Umetna inteligenca in prevajanje: kako se obnesejo veliki jezikovni modeli pri najzahtevnejših jezikovnih nalogah

Špela Vintar

1 Uvod

Veliki jezikovni modeli nas spremljajo že tri leta, v tem vznemirljivem obdobju pa so dokazali, da bo umetna inteligenca v skladu s številnimi desetletja starimi napovedmi korenito spremenila svet. Svet se je v resnici pod vplivom digitalizacije in inteligentnih tehnologij, ki selektivno usmerjajo dotok informacij do uporabnikov, spremenil že precej prej, prav tako smo se prevajalci z nevronskimi mrežami in matematičnimi reprezentacijami pomena pri programih Google Translate in DeepL srečali vsaj štiri leta pred umetnointeligenčnim valom, ki je v splošno javnost pljusknil konec leta 2022 z objavo ChatGPT-ja.

Danes strojne prevajalnike uporablja približno petina svetovnega prebivalstva,[1] v profesionalnem prevajanju pa jih po podatkih evropske raziskave o trendih v jezikovni

1 Ocena temelji na javno dostopnih številih uporabnikov prevajalskih storitev treh največjih ponudnikov – Googla, Yandexa in DeepLa.

industriji (ELIS 2024) redno uporablja približno 75 odstotkov udeleženih akterjev. Na področje jezikovnih storitev so močno prodrla tudi orodja umetne inteligence, ki se po podatkih prej omenjene raziskave za jezikovne naloge uporabljajo pri 30 odstotkih udeležencev, še precej večja pa je njihova vloga v integriranih prevajalskih okoljih, kakršna so Phrase, Trados in memoQ. Z njimi je vodenje prevajalskih projektov vse bolj avtomatizirano, samodejno se za strokovna področja izbere najustreznejši prevajalnik, z uporabniško določenimi pozivi (angl. *prompt*) pa je mogoče umetni inteligenci prepustiti tudi revidiranje in lekturo besedil, preverjanje terminologije in še marsikaj.

Veliki jezikovni modeli vse bolj vstopajo tudi na področje tolmačenja, kjer se v zadnjih letih že pojavljajo dovolj zmogljivi sistemi za simultano tolmačenje v realnem času (eden od njih je Kudo.ai, glej tudi Fantinuoli 2023). Pri prenosu govornega signala iz enega jezika v drugega je največji tehnološki izziv ravnotežje med čim večjo točnostjo modela in čim manjšim zamikom med vhodnim govornim sporočilom in pretolmačeno vsebino, saj manjši zamik pomeni, da mora model pravilno predvideti, kako se bo izjava nadaljevala.

Bliskoviti napredek, ki se je hitro prelevil v globalno tekmo za prevlado na področju vse pametnejših jezikovnih modelov ter njihove vse hitrejše implementacije v različna področja življenja in dela, ima brez dvoma tudi negativne posledice, med drugim izginjanje ali preobražanje številnih delovnih mest. Temu trendu se tudi jezikovna panoga ne more izogniti, saj kljub rastočim potrebam po jezikovnih storitvah namere delodajalcev po zaposlovanju jezikoslovcev in terminologov upadajo, raste le potreba po strokovnjakih za upravljanje podatkov in inteligentne tehnologije (ELIS 2024). Precej vprašanj se poraja tudi v zvezi z etičnimi in pravnimi vidiki umetne inteligence; na področju prevajanja in tolmačenja je najpogosteje slišati pozive za ustrezno ovrednotenje jezikovnih podatkov, na katerih se modeli učijo in ki so plod človeškega ustvarjanja.[2]

V današnjem času je posebej nehvaležno tudi pisanje člankov o umetni inteligenci. Področje jezikovnih tehnologij je v zgolj nekaj letih preraslo iz obrobne podpanoge računalništva v osrednje raziskovalno polje z neprimerno večjimi finančnimi vložki in posledično znanstveno (hiper)produkcijo. Novi raziskovalni dosežki se vrstijo tako hitro, da jih ažurno ne uspe sprocesirati niti sama umetna inteligenca, kaj šele človek. Ker se vse boljši jezikovni modeli pojavljajo praktično na dnevni bazi, so kakršne koli ugotovitve o današnjih zmogljivostih modelov že jutri zastarele. S tem se smiselnost pričujočega prispevka v monografski publikaciji kar dramatično zmanjša, vseeno pa se

2 V času pisanja tega prispevka (marec/april 2025) se je takšna polemika razvila v zvezi z akcijo zbiranja besedil za slovenski jezikovni model v okviru projekta PoVeJMo; Društvo slovenskih književnih prevajalcev (DSKP) je namreč na vodje projekta naslovilo javno pismo s številnimi vprašanji, https://www.dskp-drustvo.si/wp-content/uploads/2025/03/Vprasanja-DSKP-v-zvezi-z-zbiralno-akcijo-besedil-PovejmoSi-final.pdf.

bomo v nadaljevanju posvetili dvema temama, ki predstavljata osrednji vidik prevajalskega dela ter se pomembno navezujeta tudi na širše področje pravnega in skupnostnega tolmačenja, in sicer, prvič, strokovno terminologijo in terminološko doslednost, in drugič, jezikovno pragmatiko.

V osrednjem delu prispevka tako obravnavamo velike jezikovne modele s teh dveh vidikov, pri tem pa posebno pozornost namenjamo še vedno precejšnjim omejitvam, ki se pojavljajo pri uporabi sodobnih jezikovnih tehnologij za manjše jezike, kakršen je slovenščina.

2 Umetna inteligenca in strojno prevajanje

Preden se posvetimo pregledu trenutnega stanja na področju računalniškega prevajanja, pojasnimo temeljne izraze. Strojno prevajanje (angl. *Machine Translation*, MT) se danes skoraj izključno nanaša na nevronsko strojno prevajanje, ki se je razmahnilo po letu 2017 s premikom vseh večjih komercialnih sistemov od statističnega pristopa k nevronskemu. Za jezikovni par angleščina-slovenščina je nevronski model dostopen na prevajalniku Google Translate od leta 2018, istega leta so ta pristop uvedli tudi pri prevajalniku evropskih institucij eTranslation, prevajalnik DeepL pa je naš jezik pričel podpirati leta 2021. Nevronski strojni prevajalnik se uči za vsak jezikovni par posebej, in sicer s pomočjo globokega strojnega učenja iz velikih dvojezičnih korpusov, rezultat pa je jezikovni model. Procesiranje jezika poteka sekvenčno, tako da kodirnik zaporedoma pretvarja vhodne besede izvirnega segmenta v vektorske vložitve, dekodirnik pa ob pomoči mehanizma pozornosti ustvarja ciljno besedilo.

