

Definiranje značenja bosanskih frazema kroz prizmu vještačke inteligencije

Amela Šehović

Univerzitet u Sarajevu, Filozofski fakultet, Odsjek za bosanski, hrvatski i srpski jezik

Predmet istraživanja u radu je definiranje značenja bosanskih frazema koje vrši ChatGPT, *chatbot* vještačke inteligencije. U tu svrhu prikupljen je korpus od 50 frazema – poredbenih i neporedbenih – čija je obavezna sastavnica leksema orijentalnog porijekla (iz arapskog, perzijskog ili turskog jezika). Analizom se želi utvrditi koliko prisustvo orijentalizma u frazemi doprinosi ili odmaže u definiranju značenja frazeme, s naročitim akcentom na orijentalizme specifične za bosanski jezik i one koji su općeprihvaćeni i zastupljeni i u hrvatskom i srpskom jeziku. Pored značenja frazema, analiziran je i kontekst njihove upotrebe radi utvrđivanja uspješnosti vještačke inteligencije da ponudi adekvatne primjere.

Ključne riječi: vještačka inteligencija, ChatGPT, frazema, orijentalizam, bosanski jezik

Research conducted on a corpus of 50 comparative and non-comparative phrasemes containing lexemes of oriental origin has shown that the quality of definitions provided by artificial intelligence (AI) is high for phrasemes with generally accepted and widely used orientalisms (70%) and for comparative phrasemes with orientalisms of transparent origin (80%). However, this result is overall poorer for phrasemes with orientalisms of transparent origin that are specific to the Bosnian language (56.66%). Nevertheless, when these findings are supplemented with results on partial semantic matches of Bosnian language phrasemes, a high level of accuracy in their definitions can be established. In addition, an adequate definition of phraseme meanings is accompanied by a suitable context of their usage. However, other parameters, such as glosses, synonyms, etc., lag significantly behind the meaning of phrasemes and the context of their usage, requiring increased caution from lexicographers and dictionary users when relying on this information.

Keywords: artificial intelligence, ChatGPT, phraseme, orientalism, Bosnian language

1. Uvod

Vještačka inteligencija (VI) dobila je svoju primjenu u različitim aspektima ljudskog života. Kada je 30. novembra 2022. uveden ChatGPT, *chatbot* vještačke inteligencije, niko nije ni slutio da će ga milioni korisnika prihvatiti tako brzo – za nepuni mjesec već ga je koristilo 100 miliona korisnika (De Schryver 2023: 355), a danas je sve prisutniji i u jezičkim istraživanjima. ChatGPT je zasnovan na velikom jezičkom modelu (LLM) i »može se posmatrati kao sistem koji, kako se čini, kombinira sve alate dosad proizvedene za leksikografiju u jedan« (Jakubiček i Rundell 2023: 518). Mišljenja lingvisti i lingvistice o opravdanosti njegove upotrebe vrlo su raznorodna, što je vidljivo kada se prate debate među leksikografima i leksikografkinjama o tome da li upotreba ChatGPT-ja znači kraj leksikografije, tj. da li su leksikografi uopće više potrebni¹, i da li je ChatGPT zaslužio da se posmatra kao »osposobljeni leksikograf« (engl. *fully-fledged lexicographer*) (više o tome u De Schryver 2023: 356–357), naročito u svjetlu činjenice da je izrada kvalitetnog rječnika dugotrajan i višestruko (finansijski, stručno itd.) zahtjevan proces (više o tome u Lew 2023:1).

Pri tome se zaboravlja da je leksikografija jedna od prvih jezičkih disciplina koja je odranije imala koristi od npr. automatizacije jezičkih podataka (u procesu nastajanja rječnika), te je bilo očekivano da se taj proces automatizacije u generiranju jezičkih informacija nastavi i dalje. U tom kontekstu, postavlja se pitanje na što se on može primijeniti i koje su njegove granice. Svjesni višestrukih mogućnosti njegove primjene, u ovom radu opredijelili smo se da provjerimo stepen adekvatnosti u definiranju značenja bosanskih frazema koje nudi ChatGPT. Za definiranje značenja smo se opredijelili jer mnogi autori upravo ovaj segment smatraju »jednom od jačih tačaka ChatGPT-ja« (Jakubiček i Rundell 2023: 526), odnosno navode da su definicije koje generira ChatGPT dobre do veoma dobre (Schryver 2023: 371, Lew 2023: 4, Gantar 2024: 151), a za frazeme smo se opredijelili jer ih, za razliku od vezanih izraza, odlikuje transponirano značenje, koje ne proizlazi iz značenja pojedinih sastavnica te ga je samo po sebi teže prepoznati. Ono je prema našem mišljenju (v. i Šehović i Haverić 2017: 51) ključna karakteristika frazema, zbog čega je u fokusu analize ovog rada sposobnost vještačke inteligencije da pravilno generira to značenje.

