Original scientific paper Received: 25. III. 2025. Accepted: 5. V. 2025.

UDK 81'374:004.89
811.163.42'373.613:811.111]:004.89
https://doi.org/10.33604/10.33604/sl.19.36.3

The lexicographic potential of artificial intelligence: a case study of English loanwords in the Croatian language

Katica Balenović
University of Zadar
kbalenovic@unizd.hr
Jakov Proroković
University of Zadar
jprorokov@unizd.hr

ABSTRACT: The advent of generative artificial intelligence (AI) and large language models (LLMs) has introduced new possibilities in lexicography, particularly in defining dictionary entries with precision, while reducing the time cost compared to more traditional methods or software tools. To test AI's linguistic capabilities, our study goes beyond monolingual dictionary compilation and investigates the potential of the ChatGPT model in distinguishing between specific senses of loanwords in an L2 context. A corpus-based sampling of target English words was used to assess ChatGPT's ability to delineate different word senses in which regularly occurring loanwords can be realised in the Croatian language context. The findings indicate that AI demonstrates notable proficiency in providing definitions in general, albeit with observable flaws when responding to prompts that specifically inquire about the possible senses or word classes of targeted loanwords in their L2 setting. Its accuracy diminishes when dealing with less frequently used loanwords, often exhibiting overgeneralisation from English (L1) to Croatian (L2). The AI's tendency to produce erroneous examples, with suggested usages that lack attestation in language corpora, is discussed in detail, with the results supporting the notion that the model primarily interprets loanwords from an English perspective, regardless of the language used in the prompt. A comparison between AI responses from early 2024 and early 2025 suggests an improvement in the 2025 model, which exhibits a more nuanced handling of ambiguous cases. However, inconsistencies persist, particularly in how frequency of use correlates with the number of senses, much of which is interpreted as ChatGPT's tendency to sometimes prioritise generating a response at the cost of accuracy.

Keywords: ChatGPT; lexicography in language contact; overgeneralisation errors; corpus-based sampling; loanwords

https://orcid.org/0000-0002-9534-875X [Katica Balenović]
https://orcid.org/0000-0002-4496-8908 [Jakov Proroković]
https://ror.org/00t89vb53 [University of Zadar]

1. Introduction

While relatively recent in terms of availability to the general public, the rise of generative artificial intelligence (AI) and large language models (LLM) has quickly been recognised as a tool with immense potential in the field of lexicography, which—despite its acknowledged limitations and errors (Lew, 2023; Rees & Lew, 2023; Martin 2024, McKean & Fitzgerald, 2024; Harnad, 2025; Kalaš, 2025, etc.)—demonstrates remarkable abilities in defining dictionary entries compared to previous software tools (see Yin & Skiena, 2023; Lew et al., 2024; Klosa-Kückelhaus & Tiberius, 2025). The findings on how well the AI models do when it comes to providing definitions, examples, common phrases, related forms, and completing other dictionary compiling-related tasks are relatively consistent, but with varying assessments of AI prowess depending on the task at hand (for overviews, see De Schryver, 2023; Rundell, 2024). This study's exploration is motivated by the relative novelty of research in lexicographic applications of artificial intelligence, and the fact that its capabilities—in providing definitions and examples for foreign words verifiably employed in the observed recipient language—have yet to be tested. Moreover, the exploration of lexicographic potentials becomes even more challenging when linguistic interference is considered (cf. Chow et al., 2024; Li & Tarp, 2024; Merx et al., 2024, etc.), which seems to be the reason why the research on this topic is rather scarce, and why we delve into the issue of loanword meaning identification in L2 context.

As far as the mechanism by which they operate is concerned, the AI tools such as ChatGPT, DeepSeek and Gemini are best described as deep neural networks specialised in natural language processing (NLP) tasks, the functionality of which is based on modelling linguistic patterns and performing statistical analysis on vast amounts of textual data (see Carlini et al., 2021; Xu et al., 2022; Floridi, 2023; Min et al., 2023, etc.). More specifically, these models are trained on extensive datasets primarily through unsupervised learning (drawing from sources such as internet content, books, and articles), which is then followed by fine-tuning, a process of discriminative adjustment and optimisation, aimed at enhancing their ability to »understand« linguistic patterns, respond to queries, and generate contextually relevant text (Radford et al., 2018). Their architecture enables them to process input data, interpret contextual meaning, and generate responses or text that are both coherent and relevant to users. Apart from being able to generate text in various structural and stylistic formats, the ability of ChatGPT to refine its performance on similar tasks is what makes it impressive, although it is important to emphasise that this does not equate to reasoning or emotion-based feedback; rather, »it compares existing data to draw the most likely (e.g., the most frequent and relevant) responses« (Hong, 2023, p. 38; also see Jiang et al., 2020). On the other hand, some researchers

argue that generating conclusions and responses based on statistical analysis is, in fact, at the core of intelligent-like systems, with these models representing the first instances of something that illustrates how language comprehension and intelligence can be decoupled from the physical and emotional characteristics traditionally associated with humans and animals (Aguera-Arcas, 2022).

2. Previous Research on the Application of AI Models

The application of AI models in educational settings has been increasing at rapid rates since the breakout of highly evolved LLMs such as ChatGPT, with several key advantages being identified in the academic context, including personalised learning, lesson planning, language learning, etc. (see Kasneci et al., 2023). More specifically, in the context of language-related tasks and applications, the numerous capabilities include: emphasising key phrases, generating summaries and translations, explaining grammar and vocabulary, suggesting improvements in grammar or style, assisting with conversational practice, providing feedback to students, identifying and correcting typographical errors, and recognising opportunities to enhance writing styles tailored to specific topics (ibid., 2023, p. 3). However, when reviewing AI-generated texts, educators often notice superficial content, occasional inaccuracies, incorrect paraphrasing of existing research, or even fabricated references, while proper citations within the text and in the bibliography are frequently missing, or, what is even more concerning, the text may include entirely fictitious sources (see King & ChatGPT, 2023; Rudolph, 2023). The problematic and often fabricated information is particularly interesting in lexicography, where accuracy and the summation of relevant information play a crucial role. Two recent studies, one on 166 and the other on 223 university students, found that ChatGPT significantly outperforms the web version of the *Longman Dictionary of Contemporary English* (LDOCE) in both language reception and production tasks (see Lew et al., 2024, Ptasznik et al., 2024). The results suggest that AI-based chatbots like ChatGPT constitute strong competitors to traditional dictionaries in supporting advanced language learners, particularly in English, with their greatest advantage being in production tasks, where they can generate expected phrasal verbs but also help students convey meaning that is less deterministic in terms of employing specific lexical items. In the context of our study, it is the ability to produce a definition, not anchored in already existing repositories or dictionaries such as LDOCE, which requires a multi-layered understanding of the prompt at hand, but, more importantly, it also requires a multi-level operation on behalf of the chatbot. It is rather atypical to engage in a lexicographic endeavour in terms of analysing the meaning of loanwords (especially those retaining a form that has not been fully adopted and adapted in the

target language), and this atypicality is one of the ways to test the AI's adeptness at providing definitions that are more contextually anchored and language specific.

In this sense, it is worth noting that restricting the lexicographic research to dictionary-compiling studies observing solely the AI's feedback when it comes to the definition of the targeted entries might not be enough, for it is not sufficient to summarise the available recorded data on language use, but it is also necessary to determine its relevance regarding presence, context, potential productivity, etc., which requires a »sensorimotor grounding« of sorts (see Harnad, 2025). To test the requisite multi-level approach to meaning, the newer research has also focused on tasks going beyond the dictionary-compiling or definition-oriented tasks. For example, when it comes to addressing phrases not rooted solely in written or oral language presence, the findings indicate relatively high agreement between human and ChatGPT evaluation of neologisms such as blends or derivatives (Georgiou, 2025). In other words, the findings suggest that AI primarily captures the most common or dominant interpretation and excels in processing form-based linguistic cues, whereas when it comes to the requisite extralinguistic knowledge and multi-level approach to linguistic phenomena, it tends to struggle with meanings that require broader contextual or world knowledge (ibid.). In our research, we primarily seek to challenge its capacity to properly ascertain meanings of loanwords in the L2 language context, and more importantly, to demonstrate its ability to delineate different word senses concerning prompt-design differences. This entails that we expect it not only to provide accurate feedback on the presented question about the use of the target loanword but also to understand what exactly is being asked. Though it may seem that one cannot be realised without the other (i.e. accurate feedback without prompt-understanding), our data will demonstrate that answers vary not depending on the prompt (for it stays the same), but on the loanword at hand, thus exhibiting a nondeterministic character that goes beyond the mere outline and structure of the response, and directly influences the content itself.