Prevajanje z umetno inteligenco se običajno nanaša na uporabo velikih jezikovnih modelov (angl. *large language model*, LLM) za prevajanje, kakršni so ChatGPT, Copilot, Gemini ali odprtokodna Llama. Ti modeli so sicer prav tako naučeni na velikih količinah podatkov ter uporabljajo vektorske vložitve in arhitekturo transformer, a je med njimi in prevajalniki nekaj pomembnih razlik. Prva je, da so jezikovni modeli prevajalnikov precej manjši, saj običajno obsegajo nekaj sto milijonov parametrov, medtem ko sodobni veliki jezikovni modeli dosegajo do 500 milijard parametrov, in celo njihove zmanjšane (kvantizirane) različice uporabljajo nekaj milijard parametrov. Druga razlika, ki sledi iz prejšnjega podatka, je dejstvo, da z velikimi jezikovnimi modeli lahko opravljamo najrazličnejše naloge, saj so ustvarjeni za razmišljanje in sklepanje. Njihova jezikovna kompetenca je na neki način stranski produkt srečevanja s tako velikimi količinami (pretežno jezikovnih) podatkov, njihov enormni semantični prostor pa jim omogoča prenos jezikovnih sporočil med jeziki, in sicer tudi med tistimi jezikovnimi pari, za katere obstaja malo ali nič dvojezičnih podatkov.

Ena od posledic teh razlik je, da pri prevajanju s »klasičnimi« strojnimi prevajalniki med jezikovnimi pari, ki ne vključujejo angleščine, utegnemo zaslediti interferenco vmesnega jezika. Ta se na primer pokaže pri prevajanju ženskih samostalniških oblik (*prijateljica*, *kuharica*, *raziskovalka*) iz češčine v slovenščino ali kak drug oblikoslovno bogat jezik. Prevajalniki, kot sta Google Translate in DeepL, kot vmesni jezik uporabljajo spolno nezaznamovano angleščino, pri tem pa se ženska samostalniška oblika nadomesti z bolj generično moško (npr. *Moje nejlepší kamarádka tu nikdy nebyla.* -> Moj najboljši prijatelj ni bil nikoli tukaj.).[3]

Veliki jezikovni modeli po drugi strani pridobijo prevajalsko kompetenco s precej manjšo količino podatkov za posamezni jezik, kot bi jih potrebovali za učenje klasičnega prevajalnika, saj izkoriščajo podobnosti med jeziki in svoj enormni semantični prostor. Eno obsežnejših evalvacij velikih jezikovnih modelov pri prevajanju predstavljajo Zhu in soavtorji (2024), ki sistematično primerjajo sedem jezikovnih modelov (OPT, Llama2-7B, Falcon, XGLM, BLOOMZ, ChatGPT in GPT4) ter dva »klasična« nevronska prevajalnika, NLLB in Google Translate (GT), za kar 102 jezika in 606 jezikovnih kombinacij. Rezultati, ki jih avtorji grupirajo po jezikovnih družinah, pri večini le-teh sicer še vedno pokažejo premoč prevajalnika Google Translate, razen za nekatere kombinacije germanskih jezikov (kamor sodi tudi angleščina). GPT4 tako prekaša GT pri afrikanščini, nemščini in norveščini, pri številnih jezikovnih parih pa sta sistema skoraj izenačena. Pri slovensko-angleškem prevodu doseže GPT4 vrednost COMET[4] 88,45, GT pa 88,53.

Pri teh rezultatih velja omeniti, da so bili jezikovni modeli testirani leta 2023, tako da glede na njihov bliskoviti napredek rezultati niso več relevantni. Morda je pomembnejša ugotovitev omenjene raziskave, da se pri uporabi jezikovnih modelov za prevajanje odlično obnese metoda kontekstnega učenja (angl. *in-context learning*), kar pomeni, da v poziv dodamo nekaj primerov uspešnih prevodov za izbrani jezikovni par. Videti je, da s tem jezikovni model »pripravimo« na nalogo, ki ga čaka, pri tem pa se pokaže še, da je optimalno število primerov osem, saj pri večjem številu kakovost prevodov začne ponovno upadati. Nadalje avtorji sklepajo, da lahko jezikovni modeli razvijejo dokaj visoko prevajalsko kompetenco za jezikovni par angleščina-X, tudi če so v fazi predučenja »videli« zgolj en odstotek podatkov za jezik X v primerjavi s količino podatkov za angleščino.

Kontekstno učenje pa ni edini način, kako izboljšati prevode, ki jih generirajo veliki jezikovni modeli. He in soavtorji (2024) predlagajo metodo večnivojskega pozivanja,

3 Prevod ustvarjen s prevajalnikom DeepL dne 26. 3. 2025.

4 COMET je novejša metrika za vrednotenje nevronskih strojnih prevodov, ki izkazuje visoko stopnjo korelacije s človeškimi ocenami (Rei in dr. 2020).

imenovano MAPS (angl. *multi-aspect prompting and selection*), s katero naj bi se proces prevajanja približal človeškemu. Ljudje namreč pri prevajanju intuitivno prepoznamo ključne izraze in tematiko besedila, s tem pa aktiviramo znanje, potrebno za ustrezen prevod. Če jezikovnemu modelu podamo navodilo, naj v izvirnem besedilu najprej identificira ključne besede in teme, nato generira krajše sorodno besedilo in se šele zatem poda v prevajanje, so rezultati precej boljši od prevoda brez navodil, prav tako naj bi ta metoda za vseh enajst preskušenih jezikovnih parov delovala bolje od kontekstnega učenja. Prednost metode MAPS je, da za njeno uporabo ne potrebujemo dodatnih jezikovnih virov, kot so denimo referenčni primeri prevodov.

Veliki jezikovni modeli se v zadnjem času uporabljajo tudi za evalvacijo strojnih prevodov, o čemer piše več avtorjev (Kocmi in Federmann 2023, Wei in dr. 2024). Tudi za to nalogo se preskušajo različni načini pozivanja, zadnja dostopna raziskava (Qian in dr. 2024) pa ugotavlja, da se še najbolje obnese poziv, ki modelu sicer ne predpiše specifične tipologije napak, ampak v poziv poleg strojnega prevoda vključi tudi referenčni prevod.