1 Zapitanost nad budućnošću vlastite profesije u svjetlu sve naprednije vještačke inteligencije glavna je tema i knjige *Moj algoritam i ja* D. Kelmana, u kojoj se istražuje mogućnost vještačke inteligencije da u koautorstvu s piscem napiše priču.

2. Metodologija i ciljevi istraživanja

Kako bi se ovo istraživanje moglo provesti, najprije je trebalo odrediti korpus, a upravo je određivanje granica korpusa bio najveći metodološki problem. Prije svega, frazeme su morale biti zabilježene u nekom rječniku bosanskog jezika (frazološkom ili opisnom) kao preduslov da se opisano značenje može porediti sa značenjem frazema generiranih uz pomoć vještačke inteligencije. U radu je korišten rječnik iz knjige *Leksika orijentalnog porijekla u frazemama bosanskog jezika* A. Šehović i Đ. Haverić (2017), koji je u nastavku označen skraćenicom LOP. Imajući u vidu da je transponiranost značenja smatrana ključnom karakteristikom frazema, one su podijeljene u dvije grupe: poredbene i neporedbene frazeme. Razlog ovoj prvoj podjeli je taj što poredbene frazeme općenito sadrže manji stepen transponiranosti značenja u odnosu na neporedbene, zbog čega ih neki autori i ne smatraju frazemama (v. Kasumović i Nikolić 2018), a upravo zbog manjeg stepena transponiranosti njihova značenja moglo se očekivati da će to značenje ChatGPT bolje prepoznati te smo provjeru te hipoteze i postavili kao jedan od istraživačkih ciljeva u ovom radu.

Sljedeći kriterij za koji smo se opredijelili bio je orijentalno porijeklo makar jedne lekseme u analiziranim frazemama. Naime, lekseme orijentalnog porijekla (arabizmi, farsizmi i turcizmi) pripadaju karakterističnoj leksici bosanskog jezika, pa se mogu posmatrati i kao njegovi markeri, odnosno elementi specifični za bosanski jezik i kulturu zajednice u kojoj se on koristi. U tom smislu, provjera mogućnosti vještačke inteligencije da prepozna takve elemente jedan je od ciljeva našeg istraživanja. Ove lekseme u velikoj mjeri pripadaju razgovornom stilu bosanskog jezika, a mnoge od njih su i izrazite ekspresivnosti te doprinose promjeni intenziteta značenja pojma koji označavaju (o ovome više u Šehović i Haverić 2017), što se može provjeriti zamjenom ovih leksema sinonimnim leksemama neutralne markiranosti: *duvar* i *zid* npr. nemaju istu ekspresivnost.

Ipak, neki od ovih orijentalizama transparentnog su porijekla čak i za prosječnog govornika bosanskog jezika: *abuzemze*, *Ćaba*, *fildžan* i sl., dok su drugi neza-mjenjivi: *bubreg*, *čarapa* i sl. ili toliko općeprihvaćeni: *budala*, *šamar*, *taman* i sl. da i ne postoji svijest govornika o njihovom orijentalnom porijeklu. Na temelju toga frazeme su podijeljene u dvije velike grupe: prvu čine one u kojima je orijentalizam transparentnog porijekla, a drugu frazeme s općeraširenim orijentalizmom netransparentnog porijekla.

Nakon određivanja korpusa, vještačkoj inteligenciji je za svaku frazemu pojedinačno postavljeno pitanje: Šta znači izraz ...? Odgovor vještačke inteligencije odmah

nas je suočio sa glavnim izazovom istraživanja – njenim mogućim višestrukim odgovorima na isto pitanje. Svjesni te činjenice, opredijelili smo se za prvi dobijeni odgovor kao najvalidniji jer svaki naredni ponovljeni upit može dovesti do nešto preciznije definicije značenja frazeme, a u našem je fokusu sposobnost vještačke inteligencije da pruži najbolju definiciju značenja frazeme uz što manje lutanja. Uostalom, korpus čini samo 50 frazema, zbog čega i ne može biti osnov za neke sveobuhvatne zaključke, ali svakako može biti osnov za neka buduća opsežnije provedena istraživanja ove teme.