Our study aims to assess the capacity of artificial intelligence tools to distinguish meanings of English loanwords in the Croatian language compared to their usage in English and to evaluate the quality of responses based on the frequency of English loanwords in the Croatian language.[1] More specifically, it aims to assess the

[1] Note that the terminology surrounding linguistic borrowing has not been fully standardised when it comes to the interaction between English and Croatian, leading to various approaches in the clarification of adaptation degrees and related operationalisations of the corresponding terminology (e.g. *anglicisms, English loanwords, English words* etc.), differing from one author to another. Some use the term *anglicisms* to address those words that have undergone full adaptation to Croatian in terms of spelling, phonology, and morphology (Filipović, 1990), while others include any English-origin words that have been integrated into Croatian after some degree of adaptation (Opačić, 2012), or those borrowed from English, regardless of their level of adaptation (Drljača Margić, 2009; Patekar, 2019). In

potential of AI in addressing differences between specific senses in which L1 words can be used in L2, and how this compares to the meanings in which they are used in L1.[2] The targeted loanwords and the selection criteria are based on the work of Bogunović & Kučić (2022) and Bogunović (2023), who compiled a list of English words used in Croatian (ENGRI corpus), focusing on those that have largely retained their original orthographic and phonetic features.

3. Methodology

3.1. Research goals and questions

As stated, our research aims to assess the potential of AI in addressing differences between specific senses in which English words can be used in Croatian, and to answer the following research questions:

1. Is there a difference in AI responses when the definition of the same terms is required in Croatian and English?
2. Does AI recognise differences in the use of the same expression in Croatian and English?[3]
3. Does the frequency of a particular English loanword in Croatian affect the AI's ability to evaluate their use adequately in Croatian and provide accurate feedback?

3.2. Method

In order to address the aforementioned research questions, we designed 6 different prompts, asking the free version of ChatGPT to provide definitions of the targeted lexical items (see Section 2.3. for the sample description). The prompts were designed

---

our study, we use the term »English loanwords« and, considering that this study does not focus on terminological issues or the adaptation degree of targeted words, we hereafter refer to the observed English expressions in our research as English loanwords.

[2] Note that in our prompts, we use both the terms »*meaning*« and »*sense*«. While these two can sometimes be used interchangeably, it is important to keep in mind the distinction between them when interpreting the results obtained in the study. Unlike fully-fledged dictionary definitions, *senses* represent more contextually nuanced interpretations, whereas the term *meaning* tends to refer to the underlying concepts that constitute said *definitions*, i.e., those that should be sufficient to distinguish a concept or item from other non-synonymous concepts or items (see Geeraerts, 2010, pp. 192–199).

[3] For instance, when dealing with English expressions in Croatian, their usage should generally be more limited in the number of meanings. The word *rock* in Croatian is mainly/only a reference to a type of music, whereas in English it has multiple meanings.

as 3 questions in Croatian and 3 in English, mirroring the same requirements stated in Croatian. In other words, the first 3 prompts were designed to extrapolate relevant information on the definition of an English word in terms of its use in the Croatian language, while the other 3 prompts—which were the English translations of the first 3 prompts—required of ChatGPT to provide feedback on the use of the same lexical items, but exclusively in English (see Table 1).

Table 1. List of prompts designed to elicit information on the target English loanword where »X« is replaced by a new loanword each time / **Tablica 1.** Popis upita osmišljenih za dobivanje informacija o ciljanoj posuđenici iz engleskog jezika, pri čemu se »X« svaki put zamjenjuje novom posuđenicom

Feedback goal	Prompt designed to address it
Meaning of »X« (formulated in Cro)	*Što znači riječ »X«?*
Senses of »X« in Cro	*Ako postoje, koja su moguća značenja/načini uporabe riječi »X« u hrvatskom jeziku?*
Word class of »X« in Cro	*Kao koja vrsta riječi se riječ »X« može koristiti u hrvatskom jeziku?*
Meaning of »X« (formulated in Eng)	*What does the word »X« mean?*
Senses of »X« in Eng	*What are the possible senses of the word »X« in English?*
Word class of »X« in Eng	*Which word classes can the word »X« represent when used in English?*

An important aspect of data collection entailed engaging in a new conversation each time one of the 6 prompts was given to ChatGPT, to avoid interference with previous feedback provided by the AI tool. Since ChatGPT considers the interaction established thus far (especially the previous 3 exchanges) in a particular conversation, starting a new conversation when the prompt was given ensured that the information retrieved was always devoid of the impact the previous interaction may have caused.

In order to capture the evolving nature of AI algorithms and compare the potential improvements in targeted interactions with the chatbot, we conducted the analysis of the responses on data retrieved during two different periods (January 2024 vs February and March 2025). More precisely, we reiterated the 3 prompts of the Croatian context and the corresponding treatment of loanwords to see whether the observed idiosyncrasies of ChatGPT's feedback changed or remained the same. These interactions were carried out with the free version of ChatGPT, or more precisely ChatGPT-3.5 and ChatGPT 4o, which we hereafter refer to as ChatGPT 2024 and ChatGPT 2025 to highlight the temporal gap between two data sampling points.[4]

[4] As stated on OpenAI's official webpage, GPT-4o offers significantly improved processing power and reasoning ability compared to GPT-3.5, matching GPT-4-level intelligence while being faster

3.3. Sample

For the initial selection of English loanwords based on their frequency of use in the contemporary Croatian language, we utilised the study by Bogunović (2023), who extracted 9,452 »unadapted« English loanwords from the ENGRI corpus. The ENGRI corpus, as described by Bogunović and Kučić (2022), consists of 2,395,735 texts collected from the 12 most popular Croatian news portals (Reuters Institute for the Study of Journalism 2021), with publications ranging from 2014 to 2020. This corpus provides the advantage of newer data, although it is smaller in size compared to hrWaC or CLASSLA.[5] The texts in ENGRI primarily derive from informal and journalistic styles, which are reflective of contemporary usage trends in the Croatian language.

The data was collected during January 2024 and February and March 2025, with the retrieval procedure covering 81 different English words in Croatian language. This resulted in a database of 486 units of information for the 2024 period and 243 for the 2025 period, which was then evaluated for accuracy, the number of meanings provided, and the soundness and plausibility of examples.

The loanwords were categorised based on their frequency of occurrence in the ENGRI corpus:

1. **Highly Frequent Terms (>1000 occurrences)** including *show*, *rock*, *break*, *mail*, *party*, *reality*, *press*, *gay*, *summit*, *post*, *face*, *brand*, *cool*, *style*, *blues*, *punk*, *tablet*, *craft*, *monitor*, *stage*, *fair*, *resort*, *cloud*, *hot*, *cast*, *light*, and *story*.
2. **Relatively Frequent Terms (400 – 1,000 occurrences)** including *rank*, *pride*, *joint*, *screen*, *teaser*, *take*, *like*, *shake*, *space*, *share*, *position*, *school*, *insider*, *follow*, *round*, *deep*, *site*, *dog*, *force*, *way*, *card*, *cross*, *touch*, *name*, *capital*, *trick*, and *slow*.
3. **Less Frequent Terms (100 – 400 occurrences)** including *next*, *Bluetooth*, *index*, *resident*, *net*, *bad*, *fish*, *case*, *trip*, *extra*, *block*, *fax*, *showman*, *win*, *marker*, *unplug*, *special*, *input*, *grind*, *plank*, *budget*, *escort*, *fun*, *contact*, *tutorial*, *target*, and *relax*.

---

and more efficient. Importantly, it achieves highest scores on model evaluation when it comes to text, can handle complex tasks, multi-step reasoning, and real-time interactions more effectively than GPT-3.5 (see OpenAI, 2024a & 2024b).

[5] The hrWaC corpus, built from the .hr domain crawls in 2011 and 2014, is a shuffled, annotated, and near-deduplicated Croatian web corpus with paragraph-level metadata (see Ljubešić & Klubička, 2016).

The CLASSLA-web.hr corpus, part of the broader South Slavic CLASSLA collection, is based on a 2021–2022 crawl and includes over 2.2 billion Croatian words enriched with linguistic and genre metadata (see Ljubešić, Rupnik & Kuzman, 2024).

The selection of these terms was informed by their frequency within the ENGRI corpus, providing a representative sample of loanwords as used in Croatian media.[6] This approach ensures that the analysis is grounded in actual language use, capturing a range of terms from highly frequent to less frequent occurrences. These loanwords were then subject to further analysis using ChatGPT, which provided feedback on their definitions, senses, and word classes. This methodology enabled a comprehensive examination of the integration and adaptation of English loanwords in the contemporary Croatian language.

4. Results

4.1. Individual response analysis

ChatGPT's feedback is most interesting when it comes to the prompt requiring word-class of a particular English expression in Croatian. Although there are a number of useful and incredibly detailed replies, at times, the tool seems to be prone to using certain English loanwords in the Croatian language in some idiosyncratic, if not completely incredulous contexts. Some rather unconvincing segments of ChatGPT's responses (during both 2024 and 2025 sampling points) are portrayed in Table 2 (see the text in bold), clearly showing that the AI overgeneralises the meaning that has established itself in English by extending it into Croatian regardless of whether such application has been observed in everyday speech.[7] Note that the plausible and legitimate examples that constitute the majority of ChatGPT's responses are removed from the table for the sake of brevity.