Iz doslej napisanega je razvidno, da je umetna inteligenca že dodobra vstopila v svet prevajanja. V nadaljevanju prispevka se podrobneje posvetimo dvema zahtevnejšima vidikoma prevajalskega dela, ki sta relevantna tudi pri sodnem prevajanju in tolmačenju, in sicer najprej strokovnim izrazom oziroma terminologiji, nato pa jezikovni pragmatiki.

2.1 Veliki jezikovni modeli in terminološka doslednost

Ustrezno prevedena strokovna terminologija je najpomembnejši element kakovosti prevoda, tako po mnenju naročnikov kot prevajalcev. Kljub bliskovitemu razvoju jezikovnih modelov strojno prevajanje strokovnih besedil s specializiranim izrazjem ostaja izziv. Še v času statističnih strojnih prevajalnikov s fraznim modelom (Koehn in dr. 2003) je bilo delovanje prevajalnika mogoče eksplicitno usmerjati, bodisi z dodajanjem specializiranih glosarjev v tabelo fraz ali pa s spreminjanjem verjetnostnih uteži za določene besede ali fraze. A od nevronskih prevajalnikov dalje to ni več tako enostavno, saj uvajanje omejitev (tj. eksplicitnih terminoloških izbir) v fazo dekodiranja precej upočasni procesiranje (Anastasopoulos in dr. 2021).

Nevronski prevajalniki besedilo še vedno prevajajo po posameznih segmentih in zato ne preverjajo doslednosti svojih rešitev; zlahka se torej zgodi, da prevajalnik za isti termin v izvirniku uporabi več različnih ustreznic. Kmalu po pojavu nevronske različice GT smo v lastni raziskavi izvedli primerjavo doslednosti prevajanja terminologije s statističnim in nevronskim modelom (Vintar 2018). V kvantitativnem delu raziskave smo pri obeh sistemih zabeležili približno 70-odstotno natančnost pri prevajanju terminoloških izrazov, podrobnejši vpogled v prevode pa je pokazal, da je nevronski model

nagnjen k še večji variabilnosti terminologije kot statistični, pa tudi k halucinacijam v obliki izmišljenih besed.

Da je terminološka doslednost pomembna, je razvidno tudi iz dejstva, da se z njo posebej ukvarjajo na svetovnem tekmovanju prevajalnikov WMT[5] v obliki skupne naloge *Machine Translation using Terminologies*. Semenov in Bojar (2022) ugotavljata, da običajne metrike za ocenjevanje strojnih prevodov, kot sta BLEU in COMET, ne povedo kaj dosti o terminološki doslednosti, poleg tega se izkaže, da tudi človeški ocenjevalci, ki niso strokovnjaki na izbranem področju, težko opazijo napačno rabo specializiranega izrazja. Avtorja zato predlagata nov način merjenja doslednosti, ki posebej kaznuje terminološko variabilnost v prevodu in dvoumno rabo izrazov.

2.2 Preskus jezikovnih modelov pri zagotavljanju terminološke doslednosti

Ker se veliki jezikovni modeli danes vse pogosteje uporabljajo za preverjanje besedil in zagotavljanje kakovosti, smo izvedli kratek preskus zmogljivosti sodobnih modelov pri preverjanju terminološke doslednosti za jezikovni par angleščina-slovenščina. Za preskus smo pripravili kratko besedilo v izvirni angleščini in slovenskem prevodu, pri tem pa slovenski prevod vsebuje skupno osem primerov nedosledne rabe terminologije (slika 1). Preskus smo izvedli s štirimi velikimi jezikovnimi modeli, in sicer s ChatGPT-4o in Claude Sonnet 3.7 kot predstavnikoma komercialnih modelov ter z Llamo 3.3 in DeepSeek-R1-14B kot predstavnikoma odprtokodnih modelov. Pri obeh odprtokodnih modelih poudarjamo, da uporabljamo destilirani različici modelov z manjšim številom parametrov, in sicer jih ima DeepSeek 14 milijard, Llama 3.3 pa 70 milijard. Poleg primerjave modelov so nas zanimali tudi različni pozivi, saj iz prej omenjenih raziskav izhaja, da je uspešnost strokovnega prevajanja in – domnevno – tudi terminološkega preverjanja prevodov odvisna od dobro oblikovanega poziva.

Pri pozivanju smo tako uporabili tri strategije, in sicer je prva osnovni poziv brez dodatnega znanja (*zero-shot*), druga je uporaba miselnega procesa (*chain-of-thought*) in tretja je kontekstno učenje (*in-context learning*). Vse tri različice pozivov smo preskusili v angleščini in slovenščini. Tabela 1 navaja različice pozivov, ki so jim v vseh primerih sledili dvojezični podatki, v tabeli 2 pa podajamo rezultate, in sicer najprej število odkritih nedoslednosti, nato število napačno ugotovljenih nedoslednosti, nazadnje pa – kljub izredno majhnim številkam – izračunamo še natančnost, priklic in vrednost F1.

5 Konferenca World Machine Translation Summit je največji svetovni dogodek na temo strojnega prevajanja in poteka od leta 2006. Zasnovan je v obliki skupnih nalog (*shared task*), pri katerih raziskovalne ekipe z vsega sveta preskušajo svoje sisteme na vnaprej določenih podatkovnih zbirkah.