Glavno pitanje na koje ovo istraživanje nastoji odgovoriti je sljedeće²: Da li je vještačka inteligencija sposobna pravilno definirati značenje neke frazeme, odnosno, da li to značenje korespondira sa njenim stvarnim značenjem, naročito u kontekstu kulturološke i pragmatičke perspektive. Iz tog razloga, praćen je i kontekst upotrebe za svaku od navedenih frazema.

Također, s obzirom na kriterij transparentnosti porijekla lekseme, nametnulo se pitanje i da li će transparentnost (orijentalnog) porijekla lekseme korespondirati sa sposobnošću vještačke inteligencije da prepozna značenje frazema ili će biti upravo obrnuto.

2.1. Hipoteze istraživanja

Ovako koncipirano istraživanje provedeno je polazeći od nekoliko hipoteza:

1. Značenje frazema sa orijentalizmom transparentnog porijekla specifičnim za bosanski jezik bit će neadekvatno prepoznato ili nedovoljno precizno definirano budući da je bosanski jezik općenito manje prisutan u elektronskim korpusima. Pretpostavka je da vještačka inteligencija primarno koristi ove izvore za definiranje značenja frazema, a kada oni nisu dostatni, ni značenje frazeme ne može biti precizno formulirano.
2. Značenje frazema sa orijentalizmima koji su za prosječnoga govornika netransparentnog porijekla bit će adekvatno formulirano ili općenito bolje prepoznato u poređenju s prethodnom grupom frazema. Razlog tome je što su ove frazeme i njihova značenja zabilježeni i u hrvatskim i srpskim leksikografskim izvorima, naročito elektronskim, te se može očekivati veći stepen njihovog prepoznavanja od strane vještačke inteligencije.

2 Ova su se pitanje nametnula i prilikom pripreme izlaganja Vještačka inteligencija i njena primjena u jezičkim istraživanjima za konferenciju Saveza za esperanto *Jezik i vještačka inteligencija*, koja je održana u februaru 2025. godine. To je bio i prvi put da sam se bavila tematikom vezanom za vještačku inteligenciju. No, korpus za ovaj rad gotovo je potpuno nov, zbog čega su se pojavila i neka nova pitanja.

3. Načelno, značenja poredbenih frazema bit će bolje prepoznata u poređenju s neporedbenim frazemama, a o razlozima te pojave govorili smo u uvodnom dijelu rada.
4. Frazeme s tačno definiranim značenjem bit će ilustrirane i dobro izabranim primjerima.

2.2. Korpus frazema i leksikografska analiza

Svi prethodni pripremni koraci ovog istraživanja doveli su do stvaranja korpusa frazema podijeljenog u dvije grupe: frazeme koje sadrže orijentalizam transparentnog porijekla i frazeme s općeraširenim orijentalizmom netransparentnog porijekla.

Frazeme s orijentalizmom transparentnog porijekla kao jednu sastavnicu sadrže:

- a) lekseme iz sakralnog registra islamske provenijencije: *abdest* (perz.), *abuzemze* (ar.-perz.), *ajet* (ar.), *Āba* (ar.), *dženet*(ski) (ar.), *hafiz* (ar.), *halal*(iti) (ar.), *hodža* (perz.), *hurija* (ar.), *ramazan* (ar.), *šamija* (perz.), *tabut* (ar.), *teravija* (ar.), *tevba/toba* (ar.);
- b) lekseme iz tematskog polja orijentalne hrane i pića te posuđa u kojem se takva hrana priprema: *boza* (perz.), *ćevap* (ar.), *fildžan* (perz.), *halva* (ar.), *pilav* (perz.), *tepsija* (tur.);
- c) historijske termine: *beg* (tur.), *paša* (tur.);
- d) lekseme iz tematskog polja stanja: *ćeif* (*ćejf*) (ar.), *merak* (ar.), *rahat* (ar.), *sevdah* (ar.);
- e) ostale lekseme: *akrep* (ar.), *babo* (perz.), *mahal*(uša) (ar.), *raja*.