[6] Note that the frequencies were not calculated by us; i.e., the requisite information on frequency of occurrence, along with a hierarchically organized list of English loanwords in Croatian (sorted either by frequency or alphabetically), was already provided by the researchers who compiled the ENGRI corpus. The database is available at the following link: https://figshare.com/articles/dataset/The_database_of_English_words_in_Croatian_xlsx/20014364?file=35666141

[7] In order to corroborate the implausibility of the examples provided, we Google-searched for the exact equivalents in Croatian-language sites but found no matches when the suggested words are used in conjunction with surrounding linguistic material (e.g. English loanword + the immediate word on the right).

Table 2. Segments from ChatGPT's responses on *Word class of »X« in Cro* prompt
Tablica 2. Isječci iz odgovora ChatGPT-a o vrsti riječi »X« u upitu na hrvatskom jeziku

Prompt: Kao koja vrsta riječi se riječ »X« može koristiti u hrvatskom jeziku?	
English loanword »X«)	**ChatGPT 2024 response**
face	*U hrvatskom jeziku riječ »face« može se koristiti kao: (...) Glagol (u žargonu): U nekim slučajevima, riječ »face« može se koristiti kao glagol, opet u žargonskom kontekstu, u značenju »suočiti se« ili »gledati nekoga«. Na primjer: »**Moraš facati s time.**« (...)*
style	*U hrvatskom jeziku riječ »style« može se koristiti na različite načine, ovisno o kontekstu. Evo nekoliko primjera: (...) Glagol: U neformalnom govoru, riječ »style« može se koristiti kao glagol u značenju stilizirati ili uređivati. Primjer: »**Mogu li ti style kosu za večeras?**« (...)*
craft	*Riječ »craft« se u hrvatskom jeziku može koristiti kao imenica, pridjev ili glagol, ovisno o kontekstu: Imenica: Koristi se za označavanje ručno izrađenih predmeta ili proizvoda, obično povezanih s umjetnošću ili obrtom. Primjer: »**Volim raditi s različitim crafts.**«(...) Glagol: Kao glagol, »craft« se koristi za opisivanje procesa izrade nečega s pažnjom i vještinom. Primjer: »**Majstor je craftao drvene stolice.**« (...)*
hot	*Riječ »hot« može se koristiti u hrvatskom jeziku na nekoliko načina, uglavnom u neformalnom ili kolokvijalnom govoru:(...) Kao prilog: Rijetko, može se koristiti kao prilog, obično u frazama ili izražajima posuđenim iz engleskog jezika. Na primjer: »**On je radio hot and cold prema meni.**« (u značenju promjenjivog ponašanja) (...)*
cool	*Riječ »cool« u hrvatskom jeziku obično se koristi kao pridjev, iako se ponekad može koristiti i kao imenica. Evo kako se može koristiti u različitim kontekstima: (..) Imenica: Iako rjeđe, »cool« se može koristiti i kao imenica kako bi označio nekoga tko je opušten ili modern. Na primjer: »**On je pravi cool.**« - Ovdje se »cool« koristi kao imenica koja opisuje nekoga tko ima opušten stav ili stil. (...)*
	ChatGPT 2025 response
light	*Riječ »light« u hrvatskom jeziku može se koristiti na nekoliko načina, najčešće kao pridjev i imenica: (...) Imenica (ženskog roda, nepromjenjiva) – Označava svjetlost ili izvor svjetlosti. Primjer: **Uvečer je bio vrlo slab light iz ulične rasvjete.** (...) Glagol (rjeđe korišten, posuđenica) – Uglavnom označava upali (svjetlo) ili posvijetli. Primjer: **Možeš li light lampu?** (...)*
deep	*Riječ »deep« u hrvatskom jeziku može se koristiti kao: (...) Prilog (nepromjenjiv) – U rijetkim slučajevima može se koristiti kao prilog koji označava duboko ili u velikoj dubini. Primjer: **Zaronio je deep u more.** (Iako je ovo uobičajeno u engleskom jeziku, u hrvatskom jeziku češće će se koristiti sinonimi poput duboko.) Imenica (muškog roda, nepromjenjiv) – U specifičnom kontekstu može označavati dubinu, na primjer u kontekstu voda, bušotina ili sl. Primjer: **Mjesto ima veliku deep.** (Ovdje se koristi kao termin za dubinu.) (...)*
fax	*Riječ »fax« u hrvatskom jeziku može se koristiti kao: (...) Imenica (muškog roda, neformalno/sleng) – Skraćenica za fakultet. Primjer: **Upisao je fax u Zagrebu.** Drugi primjer: **Mora položiti još tri ispita da završi fax.** Riječ »fax« u hrvatskom jeziku najčešće se koristi kao imenica za telefaks, dok se u neformalnom govoru koristi i kao skraćenica za fakultet."*

Similarly to the prompts asking for the word-class that the target loanword can occupy when used as a loanword in the Croatian language, the prompt asking for different senses makes implausible suggestions about their use (Table 3).

Table 3. Segments from ChatGPT's responses on *Senses of »X« in Cro* prompt
Tablica 3. Isječci iz odgovora ChatGPT-a o značenjima riječi »X« u upitu na hrvatskom jeziku

Prompt: Ako postoje, koja su moguća značenja/načini uporabe riječi »X« u hrvatskom jeziku?	
English loanword (»X«)	**ChatGPT 2024 response**
rock	*Riječ »rock« može se koristiti na nekoliko načina u hrvatskom jeziku, iako se obično koristi u kontekstu glazbe. Evo nekoliko mogućih značenja i načina uporabe: (...) Kamen: Riječ »rock« može se koristiti kao sinonim za »kamen«. Na primjer: »**Pronašao sam lijepi rock na plaži.**« Pokret: U nekim slučajevima, »rock« se može koristiti kao glagol kako bi opisao kretanje naprijed-natrag ili s jedne strane na drugu. Na primjer: »**Njezino tijelo se lagano rockalo uz ritam glazbe.**« (...)*
gay	*U hrvatskom jeziku, riječ »gay« se najčešće koristi kao posuđenica iz engleskog jezika kako bi se opisale osobe koje su homoseksualne ili se identificiraju kao takve. Međutim, riječ »gay« također može imati neka druga značenja ili načine upotrebe: Gay (engl. veselo, vedro): Riječ »gay« se može koristiti u kontekstu koji označava veselost ili radost. Primjer: »**Proveli smo gay večer uz smijeh i zabavu.**« Gay (engl. neobično, šareno): Ponekad se riječ »gay« koristi kako bi se opisalo nešto što je neobično ili šareno. Primjer: »**Na njoj je nosila gay šarenu haljinu.**« Gay (engl. staromodan): Rijetko se koristi, ali ponekad se riječ »gay« može koristiti kako bi se opisala osoba ili stvar koja je staromodna ili zastarjela. Primjer: »**Njezin ukus je prilično gay.**« (...)*
light	*Riječ »light« se koristi u različitim kontekstima, a u hrvatskom jeziku može imati nekoliko različitih značenja i načina uporabe. Evo nekoliko mogućih značenja: (...) Osvjetljenje: U kontekstu rasvjete, »light« se može koristiti za opisivanje lampe ili svjetiljke. Primjer: »**Uključio je light da osvijetli sobu.**« (...) Informacija: »Light« se ponekad koristi kao sinonim za informaciju ili shvaćanje nečega. Primjer: »**On je brzo uhvatio light o tome kako riješiti problem.**« (...)*
	ChatGPT 2025 response
party	*Riječ »party« u hrvatskom jeziku dolazi iz engleskog i najčešće se koristi u sljedećim značenjima i kontekstima: (...) Politička stranka (rjeđe) – Na engleskom, »party« može značiti politička stranka (npr. Democratic Party), a u hrvatskom se taj oblik rijetko koristi jer postoji riječ »stranka«. Primjer (rjeđe): **On je član velike političke party.** (uobičajeno bi bilo reći »stranke«) Skupina ljudi s istim ciljem – U nekim specifičnim kontekstima može označavati tim ili grupu s određenom svrhom, primjerice u igrama. Primjer: **U igri smo formirali party za borbu protiv zmaja**. (...)*
stage	*Riječ »stage« u hrvatskom jeziku dolazi iz engleskog i koristi se u nekoliko značenja, uglavnom u umjetničkom, tehničkom i poslovnom kontekstu: (...) Faza, stadij razvoja – Može se koristiti u smislu određene etape razvoja, rasta ili napretka u nekom procesu. Primjer: **Projekt je u završnoj stage fazi.** Primjer: **U ovoj stage bolesti važno je što prije započeti liječenje.** (...) Internship (rjeđe, u francuskom značenju »stažiranje«) – U nekim poslovnim i akademskim krugovima može se koristiti kao sinonim za stručno osposobljavanje ili praksu, iako je u hrvatskom jeziku uobičajenija riječ »staž« ili »pripravništvo«. Primjer: **Odradio je šestomjesečni stage u poznatoj tvrtki.** (...)*

Table 4 demonstrates that ChatGPT sometimes makes the mistake of providing completely inconsistent (e.g. providing examples for a Croatian term that is the literal translation as well as for the English expression as part of the same response to the prompt; see the example for the loanword *fair*) or unrelated definitions for the term in question (not only can it be inconsistent in the feedback, but it can also make morphologically-rooted errors, providing examples for words completely lexically unrelated to the English loanword in question; see the example for the loanword *post*). In some instances, it makes part-of-speech errors (for example, see the segment from the response concerning the loanword *mail*), claiming that the loanword takes on the role of the verb whilst providing an example where it functions as a noun in the position of a direct object. The interesting aspect of the 2025 data is that, unlike before, it regularly states that the dubious examples it provides have a rare occurrence in Croatian, thus demonstrating a greater degree of data awareness (because of the content it exemplifies) and prompt understanding (demonstrating that it recognises which information it should focus on based on the request of its interlocutor).