Tabela 1: Pozivi za preverjanje terminološke doslednosti

Tip poziva	Poziv v angleščini	Poziv v slovenščini
Osnovni	In the following table, which contains ID, original, target, please check terminological consistency of translations into Slovenian. List possible inconsistencies by line IDs.	Tabela spodaj vsebuje ID, izvirnik in prevod. Prosim preveri terminološko doslednost prevodov v slovenščino. Morebitne nedoslednosti navedi po ID-jih vrstic.
Miselni proces	<system>You are a professional terminologist ensuring the consistency of term use in translations.</system> <instructions>In the following table, which contains ID, original, target, please check terminological consistency of translations into Slovenian. Proceed as follows: · Read the entire text in English and Slovenian. Identify important terms in English first. · Proceed by checking all occurrences of the original English term, and its translations. If a term is translated in more than one way, it is inconsistent. · List all inconsistencies by line IDs.</instructions>	<system>Si profesionalni terminolog, ki skrbi za terminološko doslednost v prevodih.</system> <instructions>Tabela spodaj vsebuje ID, izvirnik in prevod. Prosim preveri terminološko doslednost v prevodu v slovenščino, pri tem pa uporabi naslednji postopek: · Preberi celotno besedilo v angleščini in slovenščini. Najprej izlušči pomembne termine v angleškem besedilu. · Nato preveri vse pojavitve izvirnega angleškega termina ter njegovih prevodov. Če je termin preveden na več kot en način, gre za nedoslednost. · Naštej vse nedoslednosti skupaj z ID-jem vrstice.</instructions>
Kontekstno učenje	<system>You are a professional terminologist ensuring the consistency of term use in translations.</system> <instructions>Identify terminological inconsistencies in the translated text. Below are some examples.</instructions> <examples>1 In the northwestern part of Matarsko podolje two types of alluvial fans occur. V severozahodnem delu Matarskega podolja se pojavljata dva tipa aluvialnih vrtač. 2 One alluvial fan has an active process of alluvial sedimentation. Pri enem aluvialnem ventilatorju je na površini aktiven proces rečne sedimentacije. </examples> <expected_result>Inconsistency 1: alluvial fan – aluvialna vrtača / aluvialni ventilator Inconsistency 2: alluvial - aluvialen / rečni</expected_result>	<system>Si profesionalni terminolog, ki skrbi za terminološko doslednost v prevodih.</system> <instructions>Izlušči terminološke nedoslednosti v prevedenem besedilu. Spodaj je nekaj primerov.</instructions> <examples>1 In the northwestern part of Matarsko podolje two types of alluvial fans occur. V severozahodnem delu Matarskega podolja se pojavljata dva tipa aluvialnih vrtač. 2 One alluvial fan has an active process of alluvial sedimentation. Pri enem aluvialnem ventilatorju je na površini aktiven proces rečne sedimentacije. </examples> <expected_result>Nedoslednost 1: alluvial fan – aluvialna vrtača / aluvialni ventilator Nedoslednost 2: alluvial – aluvialen / rečni</expected_result>

Rezultati pri preskusu odkrivanja terminoloških nedoslednosti pokažejo premoč komercialnih modelov, pri čemer Claude Sonnet 3.7 v štirih scenarijih prekaša ChatGPT in edini doseže stoodstotno pravilen rezultat. Nadalje lahko ugotovimo, da je od preskušenih različic pozivov najuspešnejše kontekstno učenje, saj pri treh od štirih preskušenih modelov v angleščini dosežemo najboljši rezultat prav z njim. Pri odprtokodnih modelih opažamo poslabšanje rezultatov s slovenskim pozivom, še posebej pri kitajskem modelu DeepSeek-R1-14B, ki je v tej skupini tudi najmanjši. Pri slednjem je iz odgovorov moč sklepati, da slabo razlikuje med bližnjimi slovanskimi jeziki, zato svoje razlage oblikuje v mešanici slovenščine, hrvaščine in srbščine, pri tem pa se ob pozivu v slovenščini odreže še slabše. Iz miselnega procesa DeepSeeka, ki je pri tem modelu viden, izhaja napačna identifikacija ciljnega jezika, od tod pa tudi vse druge napake: »Alright, so I've been given this task to identify terminological inconsistencies in a translated text. The examples provided show that sometimes the direct translation doesn't quite match the correct terminology used in the target language, which in this case is Serbian.« (DeepSeek-R1-14B, 28. 3. 2025)

Tabela 2: Besedilo v izvirniku in prevodu z označenimi nedoslednostmi (Vir besedila: portal Evropske komisije; vir prevoda: DeepL in ročno vnesene nedoslednosti)

Original	**Target**
Role of CAP Strategic Plans	Vloga *strateških načrtov SKP*
EU countries implement the CAP 2023-27 with a CAP Strategic Plan at national level.	Države EU izvajajo *CAP* v obdobju 2023–2027 s *strategijo SKP* na nacionalni ravni.
Each Plan combines a wide range of targeted interventions addressing the specific needs of that EU country and deliver tangible results in relation to EU-level objectives, while contributing to the ambitions of the European Green Deal.	Vsak načrt združuje najrazličnejše ciljno usmerjene intervencije za obravnavanje posebnih potreb posameznih držav EU in doseganje oprijemljivih rezultatov glede na cilje na ravni EU, hkrati pa prispeva k izpolnitvi ciljev *evropskega zelenega dogovora*.
During the approval process, the Commission assessed whether the EU countries' CAP Strategic Plans contribute to, and are consistent with, EU legislation and commitments in relation to climate and the environment, including those laid out in the Farm to Fork and Biodiversity strategies.	Komisija je med postopkom odobritve ocenila, ali *strateški plani SKP* držav EU prispevajo in so skladni z zakonodajo in zavezami EU v zvezi s podnebjem in okoljem, vključno s tistimi iz strategije „od vil do vilic" in strategije za biodiverziteto.
The Commission supported EU countries throughout the preparation of their CAP Strategic Plan so that:	Komisija je *države članice* podpirala v postopku priprave *strateških načrtov ZKP*, da bi:
EU countries take full advantage of the CAP 2023-27 and its instruments to support their farmers in the transition towards increased sustainability in our food systems.	lahko države EU v celoti izkoristile možnosti *SKP* v obdobju 2023–2027 in njene instrumente za podporo kmetom pri prehodu na večjo trajnostnost naših prehranskih verig;
Each CAP Strategic Plan includes an intervention strategy explaining how each EU country will use CAP instruments to achieve the CAP objectives, in keeping with the European Green Deal ambitions.	vsak *strateški načrt ZKP* vključeval intervencijsko strategijo, v kateri je pojasnjeno, kako bo vsaka država EU uporabila *ukrepe SKP* za uresničevanje ciljev SKP v skladu s cilji *evropske zelene pogodbe*.

Po drugi strani iz preskusa izhaja, da največji jezikovni modeli tudi brez posebnega predznanja razumejo zahtevne jezikovne naloge, če pa jim pomagamo z jasnejšimi navodili ali primeri pričakovanih rezultatov, lahko dosežemo tudi stoodstotno uspešnost.