Detaljan popis ovih frazema dat je u sljedećoj tabeli:

Tabela 1. Popis frazema s orijentalizmom transparentnog porijekla

Poredbene frazeme sa značenjima	Neporedbene frazeme sa značenjima
1P: (<i>biti</i>) <i>bistar k(a)o boza</i> (perz.) – iron. ne biti puno pametan, inteligentan, biti glup (LOP:126)	1N: <i>biber po pilavu</i> (perz.) – vrhunac čega (LOP:220)
2P: <i>čuvati/paziti k(a)o abuzemze</i> (ar.-perz.) – razg. čuvati s posebnom pažnjom i velikom ljubavlju (LOP:110)	2N: <i>biti za muštuluka</i> (perz.-tur.) – donijeti kome radosnu, povoljnu vijest, dobar glas (LOP:208)

3P: <i>gubiti se k(a)o abdest</i> (perz.) – razg. brzo, lahko se gubiti (LOP:110)	3N: <i>doči tobe</i> (ar.) – razg. 1. pokajati se za počinjena loša djela i prestati ih dalje činiti 2. opametiti se (LOP:238)
4P: <i>ide (prodaje se) k(a)o halva</i> (ar.) – razg. odlično se prodaje, ima veliku potražnju (LOP:169)	4N: <i>duša od insana</i> (ar.) – veoma dobra, plemenita osoba (LOP:177)
5P: <i>k(a)o da je hodži</i> (perz.) s <i>tabuta</i> (ar.) <i>pobjegao</i> – razg. veoma je blijed i iscrpljen (LOP:173)	5N: <i>halali, Huso, materi</i> – razg. pomiri se sa gubitkom, propalo je, ništa od toga (LOP:169)
6P: <i>kao da je krenuo/krenula na Ćabu</i> (ar.) – iron. pretjerano dugo se spremati (za neku posjetu, odlazak, putovanje i sl.) (LOP:138)	6N: <i>ima li bujruma?</i> (tur.) – upotrebljava se kada se 1. ulazi u čiju kuću: primete li gosta? 2. naiđe na osobu koja jede: mogu li se i ja pridružiti? (LOP:127)
7P: <i>lijepa kao dženetska huriya</i> (ar.) – veoma lijepa (LOP:174)	7N: <i>nije maslo za ramazana</i> (ar.) <i>kome</i> – nije savršen, nije bez mahane (LOP:221)
8P: <i>oči k(a)o fildžani</i> (perz.) – razg. 1. veoma krupne oči 2. izbečene, razrogačene oči (ob. iz straha, čuđenja) (LOP:161)	8N: <i>nije ti ovo (to) babovina</i> (LOP: 117)/ <i>babo platio</i> – izraz kojim se kome prebacuje da se (neopravdano) ponaša kao da polaže pravo na šta (ob. na društvenu imovinu)
9P: <i>poznavati/znati (šta) kao hafiz</i> (ar.) < <i>ajete</i> > (ar.) – općenito, poznavati, znati šta veoma dobro (LOP: 112)	9N: <i>(biti) ošišan na času</i> (perz.) – razg. <i>pejor.</i> (biti) ružno ošišan, s pretjerano podrezanim šiškama (LOP:139)
10P: <i>razvući k(a)o hodža</i> (perz.) <i>teraviju</i> (ar.) – razg. <i>pejor.</i> pretjerano dugo i sporo činiti, obavljati šta, previše razvući (LOP:237)	10N: <i>padati/pasti u merak</i> (ar.) (LOP:204) (<i>sevdah</i>) (ar.) (LOP:226) – postajati/postati sve razdraganiji, veseliji
11P: <i>ružan k(a)o akrep</i> (ar.) – razg. veoma ružan (LOP: 112)	11N: <i>(govoriti, pričati, raditi i sl.) po istilahu</i> (ar.) – (govoriti, pričati, raditi i sl.) bez žurbe, sa zadržavanjem na detaljima (LOP:177)
12P: <i>skuplja pita</i> (grč.) <i>nego tepsija</i> (tur.) (<i>od tepsije</i>) – veći trošak nego korist, ne isplati se (LOP:237)	12N: <i>poslati (koga) po čevape</i> (ar.) – žarg. prevariti, obmanuti koga (ob. u šali) (LOP:140)
13P: <i>težak k(a)o tuč</i> (tur.) – razg. 1. veoma težak 2. veoma naporan, nepodnošljiv, nesnosan (LOP:240)	13N: <i>posrati se (kome) u čeif (čeff)</i> (ar.) – vulg. upropastiti kome kakvo uživanje, ugodu, nasladu (LOP:139)
14P: <i>zaboravljati k(a)o mahalusa šamiju</i> (perz.) – razg. biti zaboravan (LOP:198)	14N: <i>povaliti raju</i> (ar.) – žarg. izdati bliske osobe (prijatelje) (LOP:221)
15P: <i>živjeti kao beg</i> (tur.) (LOP:122) / <i>paša</i> (tur.) (LOP:216) – živjeti u blagostanju, bogatstvu, bez ikakvih briga, bezbrižno	15N : <i>rahat</i> (ar.) <i>od pameti</i> – razg. iron. 1. blesav, luckast 2. koji je flegmatična ponašanja, koji se ni za šta ne sekira (LOP:221)