Table 4. Segments from ChatGPT's responses indicating types of errors made
Tablica 4. Isječci iz odgovora ChatGPT-a koji upućuju na vrste pogrešaka

Types of errors made by ChatGPT in responses to prompts addressing the English loanwords *mail, fair,* and *post*	
Type of error	**ChatGPT 2024 response**
Wrong part-of-speech identification	*Riječ »mail« se u hrvatskom jeziku koristi kao imenica i kao glagol. (...)* ***Kao glagol,*** *»mail« označava slanje elektroničke pošte. Na primjer: »****Molim te, pošalji mi mail s detaljima****.« (...)*
Target language-related inconsistency in examples	*Riječ »fair« može se koristiti na više načina u hrvatskom jeziku, ovisno o kontekstu: Imenica: Koristi se za označavanje sajma ili izložbe na kojoj se izlažu i prodaju različiti proizvodi. Primjer: »****Posjetili smo književni sajam****.« Pridjev: Kao pridjev, »fair« se može koristiti za označavanje nečega što je pravedno ili pošteno. Primjer: »****Imali su fair dogovor o podjeli resursa****.« (...)*
Inconsistent and lexically unrelated examples	*Riječ »post« može imati nekoliko značenja i načina uporabe u hrvatskom jeziku: (...) Također, »post« može biti kratica za »poštanski broj«, kao u »****Koji je post tog mjesta?****« (...) Prezentsko značenje: Postojati znači biti prisutan ili biti u postojanju. Na primjer, »****Ideja postoji već stoljećima****.« Imenica »post« u smislu mjesta: Post također može biti naziv za neko geografsko mjesto, poput imena mjesta, poput »****Splitski post****«. Glagol »postati«: Postati označava prelazak iz jednog stanja u drugo. Na primjer, »****Ona je postala uspješna poduzetnica****.« (...)*

It is important to note that these are not the only implausible examples from ChatGPT's responses, but only a selection of those that belong to the highly frequent terms employed in Croatian everyday internet jargon, which makes the dubious examples all the more discouraging.

Table 5. Nondeterministic nature of ChatGPT's feedback to *Senses of »X« in Cro* prompt
Tablica 5. Nedeterministička priroda odgovora ChatGPT-a na upite o značenjima riječi »X« u hrvatskom jeziku

Variability in feedback to the prompt asking for possible senses of loanwords *rock, slow, special, screen* in the Croatian language context	
Response type	**ChatGPT 2025 response**
Providing the possible uses of the English loanword (L1 form of the word) in the Croatian language context; i.e. as intended by the prompt	*Riječ »rock« u hrvatskom jeziku najčešće se koristi kao posuđenica iz engleskog, a može imati nekoliko značenja i načina uporabe: 1. Glazbeni žanr – Najčešće označava stil glazbe, odnosno rock ('rok') glazbu. Primjer:* ***Volim slušati rock iz 70-ih godina.*** *2. Stil odijevanja i kulture – Može se odnositi na određeni stil povezan s rock-glazbom, uključujući odjeću, način života itd. Primjer:* ***Njegov rock izgled uključuje kožnu jaknu i crne čizme.*** *3. U sklopu naziva (brendovi, klubovi, događanja, pjesme) – Često se koristi u imenima barova, festivala, bendova itd. Primjer:* ***Idemo večeras u "Rock klub"****. (...)*
Providing English examples of the relevant context for the target English loanword with corresponding Croatian translations	*Riječ »slow« iz engleskog jezika rijetko se koristi u hrvatskom jeziku (...) u nekim specifičnim situacijama može se susresti u neformalnoj upotrebi. 1. Kao »spor« (brzina, tempo, reakcija) »Slow« se može koristiti za opisivanje nečega što se odvija sporo ili polako. Primjeri:* ***He is a slow runner. (On je spor trkač.)*** *The internet is very slow today. (Internet je danas jako spor.) This computer is too slow. (Ovo računalo je presporo.) 2. Kao »usporen« ili »trom« (mentalna ili fizička brzina, reakcija) »Slow« može značiti da netko sporo razmišlja, reagira ili je fizički usporen. Primjeri:* ***Sorry, I'm a bit slow in the mornings. (Oprosti, ujutro sam malo usporen.)*** *(...)*
Providing solely the Croatian translation of the target English loanword in different Croatian language use contexts	*Riječ »special« u hrvatskom jeziku koristi se kao posuđenica iz engleskog jezika i ima nekoliko mogućih značenja (...) najčešće značenje odnosi se na nešto što je iznimno, jedinstveno ili različito od drugih (...) Na primjer: »****Danas imamo poseban dan, jer slavimo obiteljski događaj.****« »Ovaj je restoran poznat po svojoj posebnoj ponudi jela.«* ***Specijalan*** *– »Special« može označavati specifičan ili poseban u smislu da je prilagođen za određeni cilj, situaciju ili osobu. Na primjer: »****Taj je proizvod specijalan jer je napravljen po narudžbi.****« »Za vas smo pripremili specijalnu ponudu.« (...)*
Alternating between the possible uses of the L1 form of the loanword in the Croatian language and the literal translation of the same loanword in the Croatian language context	*Riječ »screen« u hrvatskom jeziku koristi se u nekoliko značenja i konteksta, a najčešća su sljedeća:* ***Ekran*** *– Najčešće značenje odnosi se na zaslon ili površinu na kojoj se prikazuju slike, tekst ili video (...) Primjer:* ***Kupio sam novi ekran za računalo, sada je slika mnogo jasnija.*** *(...) Zaštita, mreža ili zaslon (u fizičkom kontekstu) –* ***»Screen«*** *se koristi i za označavanje fizičke zaštite ili mreže koja se koristi za blokiranje ili filtriranje nečega, kao što je zaštita od sunca, vjetra ili insekata. Primjer:* ***Na prozoru imamo screen protiv komaraca****. (...)*

It is also interesting to note that the stochastic nature of the ChatGPT's feedback design persisted in data retrieved during 2025; i.e. it is not only that we find variability in style and content across different prompts that are slightly different in

their design, but we find the same variability when the exact phrasing of the prompt is repeated, albeit with alternating loanwords. While it is obvious that the content differs from one loanword to another, the types of responses it provides differ in terms of data type provided (see Table 5), with possible loanword senses in the L2 context significantly differing concerning the word forms being exemplified.

4.2. The number of meanings per given prompt

As expected, the highest average number of meanings provided by ChatGPT is for prompts requesting the number of senses, and the lowest for those requesting the possible word classes (Table 6). The median and mode values are generally close to the mean, indicating a symmetric distribution of the AI's responses around the mean. When it comes to prompts formulated in Croatian asking for the meaning, the AI provides an average of 3.5 meanings for Croatian prompts and 3.9 meanings for English prompts. This suggests only a slightly higher number of meanings in English responses, thus indicating that the language of formulation, if left without further specification of the language in which the target word is to be used, is less consequential to the way in which ChatGPT responds.[8]

According to expectations when it comes to senses of the observed loanwords, the AI provides more senses in English compared to Croatian, indicating it tends to generate a richer set of senses when the prompts specify the language context in which the target word's definition is required. ChatGPT also seems to suggest fewer word classes overall, with a mean of 2.1 and 1.9 for Croatian (2024 and 2025 data respectively) and 2.9 for English. This is the lowest among the three categories for both languages, indicating that determining the word class yields fewer results compared to meanings and senses. Interestingly, the higher standard deviation (SD) values indicate variability in the English counterparts for meanings and senses (2.1 and 1.9, respectively), thus suggesting greater variability in the number of responses the AI provides for these prompts. Overall, the AI seems to provide a more extensive range of senses in English compared to Croatian, while the number of word classes remains relatively low and similar across both languages.

