing the accuracy and utility of the generated content, the research aims to clarify the practical role large language models might play in digital lexicography.

In the first experiment, ChatGPT processed 951 synonym candidates across 246 dictionary entries. The model's decisions fully matched those of the lexicographers in 41.9 % of the cases, while in 58.1 % of the cases, it made different choices. A key finding was that ChatGPT was more permissive in retaining synonym candidates that experts had excluded. In 14.6 % of the entries, synonyms were assigned to different senses than in the gold standard, and in 19.9 %, expected synonym placements were missing. These differences often stemmed from the complexity of the entries and the brevity or ambiguity of semantic indicators. Despite these issues, the system's performance suggests that it could serve as a valuable tool for the preliminary classification of synonyms, supporting rather than replacing human judgment.

The second experiment assessed ChatGPT's ability to generate complete dictionary entries for 116 headwords without human input. The model correctly identified all lexical senses in 57 % of cases. Nearly 80 % of the entries received an average quality rating of 3.5 or above, while 19 % were given the highest score by both evaluators. However, several challenges were noted, including excessive granularity in sense division, a tendency to overlook figurative meanings, and occasional mismatches between definitions and examples. Some definitions lacked precision or included minor grammatical errors, though most adhered to conventional lexicographic norms. The quality of the outputs was notably higher when the input data included clear collocations and illustrative examples, confirming the importance of structured input for effective generative processing.

A central challenge across both tasks is the unpredictability inherent in generative models such as ChatGPT. Because the model's outputs are not deterministic, results are not always repeatable or easily interpretable, complicating evaluation and

integration into structured editorial workflows. Nevertheless, the findings demonstrate that with proper monitoring and refinement, ChatGPT has real potential to accelerate routine lexicographic tasks. Future work should explore more advanced model versions, improved prompt engineering, and broader applications such as sorting user-submitted content or generating lexical suggestions. With appropriate methodology, ChatGPT could become a valuable tool in lexicography, complementing expert work with increased speed and additional insights.

PRILOGA 1: Poziv za selekcioniranje sopomenk in razvrščanje pod pomene

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], 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

- 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]

PRILOGA 2: Poziv za generiranje slovarskega gesla

You are a lexicographer. The definitions should be written in Slovene, in full-sentence form, as in Cobuild-style.

Formats used can be \“Kdor je <adjective>, je...\\“ or \“Kar je <adjective>, je...\\“ or \“Kadar je kdo <adjective>, je...\\“ but other formats can be used. Examples of good definitions:

Žival, ki je <HH>amfibijska</HH>, lahko živi tako na kopnem kot v vodi.

Kar je <HH>krvavo</HH>, je prekrito s krvjo.

Kadar je človek <HH>zaskrbljen</HH>, je zaradi nečesa živčen ali v skrbeh.

Z besedo <HH>oren</HH> opisujemo tisto, kar je povezano s pridelovanjem poljščin.

Kdor je <HH>strasten</HH> do nečesa, je za to zelo navdušen ali vnet.

Vozilo, ki je <HH>blindirano</HH>, ima trd oklep, ki potnike varuje pred morebitnimi streli in izstrelki.

Podjetje, ki je <HH>multinacionalno</HH>, ima podružnice v številnih državah.

Za človeka rečemo, da je <HH>odbijajoč</HH>, kadar se nam zdi neprijeten in ga ne želimo bolje spoznati.

Kar je <HH>oljnato</HH>, je prekrito z oljem ali ga vsebuje.

The output should follow this format:

1. short indicator

Full sentence-definition

Numbers of collocations + examples.

Example for the adjective \“prostaški\\“:

1. o komunikaciji

Govorica, ki je prostaška, vsebuje kletvice ali neotesane besede.

(5), (8), (14)