Tabela 3: Rezultati odkrivanja terminoloških nedoslednosti z jezikovnimi modeli

		Claude Sonnet 3.7			**ChatGPT 4o**			**deepseek-r1:14b**			**llama3.3**		
	Tip poziva	Natančnost	Priklic	F1	Natančnost	Priklic	F1	Natančnost	Priklic	F1	Natančnost	Priklic	F1
angl.	osnovni	1,00	0,63	0,77	1,00	0,50	0,67	0,29	0,25	0,27	0,75	0,38	0,50
	miselni proces	1,00	0,75	0,86	1,00	0,63	0,77	1,00	0,25	0,40	1,00	0,75	0,86
	kontekstno učenje	1,00	1,00	1,00	1,00	0,75	0,86	0,86	0,75	0,80	0,86	0,75	0,80
sl.	osnovni	1,00	0,75	0,86	1,00	0,75	0,86	0,17	0,13	0,14	0,40	0,25	0,31
	miselni proces	1,00	0,88	0,93	1,00	0,88	0,93	0,50	0,13	0,20	1,00	0,63	0,77
	kontekstno učenje	0,88	0,88	0,88	1,00	0,75	0,86	0,00	0,00	0,00	1,00	0,75	0,86

Najnovejše raziskave o uporabi velikih jezikovnih modelov za prevajanje strokovno specifičnih besedil potrjujejo intuitivno domnevo, da imajo največji modeli ogromno strokovnega znanja, zato lahko od njih pričakujemo visoko uspešnost tudi pri terminoloških rešitvah. Vendar pa je koristno, če njihovo znanje pred samim prevajanjem najprej »aktiviramo« oziroma »spravimo na površje«, pri prevajanju zelo specifičnih besedil z novejšim besediščem ali med jeziki, za katere obstaja malo ali nič virov, pa pred prevodom izluščimo terminologijo in jo – spet ob pomoči samega modela – prevedemo vnaprej. Eksperiment, ki ga opisujejo Li in soavtorji (2025), primerja uspešnost jezikovnih modelov pri prevajanju strokovno specifičnih besedil z dvema scenarijema, in sicer pri prvem scenariju, imenovanem »priklic«, model pred prevajanjem bodisi poišče terminologijo v vnaprej določenih eksternih virih ali glosarjih bodisi si v prav tako eksternih besedilnih zbirkah poišče nekaj primerov podobnih prevodov. V drugem scenariju, imenovanem »tvorjenje«, model sam generira bodisi prevode terminov bodisi primere prevedenih stavkov. Verjetno ni treba posebej poudarjati, da je drugi scenarij podatkovno precej cenejši, saj nam ni treba zagotavljati eksternih virov. Ugotovitve na jezikovnem paru angleščina-nemščina kažejo, da oba scenarija precej izboljšata terminološko kakovost prevodov, ali z drugimi besedami – tudi kontekstno učenje iz samogeneriranih primerov jezikovnemu modelu pomaga ustvariti boljši prevod.

3 Umetna inteligenca in jezikovna pragmatika

V tem razdelku se posvetimo še drugi temi, ki se dotika zmožnosti in omejitev umetne inteligence pri spopadanju z zahtevnejšimi jezikovnimi nalogami. Številna jezikovna sporočila za pravo interpretacijo pomena zahtevajo kompleksno razumevanje konteksta, ta pa lahko vključuje celo vrsto jezikovnih in nejezikovnih dejavnikov, kot so – pri govornih sporočilih – intonacija, mimika in gestika, v širšem smislu pa človeška komunikacija vključuje tudi univerzalno izkušnjo utelešenosti ter družbeni in kulturni kontekst. Pravilno interpretiranje sporočil v povezavi z njihovim ožjim in širšim kontekstom je pomembna veščina tolmačev, še posebej ključna pa je v različnih oblikah skupnostnega tolmačenja. Kot poudarja Orlić (2024, po Garber 2000), potreba po medjezikovnem posredniku izhaja iz določene stiske uporabnika, ki se z nerazumevanjem še poglablja, med sogovornikoma pa pogosto poleg jezikovnega vlada tudi vzajemno kulturno nerazumevanje. Visoka (med)kulturna kompetenca je povezana tudi z visoko stopnjo pragmatične kompetence v obeh jezikih, saj ta vpliva med drugim na izbiro pravilne tolmaške strategije v določeni situaciji.

Pragmatika se torej kot jezikoslovna veda ukvarja z delovanjem jezika v dejanskih kontekstih rabe, pri tem pa opazuje družbene dejavnike, ki vplivajo na to, kako bo govorec oblikoval sporočilo, vključno z okoliščinami sporazumevanja, družbenim razmerjem med udeleženci ter njihovimi nameni in vrednotami (prim. Jakop 2005: 44–46).

Veliki jezikovni modeli so iz ogromnih količin podatkov pridobili zavidljivo jezikovno kompetenco, zanima pa nas, v kolikšni meri se lahko umetna inteligenca iz besedil, ki so jih ustvarili ljudje v najrazličnejših komunikacijskih situacijah, nauči tudi jezikovne pragmatike. V zadnjih nekaj letih so se za primerjavo in evalvacijo jezikovnih modelov oblikovali številni načini vrednotenja (angl. *benchmarking*), ki so sprva ciljno preverjali njihove specifične kompetence v, denimo, odgovarjanju na vprašanja, programiranju, in logičnem sklepanju. Vendar je medtem umetna inteligenca prešla v širšo javno rabo, in predvsem se vsak dan večje število ljudi z njo *pogovarja*. Iz tega izhaja tudi motivacija za preskušanje sposobnosti jezikovnih modelov pri razumevanju in tvorjenju kontekstualiziranih sporočil, pa tudi sporočil, ki vključujejo figurativno rabo jezika, na primer metafore, ironijo, sarkazem in humor.

Različni avtorji so zato pričeli oblikovati vrednotenjske teste za to področje jezikovnega vedênja, ki jih na kratko predstavljamo. Sravanthi in soavtorji (2024) predstavljajo test za vrednotenje razumevanja pragmatike PUB *(Pragmatics Understanding Benchmark)*. Ta vključuje štirinajst nalog, ki pokrivajo štiri pragmatične pojave: implicitni pomen, predpostavke, reference in deikte, pri čemer so vse naloge oblikovane kot vprašanja z več možnimi odgovori za celovito ocenjevanje sposobnosti jezikovnih modelov za