[8] This was also corroborated by qualitative observation of individual responses, where ChatGPT tends to provide a general definition of the term that is not restricted to any particular language; i.e. if the term is part of the English vocabulary, or if it is predominantly used in English, ChatGPT will opt for the definitions appropriate in the context of everyday English language use as well.

Table 6. General descriptive statistics for the number of meanings provided by ChatGPT per given prompt / **Tablica 6.** Opća deskriptivna statistika o broju značenja koje je ChatGPT dao po pojedinom upitu

	ChatGPT 2024	Mean	Median	Mode	Min	Max	SD
Prompts in Croatian	Meaning of »X« (formulated in Cro)	3.5	4.0	5.0	1.0	7.0	1.5
	Senses of »X« in Cro	4.9	5.0	5.0	2.0	9.0	1.2
	Word class of »X« in Cro	2.1	2.0	2.0	1.0	4.0	0.9
English counterparts	Meaning of »X« (formulated in Eng)	3.9	4.0	5.0	1.0	10.0	2.1
	Senses of »X« in Eng	6.3	6.0	6.0	3.0	14.0	1.9
	Word class of »X« in Eng	2.9	3.0	2.0	1.0	5.0	1.1
	ChatGPT 2025	Mean	Median	Mode	Min	Max	SD
Prompts in Croatian	Meaning of »X« (formulated in Cro)	5.8	6	4	1	12	2.2
	Senses of »X« in Cro	5.1	5	6	2	9	1.4
	Word class of »X« in Cro	1.9	2	2	1	3	0.6

Furthermore, the analysis of meanings provided by ChatGPT 2024 depending on the frequency of the English loanword in the ENGRI corpus (i.e. everyday Croatian internet jargon) provides some interesting insights when it comes to the average number of meanings in prompts addressing senses, especially in highly frequent terms shared between English and Croatian (see Figure 1). The most striking discrepancy is in the average number of senses provided for highly frequent terms, with prompts addressing the English language context yielding a much higher average (7.6) compared to those addressing the use of loanwords in the Croatian context (5.2). In accordance with previously stated results, the difference in the number of word classes between Croatian and English is minor (less than one on average, except in the category of relatively frequent terms), indicating that the AI's part-of-speech assessment for the given terms does not vary much across languages. Interestingly, the frequency of terms does seem to affect the number of meanings and senses more prominently in English than in Croatian, especially in highly frequent terms.

Interestingly, the AI's feedback to only one of the 6 prompts correlates with the frequency of the term in ENGRI corpus: the one asking for different senses of the word in English language, i.e. the more frequent it is in Croatian, the greater the number of senses in English the term seems to have ($r = 0.33$). Furthermore, the highest correlation appears to be between the definitions provided by the AI in res-

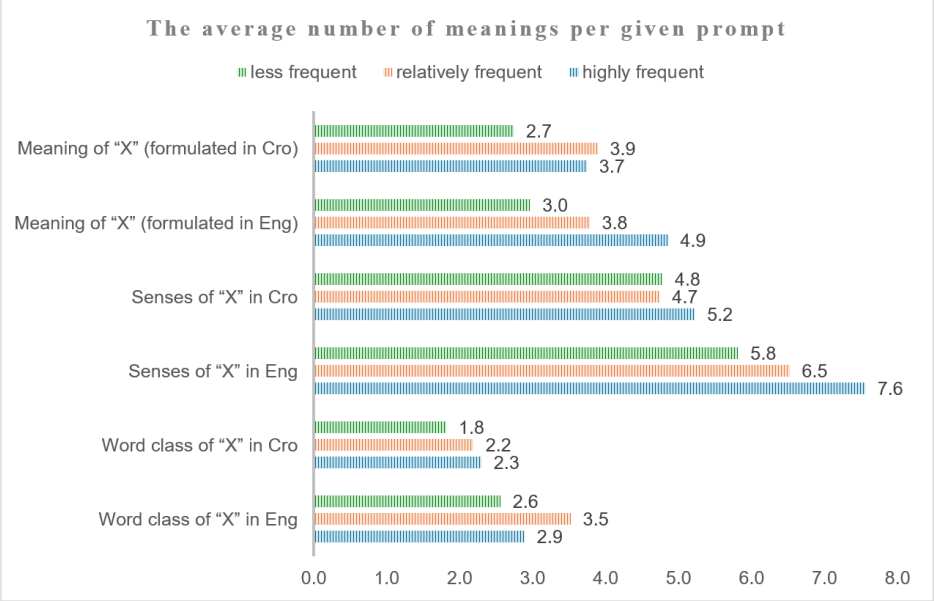

Figure 1. The average number of meanings provided by ChatGPT 2024 per given prompt depending on the frequency of the English loanword in the ENGRI corpus
Slika 1. Prosječan broj značenja koje je ChatGPT 2024 dao po pojedinom upitu, ovisno o učestalosti engleske posuđenice u korpusu ENGRI

ponse to prompts formulated in both English and Croatian, even when the prompts themselves do not specify the language context (r = 0.66). This indicates that the AI tends to generate consistent definitions across both languages (Table 7).

Additionally, there is a moderate to relatively high correlation between the AI's feedback on prompts asking for senses in both Croatian and English (r = 0.48). This suggests that the more senses a loanword has in English (according to ChatGPT), the more senses it also exhibits in Croatian (according to ChatGPT). This consistency implies that the AI's understanding of the breadth of meanings and senses of loanwords is similarly comprehensive across both languages, but also agrees with the suggestion concerning the overgeneralisation-type mistakes where the meaning that has established itself in English is extended into Croatian regardless of whether it is appropriate.

Table 7. Correlation matrix: prompts and frequency
Tablica 7. Korelacijska matrica: upiti i učestalost

	Frequency in the corpus	Meaning of "X" (formulated in Cro)	Senses of "X" in Cro	Word class of "X" in Cro	Meaning of "X" (formulated in Eng)	Senses of "X" in Eng	Word class of "X" in Eng
Frequency in the corpus	1.00	0.04	0.20	0.15	0.07	0.33**	0.00
Meaning of "X" (formulated in Cro)		1.00	0.39**	0.38**	0.66**	0.46**	0.42**
Senses of "X" in Cro			1.00	0.10	0.37**	0.48**	0.20
Word class of "X" in Cro			0.10	1.00	0.31**	0.15	0.36**
Meaning of "X" (formulated in Eng)					1.00	0.53**	0.33**
Senses of "X" in Eng						1.00	0.16
Word class of "X" in Eng							1.00

**$p<0.01$

Initially, ANOVA analysis has revealed no significant difference between the average number of meanings provided by ChatGPT in 2024 when it comes to responses to the prompts asking for the possible senses of targeted English loanwords in the Croatian language according to their frequency in the corpus (Figure 2). Although one might expect that with an increase in the frequency of a term in the Croatian language (as determined by the ENGRI corpus; Bogunović 2023), the degree of different meanings of loanwords would also increase (e.g. the more a specific term is used, the chances are that the number of senses attached to the situations in which it may be applied would increase), the situation is quite heterogeneous. There is no clear correlation between the frequency of loanwords in the Croatian language corpus and the number of meanings they can have in Croatian, based on ChatGPT's responses. This heterogeneity aligns with the findings discussed earlier, where the AI demonstrated variability in the number of meanings and senses it provides based on prompts formulated in Croatian and English. Specifically, while the AI tends to generate a richer set of meanings and senses in English, this consistency does not necessarily translate to the Croatian language in a straightforward manner. Initially, it seems that the relatively high correlation between the AI's feedback on senses in both languages suggests that the AI's comprehension is broad but not directly influenced by the frequency of terms in Croatian.