Nakon poređenja leksikografskih definicija značenja frazema iz tabele 1 sa definicijama značenja koje je ponudila vještačka inteligencija, utvrđeno je sljedeće:

- a) Podudaranje značenja ostvareno je u 17 (56,66%) frazema od 30, i to: 1P (tačno procjenjuje sarkastičnu notu), 2P (no, nema značenja polusloženice *abu-zemze* i tačnog porijekla njenih sastavnica), 3P (tačno procjenjuje humornu notu), 4P, 6P, 7P, 8P, 10P, 11P, 12P, 14P, 15P, 4N, 7N, 8N, 10N, 13N.
- b) Djelimično podudaranje značenja ostvareno je u 6 (20%) frazema od 30, i to: 9P ('znati nešto savršeno i besprijekorno napamet' – nije neophodno, akcent je na 'dobro znati/poznavati'); 13P (prvo značenje navedeno, drugo nije); 3N (drugo značenje navedeno: 'doći sebi'; 'doći pameti', prvo nije; »u srpskom jeziku«); 6N (prvo značenje navedeno, drugo nije); 9N ('biti ružno ošišan' i 'biti izigran'); 15N (drugo navedeno, prvo nije).
- c) Potpuno nepodudaranje značenja zabilježeno je u 7 (23,33%) frazema od 30, i to: 5P (VI: »Izraz *kao da je hodži s tabuta pobjegao* koristi se u kontekstu da opiše osobu koja izgleda jako iznervirano, uplašeno ili zbunjeno, kao da je prošla kroz neku tešku situaciju. *Hodža* je vjerski lider, dok je *tabut* kovčeg, pa izraz implicira da je osoba preživjela neku opasnost ili je bila u ozbiljnoj situaciji, poput opasnosti smrti, a sada se ponaša ili izgleda zbunjeno i nervozno«); 1N (VI: »Izraz se koristi da označi nešto što je suvišno, nepotrebno ili dodato na pogrešnom mjestu. Pilav (pirinač ili jelo od pirinča) i biber su začini koji se često kombinuju, ali ovaj izraz u prenesenom značenju znači da je nešto previše ili neharmonično dodato – kao da se stavi previše bibera u pilav, pa jelo postane prejako ili neukusno«); 2N (VI: »Znači 'biti spreman na pomirenje, nagodbu li dogovor, posebno nakon neke svađe ili nesuglasice'. Riječ *muštuluk* potiče iz turskog jezika i znači pomirenje, kompromis ili obostrani dogovor«); 5N (VI: »To je emotivan apel često korišten kada majka ili neko u njeno ime traži oprostaj za neku grešku ili kada neko treći pokušava da ublaži tenziju između majke i sina«); 11N (VI: »*Raditi po istilahu* znači 'raditi po dogovoru, po utvrđenom običaju, a ne nasumice ili proizvoljno'«); 12N (VI: »U prenesenom značenju, ovaj izraz se može koristiti i šaljivo ili čak kao blaga prijetnja, u smislu: 'poslati koga negdje da brzo i efikasno nešto završi'; u nekim situacijama i kao šala ili blaga opomena da ćeš dobiti batine«, »u srpskom (i širem balkanskom) kolokvijalnom govoru«); 14N (VI: »Može značiti: 'poraziti ili savladati grupu ljudi (u nekoj igri, sportu ili takmičenju)'; 'pobijediti društvo u nekoj situaciji'; može imati i seksualnu konotaciju, ali to nije najčešće značenje«).