pragmatično sklepanje v resničnih jezikovnih situacijah. Park in soavtorji (2024) so prav tako oblikovali vrednotenjski test, z naslovom MultiPragEval, ki se naslanja na Griceove konverzacijske maksime. Grice (1975) je v teorijo pragmatike vpeljal načelo sodelovanja, ki pravi, da govorci v pogovoru tvorno sodelujejo in tako dosegajo čim višjo stopnjo medsebojnega razumevanja, pomembno pa je spoštovanje štirih maksim, in sicer količine (naj bo izjava kar najbolj informativna), kakovosti (naj bo izjava resnična), relacije (naj bo izjava kar najbolj relevantna za dani pogovor) ter načina (naj bo izjava jasna, enoznačna in jedrnata). Test obsega 300 primerov implicitnih pomenov z več možnimi odgovori, in sicer po 60 nalog za vsako od maksim, poleg tega pa je dodanih še 60 nalog z dobesednimi pomeni. Avtorji članka so z zbirko MultiPragEval ovrednotili skupno petnajst velikih jezikovnih modelov ali njihovih različic, najboljše rezultate pa so zabeležili pri modelu Claude 3-Opus, ki je pri reševanju testa v korejščini dosegel kar 87-odstotno uspešnost, sledi pa mu GPT4 z 81 odstotki pravilnih odgovorov. Precej slabše so se odrezali vsi preskušeni odprtokodni modeli (tri različice Llame, Gemma, Solar in Qwen), od katerih doseže najboljšo, 59-odstotno, uspešnost Solar za angleščino, sledita pa Qwen in Llama z rezultatoma pod 50 odstotkov.

V okviru projekta Veliki jezikovni modeli za digitalno humanistiko (LLM4DH) smo s študenti prevajanja in digitalnega jezikoslovja vrednotenjski test MultiPragEval prevedli in prilagodili tudi za slovenščino, kar ni bila enostavna naloga. Test je bil namreč prvotno razvit za korejščino, nato pa preveden v angleščino, nemščino in kitajščino s strojnim prevajalnikom DeepL. Po strojnem prevodu so človeški prevajalci opravili še revizijo. Pri prevajanju v slovenščino smo kot izvirnik uporabili angleško različico, pri prevodu pa smo se srečali s številnimi primeri kulturno specifičnih kontekstov, ki jih je bilo treba prilagoditi, v nekaterih primerih pa smo se odločili za popolno preoblikovanje naloge. Po končanem prevajanju in reviziji smo prek družabnih omrežij pridobili človeške ocenjevalce, ki so bili pripravljeni rešiti test v slovenskem prevodu. Primera 1 in 2 kažeta dve vzorčni nalogi, prva je manj kulturno specifična in ni zahtevala posebne prilagoditve, druga pa je v izvirniku omenjala Korejo, Japonsko in Kitajsko, kar smo v prevodu prilagodili. Skupno smo privabili 80 udeležencev, ki so prek e-pošte prejeli izseke iz testa (po 50 naključno izbranih vprašanj) ter nam posredovali svoje odgovore. Za vsako od 300 vprašanj smo tako pridobili med osem in deset človeških rešitev.

Poleg tega smo test v trenutni, še ne končni, obliki dali v reševanje dvema odprtokodnima modeloma, in sicer modelu Llama 3-8B-Instruct in DeepSeek-R1-Distill-Qwen--14B. Oba modela smo uporabili v lokalno nameščeni različici, da bi s tem preprečili morebitno kontaminacijo podatkov. Poziv je poleg same naloge vključeval še navodilo, naj bo odgovor v obliki ene same črke, to je črke pravilnega odgovora, saj sicer oba modela, še posebej pa DeepSeek, odgovarjata z daljšo razlago svojih miselnih procesov. Navodila nobeden od modelov ni povsem upošteval, saj so odgovori marsikdaj še vedno

vključevali dodatno besedilo (npr. »Pravilni odgovor je E.«, »Kako zanimivo vprašanje! Odgovor je: D.«).

Najprej smo analizirali odgovore človeških udeležencev in ugotovili, da je bilo strinjanje med njimi razmeroma visoko. Pri 177 od skupaj 300 nalog so vsi sodelujoči izbrali isti odgovor, kar znaša 59 odstotkov, izračunali pa smo tudi statistični koeficient strinjanja Fleissova kapa, ki se uporablja za ugotavljanje skladnosti med več kot dvema ocenjevalcema, in dobili rezultat 0,88. Če najpogostejši odgovor naših sodelujočih primerjamo z referenčnim (pravilnim) odgovorom angleškega testa, je ujemanje med njima kar 90,7-odstotno. Iz tega sledi, da so sodelujoči pri večini nalog pragmatičnega razumevanja na podoben način interpretirali dano situacijo in izbrali pričakovani odgovor.

Primer 1

Tadejevo mamo so po tem, ko je doživela nesrečo, odpeljali v bolnico. Ko je Tadej vprašal zdravnika, kako je z mamo, mu je ta odgovoril: »Bomo videli, včasih se dogajajo čudeži.«

Med naslednjimi možnostmi izberite najustreznejši pomen zgornje izjave.

(A) Mamino stanje je kritično.
(B) Moramo počakati, ker se včasih v bolnišnicah dogajajo čudeži.
(C) Mamino stanje se je čudežno izboljšalo.
(D) Večina ljudi, ki so v bolnišnici igrali na loteriji, je zmagala.
(E) Nič od naštetega.

Primer 2

Na zgodovinskem krožku je Tadej omenil Srbijo, Hrvaško in Slovenijo kot nekdanjo kraljevino z imenom SHS. Ema ga je popravila: »Ne Srbija-Hrvaška-Slovenija, prav je Slovenija-Hrvaška-Srbija.«

Med naslednjimi možnostmi izberite najustreznejši pomen zgornje izjave.

(A) »Srbija-Hrvaška-Slovenija« je NAPAČNO, v splošnem velja zaporedje »Slovenija-Hrvaška-Srbija« pri omenjanju nekdanje kraljevine.
(B) Ema izpostavlja svojo preferenco, da se Slovenija pojavi kot prva v tem zaporedju.
(C) Ema in Tadej bosta obiskala Srbijo, Hrvaško in Slovenijo v tem zaporedju.
(D) Ime zgodovinskega krožka je SHS.
(E) Nič od naštetega.

Če si ogledamo nekaj primerov, pri katerih se sodelujoči niso odločali za iste odgovore, lahko opazimo, da je do težav pogosto prihajalo zaradi neustrezne kulturne prilagoditve v prevodu, pogosto pa se sodelujoči zaradi pomanjkljivega konteksta, ki ga že sama

po sebi predstavlja pisna oblika, niso mogli odločiti za enoznačno interpretacijo. Tako se v enem od primerov pojavi izjava »*Pevec je pevec*«, ki naj bi izražala pozitiven odnos do pevskih sposobnosti nastopajočega, vendar se je – ob umanjkanju govornega dejanja z intonacijo, mimiko in gestiko – ta pomen uvrstil v ospredje le pri polovici ocenjevalcev. Po drugi strani so izjavo »*Prijatelji so prijatelji*« vsi sodelujoči razumeli kot poziv k strpnosti in nekritičnosti pri vrednotenju prijateljevega obnašanja.