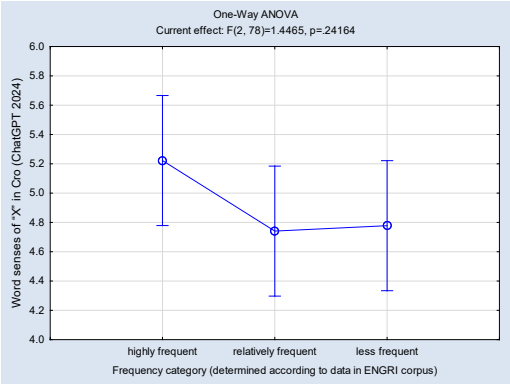

Figure 2. Difference between the average number of meanings provided by ChatGPT 2024 in response to the prompt asking for the possible senses of targeted English loanwords in the Croatian language according to their frequency in the corpora
Slika 2. Razlika u prosječnom broju značenja koje je ChatGPT 2024 dao kao odgovor na upit o mogućim značenjima ciljanih posuđenica, prema njihovoj učestalosti u korpusima

On the other hand, the ANOVA analysis conducted on data retrieved from answers provided by ChatGPT in 2025 in relation to the same prompt revealed a significant difference between the number of senses (Figure 3). More specifically, the highly frequent category had significantly less senses discerned by ChatGPT in comparison to the two others (approx. 1.5 less on average). This indicates several potential explanatory threads, all of which might be true at the same time: (1) the AI model of the chatbot has changed in the period between the two data retrieval points, (2) the nondeterministic nature inherent to the model has been proven not only across and within same prompt feedback, but also across the evolution-time span of the model, (3) there is a language-related reason why the three categories differ in the number of senses their belonging loanwords manifest, which the newer model of ChatGPT has managed to differentiate. As already stated, the intuitive supposition might be that the increase in the frequency of a term in the Croatian language would result in a greater number of different senses in which the loanword can be used, and yet the data retrieved from the 2025 model shows the opposite situation. One of the reasons why the number of senses could be fewer in the highly frequent category is that ChatGPT has been able to accurately narrow down these senses precisely due to sufficient data on their use, as opposed to the less frequent categories where it failed to do so, and thus resorted to overgeneralisation and application of L1 context senses (senses in English) on L2 ones (senses in Croatian). This tendency has been observed and already discussed in the qualitative analysis data (see previous Section), and it may suggest that the ChatGPT's flawed overgeneralised answers are often motivated by the fact that the chatbot »feels obliged« to provide some feedback, whether accurate or not. Finally, this would suggest that the model of the chatbot has evolved in the recent years but still falls short when it comes to less attested expressions in recorded language use.

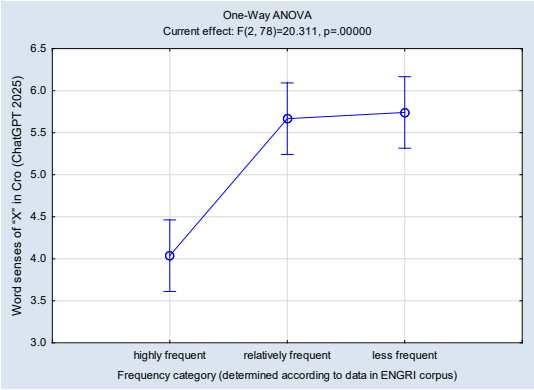

Figure 3. Difference between the average number of meanings provided by ChatGPT 2025 in response to the prompt asking for the possible senses of targeted loanwords in Croatian language according to their frequency in the corpora

Slika 3. Razlika u prosječnom broju značenja koje je ChatGPT 2025 dao kao odgovor na upit o mogućim značenjima ciljanih posuđenica u hrvatskom jeziku, prema njihovoj učestalosti u korpusima

The data on whether the frequency of the term plays a role in the ChatGPT's evaluation remains inconclusive when comparing data from 2024 and 2025. Initially, the results showed that while frequency might play a role in the AI's responses, it does not significantly impact the variety of meanings for Croatian use of English loanwords. However, upon further analysis in 2025, some difference has been ascertained with respect to frequency, suggesting that improvements in the algorithm resulted in a more constrained and precise feedback when it comes to highly frequent terms as opposed to those less frequent (corroborated by the qualitative analysis of its responses, which indicated more flaws with the decline of frequency). Finally, this indicates a complex interaction between language prompts and AI interpretations, which may be significantly impacted by AI's overextension of the loanword L1 senses into L2 language context.

5. Final discussion

Our findings remain consistent with the claims that, at least at this stage, there still needs to exist a degree of expert human oversight when it comes to AI-provided feedback. In fact, there is already a lot of data suggesting that, when going beyond the simple production of definitions and engaging in phrase or sentence-level related operations, the AI tends to underperform on certain occasions, often in terms of having to heavily rely on human oversight due to significant tendencies to invent facts, overgeneralise, or misrepresent them (cf. McKean & Fitzgerald, 2024). In this context, one of the immediate observations that become obvious in the qualitative

analysis of ChatGPT's responses is the stochastic nature of the feedback to prompts. This confirms previous research (De Schryver 2023), with our results showing answers that vary with respect to the type of information provided in them (sometimes referring to word types, and sometimes exclusively to the number of meanings without mentioning word types), even though the question was asked identically (this is especially pronounced with less detailed questions). Finally, while it may appear that accurate feedback is inherently dependent on understanding the prompt, our data illustrates that responses fluctuate—not due to changes in the prompt itself, which remains constant, but rather as a result of the specific loanword being analysed. This variability underscores the system's nondeterministic nature, which extends beyond mere structural and stylistic elements and directly influences the content of its responses.

Again, as evident from a number of different studies, there seem to exist a number of advantages and disadvantages of working with chatbots like ChatGPT. On the positive side, they tend to note that such tools can enhance productivity, lower lexicographic costs in terms of both time and money, and facilitate access to data that may otherwise be difficult to obtain, but well-documented disadvantages pertaining to the prevalence of hallucinations in AI-generated responses are just as prevalent at this stage of research (Fuertes-Olivera, 2024). The qualitative analysis of the data retrieved in our research reaffirms the claim that a noticeable amount of responses appear to be hallucinations, with examples of English loanwords provided by the ChatGPT sounding completely implausible in the Croatian language context. There is also more observable difference in data retrieved in 2025 between more frequent and less frequent loanwords, as evidenced by a greater number of flawed responses with the decline in frequency (cf. Merx et al., 2024), but this is not to be interpreted in favour of ChatGPT data from 2024; instead, in data retrieved from 2024, we have found an equal distribution of flawed answers concerning the frequency of loanwords. Again, note that we have corroborated our suspicions regarding flawed responses in view of loanword senses and definitions by consulting various phrase combinations in exact or related context formulations and found no uses resembling the ones provided by ChatGPT via Google search. This has been confirmed both in the 2024 and 2025 versions of ChatGPT, which has the more advanced GPT4o model integrated. Interestingly, we observe flawed answers both in prompts requiring the outlining of the possible senses of the loanword in L2 context and the one requiring the outlining of possible part-of-speech categories of said loanword (e.g. wrong part-of-speech identification, target language-related inconsistency in examples, inconsistent and lexically unrelated examples, etc.). On the other hand, when restricted to solely defining the term in its L1 context, there are little to no issues worth mentioning.

To be fair, in our paper, we mostly focus on the examples that, according to our standards, are either flawed or misreflect the crucial information required by the related prompts. However, it is important to emphasise that ChatGPT was mostly rather good at providing definitions, especially when they concerned the use of loanwords in their source language (which, in that case, constitute just »words«), and would almost always provide the expected or dominant interpretation of the loanword in context (cf. Georgiou, 2025). Additionally, when the chatbot is pressed for further information in cases where the answer seems flawed, it quickly tends to mend the answer in terms of factual or use-verifiable accuracy. In this context, Trap-Jensen (2025) has claimed that the results of lexicographers' experiments, due to individuals' varying attitudes toward the technology involved, may have been inconsistent, arguing that those who are enthusiastic about new technology may be impressed when a chatbot achieves 75% accuracy, while sceptics might focus on the remaining 25% of errors. At this stage of AI development, the accurate approach seems to be that probabilistic models produce outputs that are neither entirely correct nor entirely incorrect, but instead fall somewhere in between. For example, there is the issue of English bias, both linguistically and culturally, with English holding a significant advantage that makes a direct comparison impossible (see Trap-Jensen, 2025). Some researchers now propose building monolingual LLMs, including those dealing with Croatian-specific tasks; their argument is that multilingual LLMs may not perform optimally across all languages, especially lower-resourced ones, due to imbalanced training data that favours high-resource languages, and that without standardized evaluation tools for assessing multilingual performance, it remains unclear how other languages in a multilingual LLM affect its capabilities in Croatian (Štefanec et al., 2024; Thakkar et al., 2024). In the context of our research, it is interesting to observe the fact that the definition of the loanword is primarily addressed through the lens of English, which is the source language in this case, but the language of the prompt needs not be; i.e. the language by which the prompt is formulated plays little to no role in defining a specific term (in this case, an English loanword). It is not enough to formulate a prompt in the language in the context in which the definition is required, but it seems necessary to further specify the language where the term is used. This is further evidenced by the fact that, regardless of whether the prompt inquiry concerning the definition of targeted expression is designed in English or Croatian, there seems to be little difference in the average number of meanings in ChatGPT's feedback. More precisely, the only evident difference in the number of meanings provided by ChatGPT exists in the category of highly frequent English loanwords in Croatian, where one additional meaning is provided by ChatGPT on average when the question is formulated in English.