Što se tiče adekvatnosti primjera za svaku frazemu, rezultati su sljedeći:

- a) primjeri adekvatne upotrebe: 1P, 2P, 3P, 4P, 8P, 12P, 1N, 4N, 6N, 10N, 13N;
- b) primjeri djelimično adekvatne upotrebe (primjer za jedno od dva značenja nedostaje): 15N (nedostaje primjer za prvo značenje);
- c) primjeri neadekvatne upotrebe: 5P (»Obukao je odijelo za svadbu, ali s tim šeširrom izgleda kao da je hodži s tabuta pobjegao!« u značenju 'biti čudno odjeven, neuklopljen'); 2N (»Prestanimo se svađati, ja sam za muštuluka«); 5N (»Mislim da ću ujutro ići na pecanje bez dozvole. – Halali, Huso, materi! Al' ako te uhvate, ne traži mene!«); 7N (»Nije on maslo za ramazana«); 11N (»Mi to u selu radimo po istilahu – zna se ko kosi, ko nosi«); 12N (»Ako se ne središ, poslat ću te po čevape«); 14 N (»Brate, kad sam ušao u klub, povalio sam raju na igrici i svi su bili u čudu«);
- d) primjeri i adekvatne i neadekvatne upotrebe: 3N (»Kad ćeš više da dođeš tobe i prestaneš da praviš gluposti?«, »Kad je vidio račun, odmah je došao tobe« – drugi primjer u značenju 'biti u šoku'); 9N (»Ej, jesi vidio Adisa? Ošišan je na času, skoro kao vojnik«, »On im je vjerovao a sada je ošišan na času«, u značenju 'biti izigran').

Pored ove grupe frazema, u korpus su uvrštene i frazeme koje sadrže orijentalizam netransparentnog porijekla, što podrazumijeva da je leksema naširoko korištena, i to ne samo u bosanskom jeziku nego i u srpskom i hrvatskom³ jeziku: *avet* (ar.), *bubreg* (tur.), *budala* (ar.), *bunar* (tur.), *čamac* (tur.), *fitilj* (ar.), *kaiš* (tur.), *kašika* (tur.), *kičma* (tur.), *kula* (ar.), *leš* (perz.), *limun* (perz.), *majmun* (ar.), *para* (perz.), *šamar* (tur.), *šećer* (perz.), *taman* (ar.), *veresija* (tur.). Frazeme s ovim leksemama navedene su u sljedećoj tabeli:

Tabela 2. Popis frazema s orijentalizmom netransparentnog porijekla

Poredbene frazeme sa značenjima	Neporedbene frazeme sa značenjima
16P: <i>biti kao avet</i> (ar.) – biti slab, iscrpljen i blijed (LOP:116)	16N: <i>budala</i> (ar.) <i>na kvadrat</i> – žarg. velika, nepopravljiva budala (LOP:126)
17P: <i>govoriti kao iz bunara</i> (tur.) – imati dubok glas (LOP:128)	17N: <i>imati kratak fitilj</i> (ar.) – biti plahovite naravi, prznica (LOP:162)

3 Izuzev leksema *bunar*, *kaiš*, *kašika*, *kičma*, koje nisu sastavni dio leksičkog fonda hrvatskog standardnog jezika.

18P: (<i>biti</i>) <i>iscijeden kao Limun</i> (perz.) – (<i>biti</i>) premoren, potpuno iscrpljen (LOP:196)	18N: (<i>is</i>) <i>kriviti kičmu</i> (tur.) – naporno, mukotrpno raditi (LOP:190)
19P: <i>izgledati k(a)o leš</i> (perz.) – razg. veoma loše, bolesno, iscrpljeno izgledati (LOP:196)	19N: <i>nisam (nisi) od šećera</i> (perz.) – neću (nećeš) se rastopiti od nekoliko kapljica kiše (LOP:230)
20P: <i>k(a)o da se rodio u čamcu</i> (tur.) – razg. ne zatvara vrata prilikom izlaska iz prostorije (LOP:130)	20N: <i>od gotovoga napraviti veresiju</i> (tur.) – sigurno pretvoriti u nesigurno (LOP:244)
21P: <i>sjesti/pristati k(a)o šamar</i> (tur.) <i>budali</i> (ar.) – razg. biti potpuno odgovarajući; odgovarati (LOP:126)	21N : <i>parā</i> (perz.) <i>valjati</i> – veoma vrijediti (LOP:216)
22P: <i>srušiti se k(a)o kula</i> (ar.) <i>od karata</i> – naglo propasti (o kakvim planovima) (LOP:193)	22N: <i>pasti (kome) kašika</i> (tur.) <i>u med</i> – iznenadno, promijeniti se nabolje (za čiju životnu situaciju) (LOP:188)
23P: <i>stopala k(a)o čamci</i> (tur.) (<i>peraje</i>) – razg. velika stopala (LOP:130)	23N: <i>praviti od sebe majmuna</i> (ar.) – svojim postupcima izazivati podsmijeh (LOP:200)
24P: <i>tačan kao (švicarski) sat</i> (ar.) –apsolutno tačan (LOP :224)	24N: <i>stegnuti/stezati (pritegnuti/pritezati) kaiš</i> (tur.) – (početi/počinjati) štedjeti u troškovima za svakodnevni život, početi/počinjati živjeti skromnije (LOP:184)
25P: <i>živjeti kao bubreg</i> (tur.) <i>u loju</i> – živjeti u blagostanju, lagodno; uživati (LOP:126)	25N: <i>taman</i> (ar.) <i>posla!</i> – ne dolazi u obzir; nikako, nipošto (LOP:235)