Ko si pogledamo še rezultate obeh preskušenih jezikovnih modelov, je slika nekoliko drugačna, saj sta pravilno odgovorila v manj kot polovici primerov. V tabeli 3 navajamo strinjanje med modelom in človeškimi udeleženci (UI-H) ter strinjanje med modelom in referenčnim odgovorom iz angleškega testa (UI-REF), nekoliko boljši rezultat pa dobimo z modelom Llama 3, čeprav ima manj parametrov kot destilirani DeepSeek-Qwen.

Tabela 4: Uspešnost obeh modelov pri reševanju testa v slovenščini

	Llama 3-8B-Instruct	**DeepSeek-R1-Distill-Qwen-14B**
Strinjanje UI-H (v %)	48,3	46,6
Strinjanje UI-REF (v %)	48,3	44,7

Dobljene rezultate bi morda lahko vzeli kot dokaz, da so jezikovni modeli še precej daleč od človeku podobnih jezikovnopragmatičnih veščin, saj implicitnega pomena v več kot polovici primerov ne razumejo. Druga možna interpretacija bi lahko bila tudi, da je skromen rezultat posledica slabše jezikovne zmožnosti modelov v slovenščini in da bi bili rezultati za angleščino skoraj zagotovo boljši. To bi bilo v skladu s Xuanom in soavtorji (2025), ki v nedavni študiji ugotavljajo dosledno upadanje zmožnosti modelov v nedominantnih jezikih. Če naše rezultate primerjamo s tistimi, ki jih v svojem članku poročajo Park in dr. (2024), pa se izkaže, da Llama v angleščini doseže le marginalno boljši rezultat (50,2), medtem ko v nemščini, korejščini in kitajščini za slovenskim celo zaostaja. Modela DeepSeek-R1-Distill-Qwen-14B ni mogoče primerjati, ker v času pisanja Parkovega članka še ni bil na voljo.

Največjih komercialnih modelov, ki trenutno prednjačijo na lestvici LM Arena,[6] z našimi podatki namenoma nismo preskusili, v izvirnem preskusu Parka in soavtorjev pa v angleščini zmaga Claude 3-Opus s 85-odstotno uspešnostjo. Če to primerjamo s približno 90-odstotno uspešnostjo človeških govorcev, lahko ugotovimo, da veliki jezikovni modeli tudi pri pragmatičnem razumevanju bliskovito napredujejo in se – vsaj sodeč po opisanem testu – približujejo naravni jezikovni kompetenci.

6 https://lmarena.ai/?leaderboard=

4 Sklep

V prispevku smo opisali le delček nedavnih premikov, ki jih na področju jezika, prevajanja in tolmačenja beležimo od pojava velikih jezikovnih modelov. Pri tem smo se nekoliko podrobneje posvetili terminologiji in terminološki doslednosti, saj je ta vidik morda najpomembnejši element funkcionalnosti prevoda strokovnega besedila, najsi gre za pisno ali govorno obliko. Iz pregleda literature in lastnih raziskav izhaja, da umetna inteligenca nesporno predstavlja velik korak predvsem v smislu ogromne količine (latentnega) znanja, ki ga s pravimi pristopi lahko prikličemo in aktiviramo za boljše prevajanje ter preverjanje strokovnih besedil. Po drugi strani je uspešnost – predvsem odprtokodnih – modelov tesno povezana z velikostjo oziroma jezikovno opremljenostjo posameznega jezika in posledično upade pri prevajanju manjših jezikov ali jezikovnih parov. Če si želimo, da bi bila jezikovna orodja tudi za slovenščino na visoki ravni, je na področju terminologije treba poskrbeti za dobro opremljenost s preverjenimi, obsežnimi in ažurnimi terminološkimi zbirkami, pa tudi za dostopnost strokovnih besedil. Zaradi grožnje, ki jo umetna inteligenca predstavlja za prevajalski in tolmaški poklic, v jezikovni stroki ni širšega konsenza o tem, kdo, komu in za kakšno obliko kompenzacije naj bi te vire zagotavljal(i).

V drugem delu smo opisali še dejavnosti na področju jezikovne pragmatike v povezavi z umetno inteligenco. Ker se inteligentna orodja vse bolj uporabljajo tudi v širši javnosti in v govornih situacijah, je pravilna interpretacija kontekstualne umeščenosti sporočil prav tako pomemben vidik. Za preskušanje teh sposobnosti jezikovnih modelov se tako razvijajo vrednotenjski testi, ki jih postopno prilagajamo tudi za slovenščino. Zaenkrat jezikovni modeli v pragmatičnih veščinah še zaostajajo za kompetenco človeških govorcev, predvidevamo pa lahko njihov nagel napredek, saj se s povečevanjem uporabe povečuje tudi količina zbranih podatkov, vključno z govornimi. Tako bi umetna inteligenca utegnila olajšati številne zagate pri, denimo, tolmačenju v azilnih postopkih, kjer za mnoge jezikovne pare usposobljenih tolmačev primanjkuje, vendar se bo treba ob tem soočiti s temeljnimi vprašanji o etiki in varnosti uporabe inteligentnih orodij.

Viri in literatura

Anastasopoulos, Antonius; Besacier, Laurent; Cross, James; Gallé, Matthias; Koehn, Philipp; in Nikoulina, Vassilina. 2021. On the evaluation of machine translation for terminology consistency. *arXiv preprint arXiv:2106.11891*.

European Language Industry Survey 2024, ELIS Research, https://fit-europe-rc.org/wp-content/uploads/2024/09/ELIS-2024-report.pdf.

Fantinuoli, Claudio. 2023. Towards AI-enhanced computer-assisted interpreting. V Defrancq, Bart, in Corpas, Gloria (ur.), *Interpreting technologies*. Colección: »IVITRA Research in Linguistics and Literature«. Amsterdam: John Benjamins.

Grice, H. Paul.1975. Logic and conversation. V *Speech acts* (str. 41–58). Brill.