While it may seem that the main conclusion of our study regarding AI application is that sufficient training data for their respective languages needs to be im-

proved because of abundance, the results address a problem slightly more complex than the one seemingly at hand. Specifically, understanding prompt intention is sometimes a key factor in shaping the chatbot's response. This is as much an extra-linguistic issue as it is a linguistic one, and it will likely only be fully resolved with the integration of more advanced semantic networks and contextual learning mechanisms, i.e. with the development of artificial general intelligence (AGI). The types of tasks designed in our research and the data retrieved are in line with the claims that ChatGPT does not truly »understand« (at least not yet) and that it merely reproduces or mirrors language production and comprehension (cf. Harnad, 2025). For example, as predicted, the number of senses (of use) of the examined loanwords has been generally somewhat lower in the target language (Croatian) than in the source language (English). In-depth analysis has also revealed an artificial number of flawed responses, suggesting that the difference should have been higher than observed. Additionally, although one might expect that, with the increase in frequency, the degree of different meanings of English loanwords in Croatian would also increase, the situation is quite heterogeneous, as there is no clear correlation between the frequency of loanwords in the Croatian language corpus and the number of meanings they can have in Croatian when considering the responses of artificial intelligence retrieved in 2024. However, data from the 2025 model indicates the opposite trend. A possible explanation for the reduced number of senses in the high-frequency category is that ChatGPT has been able to accurately constrain these meanings due to the availability of sufficient usage data. In contrast, for less frequent terms, the model appears to struggle with this distinction, leading to overgeneralisation and the transfer of L1 contextual meanings onto L2 contexts, which has already been observed in the qualitative analysis. The chatbot model has improved over the year, but it still demonstrates limitations when handling less commonly attested expressions in language use. It would be interesting to observe whether these trends would occur on another set of English loanwords in Croatian that is not necessarily based on domain-specific corpus (i.e., our was determined according to the ENGRI corpus; Bogunović 2023), since the compiling methodology can play a role in the frequencies, skewing the results and affecting data interpretation.

Future research should examine more thoroughly the idiosyncratic examples provided by AI and compare them with recorded corpus or other text data, i.e. to corroborate on a greater scale whether each of the suggested uses truly constitute »mistakes«. Furthermore, it would be beneficial to cross-compare responses from different AI tools (e.g. Gemini, ChatGPT, DeepSeek, Grok, Perplexity, etc.) and evaluate their accuracy and style in view of understanding and ability to define loanwords in L2 context, thus further providing an insight into the difference in the quality of different AI models for language-related tasks.

6. Conclusion

While it should be noted that the language of the response always matches the language of the question, regardless of the specific nuances these tools provide in their answers, the non-deterministic nature of the feedback is most apparent when responding to questions that are not sufficiently well-formulated or explicitly defined. For instance, when the prompt requires the word's meaning and is formulated in Croatian, but does not provide additional information on the language context in which the word is used, the feedback can vary significantly on a case-by-case basis. Given that the inquiry pertains to the definition of an English loanword, the responses vary depending on whether the word is defined primarily through the lens of its meaning in the source language or whether artificial intelligence accounts for the fact that the question is asked in Croatian—presumably because the user seeks information about the word's meaning in the target language (i.e. Croatian rather than English). Additionally, in some instances, the response focuses on the grammatical category that the lexical entry can assume in a given context. In others, it provides different meanings without referencing the word's grammatical classification.

Although artificial intelligence is a practical tool in lexicography, a more detailed comparison reveals that it can serve as its counterpoint. The fundamental purpose of a dictionary is to provide a relatively stable description of the language (even though the language itself is inherently unstable), whereas the operational model of artificial intelligence follows a completely different approach. While AI does provide accurate word meanings, the number of meanings it generates often varies; more often than not, a single meaning is divided into multiple, subtly distinct senses. This variability stems from the fact that artificial intelligence is designed to generate relatively new and original text, but this artificial fluidity seems rather forced—especially when it lacks an explainable method as to why the answer would need to vary. In other words, the algorithm's tendency to always provide an answer, regardless of whether that answer is grounded in truth or language use, affects the feedback to the point where it often becomes significantly flawed—which is certainly more evident in language contact lexicography than in monolingual dictionary compiling instances. Some of the findings in our data can be summarised as follows:

1. The results demonstrate the AI's proficiency in providing accurate definitions and distinguishing between senses in which the examined expressions are used, although this is not consistently demonstrated when it comes to possible loanword senses in the L2 context. The less attested the loanword is in everyday use, the greater the chances the AI will provide a flawed response in some respect (often in view of L1 sense overgeneralisation onto L2).

2. When it comes to the inquiry concerning the definition of targeted English expression and the difference depending on whether the prompt is designed in English or Croatian, there seems to be little difference in the average number of meanings in ChatGPT's feedback. We interpret this result as a further confirmation that the definition of the loanword is primarily addressed from the English standpoint even when the prompt is formulated in L2.

3. According to predictions, the number of senses (of use) of the examined loanwords is generally somewhat lower in the target language (Croatian) than in the source language (English). However, qualitative analysis reveals that artificial intelligence often makes errors and provides flawed responses with examples not attested in the available repositories of recorded language performance, suggesting ways of use that do not align with their everyday use in Croatian, both semantically and morphologically. This is true for both the data retrieved during early 2024 and early 2025, although with qualitatively observable improvements in the 2025 model, which does address the debatable nature of the dubious examples it provides at times.

4. Although one might expect that with the increase in language use frequency, the degree of different meanings of loanwords in Croatian would also increase, the situation is quite heterogeneous, as there is no clear correlation between the frequency of loanwords in the Croatian language corpus and the number of meanings they can have in Croatian when considering the responses of ChatGPT from 2024. On the other hand, some difference has been observed in the responses of ChatGPT from 2025, with counter-intuitive findings suggesting a greater number of senses in the highly frequent category. Again, we interpret this as a by-product of ChatGPT's tendency to prioritise generating a response regardless of accuracy, with overgeneralisation and the transfer of L1 contextual meanings onto L2 contexts being responsible for the (potentially) unjustified increase in the number of overall senses for less frequent categories of targeted expressions.

In the realm of lexicography, the opaque nature of AI training texts poses challenges, colloquially described as a »black box« (Steurs et al. 2020: p. 12), but despite occasional inexplicable outputs, it appears inevitable that prompt engineering will play a crucial role in the evolving landscape of lexicographers' work as the technology progresses and becomes more transparent. Nevertheless, the main conclusion arising from our lexicographic experiments indirectly reflects that language, in a

manner consistent with the cognitivist tradition, is not an entity isolated from the contextual elements of the environment in which it manifests itself. Extralinguistic factors are integral even within lexicography, a discipline primarily focused on providing a factual, concise, yet detailed view of the lexicon.

REFERENCES

Aguera-Arcas, Blaise. 2022. »Do Large Language Models Understand Us?«. *Daedalus,* 151 (2): 183–197.

Bogunović, Irena, Kučić, Mario. 2022. *The Database of English Words in Croatian.xlsx*. figshare. Dataset. https://doi.org/10.6084/m9.figshare.20014364.v1.

Bogunović, Irena. 2023. »A Corpus-Based Approach to English Loanwords: Introducing the Database of English Loanwords in Croatian«. *Fluminensia,* 35 (2): 437–60.

Carlini, Nicholas et al. 2021. »Extracting Training Data from Large Language Models«. In: *30th USENIX Security Symposium (USENIX Security 21),* 2633–2650.

Chow, Siew Yeng, Shin, Chang-Uk, Bond, Francis. 2024. »This Word Mean What: Constructing a Singlish Dictionary with ChatGPT«. In: *Proceedings of the 2nd Workshop on Resources and Technologies for Indigenous, Endangered and Lesser-Resourced Languages in Eurasia (EURALI) @ LREC-COLING 2024*, 41–50. Torino, Italia.

De Schryver, Gilles-Maurice. 2023. »Generative AI and Lexicography: The Current State of the Art Using ChatGPT«. *International Journal of Lexicography,* 36 (4): 355–387.

Drljača Margić, Branka. 2009. »Latentno posuđivanje u hrvatskome i drugim jezicima – posljedice i otpori«. *Rasprave instituta za Hrvatski jezik i jezikoslovlje,* 35: 19–31.

Filipović, Rudolf. 1990. *Anglicizmi u hrvatskom ili srpskom jeziku: Porijeklo, razvoj, značenje.* Vol. 70. Zagreb: Jugoslavenska akademija znanosti i umjetnosti.

Floridi, Luciano. 2023. »AI as Agency without Intelligence: On ChatGPT, Large Language Models, and Other Generative Models«. *Philosophy & Technology,* 36 (1): 15.

Fuertes-Olivera, Pedro A. 2024. »Making Lexicography Sustainable: Using ChatGPT and Reusing Data for Lexicographic Purposes«. *Lexikos,* 34 (1): 123–140.

Geeraerts, Dirk. 2010. *Theories of Lexical Semantics*. Oxford: Oxford University Press.

Georgiou, Georgios P. 2025. »Can AI Mimic the Human Ability to Define Neologisms?«. *arXiv preprint* arXiv:2502.14900.

Harnad, Stevan. 2025. »Language Writ Large: LLMs, ChatGPT, Meaning, and Understanding«. *Frontiers in Artificial Intelligence,* 7.