Nakon poređenja leksikografskih definicija značenja frazema iz tabele 2 sa definicijama značenja koje je ponudila vještačka inteligencija, utvrđeno je sljedeće:

- Podudaranje značenja ostvareno je u 14 (70%) frazema od 20, i to: 18P, 19P, 21P, 22P, 23P, 24P, 25P, 16N, 17N, 20N, 22N, 23N, 24N, 25N.
- Djelimično podudaranje značenja ostvareno je u 2 (10%) frazeme od 20, i to: 16P (osnovno značenje i značenje ‘ponašati se povučeno, tiho ili neprimjetno’, i 19N (osnovno značenje i značenje »popularna fraza koja se koristi u govoru na Balkanu kako bi se izrazila otpornost ili izdržljivost osobe u situacijama koje bi druge mogle učiniti osjetljivima ili slabima«).
- Potpuno nepodudaranje značenja zabilježeno je u 4 (20%) frazeme od 20, i to: 17P (VI: »Koristi se u srpskom jeziku (i drugim srodnim jezicima) kao frazeološki izraz koji znači ‘govoriti nejasno, prigušeno, nerazgovijetno’ – kao da dolazi iz dubine, iz zatvorenog prostora, što otežava razumijevanje«); 20P (VI:

»Koristi se u kolokvijalnom govoru i ima pejorativno ili šaljivo značenje koje zavisi od konteksta, ali uglavnom znači: ‘osoba se ponaša ili izgleda nepristojno, neotesano, nevaspitano’ – kao neko ko nije naučio osnovna pravila ponašanja, posebno za stolom ili u društvu«); 18N (VI: »U prenesenom značenju, može se koristiti metaforički da označi situaciju u kojoj osoba pretrpi emocionalnu ili psihološku štetu zbog svojih postupaka ili odluka«); 21N (VI: »žargonska fraza koja se koristi u govoru na Balkanu, a znači ‘imati puno novca’ ili ‘biti bogat’«).

Što se tiče adekvatnosti primjera za svaku frazemu, rezultati su sljedeći:

- a) primjeri adekvatne upotrebe: 16P, 18P, 19P, 21P, 22P, 23P, 24P, 25P, 16N, 17N, 19N, 20N, 23N, 24N, 25N;
- b) primjeri djelimično adekvatne upotrebe (primjer za jedno od dva značenja nedostaje): /
- c) primjeri neadekvatne upotrebe: 17P (»Ne razumijem ga uopće preko telefona, govori kao iz bunara«); 20P (»Kako jedeš, kao da si se rodio u čamcu« – u značenju ‘jedeš neuredno, halapljivo, bez kulture’); 18N (»Njegovo ponašanje prema kolegama može mu iskriviti kičmu u profesionalnom smislu«); 21N (»Vidim da ti para valja, stalno ideš na putovanja«).

3. Rezultati analize i zaključci

Na temelju provedene analize obiju grupa frazema, možemo izvesti sljedeće zaključke:

1. Kvalitet definicija koje daje ChatGPT na dosta je visokom nivou kod frazema s općeprihvaćenim i naširoko korištenim orijentalizmom (70%) i poredbenih frazema s orijentalizmom transparentnog porijekla – 12 (80%) frazema od 15, što potvrđuje našu početnu hipotezu da će značenja poredbenih frazema načelno biti bolje definirana. No, kada je općenito riječ o frazemama koje sadrže orijentalizam transparentnog porijekla (56,66%) i koje su tipične za bosanski jezik, rezultat je lošiji, a može se protumačiti kao posljedica činjenice da elektronski korpusi za bosanski jezik nisu brojni i bogati kao korpusi za hrvatski i srpski jezik.
2. Djelimično podudaranje značenja ostvareno je u 20% frazema s orijentalizmom transparentnog porijekla u poređenju sa 10% frazema s općeprihvaćenim orijentalizmom.