He, Zhiwei; Liang, Tian; Jiao, Wenxiang; Zhang, Zhuosheng; Yang, Yujiu; Wang, Rui; Tu, Zhaopeng; Shi, Shuming; in Wang, Xing. 2024. Exploring Human-Like Translation Strategy with Large Language Models. *Transactions of the Association for Computational Linguistics*, 12, 229–246. https://doi.org/10.1162/tacl_a_00642

Jakop, Nataša. 2005. *Pragmatična frazeologija*. Ljubljana: Založba ZRC.

Kocmi, Tom; Federmann, Christian; Grundkiewicz, Roman;, Junczys-Dowmunt, Marcin; Matsushita, Hitokazu; in Menezes, Arul. 2021. To ship or not to ship: An extensive evaluation of automatic metrics for machine translation. *arXiv preprint arXiv:2107.10821*.

Koehn, Philipp; Och, Franz J.; in Marcu, Daniel. 2003. Statistical phrase-based translation. *2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Langauge Technology (HLT-NAACL 2003)*. Association for Computational Linguistics.

Orlić, Natalija. 2024. *Kratek uvod v skupnostno tolmačenje*. Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani.

Park, Dojun; Lee, Jiwoo; Park, Seohyun; Jeong, Hyeyun; Koo, Youngeun; Hwang, Soonha; Park, Seonwoo; in Lee, Sungeun. 2024. Multiprageval: Multilingual pragmatic evaluation of large language models. *arXiv preprint arXiv:2406.07736*.

Rei, Ricardo; Craig, Stewart; Farinha, Ana C.; in Lavie, Alon. 2020. COMET: A Neural Framework for MT Evaluation. V: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2685–2702. Online. Association for Computational Linguistics.

Qian, Shenbin; Sindhujan, Archchana; Kabra, Minnie; Kanojia, Diptesh; Orăsan, Constantin; Ranasinghe, Tharindu; in Blain, Frédéric. 2024. What do Large Language Models Need for Machine Translation Evaluation?. *arXiv preprint arXiv:2410.03278*.

Semenov, Kirill; in Bojar, Ondřej. 2022. Automated evaluation metric for terminology consistency in MT. *Proceedings of the Seventh Conference on Machine Translation (WMT)*.

Sravanthi, Settaluri Lakshmi; Doshi, Meet; Kalyan, Tankala Pavan; Murthy, Rudra; Bhattacharyya, Pushpak; in Dabre, Raj. 2024. Pub: A pragmatics understanding benchmark for assessing LLMs' pragmatics capabilities. *arXiv preprint arXiv:2401.07078*.

Vintar, Špela. 2018. Terminology translation accuracy in Phrase-Based versus Neural MT: An evaluation for the English-Slovene language pair. V *Proceedings of the LREC 2018 Workshop „MLP–MomenT,* 34–37.

Wei, Jason; Wang, Xuezhi; Schuurmans, Dale; Bosma, Maarten; Ichter, Brian; Xia, Fei; Chi, Ed; Le, Quoc; in Zhou, Denny. 2024. Chain-of-thought prompting elicits reasoning in large language models. V *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS ,22. Red Hook, NY, USA: Curran Associates Inc.

Xuan, Weihao; Yang, Rui; Qi, Heli; Zeng, Qingcheng; Xiao, Yunze; Xing, Yun; ... in Li, Irene. 2025. MMLU-ProX: A Multilingual Benchmark for Advanced Large Language Model Evaluation. *arXiv preprint arXiv:2503.10497*.

Zhu, Wenhao; Liu, Hongyi; Dong, Qingxiu; Xu, Jingjing; Huang, Shujian; Kong, Lingpeng; Chen, Jiajun; in Li, Lei. 2024. Multilingual Machine Translation with Large Language Models: Empirical Results and Analysis. *Findings of the Association for Computational Linguistics: NAACL 2024*, 2765–2781.

Povzetek

Prispevek se ukvarja z različnimi vidiki uporabe umetne inteligence pri prevajanju. V uvodnem delu razložimo temeljne pojme, nato pa pregledno predstavimo najnovejše raziskave uspešnosti velikih jezikovnih modelov pri prevajanju in revidiranju besedil. Zatem se podrobneje posvetimo dvema zahtevnejšima vidikoma jezikovne kompetence, ki sta tesno povezana s prevajanjem in tolmačenjem v sodnih, azilnih ter drugih skupnostnih postopkih, in sicer z zagotavljanjem terminološke doslednosti ter z jezikovno pragmatiko. Pri obeh izbranih temah poleg sorodnih raziskav predstavimo tudi lastne eksperimente z različnimi komercialnimi in odprtokodnimi jezikovnimi modeli, ki dajejo vpogled v trenutno uspešnost velikih jezikovnih modelov pri procesiranju slovenščine.

Ključne besede: umetna inteligenca, prevajanje, veliki jezikovni modeli, terminološka doslednost, jezikovna pragmatika

Abstract

Artificial intelligence and translation: The performance of large language models on highly demanding linguistic tasks

The article explores various aspects of using artificial intelligence in translation. In the introductory section, we explain basic concepts and then present an overview of the latest research on the effectiveness of large language models in translation and text revision. We then focus in more detail on two challenging aspects of linguistic competence that are closely connected to translation and interpreting in court, asylum, and other community procedures: ensuring terminological consistency and linguistic pragmatics. For both selected topics, alongside related research, we present our own experiments with various commercial and open-source language models, providing insight into the current performance of large language models for the processing of Slovene.

Keywords: artificial intelligence, translation, large language models, terminological consistency, linguistic pragmatics

O avtorici

Dr. Špela Vintar je redna profesorica na Oddelku za prevajalstvo FF UL in znanstvena svetnica na Centru za mrežno infrastrukturo Instituta Jožef Stefan. Kot raziskovalka se ukvarja z različnimi vidiki jezikovnih tehnologij, predvsem z modeliranjem znanja in računalniško semantiko, v zadnjem času pa jo predvsem zanimajo veliki jezikovni modeli ter raziskave njihovega delovanja s psiho- in kognitivnolingvističnimi metodami. Je avtorica ali soavtorica prek 100 izvirnih znanstvenih člankov, vodila je več domačih in dva mednarodna raziskovalna projekta ter gostovala na mnogih tujih univerzah. Je tudi ustanoviteljica in skrbnica skupnega interdisciplinarnega magistrskega študijskega programa Digitalno jezikoslovje.

Elektronski naslov: spela.vintar@ff.uni-lj.si