Hong, Wilson Cheong Hin. 2023. »The Impact of ChatGPT on Foreign Language Teaching and Learning: Opportunities in Education and Research«. *Journal of Educational Technology and Innovation,* 5 (1).

Jiang, Zhengbao, Xu, Frank F., Araki, Jun, Neubig, Graham. 2020. »How Can We Know What Language Models Know?«. *Transactions of the Association for Computational Linguistics,* 8: 423–438.

Kalaš, Filip. 2025. *Bridging Tradition and Innovation: Analysing Language Data with ChatGPT-4 in Corpus Linguistics.* Available at SSRN 5126316.

Kasneci, Enkelejda et al. 2023. »ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education«. *Learning and Individual Differences,* 103: 102274.

King, Michael R., ChatGPT. 2023. »A Conversation on Artificial Intelligence, Chatbots, and Plagiarism in Higher Education«. *Cellular and Molecular Bioengineering,* 1–2.

Klosa-Kückelhaus, Annette, Tiberius, Carole. 2025. »The Lexicographic Process Revisited«. *International Journal of Lexicography,* 38 (1): 1–12.

Lew, Robert. 2023. »ChatGPT as a COBUILD Lexicographer«. *Humanities and Social Sciences Communications,* 10: 704.

Lew, Robert, Ptasznik, Bartosz, Wolfer, Sascha. 2024. »The Effectiveness of ChatGPT as a Lexical Tool for English, Compared with a Bilingual Dictionary and a Monolingual Learner's Dictionary«. *Humanities and Social Sciences Communications,* 11 (1): 1–10.

Li, Qian, Tarp, Sven. 2024. »Using Generative AI to Provide High-Quality Lexicographic Assistance to Chinese Learners of English«. *Lexikos,* 34: 397–418.

Ljubešić, Nikola, Klubička, Filip. 2016. *Croatian Web Corpus hrWaC 2.1.* Slovenian Language Resource Repository CLARIN.SI. http://hdl.handle.net/11356/1064.

Ljubešić, Nikola, Rupnik, Peter, Kuzman, Tomaž. 2024. *Croatian Web Corpus CLASSLA-web.hr 1.0.* Slovenian Language Resource Repository CLARIN.SI. http://hdl.handle.net/11356/1929.

Martín, Antonio S. 2024. »What Generative Artificial Intelligence Means for Terminological Definitions«. *arXiv preprint* arXiv:2402.16139.

McKean, Erin, Fitzgerald, Will. 2024. »The ROI of AI in Lexicography«. *Lexicography,* 11 (1): 7–27.

Merx, Rebecca, Vylomova, Ekaterina, Kurniawan, Kemal. 2024. »Generating Bilingual Example Sentences with Large Language Models as Lexicography Assistants«. *arXiv preprint* arXiv:2410.03182.

Min, Bonan et al. 2023. »Recent Advances in Natural Language Processing via Large Pre-Trained Language Models: A Survey«. *ACM Computing Surveys,* 56 (2): 1–40.

Opačić, Nives. 2012. *Hrvatski u zagradama – globalizacijske jezične stranputice.* Hrvatska sveučilišna naklada. Zagreb.

OpenAI. 2024a. »Hello GPT-4o.« *OpenAI.* May 13, 2024. https://openai.com/index/hello-gpt-4o.

OpenAI. 2024b. »GPT-4o and More Tools to ChatGPT Free.« *OpenAI.* May 13, 2024. https://openai.com/index/gpt-4o-and-more-tools-to-chatgpt-free.

Patekar, Jakob. 2019. »Prihvatljivost prevedenica kao zamjena za anglizme«. *Fluminensia,* 31 (2): 143–179.

Ptasznik, Bartosz, Wolfer, Sascha, Lew, Robert. 2024. »A Learners' Dictionary versus ChatGPT in Receptive and Productive Lexical Tasks«. *International Journal of Lexicography,* 37 (3): 322–336.

Radford, Alec et al. 2018. »Improving Language Understanding by Generative Pre-Training«. *OpenAI Blog.*

Rees, Geraint Paul, Lew, Robert. 2023. »The Effectiveness of OpenAI GPT-Generated Definitions Versus Definitions from an English Learners' Dictionary in a Lexically Oriented Reading Task«. *International Journal of Lexicography,* ecad030.

Rudolph, Jürgen, Tan, Samson, Tan, Shannon. 2023. »ChatGPT: Bullshit Spewer or the End of Traditional Assessments in Higher Education?«. *Journal of Applied Learning and Teaching,* 6 (1): 342–363.

Rundell, Michael. 2024. »Automating the Creation of Dictionaries: Are We Nearly There?«. *Humanising Language Teaching,* 26 (1).

Steurs, Frieda, Schoonheim, Tanneke, Heylen, Kris, Vandeghinste, Vincent. 2020. *The Future of Academic Lexicography—A White Paper.* Instituut voor de Nederlandse Taal.

Štefanec, Vanja, Farkaš, Daša, Thakkar, Gaurish, Tadić, Marko. 2024. »Building a Large Language Model for Croatian.«. *Proceedings of the Conference New Trends in Translation and Technology 2024.* 204-209.

Thakkar, Gaurish, Štefanec, Vanja, Farkaš, Daša, Tadić, Marko. 2024. »Building a Large Language Model for Moderately Resourced Language: A Case of Croatian.« In: *Proceedings of the 35th*

Central European Conference on Information and Intelligent Systems - CECIS 2024, 225–229. Varaždin: University of Zagreb Faculty of Organization and Informatics.

Trap-Jensen, Lars. 2025. *The Best of Two Worlds: Exploring the Synergy between Human Expertise and AI in Lexicography.* Society for Danish Language and Literature.

Xu, Frank F. et al. 2022. »A Systematic Evaluation of Large Language Models of Code«. In: *Proceedings of the 6th ACM SIGPLAN International Symposium on Machine Programming,* 1–10.

Yin, Yunting, Skiena, Steven. 2023. »Word Definitions from Large Language Models«. *arXiv preprint* arXiv:2311.06362.

LEKSIKOGRAFSKI POTENCIJAL UMJETNE INTELIGENCIJE NA PRIMJERU ENGLESKIH POSUĐENICA U HRVATSKOM JEZIKU

Katica Balenović
Sveučilište u Zadru
kbalenovic@unizd.hr
Jakov Proroković
Sveučilište u Zadru
jprorokov@unizd.hr

SAŽETAK: Pojava generativne umjetne inteligencije (UI) i velikih jezičnih modela (VJM) otvorila je nove mogućnosti u području leksikografije, osobito kada se radi o preciznom definiranju rječničkih natuknica uz istodobno smanjenje utrošenog vremena u usporedbi s tradicionalnijim metodama ili softverskim alatima. Kako bi se ispitali jezični kapaciteti umjetne inteligencije, ovo istraživanje nadilazi izradu jednojezičnih rječnika te istražuje mogućnosti modela ChatGPT u razlikovanju specifičnih značenja posuđenica u kontekstu drugog jezika (J2). Korpusni uzorak ciljanih engleskih posuđenica korišten je za procjenu sposobnosti modela ChatGPT u razlučivanju različitih značenja koja se pojavljuju u hrvatskom jezičnom kontekstu. Rezultati pokazuju da UI pokazuje značajnu razinu preciznosti u definiranju ciljanih riječi, ali i primjetne nedostatke kada odgovara na upite koji se specifično odnose na moguća značenja ili vrste riječi spomenutih riječi (posuđenica) u kontekstu drugog jezika (J2). Točnost modela opada pri obradi manje učestalih posuđenica, pri čemu se često pojavljuje tendencija prekomjernog generaliziranja, odnosno prenošenja značenja iz engleskog (J1) u hrvatski jezik (J2). Također, model nerijetko generira pogrešne primjere upotrebe, predlažući značenja koja nisu potvrđena u jezičnim korpusima. Dobiveni rezultati potvrđuju da model prvenstveno interpretira posuđenice iz perspektive engleskog jezika, neovisno o jeziku na kojem je upit postavljen. Usporedba odgovora UI iz ranog razdoblja 2024. i početka 2025. godine upućuje na poboljšanja u novijoj verziji modela, koja pokazuje preciznije razlučivanje dvosmislenih slučajeva. Međutim, i dalje su prisutne nekonzistentnosti, osobito u korelaciji između učestalosti upotrebe i broja značenja, što se tumači tendencijom modela da ponekad daje prednost generiranju odgovora nauštrb točnosti.

Ključne riječi: ChatGPT; leksikografija u kontekstu jezičnog kontakta; prekomjerna generalizacija; korpusni uzorak; posuđenice

Članci su dostupni pod licencijom Creative Commons: Imenovanje 4.0 međunarodna (https://creativecommons.org/licenses/by/4.0/). Sadržaj se smije umnožavati, distribuirati, priopćavati javnosti, prerađivati i koristiti u bilo koju svrhu, uz obavezno navođenje autorstva i izvora.