3. Potpuno nepodudaranje značenja realizirano je u 23,33% frazema s orijentalizmom transparentnog porijekla u poređenju sa 20% frazema s općeprihvaćenim orijentalizmom, pri čemu se tri četvrtine od 23,33% netačno definiranih frazema odnosi na neporedbene frazeme, a samo jedna četvrtina na poredbene frazeme.
4. Pretpostavka da će ponovljeni pokušaji definiranja frazema koje vrši ChatGPT dovesti do tačnije i/ili preciznije definicije potvrdili su se npr. u slučaju frazema *biber po pilavu* i *biti za muštuluka*, ali ne i u slučaju frazeme *kao da je hodži s tabuta pobjegao*. Ovaj rezultat ukazuje na neophodan oprez prilikom korištenja nalaza ChatGPT-ja.
5. Netačno definirana značenja frazema očekivano su u korelaciji sa neadekvatnim kontekstom upotrebe, odnosno sa neadekvatnim primjerima⁴.
6. Pored netačno definiranih značenja frazema, prisutne su i druge netačne kvalifikacije poput one da se frazema koristi u srpskom jeziku a koristi se i u bosanskom, što je slučaj npr. kod frazema *kao da je krenuo/krenula na Čabu* i *ružan k(a)o akrep*. Iz tih razloga, treba biti veoma oprezan prilikom korištenja ovakvih izvora i pažnju obraćati na sve detalje a ne samo tražene – značenje frazeme.
7. I kod frazema čije je značenje pravilno protumačeno, često nedostaju kvalifikatori, poput *razg.* i *sl.*, ili sinonimi, a oni su veoma važni za korisnike rječnika.
8. Vještačka inteligencija pak često daje mnogo detaljnija objašnjenja motivacije za nastanak određene frazeme, kao u slučaju frazema *tačan kao švicarski sat* (objašnjava se po čemu su švicarski satovi stekli slavu) ili *živjeti kao bubreg u loju* (objašnjava se položaj bubrega u ljudskom organizmu) te se čak navode frazeme istog ili sličnog značenja u drugim jezicima (za frazemu *živjeti kao bubreg u loju*), što može biti zanimljivo korisniku rječnika ali i podsticajno u nastavi (stranih) jezika.

Na kraju, možemo ustanoviti da je istraživanje značenja frazema na ograničenom korpusu od 50 frazema potvrdilo početne hipoteze, ali će tek rezultati analize znatno većeg broja frazema ponuditi konačne odgovore i ujedno otvoriti prostor za daljnje analize ovog fenomena.

4 Ovaj zaključak ne korespondira sa zaključkom Lewa (2023:6), koji je utvrdio da su primjeri upotrebe i drugi podaci koje je navodio ChatGPT (glose i slično) bili lošije ocijenjeni u poređenju sa definicijama frazema.

Reference

- 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.
- Gantar, A. (2024). Formulisanje rečničkih definicija pomoću veštačke inteligencije na primeru slovenačkih frazeoloških jedinica. U S. Marjanović (ur.), *Leksikografski susreti 1: Moderni rečnici u funkciji prosečnoga korisnika: stari problemi, savremeni pravci i novi izazovi* (str. 151–157). Beograd: Filološki fakultet.
- Jakubiček, M., & Rundell, M. (2023). The end of lexicography? Can ChatGPT outperform current tools for post-editing lexicography? U M. Marek et al. (ur.), *Proceedings of the eLex 2023 Conference: Electronic Lexicography in the 21st Century* (str. 508–523). Brno: Lexical Computing.
- Kasumović, A., & Nikolić, M. (2018). *Rječnik frazema bosanskoga jezika*. Tuzla: Institut za humanu rehabilitaciju.
- Kelman, D. (2021). *Moj algoritam i ja*. Beograd: RED BOX.
- Lew, R. (2023). ChatGPT as a Cobuild lexicographer. *Humanities and Social Sciences Communications*, 10(1), 704. <https://doi.org/10.1057/s41599-023-02119-6>
- Šehović, A., & Haverić, Đ. (2017). *Leksika orijentalnog porijekla u frazemama bosanskog jezika [elektronsko izdanje]*. Sarajevo: Filozofski fakultet. https://ebooks.ff.unsa.ba/index.php/ebooks_ffunsa/catalog/book/36